**DISTRIBUIÇÕES ELÍPTICAS: PROPRIEDADES,
INFERÊNCIA E APLICAÇÕES A MODELOS
DE REGRESSÃO**

Este exemplar corresponde à redação
da tese devidamente corrigida
e defendida por Reinaldo Boris Arellano Valle
e aprovada pela Comissão Julgadora

São Paulo, 08 de dezembro de 1994

Banca Examinadora:

- Prof. Dr. Heleno Bolfarine (orientador)
- Prof. Dr. Gauss Moutinho Cordeiro
- Prof. Dr. Ruben Klein
- Prof. Dr. Renato Assunção
- Prof. Dr. Josemar Rodrigues

Resumo

Neste trabalho são desenvolvidos diferentes resultados em torno da distribuição t multivariada. Em primeiro lugar, propõe-se várias maneiras de caracterizar esta distribuição dentro da classe geral das distribuições elípticas. Mostra-se também que esta lei pode ser caracterizada de forma especial no contexto daquelas leis que são mistura de escala da lei normal. Em segundo lugar, aplica-se a distribuição t ao estudo estatístico de modelos de regressão linear com e sem erros nas variáveis explicativas ou independentes. Duas especificações destes modelos são aqui consideradas, digamos, os modelos de regressão- t dependente e independente. Mostra-se, em particular, que os estimadores de máxima verossimilhança e os testes de hipóteses da razão de verossimilhança e escore do primeiro modelo são similares aos do modelo normal. No segundo modelo, no entanto, estes estimadores são robustos e os testes de hipóteses devem basear-se na teoria assintótica. Alguns testes assintóticos modificados por fatores tipo Bartlett são propostos neste caso. Finalmente, no contexto da regressão com erros nas variáveis, vários resultados são desenvolvidos para um modelo elíptico geral.

Abstract

In this thesis, we develop several different results concerning the multivariate t -distribution. In the first place, we investigate the characterization of the t -distribution within the general class of the elliptical distributions. It is shown that the t -distribution can be characterized in some special ways in the context of the distributions which are scale mixtures of the normal distribution. Secondly, the t -distribution is used in the analysis of linear regression models with and without the presence of measurement errors in the independent variables. Two different specifications of these models are considered, say, the dependent and independent regression models with and without measurement errors. Specifically, we show that the maximum likelihood estimators and the likelihood ratio and score statistics for testing hypotheses about the unknown parameters are similar to the corresponding ones under normality with respect to the dependent model. In the independent model, however, the estimators are robust and hypotheses tests should rely on some asymptotic results. Some asymptotic tests modified by Bartlett correction factors are also obtained. Finally, in the context of the measurement error regression models, several results are developed for the general case of the elliptical models.

Agradecimentos

Gostaria de dizer, em primeiro lugar, que me sinto honrado de ter sido orientado pelo Professor Heleno Bolfarine para desenvolver as idéias que originaram este trabalho. Estou muito grato ao Professor Bolfarine pela confiança em mim depositada e, também, pelo constante apoio e particular interesse e respeito com que sempre analisou o conjunto de resultados que foram dando forma a esta tese. Minha admiração por ele não se restringe a suas qualidades de pesquisador e orientador, mas, principalmente pelo magnífico ser humano e pela paixão com que desenvolve e divulga a estatística.

Agradeço também à minha co-orientadora, a Professora Silvia L. P. Ferrari, cujas idéias, sugestões e contribuições foram fundamentais para obter, conferir e avaliar vários resultados deste trabalho. Também por sua particular preocupação na redação desta tese e seu indispensável apoio computacional.

Quisera, também, manifestar aqui que me sinto muito orgulhoso de ter encontrado durante este período grandes amigos, a Pilar, o Juan, a Olga e a Nancy. Em Pilar sempre encontrei apoio, amparo e alegria em meus momentos difíceis desta etapa de minha vida, sendo ela o maior “pilar” que eu tive durante este período. Além disso, a contribuição da Pilar foi indispensável para o desenvolvimento do capítulo 3 desta tese. O apoio e amizade de Juan foram, sem dúvida, essenciais para minha adaptação no meu primeiro ano no Brasil. Em Olga, além de sua agradável e sincera amizade, sempre encontrei respostas e conselhos inteligentes às minhas dúvidas e preocupações. Com Nancy compartilhei alegrias e dificuldades durante este período. Obrigado amigos por sua constante preocupação pela solução de minhas dificuldades e por terem tornado minha vida mais alegre e divertida durante todo este tempo.

Agradeço, de forma muito especial, à Flora, à Mariana e à Lúcia, por seu apoio e

amizada e por ter me facilitado a moradia aqui no Brasil. Aproveito para manifestar meus agradecimentos ao Walter por sua paciência e notável eficiência na datilografia desta tese.

Obviamente que não posso deixar de destacar o constante apoio e compreensão de minha família durante esta longa separação. Em particular, agradeço aos meus irmãos Mario, Angela e Selvia por terem me substituído, tanto em minhas obrigações para com a família, quanto em meus compromissos mais pessoais. Agradeço também aos meus amigos de sempre, Patricio, Leonardo, Ivonne e Marta, pois, apesar da distância e do tempo, não esqueceram-se de mim.

Finalmente, agradeço ao apoio financeiro do CNPq e da Pontifícia Universidade Católica de Chile, que permitiram realizar meus estudos neste maravilhoso país. Também agradeço à Universidade de São Paulo por ter me acolhido como mais um de seus inumeráveis estudantes estrangeiros, incluindo muitos compatriotas.

*Aos meus sobrinhos Daniel e Juan Carlos,
pois neles encontrei a força e a inspiração
para completar meus estudos.
À minha amiga Marisa Yadlin, por ter me
motivado a seguir estes caminhos da es-
tatística.*

Índice

Capítulo 1: Introdução	1
1.1 Formulação do problema	1
1.2 Definição dos objetivos	4
1.3 Resumo de conteúdos	5
Capítulo 2: Distribuições esféricas e elípticas	9
2.1 Introdução	9
2.2 Notações, definições e exemplos	11
2.3 Representações estocásticas	19
2.4 Existência de densidade	25
2.5 Propriedades adicionais	28
2.5.1 Momentos e medidas de curtose	28
2.5.2 Transformações lineares e distribuições marginais	32
2.5.3 Distribuições condicionais	34
2.5.4 Formas quadráticas e outras funções relacionadas com distribuições esféricas	40
2.5.5 Ferramentas para se obter certos cumulantes de derivadas da função log-verossimilhança de um modelo elíptico	47
2.6 Algumas caracterizações da distribuição normal	51
Capítulo 3: Caracterizações da distribuição t multivariada	57
3.1 Introdução	57
3.2 Definição e propriedades da distribuição t multivariada	59
3.2.1 Definição	59

3.2.2 Propriedades	66
3.2.3 Algumas caracterizações da distribuição t dentro da classe elíptica	73
3.3 Caracterização da distribuição t na classe das distribuições normais	
compostas: uma interpretação preditivista	78
3.3.1 Preliminares	78
3.3.2 Caracterização da distribuição t na classe das distribuições normais	
compostas	81
3.3.3 Caracterização da distribuição t multivariada a partir de seqüências de	
variáveis aleatórias	84
3.3.4 Caracterização da matriz t variada a partir de seqüências de vetores	
aleatórios	88

Capítulo 4: Inferência no modelo de regressão linear usando a

distribuição t	96
4.1 Introdução	96
4.2 O modelo de regressão t com erros dependentes	101
4.2.1 Estimação de máxima verossimilhança	103
4.2.2 Testes da hipótese linear $H_0 : A\beta = \gamma_0$	106
4.3 O modelo de regressão t com erros independentes	111
4.3.1 Estimação de máxima verossimilhança	115
a) O algoritmo de scoring	117
b) O algoritmo EM	118
4.3.2 Robustez e outras propriedades do estimador de máxima verossimilhança	120
4.3.3 Parametrização ortogonal de $(\tau^2, \nu)^\top$	122
4.3.4 Testes assintóticos da hipótese linear $H_0 : A\beta = \gamma_0$	125

4.4 Testes assintóticos modificados por fatores tipo Bartlett no modelo de regressão t independente	129
4.4.1 Resultados teóricos gerais	129
4.4.2 Fatores de correção supondo τ^2 e ν conhecidos	134
4.4.3 Fatores de correção supondo τ^2 conhecido e ν desconhecido	141
4.4.4 Fatores de correção supondo τ^2 e ν desconhecido	145
4.4.5 Alguns estudos de simulação	151

Capítulo 5: Inferência no modelo de regressão linear simples com erros

nas variáveis usando a distribuição t	157
5.1 Introdução	157
5.1.1 Estimação de mínimos quadrados ordinários e generalizado de β	159
5.1.2 Inferência no modelo estrutural normal supondo razões de variâncias conhecidas	163
a) Matriz de informação e parametrização ortogonal	166
b) Estimação de máxima verossimilhança usando a parametrização ortogonal	172
c) Propriedades dos EMV	174
d) Testes da razão de verossimilhança e escore da hipótese $H_0 : \beta = \beta_0$...	179
5.2 Modelos estruturais elípticos	185
5.2.1 Introdução	185
5.2.2 Especificação de um modelo estrutural elíptico	186
5.2.3 Identificabilidade e estimação consistente dos parâmetros estruturais	190
5.2.4 Matriz de informação e parametrização ortogonal	194
5.3 Inferência no modelo estrutural- t dependente	198
5.3.1 Matriz de informação e EMV da parametrização ortogonal	199

a) Derivação da matriz de informação	199
b) Derivação dos EMV	200
5.3.2 Propriedades dos EMV	201
5.3.3 Testes da razão de verossimilhança e escore de $H_0 : \beta = \beta_0$	202
5.4 Inferência no modelo estrutural t independente	203
5.4.1 EMV e matriz de informação da parametrização ortogonal	206
a) Derivação dos EMV	206
b) Derivação da matriz de informação	208
5.4.2 Testes da razão de verossimilhança e escore modificado da hipótese	
$H_0 : \beta = \beta_0$	211
a) Teste da razão de verossimilhança modificado	214
b) Teste do escore modificado	220
c) Um estudo de simulação	223

Apêndice A

A.1 Cumulantes conjuntos de derivadas da função log-verossimilhança de um	
modelo t -elíptico p -variado	225
A.1.1 Alguns cumulantes para parâmetros de locação	230
A.1.2 Alguns cumulantes para parâmetros de escala	231
A.1.3 Alguns cumulantes mistos	232
A.2 Cumulantes para o modelo de regressão linear t independentes	234
A.2.1 Alguns cumulantes para os parâmetros de regressão	235
A.2.2 Alguns cumulantes para o parâmetro de escala	235
A.2.3 Alguns cumulantes mistos	235
A.2.4 Alguns cumulantes mistos que envolvem o parâmetro ν	235
A.3 Cumulantes para o modelo estrutural t independente	236

A.3.1 Alguns cumulantes para os parâmetros de locação	239
A.3.2 Alguns cumulantes para os parâmetros de escala	239
A.3.3 Alguns cumulantes mistos que envolvem o parâmetro β	240
Bibliografia	242

Capítulo 1

Introdução

1.1 Formulação do problema

As boas propriedades da distribuição normal (univariada e multivariada) juntamente com o fato de que a normalidade pode ser habitualmente justificada apelando ao Teorema Central do Limite, têm sido as fundamentações básicas da aplicação do modelo normal na análise estatística de dados contínuos. Assim, como é bem sabido, a maior parte da teoria de inferência estatística para variáveis reais contínuas está desenvolvida em torno deste modelo. Este desenvolvimento encontra-se particularmente enfatizado no contexto dos modelos lineares. Contudo, em muitas situações este modelo é claramente impróprio, por exemplo, quando os dados provêm de uma distribuição com caudas mais pesadas que a distribuição normal. De fato, a vulnerabilidade da inferência baseada no modelo normal com respeito a observações aberrantes (“outliers”) têm sido o motivo de vários estudos de robustez. Em particular, a detecção e tratamento de pontos “outliers” têm representado uma importante área de pesquisa da teoria estatística robusta. Neste sentido, alguns modelos alternativos ao normal e que são baseados em distribuições simétricas que permitem reduzir a influência dos “outliers”, têm sido sugeridos por diferentes autores. Por exemplo, Lange, Little e Taylor (1989) propõem o modelo t com $\nu > 0$ graus de liberdade como uma extensão paramétrica robusta do modelo normal, e ilustram o desempenho do novo modelo numa grande variabilidade de aplicações.

Por outro lado, trabalhos recentes sobre robustez mostram que a inferência estatística clássica, desenvolvida para o modelo normal, é basicamente correta, não somente para uma população normal, mas também para aquelas com simetria elíptica, desde que um certo ajuste pela curva da distribuição destas populações elípticas seja feito em caso de ser necessário (veja, por exemplo, Muirhead e Waternaux (1980); Browne (1982, 1984); Bentler (1983); Shapiro e Browne (1987) e Kano, Berkane e Bentler (1993)). Contudo, Tyler (1983) nota que embora a distribuição da população seja elíptica, a inferência baseada na teoria estatística normal é significativamente ineficiente quando a curtose da distribuição dos dados assume valores bem mais altos que moderados. Isto sugere a necessidade de sustentar as inferências numa determinada classe de modelos elípticos que possam cobrir distribuições com curtose alta, ao invés de realizar ajustes na teoria de inferência normal. Para estudar estas populações elípticas de caudas pesadas, novamente vários autores têm enfatizado a conveniência de uma determinada distribuição elíptica. Por exemplo, assim como Lange et al. (1989) recomendam a distribuição t -multivariada, Little (1988) e Yamaguchi (1990) usam a distribuição normal contaminada. Ambos os modelos incorporam parâmetros adicionais que permitem ajustar a curtose da distribuição dos dados. No caso da distribuição t com $\nu > 0$ graus de liberdade, a curtose é modelada através de ν . Em particular, as distribuições Cauchy e normal são obtidas colocando $\nu = 1$ e $\nu = \infty$, respectivamente.

O estudo tanto teórico quanto aplicado destas distribuições elípticas, como é o caso da distribuição t , começou desenvolver-se com um interesse crescente a partir do trabalho devido a Kelker (1970). No aspecto mais teórico os autores em geral têm concentrado seus esforços em pesquisar as propriedades desta classe geral de distribuições elípticas. Muitos são os trabalhos que podem ser referidos aqui, por exemplo, os devidos a Cambanis, Huang e Simons (1981) e Anderson e Fang (1987). A maioria destes trabalhos encontram-se unificados em Fang, Kotz e Ng (1990). Um esforço complementar e que incorpora alguns

resultados mais recentes e algumas extensões deles é dado no Capítulo 2 desta tese. No contexto aplicado, os autores em geral têm dirigido seus trabalhos a estudos sobre robustez, novamente, inúmeras referências podem ser dadas, por exemplo, veja o trabalho de Lange et al. (1989). Na etapa atual, a classe geral das distribuições elípticas proporcionam uma grande variedade de modelos estatísticos alternativos ao modelo normal. Uma subclasse destas distribuições que tem sido particularmente enfatizada, especialmente na estatística Bayesiana, é dada por aquelas distribuições que são mistura nos parâmetros de escala da distribuição normal. Esta subclasse é denominada de distribuições normais compostas e contém as distribuições t e normal contaminada. Como tem sido provado por diferentes autores, a maioria das propriedades da distribuição normal também vale na classe geral das distribuições elípticas e, em particular, na subclasse das distribuições normais compostas. Contudo, algumas propriedades tais como a equivalência entre independência e covariância nula de variáveis aleatórias que têm distribuição conjunta normal, são próprios desta distribuição e permitem caracterizá-la dentro da classe geral das distribuições elípticas. Outras caracterizações da distribuição normal têm sido dadas por vários autores. Tais resultados seguem do fato de que toda distribuição elíptica n -variada pode ser obtida misturando a distribuição uniforme sobre a superfície da n -esfera unitária com uma distribuição radial (definida sobre $[0, \infty)$). Embora esta representação esteja estabelecida em Kelker (1970), sua versão estocástica é mostrada mais tarde por Cambanis et al. (1981). Assim, cada distribuição elíptica ou, equivalentemente, sua versão padronizada, chamada comumente de esférica, é completamente caracterizada pela distribuição radial. Em particular, a distribuição elíptica (ou esférica) normal n -variada é caracterizada pela distribuição radial qui-quadrado com n graus de liberdade. Similarmente, a distribuição elíptica ou esférica t n -variada com ν graus de liberdade é caracterizada pela distribuição radial Fisher com n e ν graus de liberdade (mais formalmente, pela distribuição beta de tipo II com parâmetros $n/2$ e $\nu/2$).

Finalmente, tanto na teoria estatística desenvolvida em torno desta classe geral das distribuições elípticas quanto as aplicações que têm sido derivadas a partir de um membro particular delas, encontram-se dois enfoques para especificar o modelo estatístico. O primeiro deles assume que a distribuição conjunta das variáveis de interesse (sejam estas unidimensionais ou multidimensionais) é elíptica, de modo que estas variáveis podem estar não-correlacionadas sem serem independentes, a menos de que suponha-se normalidade. Tal enfoque é seguido entre outros autores por Anderson, Fang e Hsu (1986), Ghosh e Sinha (1980) e Zellner (1976). Os resultados destes autores mostram que grande parte da inferência estatística baseada no modelo normal permanece válida para estes modelos elípticos conjuntos ou dependentes, de modo que tais procedimentos são robustos com respeito a este tipo de não normalidade. O segundo enfoque, no entanto, supõe que as variáveis de interesse são independentes e que têm distribuições marginais elípticas. Este enfoque é seguido, por exemplo, por Kano et al. (1993) e Lange et al. (1989). Neste caso, em geral, a inferência estatística deve basear-se na teoria assintótica. Contudo, a vantagem destes modelos elípticos independentes é que muitos deles, tais como o modelo t -independente, implicam numa inferência robusta com respeito a observações aberrantes. Por último, convém notar aqui que sob normalidade ambos os enfoque são equivalentes.

Em resumo, os antecedentes acima mostram que a inferência estatística normal pode ser generalizada para cobrir uma classe mais ampla de modelo paramétricos que preservam a simetria elíptica do modelo normal, ou pode ser fornecida considerando modelos paramétricos alternativos ao normal e que, além de conter este último modelo, também superam sua falta de robustez com respeito a observações aberrantes, as que podem ser geradas por distribuições com caudas mais pesadas do que as da distribuição normal. O trabalho desenvolvido neste tese é motivado por estes fatos.

1.2 Definição dos objetivos

O objetivo geral deste trabalho está relacionado com a análise estatística de modelos de regressão linear usando a distribuição t . O estudo cobre tanto o caso em que as variáveis independentes ou explicativas são medidas sem erro quanto aquela situação em que estas variáveis podem estar medidas com erro. Neste último caso, contempla-se somente o modelo estrutural, o qual supõe que o verdadeiro valor destas variáveis é aleatório. Na especificação dos modelos considera-se ambos os enfoques mencionados acima. Assim, segundo Zellner (1986) e Lange, Little e Taylor (1989) definem-se os modelos de regressão t dependente e independente, respectivamente, e estabelecem-se as vantagens e desvantagens de ambos os modelos. No caso da regressão com variáveis medidas com erro, estende-se grande parte dos resultados obtidos para o contexto geral dos modelos elípticos. Finalmente, e como a distribuição latente deste estudo é a distribuição t multivariada, alguns resultados teóricos relativos a esta distribuição são também desenvolvidos. Desta forma, os objetivos específicos desta tese podem ser resumidos como segue:

- (1) propor algumas caracterizações da distribuição t multivariada;
- (2) desenvolver de forma unificada a inferência estatística para modelos de regressão com e sem erros nas variáveis, usando a distribuição t ;
- (3) derivar alguns testes assintóticos modificados por fatores tipo Bartlett para hipóteses dos parâmetros de regressão no modelo t independente;
- (4) estabelecer aqueles resultados que podem ser estendidos para a classe geral dos modelos elípticos.

1.3 Resumo de conteúdos

Em termos de atingir os objetivos definidos acima, este trabalho é organizado em quatro capítulos fundamentais mais um apêndice que resume, talvez de forma exageradamente sintética, o grande esforço investido na obtenção de alguns dos resultados dados nos Capítulos 4 e 5 desta tese. Assim, o conteúdo destes quatro capítulos podem ser resumidos como segue.

No Capítulo 2 são definidas as distribuições esféricas e elípticas multivariadas e proporciona-se um desenvolvimento teórico unificado das propriedades básicas destas distribuições. Diferentes exemplos que permitem construir famílias de distribuições esféricas ou elípticas são discutidos aqui. Neste capítulo também são estendidos alguns resultados de Mitchell (1989) para obter cumulantes conjuntos das derivadas da função de log-verossimilhança de um modelo elíptico multivariado. Tais resultados são complementados no apêndice deste trabalho. Finaliza-se este capítulo com uma discussão sobre as diferentes maneiras de caracterizar a distribuição normal dentro da classe geral das distribuições elípticas. As referências básicas utilizadas aqui são devidas a Kelker (1970); Cambanis, Huang e Simons (1981); Anderson e Fang (1987); Mitchell (1989); Muirhead (1982) e Fang, Kotz e Ng (1990).

O Capítulo 3 é dedicado ao estudo da distribuição t multivariada. Devido ao interesse que tem surgido na análise estatística por esta distribuição, que é motivo principal deste trabalho, suas propriedades são estudadas de maneira especial neste capítulo. Uma versão generalizada desta distribuição é considerada aqui e, a partir da qual, diferentes formas de caracterizá-la dentro da classe geral das distribuições elípticas são propostas. Mostra-se também como esta distribuição pode ser caracterizada na subclasse das distribuições normais compostas. Um resultado bem geral devido a Diaconis e Ylvisaker (1979, 1985) é aqui utilizado. Como consequência desta última caracterização, discute-se uma justi-

ficativa preditivista da distribuição t para seqüências de variáveis ou vetores aleatórios, considerando apenas condições de invariância frente a certos grupos de transformações ortogonais e uma certa linearidade nas variâncias ou matrizes de covariâncias condicionais. Alguns resultados prévios devidos a Kingman (1972); Smith (1981); Dawid (1978) e Diaconis, Eaton e Lauritzen (1992) são aqui utilizados.

No Capítulo 4 aplica-se a distribuição t no contexto do modelo de regressão linear usual. A análise estatística de duas versões deste modelo é desenvolvida. Na primeira versão segue-se os trabalhos de Zellner (1976) e especifica-se o modelo supondo que a distribuição conjunta dos erros é t multivariada. Mostra-se que vários resultados obtidos para o modelo normal são válidos também neste modelo. Em particular, os estimadores de máxima verossimilhança coincidem com os respectivos estimadores de mínimos quadrados e os testes usuais da hipótese linear $H_0 : A\beta = \gamma_0$, onde β é o vetor de parâmetros de regressão, se reduzem ao teste exato F de Fisher. A segunda versão é baseada nos estudos de Lange, Little e Taylor (1989), de modo que o modelo é especificado supondo que os erros são independentes e que têm distribuição comum t . Mostra-se que neste caso as estimativas de máxima verossimilhança são robustas, mas elas devem ser obtidas usando algoritmos numéricos. Ilustra-se a aplicação do algoritmo EM para obter estas estimativas considerando o fato de que a distribuição t é mistura de escala da distribuição normal. Finalmente, derivam-se as estatísticas da razão de verossimilhança e escore modificadas para o teste da hipótese linear $H_0 : A\beta = \gamma_0$.

No Capítulo 5, estuda-se o modelo de regressão linear simples com erros nas variáveis seguindo o enfoque estrutural e considerando-se condições de identificabilidade que supõem razões de variâncias conhecidas. Este capítulo é introduzido com um desenvolvimento unificado do modelo estrutural normal ou clássico. Alguns resultados novos são aqui estabelecidos considerando a parametrização ortogonal deste modelo. Em particular, as distribuições exatas de certas funções do estimador de máxima verossimilhança

são deduzidas. Também mostra-se que o fator de Bartlett da estatística da razão de máxima verossimilhança para hipóteses sobre o coeficiente de regressão pode ser obtido de forma fácil, avaliando diretamente o valor esperado desta estatística. Posteriormente, estabelecem-se critérios para formular modelos estruturais elípticos dependentes e independentes, mostrando-se que tanto o problema de identificabilidade quanto a parametrização ortogonal prevalecem nesta classe geral de modelos. Vários outros resultados obtidos supondo normalidade são também estendidos para os modelos elípticos dependentes. Finaliza-se este capítulo com o estudo detalhado dos modelos estruturais t dependente e independente. Para este último modelo, grande parte deste trabalho foi dedicado à derivação dos testes assintóticos modificados da razão de verossimilhança e escore para hipóteses sobre o coeficiente de regressão. Um trabalho similar devido a Wong (1989) no contexto do modelo normal e considerando somente o teste da razão de verossimilhança é estendido com este último resultado. Além disso, os elementos necessários para obter tal resultado são organizados de forma tal que eles possam ser aplicados a qualquer modelo estrutural elíptico independente. De maneira similar são proporcionadas as ferramentas no apêndice para a obtenção de cumulantes conjuntos das derivadas da log-verossimilhança de um modelo elíptico multivariado regular.

Capítulo 2

Distribuições esféricas e elípticas

2.1 Introdução

A partir da década de 70, as distribuições esféricas e elípticas vêm ocupando um importante e crescente lugar na teoria estatística e suas aplicações. As distribuições esféricas são caracterizadas por sua invariância com respeito a transformações ortogonais e formam uma classe geral de distribuições com a mesma simetria esférica da distribuição normal padrão. As distribuições elípticas, entretanto, podem ser geradas mediante transformações (lineares) de locação e escala das distribuições esféricas. Assim, cada distribuição esférica gera uma família paramétrica de distribuições elípticas através de transformações lineares. Reciprocamente, cada distribuição esférica é um elemento particular de alguma família de distribuições elípticas. Conseqüentemente, as distribuições elípticas formam uma classe generalizada de famílias paramétricas de distribuições probabilísticas que preservam a estrutura simétrica (simetria elíptica) das distribuições normais. De fato, a família das distribuições normais é um elemento particular desta classe e como será mostrado neste capítulo, a maior parte das propriedades da distribuição normal pode ser generalizada para as distribuições elípticas. Além da família de distribuições normais, outros elementos típicos na classe de distribuições elípticas são, por exemplo, a família de distribuições normais contaminadas (veja o Exemplo 2.8) e a família de distribuições t (veja o Exemplo 2.9). Contudo, um espectro amplo de modelos paramétricos alternativos ao modelo

normal pode ser obtida a partir da classe das distribuições elípticas (veja a Tabela 2.1).

Embora o interesse por estas distribuições seja relativamente recente, muitas de suas propriedades têm sido bem estudadas por diferentes autores. O primeiro desenvolvimento teórico e bem organizado das distribuições elípticas é devido a Kelker (1970). Desde então, estas distribuições têm sido o objetivo de diversos estudos em termos de pesquisar tanto suas propriedades (teóricas) como suas áreas de aplicação. Alguns resultados neste sentido são dados, por exemplo, em Devlin, Gnadesikan e Kettenring (1975), Kariya e Eaton (1977), Muirhead e Waternaux (1980) e Muirhead (1980). Estes resultados juntamente com os obtidos em Kelker (1970) encontram-se resumidos em Muirhead (1982). Destaca-se ainda o excelente trabalho de Cambanis, Huang e Simons (1981) que proporciona um rigoroso e sistemático tratamento das distribuições elípticas. Uma completa bibliografia, até o ano de 1981, com ampla discussão sobre as áreas de aplicação destas distribuições, pode ser encontrada em Chmielewski (1981). Em trabalhos mais recentes alguns estudos interessantes das propriedades das distribuições elípticas podem ser encontrados, entre outros, em Mitchell e Krzanowski (1985), Berkane e Bentler (1986, 1987), Anderson e Fang (1987) e Mitchell (1988, 1989). Um desenvolvimento resumido e bem organizado das distribuições esféricas e elípticas encontra-se em Fang, Kotz e Ng (1990, cap. 2).

Neste capítulo unificamos, com adaptações, o material encontrado nos trabalhos citados acima juntamente com alguns resultados mais recentes. Assim, as distribuições esféricas e elípticas são definidas na Seção 2.2, onde inclui-se também algumas propriedades preliminares e alguns exemplos ilustrativos. Na Seção 2.3 descreve-se uma representação estocástica das distribuições esféricas e elípticas que é de grande utilidade no desenvolvimento de suas propriedades. Na Seção 2.4 são estabelecidas as condições sobre as quais as distribuições esféricas e elípticas admitem uma densidade (com respeito à medida de Lebesgue). A Seção 2.5 é dedicada ao desenvolvimento de diferentes propriedades destas distribuições e também incorpora extensões e demonstrações mais simplificadas de

alguns resultados. Finalmente, na Seção 2.6 são dados diferentes resultado que permitem caracterizar a distribuição normal dentro da classe das distribuições elípticas.

Por outro lado, alguns procedimentos estatísticos têm sido desenvolvidos no contexto da classe geral das distribuições elípticas ou em torno de um elemento particular desta classe (por exemplo, a família t). O estudo da robustez de diferentes métodos estatísticos tem sido o objetivo de recentes trabalhos tanto teóricos como aplicados (em particular, em análise multivariada). Algumas contribuições iniciais desta natureza incluem, por exemplo, os trabalhos de Maronna (1976), Zellner (1976), Kariya e Eaton (1977), Chmielewski (1980), Ghosh e Sinha (1980), King (1980), Muirhead (1980) e Muirhead e Waternaux (1980) (veja, também, as referências dadas em Chmielewski (1981)). Resultados mais recentes são obtidos, entre outros, por Tylor (1982, 1983, 1987), Andersons e Fang (1986), Shapiro e Browne (1987), Lange, Little e Taylor (1989) e Kano, Berkane e Bentler (1993). Alguns destes resultados são utilizados nos capítulos finais deste trabalho. Embora a unificação destes resultados requeira um esforço considerável, sua realização seria uma importante contribuição para a teoria estatística.

2.2 Notações, definições e exemplos

Como ocorre com a distribuição normal multivariada, a definição de uma distribuição elíptica multivariada pode ser feita de diversas maneiras. Uma maneira é restringir-se ao caso não singular e basear a definição na forma de sua função de densidade (veja, por exemplo, Muirhead (1982); Mitchell (1989) e Berkane e Bentler (1987)). Se desejamos incorporar o caso singular na análise, então a definição pode ser feita em termos da forma da função característica (veja, por exemplo, Kelker (1970); Cambanis, Huang e Simons (1981) e Anderson, Fang e Hsu (1986)). Dada uma definição para a distribuição elíptica multivariada, a distribuição esférica multivariada pode ser estabelecida como um caso particular. Contudo, neste trabalho começamos definindo as distribuições esféricas

e então estendemos a definição para as distribuições elípticas n -dimensionais. Para $n = 1$, a classe das distribuições elípticas coincide com a classe das distribuições simétricas unidimensionais.

No que segue, se X e Y são variáveis ou vetores aleatórios n -dimensionais, a notação $X \stackrel{d}{=} Y$ indicará que X e Y têm a mesma distribuição. Se A é uma matriz simétrica de dimensão $n \times n$, as notações $A \geq 0$ e $A > 0$ serão usadas para indicar que A é definida não-negativa e definida positiva, respectivamente. O posto de uma matriz A será denotado por $r(A)$ e a notação A^- indicará uma inversa generalizada da matriz A , isto é, $AA^-A = A$. Além disto, I_n denotará a matriz identidade de dimensão $n \times n$ e a notação O_n será usada para denotar o conjunto das matrizes $\Gamma(n \times n)$ que são ortogonais, isto é, $\Gamma^T \Gamma = I_n$, onde Γ^T denota a transposta da matriz Γ . O_n é um grupo, chamado grupo ortogonal, e contém as matrizes de permutação $n \times n$, $P = [p_{ij}]$, com $p_{ij} = 0$ ou 1 (veja, por exemplo, a Seção 1.3 em Fang et al. (1990)).

Definição 2.1 *O vetor aleatório $X(n \times 1)$ é dito ter uma distribuição (com simetria) esférica (n -variada) se para cada matriz Γ em O_n , tem-se que*

$$X \stackrel{d}{=} \Gamma X . \quad (2.1)$$

De (2.1) é claro que um vetor aleatório $X(n \times 1)$ tem distribuição esférica se e somente se a função característica de X , $\phi_X(t) = E(e^{it^T X})$, $t \in \mathbb{R}^n$, é da forma

$$\phi_X(t) = \phi(t^T t) , \quad t \in \mathbb{R}^n , \quad (2.2)$$

para alguma função ϕ , com $\phi(u) \in \mathbb{R}$, $u \geq 0$. De (2.1) segue também que se X tem densidade (com respeito à medida de Lebesgue em $\mathbb{R}^n$), ela é da forma

$$p_X(x) = f(x^T x) , \quad x \in \mathbb{R}^n , \quad (2.3)$$

para alguma função f , com $f(u) \geq 0$, $u \geq 0$. Como é notado em Kelker (1970), a expressão $r^{n-1}f(r^2)$ deve ser integrável sobre $[0, \infty)$ para que (2.3) seja uma densidade n -dimensional (veja a Seção 2.4).

Alguns exemplos de distribuições esféricas são (Muirhead (1980, 1982), Fang, Kotz e Ng (1990)):

Exemplo 2.1 Seja $U^{(n)}(n \times 1)$ um vetor aleatório com distribuição uniforme sobre o conjunto

$$S_n = \{x \in \mathbb{R}^n : \|x\| = 1\} , \quad (2.4)$$

a superfície da esfera unitária de $\mathbb{R}^n$, onde $\|x\| = (x^\top x)^{\frac{1}{2}}$, $x \in \mathbb{R}^n$. Isto é,

$$U^{(n)} \sim \mathcal{U}(S_n) .$$

É óbvio que $U^{(n)} \stackrel{d}{=} \Gamma U^{(n)}$ para toda matriz Γ em O_n . Além disto, Kariya e Eaton (1977) e Eaton (1977, 1986) notam que a distribuição uniforme é a única distribuição sobre S_n que é invariante por transformações ortogonais.

A distribuição de $U^{(n)}$ é singular sobre $\mathbb{R}^n$ e sua função característica é denotada por $\Omega_n(t^\top t)$, $t \in \mathbb{R}^n$. A função $\Omega_n(u)$, $u \geq 0$, pode ser expressa de diversas maneiras, por exemplo (veja Fang et al. (1990, p.70)),

$$\Omega_n(u) = \frac{\Gamma\left(\frac{1}{2}n\right)}{\Gamma\left[\frac{1}{2}(n-1)\right] \pi^{\frac{1}{2}}} \int_0^\pi e^{i\sqrt{u}\cos\theta} (\sin\theta)^{n-2} d\theta , \quad u \geq 0 , \quad (2.5)$$

onde $\Gamma(a)$, $a > 0$, denota a função gama.

Por outro lado, seja $(V_1, V_2, \dots, V_m)^\top \sim D_m(\alpha_1, \alpha_2, \dots, \alpha_m; \alpha_{m+1})$ a distribuição Dirichlet (m -variada) com parâmetros $\alpha_1, \alpha_2, \dots, \alpha_{m+1} > 0$. Então $(V_1, V_2, \dots, V_m)^\top$ tem densidade dada por (veja a seção 1.4 em Fang et al. (1990))

$$p_{v_1, v_2, \dots, v_m}(v_1, v_2, \dots, v_m) = \frac{\Gamma\left(\sum_{i=1}^{m+1} \alpha_i\right)}{\prod_{i=1}^{m+1} \Gamma(\alpha_i)} \left(\prod_{i=1}^m v_i^{\alpha_i-1}\right) \left(1 - \sum_{i=1}^m v_i\right)^{\alpha_{m+1}-1} , \quad (2.6)$$

onde $\sum_{i=1}^m v_i < 1$, $v_1, v_2, \dots, v_m \geq 0$. Para $m = 1$, tem-se que

$$D_1(\alpha_1; \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2) . \quad (2.7)$$

Sendo $(U_1^2, U_2^2, \dots, U_m^2)^\top \sim D_m(1/2, 1/2, \dots, 1/2; (n-m)/2)$, de (2.6) é fácil ver que a distribuição do vetor $U^{(n)} = (U_1, U_2, \dots, U_n)^\top$ define a densidade marginal m -variada, $1 \leq m < n$, dada por

$$p_{u_1, u_2, \dots, u_m} = \frac{\Gamma\left(\frac{1}{2}n\right)}{\Gamma\left[\frac{1}{2}(n-m)\right] \pi^{\frac{1}{2}m}} \left(1 - \sum_{i=1}^m u_i^2\right), \quad \sum_{i=1}^m u_i^2 < 1. \quad (2.8)$$

Note-se que a distribuição de $(U_1, U_2, \dots, U_m)^\top$ é esférica mas não é necessariamente uniforme.

Exemplo 2.2 Seja $X(n \times 1)$ um vetor aleatório com distribuição uniforme na esfera unitária em $\mathbb{R}^n$. Isto é, $X \sim \mathcal{U}(E_n)$, onde

$$E_n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}.$$

Então X tem função característica da forma (2.2) com ϕ dada por

$$\phi(u) = \frac{2\Gamma\left[\frac{1}{2}(n+2)\right]}{\Gamma\left[\frac{1}{2}(n+1)\right] \pi^{\frac{1}{2}}} \int_0^\infty (1-r^2)^{\frac{1}{2}(n-1)} \cos(ur) dr, \quad u \geq 0,$$

e com densidade

$$p_X(x) = f(x^\top x) = \frac{n\Gamma\left(\frac{1}{2}n\right)}{2\pi^{\frac{1}{2}n}} I_{E_n}(x),$$

onde

$$I_A(x) = \begin{cases} 1 & , \text{ se } x \in A \\ 0 & , \text{ caso contrário,} \end{cases}$$

é a função indicadora do conjunto A .

Exemplo 2.3 Seja $X \sim N_n(0, I_n)$, a distribuição normal n -variada padrão. Isto é, X tem função característica

$$\phi_X(t) = \phi(t^\top t) = e^{-\frac{1}{2}t^\top t}, \quad t \in \mathbb{R}^n, \quad (2.9)$$

e função de densidade

$$p_X(x) = f(x^\top x) = (2\pi)^{-\frac{1}{2}n} e^{-\frac{1}{2}x^\top x}, \quad x \in \mathbb{R}^n. \quad (2.10)$$

Exemplo 2.4 Seja $X \sim NC_n(0, I_n; \varepsilon, \sigma)$, $0 \leq \varepsilon \leq 1$, $\sigma > 0$, a distribuição normal n -variada padrão “ ε -contaminada”. Isto é, X tem densidade dada por

$$p_X(x|\varepsilon, \sigma) = f(x^\top x) = (1 - \varepsilon)(2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}x^\top x} + \varepsilon(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}x^\top x}, \quad x \in \mathbb{R}^n. \quad (2.11)$$

Exemplo 2.5 Seja $X \sim t_n(0, I_n; \nu)$, a distribuição t (Student) n -variada padrão com ν ($\nu > 0$) graus de liberdade. Isto é, X tem densidade da forma

$$p_X(x|\nu) = f(x^\top x) = \frac{\Gamma[\frac{1}{2}(\nu + n)]\nu^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)\pi^{\frac{1}{2}n}} \{\nu + x^\top x\}^{-\frac{1}{2}(\nu+n)}, \quad x \in \mathbb{R}^n. \quad (2.12)$$

Observe que (2.12) se reduz a (2.10) quando $\nu = \infty$ ($\nu \rightarrow \infty$), e a distribuição Cauchy n -variada padrão quando $\nu = 1$.

Uma forma conveniente de gerar algumas distribuições esféricas é dada no exemplo que segue.

Exemplo 2.6 Distribuições normais n -variadas compostas.

Sejam $X_1, X_2, \dots, X_n$ e V , ($V > 0$), variáveis aleatórias tais que, dado $V = v$, $X_1, X_2, \dots, X_n$ são independentes e identicamente distribuídas (iid) com distribuição $N(0, v)$. Se V tem função de distribuição F_V , então a função de densidade marginal de $X = (X_1, \dots, X_n)^\top$ é

$$p_X(x) = f(x^\top x) = \int_0^\infty (2\pi v)^{-\frac{1}{2}n} e^{-\frac{1}{2v}x^\top x} dF_V(v), \quad x \in \mathbb{R}^n. \quad (2.13)$$

A distribuição esférica assim construída é chamada mistura de escala da distribuição normal (veja, por exemplo, Andrews e Mallows (1974)). A classe de distribuições esféricas obtida variando F_V em (2.13) é denominada classe das distribuições normais compostas. Desta construção segue a representação estocástica

$$X =^d \sqrt{V}Z,$$

onde $Z = (Z_1, \dots, Z_n)^\top$ tem distribuição $N_n(0, I_n)$ e é independente de V . Esta representação é útil na geração de valores para X . Se V assume valores 1 e σ^2 com probabilidades $1 - \varepsilon$ e ε , respectivamente, então X tem a distribuição normal n -variada

“ ε -contaminada” dada em (2.11). Entretanto, se V tem distribuição gama inversa, com parâmetros $(\frac{1}{2}\nu, \frac{1}{2}\nu)$, $\nu > 0$, isto é, se sua densidade é

$$p_\nu(v|\nu) = \frac{(\frac{1}{2}\nu)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} v^{-\frac{1}{2}(\nu+2)} e^{-\frac{1}{2}\frac{\nu}{v}}, \quad v > 0,$$

então a distribuição de X é como em (2.12) (veja o Capítulo 3).

A definição das distribuições elípticas é dada seguindo Cambanis et al. (1981) (veja, também, Kelker (1970) e Anderson e Fang (1986)). Para motivar a definição, suponha que $Z(m \times 1)$ é um vetor aleatório tendo distribuição esférica m -variada com função característica $\phi(s^\top s)$, $s \in \mathbb{R}^m$. Considere o vetor aleatório $X(n \times 1)$ definido pela transformação linear

$$X = \mu + AZ,$$

onde A é uma matriz $n \times m$ e $\mu(n \times 1) \in \mathbb{R}^n$. Então, a função característica de X é dada por

$$\phi_X(t) = E\left(e^{it^\top(\mu + AZ)}\right) = e^{it^\top\mu} \phi(t^\top \Sigma t), \quad t \in \mathbb{R}^n,$$

onde $\Sigma = AA^\top$. Com este resultado em mente, a seguinte definição da distribuição elíptica surge de forma natural.

Definição 2.2 *O vetor aleatório $X(n \times 1)$ é dito ter distribuição (com simetria) elíptica (n -variada) com vetor de locação $\mu(n \times 1) \in \mathbb{R}^n$ e matriz de dispersão $\Sigma(n \times n) \geq 0$, se a função característica de X é da forma*

$$\phi_X(t) = e^{it^\top\mu} \phi(t^\top \Sigma t), \quad t \in \mathbb{R}^n, \quad (2.14)$$

para alguma função ϕ ($\phi(u) \in \mathbb{R}$, $u \geq 0$). Em tal caso dizemos que X tem distribuição elíptica (n -variada) com parâmetros μ , Σ (e ϕ) e escrevemos

$$X \sim El_n(\mu, \Sigma),$$

ou

$$X \sim El_n(\mu, \Sigma; \phi) , \quad (2.15)$$

quando é necessário indicar a dependência de ϕ na representação paramétrica.

Uma forma equivalente de definir a distribuição elíptica é dada por: $X \sim El_n(\mu, \Sigma; \phi)$

se

$$X =^d \mu + AZ , \quad (2.16)$$

onde A é uma matriz $n \times k$ tal que $r(A) = k$ e $AA^T = \Sigma$ e $Z(k \times 1)$ é um vetor aleatório tendo distribuição esférica k -variado com função característica $\phi(s^T s)$, $s \in \mathbb{R}^k$ (veja Fang et al. (1990)).

Se a distribuição de X tem densidade (com respeito à medida de Lebesgue em $\mathbb{R}^n$), então $\Sigma > 0$ (Arnold (1981, p.41)) e sua densidade é da forma

$$p_X(x|\mu, \Sigma) = |\Sigma|^{-\frac{1}{2}} f((x - \mu)^T \Sigma^{-1} (x - \mu)) , \quad x \in \mathbb{R}^n , \quad (2.17)$$

para alguma função f , ($f(u) \geq 0$, $u \geq 0$). Neste caso, costuma-se substituir ϕ por f na representação paramétrica dada em (2.15).

Note-se que (2.17) pode ser obtida usando a transformação linear (2.16) com $k = n$, isto é,

$$X = \mu + \Sigma^{1/2} Z ,$$

onde $Z(n \times 1)$ tem distribuição esférica com densidade $f(z^T z)$, $z \in \mathbb{R}^n$. De fato, desde que $Z = \Sigma^{-1/2}(X - \mu)$, o jacobiano da transformação é $|\Sigma|^{-1/2}$ e (2.17) segue avaliando a função $f(z^T z)$ em $z = \Sigma^{-1/2}(x - \mu)$.

Por outro lado, se em (2.15) colocamos $\mu = 0$ e $\Sigma = I_n$, isto é,

$$X \sim El_n(0, I_n; \phi) ,$$

então (2.14) e (2.17) se reduzem a (2.2) e (2.3), respectivamente, e a distribuição de X é esférica.

Se $X \sim El_n(0, I_n; \phi)$ e σ é uma constante positiva, é fácil ver que $\sigma X \sim El_n(0, \sigma^2 I_n; \phi)$ também é esférica (veja a Seção 2.5.2).

Exemplos típicos de distribuições elípticas são (Muirhead (1980)):

Exemplo 2.7 Seja $X \sim N_n(\mu, \Sigma)$, a distribuição normal n -variada com função característica

$$\phi_X(t) = e^{it^\top \mu} e^{-\frac{1}{2}t^\top \Sigma t}, \quad t \in \mathbb{R}^n.$$

Se $\Sigma > 0$, então X tem densidade

$$p_X(x|\mu, \Sigma) = |\Sigma|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}, \quad x \in \mathbb{R}^n.$$

Exemplo 2.8 Seja $X \sim NC_n(\mu, \Sigma; \varepsilon, \sigma)$, a distribuição normal n -variada “ ε -contaminada”. Se X tem densidade ($\sigma^2 > 0$, $\Sigma > 0$), então

$$p_X(x|\mu, \Sigma, \sigma, \varepsilon) = \varepsilon p_Z(x|\mu, \Sigma) + (1 - \varepsilon) p_Z(\sigma x|\mu, \Sigma), \quad x \in \mathbb{R}^n, \quad (0 \leq \varepsilon \leq 1),$$

onde $p_Z(z|\mu, \Sigma)$ é a densidade de $Z \sim N_n(\mu, \Sigma)$.

Exemplo 2.9 Seja $X \sim t_n(\mu, \Sigma; \nu)$, a distribuição t n -variada com ν graus de liberdade e parâmetros de locação μ e dispersão Σ (veja o Capítulo 3). Se $\Sigma > 0$, então X tem densidade dada por

$$p_X(x|\mu, \Sigma; \nu) = |\Sigma|^{-\frac{1}{2}} \frac{\Gamma\left[\frac{1}{2}(\nu + n)\right] \nu^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu) \pi^{\frac{1}{2}n}} \left\{ \nu + (x - \mu)^\top \Sigma^{-1}(x - \mu) \right\}^{-\frac{1}{2}(\nu+n)}, \quad x \in \mathbb{R}^n.$$

A função característica desta distribuição é derivada em Sutradhar (1986).

Exemplo 2.10 Consideremos agora as distribuições normais n -variadas compostas, com função característica (veja (2.40) na Seção 2.4)

$$\phi_X(t) = e^{it^\top \mu} \int_{[0, \infty)} e^{-\frac{1}{2}vt^\top \Sigma t} dF(v), \quad t \in \mathbb{R}^n,$$

para alguma função distribuição F sobre $[0, \infty)$. Se $F(0) = 0$ e $\Sigma > 0$, então X tem densidade (veja (2.41) na Seção 2.4)

$$p_X(x|\mu, \Sigma) = |\Sigma|^{-\frac{1}{2}} \int_{(0, \infty)} (2\pi v)^{-\frac{1}{2}n} e^{-\frac{1}{2v}(x-\mu)^\top \Sigma^{-1}(x-\mu)} dF(v), \quad x \in \mathbb{R}^n.$$

Os Exemplos 2.8 e 2.9 podem ser gerados a partir do Exemplo 2.10 na forma descrita no Exemplo 2.6. Maiores detalhes destes e outros exemplos podem ser encontrados em Fang et al. (1990).

2.3 Representações estocásticas

Em termos de caracterizar a classe de funções $\phi : [0, \infty) \rightarrow \mathbb{R}$, para que $\phi(t^\top t)$, $t \in \mathbb{R}^n$, seja a função característica de uma distribuição esférica n -dimensional, a seguinte notação será incorporada (veja Cambanis et al. (1981)).

Seja,

$$\Phi_k = \left\{ \phi(\cdot) : \phi(t^\top t), t \in \mathbb{R}^k, \text{ é uma função característica} \right\}, \quad k \geq 1. \quad (2.18)$$

É fácil ver que $\Phi_1 \supset \Phi_2 \supset \cdots \supset \Phi_k \supset \Phi_{k+1} \supset \cdots$ e, portanto, que

$$\Phi_k \downarrow \Phi_\infty = \bigcap_{m=1}^{\infty} \Phi_m. \quad (2.19)$$

Schoenberg (1938) mostra que $\phi \in \Phi_k$ se e somente se

$$\phi(u) = \int_{[0, \infty)} \Omega_k(r^2 u) dF(r), \quad u \geq 0, \quad (2.20)$$

onde F é uma função de distribuição sobre $[0, \infty)$, e

$$\Omega_k(s^\top s) = E(e^{is^\top U^{(k)}}), \quad s \in \mathbb{R}^k,$$

é a função característica de $U^{(k)} \sim \mathcal{U}(S_k)$ (veja o Exemplo 2.1). Além disso, $\phi \in \Phi_\infty$ se e somente se

$$\phi(u) = \int_{[0, \infty)} e^{-\frac{1}{2}r^2 u} dF(r), \quad (2.21)$$

com F sendo uma função distribuição sobre $[0, \infty)$.

A prova destes resultados pode ser encontrada em Kelker (1970) ou em Fang et al. (1990). Uma forma alternativa de enunciar (2.21), em termos de seqüências de variáveis aleatórias, é dada no Capítulo 3 (veja, também, Kelker (1970); Kingman (1972) e Vitale (1993)).

O seguinte fato é consequência direta de (2.20) e estabelece uma representação estocástica da distribuição elíptica e que será de grande utilidade na derivação das propriedades desta distribuição. Este fato é devido a Cambanis et al. (1981) e sua prova pode ser vista também em Fang et al. (1990). Para enunciar este fato, a notação do Exemplo 2.1 é requerida.

Seja $U^{(k)} \sim \mathcal{U}(S_k)$ a distribuição uniforme sobre S_k , a superfície da esfera unitária k -dimensional E_k , e seja X um vetor aleatório $n \times 1$.

Fato 2.1 $X \sim El_n(\mu, \Sigma; \phi)$ com $r(\Sigma) = k$ se e somente se

$$X =^d \mu + RAU^{(k)}, \quad (2.22)$$

onde A é uma matriz $n \times k$ com $r(A) = k$ e tal que $\Sigma = AA^T$, e R é uma variável aleatória não-negativa, sendo independente de $U^{(k)}$ e tendo função de distribuição F relacionada com a função ϕ através de (2.20).

Do Fato 2.1 é claro que se $r(\Sigma) = k$, então a classe de funções ϕ que são admissíveis para definir uma função característica de alguma distribuição elíptica k -variada é Φ_k .

Por outro lado, usando as propriedades de uma inversa generalizada, é fácil ver que (2.22) implica que

$$Q(X) \equiv (X - \mu)^T \Sigma^- (X - \mu) =^d R^2, \quad (2.23)$$

isto é, $Q(X)$ tem função distribuição $F(\sqrt{\cdot})$. Se $k = n$ ($r(\Sigma) = n$), então $A = \Sigma^{1/2}$ e $\Sigma^- = \Sigma^{-1}$. Em particular, $X \sim El_n(0, I_n; \phi)$ se e somente se

$$X =^d RU^{(n)}, \quad (2.24)$$

e, portanto,

$$\|X\| =^d R, \quad (2.25)$$

onde R e $U^{(n)}$ são definidas como no Fato 2.1.

O fato seguinte é também provado em Cambanis et al. (1981) e estabelece que em (2.20) a relação entre ϕ e F é um a um.

Fato 2.2 $\phi \in \Phi_k$ se e somente se ϕ é contínua e

$$\int_0^\infty \phi(2rv)q_k(v)dv, \quad r \geq 0,$$

é a transformada de Laplace de uma variável aleatória não-negativa (a saber, $Q(X)$ dada em (2.23)) cuja função distribuição é $F(\sqrt{\cdot})$ quando ϕ é dada por (2.20). Aqui, q_k denota a densidade da distribuição qui-quadrado com k graus de liberdade (χ_k^2).

Os seguintes fatos serão usados mais adiante e são consequência direta de (2.24) e (2.25).

Fato 2.3 Suponha que $X =^d RU^{(n)} \sim El_n(0, I_n; \phi)$ com $P(X = 0) = 0$. Então,

$$\frac{X}{\|X\|} =^d U^{(n)} \quad (2.26)$$

e é independente de $\|X\| =^d R$.

Com efeito, note que a aplicação $x \rightarrow (\|x\|, x/\|x\|)^T$ é (Borel) mensurável sobre $\mathbb{R}^n - \{0\}$. Logo, de (2.24) e (2.25) segue que

$$\left(\|X\|, \frac{X}{\|X\|} \right)^T =^d (R, U^{(n)})^T,$$

o que prova o fato. □

Uma prova alternativa é dada em Muirhead (1982, p.38) (veja, também, Kariya e Eaton (1977) e Fang et al. (1990)).

De (2.26) é claro que se $X \sim El_n(0, I_n; \phi)$ com $P(X = 0) = 0$ e $Z \sim N_n(0, I_n)$, então (Kariya e Eaton (1977)),

$$\frac{X}{\|X\|} \stackrel{d}{=} \frac{Z}{\|Z\|}, \quad (2.27)$$

desde que, em particular, $Z \sim El_n(0, I_n; \phi)$ com $P(Z = 0) = 0$ e com ϕ dada em (2.9).

Fato 2.4 *Suponha que $U^{(n)} = (U_1^\top, U_2^\top)^\top$, onde U_1 é um vetor $m \times 1$, $1 \leq m < n$. Então,*

$$(U_1^\top, U_2^\top)^\top \stackrel{d}{=} (\sqrt{V_1}U^{(m)\top}, \sqrt{V_2}U^{(n-m)\top})^\top, \quad (2.28)$$

onde $V_1, V_2 \geq 0$, $V_1 + V_2 = 1$; $(V_1, V_2)^\top$, $U^{(m)}$ e $U^{(n-m)}$ são independentes, com

$$V_1 \sim \text{Beta}\left(\frac{1}{2}m, \frac{1}{2}(n-m)\right). \quad (2.29)$$

Para provar este fato, suponha que $X = (X_1^\top, X_2^\top)^\top \sim N_n(0, I_n)$, onde X_1 é um vetor $m \times 1$, $1 \leq m < n$. Sejam

$$U^{(m)} = \frac{X_1}{\|X_1\|} \quad \text{e} \quad U^{(n-m)} = \frac{X_2}{\|X_2\|}.$$

Então, da independência entre X_1 e X_2 e do Fato 2.3, segue que $U^{(m)}$, $U^{(n-m)}$, $\|X_1\|$ e $\|X_2\|$ são conjuntamente independentes. Definindo

$$V_1 = \frac{\|X_1\|^2}{\|X\|^2} \quad \left(V_2 = 1 - V_1 = \frac{\|X_2\|^2}{\|X\|^2} \right),$$

segue-se que V_1 , $U^{(m)}$ e $U^{(n-m)}$ são independentes. Além disto, V_1 tem a distribuição desejada (isto é, (2.29)), já que $\|X_1\|^2 \sim \chi_m^2$ e $\|X_2\|^2 \sim \chi_{(n-m)}^2$ são independentes (veja, também, a Seção 2.5.5). Logo, aplicando (2.26), tem-se que

$$(U_1^\top, U_2^\top) = U^{(n)} \stackrel{d}{=} \frac{X}{\|X\|} = \left(\frac{X_1^\top}{\|X\|}, \frac{X_2^\top}{\|X\|} \right)^\top = \left(\sqrt{V_1}U^{(m)\top}, \sqrt{1-V_1}U^{(n-m)\top} \right)^\top,$$

o que prova o Fato 2.4. □

De (2.28) segue que se $X \stackrel{d}{=} RU^{(n)} \sim El_n(0, I_n; \phi)$, então

$$X = (X_1^\top, X_2^\top)^\top \stackrel{d}{=} (R\sqrt{V_1}U^{(m)\top}, R\sqrt{V_2}U^{(n-m)\top})^\top, \quad (2.30)$$

com $V_1, V_2 \geq 0$, $V_1 + V_2 = 1$; R , (V_1, V_2) , $U^{(m)}$ e $U^{(n-m)}$ sendo independentes e V_1 tendo a distribuição da forma (2.29). Além disto, usando (2.30) e o fato de que (veja a Seção 2.5.2) $X_1 \stackrel{d}{=} R_1 U^{(m)} \sim El_m(0, I_m; \phi)$, onde $R_1 \stackrel{d}{=} \|X_1\|$ e $U^{(m)}$ são independentes, obtém-se que

$$R_1 U^{(m)} \stackrel{d}{=} X_1 \stackrel{d}{=} R \sqrt{V_1} U^{(m)} .$$

A correspondência um a um entre ϕ e R em (2.20) (estabelecida no Fato 2.2) implica que

$$R_1 \stackrel{d}{=} \|X_1\| \stackrel{d}{=} R \sqrt{V_1} , \quad (2.31)$$

isto é, se R_1 tem função distribuição F_1 , então F_1 coloca massa $F(0)$ no ponto zero e sobre $(0, \infty)$ tem densidade dada por

$$p_{R_1}(r_1) = \frac{2\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}m)\Gamma[\frac{1}{2}(n-m)]} r_1^{m-1} \int_{r_1}^{\infty} r^{2-n} (r^2 - r_1^2)^{\frac{1}{2}(n-m)-1} dF(r) , \quad 0 < r_1 < \infty . \quad (2.32)$$

Se $P(X = 0) = 0$, então $P(R = 0) = F(0) = 0$. Em tal caso, R_1 tem a densidade (2.32).

Note-se agora que se $\phi \in \Phi_k$, então para cada $1 \leq n \leq k$, $\phi \in \Phi_n$ e, de acordo com (2.20) e com o Fato 2.2,

$$\phi(u) = \int_0^{\infty} \Omega_n(r^2 u) dF_n(r) , \quad u \geq 0 ,$$

para uma única função de distribuição F_n sobre $[0, \infty)$. A relação precisa entre as diferentes F_n , $1 \leq n < k$, é especificada em (2.31) colocando $R = R_n \sim F_n$, $R_1 = R_m \sim F_m$ e $V_1 = V_{mn} \sim \text{Beta}(m/2, (n-m)/2)$, $1 \leq m < n \leq k$.

Os resultados derivados do Fato 2.4 podem ser generalizados como se segue (veja, também, Anderson e Fang (1986) e Fang et al. (1990)).

Suponha que $X \stackrel{d}{=} R U^{(n)} \sim El_n(0, I_n; \phi)$ é particionado como $X = (X_1^\top, X_2^\top, \dots, X_k^\top)^\top$, onde $X_1, X_2, \dots, X_k$ são vetores com $n_1, n_2, \dots, n_k$ ($n_1 + n_2 + \dots + n_k = n$) componentes do vetor X , respectivamente. Então, a relação (2.30) é estendida da seguinte maneira:

$$X = (X_1^\top, X_2^\top, \dots, X_k^\top)^\top \stackrel{d}{=} (R \sqrt{V_1} U^{(n_1)}, R \sqrt{V_2} U^{(n_2)\top}, \dots, R \sqrt{V_k} U^{(n_k)\top})^\top , \quad (2.33)$$

onde $V_1, V_2, \dots, V_k \geq 0$, $V_1 + V_2 + \dots + V_k = 1$, $(V_1, V_2, \dots, V_k)^\top$, $U^{(n_1)}, U^{(n_2)}, \dots, U^{(n_k)}$ e R são independentes, com (veja (2.6) no Exemplo 2.1)

$$(V_1, V_2, \dots, V_{k-1})^\top \sim D_{k-1} \left(\frac{1}{2}n_1, \frac{1}{2}n_2, \dots, \frac{1}{2}n_{k-1}; \frac{1}{2}n_k \right). \quad (2.34)$$

Além disto, como para cada $i = 1, \dots, k$, tem-se que $X_i =^d R_i U^{(n_i)} \sim E\ell_{n_i}(0, I_{n_i}; \phi)$ (veja a Seção 2.3.2), onde $R_i =^d \|X_i\|$ e $U^{(n_i)}$ são definidos como no Fato 2.1, então (2.31) é estendida como

$$(R_1, R_2, \dots, R_k)^\top =^d (\|X_1\|, \|X_2\|, \dots, \|X_k\|)^\top =^d (R\sqrt{V_1}, R\sqrt{V_2}, \dots, R\sqrt{V_k})^\top. \quad (2.35)$$

Quando $P(X = 0) = 0$ e $n_k \geq 1$, é fácil ver que para $1 \leq m < k$, o vetor aleatório $(R_1, R_2, \dots, R_m)^\top =^d (\|X_1\|, \|X_2\|, \dots, \|X_m\|)^\top$ tem densidade dada por (veja, também, (2.38) e (2.39) na Seção 2.4)

$$p_{R_1, R_2, \dots, R_m}(r_1, r_2, \dots, r_m) = \frac{2^m \Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}N) \prod_{i=1}^m \Gamma(\frac{1}{2}n_i)} \prod_{i=1}^m r_i^{n_i-1} \int_{(\sum_{i=1}^m r_i^2)^{\frac{1}{2}}}^{\infty} r^{2-n} \left(r^2 - \sum_{i=1}^m r_i^2 \right)^{\frac{1}{2}N-1} dF(r), \quad (2.36)$$

onde $0 < r_1, r_2, \dots, r_m < \infty$ e $N = n - \sum_{i=1}^m n_i$. Este último resultado generaliza (2.32) no caso em que a distribuição de X estabelece massa nula na origem.

Para finalizar esta seção, considere o seguinte resumo de alguns dos resultados relacionados com as distribuições esféricas.

Seja $X = (X_1, X_2, \dots, X_n)^\top$ um vetor aleatório $n \times 1$. Então, as seguintes afirmações são equivalentes:

- (i) $X =^d \Gamma X$, para cada $\Gamma \in O_n$;
- (ii) $\phi_X(t) = \phi(t^\top t)$, $t \in \mathbb{R}^n$, para alguma função ϕ e Φ_n ;
- (iii) $X =^d R U^{(n)}$, para alguma variável aleatória $R \geq 0$, independente de $U^{(n)}$;
- (iv) $a^\top X =^d \|a\| X_1$, para cada $a \in \mathbb{R}^n$.

A equivalência de (i) e (ii) é direta. A prova de que (ii) e (iii) são equivalentes pode ser encontrada em Cambanis et al. (1981) (veja o Fato 2.1). Finalmente, é fácil ver que (ii) e (iv) são também equivalentes (veja Fang et al. (1990, p.31)). Note-se que em (iv) é suficiente tomar $a \in S_n$, isto é, $a \in \mathbb{R}^n$ com $\|a\| = 1$ (veja Fang, Zhu e Bentler (1993)).

Com estes resultados e considerando as propriedades da distribuição normal padrão, um enfoque útil para desenvolver a teoria das distribuições elípticas pode ser representado esquematicamente como segue:

$$X_i \Rightarrow X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \Rightarrow U^{(n)} = \frac{X}{\|X\|} \Rightarrow RU^{(n)} \Rightarrow \mu + RAU^{(n)}$$

$$N(0, 1) \Rightarrow N_n(0, I_n) \Rightarrow \text{Uniforme} \Rightarrow \text{Esférica} \Rightarrow \text{Elíptica}.$$

Usando este enfoque, as propriedades da distribuição normal podem ser estendidas para as distribuições esféricas e elípticas.

2.4 Existência de densidade

O vetor aleatório $X \sim El_n(0, I_n; \phi)$ não possui necessariamente uma densidade (veja o Exemplo 2.1). Contudo, se X tem densidade, de acordo com (2.1), esta deve ser invariante frente a transformações ortogonais e, conseqüentemente, deve ser da forma $f(x^\top x)$, $x \in \mathbb{R}^n$, para alguma função f não-negativa. Neste caso, tem-se que

$$\int_{\mathbb{R}^n} f(x^\top x) dx = s_n^{-1} \int_0^\infty r^{n-1} f(r^2) dr = 1 ,$$

onde

$$s_n = \frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}$$

é a superfície de S_n (veja (2.4)). Daqui segue que uma função não-negativa f pode ser usada para definir a densidade de alguma distribuição esférica se e somente se

$$\int_0^\infty r^{n-1} f(r^2) dr < \infty .$$

Neste caso, escrevemos $X \sim El_n(0, I_n; f)$ ao invés de $X \sim El_n(0, I_n; \phi)$.

Como é notado em Kelker (1970) e em Cambanis et al. (1981), se $X \stackrel{d}{=} RU^{(n)} \sim El_n(0, I_n; \phi)$, então $X \sim El_n(0, I_n; f)$ se e somente se R tem densidade p_R , relacionada com f através de

$$p_R(r) = s_n r^{n-1} f(r^2), \quad r \geq 0. \quad (2.37)$$

Além disto, se $P(X = 0) = 0$, então todas as distribuições marginais de X têm densidade (veja a Seção 2.5.2). A demonstração destes resultados pode ser encontrada em Kelker (1970) ou em Fang et al. (1990).

Com estes resultados e usando o fato de que se $X \sim El_n(\mu, \Sigma)$ e $r(\Sigma) = n$, então

$$\Sigma^{-\frac{1}{2}}(X - \mu) \sim El_n(0, I_n)$$

(veja a Seção 2.5.2), Cambanis et al. (1981) notam que a densidade de X existe e assume a forma (2.17) se e somente se a variável aleatória R na representação estocástica (2.22) (com $k = n$ e $A = \Sigma^{\frac{1}{2}}$) tem a densidade dada em (2.37). Maiores detalhes destes e outros resultados adicionais podem ser encontrados em Fang et al. (1990) (veja, também, a Seção 2.5.2).

Por outro lado, se $X \stackrel{d}{=} RU^{(n)} \sim El_n(0, I_n; f)$, então usando o fato de que a densidade $f(x^\top x)$, $x \in \mathbb{R}^n$, pode ser escrita como

$$f(x^\top x) = f\left(\sum_{i=1}^k x_i^\top x_i\right), \quad x_i \in \mathbb{R}^{n_i}, \quad i = 1, \dots, k,$$

tem-se que o vetor aleatório definido por $(R_1^2, R_2^2, \dots, R_k^2)^\top \stackrel{d}{=} (\|X_1\|^2, \|X_2\|^2, \dots, \|X_k\|^2)^\top$ tem densidade dada por (veja, também, Anderson e Fang (1987))

$$p_{R_1^2, R_2^2, \dots, R_k^2}(u_1, u_2, \dots, u_k) = \frac{\pi^{\frac{1}{2}n}}{\prod_{i=1}^k \Gamma(\frac{1}{2}n_i)} \prod_{i=1}^k u_i^{\frac{1}{2}n_i-1} f\left(\sum_{i=1}^k u_i\right), \quad u_i \geq 0, \quad i = 1, 2, \dots, k, \quad (2.38)$$

de onde é fácil ver que em (2.35) $(R_1, R_2, \dots, R_k)^\top \stackrel{d}{=} (\|X_1\|, \|X_2\|, \dots, \|X_k\|)^\top$ tem densidade

$$p_{R_1, R_2, \dots, R_k}(r_1, r_2, \dots, r_k) = \frac{2^k \pi^{\frac{1}{2}n}}{\prod_{i=1}^k \Gamma(\frac{1}{2}n_i)} \prod_{i=1}^k r_i^{n_i-1} f\left(\sum_{i=1}^k r_i^2\right), \quad r_i \geq 0, \quad i = 1, 2, \dots, k. \quad (2.39)$$

Por exemplo, se $R^2 \sim \chi_n^2$, então $X \stackrel{d}{=} RU^{(n)} \sim N_n(0, I_n)$. Além disto, $\|X_1\|^2, \|X_2\|^2, \dots, \|X_k\|^2$ são independentes com $\|X_i\|^2 \sim \chi_{n_i}^2, i = 1, 2, \dots, k$ (veja os Fatos (g) e (h) na Seção 2.6).

Finalmente, suponha que $X \sim El_n(u, \Sigma; \phi)$ com $r(\Sigma) = n$ e $\phi \in \Phi_\infty$. Pela caracterização de Φ_∞ , dada em (2.19), existe uma função de distribuição F_∞ tal que

$$\phi(x) = \int_{[0, \infty)} e^{-\frac{1}{2}r^2 u} dF_\infty(r), \quad u \geq 0. \quad (2.40)$$

Como $\phi(t^\top \Sigma t), t \in \mathbb{R}^n$, é função característica de $X - \mu$, então X tem a densidade dada em (2.17), com

$$f(u) = \int_{(0, \infty)} (2\pi r^2)^{-\frac{1}{2}n} e^{-\frac{1}{2r^2}u} dF_\infty(r), \quad u \geq 0, \quad (2.41)$$

se e somente se $F_\infty(0) = 0$ (veja o Exemplo 2.4). Em caso contrário, X coloca massa $F_\infty(0)$ no ponto μ e fora de μ tem a densidade (2.17) com f dada por (2.41) (veja, também, o Exemplo 2.10).

A tabela a seguir é tomada de Fang et al. (1990, p.69) e proporciona algumas subclasses de distribuições esféricas n -dimensionais com função densidade $f(x^\top x), x \in \mathbb{R}^n$. As duas últimas subclasses nesta tabela são representadas pela função característica $\phi(t^\top t), t \in \mathbb{R}^n$.

Tabela 2.1: Subclasses de distribuições esféricas n -variadas

Tipo	$f(u) \quad (u = x^\top x, x \in \mathbb{R}^n) \quad \text{ou} \quad \phi(u) \quad (u = t^\top t, t \in \mathbb{R}^n)$
Kotz	$f(u) = cu^{N-1} \exp(-ru^s), r, s > 0, 2N + n > 2$
Normal	$f(u) = c \exp(-u/2)$
Pearson VII	$f(u) = c(1 + u/s)^{-N}, N > n/2, s > 0$
t generalizada	$f(u) = c(1 + u/s)^{-(n+m)/2}, s, m > 0$
Cauchy generalizada	$f(u) = c(1 + u/s)^{-(n+1)/2}, s > 0$
Pearson II	$f(u) = c(1 - u)^m, m > 0, 0 \leq u < 1$
Logística	$f(u) = c \exp(-u)[1 + \exp(-u)]^{-2}$
Bessel	$f(u) = c(\sqrt{u}/\beta)^a K_a(\sqrt{u}/\beta), a > -n/2, \beta > 0$, onde $K_a(\cdot)$ denota a função de Bessel modificado de terceira ordem
Normal composta	$f(u) = c \int_0^\infty v^{-n/2} \exp(-u/2v) dF(v)$, onde F é uma F.D.
Lei estável	$\phi(u) = \exp(ru^{\alpha/2}), 0 < \alpha \leq 2, r < 2$
Uniforme	$\phi(u) = {}_0F_1(n/2; -u^2/4)$, onde ${}_0F_1(\cdot; \cdot)$ é a função hipergeométrica generalizada

Nota: c é uma constante de normalização.

2.5 Propriedades adicionais

As propriedades dadas a seguir podem ser derivadas supondo que $X \sim El_n(\mu, \Sigma)$ com $r(\Sigma) = k \leq n$, excluindo o caso degenerado com $\Sigma = 0$. Contudo, alguns resultados serão estabelecidos assumindo que X tem densidade.

2.5.1 Momentos e medidas de curtose

Seja $X \sim El_n(\mu, \Sigma)$. Então, desde que existam,

$$(a) E(X) = \mu,$$

$$(b) \text{Cov}(X) = \alpha \Sigma,$$

onde $\text{Cov}(X) = E[(X - E(X))(X - E(X))^\top]$ é a matriz de covariâncias de X e $\alpha > 0$ é uma constante.

Se $X \sim El_n(\mu, \Sigma; \phi)$, então a constante α é dada por $\alpha = -2\phi'(0)$, onde $\phi'(0) = (d\phi(u)/du)|_{u=0}$. Se $X \sim N_n(\mu, \Sigma)$, então $\alpha = 1$. Se $X \sim El_n(0, I_n)$, então α é a variância

das distribuições marginais univariadas. Se $\Gamma = [\rho_{ij}]$ denota a matriz de correlações de $X = (X_1, X_2, \dots, X_n)^\top$, então

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\{\text{Var}(X_i)\text{Var}(X_j)\}^{\frac{1}{2}}} = \frac{\alpha\sigma_{ij}}{\{\alpha\sigma_{ii}\alpha\sigma_{jj}\}^{\frac{1}{2}}} = \frac{\sigma_{ij}}{\{\sigma_{ii}\sigma_{jj}\}^{\frac{1}{2}}},$$

onde $\Sigma = [\sigma_{ij}]$. Isto é, Γ independe de ϕ , de modo que para cada $\mu \in \mathbb{R}^n$ e $\Sigma(n \times n) > 0$, fixados, tem-se que todas as distribuições na classe $\{E\ell_n(\mu, \Sigma; \phi); \phi \in \Phi_n\}$ com segundo momento finito, além de ter a mesma média elas também têm a mesma matriz de correlações. Algumas caracterizações adicionais sobre a forma da matriz de covariâncias de uma distribuição elíptica são dadas em Nomakuchi e Sokata (1988).

A prova de (a) e (b) pode ser feita usando a representação (2.22). Nesta representação é fácil ver que (veja (2.73) e (2.74) na Seção 2.5.5)

$$E(U^{(k)}) = 0 \quad \text{e} \quad \text{Cov}(U^{(k)}) = E(U^{(k)}U^{(k)\top}) = k^{-1}I_k.$$

Logo, como R e $U^{(k)}$ são independentes, é claro que as componentes de $X =^d \mu + RAU^{(k)} \sim E\ell_n(\mu, \Sigma)$ têm o primeiro momento finito se e somente se $E(R) < \infty$, sendo, em tal caso, o parâmetro de locação μ a média de X . Da mesma maneira, o segundo momento das componentes de X é também finito se e somente se $E(R^2) < \infty$, sendo neste caso o parâmetro de dispersão Σ proporcional à matriz de covariâncias de X , isto é,

$$E[(X - \mu)(X - \mu)^\top] = E[R^2 AU^{(k)}U^{(k)\top} A^\top] = \alpha\Sigma,$$

com $\alpha = k^{-1}E(R^2)$. □

Kelker (1970) nota que se $X \sim E\ell_n(\mu, \Sigma; f)$, isto é, se X tem densidade dada por (2.17), então as componentes de X têm k momentos finitos se e somente se

$$\int_0^\infty r^{k+n-1} f(r^2) dr < \infty.$$

Para $k \geq 1$, vale (a) e se $k \geq 2$, (b) também vale. No caso esférico, com $X \sim E\ell_n(0, I_n; f)$, é fácil ver que (veja (2.76) na Seção 2.5.5) os momentos de X , desde que existam, são

dados por

$$E\left(\prod_{i=1}^n X_i^{k_i}\right) = \begin{cases} \frac{\prod_{i=1}^n [(k_i-1)(k_i-3)\dots 1]}{n(n+2)\dots(n+k-2)} E(\|X\|^k), & \text{se } k_i \text{ é par, } i = 1, 2, \dots, n \\ 0, & \text{se } k_i \text{ é ímpar para algum } i = 1, 2, \dots, n, \end{cases} \quad (2.42)$$

onde $k = k_1 + k_2 + \dots + k_n$, com $k_1, k_2, \dots, k_n$ sendo inteiros não-negativos.

Considerando o caso univariado com $X \sim El(0, \sigma^2; \phi)$, Berkane e Bentler (1986) mostram que se a função característica de X pode ser expandida como

$$\phi_X(t) = \sum_{k=0}^{\infty} i^k \mu_k \frac{t^k}{k!} = \sum_{k=0}^{\infty} \phi_X^{(k)}(0) \frac{t^k}{k!},$$

onde $\mu_k = E(X^k)$ e $\phi_X^{(k)}(0)$ denota a k -ésima derivada de $\phi_X(t)$ avaliada em $t = 0$. Então $\mu_k = 0$ para k sendo ímpar e para $k = 2m$, $m = 1, 2, \dots$, vale a relação

$$\mu_{2m} = \frac{(2m)!}{2^m m!} (\kappa(m) + 1) \mu_2^m, \quad (2.43)$$

onde

$$\kappa(m) + 1 = \frac{\phi^{(m)}(0)}{[\phi'(0)]^m}. \quad (2.44)$$

Os coeficientes $\kappa(m)$, $m = 1, 2, \dots$ são chamados parâmetros de momentos e generalizam o parâmetro de curtose $\kappa = \kappa(2)$ definido em Muirhead (1982) para uma distribuição elíptica.

Se $X \sim N(0, 1)$, tem-se que $\mu_2 = \sigma^2 = 1$ e (2.42) se reduz a

$$\mu_{2m} = \frac{(2m)!}{2^m m!}.$$

Consequentemente, $\kappa(m) = 0$, $m = 1, 2, \dots$

Além disto, Berkane e Bentler (1986) estabelecem que se uma variável aleatória Y tem densidade $p_Y(y)$, $y \in \mathbb{R}$, com momentos ímpares nulos e com o $(2m)$ -ésimo momento finito e da forma $\mu_{2m} = c_m(\mu_2)^m$, $m = 1, 2, \dots$, para alguma constante c_m que dependa de m , então Y tem distribuição elíptica. Estes fatos são estendidos para o caso n -dimensional

em Berkane e Bentler (1987), usando condições similares sobre a função característica. O coeficiente de curtose multivariado é definido em Mardia (1970) como

$$\beta_{2,n} = E\{[(X - \mu_X)^\top \Sigma_X^{-1}(X - \mu_X)]^2\} ,$$

onde $\mu_X = E(X)$ e $\Sigma_X = \text{Cov}(X)$, é generalizado como

$$\beta_{m,n} = E\{[(X - \mu_X)^\top \Sigma_X^{-1}(X - \mu_X)]^m\} .$$

Se $X = (X_1, X_2, \dots, X_n)^\top \sim E\ell_n(0, \Sigma; \phi)$ e denotando por

$$\mu_{i_1, i_2, \dots, i_{2m}} = E(X_{i_1} X_{i_2} \cdots X_{i_{2m}}) ,$$

então as seguintes relações são obtidas:

$$\frac{\beta_{m,n}}{n(n+2) \cdots (n+2k) \cdots (n+2(m-1))} = \kappa(m) + 1 , \quad (2.45)$$

e

$$\mu_{i_1 i_2 \cdots i_{2m}} = (\kappa(m) + 1) \sum \mu_{i_1 i_2} \cdots \mu_{i_{2m-1} i_{2m}} , \quad (2.46)$$

onde $\kappa(m)$ é definido em termos da função característica como em (2.44). Em (2.46) a soma estende-se sobre as $(2m)!/2^m m!$ maneiras de reagrupar os subscritos. Por exemplo, para $m = 2$, tem-se que

$$\frac{\beta_{2,n}}{n(n+2)} = \kappa(2) + 1 \quad \text{e} \quad \mu_{ijkl} = (\kappa(2) + 1)(\mu_{ij}\mu_{kl} + \mu_{il}\mu_{jk} + \mu_{ik}\mu_{jl}) .$$

Desta última relação, é fácil ver que para os modelos considerados nos Exemplos 2.8 e 2.9

$$\kappa(2) + 1 = \frac{1 + \varepsilon(\sigma^4 - 1)}{[1 - \varepsilon(\sigma^2 - 1)]^2} \quad \text{e} \quad \kappa(2) + 1 = \frac{\nu - 2}{\nu - 4} , \quad \nu > 4 ,$$

respectivamente.

2.5.2 Transformações lineares e distribuições marginais

Seja $X \sim El_n(\mu, \Sigma; \phi)$ não degenerada (isto é, $\Sigma \neq 0$) e com representação estocástica dada em (2.22). Então, existe uma constante $c > 0$ tal que:

(i) se $X \sim El_n(\mu^*, \Sigma^*; \phi^*)$, tem-se que

$$\mu^* = \mu, \Sigma^* = c\Sigma \text{ e } \phi^*(u) = \phi(c^{-1}u), \quad u \geq 0;$$

(ii) se $X = {}^d \mu^* + R^* A^* U^{(k^*)}$, com $k \geq k^*$, tem-se que

$$\mu^* = \mu, A^* A^{*\top} = c A A^\top \text{ e } R^* = {}^d c^{-1} R \sqrt{V},$$

onde $V \sim \text{Beta}(k^*/2, (k - k^*)/2)$ se $k > k^*$ e $V \equiv 1$ se $k = k^*$.

Estes resultados são provados em Cambanis et al. (1981) e mostram que Σ , A , ϕ e R não são únicos a menos que a condição $|\Sigma| = 1$ (ou $|A A^\top| = 1$) seja imposta.

A próxima propriedade estabelece que a distribuição de qualquer transformação linear de um vetor aleatório com distribuição elíptica é também elíptica.

Seja $X \sim El_n(\mu, \Sigma; \phi)$. Se $\eta(m \times 1) \in \mathbb{R}^m$ e A é uma matriz $m \times n$, então

$$Y = \eta + AX \sim El_m(\eta + A\mu, A\Sigma A^\top; \phi). \quad (2.47)$$

A prova deste resultado é direta, desde que a função característica de Y é dada por

$$\phi_Y(s) = e^{is^\top \eta} \phi_X(A^\top s), \quad s \in \mathbb{R}^m.$$

Usando (2.8), tem-se que

$$\phi_Y(s) = e^{is^\top (\eta + A\mu)} \phi(s^\top (A\Sigma A^\top)s), \quad s \in \mathbb{R}^m,$$

e (2.47) segue da Definição 2.2. □

Em particular, se $X \sim El_n(\mu, \Sigma; \phi)$, com $\Sigma = A A^\top$, $A(n \times k)$ e $r(A) = k$, então

$$Z = A^-(X - \mu) \sim El_k(0, I_k; \phi), \quad (2.48)$$

onde $A^- = (A^\top A)^{-1} A^\top$ ($A^- = \Sigma^{-\frac{1}{2}}$ para $k = n$).

Por outro lado, usando (2.47), é fácil ver que se $X \sim El_n(\mu, \Sigma; \phi)$, então todas as distribuições marginais são elípticas. Por exemplo, particionando X, μ e Σ como

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad (2.49)$$

onde X_1 e μ_1 são vetores $m \times 1$ e Σ_{11} é uma matriz $m \times m$, $1 \leq m < n$. Então

$$X_1 \sim El_m(\mu_1, \Sigma_{11}; \phi),$$

com função característica

$$\phi_{X_1}(t_1) = e^{it_1^\top \mu_1} \phi(t_1^\top \Sigma_{11} t_1), \quad t_1 \in \mathbb{R}^m.$$

De fato, colocando em (2.47) $A = [I_m, 0]$ e $\eta = 0$, tem-se a prova. Uma forma mais direta é derivar $\phi_{X_1}(t_1)$ colocando $t = (t_1^\top, 0^\top)^\top$ em (2.14). A distribuição de X_2 é obtida por simetria.

No caso em que $X \sim El_n(\mu, \Sigma; \phi)$ tem densidade, isto é, $X \sim El_n(\mu, \Sigma; f)$, então todas as densidades marginais de dimensão $m < n$ não diferem de sua forma funcional. Por exemplo, se na partição dada em (2.49) se definem

$$\mu_2(x_1) = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \quad \text{e} \quad \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}, \quad (2.50)$$

sabe-se que (veja, por exemplo, Muirhead (1982, p.581)),

$$|\Sigma| = |\Sigma_{11}| \cdot |\Sigma_{22.1}|$$

e

$$(x - \mu)^\top \Sigma^{-1} (x - \mu) = (x_1 - \mu_1)^\top \Sigma_{11}^{-1} (x_1 - \mu_1) + (x_2 - \mu_2(x_1))^\top \Sigma_{22.1}^{-1} (x_2 - \mu_2(x_1)).$$

Reescrevendo (2.17) em termos destas últimas identidades e integrando com relação a $x_2 \in \mathbb{R}^{n-m}$, tem-se a densidade marginal de X_1 , isto é,

$$\begin{aligned}
p_{X_1}(x_1|\mu_1, \Sigma_{11}) &= \int_{\mathbb{R}^{n-m}} |\Sigma_{11}|^{-\frac{1}{2}} |\Sigma_{22.1}|^{-\frac{1}{2}} f\left((x_1 - \mu_1)^\top \Sigma_{11}^{-1} (x_1 - \mu_1) \right. \\
&\quad \left. + (x_2 - \mu_2(x_1))^\top \Sigma_{22.1}^{-1} (x_2 - \mu_2(x_1))\right) dx_2 \\
&= |\Sigma_{11}|^{-\frac{1}{2}} \int_{\mathbb{R}^{n-m}} f\left((x_1 - \mu_1)^\top \Sigma_{11}^{-1} (x_1 - \mu_1) + y_2^\top y_2\right) dy_2 \\
&\equiv |\Sigma_{11}|^{-\frac{1}{2}} f_1((x_1 - \mu_1)^\top \Sigma_{11}^{-1} (x_1 - \mu_1)) , \quad x_1 \in \mathbb{R}^m , \quad (2.51)
\end{aligned}$$

onde

$$f_1(u) = s_{(n-m)} \int_0^\infty f(u + r^2) r^{n-m-1} dr , \quad u \geq 0 . \quad (2.52)$$

Consequentemente, a função f_1 é determinada apenas pela forma da função f e pela dimensão do vetor X . A densidade marginal de X_2 segue por simetria. Além disto, se $X_1 \sim El_m(0, I_m; \phi)$, então do Fato 2.4 tem-se que $X_1 =^d R_1 U^{(m)} =^d R \sqrt{V_1} U^{(m)}$ tem densidade se e somente se $P(R = 0) = 0$. Neste caso, a densidade de R_1 é dada por (2.32), de onde segue-se que a densidade de X_1 é da forma $f_1(x_1^\top x_1)$, $x_1 \in \mathbb{R}^m$, com

$$f_1(u) = s_n^{-1} s_{(n-m)} \int_u^\infty r^{2-n} (r^2 - u)^{\frac{1}{2}(n-m)-1} dF(r) , \quad u \geq 0 . \quad (2.53)$$

Quando R tem densidade, (2.53) se reduz a (2.52).

2.5.3 Distribuições condicionais

Suponha, inicialmente, que $X \sim El_n(\mu, \Sigma; f)$, isto é, X tem a densidade (2.17), e considere a partição dada em (2.49). Então,

(a) a distribuição condicional de X_1 , dado $X_2 = x_2$, é elíptica m -variada, a saber,

$$(X_1|X_2 = x_2) \sim El_m(\mu_1(x_2), \Sigma_{11.2}; f_{q(x_2)}) , \quad (2.54)$$

onde f_a é definida em (2.55) (abaixo);

(b) assumindo que existem

$$E(X_1|X_2 = x_2) = \mu_1(x_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) ,$$

e

$$\text{Cov}(X_1|X_2 = x_2) = \alpha_{q(x_2)} \Sigma_{11.2} = \alpha_{q(x_2)} (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}) ,$$

onde, como é indicado na notação utilizada acima, $f_{q(x_2)}$ e $\alpha_{q(x_2)}$ são funções de x_2 através da forma quadrática $q(x_2) = (x_2 - \mu_2)^\top \Sigma_{22}^{-1} (x_2 - \mu_2)$.

A prova de (a) segue de observar que (2.17) pode ser escrita como

$$p(x_1, x_2 | \mu, \Sigma) = |\Sigma_{22}|^{-\frac{1}{2}} |\Sigma_{11.2}|^{-\frac{1}{2}} f((x_1 - \mu_1(x_2))^\top \Sigma_{11.2}^{-1} (x_1 - \mu_1(x_2)) + q(x_2)) ,$$

de onde se tem que a densidade condicional de X_1 , dado $X_2 = x_2$, é dada por

$$\begin{aligned} p_{X_1}(x_1|x_2; \mu, \Sigma) &= |\Sigma_{11.2}|^{-\frac{1}{2}} \frac{f((x_1 - \mu_1(x_2))^\top \Sigma_{11.2}^{-1} (x_1 - \mu_1(x_2)) + q(x_2))}{s_m \int_0^\infty f(r^2 + q(x_2)) r^{m-1} dr} \\ &\equiv |\Sigma_{11.2}|^{-\frac{1}{2}} f_{q(x_2)}((x_1 - \mu_1(x_2))^\top \Sigma_{11.2}^{-1} (x_1 - \mu_1(x_2))), \quad x_1 \in \mathbb{R}^m, \end{aligned} \quad (2.55)$$

onde para cada $a \geq 0$,

$$f_a(u) = \frac{f(u + a)}{s_m \int_0^\infty f(r^2 + a) r^{m-1} dr} , \quad u \geq 0 ,$$

é uma função não-negativa. Isto conclui a prova de (a). $\square$

Note-se que para cada $a \geq 0$, $f_a(x^\top x)$, $x \in \mathbb{R}^m$, é a densidade de uma distribuição esférica m -variada. Além disso, se $\phi_{X_1}(t_1|x_2)$, $t_1 \in \mathbb{R}^m$, denota a função característica de $(X_1|X_2 = x_2)$, então

$$\phi_{X_1}(t_1|x_2) = E(e^{it_1^\top X_1} | X_2 = x_2) = e^{it_1^\top \mu_1(x_2)} E(e^{it_1^\top (X_1 - \mu_1(x_2))} | X_2 = x_2) .$$

Mas, usando (2.55), tem-se que

$$E(e^{it_1^\top (X_1 - \mu_1(x_2))} | X_2 = x_2) = \frac{\int_{\mathbb{R}^m} e^{i(\Sigma_{11.2}^{-\frac{1}{2}} t_1)^\top z_1} f(z_1^\top z_1 + q(x_2)) dz_1}{s_m \int_0^\infty f(r^2 + q(x_2)) r^{m-1} dr} = \phi_{q(x_2)}(t_1^\top \Sigma_{11.2} t_1) ,$$

e, conseqüentemente,

$$\Phi_{X_1}(t_1|X_2) = e^{it_1^\top \mu_1(x_2)} \phi_{q(x_2)}(t_1^\top \Sigma_{11.2} t_1) , \quad t_1 \in \mathbb{R}^m . \quad (2.56)$$

Para cada $a \geq 0$, tem-se que $\phi_a(t^\top t)$, $t \in \mathbb{R}^m$, é a função característica de uma distribuição esférica m -variada.

A prova de (b) pode ser construída de forma análoga ao caso normal e pode ser encontrada, por exemplo, em Kelker (1970).

Em (b), observe-se que $\alpha_{q(x_2)} = -2\phi'_{q(x_2)}(0)$. Se $X \sim N_n(\mu, \Sigma)$, então $\phi_{q(x_2)}(x) = e^{-\frac{1}{2}u}$, $u \geq 0$, e, conseqüentemente, $\alpha_{q(x_2)} = 1$, qualquer que seja o valor de x_2 . Além disto, se $X \sim El_n(0, I_n; \phi)$, então $q(x_2) = \|x_2\|^2$ e $\alpha_{\|x_2\|^2}$ corresponde à variância das distribuições marginais univariadas de $(X_1|X_2 = x_2)$.

A derivação da distribuição condicional de X_1 , dado $X_2 = x_2$, no caso geral em que $X \sim El_n(\mu, \Sigma; \phi)$, com $r(\Sigma) = k$, será feita seguindo Cambanis et al. (1981). O seguinte fato é requerido para tais efeitos, cuja prova é direta e, portanto, será omitida.

Fato 2.5 *Sejam R e S variáveis aleatórias não-negativas e independentes, com R tendo função de distribuição F e S sendo absolutamente contínua com densidade p_S . Então, se $F(0) > 0$, a distribuição da variável aleatória $T = RS$ coloca massa $F(0)$ em zero e sobre $(0, \infty)$ é absolutamente contínua com densidade*

$$p_T(t) = \int_{(0, \infty)} r^{-1} p_S(t/r) dF(r) .$$

Além disso, uma versão regular da distribuição condicional de R , dado $T = t$, é dada por

$$P(R \leq \rho | T = t) = \begin{cases} 0 , & \text{para } \rho < 0; \\ 1 , & \text{para } t=0 \text{ ou } t>0, \text{ com } p_T(t)=0, \rho \geq 0; \\ \frac{1}{p_T(t)} \int_{(0, \rho]} r^{-1} p_S(t/r) dF(r) , & \text{para } t>0 \text{ com } p_T(t) \neq 0, \rho \geq 0. \end{cases} \quad (2.57)$$

Seja, agora, $X = RU^{(n)} \sim El_n(0, I_n; \phi)$ e $X = (X_1^\top, X_2^\top)^\top$, onde X_1 é $m \times 1$, $1 \leq m < n$. Então

$$(X_1|X_2 = x_2) \sim El_m(0, I_m; \phi_{\|x_2\|^2}) , \quad (2.58)$$

com representação estocástica

$$(X_1|X_2 = x_2) =^d R_{\|x_2\|^2} U^{(m)} , \quad (2.59)$$

onde, para cada $a \geq 0$, R_a é independente de $U^{(m)}$ e sua distribuição é apresentada em (2.62) (abaixo). Além disto,

$$R_{\|x_2\|^2} =^d \left(\sqrt{R^2 - \|x_2\|^2} | X_2 = x_2 \right) , \quad (2.60)$$

e a função ϕ_a é dada por (2.20) com $k = m$ e substituindo F por F_a , a função distribuição de R_a .

Com efeito, usando (2.26) (veja o Fato 2.4), tem-se que

$$(X_1|X_2 = x_2) =^d \left(R\sqrt{V_1}U^{(m)} \mid R\sqrt{1-V_1}U^{(n-m)} = x_2 \right) ,$$

onde R , V_1 , $U^{(m)}$ e $U^{(n-m)}$ são independentes. Notando que $R\sqrt{V_1} = \sqrt{R^2 - \|x_2\|^2}$ quando $R\sqrt{1-V_1}U^{(n-m)} = x_2$, obtém-se (2.59) e (2.60). O fato de que $R_{\|x_2\|^2}$ depende de x_2 somente através de $\|x_2\|^2$ (como é definido em (2.60) e indicado na notação) é óbvio quando $x_2 = 0$; para $x_2 \neq 0$, segue-se de observar que

$$\begin{aligned} \left(\sqrt{R^2 - \|x_2\|^2} \mid X_2 = x_2 \right) &=^d \left(\sqrt{R^2 - \|x_2\|^2} \mid R\sqrt{1-V_1}U^{(n-m)} = x_2 \right) \\ &=^d \left(\sqrt{R^2 - \|x_2\|^2} \mid R^2(1-V_1) = \|x_2\|^2 \right) , \end{aligned} \quad (2.61)$$

desde que $U^{(n-m)}$ é independente de R e V_1 . Considerando (2.59), (2.58) é consequência do Fato 2.1 (mais precisamente, de (2.24)), com $\phi_{\|x\|^2}$ definida da maneira já descrita. Aplicando o Fato 2.5 para $R_2 = R\sqrt{1-V_1}$, com $R \sim F$ e $S = \sqrt{1-V_1}$ tendo densidade dada por (veja (2.29))

$$p_S(s) = \frac{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}(n-m))}{\Gamma(\frac{1}{2}n)} s^{n-m-1} (1-s^2)^{\frac{1}{2}m-1} , \quad 0 < s < 1 ,$$

e usando (2.60) e (2.61), obtém-se finalmente que

$$R_a = 0 \quad \text{com probabilidade 1 quando } a = 0 \text{ ou } F(\sqrt{a}) = 1 ,$$

e

$$P(R_a \leq \rho) = \frac{\int_{(\sqrt{a}, \sqrt{\rho^2 + a}]} (R^2 - a)^{\frac{1}{2}m-1} r^{2-n} dF(r)}{\int_{(\sqrt{a}, \infty)} (r^2 - a)^{\frac{1}{2}m-1} r^{2-n} dF(r)}, \quad \rho \geq 0, \quad a > 0 \text{ e } F(\sqrt{a}) < 1. \quad (2.62)$$

□

Considere, finalmente, o caso geral em que $X \stackrel{d}{=} \mu + RAU^{(k)} \sim El_n(\mu, \Sigma; \phi)$, com $AA^\top = \Sigma$, $r(A) = r(\Sigma) = k \geq 1$, onde X , μ e Σ são particionados como em (2.49). Suponha que $k_2 = r(\Sigma_{22}) \geq 1$ e $k_1 = k - k_2 \geq 1$ e denote por M_2 o espaço gerado pelas colunas de Σ_{22} . Então, uma versão regular da distribuição condicional de X_1 , dado $X_2 = x_2$, é dada por

$$(X_1|X_2 = x_2) \sim El_m(\mu_1^*(x_2), \Sigma_{11.2}^*; \phi_{q^*(x_2)}), \quad \text{para } x_2 \in \mu_2 + M_2, \quad (2.63)$$

e

$$(X_1|X_2 = x_2) \stackrel{d}{=} \mu_1, \quad \text{para } x_2 \notin \mu_2 + M_2, \quad (2.64)$$

com representação estocástica

$$(X_1|X_2 = x_2) \stackrel{d}{=} \mu_1^*(x_2) + R_{q^*(x_2)} A_1 U^{(k_1)}, \quad \text{para } x_2 \in \mu_2 + M_2, \quad (2.65)$$

onde

$$\begin{aligned} \mu_1^*(x_2) &= \mu_1 + \Sigma_{12} \Sigma_{22}^- (x_2 - \mu_2), \quad \Sigma_{11.2}^* = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^- \Sigma_{21}, \\ q^*(x_2) &= (x_2 - \mu_2)^\top \Sigma_{22}^- (x_2 - \mu_2), \end{aligned}$$

e $\Sigma_{11.2}^* = A_1 A_1^\top$, com $A_1(m \times k_1)$ e $r(A_1) = k_1$. Além disto, para cada $a \geq 0$, R_a é independente de $U^{(k_1)}$ e sua distribuição é dada por (2.62) com $n = k$ e $m = k_1$;

$$R_{q^*(x_2)} \stackrel{d}{=} (\sqrt{R^2 - q^*(x_2)} | X_2 = x_2), \quad \text{para } x_2 \in \mu_2 + M_2,$$

e a função ϕ_a é dada por (2.20) com $k = k_1$ e substituindo F pela distribuição de R_a (veja (2.62)).

A demonstração do resultado anterior segue de escrever $RU^{(k)} = (Z_1^\top, Z_2^\top)^\top$, onde Z_1 é um vetor $k_1 \times 1$ e expressar $\Sigma_{22} = A_2 A_2^\top$, com $A_2((n - m) \times k_2)$ e $r(A_2) = k_2$. Usando as respectivas funções características é fácil verificar que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} =^d \begin{pmatrix} \mu_1 + A_1 Z_1 + \Sigma_{12} \Sigma_{22}^{-1} A_2 Z_2 \\ \mu_2 + A_2 Z_2 \end{pmatrix}.$$

Daqui tem-se que

$$(X_1 | X_2 = x_2) =^d \mu_1(x_2) + (A_1 Z_1 | A_2 Z_2 = x_2 - \mu_2),$$

de onde segue o resultado desejado, já que

$$(Z_1 | Z_2 = z_2) \sim E\ell_{k_1}(0, I_{k_1}; \phi_{||z_2||^2}).$$

□

Note-se que (2.63) é a representação de maior interesse para $(X_1 | X_2 = x_2)$, desde que $X_2 \in \mu_2 + M_2$, com probabilidade 1.

Os casos triviais com $k_1 = 0$ e $k_2 = 0$ foram excluídos, já que se $k_2 = 0$, então $X_2 = \mu_2$, com probabilidade 1, e se $k_1 = 0$, então $X_1 = \mu_1^*(X_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2)$, com probabilidade 1.

Por outro lado, usando (2.65), segue-se que $\text{Cov}(X_1 | X_2)$ existe se e somente se $E(R_a^2) < \infty$. Neste caso, tem-se que

$$\text{Cov}(X_1 | X_2 = x_2) = \alpha_{q^*(x_2)} \Sigma_{11.2}^*,$$

onde $\alpha_{q^*(x_2)} = k_1^{-1} E(R_{q^*(x_2)}^2)$.

Usando (2.62) pode-se mostrar que $E(R_a^2) < \infty$ se $k_1 \leq k - 2$ ou se $E(R) < \infty$ e $k_1 \leq k - 1$ ou, obviamente, se $E(R^2) < \infty$. Finalmente, quando Σ_{22} é não-singular ($k_2 = n - m$), então

$$\Sigma_{22}^{-1} = \Sigma_{22}^{-1}, \quad \mu_1^*(x_2) = \mu_1(x_2), \quad \Sigma_{11.2}^* = \Sigma_{11.2} \quad \text{e} \quad q^*(x_2) = q(x_2),$$

o que sucede, por exemplo, se Σ é não-singular ($k = n$). Neste caso, tem-se que

$$(X_1 | X_2 = x_2) =^d \mu_1(x_2) + R_{q(x_2)} \Sigma_{11.2}^{-\frac{1}{2}} U^{(m)} \sim E\ell_m(\mu_1(x_2), \Sigma_{11.2}; \phi_{q(x_2)}). \quad (2.66)$$

2.5.4 Formas quadráticas e outras funções relacionadas com distribuições esféricas

No que segue, é suposto que $X \sim El_n(0, I_n)$ e $P(X = 0) = 0$ para derivar as distribuições e outras propriedades de algumas funções de interesse do vetor X . Neste sentido, Kelker (1970) obtém as distribuições de $\|X\|$, $\|X\|^2$ e $(\|X_1\|, \|X_2\|)^T$ quando $X = (X_1^T, X_2^T)^T \sim El_n(0, I_n)$ tem densidade. Ele também estende o teorema de Cochran ao caso elíptico assumindo a existência de densidade com o quarto momento finito. Kariya e Eaton (1977) derivam as distribuições das função $a^T X / \|X\|$, $a^T \in \mathbb{R}^n$, com $\|a\| = 1$ e $X^T A X / \|X\|^2$, com $A(n \times n)$ sendo simétrica. Estes e outros resultados são resumidos em Muirhead (1982). Entretanto, Anderson e Fang (1986), seguindo Cambanis et al. (1981), conseguem estender o teorema de Cochran para o caso elíptico com menos restrições que Kelker (1970). Alguns resultados desta natureza foram dados nos Fatos 2.3 e 2.4 (veja, também, (2.33) a (2.39)). Estes e outros resultados também podem ser vistos em Fang et al. (1990).

Suponha que $X \sim El_n(0, I_n)$ e $P(X = 0) = 0$. Então,

(a) se $A(n \times n)$ é uma matriz simétrica com valores próprios $\lambda_1, \lambda_2, \dots, \lambda_n$, tem-se que

$$\frac{X^T A X}{\|X\|^2} =^d \sum_{i=1}^n \lambda_i V_i, \quad (2.67)$$

onde $(V_1, V_2, \dots, V_{n-1})^T \sim D_{n-1}(1/2, 1/2, \dots, 1/2)$ (veja (2.28)). Em particular, se $A^2 = A$ e $r(A) = m$, então

$$\frac{X^T A X}{\|X\|^2} \sim \text{Beta} \left(\frac{1}{2}m, \frac{1}{2}(n-m) \right); \quad (2.68)$$

(b) se $W = a^T X / \|X\|$, $a(n \times 1) \in \mathbb{R}^n$ com $\|a\| = 1$, tem-se que

$$\frac{\sqrt{n-1}W}{\sqrt{1-W^2}} \sim t_{n-1},$$

onde $t_\nu \equiv t(0, 1; \nu)$ denota a distribuição t (de Student) univariada padrão com ν graus de liberdade.

Com efeito, como $A(n \times n)$ é simétrica, então existe uma matriz ortogonal $\Gamma(n \times n)$ tal que

$$\Gamma^T A \Gamma = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) .$$

Usando que $\Gamma X =^d X$, tem-se que

$$\frac{X^T A X}{\|X\|^2} =^d \frac{X^T \Lambda X}{\|X\|^2} =^d U^{(n)T} \Lambda U^{(n)} = \sum_{i=1}^n \lambda_i U_i^2 ,$$

já que $U^{(n)} = (U_1, U_2, \dots, U_n)^T$ e $X/\|X\|$ tem a mesma distribuição (veja (2.26)). Observando que

$$(U_1^2, U_2^2, \dots, U_n^2)^T =^d (V_1, V_2, \dots, V_n)^T ,$$

onde $(V_1, V_2, \dots, V_{n-1})^T \sim D_{n-1}(1/2, 1/2, \dots, 1/2; 1/2)$ (veja (2.6) e (2.8) no Exemplo 2.1), tem-se (2.67).

Se A é idempotente com $r(A) = m$, então existe $\Gamma(n \times n)$ ortogonal tal que

$$\Gamma^T A \Gamma = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} .$$

Escrevendo $X = (X_1^T, X_2^T)^T$, com X_1 sendo $m \times 1$, tem-se que

$$\frac{X^T A X}{\|X\|^2} =^d \frac{\|X_1\|^2}{\|X\|^2} =^d \sum_{i=1}^m V_i \sim \text{Beta} \left(\frac{1}{2}m, \frac{1}{2}(n-m) \right) ,$$

o que prova (2.68). □

Para uma prova alternativa de (a), basta supor que $X \sim N_n(0, I_n)$ e usar (2.27).

A prova de (b) pode ser feita, sem perda de generalidade, tomando $a^T = (1, 0, \dots, 0)^T$ e assumindo que $X \sim N_n(0, I_n)$. Neste caso, tem-se que

$$\frac{\sqrt{n-1}W}{\sqrt{1-W^2}} = \frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^n X_i^2}} \sim t_{n-1} .$$

□

Os exemplos dados a seguir mostram que dentro da classe das distribuições esféricas, a suposição de normalidade não é importante na derivação da distribuição de algumas estatísticas usuais.

Exemplo 2.11 (A estatística t). Suponha que $X = (X_1, X_2, \dots, X_n)^\top \sim El_n(0, I_n)$ e $P(X = 0) = 0$. Então (Efrom (1969), Kariya e Eaton (1977)),

$$t = \frac{\sqrt{n}\bar{X}_n}{S_n} \sim t_{n-1} ,$$

onde $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

De fato, colocando $a = n^{-\frac{1}{2}} \mathbf{1}_n$, onde $\mathbf{1}_n = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$, a prova deste resultado vem diretamente de (b). $\square$

Exemplo 2.12 (O coeficiente de correlação). Suponha que $X = (X_1, X_2, \dots, X_n)^\top \sim El_n(0, I_n)$ e $P(X = 0) = 0$. Seja $Y = (Y_1, Y_2, \dots, Y_n)^\top$ com distribuição tal que

$$P(Y = k\mathbf{1}_n , \text{ para algum } k) = 0 .$$

Se, além disso, X e Y são independentes, tem-se que (Kariya e Eaton (1977)),

$$\frac{\sqrt{n-2}r_{XY}}{\sqrt{1-r_{XY}^2}} \sim t_{n-2} ,$$

onde

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\left\{ \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \right\}^{\frac{1}{2}}} ,$$

é a correlação amostral de $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.

A prova do Exemplo 2.12 segue de observar que

$$r_{XY} = \frac{Y^\top M X}{\{(Y^\top M Y)(X^\top M X)\}^{1/2}} ,$$

onde $M = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$ é uma matriz idempotente com $r(M) = \text{tr}(M) = n - 1$. Seja $\Gamma(n \times n)$ uma matriz ortogonal tal que

$$M = \Gamma^\top \Lambda \Gamma , \text{ com } \Lambda = \begin{pmatrix} I_{n-1} & 0 \\ 0 & 0 \end{pmatrix} .$$

Escrevendo $\widetilde{X} = (X_1, X_2, \dots, X_{n-1})^\top$ e $\widetilde{Z} = (Z_1, Z_2, \dots, Z_{n-1})^\top$, onde $Z = \Gamma Y$, tem-se que

$$r_{XY} =^d \frac{Z^\top \Lambda X}{\{(Z^\top \Lambda Z)(X^\top \Lambda X)\}^{1/2}} = \frac{\widetilde{Z}^\top \widetilde{X}^\top}{\|\widetilde{Z}\| \|\widetilde{X}\|}.$$

Considerando que $\widetilde{X} \sim E\ell_{n-1}(0, I_{n-1})$, então, condicionando em $\widetilde{Z}$ e aplicando (b), tem-se a prova. $\square$

A distribuição de r_{YX} no caso em que (X_i, Y_i) , $i = 1, 2, \dots, n$ têm distribuição conjunta elíptica é derivada em Ali e Joarder (1994).

Exemplo 2.13 (A estatística F). Assuma que $X \sim E\ell_n(0, I_n)$ e $P(X = 0) = 0$. Sejam A e B duas matrizes $n \times n$ simétricas, idempotentes e tais que $AB = 0$, $r(A) = a$, $r(B) = b$ e $a + b \leq n$. Então,

$$F = \frac{X^\top AX/a}{X^\top BX/b} \sim F_{a,b},$$

onde $F_{a,b}$ denota a distribuição F (de Fisher) com a e b graus de liberdade.

Com efeito, usando o Fato 2.3 junto com (2.27), tem-se que

$$F =^d \frac{U^{(n)} A U^{(n)} / a}{U^{(n)} B U^{(n)} / b} =^d \frac{Z^\top A Z / a}{Z^\top B Z / b},$$

onde $Z \sim N_n(0, I_n)$. As características das matrizes A e B implicam que $Z^\top A Z \sim \chi_a^2$ e $Z^\top B Z \sim \chi_b^2$ são independentes. Isto prova o resultado desejado. $\square$

Para uma prova alternativa do Exemplo 2.13, veja (2.69) (abaixo).

O resultado dado a seguir é provado em Anderson e Fang (1987) e constitui uma adaptação do teorema de Cochran ao caso esférico.

Suponha que $X =^d R U^{(n)} \sim E\ell_n(0, I_n; \phi)$, $P(X = 0) = 0$ e $(V_1, V_2, \dots, V_k)^\top \sim D_k(m_1/2, m_2/2, \dots, m_k/2; m_{k+1}/2)$, onde R , $U^{(n)}$ e $(V_1, V_2, \dots, V_k)^\top$ são independentes. Então, se $A_1, A_2, \dots, A_k$ são matrizes simétricas, tem-se que

$$(X^\top A_1 X, X^\top A_2 X, \dots, X^\top A_k X)^\top =^d R^2 (V_1, V_2, \dots, V_k)^\top, \quad (2.69)$$

se e somente se $A_i A_j = \delta_{ij} A_i$ e $r(A_i) = m_i$, onde $\delta_{ii} = 1$ e $\delta_{ij} = 0$, $i \neq j$.

Em particular, se $X =^d RU^{(n)} \sim El_n(0, I_n; \phi)$, $P(X = 0) = 0$ e $V_1 \sim \text{Beta}(m/2, (n - m)/2)$, onde V_1 , R e $U^{(n)}$ são independentes, então,

$$X^\top AX =^d R^2 V_1, \quad (2.70)$$

se e somente se $A^2 = A$ e $r(A) = m$ ($1 \leq m < n$).

Note-se que a parte “se” deste resultado é consequência direta de (2.68).

Contudo, se $X \sim El_n(0, I_n)$ e A e B são matrizes $n \times n$ simétricas e idempotentes com $AB = 0$, então $X^\top AX$ e $X^\top BX$ não são independentes, a menos que $X \sim N_n(0, cI_n)$, para algum $c > 0$ (veja os Fatos (g) e (h) na Seção 2.6).

Por outro lado, se $X =^d RU^{(n)} \sim El_n(0, I_n; \phi)$ com $P(X = 0) = 0$ e se $A(n \times n)$ é uma matriz simétrica e idempotente com $r(A) = m$ e $E(R^{2k}) < \infty$, então de (2.70) segue que

$$E[(X^\top AX)^k] = E(V_1^k)E(R^{2k}) = \frac{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}m + k)}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n + k)} E(\|X\|^{2k}), \quad (2.71)$$

uma vez que $R =^d \|X\|$.

Resultados adicionais deste tipo serão obtidos na Seção 2.5.5, os quais serão utilizados nos capítulos posteriores. A derivação destes resultados será feita inicialmente considerando a distribuição normal esférica e colocando (por conveniência) $n = p$.

Suponha que $Z = (Z_1, Z_2, \dots, Z_p)^\top \sim N_p(0, I_p)$ e sejam

$$R = \|Z\| \quad \text{e} \quad U^{(p)} = U = \frac{Z}{\|Z\|}.$$

Então (Anderson e Stephens (1972)),

$$E(R^m) = E[(R^2)^{\frac{1}{2}m}] = \frac{\Gamma[\frac{1}{2}(m + p)]}{2^{\frac{1}{2}m} \Gamma(\frac{1}{2}p)}, \quad (2.72)$$

$$E(U) = 0, \quad (2.73)$$

$$E(U^\top U) = \frac{1}{p} I_p. \quad (2.74)$$

Com efeito, usando o fato de que $R^2 \sim \chi_p^2$, tem-se (2.72). Como R e U são independentes, (2.73) segue de observar que $0 = E(Z) = E(RU) = E(R)E(U)$. Procedendo de forma análoga, obtém-se (2.74), uma vez que

$$E(UU^\top) = \frac{1}{E(R^2)} E(Z^\top Z) .$$

□

Note-se que (2.73) e (2.74) são consequência direta do fato de que $U \sim \mathcal{U}(S_p)$, onde $S_p = \{x \in \mathbb{R}^p : \|x\| = 1\}$ (veja o Exemplo 2.1).

Uma extensão destes resultados é obtida seguindo o mesmo procedimento e considerando que para qualquer subvetor $(U_{j_1}, U_{j_2}, \dots, U_{j_r})^\top$, $1 \leq r \leq p$, de $U = (U_1, U_2, \dots, U_p)^\top$ e qualquer conjunto de inteiros não-negativos $\{k_1, k_2, \dots, k_r\}$, com $k_1 + k_2 + \dots + k_r = k$, tem-se que

$$\begin{aligned} E\left(\prod_{i=1}^r U_{j_i}^{k_i}\right) &= \frac{1}{E(R^k)} E\left(\prod_{i=1}^r Z_{j_i}^{k_i}\right) = \frac{1}{E(R^k)} \prod_{i=1}^r E(Z_{j_i}^{k_i}) \\ &= \begin{cases} \frac{\Gamma(\frac{1}{2}p)}{\Gamma[\frac{1}{2}(k+p)]} \frac{1}{[\Gamma(\frac{1}{2})]^k} \prod_{i=1}^r \Gamma[\frac{1}{2}(k_i+1)] , & \text{para } k_i \text{ par, } i = 1, 2, \dots, r \\ 0 , & \text{caso contrário,} \end{cases} \end{aligned}$$

já que $Z_{j_1}, Z_{j_2}, \dots, Z_{j_n}$ são iid $N(0, 1)$. Usando as propriedades da função gama, este último resultado pode ser expresso como

$$E\left(\prod_{i=1}^r U_{j_i}^{k_i}\right) = \begin{cases} \frac{\prod_{i=1}^r [(k_i-1)(k_i-3)\dots 1]}{p(p+2)\dots(p+k-2)} , & \text{se } k_i \text{ é par, } i = 1, 2, \dots, r, \\ 0 , & \text{caso contrário.} \end{cases} \quad (2.75)$$

Para $r = 3$, (2.75) é derivado em Mitchell (1989) para $U = X/\|X\|$, com $X \sim \text{El}_p(0, I_p)$ e assumindo que X tem a densidade (2.3). Contudo, de (2.27) (ou do Fato 2.3), é claro que se $X = (X_1, X_2, \dots, X_p)^\top \sim \text{El}_p(0, I_p)$ e $P(X = 0) = 0$, então (2.73) – (2.75) também valem para $U = X/\|X\|$. Além disto, é fácil ver também que

$$E\left(\prod_{i=1}^r X_{j_i}^{k_i}\right) = E\left(\prod_{i=1}^r U_{j_i}^{k_i}\right) E(\|X\|^r) , \quad (2.76)$$

desde que os momentos envolvidos existam. De (2.75) e (2.76), a expressão dada em (2.42) é direta.

Por outro lado, Magnus e Neudecker (1979) mostram que se $Z \sim N_p(0, I_p)$ e A , B e C são matrizes simétricas de dimensão $p \times p$, então

$$E[Z^T AZ] = \text{tr}(A) ; \quad (2.77)$$

$$E[(Z^T AZ)(Z^T BZ)] = 2\text{tr}(AB) + \text{tr}(A)\text{tr}(B) \quad (2.78)$$

e

$$\begin{aligned} E[(Z^T AZ)(Z^T BZ)(Z^T CZ)] &= 8\text{tr}(ABC) + 2\text{tr}(A)\text{tr}(BC) + 2\text{tr}(B)\text{tr}(AC) \\ &+ 2\text{tr}(C)\text{tr}(AB) + \text{tr}(A)\text{tr}(B)\text{tr}(C) , \end{aligned} \quad (2.79)$$

onde $\text{tr}(\cdot)$ representa o traço de uma matriz.

A prova destes resultados pode ser encontrada também em Graybill (1983, pp. 366–368). Extensões ao caso em que Z é uma matriz aleatória $p \times q$ com $Z \sim N_{pq}(0, \Sigma, \Omega)$ (ou $N_{pq}(0, \Sigma \otimes \Omega)$) onde Σ e Ω são matrizes de dimensões $p \times p$ e $q \times q$, respectivamente, são dadas em Neudecker e Wansbeek (1987), Rosen (1988) e Tracy e Sultan (1993).

Usando (2.72) e o fato de que $R = \|Z\|$ e $U = Z/\|Z\|$ são independentes, então de (2.77), (2.78) e (2.79) tem-se, respectivamente, que

$$E[U^T AU] = \frac{1}{p} \text{tr}(A) ; \quad (2.80)$$

$$E[(U^T AU)(U^T BU)] = \frac{1}{p(p+2)} \{2\text{tr}(AB) + \text{tr}(A)\text{tr}(B)\} \quad (2.81)$$

e

$$\begin{aligned} E[(U^T AU)(U^T BU)(U^T CU)] &= \frac{1}{p(p+2)(p+4)} \{8\text{tr}(ABC) + 2\text{tr}(A)\text{tr}(BC) \\ &+ 2\text{tr}(B)\text{tr}(AC) + 2\text{tr}(C)\text{tr}(AB) + \text{tr}(A)\text{tr}(B)\text{tr}(C)\} . \end{aligned} \quad (2.82)$$

De acordo com (2.27) (ou do Fato 2.3) é claro que (2.80) – (2.82) também valem para $U = X/\|X\|$, quando $X \sim E\ell_p(0, I_p)$ e $P(X = 0) = 0$. Neste caso, Mitchell (1989) deriva (2.80) – (2.82) assumindo que X tem densidade (2.3).

2.5.5 Ferramentas para se obter certos cumulantes de derivadas da função log-verossimilhança de um modelo elíptico

Alguns dos resultados apresentados nesta seção também são devidos a Mitchell (1989) e têm relevância, por exemplo, na obtenção dos cumulantes de derivadas da função log-verossimilhança de uma distribuição elíptica. Contudo, os resultados de Mitchell (1989) são aqui provados em forma mais simples e uma extensão deles é apresentada.

Suponha que $X \sim El_p(0, I_p; f)$, isto é, X tem densidade (2.3), e considere a variável aleatória

$$W = \frac{d \log f(\|X\|^2)}{d(\|X\|^2)}. \quad (2.83)$$

Então, se A , B e C são matrizes simétricas de dimensão $p \times p$ e desde que os momentos envolvidos existam, tem-se que

$$E[(X^\top AX)W^m] = \frac{1}{p} \text{tr}(A) E[\|X\|^2 W^m], \quad (2.84)$$

$$E[(X^\top AX)(X^\top BX)W^m] = \frac{1}{p(p+2)} \{2\text{tr}(AB) + \text{tr}(A)\text{tr}(B)\} E[\|X\|^4 W^m] \quad (2.85)$$

e

$$\begin{aligned} E[(X^\top AX)(X^\top BX)(X^\top CX)W^m] &= \frac{1}{p(p+2)(p+4)} \{8\text{tr}(ABC) \\ &+ 2\text{tr}(A)\text{tr}(BC) + 2\text{tr}(B)\text{tr}(AC) + 2\text{tr}(C)\text{tr}(AB) \\ &+ \text{tr}(A)\text{tr}(B)\text{tr}(C)\} E[\|X\|^6 W^m]. \end{aligned} \quad (2.86)$$

Com efeito, seja $U = X/\|X\|$. Como U e $\|X\|$ são independentes, segue-se que U e W são independentes. Então, para (2.84), tem-se que

$$E[(X^\top AX)W^m] = E[(U^\top AU)\|X\|^2 W^m] = E[U^\top AU] E[\|X\|^2 W^m]. \quad (2.87)$$

Usando (2.27) e (2.80) tem-se a prova. A demonstração de (2.85) e (2.86) é análoga. $\square$

Mitchell (1989) prova estes fatos aplicando sucessivamente (2.75) com $r = 1, 2, 3$, isto é, sem considerar os resultados prévios obtidos em (2.77) – (2.79). Portanto, um trabalho algébrico maior foi requerido na prova de Mitchell (1989).

Os resultados apresentados anteriormente permitem obter cumulantes até terceira ordem para a derivada da função log-verossimilhança de um modelo elíptico. No entanto, cumulantes de ordem quarta são requeridos, por exemplo, na obtenção das correções tipo Bartlett dos testes usuais para hipóteses sobre os parâmetros do modelo (veja os Capítulos 4 e 5). Na derivação destes cumulantes é necessária a avaliação de expressões da forma $E[(X^\top AX)^4 W^m]$. Se A é idempotente com $r(A) = q$, então de (2.68) e (2.71) com $n = p$, tem-se que

$$E[(X^\top AX)^k W^m] = \frac{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q + k)}{\Gamma(\frac{1}{2}q)\Gamma(\frac{1}{2}p + k)} E[\|X\|^{2k} W^m] . \quad (2.88)$$

Se A não é idempotente, pode-se usar a propriedade (2.67) junto com (2.75). Neste caso, uma forma geral é aqui derivada usando polinômios zonais.

Seja A uma matriz simétrica de dimensão $p \times p$ com raízes características $\lambda_1, \lambda_2, \dots, \lambda_p$. Seja, também, $(k) = (k_1, k_2, \dots, k_r)$, onde $k_i \geq 0$, $i = 1, 2, \dots, r$, $k_1 + k_2 + \dots + k_r = k$, uma partição de ordem r , $1 \leq r \leq p$, do inteiro positivo k , $1 \leq k \leq p$. O polinômio zonal da matriz A correspondente à partição (k) , denotado por $C_{(k)}(A)$, é um polinômio homogêneo, simétrico e de grau k em $\lambda_1, \lambda_2, \dots, \lambda_p$ e tal que (James (1961, 1964), Constantine (1963); veja, também, Muirhead (1982, cap. 7)),

$$(\text{tr}(A))^k = (\lambda_1 + \lambda_2 + \dots + \lambda_p)^k = \sum_{(k)}^p C_{(k)}(A) .$$

James (1961) estabelece a seguinte propriedade fundamental destes polinômios,

$$\int_{O_p} C_{(k)}(A\Gamma B\Gamma^\top) d\Gamma = \frac{C_{(k)}(A)C_{(k)}(B)}{C_{(k)}(I_p)} ,$$

onde A e B são matrizes simétricas de dimensão $p \times p$ e $d\Gamma$ denota a medida de Haar que é invariante sobre O_p , o grupo das matrizes $\Gamma(p \times p)$ ortogonais (veja Muirhead (1982, seção 2.1.4)). Constantine (1963) mostra que

$$C_{(k)}(\Gamma A \Gamma^\top) = C_{(k)}(A) , \quad \Gamma \in O_p ,$$

e

$$C_{(k)}(I_p) = p(p+2) \cdots (p+2(k-1)) .$$

Os coeficientes de $C_{(k)}(\cdot)$, até $k = 6$, são dados em James (1964).

Com estes fatos tem-se que se $Z \sim N_p(0, I_p)$ e $A(p \times p)$ é simétrica, então (veja Hayakawa e Puri (1985)),

$$\begin{aligned} E[\{Z^\top A Z\}^k] &= E[\{\text{tr}(AZZ^\top)\}^k] \\ &= \int_{\mathbb{R}} \{\text{tr}(Azz^\top)\}^k p_z(z) dz \\ &= \int_{\mathbb{R}} \left\{ \int_{O_p} \{\text{tr}(A\Gamma zz^\top \Gamma^\top)\}^k d\Gamma \right\} p_z(z) dz \\ &= \int_{\mathbb{R}} \left\{ \int_{O_p} C_{(k)}(A\Gamma zz^\top \Gamma^\top) d\Gamma \right\} p_z(z) dz \\ &= \frac{C_{(k)}(A)}{C_{(k)}(I_p)} \int_{\mathbb{R}} \{\text{tr}(zz^\top)\}^k p_z(z) dz \\ &= \frac{C_{(k)}(A)}{C_{(k)}(I_p)} E[\|Z\|^{2k}] . \end{aligned} \tag{2.89}$$

Mas, $\|Z\|^2 \sim \chi_p^2$ e, portanto, $E(\|Z\|^{2k}) = C_{(k)}(I_p)$. Logo, se $U = Z/\|Z\|$, de (2.89) tem-se que

$$E[\{U^\top A U\}^k] = \frac{C_{(k)}(A)}{C_{(k)}(I_p)} . \tag{2.90}$$

Usando o apêndice dado em James (1964), pode-se deduzir que

$$C_{(1)}(A) = \text{tr}(A) ,$$

$$C_{(2)}(A) = (\text{tr}(A))^2 + 2\text{tr}(A^2) ,$$

$$C_{(3)}(A) = (\text{tr}(A))^3 + 6\text{tr}(A)\text{tr}(A^2) + 8\text{tr}(A^3) ,$$

$$C_{(4)}(A) = (\text{tr}(A))^4 + 12(\text{tr}(A))^2\text{tr}(A^2) + 12(\text{tr}(A^2))^2 + 32\text{tr}(A)\text{tr}(A^3) + 48\text{tr}(A^4) ,$$

$$\begin{aligned} C_{(5)}(A) &= (\text{tr}(A))^5 + 20(\text{tr}(A))^3\text{tr}(A^2) + 60\text{tr}(A)(\text{tr}(A^2))^2 + 80(\text{tr}(A))^2\text{tr}(A^3) \\ &\quad + 160\text{tr}(A^2)\text{tr}(A^3) + 240\text{tr}(A)\text{tr}(A^4) + 384\text{tr}(A^5) , \end{aligned}$$

$$C_{(6)}(A) = (\text{tr}(A))^6 + 30(\text{tr}(A))^4\text{tr}(A^2) + 180(\text{tr}(A))^2(\text{tr}(A^2))^2 + 160(\text{tr}(A))^3\text{tr}(A^3)$$

$$\begin{aligned}
& +120(\text{tr}(A^2))^3 + 960\text{tr}(A)\text{tr}(A^2)\text{tr}(A^3) + 720(\text{tr}(A))^2\text{tr}(A^4) \\
& +640(\text{tr}(A^3))^2 + 1440\text{tr}(A^2)\text{tr}(A^4) + 2304\text{tr}(A)\text{tr}(A^5) + 3840\text{tr}(A^6) .
\end{aligned}$$

De (2.27) (ou do Fato 2.3), segue que (2.90) também vale para $U = X/\|X\|$ quando $X \sim E\ell_p(0, I_p)$ com $P(X = 0) = 0$.

Finalmente, suponha que $X \sim E\ell_p(0, I_p; f)$ tenha a densidade dada em (2.3) e considere a variável aleatória W definida em (2.83). Então, se $A(p \times p)$ é simétrica e supondo que $E[\|X\|^{2k}W^m] < \infty$, tem-se que

$$E[(X^\top AX)^k W^m] = \frac{C_{(k)}(A)}{C_{(k)}(I_p)} E[\|X\|^{2k} W^m] . \quad (2.91)$$

A prova de (2.91) segue da independência entre U e $\|X\|$ e de considerar (2.27) junto com (2.90). $\square$

Os resultados dados no exemplo seguinte são devidos a Mitchell (1989).

Exemplo 2.14 Suponha $X \sim E\ell_p(0, I_p; f)$. Então:

(i) Para $X \sim N_p(0, I_p)$ tem-se que (veja o Exemplo 2.3),

$$W = \frac{d \log f(\|X\|^2)}{d(\|X\|^2)} = -\frac{1}{2} .$$

Logo, usando que $\|X\|^2 \sim \chi_p^2$, tem-se que

$$E[\|X\|^{2k} W^m] = (-1)^m 2^{-m} E[\|X\|^{2k}] = \frac{(-1)^m \Gamma[\frac{1}{2}(p+2k)]}{2^{m-k} \Gamma(\frac{1}{2}p)} .$$

(ii) Para $X \sim t_p(0, I_p; \nu)$ tem-se que (veja o Exemplo 2.5),

$$W = \frac{d \log f(\|X\|^2)}{d(\|X\|^2)} = -\frac{\nu+p}{2} (\nu + \|X\|^2)^{-1} .$$

Logo, usando que $\|X\|^2 \sim pF_{p,\nu}$, é fácil ver que (veja (3.27) no Capítulo 3)

$$\begin{aligned}
E[\|X\|^{2k} W^m] &= (-1)^m (\nu+p)^m E[\|X\|^{2k} (\nu + \|X\|^2)^{-m}] \\
&= \frac{(-1)^m (\nu+p)^m \Gamma[\frac{1}{2}(\nu+p)]}{2^m \nu^{m-k} \Gamma(\frac{1}{2}p) \Gamma(\frac{1}{2}\nu)} \times \frac{\Gamma[\frac{1}{2}(p+2k)] \Gamma[\frac{1}{2}(\nu+2m-2k)]}{\Gamma[\frac{1}{2}(\nu+p+2m)]} .
\end{aligned}$$

2.6 Algumas caracterizações da distribuição normal

Os resultados obtidos nas seções prévias mostram que muitas das propriedades clássicas da distribuição normal são válidas para todas as distribuições elípticas. Contudo, a distribuição normal possui várias propriedades que permitem caracterizá-la dentro da classe das distribuições elípticas. Alguns resultados neste sentido são obtidos primeiramente por Kelker (1970) e são resumidos em Muirhead (1982). Uma interessante extensão deles é estabelecida em Cambanis et al. (1981). Além disso, algumas caracterizações em termos de formas quadráticas é dada em Anderson e Fang (1986). Finalmente, Fang et al. (1990) proporcionam um resumo dos resultados mais relevantes.

Da Seção 2.5.2 segue-se o conhecido fato de que se $X = (X_1, X_2, \dots, X_n)^\top \sim N_n(\mu, \Sigma)$, então todas as distribuições marginais são normais. É sabido também que, sob normalidade, o fato de Σ ser diagonal implica em independência para $X_1, X_2, \dots, X_n$. Deste fato derivam-se os seguintes resultados, que são devidos a Kelker (1970).

- (a) Seja $X \sim El_n(\mu, \Sigma; \phi)$. Então cada distribuição marginal é normal se e somente se X tem distribuição normal.

Com efeito, se cada distribuição marginal de X é normal, da Seção 2.5.2 é imediato que X é normal desde que a função característica de X deve ter a mesma forma funcional que as funções características das distribuições marginais, isto é, normal. Se X é normal, segue das propriedades da distribuição normal que todas suas marginais também são normais.

□

- (b) Seja $X = (X_1, X_2, \dots, X_n)^\top \sim El_n(\mu, \Sigma; \phi)$, com Σ sendo uma matriz diagonal. Então, $X_1, X_2, \dots, X_n$ são independentes se e somente se X é normal.

Para provar a parte “somente se”, suponha sem perda de generalidade que $\mu = 0$.

Como, por hipótese, $\Sigma = \text{diag} \{\sigma_{11}, \sigma_{22}, \dots, \sigma_{nn}\}$, de (2.14) tem-se que

$$\phi_X(t) = \phi(t^\top \Sigma t) = \phi \left(\sum_{i=1}^n t_i^2 \sigma_{ii} \right), \quad t \in \mathbb{R}^n. \quad (2.92)$$

Se, além disso, $X_1, X_2, \dots, X_n$ são independentes, então

$$\phi_X(t) = \phi(t^\top \Sigma t) = \prod_{i=1}^n \phi(t_i^2 \sigma_{ii}), \quad t \in \mathbb{R}^n. \quad (2.93)$$

Escrevendo $u_i = t_i \sqrt{\sigma_{ii}}$, $i = 1, \dots, n$, de (2.92) e (2.93) segue que

$$\phi \left(\sum_{i=1}^n u_i^2 \right) = \prod_{i=1}^n \phi(u_i^2).$$

Esta equação é chamada de equação de Hamel e sua solução é dada por (veja, por exemplo, Feller (1970, p.305)) $\phi(u) = e^{-\frac{1}{2}cu}$, para alguma constante c . Como ϕ é uma função característica (isto é, ϕ é contínua com $\phi(0) = 1$ e $|\phi(u)| \leq 1$), então $c \geq 0$. Logo,

$$\phi(t^\top \Sigma t) = e^{-\frac{1}{2}ct^\top \Sigma t}, \quad t \in \mathbb{R}^n$$

é a função característica de uma distribuição normal. A implicação no sentido inverso segue das propriedades da distribuição normal. $\square$

Observe-se que a extensão de (b) no caso em que $X_1, X_2, \dots, X_n$ são vetores aleatórios $p \times 1$ com distribuição conjunta elíptica é imediata. Além disso, (b) é relevante no contexto estatístico, desde que supor que $X_1, X_2, \dots, X_n$ têm distribuição conjunta elíptica junto com independência equivale a assumir normalidade.

O material que segue é devido a Cambanis et al. (1981).

A distribuição normal também pode ser caracterizada em termos das distribuições condicionais (veja Kelker (1970) e Cambanis et al. (1981)). Com efeito, suponha que $X \sim El_n(\mu, \Sigma; \phi)$, com $r(\Sigma) = k$ e X, μ e Σ sendo particionados como em (2.49). Assuma também que $k_2 = r(\Sigma_{22})$ e $k_1 = k - k_2$ são estritamente positivos. Então:

(c) $\phi_{q^*(X_2)}(u)$ é degenerada para cada $u \geq 0$ se e somente se X é normal;

(d) se, ainda, X é não degenerado, X tem distribuição normal se e somente se, com probabilidade 1, a distribuição de X_1 dado X_2 é normal não degenerada.

Para provar a parte “somente se” de (c), note-se, em primeiro lugar, que de (2.63) segue que

$$\phi_{q^*(X_2)}(t_1^\top \Sigma_{11.2} t_2) = E \left[e^{it_1^\top (X_1 - \mu_1^*(X_2))} | X_2 \right], \quad t_1 \in \mathbb{R}^m.$$

Mas, para todo $t = (t_1^\top, t_2^\top)^\top \in \mathbb{R}^n$, tem-se que

$$\begin{aligned} \phi(t^\top \Sigma t) &= E[e^{it^\top (X - \mu)}] \\ &= E[e^{it_2^\top (X_2 - \mu_2)} E(e^{it_1^\top (X_1 - \mu_1)} | X_2)] \\ &= E[e^{it_2^\top (X_2 - \mu_2) + it_1^\top (\mu_1^*(X_2) - \mu_1)} E(e^{it_1^\top (X_1 - \mu_1^*(X_2))} | X_2)] \\ &= E[e^{i(t_2 + \Sigma_{22}^- \Sigma_{21} t_1)^\top (X_2 - \mu_2)} \phi_{q^*(X_2)}(t_1^\top \Sigma_{11.2}^* t_1)] . \end{aligned}$$

Logo, se para cada $u \geq 0$, $\phi_{q^*(X_2)}(u) = \psi(u)$ com probabilidade 1, segue-se que

$$\begin{aligned} \phi(t^\top \Sigma t) &= \psi(t_1^\top \Sigma_{11.2}^* t_1) E[e^{i(t_2 + \Sigma_{22}^- \Sigma_{21} t_1)^\top (X_2 - \mu_2)}] \\ &= \psi(t_1^\top \Sigma_{11.2}^* t_1) \phi[(t_2 + \Sigma_{22}^- \Sigma_{21} t_1)^\top \Sigma_{22} (t_2 + \Sigma_{22}^- \Sigma_{21} t_1)] . \end{aligned}$$

Agora, usando a identidade

$$t^\top \Sigma t = t_1^\top \Sigma_{11.2}^* t_1 + (t_2 + \Sigma_{22}^- \Sigma_{21} t_1)^\top \Sigma_{22} (t_2 + \Sigma_{22}^- \Sigma_{21} t_1) ,$$

tem-se que

$$\phi(u + v) = \psi(u) \phi(v) , \quad u, v \geq 0$$

(aqui se usa que $k_1, k_2 \geq 1$). Colocando $v = 0$, obtém-se que $\phi = \psi$ e

$$\phi(v + u) = \phi(u) \phi(v) , \quad u, v \geq 0 .$$

Como ϕ é uma função característica (ou seja, ϕ é contínua, $\phi(0) = 1$ e $|\phi(u)| \leq 1$), segue-se que

$$\phi(u) = e^{-\frac{1}{2}cu} , \quad u \geq 0 ,$$

para algum $c \geq 0$, isto é, $X \sim N_n(\mu, c\Sigma)$. A recíproca segue das propriedades da distribuição normal. $\square$

Para provar a parte “se” de (d), observe-se que $q^*(X_2) =^d \|X_2\|^2$, onde $(Z_1^\top, Z_2^\top)^\top = RU^{(k)} \sim El_k(0, I_k; \phi)$, o que junto ao fato de que a distribuição condicional de X_1 dado X_2 é normal não degenerada implica que $q^*(X_2) > 0$, com probabilidade 1. É suficiente, então, mostrar que $\phi_{q^*(X_2)}(u)$ é degenerada (não depende de X_2), para cada $u \geq 0$, e aplicar (c) (veja Cambanis et al. (1981)). A parte “somente se” segue das propriedades da distribuição normal. $\square$

Por outro lado, quando $X = (X_1^\top, X_2^\top)^\top$ é normal, os momentos centrados de qualquer ordem da distribuição condicional de X_1 dado X_2 são finitos e não dependem de X_2 . Além disso, se X é normal, todos os momentos condicionais de $R_{q^*(X_2)}$ dado X_2 são não estocásticos (degenerados). Daqui segue que se $X = (X_1^\top, X_2^\top)^\top \sim El_n(\mu, \Sigma; \phi)$, com $r(\Sigma) = k$ e $k_2 = r(\Sigma_{22})$ e $k_1 = k - d_2$ são estritamente positivos, então para cada inteiro positivo p , tem-se que:

- (e) $E[(R_{q^*(X_2)})^p | X_2]$ é finita e degenerada (isto é, não depende de X_2) se e somente se X é normal;
- (f) o p -ésimo momento central de X_1 dado X_2 é finito e degenerado (independe de X_2) se e somente se X tem distribuição normal.

Para provar a parte “somente se” de (f), considere (2.65) com (abusando da notação) $X_1 = (X_1, X_2, \dots, X_n)^\top$, $\mu_1^*(x_2) = (\mu_1^*, \mu_2^*, \dots, \mu_m^*)^\top$ e seja $a_i^\top$ a i -ésima linha da matriz A_1 . Escrevendo $p = p_1 + p_2 + \dots + p_m$, onde $p_1, p_2, \dots, p_m$ são inteiros não-negativos, tem-se que

$$\left(\prod_{i=1}^m (X_i - \mu_i^*)^{p_i} | X_2 \right) =^d (R_{q^*(X_2)})^p \prod_{i=1}^m (a_i^\top U^{(k_i)})^{p_i}.$$

Tomando valor esperado e usando que $R_{q^*}(X_2)$ e $U^{(k_1)}$ são independentes, a normalidade de X segue de (d). O sentido inverso segue das propriedades da distribuição normal. $\square$

No caso em que $X = (X_1^\top, X_2^\top)^\top \sim El_n(\mu, \Sigma; f)$, isto é, se X tem densidade (2.17), Kelker (1970) mostra que $\text{Cov}(X_1|X_2)$ é degenerada (não depende de X_2) se e somente se X é normal. Berkane e Bentler (1986, 1987) mostram também que os momentos centrais de primeira ordem de uma distribuição elíptica são determinados por Σ e pelos parâmetros de momentos $\kappa(m)$, os quais são nulos na distribuição normal (veja a Seção 2.5.1).

Finalmente, como é notado em Anderson e Fang (1986), a distribuição normal também pode ser caracterizada, na classe das distribuições esféricas, em termos de formas quadráticas como se segue (veja, também, Fang et al. (1990)):

(g) Seja $X = (X_1^\top, X_2^\top)^\top \sim El_n(0, I_n; \phi)$, onde X_1 é um vetor $m \times 1$, $1 \leq m < n$. Então $\|X_1\|^2$ e $\|X_2\|^2$ são independentes se e somente se $X \sim N_n(0, cI_n)$, para algum $c \geq 0$. Além disso, se $P(X = 0) = 0$ (ou $P(X = 0) < 1$) e $A = I_n - B$ é uma matriz idempotente com $1 \leq r(A) < n$, então $X^\top AX$ e $X^\top BX$ são independentes se e somente se $X \sim N_n(0, cI_n)$, para algum $c > 0$;

(h) Suponha que $X \sim El_n(0, I_n; \phi)$ se seja $A(n \times n)$ uma matriz simétrica. Então, $X^\top AX \sim \chi_k^2$ se e somente se $X \sim N_n(0, cI_n)$, para algum $c > 0$, $A^2 = c^{-1}A$ e $r(A) = k$.

A prova de (g) pode ser vista em Fang et al. (1990) e em Anderson e Fang (1987). A prova de (h) pode ser encontrada em Khatri e Murkejee (1987). Contudo, no caso em que $P(X = 0) = 0$, este último resultado segue de (2.70) (veja Anderson e Fang (1987)).

Exemplo 2.15 Suponha que $X = (X_1, X_2, \dots, X_n)^\top \sim El_n(0, I_n; \phi)$. Sejam $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ e $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$. Então, $\bar{X}^2 = X^\top AX$ e $S^2 = X^\top BX$ com $A = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$

e $B = I_n - A$, onde $\mathbf{1}_n = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$. De (g) segue que $\overline{X}^2$ e S^2 são independentes se e somente se $X \sim N_n(0, cI_n)$ para algum $c > 0$ (formas diferentes de provar este exemplo são dadas em Ali (1980) e em Arnold e Lynch (1982)).

Capítulo 3

Caracterizações da distribuição t multivariada

3.1 Introdução

Uma primeira versão da distribuição t multivariada é introduzida na literatura estatística por Cornish (1954) e por Dunnett e Sobel (1954). Tal versão foi obtida de forma independente por esses autores e é construída como o análogo multivariado da distribuição t_ν , a distribuição t de Student com ν graus de liberdade, isto é, considerando a distribuição do vetor aleatório $n \times 1$, definido por $X = Z/S$, onde $Z \sim N_n(0, \Sigma)$ e $\nu S^2/\sigma^2 \sim \chi_\nu^2$ são independentes. Aqui $\Sigma = \sigma^2 \Gamma$, onde $\Gamma = [\rho_{ij}]$ representa a matriz de correlações entre as componentes do vetor Z , e é assumida não-singular e conhecida. Por seu lado, σ^2 corresponde à variância das componentes de Z e é suposta desconhecida. Consequentemente, S^2 representa uma estimativa baseada em ν graus de liberdade do parâmetro σ^2 . Desta forma, é obtida uma versão centrada na origem da distribuição t multivariada não-singular. Com um procedimento similar, Kshirsagar (1960) estende esta definição ao caso não-central. Contudo, a distribuição t n -variada não-singular com locação $\mu \in \mathbb{R}^n$, dispersão $\Sigma(n \times n) > 0$ e ν graus de liberdade, $t_n(\mu, \Sigma; \nu)$, é definida formalmente em termos de uma densidade em Bennett (1961). Bennett também estabelece algumas propriedades desta distribuição e descreve o procedimento de máxima verossimilhança para

estimar μ e Σ , quando ν é conhecido. Posteriormente, Dickey (1967) introduz uma versão matricial desta distribuição (análoga à matriz normal) considerando a matriz aleatória $n \times p$ definida por $X = ZS^{-1/2}$, onde Z é a matriz normal $n \times p$ e S é a matriz Wishart $p \times p$ não-singular, com Z e S sendo independentes. Resultados adicionais sobre a matriz t e outras distribuições matriciais podem ser encontradas em Dawid (1977, 1981).

Por outro lado, nas três últimas décadas, tem sido observado um crescente interesse pela distribuição $t_n(\mu, \Sigma; \nu)$ na literatura estatística, tanto em teoria como em aplicações. Este interesse é devido basicamente ao fato de que as distribuições da família $\{t_n(\mu, \Sigma; \nu); \nu > 0\}$ têm caudas mais pesadas que a distribuição normal $N_n(\mu, \Sigma)$. Além disto, a distribuição $t_n(\mu, \Sigma; \nu)$ reduz-se (aproxima-se) à distribuição $N_n(\mu, \Sigma)$ quando $\nu = \infty$ ($\nu \rightarrow \infty$). Consequentemente, o interesse pela distribuição t tem sido particularmente enfatizado em estudos de robustez. Diferentes trabalhos têm sido feitos neste contexto. Alguns resultados iniciais são dados em Maronna (1976). Mais recentemente, Lange, Little e Taylor (1989) ilustram com diferentes aplicações a conveniência de usar a família t na modelagem de dados (veja, também, Taylor (1992)). Outros estudos interessantes baseados na distribuição t multivariada podem ser encontrados em Zellner (1976), Sutradhar e Ali (1981), Sing (1991), Delampady (1992) e Sutradhar (1993).

Como é ilustrado no Exemplo 2.9 (veja também o Exemplo 2.4), a distribuição t multivariada é um membro particular da classe das distribuições elípticas. Consequentemente, suas propriedades podem ser derivadas a partir da teoria desenvolvida no Capítulo 2. Contudo, esta distribuição tem sido aceita como uma alternativa robusta à distribuição normal. Tal importância motiva um desenvolvimento teórico individual em torno da distribuição t . Neste sentido, alguns resultados são dados em Tong (1980), Sutradhar (1986) e resumidos em Fang, Kotz e Ng (1990). Seguindo esta linha, este capítulo tem por objetivo contribuir com este desenvolvimento. Assim, na Seção 3.2.1, são discutidas as diferentes formas de definir a distribuição t multivariada. Considerando que esta distri-

buição pertence à classe das distribuições normais compostas (veja os Exemplos 2.6 e 2.10), a definição desta distribuição é dada em termos deste fato. Uma generalização da distribuição t multivariada é também estabelecida na Seção 3.2.1. Usando esta generalização, as propriedades da distribuição t são desenvolvidas na Seção 3.2.2. Na Seção 3.2.3 são discutidas algumas formas de se caracterizar a distribuição t generalizada, dentro da classe das distribuições elípticas (ou esféricas). Finalmente, seguindo Diaconis e Ylvisaker (1979, 1985), na Seção 2.3 estabelece-se uma caracterização preditivista da distribuição e a matriz t variadas. Resultados devidos a Kingman (1972) e a Smith (1981) e estendidos em Dawid (1978) e em Diaconis, Eaton e Lauritzen (1992), são aqui utilizados.

3.2 Definição e propriedades da distribuição t multivariada

3.2.1 Definição

A distribuição t n -variada padrão com ν graus de liberdade (g.l.), $t_n(0, I_n; \nu)$, está definida no Exemplo 2.5 como sendo a distribuição esférica $El_n(0, I_n; f)$ com f dada por

$$f(u) = k(n, \nu) \nu^{\frac{1}{2}\nu} \{\nu + u\}^{-\frac{1}{2}(\nu+n)}, \quad u \geq 0, \quad (3.1)$$

onde

$$k(n, \nu) = \frac{\Gamma[\frac{1}{2}(\nu + n)]}{\pi^{\frac{1}{2}n} \Gamma(\frac{1}{2}\nu)}. \quad (3.2)$$

Isto é, se X é um vetor aleatório $n \times 1$, então $X \sim t_n(0, I_n; \nu)$ se tem densidade dada por

$$p_X(x|\nu) = f(x^\top x) = k(\nu, n) \nu^{\frac{1}{2}\nu} \{\nu + x^\top x\}^{-\frac{1}{2}(\nu+n)}, \quad x \in \mathbb{R}^n. \quad (3.3)$$

Por outro lado, do Exemplo 2.6 segue que se X é um vetor aleatório $n \times 1$ definido por

$$X = V^{\frac{1}{2}} Z, \quad (3.4)$$

onde $V \sim \nu/\chi_\nu^2 \equiv GI(\nu/2, \nu/2)$ e $Z \sim N_n(0, I_n)$ são independentes, então $X \sim t_n(0, I_n; \nu)$. Isto é, usando que $(X|V = v) \stackrel{d}{=} vZ \sim N_n(0, vI_n)$, tem-se que $X \sim El_n(0, I_n; f)$ com f dada em (3.1).

Suponha, agora, que em (3.4) a variável aleatória V tenha função distribuição F_V sobre $[0, \infty)$ e seja independente de $Z \sim N_n(0, I_n)$. Em tal caso, Andrews e Mallows (1974) mostram que se $X = V^{\frac{1}{2}}Z \sim t_n(0, I_n; \nu)$, então F_V é $GI(\nu/2, \nu/2)$. Em outras palavras, $X \sim t_n(0, I_n; \nu)$ se e somente se $X \stackrel{d}{=} V^{\frac{1}{2}}Z$, onde $V \sim \nu/\chi_\nu^2$ e $Z \sim N_n(0, I_n)$ são independentes.

Usando (3.4) tem-se também que se $X = V^{\frac{1}{2}}Z \sim t_n(0, I_n; \nu)$, então

$$\|X\|^2 = V\|Z\|^2 \sim nF_{n,\nu}, \quad (3.5)$$

onde $F_{n,\nu}$ denota a distribuição F (de Fisher) com n e ν graus de liberdade (veja a Nota 3.1 abaixo). Além disto, do Fato 2.1 e (3.5) tem-se que $X \sim t_n(0, I_n; \nu)$ se e somente se

$$X \stackrel{d}{=} RU^{(n)}, \quad (3.6)$$

onde $R^2 \stackrel{d}{=} \|X\|^2 \sim nF_{n,\nu}$ e $U^{(n)} \sim \mathcal{U}(S_n)$ são independentes.

Note-se que de (3.6) segue que na classe das distribuições esféricas a distribuição t é determinada pela distribuição $nF_{n,\nu}$. Entretanto, de (3.4) segue que na subclasse das distribuições normais compostas esféricas a distribuição t é determinada pela distribuição ν/χ_ν^2 .

Nota 3.1 Sejam $S_1 \sim \chi_r^2/r$ e $S_2 \sim s/\chi_s^2$ ($S_2^{-1} \sim \chi_s^2/s$) independentes. Então, $F = S_1S_2 \sim F_{r,s}$ e tem densidade dada por

$$p_F(t) = \frac{\Gamma[\frac{1}{2}(r+s)]}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}s)} \left(\frac{r}{s}\right)^{\frac{1}{2}r} t^{\frac{1}{2}(r-2)} \left\{1 + \frac{r}{s}t\right\}^{-\frac{1}{2}(r+s)}, \quad t > 0,$$

de onde tem-se que

$$E(F^{\frac{1}{2}k}) = \left(\frac{s}{r}\right)^{\frac{1}{2}k} \frac{\Gamma[\frac{1}{2}(r+k)]\Gamma[\frac{1}{2}(s-k)]}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}s)}, \quad k < s.$$

É fácil ver também que se $S \sim pF_{p,s}$ e $T \sim \text{Beta}(r/2, (p-r)/2)$, $0 < r < p$, são independentes, então $Q = ST \sim rF_{r,s}$.

Os resultados dados acima sugerem que (3.1), (3.4) e (3.6) são formas equivalentes de definir a distribuição $t_n(0, I_n; \nu)$. O fato que segue é motivado por tais resultados.

Fato 3.1 *Seja X um vetor aleatório $n \times 1$. Então as seguintes afirmações são equivalentes:*

- (i) X tem densidade dada por (3.3);
- (ii) $X = {}^d V^{\frac{1}{2}} Z$, onde $V \sim \nu/\chi_\nu^2$ e $Z \sim N_n(0, I_n)$ são independentes;
- (iii) $X = {}^d R U^{(n)}$, onde $R^2 \sim nF_{n,\nu}$ e $U^{(n)} \sim \mathcal{U}(S_n)$ são independentes.

Com efeito, é fácil ver que (ii) implica em (i), desde que $(X|V = v) \sim N_n(0, vI_n)$ (veja (2.12) no Exemplo 2.6). Que (i) implica em (ii) é mostrado em Andrews e Mallows (1974). A equivalência de (ii) com (iii) segue do Fato 2.1, desde que $Z = {}^d R^* U^{(n)}$, com $R^{*2} = {}^d \|Z\|^2 \sim \chi_n^2$ sendo independente com V e $U^{(n)}$. Daqui segue que $R^2 = {}^d R^{*2} V \sim nF_{n,\nu}$. $\square$

Com qualquer destas alternativas, a distribuição $t_n(\mu, \Sigma; \nu)$ não-singular ($\Sigma > 0$) pode ser introduzida considerando a transformação linear dada por

$$X = \mu + \Sigma^{\frac{1}{2}} Y, \quad (3.7)$$

onde $Y \sim t_n(0, I_n; \nu)$. De fato, aplicando o método do Jacobiano junto com a Definição 3.3 tem-se que

$$\begin{aligned} p_X(x|\mu, \Sigma, \nu) &= |\Sigma|^{-\frac{1}{2}} p_Y(\Sigma^{-\frac{1}{2}}(x - \mu)|\nu) \\ &= |\Sigma|^{-\frac{1}{2}} f((x - \mu)^\top \Sigma^{-1}(x - \mu)) \\ &= k(n, \nu) |\Sigma|^{-\frac{1}{2}} \nu^{\frac{1}{2}} \{\nu + (x - \mu)^\top \Sigma^{-1}(x - \mu)\}^{-\frac{1}{2}(n+\nu)}, \quad x \in \mathbb{R}^n, \end{aligned} \quad (3.8)$$

onde $k(n, \nu)$ é definido em (3.2).

Similarmente, de acordo com a definição (3.4), a relação dada em (3.7) pode ser expressa como

$$X = \mu + V^{\frac{1}{2}} \Sigma^{\frac{1}{2}} Z, \quad (3.9)$$

onde $V \sim \nu/\chi_\nu^2$ e $Z \sim N_n(0, I_n)$ são independentes. Condiçãoando em $V = v$, tem-se que

$$(X|V = v) \sim N_n(\mu, v\Sigma). \quad (3.10)$$

Daqui segue que

$$p_X(x|\mu, \Sigma, \nu) = \int_0^\infty |\Sigma|^{-\frac{1}{2}} (2\pi v)^{-\frac{1}{2}n} e^{-\frac{1}{2v}(x-\mu)^\top \Sigma^{-1}(x-\mu)} dF_V(v|\nu), \quad (3.11)$$

onde $dF_V(v|\nu) = p_V(v|\nu) dv$, com $p_V(v|\nu)$ sendo a função de densidade de $V \sim \nu/\chi_\nu^2$. Isto é,

$$p_V(v|\nu) = \frac{(\nu/2)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} v^{-\frac{1}{2}(\nu+2)} e^{-\frac{\nu}{2v}}, \quad v > 0. \quad (3.12)$$

Substituindo (3.12) em (3.11) tem-se finalmente (3.8).

Da mesma maneira, aplicando agora (3.6) para o vetor aleatório Y , envolvido em (3.7), tem-se que

$$X =^d \mu + R \Sigma^{\frac{1}{2}} U^{(n)}, \quad (3.13)$$

onde $R^2 \sim nF_{n,\nu}$ e $U^{(n)} \sim \mathcal{U}(S_n)$ são independentes. Da Nota 3.1 segue que nesta representação a variável aleatória R pode ser fatorada como $R = V^{\frac{1}{2}} R^*$, onde $V \sim x/\chi_\nu^2$ e $R^{*2} \sim \chi_n^2$ são independentes (digamos, $R^* = \|Z\|$, com $Z \sim N_n(0, I_n)$ e sendo independente de V). Usando este fato e observando que $Z = R^* U^{(n)} \sim N_n(0, I_n)$, (3.13) se reduz a (3.9), de onde segue (3.8).

Estes resultados são resumidos no fato que segue, cuja prova vem diretamente da aplicação do Fato 3.1 como foi descrito acima.

Fato 3.2 *Seja $X(n \times 1)$ um vetor aleatório com vetor de locação $\mu(n \times 1) \in \mathbb{R}^n$ e matriz de dispersão $\Sigma(n \times 1) > 0$. Então as seguintes afirmações são equivalentes:*

- (i) *X tem densidade dada por (3.8);*
- (ii) *$X =^d \mu + V^{\frac{1}{2}} \Sigma^{\frac{1}{2}} Z$, onde $V \sim \nu/\chi_\nu^2$ e $Z \sim N_n(0, I_n)$ são independentes;*
- (iii) *$X =^d \mu + R \Sigma^{\frac{1}{2}} U^{(n)}$, onde $R^2 \sim nF_{n,\nu}$ e $U^{(n)} \sim \mathcal{U}(S_n)$ são independentes;*
- (iv) *$X =^d \mu + \Sigma^{\frac{1}{2}} Y$, onde $Y \sim t_n(0, I_n; \nu)$.*

Além disto, é fácil ver que se $X \sim t_n(\mu, \Sigma; \nu)$, $\Sigma > 0$, então

$$Q(X) = (X - \mu)^T \Sigma^{-1} (X - \mu) \sim nF_{n,\nu} . \quad (3.14)$$

A distribuição t multivariada não-singular pode ser definida usando qualquer das alternativas equivalentes dadas no Fato 2.1. A primeira alternativa assume explicitamente a existência de uma função densidade e, portanto, não permite incorporar uma versão singular da distribuição t multivariada. Esta versão pode ser incorporada na definição usando-se qualquer das alternativas restantes. Contudo, uma vez que a distribuição t está contida nas distribuições normais compostas (veja o Exemplo 2.6), a qual é uma subclasse das distribuições esféricas, neste trabalho a distribuição t multivariada será definida em termos da segunda alternativa.

Sejam $Z \sim N_n(0, \Sigma)$, com $r(\Sigma) = k$, e $V \sim \nu/\chi_\nu^2$ independentes. Note-se que Z pode ser expresso como $Z = AZ^{(k)}$, onde $AA^T = \Sigma$, $r(A) = k$ e $Z^{(k)} \sim N_k(0, I_k)$. Com esta notação, a distribuição t n -variada pode ser definida como segue.

Definição 3.1 *Um vetor aleatório $X(n \times 1)$ é dito ter distribuição t n -variada com parâmetros $\mu(n \times 1) \in \mathbb{R}^n$ (locação), $\Sigma(n \times n)$ (dispersão) e $\nu > 0$ (graus de liberdade ou parâmetro de forma ou curtose), isto é, $X \sim t_n(\mu, \Sigma; \nu)$, se*

$$X =^d \mu + V^{\frac{1}{2}} \Sigma^{\frac{1}{2}} Z . \quad (3.15)$$

Se $r(\Sigma) = k = n$, isto é, $\Sigma > 0$, de (3.15) segue que X tem densidade dada por (3.8). Se $r(\Sigma) = k < n$, então a distribuição de $(X|V = v)$ é $N_n(0, v\Sigma)$ singular. Consequentemente, a distribuição incondicional de X é também singular. Este caso pode ser interpretado como a versão singular da distribuição t n -variada. Contudo, para ambos os casos, de (3.15) juntamente com (3.10) e (3.12), segue que $X \sim t_n(\mu, \Sigma; \nu)$ tem função característica dada por

$$\phi_X(t|\mu, \Sigma, \nu) = e^{it^\top \mu} \phi(t^\top \Sigma t), \quad t \in \mathbb{R}^n, \quad (3.16)$$

onde

$$\phi(u) = E\left(e^{-\frac{1}{2}uV}\right) = \frac{(\nu/2)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \int_0^\infty v^{-\frac{1}{2}(\nu+2)} e^{-\frac{1}{2}(uv+\nu v^{-1})} dv, \quad u \geq 0. \quad (3.17)$$

Isto é, se $X \sim t_n(\mu, \Sigma; \nu)$, então $X \sim El_n(\mu, \Sigma; \phi)$ com ϕ dada por (3.17).

Usando o fato de que se $Y = (Y_1, Y_2, \dots, Y_n)^\top \sim El_n(0, I_n; \phi)$, então $Y_1 \sim El_1(0, 1; \phi)$, onde $\phi(t^2) = E(e^{itY_1})$, $t \in \mathbb{R}$, uma forma alternativa de expressar (3.17) segue-se de supor que $Y = (Y_1, Y_2, \dots, Y_n) \sim t_n(0, I_n; \nu)$ e de aplicar (3.3). Isto é,

$$\phi(u) = E\left(e^{i\sqrt{u}Y_1}\right) = 2k(1, \nu) \int_0^\infty \cos(\sqrt{\nu}ut)(1+t^2)^{-\frac{1}{2}(\nu+1)} dt, \quad u \geq 0. \quad (3.18)$$

Esta integral é desenvolvida em Sutradhar (1986, 1988), para ν sendo um inteiro, par e ímpar, e para ν sendo um valor fracionário. Assim, três expressões diferentes da função característica de $X \sim t_n(\mu, \Sigma; \nu)$ são obtidas em Sutradhar. Estas expressões também podem ser encontradas em Fang et al. (1990, p. 87).

Uma extensão da distribuição $t_n(\mu, \Sigma; \nu)$ pode ser obtida, assumindo que em (3.15) a variável aleatória V tem distribuição $\lambda/\chi_\nu^2 \equiv GI(\lambda/2, \nu/2)$, isto é, V tem densidade dada por

$$p_V(v|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} v^{-(\alpha+1)} e^{-\beta v^{-1}}, \quad v > 0, \quad (3.19)$$

com parâmetro de escala $\beta = \lambda/2 > 0$ e parâmetro de forma $\alpha = \nu/2 > 0$. Neste caso, em (3.15) tem-se que $X \sim E\ell_n(\mu, \Sigma; \phi)$ com ϕ dada por

$$\phi(u) = E\left(e^{-\frac{1}{2}uV}\right) = \frac{(\lambda/2)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \int_0^\infty v^{-\frac{1}{2}(\nu+2)} e^{-\frac{1}{2}(uv+\lambda v^{-1})} dv, \quad u \geq 0. \quad (3.20)$$

Esta integral pode ser escrita como (3.18) substituindo ν por λ em $\cos(\sqrt{\nu}ut)$. Além disso, usando (3.10) junto com (3.19), é fácil ver que neste caso X tem densidade dada por

$$p_X(x|\mu, \Sigma, \lambda, \nu) = k(n, u) |\Sigma|^{-\frac{1}{2}} \lambda^{\frac{1}{2}V} \left\{ \lambda + (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\}^{-\frac{1}{2}(\nu+n)}, \quad x \in \mathbb{R}^n, \quad (3.21)$$

onde $k(n, \nu)$ é dada em (3.2).

A notação $X \sim t_n(\mu, \Sigma; \lambda, \nu)$ será incorporada para referir-nos a esta extensão da distribuição t n -variada. Neste caso, diremos que X tem distribuição t n -variada generalizada. De (3.20) e (3.21), é claro que quando o parâmetro extra $\lambda = \nu$ tem-se a versão usual da distribuição t n -variada, isto é,

$$t_n(\mu, \Sigma; \nu, \nu) \equiv t_n(\mu, \Sigma; \nu).$$

A definição dada em (3.21) da distribuição t multivariada é usada em Dickey (1967) para derivar a matriz t -variada. Uma variação adicional é obtida colocando em (3.21) $\nu = 2\gamma - n$ (ou $\gamma = (\nu + n)/2$, com $\gamma > n/2$). Neste último caso, a distribuição de X é do tipo Pearson VII e é denotada por $PVII_n(\mu, \Sigma; \lambda, \gamma)$ (para maiores detalhes, veja Fang et al. (1990), seção 3.3). Contudo, a menos deste capítulo, neste trabalho nos restringiremos ao caso particular da família de distribuições t n -variadas não-singulares dada por

$$\{t_n(\mu, \Sigma; \nu); \mu \in \mathbb{R}^n, \Sigma > 0, \nu > 0\}, \quad (3.22)$$

cujas funções de densidades são dadas em (3.8).

Como já foi mencionado anteriormente, quando $\nu = 1$ e $\nu = \infty$ ($\nu \rightarrow \infty$), (3.22) se reduz às famílias de distribuições n -variadas não-singulares de Cauchy, $\{C_n(\mu, \Sigma), \mu \in \mathbb{R}^n, \Sigma > 0\}$ e normal, $\{N_n(\mu, \Sigma), \mu \in \mathbb{R}^n, \Sigma > 0\}$, respectivamente.

3.2.2 Propriedades

Uma vez que a distribuição t multivariada é um membro particular da classe das distribuições elípticas, suas propriedades podem ser obtidas aplicando a teoria desenvolvida no Capítulo 2. Contudo, e com a finalidade de ilustração, algumas das propriedades que serão dadas no que segue serão provadas aplicando a Definição 3.1. Consideremos, em primeiro lugar, alguns resultados preliminares.

Seja $V \sim GI(\beta, \alpha)$, isto é, V tem densidade dada por (3.19). Então, notando que $V =^d Y^{-1}$, onde $Y \sim Ga(\beta, \alpha)$, a distribuição gama com parâmetros de escala β e de forma α , para $s < \alpha$ tem-se que

$$E(V^s) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{(\alpha-s)-1} e^{-\beta y} dy = \frac{\beta^s \Gamma(\alpha - s)}{\Gamma(\alpha)}. \quad (3.23)$$

Suponha agora que $V \sim \lambda/\chi_\nu^2$ e $Z \sim N_p(0, I_p)$ são independentes. Do fato que o vetor aleatório definido por $X = V^{\frac{1}{2}}Z \sim t_p(0, I_p; \lambda, \nu)$ juntamente com o teorema de Bayes e de aplicar o Fato 2.5 para $\|X\|^2 = V\|Z\|^2$, segue que

$$(V|X = x) =^d (V \mid \|X\|^2 = \|x\|^2) \sim (\lambda + \|x\|^2)/\chi_{\nu+p}^2. \quad (3.24)$$

Daqui tem-se também que

$$E(V^{-1}|X = x) = E(V^{-1} \mid \|X\|^2 = \|x\|^2) = \frac{\nu + p}{\lambda + \|x\|^2}. \quad (3.25)$$

Além disto, usando que $(\nu/\lambda)V \sim \nu/\chi_\nu^2$ e $(1/p)\|Z\|^2 \sim \chi_p^2/p$, da Nota 3.1 segue que

$$F = \frac{\nu}{\lambda p} \|X\|^2 = \underbrace{\left(\frac{\nu}{\lambda} V\right)}_1 \left(\frac{1}{p} \|Z\|^2\right) \sim F_{p, \nu}. \quad (3.26)$$

Notando agora que

$$\|X\|(\lambda + \|X\|^2)^{-1} = \lambda^{-\frac{1}{2}} \left(\frac{p}{\nu} F\right)^{\frac{1}{2}} \left(1 + \frac{p}{\nu} F\right)^{-1}$$

e de aplicar novamente a Nota 3.1, segue-se que (veja o Exemplo 2.14(ii))

$$\begin{aligned}
E[\|X\|^k (\lambda + \|X\|^2)^{-m}] &= \lambda^{\frac{1}{2}k-m} E \left[\left(\frac{p}{\nu} F \right)^{\frac{1}{2}k} \left(1 + \frac{p}{\nu} F \right)^{-m} \right] \\
&= \frac{B \left[\frac{1}{2}(p+k), \frac{1}{2}(\nu+2m-k) \right]}{\lambda^{\frac{1}{2}(2m-k)} B(\frac{1}{2}p, \frac{1}{2}\nu)} , \text{ se } (\nu+2m-k > 0), \quad (3.27)
\end{aligned}$$

onde $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ é a função beta. Note-se que de (3.26) também segue que

$$T = \frac{\|X\|^2}{\lambda + \|X\|^2} = \frac{(p/\nu)F}{1 + (p/\nu)F} \sim \text{Beta} \left(\frac{1}{2}p, \frac{1}{2}\nu \right) .$$

Por outro lado, colocando $X = (X_1, X_2, \dots, X_p)^\top$ e $Z = (Z_1, Z_2, \dots, Z_p)^\top$, tem-se que $X_i = V^{\frac{1}{2}} Z_i$, onde $Z_1, Z_2, \dots, Z_p$ e V são independentes, com $Z_i \sim N(0, 1)$ e $V \sim GI(\lambda/2, \nu/2)$. Destes resultados juntamente com (3.23) e (2.72), tem-se que se $k_1, k_2, \dots, k_p$ são inteiros não-negativos tais que $k = k_1 + k_2 + \dots + k_p < \nu$, então

$$\begin{aligned}
E \left[\prod_{i=1}^p X_i^{k_i} \right] &= E \left[V^{\frac{1}{2}k} \right] \prod_{i=1}^p E \left[Z_i^{k_i} \right] \\
&= \begin{cases} \left(\frac{\lambda}{\pi} \right)^{\frac{1}{2}k} \frac{\Gamma \left[\frac{1}{2}(\nu - k) \right]}{2^k \Gamma(\frac{1}{2}\nu)} \prod_{i=1}^k \Gamma \left[\frac{1}{2}(k_i + 1) \right] , & \text{se } k_i \text{ é par, } i = 1, 2, \dots, p , \\ 0 , & \text{se } k_i \text{ é ímpar, para algum } i = 1, 2, \dots, p . \end{cases} \quad (3.28)
\end{aligned}$$

Este resultado também pode ser obtido aplicando (2.76) junto com (2.75) e a Nota 3.1.

As propriedades dadas a seguir serão derivadas a partir da distribuição t multivariada generalizada, $t_n(\mu, \Sigma; \lambda, \nu)$, definida em (3.21). Estas propriedades também valem no caso singular com $r(\Sigma) = k < n$. Para $\lambda = \nu$, tem-se as correspondentes propriedades da distribuição t multivariada usual, $t_n(\mu, \Sigma; \nu)$. Na derivação destas propriedades, a seguinte notação será usada.

Seja $X = \mu + V^{\frac{1}{2}} Z \sim t_n(\mu, \Sigma; \lambda, \nu)$, isto é, aqui tem-se que $V \sim \lambda/\chi_\nu^2$, $Z \sim N_n(0, \Sigma)$ e são independentes. Com esta distribuição, o vetor aleatório X verifica as seguintes propriedades:

(a) $E(X)$ e $\text{Cov}(X)$ são dadas por

$$E(X) = \mu, \text{ se } \nu > 1, \quad (3.29)$$

e

$$\text{Cov}(X) = \frac{\lambda}{\nu - 2} \Sigma, \text{ se } \nu > 2. \quad (3.30)$$

Com efeito, notando que $E(X) = E(V^{\frac{1}{2}})E(Z)$ e $\text{Cov}(X) = E(V)E(Z^T Z)$, a prova segue de aplicar (3.23) com $\beta = \lambda/2$ e $\alpha = \nu/2$. $\square$

(b) Se $A(m \times n)$ e $\eta(m \times 1)$ são não estocásticos, tem-se que

$$\eta + AX \stackrel{d}{=} \eta + A\mu + V^{\frac{1}{2}}AZ \sim t_m(\eta + A\mu, A\Sigma A^T; \lambda, \nu), \quad (3.31)$$

desde que $AZ \sim N_m(0, A\Sigma A^T)$ é independente de $V \sim \lambda/\chi_\nu^2$.

Em particular, para $\Sigma > 0$, tem-se que

$$\Sigma^{-\frac{1}{2}}(X - \mu) \sim t_n(0, I_n; \lambda, \nu). \quad (3.32)$$

Por outro lado, considerando a partição

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{e } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad (3.33)$$

onde X_1, Z_1 e μ_1 são vetores $m \times 1$ e Σ_{11} é uma matriz $m \times m$, de (3.31) também segue que

$$X_1 \stackrel{d}{=} \mu_1 + V^{\frac{1}{2}}Z_1 \sim t_m(\mu_1, \Sigma_{11}; \lambda, \nu), \quad (3.34)$$

onde $Z_1 \sim N_m(0, \Sigma_{11})$ é independente de $V \sim \lambda/\chi_\nu^2$. A distribuição marginal de X_2 é obtida por simetria.

(c) Se X é particionado como em (3.33), tem-se que a distribuição condicional de X_1 , dado $X_2 = x_2$, é t m -variada generalizada.

Este fato pode ser provado usando os resultados desenvolvidos na Seção 2.5.3, desde que se $X \sim t_n(\mu, \Sigma; \lambda, \nu)$, então $X \sim El_n(\mu, \Sigma; \phi)$ com ϕ dada em (3.19).

Com a idéia de ilustrar e de derivar alguns resultados que serão úteis nas seções que seguem, consideremos (por simplicidade) o caso não-singular com $\Sigma > 0$. Mostraremos, portanto, que se $X = (X_1^\top, X_2^\top)^\top \sim t_n(\mu, \Sigma; \lambda, \nu)$, então

$$(X_1|X_2 = x_2) \sim t_m(\mu_1(x_2), \Sigma_{11.2}; \lambda_{q(x_2)}, \nu_m), \quad (3.35)$$

onde (de acordo com a partição (3.33); veja também (2.66))

$$\left. \begin{aligned} \mu_1(x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \\ \Sigma_{11.2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \\ q(x_2) &= (x_2 - \mu_2)^\top \Sigma_{22}^{-1}(x_2 - \mu_2), \end{aligned} \right\} \quad (3.36)$$

e os parâmetros λ_a e ν_m são dados por

$$\left. \begin{aligned} \lambda_a &= \lambda + a, \quad a \geq 0, \\ \nu_m &= \nu + n - m, \quad 1 \leq m \leq n. \end{aligned} \right\} \quad (3.37)$$

De fato, como $(X_1|X_2 = x_2) \sim El_m(\mu_1(x_2), \Sigma_{11.2}; \phi_{q(x_2)})$ (veja (2.66)), então para provar (3.35) é suficiente mostrar que para cada $a \geq 0$, a função ϕ_a é da forma dada em (3.19).

Para tais efeitos, basta considerar o caso esférico com $\mu = 0$ e $\Sigma = I_n$. Seja então

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = V^{\frac{1}{2}} \begin{pmatrix} Z^{(m)} \\ Z^{(n-m)} \end{pmatrix},$$

onde Y_1 é um vetor $m \times 1$; $V \sim \lambda|\chi_V^2$, $Z^{(m)} \sim N_m(0, I_m)$ e $Z^{(n-m)} \sim N_{n-m}(0, I_{n-m})$ são independentes. Isto é,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = V^{\frac{1}{2}} \begin{pmatrix} Z^{(m)} \\ Z^{(n-m)} \end{pmatrix} \sim t(0, I_n; \lambda, \nu).$$

É claro que Y_1 e Y_2 são condicionalmente independentes dado $V = v$, com $(Y_1|V = v) \sim N_m(0, vI_m)$. Deste resultado juntamente com o fato de que $(Y_1|Y_2 = y_2) \sim$

$E\ell_m(0, I_m; \phi_{\|y_2\|^2})$, tem-se que

$$\begin{aligned}\phi_{\|Y_2\|^2}(t_1^\top t_1) &= E\left(e^{t_1^\top Y_1} \mid Y_2\right) = E\left\{E\left(e^{t_1^\top Y_1} \mid V, Y_2\right) \mid Y_2\right\} \\ &= E\left\{E\left(e^{t_1^\top Y_1} \mid V\right) \mid Y_2\right\} = E\left\{e^{-\frac{1}{2}V t_1^\top t_1} \mid Y_2\right\}, \quad t_1 \in \mathbb{R}^m.\end{aligned}$$

Colocando $\|y_2\|^2 = a$, de (3.24) segue que

$$(V \mid Y_2 = y_2) =^d (V \mid \|Y_2\|^2 = a) \sim \lambda_a / \chi_{\nu_m}^2,$$

onde λ_a e ν_m são definidos em (3.37). Daqui segue que

$$\begin{aligned}\phi_a(u) &= E\left(e^{-\frac{1}{2}uV} \mid \|Y_2\|^2 = a\right) \\ &= \frac{(\lambda_a/2)^{\frac{1}{2}\nu_m}}{\Gamma(\frac{1}{2}\nu_m)} \int_0^\infty v^{-\frac{1}{2}(\nu_m+2)} e^{-\frac{1}{2}(uv+\lambda_a v^{-1})} dv, \quad u \geq 0.\end{aligned}\quad (3.38)$$

Isto prova que $(Y_1 \mid Y_2 = y_2) \sim t_n(0, I_n; \lambda_{\|y_2\|^2}; \nu_m)$. A prova de (3.35) segue de observar que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} =^d \begin{pmatrix} \mu_1 + \Sigma_{11.2}^{\frac{1}{2}} Y_1 + \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} Y_2 \\ \mu_2 + \Sigma_{22}^{\frac{1}{2}} Y_2 \end{pmatrix},$$

de onde tem-se que

$$(X_1 \mid X_2 = x_2) =^d \mu_1(x_2) + \Sigma_{11.2}^{\frac{1}{2}} (Y_1 \mid Y_2 = \Sigma_{22}^{-\frac{1}{2}}(x_2 - \mu_2)) \sim t_m(\mu_1(x_2), \Sigma_{11.2}; \lambda_{q(x_2)}, \nu_m).$$

□

Por outro lado, é fácil verificar que

$$(X_1 \mid X_2 = x_2) =^d \mu_1(x_2) + V_{q(x_2)}^{\frac{1}{2}} \Sigma_{11.2}^{\frac{1}{2}} Z^{(m)}, \quad (3.39)$$

onde, para cada $x_2 \in \mathbb{R}^{n-m}$,

$$V_{q(x_2)} =^d (V \mid X_2 = x_2) \sim \lambda_{q(x_2)} / \chi_{\nu_m}^2, \quad (3.40)$$

e é independente de $Z^{(m)} \sim N_m(0, I_m)$.

Destas últimas relações segue também que

$$(X_1|X_2 = x_2) =^d \mu_1(x_2) + R_{q(x_2)} \Sigma_{11.2}^{\frac{1}{2}} U^{(m)} , \quad (3.41)$$

onde, para cada $x_2 \in \mathbb{R}^{n-m}$,

$$R_{q(x_2)}^2 =^d V_{q(x_2)} \|Z^{(m)}\|^2 \sim \frac{m \lambda_{q(x_2)}}{\nu_m} F_{m, \nu_m} , \quad (3.42)$$

é independente de $U^{(m)}$.

De (3.39) ou (3.41) (ou aplicando (3.29) e (3.30) com λ e ν substituídos por λ_a e λ_m , respectivamente) segue que

$$\begin{aligned} E(X_1|X_2) &= \mu_1(x_2) \\ &= \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2) , \end{aligned} \quad (3.43)$$

para $\nu_m = \nu + n - m > 1$, e

$$\begin{aligned} \text{Cov}(X_1|X_2) &= \frac{\lambda_{q(X_2)}}{\nu_m - 2} \Sigma_{11.2} \\ &= \frac{\lambda + (X_2 - \mu_2)^\top \Sigma_{22}^{-1} (X_2 - \mu_2)}{\nu + n - m - 2} (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}) , \end{aligned} \quad (3.44)$$

para $\nu_m = \nu + n - m > 2$.

Note-se que $\text{Cov}(X_1|X_2)$ é linear em $q(X_2) = (X_2 - \mu_2)^\top \Sigma_{22}^{-1} (X_2 - \mu_2)$.

Finalmente, se $X \sim t_n(\mu, \Sigma; \lambda, \nu)$ e $\Sigma > 0$, então X tem a densidade dada em (3.20). Isto é, $X \sim El_n(\mu, \Sigma; f)$, onde a função f é definida por

$$f(u) = k(n, \nu) \lambda^{\frac{1}{2}\nu} \{\lambda + u\}^{-\frac{1}{2}(\nu+n)} , \quad u \geq 0 , \quad (3.45)$$

com $k(n, \nu)$ sendo como em (3.2). Considerando a partição dada em (3.33) e aplicando (2.55), segue que $(X_1|X_2 = x_2)$ tem densidade dada por

$$p_{X_1}(x_1|x_2; \mu, \Sigma, \lambda, \nu) = |\Sigma_{11.2}|^{-\frac{1}{2}} f_{q(x_2)}((x_1 - \mu_1(x_2))^\top \Sigma_{11.2}^{-1} (x_1 - \mu_1(x_2))) , \quad x_1 \in \mathbb{R}^m , \quad (3.46)$$

onde

$$f_a(u) = k(m, \nu_m) \lambda_a^{\frac{1}{2}\nu_m} \{\lambda_a + u\}^{-\frac{1}{2}(\nu_m+m)}, \quad u \geq 0. \quad (3.47)$$

É claro que estes últimos resultados são diretos, uma vez que se sabe que a distribuição de $(X_1|X_2 = x_2)$ é $t_m(\mu_1(x_2), \Sigma_{11.2}; \lambda_{q(x_2)}, \nu_m)$.

Note-se que a distribuição t usual ($\lambda = \nu$) não preserva suas distribuições condicionais, isto é, se $X = (X_1^\top, X_2^\top)^\top \sim t_n(\mu, \Sigma; \nu)$, onde X_1 é um vetor $m \times 1$, então $(X_1|X_2 = x_2) \sim t_m(\mu_1(x_2), \Sigma_{11.2}; \nu_{q(x_2)}, \nu_m)$, onde $\nu_a = \nu + a$, que é uma versão generalizada da distribuição t m -variada, já que $\nu_a = \nu + a \neq \nu + n - m = \nu_m$ para $a \neq n - m$.

Algumas propriedades adicionais são dadas no que segue.

Seja X um vetor aleatório $n \times 1$. Então:

(d) $X \sim t_n(\mu, \Sigma; \lambda, \nu)$ se e somente se para todo $a(n \times 1) \in \mathbb{R}^n$, $a^\top X \sim t_1(a^\top \mu, a^\top \Sigma a; \lambda, \nu)$.

A prova deste fato segue de observar que para todo $a \in \mathbb{R}^n$, tem-se que $\phi_X(a) = E(e^{ia^\top X}) = \phi_{a^\top X}(1)$. $\square$

(e) Se $X \sim t_n(0, I_n; \lambda, \nu)$ e $A(n \times n)$ é uma matriz simétrica, tem-se que

$$X^\top A X \sim m \frac{\lambda}{\nu} F_{m, \nu}, \quad (3.48)$$

se e somente se $A^2 = A$ e $r(A) = m$.

Com efeito, como $X \stackrel{d}{=} V^{\frac{1}{2}} Z$, onde $V \sim \lambda/\chi_\nu^2$ e $Z \sim N_n(0, I_n)$ são independentes, tem-se que $X^\top A X \stackrel{d}{=} V Z^\top A Z$, com V e $Z^\top A Z$ sendo independentes. Lembrando que $Z^\top A Z \sim \chi_m^2$ se e somente se A é uma matriz idempotente e $r(A) = m$, tem-se a prova.

Alternativamente, como $X \stackrel{d}{=} R U^{(n)} \sim E \ell_n(0, I_n; \phi)$ e $P(X = 0) = 0$. De (2.70) segue que $X^\top A X \stackrel{d}{=} R^2 V_1$ se e somente se $A^2 = A$ e $r(A) = m$, onde $R^2 \stackrel{d}{=} \|X\|^2 \sim (n\lambda/\nu) F_{n, \nu}$ e $V_1 \sim \text{Beta}(m/2, (n-m)/2)$ são independentes. A prova segue de observar que $R^2 V_1 \sim (m\lambda/\nu) F_{m, \nu}$. $\square$

Além disso, se $X \sim t_n(0, I_n; \lambda, \nu)$ e $A = I_n - B$, onde $B^2 = B$ e $r(B) = k$, então, aplicando o Exemplo 2.14, tem-se que

$$\frac{X^\top AX/(n-k)}{X^\top BX/k} \sim F_{n-k,k} . \quad (3.49)$$

Note-se que (3.48) é válida somente na distribuição t esférica multivariada. Entretanto, (3.49) vale para todas as distribuições na classe esférica com massa zero na origem.

3.2.3 Algumas caracterizações da distribuição t dentro da classe elíptica

Como a distribuição normal, a distribuição t generalizada pode ser caracterizada dentro da classe elíptica (ou esférica) em termos de algumas de suas propriedades. Três resultados desta natureza são aqui desenvolvidos. O primeiro deles caracteriza a distribuição t generalizada em termos de suas marginais. O segundo resultado caracteriza esta distribuição através de suas distribuições condicionais. A prova deste fato é feita assumindo a existência de densidade e é similar à prova dada em Kelker (1970) para o caso da distribuição normal. Contudo, o resultado de Kelker (1970) é estendido em Cambanis et al. (1981) para o caso da distribuição normal não degenerada (veja Seção 2.6). Conjeturamos que ele também pode ser mostrado no caso da distribuição t generalizada não degenerada com hipóteses similares às usadas por estes autores. Finalmente, o terceiro resultado mostra que a distribuição t generalizada também pode ser caracterizada em termos de formas quadráticas. Este resultado pode ser derivado usando um fato preliminar estabelecido em Anderson e Fang (1987) para as distribuições esféricas que colocam massa zero na origem. Mas, este fato é provado para distribuição normal sem a última condição em Khatri e Mukerjee (1987) (veja a Seção 2.6). Mostra-se que este último resultado também vale para a distribuição t sem tal condição. A prova de um fato preliminar é requerida para tais efeitos.

Fato 3.3 *Seja $X \sim El(\mu, \Sigma; \phi)$. Então cada distribuição marginal é t generalizada se e somente se X tem distribuição t generalizada.*

A parte “somente se” deste fato está provada no fato (b) da Seção 3.2.2 (veja (3.34)). Para provar a recíproca, consideremos (sem perda de generalidade) o caso esférico com $\mu = 0$ e $\Sigma = I_n$. Isto é, $X \sim El_n(0, I_n; \phi)$ e tem função característica $\phi_X(t) = \phi(t^\top t)$, $t \in \mathbb{R}^n$, para alguma função ϕ . Seja X_1 um subvetor de dimensão $m \times 1$ ($1 \leq m < n$) de X . Então $X_1 \sim El_m(0, I_m; \phi)$ e tem função característica $\phi_{X_1}(t_1) = \phi(t_1^\top t_1)$, $t_1 \in \mathbb{R}^m$, onde a função ϕ é a mesma função que determina a função característica de X . Assumindo que X_1 tem distribuição t generalizada m -variada, em cujo caso $\phi(u)$, $u \geq 0$, é dada por (3.20), tem-se a prova. $\square$

Fato 3.4 *Seja $X \sim El_n(\mu, \Sigma; f)$ e suponha que X , μ e Σ são particionados como em (3.33). Então a distribuição condicional de X_1 , dado $X_2 = x_2$, é t generalizada m -variada ($1 \leq m < n$) se e somente se X tem distribuição t generalizada n -variada ($n \geq 2$).*

A parte “somente se” deste fato está provada no fato (c) da Seção 3.2.2. Para provar a parte “se”, é suficiente considerar o caso esférico com $\mu = 0$ e $\Sigma = I_n$, uma vez que a forma geral da densidade condicional independe de μ e Σ . Em tal caso a densidade condicional de X_1 , dado $X_2 = x_2$, é dada por (veja (2.55))

$$p_{X_1}(x_1|x_2) = f_{\|x_2\|^2}(x_1^\top x_1), \quad x_1 \in \mathbb{R}^m,$$

onde

$$f_a(u) = \frac{f(u+a)}{f_2(a)}, \quad u \geq 0, \quad (a \geq 0), \quad (3.50)$$

com $f_2(x_2^\top x_2)$, $x_2 \in \mathbb{R}^{n-m}$, sendo a densidade marginal de X_2 .

Suponha, então, que $(X_1|X_2 = x_2) \sim t_m(0, I_m; \lambda_{\|x_2\|^2}, \nu_m)$, onde $\lambda_a = \lambda + a$ e $\nu_m = \nu + n - m$. Isto é, $(X_1|X_2 = x_2) \sim El_m(0, I_m; f_{\|x_2\|^2})$, onde $f_a(u)$, $u \geq 0$, é dada em (3.47).

Logo, para cada $a \geq 0$ em (3.50) tem-se que

$$\begin{aligned} f(u+a) &= f_2(a)f_a(u) \\ &= f_2(a)k(m, \nu+n-m)\{\lambda+a\}^{\frac{1}{2}(\nu+n-m)}\{\lambda+a+u\}^{-\frac{1}{2}(\nu+m)} \end{aligned}$$

ou

$$f(v) = f_2(a)k(m, \nu+n-m)\{\lambda+a\}^{\frac{1}{2}(\nu+n-m)}\{\lambda+v\}^{-\frac{1}{2}(\nu+n)}, \quad v \geq 0.$$

Em particular, para $v = a$, tem-se que

$$f_2(a) = \frac{f(a)\{\lambda+a\}^{\frac{1}{2}m}}{k(m, \nu+n-m)},$$

de onde segue que

$$f(v) = f(a)\{\lambda+a\}^{\frac{1}{2}(\nu+n)}\{\lambda+v\}^{-\frac{1}{2}(\nu+n)}, \quad v \geq 0.$$

Mas como $a = \|x_2\|^2$ é fixo e $f(x^\top x)$, $x \in \mathbb{R}^n$, é a densidade de X , nesta última expressão tem-se que

$$1 = \int_{\mathbb{R}^n} f(x^\top x) dx = f(a)\{\lambda+a\}^{\frac{1}{2}(\nu+n)} \int_{\mathbb{R}^n} \{\lambda+x^\top x\}^{-\frac{1}{2}(\nu+n)} dx,$$

o que implica que

$$\begin{aligned} f(a)\{\lambda+a\}^{\frac{1}{2}(\nu+n)} &= \left\{ \int_{\mathbb{R}^n} \{\lambda+x^\top x\}^{-\frac{1}{2}(\nu+n)} dx \right\}^{-1} \\ &= \frac{\Gamma[\frac{1}{2}(\nu+n)]}{\Gamma(\frac{1}{2}\nu)\pi^{\frac{1}{2}n}} \lambda^{\frac{1}{2}\nu} \\ &= k(n, \nu)\lambda^{\frac{1}{2}\nu}. \end{aligned}$$

Portanto,

$$f(v) = k(n, \nu)\lambda^{\frac{1}{2}\nu}\{\lambda+v\}^{-\frac{1}{2}(\nu+n)}, \quad v \geq 0,$$

de onde segue que $X \sim t_n(0, I_n; \lambda, \nu)$. □

No próximo resultado a distribuição t é caracterizada dentro da classe das distribuições esféricas em termos de formas quadráticas. Para provar este resultado o fato que segue será requerido.

Fato 3.5 Seja $Y \sim El_p(0, I_p; \phi)$ e $\Lambda = \text{Diag} \{\lambda_1, \lambda_2, \dots, \lambda_p\}$ com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. Suponha que $Y^\top \Lambda Y \sim (q\lambda/\nu)F_{q,\nu}$. Então, $p = q$ e Y tem densidade $f(y^\top y)$, $y \in \mathbb{R}^p$, isto é, $Y \sim El_n(0, I_p; f)$, com $f(u)$, $u \geq 0$, sendo infinitamente diferenciável.

Com efeito, do fato que $Y^\top \Lambda Y \sim (q\lambda/\nu)F_{q,\nu}$, com $\Lambda > 0$, segue que $P(Y = 0) = 0$. Como $Y \sim El_p(0, I_p; \phi)$ e $P(Y = 0) = 0$, então $S = \|X\|^2$ e $U = X/S^{\frac{1}{2}}$ são independentes (veja o Fato 2.3). Sejam F_S e F_T as funções de distribuições de S e $T = U^\top \Lambda U$, respectivamente. Note-se que $\lambda_1 \geq T \geq \lambda_p > 0$. Como $ST = Y^\top \Lambda Y$ tem distribuição $(q\lambda/\nu)F_{q,\nu}$, tem-se que

$$\int_{\lambda_p}^{\lambda_1} F_S(u/t) dF_T(t) = \int_0^u \frac{1}{\lambda B(\frac{1}{2}q, \frac{1}{2}\nu)} \left\{ \frac{v}{\lambda} \right\}^{\frac{1}{2}q-1} \left\{ 1 + \frac{v}{\lambda} \right\}^{-\frac{1}{2}(1+\nu)} dv, \quad (3.51)$$

onde

$$B = (\frac{1}{2}q, \frac{1}{2}\nu) = \Gamma[\frac{1}{2}(q + \nu)] / \Gamma(\frac{1}{2}q)\Gamma(\frac{1}{2}\nu).$$

Note-se que o lado direito de (3.51) é infinitamente diferenciável em relação a u . Portanto o lado esquerdo de (3.51) também deve ser infinitamente diferenciável em relação a u . Além disto, como o domínio integração do lado esquerdo de (3.51) é finito, é óbvio que $F_S(u/t)$ deve ser infinitamente diferenciável em relação a u . Isto mostra que a densidade de S deve existir e deve ser infinitamente diferenciável. Logo, a densidade de Y a saber, $f(y^\top y)$, $y \in \mathbb{R}^p$, existe e é infinitamente diferenciável, já que a densidade de S é dada por (veja (2.36) na Seção 2.4)

$$p_S(s) = \frac{\pi^{\frac{1}{2}p}}{\Gamma(\frac{1}{2}p)} s^{\frac{1}{2}p-1} f(s), \quad s > 0.$$

Por outro lado, desta última relação juntamente com (3.51) segue que

$$\int_{\lambda_p}^{\lambda_1} \frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{1}{2}p)} u^{\frac{1}{2}p-1} f(u/t) t^{-\frac{1}{2}p} dF_T(t) = \frac{1}{\lambda B(\frac{1}{2}q, \frac{1}{2}\nu)} \left\{ \frac{u}{\lambda} \right\}^{\frac{1}{2}q-1} \left\{ 1 + \frac{u}{\lambda} \right\}^{-\frac{1}{2}(q+\nu)}, \quad (3.52)$$

para todo $u > 0$. Suponha que $q > p$. Então,

$$\int_{\lambda_p}^{\lambda_1} f(u/t) t^{-\frac{1}{2}p} dF_T(t) = \frac{\Gamma[\frac{1}{2}(q + \nu)]}{\Gamma(\frac{1}{2}\nu)\pi^{\frac{1}{2}p}} \lambda^{-\frac{1}{2}p} \left\{ \frac{v}{\lambda} \right\}^{\frac{1}{2}(q-p)} \left\{ 1 + \frac{v}{\lambda} \right\}^{-\frac{1}{2}(q+\nu)}, \quad (3.53)$$

para todo $u > 0$. Note-se que, para $u = 0$, em (3.53) tem-se que $f(0)E(T^{-\frac{1}{2}p}) = 0$, o que é impossível uma vez que $f(0)E(T^{-\frac{1}{2}p}) > 0$. Assim, (3.53) não é possível para $u \rightarrow 0$ e, conseqüentemente, $q \leq p$. Procedendo de forma similar, mostra-se que $q < p$ é também impossível. Logo, deve ter-se que $q = p$. Isto prova o fato. $\square$

Convém notar aqui que o fato anterior é mostrado por Khatri e Murkerjee (1987), mas supondo que $Y^\top \Lambda Y \sim \chi_q^2$ em lugar de que $Y^\top \Lambda Y \sim (q\lambda/\nu)F_{q,\nu}$. Estes autores usam seu resultado para caracterizar a distribuição normal em termos de formas quadráticas (veja o fato (h) na Seção 2.6). Para desenvolver a prova desta caracterização, tais autores utilizam o fato de que a função $f(u)$, $u \geq 0$, seja infinitamente diferenciável. Contudo, da demonstração do fato que segue depreende-se que esta última condição não é requerida.

Fato 3.6 *Suponha que $X \sim El_n(0, I_n; \phi)$ e seja $A(n \times n)$ uma matriz simétrica. Então $X^\top AX \sim (m\lambda/\nu)F_{m,\nu}$ se e somente se $X \sim t_n(0, I_n; \lambda, \nu)$, $A^2 = A$ e $r(A) = m$.*

Com efeito, a condição “se” segue do fato (f) dado na Seção 2.3.2 (veja (3.48)). Para provar a condição “somente se”, note-se em primeiro lugar que $x^\top Ax \geq 0$ para todo $x \in \mathbb{R}^n$, uma vez que $X^\top AX \sim (m\lambda/\nu)F_{m,\nu}$. Seja $k = r(A)$ e seja $\Gamma(n \times k)$ tal que

$$A = \Gamma \Lambda \Gamma^\top \quad \text{e} \quad \Gamma^\top \Gamma = I_k,$$

onde $\Lambda = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$. Seja $Y = \Gamma^\top X$. Então $Y \sim El_k(0, I_k; \phi)$ e $Y^\top \Lambda Y = X^\top AX \sim (m\lambda/\nu)F_{m,\nu}$. Aplicando o Fato 3.5, tem-se que $Y \sim El_k(0, I_k; f_1)$, isto é, Y tem densidade $f_1(y^\top y)$, $y \in \mathbb{R}^k$ e $k = m$. Como $P(Y = 0) = 0$ e $Y^\top \Lambda Y \sim (m\lambda/\nu)F_{m,\nu}$, então (veja Anderson e Fang (1987)) $Y^\top \Lambda Y =^d R_1^2 V_1$, onde $R_1 =^d \|Y\|$ e $V_1 \sim \text{Beta}(m/2, (k-m)/2)$ são independentes. Mas, $k = m$ implica que $V_1 \equiv 1$ e o fato de que $Y^\top \Lambda Y =^d R_1^2 V_1$ implica que $\Lambda^2 = \Lambda$ e $r(\Lambda) = m$ (veja (2.70)), isto é, $\Lambda = I_m$ e $Y^\top \Lambda Y = \|Y\|^2 =^d R_1^2 \sim (m\lambda/\nu)F_{m,\nu}$. Isto mostra que

$Y = \|Y\|(Y/\|Y\|) \stackrel{d}{=} R_1^2 U^{(m)} \sim t_m(0, I_m; \lambda, \nu)$ é uma marginal de $X \sim E\ell_n(0, I_n, \phi)$. Como m é arbitrário, do Fato 3.3 dado acima segue que $X \sim t_n(0, I_n; \lambda, \nu)$. $\square$

Uma prova alternativa segue de usar (2.70) juntamente com o Lema 1 dado em Anderson e Fang (1987). Em tal caso, o Fato 3.5 não é requerido.

Finalmente, note-se que no caso das distribuições normais compostas, o vetor aleatório $X(m \times 1)$ pode ser definido como $X = V^{\frac{1}{2}}Z$, onde $V \sim F_V$ (sobre $[0, \infty)$) e $Z \sim N_n(0, I_n)$ são independentes. Em tal caso, os fatos anteriores podem derivar-se considerando os resultados análogos estabelecidos para a distribuição normal (veja a Seção 2.6).

3.3 Caracterização da distribuição t na classe das distribuições normais compostas: uma interpretação preditivista

3.3.1 Preliminares

Consideremos a classe das distribuições normais compostas descritas no Exemplo 2.6 (veja também o Exemplo 2.14). Cada distribuição nesta classe pode ser obtida como uma mistura da distribuição $N_n(0, VI_n)$ no “parâmetro” de escala V . Assim, se X é um vetor aleatório n -dimensional com distribuição nesta classe, a distribuição de X é determinada especificando uma distribuição para o “parâmetro” de escala V , uma vez que a distribuição condicional de X , dado V , é $N_n(0, VI_n)$. Por exemplo, se $(X|V) \sim N_n(0, VI_n)$ e $V \sim D/\chi_d^2$ (veja (3.19)), então $X \sim t_n(0, I_n; D, d)$. Em geral a especificação de um modelo mediante um procedimento como o descrito acima requer dois estágios: (i) a especificação da distribuição condicional das observáveis (X), dado o “parâmetro” (V); e (ii) a especificação da distribuição do “parâmetro”. Muitos modelos estatísticos são construídos segundo este enfoque (em particular na inferência Bayesiana). Como consequência deste fato, as distribuições que podem ser representadas como mistura de

distribuições paramétricas usuais, em particular, as distribuições normais compostas, têm sido motivo de diversos trabalhos tanto teóricos como aplicados. Na linha teórica, um dos interesses tem sido o estabelecimento de condições (ou postulados) adicionais que permitam identificar a medida da mistura. No contexto mais aplicado encontram-se, por exemplo, estudos de robustificação.

As distribuições normais compostas encontram-se caracterizadas como elementos da classe elíptica em (2.40) (ou em (2.41) no caso de existir densidade; veja também os Exemplos 2.6 e 2.10) e suas propriedades podem derivar-se a partir dos resultados apresentados no Capítulo 2. Algumas representações adicionais em termos de seqüências infinitas de variáveis e vetores aleatórios serão dadas nesta seção. Condições de invariância da seqüência com respeito a certos grupos de transformações ortogonais implicam em que a lei da seqüência é mistura de escala da lei normal, isto é, as leis n -dimensionais são normais compostas (veja Kingman (1972), Smith (1981), Dawid (1978) e Diaconis, Eaton e Lauritzen (1992)). Contudo, este tipo de considerações não permite caracterizar a medida da mistura. Condições adicionais podem ser examinadas para determinar a forma desta medida. Diaconis e Ylvisaker (1985) estabelecem alguns resultados nesta direção. Mediante postulados de simetria e condições de como prever X_{n+1} , dado $X_1, X_2, \dots, X_n$, eles caracterizam as distribuições que são misturas de escala da lei exponencial na distribuição gama. Usando esta idéia juntamente com um fato apresentado em Diaconis e Ylvisaker (1979), no qual caracteriza-se a lei da mistura (ou distribuição a priori) no contexto das famílias exponenciais (regulares) através da condição de linearidade na média a posteriori. Este fato é estabelecido na Seção 3.3.2 e é aplicado para caracterizar a distribuição t dentro da classe das distribuições normais compostas. Na Seção 3.3.3, usa-se este fato juntamente com um fato devido a Kingman (1972) para caracterizar a medida da mistura na representação da lei de uma seqüência infinita de variáveis aleatórias ortogonalmente invariantes (veja a Definição 3.2). Mostra-se que se $X_1, X_2, \dots$

é uma seqüência de variáveis aleatórias ortogonalmente invariantes com $P(X_1 = 0) = 0$ e $E(X_2|X_1) = b + aX_1^2$, para algum $a \in (0, 1)$ e $b > 0$, então para cada n , a lei de $(X_1, X_2, \dots, X_n)^\top$ é $t_n(0, I_n; a^{-1}b, a^{-1}(a+1))$. Posteriormente, considerando seqüências aleatórias que são invariantes sobre transformações ortogonais que preservam o vetor $1_n = (1, \dots, 1)^\top \in \mathbb{R}^n$, mostra-se que condições de linearidade similares às dadas acima implicam que as leis n -dimensionais são mistura de locação da distribuição t . Um resultado devido a Smith (1981) é também usado aqui. Finalmente, considerando alguns resultados estabelecidos em Dawid (1977) e Diaconis, Eaton e Lauritzen (1992), os fatos provados nas seções prévias são estendidos para cobrir o caso de seqüências de vetores aleatórios p -dimensionais ortogonalmente invariantes. Neste caso mostra-se que condições similares implicam que a medida da mistura é a distribuição Wishart invertida.

Para o desenvolvimento destes resultados, se introduz (por conveniência) uma mudança notacional.

Seja $X \sim t_n(m, C; D, d)$, onde m , C , D e d correspondem aos parâmetros μ , Σ , λ e ν , respectivamente, na notação usada nas seções prévias. Suponha que $C(n \times n)$ é definida positiva e que X tem densidade, a qual é dada em (3.21) (com $\mu = m$, $\Sigma = C$, $\lambda = D$ e $\nu = d$). Note-se que para $d > 2$, de (3.29) e (3.30) segue que $E(X) = m$ e $\Sigma_X \equiv \text{Cov}(X) = (d-2)^{-1}DC$. Em tal caso, a densidade de X pode ser expressa como

$$p_X(x|m, C, D, d) = \tilde{k}(n, d)|\Sigma|^{-\frac{1}{2}} \left\{ 1 + \frac{(x-m)^\top C^{-1}(x-m)}{d-2} \right\}^{-\frac{1}{2}(d+m)}, \quad x \in \mathbb{R}^n,$$

onde $\tilde{k}(n, d) = \Gamma[\frac{1}{2}(d+n)]/\Gamma(\frac{1}{2}d)(\pi(d-2))^{\frac{1}{2}n}$. Além disso, se $X = (X_1, X_2, \dots, X_n)^\top \sim t_n(0, I_n; D, d)$, então de (3.43) e (3.44) segue que (desde que existam)

$$E(X_{k+1}|X_k) = 0 \quad \text{e} \quad \text{Var}(X_{k+1}|X_k) = b + aX_k^2, \quad (3.54)$$

onde $a = 1/(d-1)$ e $b = D/(d-1)$.

3.3.2 Caracterização da distribuição t na classe das distribuições normais compostas

Diaconis e Ylvisaker (1979) estabelecem uma caracterização de prioris conjugadas para o parâmetro natural de uma distribuição da forma exponencial regular. Eles mostram que a linearidade na esperança a posteriori implica que a distribuição a priori deve ser conjugada. Este resultado é enunciado no que segue.

Seja μ uma medida σ -finita sobre os conjuntos borelianos de $\mathbb{R}^k$. Seja $\mathcal{I}$ o suporte de μ e assumamos que $\mathcal{I}$ é um subconjunto não vazio e aberto em $\mathbb{R}^k$. Para $\theta \in \mathbb{R}^k$, defina

$$K(\theta) = \log \int e^{\theta y} \mu(dy)$$

e seja $\Theta = \{\theta : K(\theta) < \infty\}$ um subconjunto não vazio e aberto em $\mathbb{R}^k$. A família exponencial $\{P(\cdot|\theta), \theta \in \Theta\}$ de medidas de probabilidades com respeito a μ é definida por

$$P(dy|\theta) = e^{\theta y - K(\theta)} \mu(dy), \quad \theta \in \Theta. \quad (3.55)$$

O parâmetro θ é chamado parâmetro natural. Além disto, se Y é um vetor aleatório k -dimensional com distribuição dada por (3.55), então, derivando a identidade $\int_{\mathcal{I}} P(dy|\theta) = 1$, com respeito a θ e fazendo admissível a troca de derivadas com integrais, tem-se que

$$E(Y|\Theta) = K'(\Theta),$$

onde $K'(\theta) = \left(\frac{\partial K(\theta)}{\partial \theta_1}, \dots, \frac{\partial K(\theta)}{\partial \theta_k} \right)^T$. Aqui Θ denota o vetor aleatório e θ um valor particular de Θ .

Fato 3.7 (Diaconis e Ylvisaker (1979)). *Suponha que Θ é aberto em $\mathbb{R}^k$. Seja Y uma amostra de tamanho um de $P(\cdot|\theta)$ em (3.55) e suponha que o suporte de μ contém um intervalo aberto em $\mathbb{R}^k$. Se Θ tem distribuição (a priori) Π , a qual não é concentrada num ponto e se*

$$E(K'(\Theta)|Y) = b + aY, \quad (3.56)$$

para alguma constante a e um vetor de constantes $b \in \mathbb{R}^k$, então $a \neq 0$, Π é absolutamente contínua com respeito à medida de Lebesgue e tem densidade

$$\Pi(d\theta) = ce^{\frac{1}{a}b \cdot \theta - \frac{(1-a)}{a}K(\theta)} d\theta .$$

Uma primeira aplicação deste fato é dada como um exemplo no que segue.

Exemplo 3.1 Seja X uma variável aleatória. Suponha que $P(X = 0) = 0$ e $(X|V = v) \sim N(0, v)$, onde $V \sim F_V$ é definida sobre $[0, \infty)$. É fácil ver que a condição $P(X = 0) = 0$ implica que $F_V(0) = 0$ e, portanto, X tem densidade (veja (2.41) e o Exemplo 2.6). Seja

$$Y = X^2 \quad \text{e} \quad \Theta = -\frac{1}{2V} .$$

Como a relação entre V e Θ é um a um, então do fato de que $(Y|V = v) \sim v\chi_1^2$, com $0 < v < \infty$, segue que

$$(Y|\Theta = \theta) \sim \chi_1^2/(-2\theta) , \quad \text{com} \quad -\infty < \theta < 0 ,$$

isto é, $(Y|\Theta = \theta) \sim P(\cdot|\theta)$ dada por (3.55) com

$$\left. \begin{aligned} K(\theta) &= -\frac{1}{2} \log(-\theta) , \quad -\infty < \theta < 0 , \\ \mu(dy) &= (\pi y)^{-\frac{1}{2}} I_{(0, \infty)}(y) dy . \end{aligned} \right\} \quad (3.57)$$

É fácil verificar que a medida μ dada em (3.57) é σ -finita sobre os conjuntos borelianos em $\mathbb{R}$. Além disto, $\Theta = (-\infty, 0) \subset \mathbb{R}$ é não vazio e aberto e

$$E(Y|\Theta) = K'(\Theta) = -\frac{1}{2\Theta} = V .$$

Se, adicionalmente, incorporamos a condição (3.56), isto é, se

$$E\{E(Y|\Theta)|Y\} = E\{(-2\Theta)^{-1}|Y\} = E\{V|Y\} = b + aY ,$$

então, aplicando o Fato 3.7 tem-se que

$$\Pi(d\theta) = c(-\theta)^{\{(1+a)/2a\}-1} e^{(b/a)\theta} d\theta , \quad -\infty < \theta < 0 .$$

Em outras palavras, $(-\Theta) \sim (2a/b)\chi_{\{(a+1)/a\}}^2$, de onde segue que $V = 1/(-2\Theta) \sim (b/a)/\chi_{\{(a+1)/a\}}^2$. Isto é,

$$X \sim t_1(0, 1; b/a, (a+1)/a) ,$$

onde é fácil ver que $0 < a < 1$ e $b > 0$. Note-se que para $b = a + 1$, tem-se a distribuição t usual.

O Exemplo 3.1 mostra que se $X_1, X_2, \dots, X_n$ e V são variáveis aleatórias tais que, dado $V = v$, $X_1, X_2, \dots, X_n$ são iid $N(0, v)$ e $P(X_1 = 0) = 0$, onde a variável aleatória V tem distribuição F_V sobre $[0, \infty)$ tal que

$$E(V|X_1) = b + aX_1^2 , \quad 0 < a < 1 \quad \text{e} \quad b > 0 , \quad (3.58)$$

então

$$X = (X_1, X_2, \dots, X_n)^\top \sim t_n(0, I_n; b/a, (a+1)/a) .$$

A prova deste resultado segue de aplicar o Exemplo 3.1 (ou o Fato 3.7) para $Y_1 = X_1^2$, uma vez que $(Y_1|V = v) \sim v\chi_1^2$. De (3.34) segue que a recíproca também vale. Consequentemente, este resultado caracteriza a distribuição t generalizada n -variada dentro da classe das distribuições normais compostas.

Por outro lado, como $(X_i|V = v)$, $i = 1, 2, \dots, n$, são iid $N(0, v)$, então $E(X_i^2|V = v) = v$, $i = 1, 2, \dots, n$, de onde segue que

$$E(V|X_1) = E\{E(X_2^2|V)|X_1\} = E\{E(X_2^2|V, X_1)|X_1\} = E(X_2^2|X_1) . \quad (3.59)$$

Este fato permite estabelecer a condição (3.58) em termos das observáveis. Assim, a caracterização dada acima da distribuição t generalizada n -variada pode ser expressa como segue.

Fato 3.8 *Seja $X = (X_1, X_2, \dots, X_n)^\top$ um vetor aleatório com distribuição normal composta. Então $P(X_1 = 0) = 0$ e*

$$E(X_2^2|X_1) = b + aX_1^2 , \quad (3.60)$$

se e somente se $X \sim t_n(0, I_n; b/a, (a+1)/a)$, com $0 < a < 1$ e $b > 0$.

Com efeito, suponha que $X = (X_1, X_2, \dots, X_n)^\top \sim t_n(0, I_n; b/a, (a+1)/a)$, $0 < a < 1$, $b > 0$. Como X tem densidade, então $P(X_1 = 0) = 0$. A relação (3.60) segue de (3.54). Reciprocamente, como $X = (X_1, X_2, \dots, X_n)^\top$ tem distribuição normal composta, então $(X|V = v) \sim N_n(0, vI_n)$ para alguma variável aleatória $V \geq 0$. Como $P(X_1 = 0) = 0$, então $V > 0$ com probabilidade um. Observando finalmente que (3.60) implica que

$$\begin{aligned} E\{E(X_1^2|V)|X_1^2\} &= E\{E(X_2^2|V)|X_1^2\} = E\{E(X_2^2|V, X_1)|X_1^2\} \\ &= E\{X_2^2 | X_1^2\} = E\{E(X_2^2|X_1)|X_1^2\} \\ &= b + aX_1^2, \end{aligned}$$

a prova segue de aplicar o Fato 3.7. □

Do Fato 3.8 segue, também, que se $Y_i = X_i^2$, $i = 1, 2, \dots, n$, então a distribuição de $Y = (Y_1, Y_2, \dots, Y_n)^\top$ pode ser representada como mistura de escala da distribuição gama na distribuição gama invertida. Além disto, Y tem distribuição n -variada com densidade dada por

$$p_Y(y) = \frac{\Gamma[(a+1)/2a + n/2]}{\Gamma[(a+1)/2a] \Pi^{\frac{1}{2}n}} (b/a)^{(a+1)/2a} \prod_{i=1}^n y_i^{\frac{1}{2}-1} \left\{ b/a + \sum_{i=1}^n y_i \right\}^{-\frac{1}{2}((a+1)/a+n)},$$

para $y_1, y_2, \dots, y_n > 0$.

Esta distribuição é considerada como uma versão multivariada da distribuição Beta II (veja Olkin e Rubin (1964) e Tan (1969)), é denominada Dirichlet invertida n -variada e é denotada por $(b/a)DI_n(1/2, 1/2, \dots, 1/2; (a+1)/2a)$ em Tiao e Guttman (1963).

3.3.3 Caracterização da distribuição t multivariada a partir de seqüências de variáveis aleatórias

Seja O_n o grupo das matrizes ortogonais de dimensão $n \times n$. Uma seqüência infinita de variáveis aleatórias, digamos, $X_1, X_2, \dots$, é dita ser ortogonalmente invariante se para cada

n , $X^{(n)} = (X_1, X_2, \dots, X_n)^\top$ é O_n -invariante, isto é, $\Gamma X^{(n)} \stackrel{d}{=} X^{(n)}$ para todo $\Gamma \in O_n$. Por exemplo, se $Z_1, Z_2, \dots$ é uma seqüência de variáveis aleatórias iid $N(0, \sigma^2)$, então $Z_1, Z_2, \dots$ é ortogonalmente invariante. Se P_σ denota a medida de probabilidade sobre $\mathbb{R}^\infty$ de $Z_1, Z_2, \dots$ e Π é uma medida de probabilidade em $[0, \infty)$, então

$$P_\Pi = \int_{[0, \infty)} P_\sigma \Pi(d\sigma)$$

é ortogonalmente invariante. Reciprocamente, toda seqüência infinita de variáveis aleatórias ortogonalmente invariante podem ser representadas desta maneira. Esta representação tem sido provada em diferentes contextos (Schoenberg (1938), Kelker (1970), Kingman (1972), Diaconi e Freedman (1987) e Vitale (1993)). Uma interessante revisão bibliográfica é dada em Eaton (1989). Uma representação equivalente deste resultado está dada em (2.40). O resultado que segue é devido a Kingman (1972).

Fato 3.9 (Kingman (1972)) *Seja $X_1, X_2, \dots$ uma seqüência infinita de variáveis aleatórias ortogonalmente invariantes, então existe uma variável aleatória não-negativa V tal que, dado V , $X_1, X_2, \dots$ são iid $N(0, V)$.*

A prova deste resultado encontra-se também em Fang et al. (1990). Uma prova diferente é dada em Vitale (1993).

A variável aleatória V definida no Fato 3.9 pode ser interpretada de diferentes maneiras, por exemplo, da lei forte dos grandes números segue que $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^2 = V$, com probabilidade um.

Nota 3.2 Se no Fato 3.9 coloca-se a condição que $X_1 = 0$ com probabilidade zero, então $\Pi\{\sigma : \sigma > 0\} = 1$ (Eaton (1981)), onde Π é a lei de V . Este resultado segue da identidade

$$P(X_1 = 0) = \int P_\sigma(X_1 = 0) \Pi(d\sigma)$$

com $P_\sigma(X_1 = 0) = I_{\{0\}}(\sigma)$.

Considerando os Fatos 3.8 e 3.9, o seguinte resultado vem direto.

Fato 3.10 *Seja $X_1, X_2, \dots$ uma seqüência infinita de variáveis aleatórias ortogonalmente invariante tal que $X_1 = 0$ com probabilidade um. Se $\text{Var}(X_2|X_1) = b + aX_1^2$, então, para cada n , $X^{(n)} = (X_1, X_2, \dots, X_n)^\top \sim t_n(0, I_n; a/b, (a+1)/a)$, com $0 < a < 1$ e $b > 0$. A recíproca também vale.*

Com efeito, do Fato 3.9 segue que, para cada n , a distribuição condicional de $X^{(n)}$, dado $V = v$, é $N_n(0, vI_n)$, isto é, $X^{(n)}$ tem distribuição normal composta. Usando o Fato 3.8 tem-se a prova. $\square$

Os resultados e a notação que seguem serão usados para caracterizar a classe de variáveis aleatórias ortogonalmente invariante que são mistura de locação da lei t multi-variada não-centrada.

Seja $O_n(1_n) = \{\Gamma \in O_n : \Gamma 1_n = 1_n\}$, onde $1_n = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$.

Fato 3.11 *(Smith (1981)) Seja $X_1, X_2, \dots$ uma seqüência de variáveis aleatórias reais tal que, para cada n , $\Gamma X^{(n)} \stackrel{d}{=} X^{(n)}$ para todo $\Gamma \in O_n(1_n)$, onde $X^{(n)} = (X_1, X_2, \dots, X_n)^\top$, então, existe variáveis aleatórias reais M e V , com V sendo não-negativa, tal que, dados M e V , $X_1, X_2, \dots$ são iid $N(M, V)$.*

Se $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ e $S_n^2 = (1/n) \sum_{i=1}^n (X_i - \bar{X}_n)^2$, então os “parâmetros” M e V podem ser interpretados como

$$\lim_{n \rightarrow \infty} \bar{X}_n = M \quad \text{e} \quad \lim_{n \rightarrow \infty} S_n^2 = V,$$

com probabilidade um. Note-se que para cada n , $\bar{X}_n$ e S_n^2 são condicionalmente independentes, dado (M, V) .

Nota 3.3 Usando argumentos similares à Nota 3.2, pode-se mostrar que se no Fato 3.11 se introduz a condição adicional que $X_1 = X_2$ com probabilidade zero, então $V > 0$ com probabilidade um.

Por outro lado, sejam $X_1, X_2, \dots, M$ e V definidas como no Fato 3.11 e considere a seqüência definida por $Y_i = X_i - M$, $i = 1, 2, \dots$. Então Y_1 e M são independentes se e somente se V e M são independentes. Com efeito, suponha (por simplicidade) que $V = 0$ com probabilidade zero (veja as Notas 3.2 e 3.3). Neste caso, Y_1 tem densidade dada por

$$p_{Y_1}(y_1) = \int_0^\infty (2\pi v)^{-\frac{1}{2}} e^{-\frac{1}{2v}y_1^2} \Pi(dv), \quad (3.61)$$

onde $\Pi(\cdot)$ denota a distribuição de V . Similarmente, dado $M = m$, Y_1 tem densidade dada por

$$p_{Y_1}(y_1|m) = \int_0^\infty (2\pi v)^{-\frac{1}{2}} e^{-\frac{1}{2v}y_1^2} \Pi_m(dv), \quad (3.62)$$

onde $\Pi_m(\cdot) = \Pi(\cdot|m)$ denota a distribuição condicional de V dado $M = m$. Suponha que V é independente de M . Então (3.62) se reduz a (3.61), isto é, $p_{Y_1}(y_1|m) = p(y_1)$, para todo m em $\mathbb{R}$. Isto implica que Y_1 é independente de M . Reciprocamente, se Y_1 é independente de M , então as expressões da direita de (3.61) e (3.62) devem ser iguais. Isto se cumpre somente se $\Pi_m(\cdot) = \Pi(\cdot)$, para todo m em $\mathbb{R}$, o qual implica que V é independente de M . Contudo, do fato que, dado M e V , $Y_1 \sim N(0, V)$, a qual não depende de M , segue que, dado V , Y_1 e M são independentes. Similarmente, observando que (3.61) implica que

$$Y_1 = {}^d V^{\frac{1}{2}} Z \quad \text{ou} \quad Y_1^2 = {}^d V Z^2,$$

onde $Z \sim N(0, 1)$ ou $Z^2 \sim \chi_1^2$ é independente de V e M , é fácil ver que (veja o Fato 2.5) a distribuição condicional de V , dado Y_1 e M , não depende de M e depende de Y_1 somente através de Y_1^2 . Isto é, dado Y_1 , V é independente de M e $(V|Y_1 = y_1) = {}^d (V|Y_1^2 = y_1^2)$.

Fato 3.12 *Sejam $X_1, X_2, \dots, M$ e V definidos como no Fato 3.11 e suponha que $X_1 = X_2$ com probabilidade zero. Considere a seqüência definida por $Y_i = X_i - M$, $i = 1, 2, \dots$ e assuma que $\text{Var}(Y_2|Y_1) = b + aY_1^2$. Então a lei de $X^{(n)} = (X_1, X_2, \dots, X_n)^\top$ é mistura no parâmetro de locação da lei $t_n(m1_n, I_n; b/a, (a+1)/a)$. Além disto, M e V são independentes.*

Com efeito, suponha que $\text{Var}(Y_2|Y_1) = E(Y_2^2|Y_1) = b + aY_1^2$. Notando que $E(V|Y_1, M) = E(V|Y_1^2)$ e usando as propriedades de esperança condicional, tem-se que

$$E\{E(Y_1^2|V, M)|Y_1^2, M\} = b + aY_1^2. \quad (3.63)$$

Seja P_m a lei condicional de $Y_1, Y_2, \dots$, dado $M = m$, e denote por E_m ao operador esperança com relação à lei P_m . Então, de (3.63), segue que

$$E_m\{E_m(Y_1^2|V)|Y_1^2\} = b + aY_1^2.$$

Seja, agora, $P_m^{(n)}(\cdot|v)$ a P_m -lei de $X^{(n)}$. Isto é, $P_m^{(n)}(\cdot|v)$ é a lei $N_n(m1_n, vI_n)$. Como, para cada m , a seqüência $X_1 - m, X_2 - m, \dots$ é ortogonalmente invariante em relação a P_m , então do Fato 3.10 tem-se que

$$\Pi_m(dv) = cv^{-(a+1)/2a-1} e^{-b/2a} I_{(0,\infty)}(v) dv.$$

Em outras palavras, dado $M = m$, V tem distribuição $(b/a)/\chi_{(a+1)/a}^2$. Isto é,

$$\int P_m^{(n)}(\cdot|v) \Pi_m(dv),$$

é a lei $t_n(m1_n, I_n; b/a, (a+1)/a)$. Finalmente, do fato que a lei condicional de V , dado $M = m$, não depende de m , segue que M e V são independentes. $\square$

3.3.4 Caracterização da matriz t variada a partir de seqüências de vetores aleatórios

Nesta seção estende-se os resultados da seção anterior para o caso matricial. Isto é, caracteriza-se a distribuição conjunta de n vetores aleatórios p -dimensionais com condições similares às usadas no caso 1-dimensional, desenvolvido na Seção 3.3.3. A notação matricial das distribuições normal (np) -variada (ou matriz normal) e Wishart $(p(p+1)/2)$ -variada (ou matriz Wishart) são incorporadas (veja, por exemplo, Muirhead (1982)). Seguindo Dickey (1967) (veja, também, Dawid (1981)) uma definição da distribuição t (np) -variada (ou matriz t) é dada aqui assumindo existência de densidade.

Sejam $X_1, X_2, \dots, X_n$ vetores aleatórios de dimensão $p \times 1$ e considere-se a matriz aleatória definida por

$$\mathbf{X} = \begin{pmatrix} X_1^\top \\ X_2^\top \\ \vdots \\ X_n^\top \end{pmatrix}.$$

A distribuição da matriz aleatória $\mathbf{X}$ é dada pela distribuição conjunta de $X_1, X_2, \dots, X_n$, isto é, pela distribuição do vetor definido por

$$\text{vec}(\mathbf{X}^\top) = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

Sejam $\mathbf{C} = [c_{ij}]$, $\mathbf{D} = [d_{ij}]$ duas matrizes de dispersão de dimensões $n \times n$ e $p \times p$, respectivamente, e seja $\mathbf{M}$ uma matriz de locação de dimensão $n \times p$. A matriz $\mathbf{M}$ pode apresentar-se como

$$\mathbf{M} = \begin{pmatrix} m_1^\top \\ m_2^\top \\ \vdots \\ m_n^\top \end{pmatrix},$$

onde $m_1, m_2, \dots, m_n$ são vetores de locação de dimensão $p \times 1$. Com esta notação a distribuição de $\mathbf{X}$ é dita ser a matriz normal, e se escreve $\mathbf{X} \sim N(\mathbf{M}, \mathbf{C}, \mathbf{D})$, se $\text{vec}(\mathbf{X}^\top) \sim N_{np}(\text{vec}(\mathbf{M}^\top), \mathbf{C} \otimes \mathbf{D})$, onde $(\mathbf{C} \otimes \mathbf{D})_{ir, js} = c_{ij}d_{rs}$.

Se $\mathbf{C}$ e $\mathbf{D}$ são matrizes definidas positivas, então $\mathbf{X}$ tem densidade dada por

$$p_{\mathbf{X}}(\mathbf{X}|\mathbf{C}, \mathbf{D}) = (2\pi)^{-\frac{1}{2}np} |\mathbf{C}|^{-\frac{1}{2}p} |\mathbf{D}|^{-\frac{1}{2}n} e^{\text{tr}\{-\frac{1}{2}\mathbf{C}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{D}^{-1}(\mathbf{X}-\mathbf{M})^\top\}}.$$

Similarmente, a distribuição de uma matriz aleatória $\mathbf{X}(n \times p)$ é dita ser a matriz t , com matriz de locação $\mathbf{M}(n \times p)$, matrizes de dispersão $\mathbf{C}(n \times n)$ e $\mathbf{D}(p \times p)$ ($\mathbf{C}$ e $\mathbf{D}$ sendo matrizes definidas positivas) e parâmetro de forma d ($d + (n-1)(p-1) > 0$), se $\mathbf{X}$ tem densidade dada por

$$p_{\mathbf{X}}(\mathbf{X}|\mathbf{M}, \mathbf{C}, \mathbf{D}, d) = k(n, d) |\mathbf{D}|^{\frac{1}{2}(d+n(p-1))} |\mathbf{C}|^{-\frac{1}{2}p} |\mathbf{D} + (\mathbf{X} - \mathbf{M})^\top \mathbf{C}^{-1}(\mathbf{X} - \mathbf{M})|^{-\frac{1}{2}(d+np)},$$

onde

$$k(n, d) = \frac{\Gamma_p[\frac{1}{2}(d + np)]}{\Gamma_p[\frac{1}{2}(d + n(p - 1))]\pi^{\frac{1}{2}np}} \quad \text{e} \quad \Gamma_p(\frac{1}{2}\alpha) = \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^p \Gamma[\frac{1}{2}(\alpha - i + 1)] .$$

Neste caso, se escreve $\mathbf{X} \sim t(\mathbf{M}, \mathbf{C}; \mathbf{D}, d)$. Esta notação difere da notação de Dickey (1967) e Dawid (1981). Com a notação de Dickey (1967) e Dawid (1981), a distribuição definida acima deve denotar-se por $t(\mathbf{C}, \mathbf{D}, \mathbf{M}, d + np)$ e $t(d + (n - 1)(p - 1); \mathbf{M}, \mathbf{C}, \mathbf{D})$, respectivamente. A maior diferença vem da forma de escrever os graus de liberdade.

Seja agora $\mathbf{W} = [W_{ij}]$ uma matriz aleatória simétrica de dimensão $p \times p$. Neste caso, a distribuição de $\mathbf{W}$ é dada pela distribuição conjunta de seus $p(p + 1)/2$ elementos diferentes, digamos, $W_{11}, W_{12}, \dots, X_{1p}, W_{2p}, \dots, W_{pp}$. Se $X(n \times p) \sim N(0, \mathbf{I}_m, \mathbf{D})$ e $\mathbf{W} = \mathbf{X}^T \mathbf{X}$, a distribuição de $\mathbf{W}$ é chamada matriz Wishart e se escreve $\mathbf{W} \sim W_p(\mathbf{D}, n)$. Se $\mathbf{W} \sim W_p(\mathbf{D}, \alpha)$ com $D > 0$ e $p \leq \alpha$, então $\mathbf{W}$ tem densidade dada por

$$p_{\mathbf{W}}(\mathbf{W}) = \frac{|\mathbf{D}|^{-\frac{1}{2}\alpha}}{2^{\frac{1}{2}\alpha p} \Gamma_p(\frac{1}{2}\alpha)} |\mathbf{W}|^{\frac{1}{2}(\alpha - p - 1)} e^{\text{tr}\{-\frac{1}{2}\mathbf{D}^{-1}\mathbf{W}\}} , \quad \mathbf{W} > 0 .$$

Se $\mathbf{W} \sim W_p(\mathbf{D}, \alpha)$ é não singular, a distribuição de $\mathbf{V} = \mathbf{W}^{-1}$ é chamada Wishart invertida. Tal distribuição é denotada por $WI_p(\mathbf{D}, \alpha)$ e tem densidade dada por

$$p_{\mathbf{V}}(\mathbf{V}) = \frac{|\mathbf{D}|^{\frac{1}{2}\alpha}}{2^{\frac{1}{2}\alpha p} \Gamma_p(\frac{1}{2}\alpha)} |\mathbf{V}|^{-\frac{1}{2}(\alpha + p + 1)} e^{\text{tr}\{-\frac{1}{2}\mathbf{D}\mathbf{V}^{-1}\}} , \quad \mathbf{V} > 0 . \quad (3.64)$$

Em caso contrário, esta densidade é igual a zero.

É fácil ver que (veja Dickey (1967)) a matriz $t(0, \mathbf{I}_n; \mathbf{D}, d)$ é mistura da matriz $N(0, \mathbf{I}_n; \mathbf{V})$ na matriz de escala $\mathbf{V}$, onde $\mathbf{V} \sim WI_p(\mathbf{D}, d + n(p - 1))$.

Com a finalidade de estender os resultados apresentados na Seção 3.3.3, a seguinte notação adicional é requerida.

Seja $X_1, X_2, \dots$ uma sequência infinita de vetores aleatórios com dimensão $p \times 1$. Para cada k , considere a matriz aleatória $k \times p$ definida por

$$\mathbf{X}^{(k)} = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_k^T \end{pmatrix} .$$

Seja O_k o grupo das matrizes ortogonais $k \times k$ que atuam sobre $\mathbf{X}^{(k)}$ pela esquerda. A seqüência de vetores aleatórios $X_1, X_2, \dots$ é dita ser ortogonalmente invariante se para cada k , a lei de $\mathbf{X}^{(k)}$, digamos $P^{(k)}$, é O_k invariante, isto é, $\Gamma \mathbf{X}^{(k)} =^d \mathbf{X}^{(k)}$ para todo Γ em O_k . Sejam, agora, P a lei da seqüência $X_1, X_2, \dots$, Q_Σ a lei da seqüência $Z_1, Z_2, \dots$, onde os Z_i 's são iid $N_p(0, \Sigma)$ e $Q_\Sigma^{(n)}$ a lei $N(0, I_n, \Sigma)$. Com esta notação, Dawid (1978) prova o seguinte fato.

Fato 3.13 (Dawid (1978)) *Se $X_1, X_2, \dots$ é uma seqüência de vetores aleatórios $p \times 1$ ortogonalmente invariantes, então existe uma única medida de probabilidade μ definida sobre os conjuntos borelianos em $S_p^+ = \{\mathbf{A}(p \times p) : \mathbf{A} \geq 0\}$, tal que*

$$P = \int Q_\Sigma \mu(d\Sigma) .$$

Em particular,

$$P^{(k)} = \int Q_\Sigma^{(k)} \mu(d\Sigma) .$$

A recíproca também é verdadeira.

Em outras palavras, se $X_1, X_2, \dots$ é ortogonalmente invariante, então existe uma matriz aleatória $\mathbf{V}(p \times p) \geq 0$, tal que dado $\mathbf{V}$, $X_1, X_2, \dots$ são iid $N_p(0, \mathbf{V})$. Da lei forte dos grandes números, segue que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i X_i^\top = \mathbf{V} ,$$

com probabilidade um.

Nota 3.4 Se no Fato 3.13 se inclui a condição que $X_1, X_2, \dots, X_p$ sejam linearmente dependentes com probabilidade zero, então

$$\mu(\{\Sigma; \Sigma > 0\}) = 1 ,$$

onde μ é a lei de $\mathbf{V}$. Este resultado é similar ao dado na Nota 3.2, e vem do fato que se $Z_1, Z_2, \dots, Z_n$ são vetores aleatórios iid $N_p(0, \Sigma)$, onde $\Sigma > 0$, então $Z_1, Z_2, \dots, Z_p$ são

linearmente dependentes com probabilidade zero (veja Muirhead (1982, pp. 82, 83)). O seguinte fato estende o Fato 3.11.

Fato 3.14 *Sejam $X_1, X_2, \dots$ e $\mathbf{V}$ definidos como no Fato 3.13 e para cada k fixo, seja*

$$\mathbf{X}_i^{(k)} = \begin{pmatrix} X_{(i-1)k+1}^\top \\ \vdots \\ X_{ik}^\top \end{pmatrix}, \quad i = 1, 2, \dots$$

Se $X_1, X_2, \dots, X_p$ são linearmente dependentes com probabilidade zero e

$$E(\mathbf{X}_2^{(p)\top} \mathbf{X}_2^{(p)} | \mathbf{X}_1^{(p)}) = \mathbf{B} + a \mathbf{X}_1^{(p)\top} \mathbf{X}_1^{(p)}, \quad (3.65)$$

então $\mathbf{X}^{(n)} \sim t(0, I_n; a^{-1}\mathbf{B}, a^{-1}(a+p) - n(p-1))$, $0 < a < 1$ e $\mathbf{B}(p \times p) > 0$.

Com efeito, do Fato 3.13 e a primeira condição segue que, para cada n , a distribuição condicional de $\mathbf{X}^{(n)}$, dado $\mathbf{V}$, é $N(0, I_n, \mathbf{V})$, onde $\mathbf{V} > 0$ com probabilidade um. Seja $\mathbf{Y}_i = \mathbf{X}_i^{(p)\top} \mathbf{X}_i^{(p)}$, $i = 1, 2, \dots$. Então, dado $\mathbf{V}$, $\mathbf{Y}_i$'s são iid $W_p(\mathbf{V}, p)$. Isto é, dado $\mathbf{V}$, $\mathbf{Y}_1 > 0$ com probabilidade um (veja a Nota 3.4) e tem densidade dada por (3.64) com $\mathbf{W} = \mathbf{Y}_1$, $\mathbf{D} = \mathbf{V}$ e $\alpha = p$. Seja $\boldsymbol{\Theta} = (-2\mathbf{V})^{-1}$. Então, a distribuição de $\mathbf{Y}_1$ dado $\boldsymbol{\Theta}$ pode ser expressa como

$$P_{\boldsymbol{\Theta}}(d\mathbf{Y}_1) = e^{\text{tr}\{\boldsymbol{\Theta}\mathbf{Y}_1\} + \frac{1}{2}p \log |-\boldsymbol{\Theta}|} \mu(d\mathbf{Y}_1),$$

onde

$$\mu(d\mathbf{Y}_1) = \frac{|\mathbf{Y}_1|^{-\frac{1}{2}}}{\Gamma(\frac{1}{2}p)} d\mathbf{Y}_1.$$

Por outro lado, das propriedades da esperança condicional juntamente com (3.65), segue que

$$\begin{aligned} E\{E[\mathbf{Y}_1 | \boldsymbol{\Theta}] | \mathbf{Y}_1\} &= E\{E[\mathbf{Y}_1 | \mathbf{V}] | \mathbf{Y}_1\} \\ &= E\{E[\mathbf{Y}_2 | \mathbf{V}] | \mathbf{Y}_1\} \\ &= E\{E[\mathbf{Y}_2 | \mathbf{Y}_1, \mathbf{V}] | \mathbf{Y}_1\} \\ &= E\{\mathbf{Y}_2 | \mathbf{Y}_1\} \\ &= \mathbf{B} + a\mathbf{Y}_1. \end{aligned}$$

Aplicando o Fato 3.7 (com $k = p(p+1)/2$), tem-se que

$$\Pi(d\Theta) = c|-\Theta|^{\frac{1}{2}pa^{-1}(a-1)}e^{\text{tr}\{a^{-1}\mathbf{B}\Theta\}}d\Theta,$$

o qual implica que $(-\Theta) \sim W_p(2a^{-1}\mathbf{B}, a^{-1}(a+p))$ e, portanto, que $\mathbf{V} = (-2\Theta)^{-1} \sim WI_p(a^{-1}\mathbf{B}, a^{-1}(a+p))$ (as condições $0 < a < 1$ e $\mathbf{B} > 0$ são diretas). Isto prova o fato. $\square$

O análogo multivariado do resultado dado no Fato 3.12 é dado no Fato 3.16 (abaixo). Alguns resultados preliminares devidos a Diaconis, Eaton e Lauritzen (1992) são requeridos.

Sejam $X_1, X_2, \dots, X_k$ vetores aleatórios $p \times 1$ e seja $\mathbf{X}^{(k)}$ a matriz aleatória $k \times p$, cujas filas são dadas por $X_1^\top, X_2^\top, \dots, X_k^\top$. A lei de $\mathbf{X}^{(k)}$ é dita ser $O_k(1_k)$ -invariante se $\Gamma\mathbf{X}^{(k)} =^d \mathbf{X}^{(k)}$ para todo Γ em $O_k(1_k)$. Seja $P^{(n)}$ a lei de $\mathbf{X}^{(n)}$ e

$$Q_\mu^{(k)} = \int N(1_k m^\top, I_k, \Sigma) \mu(dm, d\Sigma),$$

onde $m(p \times 1)$ é um vetor fixo em R^P .

Fato 3.15 (Diaconis, Eaton e Lauritzen (1992)) *Sejam $P^{(k)}$ e $Q^{(k)}$ as medidas definidas acima. Então existe uma medida de probabilidade μ sobre $R^p \times S_p^+$ tal que para cada k , com $p+k+2 < n-1$, a distância de variação entre $P^{(k)}$ e $Q_\mu^{(k)}$ é limitada superiormente por*

$$2 \left\{ \left(1 - \frac{k+p+2}{n-1} \right)^{-c} - 1 \right\} + 2 \left\{ \left(\frac{n}{n-2} \right)^{\frac{1}{2}p} - 1 \right\},$$

onde $c = t^2/2$ e $t = \min\{k, p\}$.

Seja agora $X_1, X_2, \dots$ uma seqüência infinita de vetores aleatórios de dimensão $p \times 1$, tal que $P^{(n)}$, a lei de $\mathbf{X}^{(n)}$, é $O_n(1_n)$ -invariante. Seja P a lei de $X_1, X_2, \dots$. Suponha que $E(X_1) < \infty$ e $E(X_1 X_1^\top) < \infty$. Com estas condições, uma versão infinita do Fato 3.15 pode ser obtida usando argumentos similares ao dado em Diaconis et al. (1992). De fato,

como a seqüência $\{X_n : n \in \mathbb{N}\}$ é permutável e como $E(X_1) < \infty$ e $E(X_1 X_1^\top) < \infty$, então existe uma σ -álgebra τ de eventos tal que (veja Chow e Taicher (1978))

$$\lim_{n \rightarrow \infty} \bar{X}_n = M \equiv E(X_1 | \tau)$$

e

$$\lim_{n \rightarrow \infty} S_n = V \equiv E(X_1 X_1^\top | \tau) - E(X_1 | \tau)(E(X_1 | \tau))^\top,$$

com probabilidade um, onde

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad S_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^\top.$$

Denotando por μ_n a P -lei de $(\bar{X}_n, S_n)$ e usando o Fato 3.15, segue que $P^{(k)} = Q_\mu^{(k)}$, para todo k , onde μ é a lei de (M, V) .

Sejam, agora,

$$\mathbf{X}_k^{(p)} = \begin{pmatrix} X_{(k-1)p}^\top \\ \vdots \\ X_{kp}^\top \end{pmatrix} \quad \text{e} \quad \mathbf{Z}^{(p)} = \mathbf{X}_1^{(p)} - \mathbf{X}_p^{(p)} = \begin{pmatrix} Z_1^\top \\ Z_2^\top \\ \vdots \\ Z_p^\top \end{pmatrix},$$

onde $Z_i = X_i - X_{i+1}$, $i = 1, 2, \dots, p$. Com argumentos similares aos usados na Nota 3.4, segue que se $Z_1, Z_2, \dots, Z_p$ são linearmente dependentes com a probabilidade zero, então

$$\mu(R^P \times \{\Sigma : \Sigma > 0\}) = 1.$$

Fato 3.16 *Seja $X_1, X_2, \dots$ uma seqüência infinita de vetores aleatórios de dimensão $p \times 1$ tal que para cada n , $\mathbf{X}^{(n)}$ é $O_n(1_n)$ -invariante. Sejam M , V e $\mathbf{Z}^{(p)}$ definidos como acima e considere a seqüência de matrizes $p \times p$ definidas por $\mathbf{Y}_i^{(p)} = (\mathbf{X}_i^{(p)} - 1_p M^\top)^\top (\mathbf{X}_i^{(p)} - 1_p M^\top)$, $i = 1, 2, \dots$. Se*

$$E(\mathbf{Y}_2^{(p)\top} \mathbf{Y}_2^{(p)} | \mathbf{Y}_1^{(p)}) = \mathbf{B} + a \mathbf{Y}_1^{(p)\top} \mathbf{Y}_1^{(p)},$$

então $\mathbf{X}^{(n)}$ é mistura no parâmetro de locação (M) da matriz $t(1_n M^\top, I_n; a^{-1} \mathbf{B}, a^{-1}(a + p) - n(p - 1))$, $0 < a < 1$ e $\mathbf{B}(p \times p) > 0$.

Com efeito, com os mesmos argumentos considerados na prova do Fato 3.12, é fácil ver que

$$E\{E(\mathbf{Y}_1^{(p)\top} \mathbf{Y}_1^{(p)} | M, \mathbf{V}) | \mathbf{Y}_1^{(p)\top} \mathbf{Y}_1^{(p)}\} = \mathbf{B} + a \mathbf{Y}_1^{(p)\top} \mathbf{Y}_1^{(p)} . \quad (3.66)$$

Sejam P_m a lei de $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, dado $M = m$ e $E_m(\cdot)$ o operador esperança relativo a P_m . Então, a P_m -lei de $\mathbf{Y}_1^{(p)} = \mathbf{X}_1^{(p)} - \mathbf{1}_p m^\top$, dado $\mathbf{V}$ é $N(\mathbf{1}_p m^\top, I_p, \mathbf{V})$. Mas a identidade dada em (3.66) implica que

$$E_m\{E_m(Y_1^{(p)\top} Y_1^{(p)} | V) | Y_1^{(p)\top} Y_1^{(p)}\} = B + a Y_1^{(p)\top} Y_1^{(p)} .$$

Deste resultado juntamente com o Fato 3.14 tem-se a prova. □

Capítulo 4

Inferência no modelo de regressão linear usando a distribuição t

4.1 Introdução

Considere o modelo de regressão linear definido por

$$Y = X\beta + \epsilon, \quad (4.1)$$

onde $Y = (Y_1, Y_2, \dots, Y_n)^\top$ é um vetor aleatório, $X = [x_{ij}]$ é uma matriz $n \times k$ não aleatória de posto $k < n$, $\beta = (\beta_1, \beta_2, \dots, \beta_k)^\top$ é um vetor de parâmetros (de regressão) desconhecidos e $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ é um vetor de erros aleatórios. Sabe-se que se

$$E(\epsilon) = 0 \quad \text{e} \quad \text{Cov}(\epsilon) = E(\epsilon\epsilon^\top) = \sigma^2 I_n, \quad (4.2)$$

então o estimador de mínimos quadrados (EMQ) de β , isto é,

$$\hat{\beta}_{MQ} = (X^\top X)^{-1} X^\top Y, \quad (4.3)$$

verifica as seguintes propriedades (veja, por exemplo, Arnold (1981)):

- (1) $\hat{\beta}_{MQ} = CY$, onde $C = (X^\top X)^{-1} X^\top$ ($\hat{\beta}_{MQ}$ é linear em Y);
- (2) $E(\hat{\beta}_{MQ}) = \beta$ ($\hat{\beta}_{MQ}$ é um estimador não viciado de β);

- (3) se $\tilde{\beta}$ é qualquer outro estimador de β que satisfaz (1) e (2), então pelo teorema de Gauss-Markov tem-se que

$$\text{Cov}(\tilde{\beta}) - \text{Cov}(\hat{\beta}_{MQ}) \geq 0, \quad (\text{n\~ao-negativa definida}),$$

onde é fácil ver que

$$\text{Cov}(\hat{\beta}_{MQ}) = \sigma^2(X^\top X)^{-1}.$$

Em outras palavras, $\hat{\beta}_{MQ}$ é o estimador de variância mínima na classe dos estimadores lineares e não-viciados de β .

Além disso, se a matriz X verifica a condição

$$\lim_{n \rightarrow \infty} \frac{1}{n} X^\top X = \Omega, \quad (4.4)$$

com $\Omega(k \times k)$ sendo uma matriz finita e não-singular, então:

- (4) $\hat{\beta}_{MQ}$ converge em probabilidade para β , isto é, para todo $\varepsilon > 0$ tem-se que

$$\lim_{n \rightarrow \infty} P(\|\hat{\beta}_{MQ} - \beta\|^2 > \varepsilon) = 0,$$

de modo que $\hat{\beta}_{MQ}$ é um estimador consistente de β (no que segue será usada a notação $\text{plim} \hat{\theta}_n = \theta$ para indicar que $\hat{\theta}_n$ converge em probabilidade para θ).

A prova da propriedade (4) segue de (2) e de observar que (4.4) implica que

$$\lim_{n \rightarrow \infty} \text{Cov}(\hat{\beta}_{MQ}) = 0.$$

Se, adicionalmente, se tem que os erros $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são i.i.d., então:

- (5) $\sqrt{n}(\hat{\beta}_{MQ} - \beta)$ converge em distribuição para $Z \sim N_k(0, \sigma^2 \Omega^{-1})$. Isto é, $\hat{\beta}_{MQ}$ tem distribuição assintótica normal, a saber,

$$\hat{\beta}_{MQ} \sim AN_k(\beta, \sigma^2(X^\top X)^{-1}). \quad (4.5)$$

A prova deste resultado segue de observar que

$$\sqrt{n}(\hat{\beta}_{MQ} - \beta) = \left(\frac{X^T X}{n} \right)^{-1} \frac{X^T \epsilon}{\sqrt{n}} .$$

Então, de acordo com a condição (4.4) e o teorema de Slutsky, é suficiente mostrar que $\epsilon_n = X^T \epsilon / \sqrt{n} \xrightarrow{d} \epsilon \sim N_k(0, \sigma^2 \Omega)$, o que pode ser encontrado, por exemplo, em Arnold (1981, p.147).

Seja $\hat{\epsilon}_{MQ} = Y - X\hat{\beta}_{MQ}$ o vetor de resíduos. Então usando o fato de que $\hat{\epsilon}_{MQ} = M\epsilon$, onde $M = I_n - P$ e $P = X(X^T X)^{-1}X^T$ são matrizes idempotentes com $r(M) = n - r(P) = n - k$, é fácil ver que (veja Arnold (1981, p.148))

$$\hat{\sigma}_{MQ}^2 = \frac{1}{n - k} \hat{\epsilon}_{MQ}^T \hat{\epsilon}_{MQ} \quad (4.6)$$

é um estimador não-viciado e consistente de σ^2 .

Finalmente, se é incorporada a suposição de normalidade na distribuição do vetor de erros, isto é, se

$$\epsilon \sim N_n(0, \sigma^2 I_n) ,$$

então em (4.1) tem-se que

$$Y \sim N_n(X\beta, \sigma^2 I_n) . \quad (4.7)$$

Note que a suposição de normalidade implica que as componentes de ϵ , e portanto de Y , são independentes, uma vez que $\text{Cov}(\epsilon) = \text{Cov}(Y) = \sigma^2 I_n$. A função de verossimilhança do modelo (4.7) é dada por

$$p_Y(y|\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}n} e^{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta)} ,$$

de onde segue que os estimadores de máxima verossimilhança (EMV) de β e σ^2 no modelo (4.7), digamos, $\hat{\beta}_N$ e $\hat{\sigma}_N^2$, são dados por

$$\hat{\beta}_N = \hat{\beta}_{MQ} \quad \text{e} \quad \hat{\sigma}_N^2 = \frac{n - k}{n} \hat{\sigma}_{MQ}^2 ,$$

onde $\hat{\beta}_{MQ}$ e $\hat{\sigma}_{MQ}^2$ são os EMQ de β e σ^2 definidos em (4.3) e (4.6), respectivamente. Neste caso é fácil ver também que $(\hat{\beta}_{MQ}^\top, \hat{\sigma}_{MQ}^2)^\top$ (ou $(\hat{\beta}_N^\top, \hat{\sigma}_N^2)^\top$) é uma estatística suficiente e completa para $(\beta^\top, \sigma^2)^\top$, de modo que a propriedade (3) de $\hat{\beta}_{MQ}$ se estende para a classe de estimadores não-viciados de β . A mesma propriedade se verifica para o estimador $\hat{\sigma}_{MQ}^2$ de σ^2 . Contudo, a principal desvantagem deste procedimento é o fato de que $\hat{\beta}_N = \hat{\beta}_{MQ}$ não é robusto, no sentido que sua função de influência não é uma função limitada (veja, por exemplo, Hampell et al. (1986) e Huber (1981)). Esta limitação tem motivado diversos estudos de robustez, tanto em quantidade quanto em natureza, com o objetivo de conseguir procedimentos adequados que permitam superá-la. Na etapa atual, já está desenvolvida toda uma teoria sobre robustez que vai além do caso particular do modelo de regressão. Textos clássicos como os de Hampell et al. (1986) e Huber (1981) são claros exemplos deste desenvolvimento. De qualquer modo, neste trabalho a noção de robustez é utilizada em seu sentido mais simples (e intuitivo) da vulnerabilidade de um certo estimador com respeito à presença de observações extremas (ou “outliers”). Neste sentido, o modelo normal dado em (4.7) é inadequado se em (4.1) as distribuições marginais do vetor de erros ϵ têm caudas mais pesadas que a distribuição normal.

Neste capítulo, se discute a análise estatística do modelo (4.1) quando a suposição de normalidade na distribuição marginal dos erros $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ é substituída pela suposição de que esta distribuição é $t_1(0, \tau^2; \nu)$. Duas especificações diferentes para a distribuição conjunta dos erros que preservam tais marginais serão discutidas. A primeira delas considera que os erros $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ têm distribuição conjunta $t_n(0, \tau^2 I_n; \nu)$ e, portanto, não são independentes (Seção 4.2). Por outro lado, a segunda especificação estabelece independência entre os erros, de modo que sua distribuição conjunta, digamos, $t_n^{(I)}(0, \tau^2 I_n; \nu)$, é o produto das n distribuições marginais $t_1(0, \tau^2; \nu)$ (Seção 4.3). Para ambos os casos tem-se que o modelo (4.1) se reduz ao modelo normal dado em (4.7) quando $\nu = \infty$. Além disto, se $\nu > 2$, as duas especificações do modelo t verificam (4.2) com $\sigma^2 = \{\nu/(\nu - 2)\}\tau^2$

e, conseqüentemente, $\hat{\beta}_{MQ}$ satisfaz as propriedades (1) a (3) e a consistência de $\hat{\beta}_{MQ}$ se verifica para todo $\nu > 0$. A diferença entre ambas as especificações vem do fato que no primeiro modelo o estimador de máxima verossimilhança também coincide com $\hat{\beta}_{MQ}$ e, portanto, não é robusto. Além disto, como é provado na Seção 4.2.2 (veja também Zellner (1976), Gosh e Sinha (1980) e Ullah e Zinde-Walsh (1984, 1985)), no primeiro modelo os testes da razão de verossimilhança, escore e de Wald da hipótese nula $H_0 : A\beta = \gamma_0$, onde A é uma matriz $q \times k$ com $r(A) = q \leq k$, se reduzem ao mesmo teste usual F relativo ao modelo normal. O segundo modelo, entretanto, nos conduz a um estimador robusto de β . Contudo, tal estimador é a solução não-explicita da respectiva equação de (log-) verossimilhança, de modo que são necessários algoritmos numéricos para determinar seu valor em uma dada amostra. Como conseqüência disto, para este modelo as inferências sobre β devem se basear na teoria assintótica. Em particular, os testes de H_0 mencionados acima são baseados na distribuição assintótica qui-quadrado. Tal aproximação não é garantida para amostras de tamanhos finitos.

A conveniência do modelo de regressão t com marginais independentes é fundamentada em Lange, Little e Taylor (1989) (veja também Little (1988) e Taylor (1992)), o qual também será o principal tema de análise deste capítulo. Embora a maior parte dos resultados que serão desenvolvidos nas próximas seções sejam próprios deste trabalho, a contribuição que aqui se enfatiza é a derivação das estatísticas da razão de verossimilhança e de escore modificadas por fatores tipo Bartlett, para testar uma hipótese nula $H_0 : A\beta = \gamma_0$ (Seção 4.4.2).

Como é mostrado, respectivamente, por Lawley (1956) e por Cordeiro e Ferrari (1991) (veja também Hayakawa (1979), Cordeiro (1987) e Barndorff-Nielsen e Cox (1984)), para um modelo regular, as estatísticas da razão de verossimilhança e escore modificadas têm distribuição assintótica qui-quadrado até ordem n^{-1} sob a hipótese nula, isto é, se os termos de ordem inferior a n^{-1} são ignorados, onde n denota o tamanho da amostra. Os

fatores de correção de ambas as estatísticas são funções de cumulantes de derivadas da log-verossimilhança, de modo que sua determinação envolve um processo algébrico complexo na maioria dos casos. Esta complexidade aumenta com o número de parâmetros incluídos no modelo e, em particular, com a presença de parâmetros de incômodo (“nuisance parameters”). No caso com mais de um parâmetro, a situação mais simples acontece quando a matriz de informação é diagonal. Tal situação se produz, por exemplo, com os parâmetros de locação e de escala de um modelo elíptico (ou simétrico), como é o caso do modelo t aqui considerado quando supõe-se que o parâmetro de incômodo (os graus de liberdade) é conhecido. Para ν desconhecido tem-se que os elementos cruzados da matriz de informação que ligam o parâmetro de locação η com ν também são nulos, de modo que η e ν são ortogonais segundo a definição de Cox e Reid (1987). Mas, ν e o parâmetro de escala τ^2 não são ortogonais. Este último fato dificulta a derivação dos fatores de correção dos testes mencionados acima no caso que ν seja desconhecido. Contudo, como será mostrado, é possível obter uma reparametrização ortogonal de $(\tau^2, \nu)^T$ sem alterar a ortogonalidade de η com $(\tau^2, \nu)^T$ e as inferências sobre η e ν . Com esta parametrização ortogonal, a derivação dos testes modificados quando ν é desconhecido é simplificada consideravelmente. Além disso, é possível mostrar que os fatores de correção não dependem da parametrização utilizada em sua derivação. Notamos que no modelo de regressão t , estes fatores dependem somente de ν e da matriz X .

4.2 O modelo de regressão t com erros dependentes

Seja $Y = X\beta + \epsilon$ o modelo de regressão definido em (4.1). Como o vetor de locação $\eta = X\beta$ é fixo, das propriedades da distribuição t multivariada (veja (3.31)) segue que se

$$\epsilon \sim t_n(0, \tau^2 I_n; \nu) ,$$

então tem-se o modelo de regressão t dado por

$$Y \sim t_n(X\beta, \tau^2 I_n; \nu), \quad (4.8)$$

com função de verossimilhança dada por

$$p_Y(y | \beta, \tau^2, \nu) = k(\nu, n)(\nu\tau^2)^{\frac{1}{2}\nu} \{\nu\tau^2 + (y - X\beta)^\top (y - X\beta)\}^{-\frac{1}{2}(\nu+n)}, \quad (4.9)$$

onde

$$k(\nu, n) = \frac{\Gamma[\frac{1}{2}(n + \nu)]}{\Gamma(\frac{1}{2}\nu)\pi^{\frac{1}{2}n}}.$$

De (3.29) e (3.30) é imediato que neste modelo

$$E(Y) = X\beta \quad (\text{ou } E(\epsilon) = 0), \quad \text{para } \nu > 1 \quad (4.10)$$

e

$$\text{Cov}(Y) = \text{Cov}(\epsilon) = \frac{\nu}{\nu - 2}\tau^2 I_n, \quad \text{para } \nu > 2. \quad (4.11)$$

Consequentemente, para $\nu > 2$, o estimador $\hat{\beta}_{MQ}$ dado em (4.3) verifica as propriedades

(1) a (3) e, sobre a condição (4.4), $\hat{\beta}_{MQ}$ é também um estimador consistente de β .

Escrevendo $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$, $Y = (Y_1, Y_2, \dots, Y_n)^\top$ e denotando as linhas da matriz X por $x_1^\top, x_2^\top, \dots, x_n^\top$, de (3.34) segue que (4.8) tem regressões marginais dadas por

$$Y_i = x_i^\top \beta + \epsilon_i \sim t_1(x_i^\top \beta, \tau^2; \nu), \quad i = 1, 2, \dots, n,$$

com densidades

$$p_{Y_i}(y_i | \beta, \tau^2, \nu) = k(1, \nu)(\nu\tau^2)^{\frac{\nu}{2}} \{\nu\tau^2 + (y_i - x_i^\top \beta)^2\}^{-\frac{1}{2}(\nu+1)}, \quad i = 1, 2, \dots, n, \quad (4.12)$$

tais que $E(Y_i) = x_i^\top \beta$ (ou $E(\epsilon_i) = 0$), quando $\nu > 1$, e

$$\text{Cov}(Y_i, Y_j) = \text{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} \frac{\nu}{\nu-2}\tau^2, & \text{se } i = j, \\ 0, & \text{se } i \neq j, \end{cases}$$

quando $\nu > 2$, $i = 1, 2, \dots, n$. De qualquer modo, $Y_1, Y_2, \dots, Y_n$ (ou $\epsilon_1, \epsilon_2, \dots, \epsilon_n$) não são independentes, uma vez que (4.9) é diferente do produto das n densidades marginais dadas em (4.12).

4.2.1 Estimação de máxima verossimilhança

É claro que para cada $\tau^2 > 0$ e $\nu > 0$ fixados, maximizar (4.9) com respeito a β equivale a minimizar $Q(\beta) = (y - X\beta)^\top(y - X\beta)$ com respeito a β , uma vez que (4.9) é uma função monótona decrescente em $Q(\beta)$. Consequentemente, o estimador de máxima verossimilhança de β relativo ao modelo (4.8), digamos $\hat{\beta}$, coincide com $\hat{\beta}_{MQ}$, isto é,

$$\hat{\beta} = \hat{\beta}_{MQ} = (X^\top X)^{-1} X^\top Y .$$

Procedendo de forma similar, é fácil ver também que para cada $\nu > 0$, o estimador de máxima verossimilhança de τ^2 é dado por

$$\hat{\tau}^2 = \frac{1}{n} \hat{\epsilon}^\top \hat{\epsilon} ,$$

onde $\hat{\epsilon} = Y - X\hat{\beta}$ é o vetor de resíduos. Note-se que $(\hat{\beta}^\top, \hat{\tau}^2)^\top$ independe do valor de ν e é uma estatística suficiente para $(\beta^\top, \tau^2)^\top$ quando ν é conhecido. Isto é, do fato de que

$$\begin{aligned} (Y - X\beta)^\top(Y - X\beta) &= (X\beta - X\hat{\beta})^\top(X\beta - X\hat{\beta}) + \hat{\epsilon}^\top \hat{\epsilon} \\ &= (\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) + n\hat{\tau}^2 , \end{aligned}$$

mostra-se que (4.9) pode ser expressa como

$$p_Y(y \mid \beta, \tau^2, \nu) = k(\nu, n)(\nu\hat{\tau}^2)^{\frac{1}{2}\nu} \{ \nu\hat{\tau}^2 + n\hat{\tau}^2 + (\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) \}^{-\frac{1}{2}(\nu+n)} ,$$

e a suficiência de $(\hat{\beta}^\top, \hat{\tau}^2)^\top$ segue do critério da fatoração, uma vez que ν seja conhecido. É fácil ver também que $T = (\hat{\beta}^\top, \hat{\tau}^2)^\top$ é uma estatística completa para $\theta = (\beta^\top, \tau^2)^\top$. De fato, lembramos que (4.8) pode ser expressa em forma equivalente como (veja o Capítulo 3)

$$(Y \mid V = v) \sim N_n(X\beta, v\tau^2 I_n) , \quad \text{onde } V \sim \nu/\chi_\nu^2 . \quad (4.13)$$

Seja então g uma função tal que $E_\theta(g(T)) = 0$ para todo θ . De acordo com a definição de completividade, prova-se que isto implica que $P_\theta(g(T) = 0) = 1$ para todo θ (veja Casella

e Berger (1990, p.260)). Seja $h_\theta(V) = E_\theta(g(T) | V)$. Das propriedades da esperança condicional tem-se que $0 = E_\theta(g(T)) = E(h_\theta(V))$, isto é,

$$0 = \int_0^\infty h_\theta(v) p_v(v) dv ,$$

de onde segue que, para todo θ ,

$$h_\theta(V) = E_\theta(g(T) | V) = 0 ,$$

com probabilidade um. Mas isto implica que, dado V , $g(T) = 0$ com probabilidade um, uma vez que sobre o modelo normal $T = (\hat{\beta}^\top, \hat{\tau}^2)^\top$ é uma estatística completa para $\theta = (\beta^\top, \tau^2)^\top$. Notando, agora, que $P_\theta(g(T) = 0) = E_\theta(P_\theta(g(T) = 0 | V)) = 1$, tem-se a completividade de T . Daqui segue também que $(\hat{\beta}^\top, \hat{\tau}^2)^\top$ é a estatística suficiente minimal de $(\beta^\top, \tau^2)^\top$. Além disto, usando (4.10) e (4.11) tem-se que

$$E(\hat{\beta}) = \beta , \quad \nu > 1 , \quad \text{e} \quad \text{Cov}(\hat{\beta}) = \{\nu/(\nu - 2)\} \tau^2 (X^\top X)^{-1} , \quad \nu > 2 .$$

Logo, para $\nu > 2$, $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ é também o estimador de variância mínima na classe dos estimadores não-viciados de β , uma vez que $T = (\hat{\beta}^\top, \hat{\tau}^2)^\top$ é suficiente e completa para $\theta = (\beta^\top, \tau^2)^\top$. Lembrando agora que $\hat{\epsilon} = MY$, onde $M = I_n - P$ e $P = X(X^\top X)^{-1} X^\top$ são matrizes idempotentes (de projeção), é fácil ver que $E(\hat{\epsilon}^\top \hat{\epsilon}) = (n - k) \{\nu/(\nu - 2)\} \tau^2$, para $\nu > 2$. Logo, para $\nu > 2$, o estimador não viciado de variância mínima de τ^2 é dado por

$$\hat{\tau}^2 = \left(\frac{\nu - 2}{\nu} \right) \hat{\sigma}_{MQ}^2 = \left(\frac{\nu - 2}{\nu} \right) \left(\frac{n}{n - k} \right) \hat{\tau}^2 .$$

Por outro lado, o fato de que $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ é linear em Y implica que (veja (3.31))

$$\hat{\beta} \sim t_k(\beta, \tau^2 (X^\top X)^{-1}; \nu) . \quad (4.14)$$

Considerando a condição (4.4), a consistência de $\hat{\beta}$ para todo $\nu > 0$ vem diretamente da aplicação de (4.13) em (4.14). De fato, seja $\varepsilon > 0$. Então,

$$P(\|\hat{\beta} - \beta\| > \varepsilon) = \int_0^\infty P(\|\hat{\beta} - \beta\| > \varepsilon | V = v) dF_V(v) ,$$

onde $V \sim F_V = GI(\nu/2, \nu/2)$. Mas, dado $V = v$, $\hat{\beta} \sim N_k(\beta, v\tau^2(X^\top X)^{-1})$ e $P(\|\hat{\beta} - \beta\| > \varepsilon \mid V = v) = h_n(v) \rightarrow 0$, quando $n \rightarrow \infty$. Então, do teorema da convergência dominada segue que

$$\lim_{n \rightarrow \infty} P(\|\hat{\beta} - \beta\| > \varepsilon) = \int_0^\infty \lim_{n \rightarrow \infty} P(\|\hat{\beta} - \beta\| > \varepsilon \mid V = v) dF_V(v) = 0 .$$

Como $\varepsilon > 0$ é arbitrário, tem-se que $\text{plim } \hat{\beta} = \beta$. Procedendo de forma similar, é fácil ver que

$$\text{plim } \hat{\tau}^2 = E(\epsilon_1^2 \mid V) = V\tau^2 .$$

De (4.14) segue também que (veja (3.14))

$$\frac{(\hat{\eta} - \eta)^\top (\hat{\eta} - \eta)}{k\tau^2} = \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{k\tau^2} \sim F_{k,\nu} , \quad (4.15)$$

onde $\eta = X\beta$ e $\hat{\eta} = X\hat{\beta} = PY$. Similarmente, como $\hat{\epsilon} = MY$, com $M^2 = M$ e $\text{tr}(M) = n - k$, então

$$\hat{\epsilon} \sim t_n(0, \tau^2 M; \nu)$$

e, portanto (veja (3.48)),

$$\frac{\hat{\epsilon}^\top \hat{\epsilon}}{(n - k)\tau^2} \sim F_{n-k,\nu} . \quad (4.16)$$

Finalmente, note-se que

$$\begin{pmatrix} \hat{\eta} \\ \hat{\epsilon} \end{pmatrix} = \begin{pmatrix} P \\ M \end{pmatrix} Y \sim t_{2n} \left(\begin{pmatrix} \eta \\ 0 \end{pmatrix}, \tau^2 \begin{pmatrix} P & 0 \\ 0 & M \end{pmatrix}; \nu \right) ,$$

de onde segue que (veja (3.49))

$$\frac{(\hat{\eta} - \eta)^\top (\hat{\eta} - \eta)}{\hat{\epsilon}^\top \hat{\epsilon}} \left(\frac{n - k}{k} \right) = \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{\hat{\epsilon}^\top \hat{\epsilon}} \left(\frac{n - k}{k} \right) \sim F_{k,n-k} . \quad (4.17)$$

Contudo, $\hat{\eta}$ (ou $\hat{\beta}$) e $\hat{\epsilon}$ não são independentes. Os resultados dados em (4.15) e (4.17) são úteis, por exemplo, na construção de regiões de confiança para β ou η . Assim, quando τ^2 é conhecido deve ser usado (4.15) e quando τ^2 é desconhecido usa-se (4.17). Neste último caso, utilizando (4.16) pode ser construído também um intervalo de confiança

para o parâmetro τ^2 . Resultados similares podem ser obtidos para funções paramétricas da forma $a^\top \beta$ ($a \in \mathbb{R}^k$) e $A\beta$ ($A(q \times k)$ com $r(A) = q \leq k$), uma vez que

$$a^\top \hat{\beta} \sim t_1(a^\top \beta, \tau^2 a^\top (X^\top X)^{-1} a; \nu)$$

e

$$A\hat{\beta} \sim t_q(A\beta, \tau^2 A(X^\top X)^{-1} A^\top; \nu). \quad (4.18)$$

Para concluir esta seção, note-se que se ν é desconhecido, então na função de verossimilhança dada em (4.9) tem-se que $\theta = (\beta^\top, \tau^2, \nu)^\top$. É fácil ver que em tal caso a função de verossimilhança não satisfaz as condições necessárias para ser maximizada com respeito a θ (veja Zellner (1976) e Anderson et al. (1986)). Em outras palavras, embora o EMV de $(\beta^\top, \tau^2)^\top$ exista para cada ν fixado (e não depende de ν), o EMV de $(\beta^\top, \tau^2, \nu)^\top$ não existe. Contudo, ν pode ser estimado a partir dos dados com observações replicadas (veja Sutradhar e Ali (1986), Lange et al. (1989) e Kano et al. (1993)). Uma análise Bayesiana sobre este modelo é feita em Zellner (1976). Os resultados de Zellner (1976) são estendidos por Chib et al. (1988) para a classe dos modelos elípticos que são mistura de escala do modelo normal.

4.2.2 Testes da hipótese linear $H_0 : A\beta = \gamma_0$

Assumindo o modelo de regressão t dado em (4.8), nesta seção se derivam os testes de Wald, da razão de verossimilhança e escore para a hipótese linear $H_0 : A\beta = \gamma_0$, onde o vetor $\gamma_0(q \times 1)$ e a matriz $A(q \times k)$ são conhecidos e $r(A) = q \leq k$.

Como se sabe, o estimador de β que minimiza $Q(\beta) = (Y - X\beta)^\top(Y - X\beta)$ quando β está sujeito à restrição linear $A\beta = \gamma_0$ (A e γ_0 definidos como acima) é dado por (veja Seber (1977, p.85))

$$\tilde{\beta} = \hat{\beta} - (X^\top X)^{-1} A^\top [A(X^\top X)^{-1} A^\top]^{-1} (A\hat{\beta} - \gamma_0), \quad (4.19)$$

onde $\hat{\beta}$ é o EMQ usual (não restrito) de β e é dado em (4.3), o qual coincide com o respectivo EMV no modelo (4.8). É fácil ver também que, em (4.19), $\tilde{\beta}$ coincide com o EMV de β do modelo (4.8) quando se incorpora a restrição linear $A\beta = \gamma_0$, e $\tilde{\beta}$ possui todas as propriedades mencionadas na seção anterior para o estimador não-restrito $\hat{\beta}$. Considerando estes resultados, mostra-se no que segue que os testes de $H_0 : A\beta = \gamma_0$ mencionados acima se reduzem ao teste usual F quando usa-se o modelo t dado em (4.8).

Seja então $Y \sim t_n(X\beta, \tau^2 I_n; \nu)$ o modelo de regressão t e considere a hipótese nula $H_0 : A\beta = \gamma_0$, com $A(q \times k)$ e $\gamma_0(q \times 1)$ sendo conhecidos e $r(A) = q$. Sob o modelo não restrito a componente residual da variância de Y é dada por

$$Q(\hat{\beta}) = (Y - X\hat{\beta})^\top (Y - X\hat{\beta}) = \hat{\epsilon}^\top \hat{\epsilon} .$$

Similarmente, para o modelo restrito esta componente é dada por

$$Q(\tilde{\beta}) = (Y - X\tilde{\beta})^\top (Y - X\tilde{\beta}) = \tilde{\epsilon}^\top \tilde{\epsilon} ,$$

onde $\tilde{\beta}$ é dado em (4.19). Usando (4.19) é fácil ver que

$$Q(\tilde{\beta}) - Q(\hat{\beta}) = (A\hat{\beta} - \gamma_0)^\top [A(X^\top X)^{-1}A^\top]^{-1} (A\hat{\beta} - \gamma_0) . \quad (4.20)$$

Note-se que (4.18) implica que sob $H_0 : A\beta = \gamma_0$, tem-se que

$$\frac{Q(\hat{\beta}) - Q(\tilde{\beta})}{q\tau^2} \sim F_{q, \nu} .$$

Seja $B = A(X^\top X)^{-1}A^\top$ e lembremos que $\hat{\epsilon} = Y - X\hat{\beta} = MY$, onde $M = I_n - P$, $P = X(X^\top X)^{-1}X^\top$ e $X^\top M = MX = 0$. Do fato de que

$$\begin{pmatrix} B^{-\frac{1}{2}}(A\hat{\beta} - \gamma_0) \\ \hat{\epsilon} \end{pmatrix} = \begin{pmatrix} B^{-\frac{1}{2}}A(X^\top X)^{-1}X^\top \\ M \end{pmatrix} Y - \begin{pmatrix} B^{-\frac{1}{2}}\gamma_0 \\ 0 \end{pmatrix} ,$$

onde $B^{-\frac{1}{2}}A(X^\top X)^{-1}X^\top M = 0$, segue que sob $H_0 : A\beta = \gamma_0$,

$$\begin{pmatrix} B^{-\frac{1}{2}}(A\hat{\beta} - \gamma_0) \\ \hat{\epsilon} \end{pmatrix} \sim t_{q+n} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tau^2 \begin{pmatrix} I_q & 0 \\ 0 & M \end{pmatrix}; \nu \right) ,$$

o que implica que (veja (3.49)), sob $H_0 : A\beta = \gamma_0$,

$$F = \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} \left(\frac{n-k}{q} \right) \sim F_{q, n-k} . \quad (4.21)$$

Portanto, um teste de nível α rejeita $H_0 : A\beta = \gamma_0$ se em (4.21) tem-se que

$$F > F_{q, n-k; 1-\alpha} , \quad (4.22)$$

onde $F_{q, n-k; 1-\alpha}$ denota o percentil $(1 - \alpha)$ 100% da distribuição $F_{q, n-k}$. Note-se que em (4.21) $F =^d F_\infty$, onde F_∞ é a estatística dada por (4.21) para $\nu = \infty$, isto é, sob o modelo normal dado em (4.6).

Sejam $\Lambda = \Lambda(Y)$, $S = S(Y)$ e $W = W(Y)$ as estatísticas da razão de verossimilhança, escore e de Wald, respectivamente, onde Y é assumido como em (4.8). Então, para a hipótese nula $H_0 : A\beta = \gamma_0$ (contra a hipótese alternativa $H_1 : A\beta \neq \gamma_0$), estas estatísticas são dadas por

$$\begin{aligned} \Lambda &= \left\{ 1 + \frac{q}{n-k} F \right\}^{\frac{1}{2}n} , \\ S &= \frac{(\nu + n + 2)n}{\nu + n} \left\{ \left(\frac{q}{n-k} F \right) \left(1 + \frac{q}{n-k} F \right)^{-1} \right\} , \\ W &= \frac{(\nu + n)n}{\nu + n + 2} \left(\frac{q}{n-k} F \right) , \end{aligned} \quad (4.23)$$

onde $F = F(Y)$ é a estatística definida em (4.21). De fato, sejam $\tilde{\theta} = (\tilde{\beta}^\top, \tilde{\tau}^2)^\top$ e $\hat{\theta} = (\hat{\beta}^\top, \hat{\tau}^2)^\top$ os EMV de $\theta = (\beta^\top, \tau^2)^\top$ sob $H_0 : A\beta = \gamma_0$ e sob $H_1 : A\beta \neq \gamma_0$, respectivamente. Então, de (4.9) segue que

$$\Lambda = \left\{ \frac{\hat{\tau}^2}{\tilde{\tau}^2} \right\}^{\frac{1}{2}\nu} \left\{ \frac{\nu \hat{\tau}^2 + Q(\hat{\beta})}{\nu \tilde{\tau}^2 + Q(\tilde{\beta})} \right\}^{-\frac{1}{2}(\nu+n)} ,$$

onde $Q(\beta) = (Y - X\beta)^\top (Y - X\beta)$. Notando que $n\tilde{\tau}^2 = Q(\tilde{\beta})$ e $n\hat{\tau}^2 = Q(\hat{\beta})$, tem-se que

$$\Lambda = \left\{ \frac{Q(\tilde{\beta})}{Q(\hat{\beta})} \right\}^{\frac{1}{2}n} = \left\{ 1 + \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} \right\}^{\frac{1}{2}n} = \left\{ 1 + \frac{q}{n-k} F \right\}^{\frac{1}{2}n} ,$$

o que prova a primeira igualdade em (4.23). É claro que Λ (ou $G = 2 \log \Lambda$) é uma função monótona crescente com F . Além disto, observe-se que Λ não depende de ν .

Consideremos agora a estatística escore, S . Seja $L = L(\theta)$ função de log-verossimilhança do modelo (4.8), onde $\theta = (\beta^\top, \tau^2)^\top$. Então a estatística S é definida por

$$S = \left\{ \left(\frac{\partial L}{\partial \theta} \right)^\top K^{-1} \left(\frac{\partial L}{\partial \theta} \right) \right\} \Big|_{\theta=\tilde{\theta}}, \quad (4.24)$$

onde $K = K(\theta)$ denota a matriz de informação e $\tilde{\theta}$ é o EMV de θ sob H_0 . De (4.9) segue que

$$L = \log k(n, \nu) + \frac{\nu}{2} \log(\nu \tau^2) - \frac{(\nu + n)}{2} \log\{\nu \tau^2 + Q(\beta)\}, \quad (4.25)$$

onde $Q(\beta) = (Y - X\beta)^\top (Y - X\beta)$. Daqui é fácil ver que

$$\begin{aligned} \frac{\partial L}{\partial \theta} \Big|_{\theta=\tilde{\theta}} &= \left(\frac{\partial L}{\partial \beta^\top}, \frac{\partial L}{\partial \tau^2} \right)^\top \Big|_{\theta=\tilde{\theta}} \\ &= \left(\frac{\nu + n}{\nu \tilde{\tau}^2 + Q(\tilde{\beta})} (Y - X\tilde{\beta})^\top X, \frac{\nu}{2\tilde{\tau}^2} - \frac{\nu(\nu + n)}{2(\nu \tilde{\tau}^2 + Q(\tilde{\beta}))} \right)^\top \\ &= \left(\frac{\tilde{\epsilon}^\top X}{\tilde{\tau}^2}, 0 \right)^\top, \end{aligned}$$

onde $\tilde{\epsilon} = Y - X\tilde{\beta}$ é o vetor de resíduos sob H_0 . Do Apêndice A.1 segue que $K = \text{Diag}\{K_1, K_2\}$, onde

$$K_1 = E \left\{ \left(\frac{\partial L}{\partial \beta} \right) \left(\frac{\partial L}{\partial \beta} \right)^\top \right\} = \frac{1}{\tau^2} \frac{\nu + n}{\nu + n + 2} X^\top X$$

e

$$K_2 = E \left\{ \left(\frac{\partial L}{\partial \tau^2} \right)^2 \right\} = \frac{n}{\tau^4} \frac{\nu + n - 1}{\nu + n + 2}.$$

Logo, em (4.24), tem-se que

$$S = \left\{ \left(\frac{\partial L}{\partial \beta} \right)^\top K^{-1}(\beta) \left(\frac{\partial L}{\partial \beta} \right) \right\} \Big|_{\beta=\tilde{\beta}} = \frac{\nu + n + 2}{\nu + n} \frac{\tilde{\epsilon}^\top P \tilde{\epsilon}}{\tilde{\tau}^2},$$

onde $P = X(X^\top X)^{-1}X^\top$. Usando (4.19) e (4.20), é fácil ver que

$$\tilde{\epsilon}^\top P \tilde{\epsilon} = (\hat{\beta} - \tilde{\beta})^\top X^\top X (\hat{\beta} - \tilde{\beta}) = (A\hat{\beta} - \gamma_0)^\top [A(X^\top X)^{-1}A^\top]^{-1} (A\hat{\beta} - \gamma_0) = Q(\tilde{\beta}) - Q(\hat{\beta}),$$

de onde segue que

$$\begin{aligned} S &= \frac{(\nu + n + 2)n}{\nu + n} \left\{ \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\tilde{\beta})} \right\} \\ &= \frac{(\nu + n + 2)n}{\nu + n} \left\{ \left(\frac{q}{n - k} F \right) \left(1 + \frac{q}{n - k} F \right)^{-1} \right\}, \end{aligned}$$

que é uma função monótona crescente com F . Note-se também que S depende de ν , digamos, $S = S_\nu$.

Consideremos, finalmente, a estatística de Wald, W . Esta estatística é definida por

$$W = (A\hat{\beta} - \gamma_0)^\top [AK_1^{-1}(\hat{\theta})A^\top]^{-1} (A\hat{\beta} - \gamma_0), \quad (4.26)$$

onde $\hat{\beta}$ é o EMV de β sob $H_1 : A\beta \neq \gamma_0$. Notando que $K_1(\hat{\theta}) = \{(\nu + n)/\hat{\tau}^2(\nu + n + 2)\} X^\top X$ e usando (4.20), em (4.26) tem-se que

$$\begin{aligned} W &= \frac{\nu + n}{\nu + n + 2} \frac{(A\hat{\beta} - \gamma_0)^\top [A(X^\top X)^{-1}A^\top]^{-1} (A\hat{\beta} - \gamma_0)}{\hat{\tau}^2} \\ &= \frac{(\nu + n)n}{\nu + n + 2} \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} \\ &= \frac{(\nu + n)n}{\nu + n + 2} \left(\frac{q}{n - k} F \right), \end{aligned}$$

a qual também é uma função monótona crescente em F e $W = W_\nu$.

Consequentemente, o teste de nível α para $H_0 : A\beta = \gamma_0$ baseado em qualquer uma das estatísticas dadas em (4.23) é equivalente ao teste dado em (4.22). Além disso, de (4.23) também se observa que $\Lambda \stackrel{d}{=} \Lambda_\infty$, $S_\nu \stackrel{d}{=} \{(\nu + n + 2)/(\nu + n)\} S_\infty$ e $W_\nu \stackrel{d}{=} \{(\nu + n)/(\nu + n + 2)\} W_\infty$, onde $T_\infty = \lim_{\nu \rightarrow \infty} T_\nu$ é a respectiva estatística sob o modelo normal.

Discussões sobre estes testes na classe dos modelos elípticos podem ser encontrados em Ghosh e Sinha (1980) e em Ullah e Zinde-Walsh (1985).

4.3 O modelo de regressão t com erros independentes

Sejam $Y = (Y_1, Y_2, \dots, Y_n)^\top$, $X = [x_{ij}]$, $\beta = (\beta_1, \beta_2, \dots, \beta_k)^\top$ e $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ relacionados como em (4.1) e suponha que $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são iid com

$$\epsilon_i \sim t_1(0, \tau^2; \nu) .$$

Então, se $x_1^\top, x_2^\top, \dots, x_n^\top$ denotam as linhas da matriz X , tem-se que $Y_1, Y_2, \dots, Y_n$ são independentes com

$$Y_i \sim t_1(x_i^\top \beta, \tau^2; \nu) ,$$

de modo que a função de verossimilhança deste modelo é dado por

$$\begin{aligned} p_Y(y \mid \beta, \tau^2, \nu) &= \prod_{i=1}^n \left[k(1, \nu) (\nu \tau^2)^{\frac{1}{2}\nu} \{ \nu \tau^2 + (y_i - x_i^\top \beta)^2 \}^{-\frac{1}{2}(\nu+1)} \right] \\ &= (k(1, \nu))^n (\nu \tau^2)^{\frac{1}{2}n\nu} \prod_{i=1}^n \{ \nu \tau^2 + (y_i - x_i^\top \beta)^2 \}^{-\frac{1}{2}(\nu+1)} . \end{aligned} \quad (4.27)$$

Uma expressão equivalente de (4.27) é dada no que se segue.

Seja

$$t_i = \frac{y_i - x_i^\top \beta}{\tau} , \quad (4.28)$$

o erro “padronizado” relativo à i -ésima observação, $i = 1, 2, \dots, n$. Então, usando o fato de que

$$\begin{aligned} \prod_{i=1}^n \left\{ 1 + \frac{1}{\nu} t_i^2 \right\}^{-\frac{1}{2}(\nu+1)} &= \left\{ \prod_{i=1}^n \left(1 + \frac{1}{\nu} t_i^2 \right) \right\}^{-\frac{1}{2}(\nu+1)} \\ &= \left\{ 1 + \frac{1}{\nu} \sum_{i=1}^n t_i^2 + \frac{1}{\nu^2} \sum_{i < j}^n t_i^2 t_j^2 + \dots + \frac{1}{\nu^n} \prod_{i=1}^n t_i^2 \right\}^{-\frac{1}{2}(\nu+1)} , \end{aligned}$$

tem-se que (4.27) também pode ser expressa como

$$p_Y(y \mid \beta, \tau^2, \nu) = \left\{ \frac{k^2(1, \nu)}{\nu \tau^2} \right\}^{\frac{1}{2}n} \left\{ 1 + \frac{1}{\nu} \sum_{i=1}^n t_i^2 + \frac{1}{\nu^2} \sum_{i < j}^n t_i^2 t_j^2 + \dots + \frac{1}{\nu^n} \prod_{i=1}^n t_i^2 \right\}^{-\frac{1}{2}(\nu+1)} .$$

O modelo definido acima será denominado aqui como o modelo de regressão t independente e será denotado (matricialmente) como se segue.

$$Y = (Y_1, Y_2, \dots, Y_n)^T \sim \prod_{i=1}^n t_1(x_i^T \beta, \tau^2; \nu) \equiv t_n^{(I)}(X\beta, \tau^2 I_n; \nu) . \quad (4.29)$$

Sob este modelo estudaremos agora as propriedades dos EMQ de β e τ^2 . É claro que, para este modelo, (4.10) e (4.11) também valem, de modo que $\hat{\beta}_{MQ}$ (dado em (4.3)) verifica as propriedades (1) a (3) para $\nu > 2$. Além disto, sob a condição dada em (4.4), para $\nu > 2$ tem-se também que $\hat{\beta}_{MQ}$ tem distribuição assintótica como em (4.5) com $\sigma^2 = \{\nu/(\nu-2)\}\tau^2$, uma vez que $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são iid com $E(\epsilon_i) = 0$ e $\text{Var}(\epsilon_i) = \{\nu/(\nu-2)\}\tau^2$, sendo ambas finitas quando $\nu > 2$. A consistência de $\hat{\beta}_{MQ}$ para $\nu > 2$ segue deste último fato. Similarmente, o estimador $\hat{\sigma}_{MQ}^2$ (definido em (4.6)) é um estimador não-viciado e consistente de σ^2 quando $\nu > 2$, de onde segue que se $\nu > 2$ é conhecido, então um estimador consistente e não-viciado de τ^2 é dado por

$$\hat{\tau}_{MQ}^2 = \frac{\nu-2}{\nu} \hat{\sigma}_{MQ}^2 . \quad (4.30)$$

Se ν é desconhecido e assume-se que $\nu > 4$, então um estimador consistente de ν pode ser obtido a partir dos resíduos MQ $\hat{\epsilon}_{iMQ} = Y_i - x_i^T \hat{\beta}_{MQ}$, $i = 1, 2, \dots, n$. De fato, observando que (veja (2.45))

$$\beta_{2,1} \equiv E \left\{ \left(\frac{\epsilon_i}{\sigma} \right)^4 \right\} = 3 \frac{\nu-2}{\nu-4} ,$$

onde $\sigma^2 = \text{Var}(\epsilon_i)$, tem-se que um estimador de ν , digamos, $\hat{\nu}_{MQ}$, pode ser construído resolvendo a equação (“de momentos”) dada por

$$3 \frac{\hat{\nu}_{MQ} - 2}{\hat{\nu}_{MQ} - 4} = \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^4 / \left(\frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^2 \right)^2 \right\} \equiv \hat{\beta}_{2,1} ,$$

isto é,

$$\hat{\nu}_{MQ} = \frac{4\hat{\beta}_{2,1} - 6}{\hat{\beta}_{2,1} - 3} . \quad (4.31)$$

Para provar que $\hat{\nu}_{MQ}$ é um estimador consistente de ν , é suficiente mostrar que $\text{plim} \hat{\beta}_{2,1} = \beta_{2,1}$ ou, equivalentemente, que

$$\text{plim} \left(\frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^4, \frac{n-k}{n} \hat{\sigma}_{MQ}^2 \right) = \left(3 \frac{\nu-2}{\nu-4} \sigma^4, \sigma^2 \right),$$

uma vez que o denominador de (4.31) é distinto de zero com probabilidade um e, portanto, $\hat{\nu}_{MQ}$ é uma função contínua em $\left(\frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^4, \frac{n-k}{n} \hat{\sigma}_{MQ}^2 \right)$ a menos de um conjunto com medida zero. Sabemos que $\text{plim} \left\{ \frac{n-k}{n} \hat{\sigma}_{MQ}^2 \right\} = \text{plim} \hat{\sigma}_{MQ}^2 = \sigma^2$. Para mostrar que $\text{plim} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^4 \right\} = 3 \frac{\nu-2}{\nu-4} \sigma^4$, a condição adicional de que $|x_{ij}| \leq c$ (para algum $c > 0$) para todo i e j é requerida. De fato, notando que $\hat{\epsilon}_{iMQ} = \epsilon_i + x_i^\top (\hat{\beta}_{MQ} - \beta)$, tem-se que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{iMQ}^4 &= \frac{1}{n} \sum_{i=1}^n \{ \epsilon_i + x_i^\top (\hat{\beta}_{MQ} - \beta) \}^4 \\ &= \frac{1}{n} \sum_{i=1}^n \epsilon_i^4 - \frac{4}{n} \sum_{i=1}^n \epsilon_i^3 x_i^\top (\hat{\beta}_{MQ} - \beta) + \frac{6}{n} \sum_{i=1}^n \epsilon_i^2 [x_i^\top (\hat{\beta}_{MQ} - \beta)]^2 \\ &\quad - \frac{4}{n} \sum_{i=1}^n \epsilon_i [x_i^\top (\hat{\beta}_{MQ} - \beta)]^3 + \frac{1}{n} \sum_{i=1}^n [x_i^\top (\hat{\beta}_{MQ} - \beta)]^4. \end{aligned}$$

Da lei fraca dos grandes números (para variáveis aleatórias iid) segue que $\text{plim} \frac{1}{n} \sum_{i=1}^n \epsilon_i^4 = 3\{(\nu-2)/(\nu-4)\}\sigma^4$. Os termos restantes da expressão dada acima convergem em probabilidade para zero. Por exemplo, usando agora a condição de que $|x_{ij}| \leq c$, para todo i e j , tem-se que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 [x_i^\top (\hat{\beta}_{MQ} - \beta)]^2 &= \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \left[\sum_{j=1}^k x_{ij} (\hat{\beta}_{jMQ} - \beta_j) \right]^2 \\ &\leq \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \left[\sum_{j=1}^k |x_{ij} (\hat{\beta}_{jMQ} - \beta_j)| \right]^2 \\ &\leq c^2 \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \left[\sum_{j=1}^k |\hat{\beta}_{jMQ} - \beta_j| \right]^2, \end{aligned}$$

de onde segue que

$$0 \leq \text{plim} \left\{ \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 [x_i^\top (\hat{\beta}_{MQ} - \beta)]^2 \right\}$$

$$\begin{aligned} &\leq c^2 \text{plim} \left\{ \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \right\} \text{plim} \left\{ \left[\sum_{j=1}^k |\hat{\beta}_{jMQ} - \beta_j| \right]^2 \right\} \\ &= 0 , \end{aligned}$$

uma vez que $\text{plim} \left\{ \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \right\} = \sigma^2 < \infty$ e $\text{plim} \hat{\beta}_{jMQ} = \beta_j$, $j = 1, \dots, k$. Assim, quando ν é desconhecido e $\nu > 4$, um estimador consistente de τ^2 segue de substituir ν por $\hat{\nu}_{MQ}$ em (4.30).

Por outro lado, suponha que $V_1, V_2, \dots, V_n$ são variáveis aleatórias iid com distribuição $GI(\nu/2, \nu/2)$ e tais que

$$(Y_i \mid V_i = v_i) \sim N_1(x_i^\top \beta, v_i \tau^2) , \quad i = 1, 2, \dots, n , \quad (4.32)$$

são independentes. Então, é fácil ver que $Y = (Y_1, Y_2, \dots, Y_n)^\top \sim t_n^{(I)}(X\beta, \tau^2 I_n; \nu)$. Definindo o vetor $V = (V_1, V_2, \dots, V_n)^\top$ e a matriz $D(V) = \text{Diag} \{V_1, V_2, \dots, V_n\}$ e usando o fato de que $(Y_i \mid V = v) =^d (Y_i \mid V_i = v_i)$, $i = 1, 2, \dots, n$, então a suposição (4.32) pode ser expressa (matricialmente) como

$$(Y \mid V = v) \sim N_n(X\beta, \tau^2 D(v)) , \quad (4.33)$$

onde $V \sim F_V^{(I)}(v) = \prod_{i=1}^n F_{V_i}(v_i)$, com F_{V_i} sendo a distribuição $GI(\nu/2, \nu/2)$. Com esta notação, a distribuição exata de $\hat{\beta}_{MQ} = CY$, onde $C = (X^\top X)^{-1} X^\top$, pode ser obtida usando o fato de que

$$(\hat{\beta}_{MQ} \mid V = v) \sim N_k(\beta, CD(v)C^\top) .$$

Para ilustrar, suponha que $k = 1$ com $X = (x_1, x_2, \dots, x_n)^\top$. Então $\hat{\beta}_{MQ} = \sum_{i=1}^n c_i Y_i$ e $CD(v)C^\top = \sum_{i=1}^n c_i^2 v_i$, onde $c_i = x_i / \sum_{i=1}^n x_i^2$. Seja $U = \sum_{i=1}^n U_i$, onde $U_i = c_i^2 V_i$, $i = 1, 2, \dots, n$. Observa-se que $U_1, U_2, \dots, U_n$ são variáveis aleatórias independentes com $U_i \sim GI(\nu/2, \nu c_i^2/2)$. Então, notando que $(\hat{\beta}_{MQ} \mid V = v) =^d (\hat{\beta}_{MQ} \mid U = u) \sim N_1(\beta, u\tau^2)$, tem-se que

$$p_{\hat{\beta}_{MQ}}(z) = \int_0^\infty (2\pi u \tau^2)^{-\frac{1}{2}} e^{-\frac{1}{2u\tau^2}(z-\beta)^2} dF_U(u) .$$

A dificuldade aqui é determinar a distribuição da variável aleatória $U = \sum_{i=1}^n c_i^2 V_i$ (ou da matriz aleatória $U = CD(V)C^\top$ no caso geral). Contudo, no caso geral, a função característica de $\hat{\beta}_{MQ} = CY$ é dada por

$$\phi_{\hat{\beta}_{MQ}}(s) = \prod_{i=1}^n \phi_{GI} \left\{ \left(\sum_{j=1}^k c_{ij} s_j \right)^2 \right\}, \quad s = (s_1, s_2, \dots, s_k)^\top \in \mathbb{R}^k,$$

onde ϕ_{GI} é a função característica da distribuição $GI(\nu/2, \nu/2)$ e $C = [c_{ij}] = (X^\top X)^{-1} X^\top$. De qualquer modo, como será provado na próxima seção, $\hat{\beta}_{MQ}$ não coincide com o EMV do modelo (4.29) para $\nu < \infty$.

4.3.1 Estimação de máxima verossimilhança

Considere-se primeiramente o caso geral quando o parâmetro extra ν é desconhecido. Seja $L = L(\theta)$, onde $\theta = (\beta^\top, \tau^2, \nu)^\top$, a função de verossimilhança do modelo (4.29). Então, de (4.27) segue que

$$L = n \left\{ \log k(1, \nu) + \frac{1}{2} \nu \log \nu - \frac{1}{2} \log \tau^2 \right\} - \frac{1}{2} (\nu + 1) \sum_{i=1}^n \log \{ \nu + t_i^2 \},$$

onde $t_i = t_i(\theta)$ está definido em (4.28) e

$$\log k(1, \nu) = \log \Gamma \left[\frac{1}{2} (\nu + 1) \right] - \log \Gamma \left(\frac{1}{2} \nu \right) - \frac{1}{2} \log \pi.$$

Escrevendo $\theta = (\beta_1, \beta_2, \dots, \beta_k, \tau^2, \nu)^\top$, $x_i^\top = (x_{i1}, x_{i2}, \dots, x_{ik})$ e observando que

$$\frac{\partial t_i^2}{\partial \beta_j} = -\frac{2}{\tau} t_i x_{ij}, \quad j = 1, 2, \dots, k \quad \text{e} \quad \frac{\partial t_i^2}{\partial \tau^2} = -\frac{t_i^2}{\tau^2},$$

tem-se que o EMV de $\theta = (\beta_1, \beta_2, \dots, \beta_k, \tau^2, \nu)^\top$, digamos $\hat{\theta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k, \hat{\tau}^2, \hat{\nu})^\top$, é a solução do sistema de equações dado por

$$\frac{\partial L}{\partial \beta_j} \Big|_{\theta=\hat{\theta}} = \frac{1}{\hat{\tau}} \sum_{i=1}^n \hat{\omega}_i \hat{t}_i x_{ij} = 0, \quad j = 1, 2, \dots, k, \quad (4.34)$$

$$\frac{\partial L}{\partial \tau^2} \Big|_{\theta=\hat{\theta}} = -\frac{n}{2\hat{\tau}^2} + \frac{1}{2\hat{\tau}^2} \sum_{i=1}^n \hat{\omega}_i \hat{t}_i^2 = 0, \quad (4.35)$$

e

$$\frac{\partial L}{\partial \nu} \Big|_{\theta=\hat{\theta}} = \frac{n}{2} \left\{ \psi \left[\frac{1}{2}(\hat{\nu} + 1) \right] - \psi \left(\frac{1}{2}\hat{\nu} \right) - \log(\hat{\nu} + 1) + \log \hat{\nu} + \log \hat{\omega}_g - \hat{\omega}_a + 1 \right\} = 0 , \quad (4.36)$$

onde $\psi(z) = d \log \Gamma(z) / dz$ é a função digama, $\bar{\omega}_g = (\prod_{i=1}^n \omega_i)^{1/n}$, $\bar{\omega}_a = (1/n) \sum_{i=1}^n \omega_i$ e $\omega_i = \omega_i(\theta)$, sendo definido por

$$\omega_i = \frac{\nu + 1}{\nu + t_i^2} , \quad i = 1, 2, \dots, n , \quad (4.37)$$

com $t_i = t_i(\theta)$ dado em (4.28) (aqui se $f = f(\theta)$ é uma função do parâmetro θ , então a notação $\hat{f}$ indica que $\hat{f} = f(\hat{\theta})$, onde $\hat{\theta}$ é o EMV de θ). Seja $\omega = \omega(\theta) = (\omega_1, \omega_2, \dots, \omega_n)^T$ e considere-se a matriz diagonal definida por

$$D(\omega) = \text{Diag} \{ \omega_1, \omega_2, \dots, \omega_n \} .$$

Então, é fácil ver que as equações (4.34) e (4.35) podem ser expressas de forma equivalente como

$$\hat{\beta} = (X^T D(\hat{\omega}) X)^{-1} X^T D(\hat{\omega}) y \quad \text{e} \quad \hat{\tau}^2 = \frac{1}{n} \hat{\epsilon}^T D(\hat{\omega}) \hat{\epsilon} , \quad (4.38)$$

respectivamente, onde $\hat{\omega} = \omega(\hat{\theta})$ e $\hat{\epsilon} = y - X\hat{\beta}$ denota o vetor de resíduos. De qualquer modo, as equações de log-verossimilhança dadas acima devem ser resolvidas iterativamente usando, por exemplo, o algoritmo de scoring ou o algoritmo EM. A implementação do algoritmo de scoring envolve a matriz de informação $K = K(\theta)$ do modelo (4.29). Esta matriz é derivada no Apêndice A.2 e é bloco diagonal entre o vetor de parâmetros de regressão β e $(\tau^2, \nu)^T$, digamos $K = \text{Diag} \{ K_1, K_2 \}$, onde

$$K_1 = [\kappa_{i,j}] = \frac{1}{\tau^2} \frac{\nu + 1}{\nu + 3} X^T X \quad \text{e} \quad K_2 = \begin{pmatrix} \kappa_{\tau^2, \tau^2} & \kappa_{\tau^2, \nu} \\ \kappa_{\nu, \tau^2} & \kappa_{\nu, \nu} \end{pmatrix} , \quad (4.39)$$

com

$$\kappa_{\tau^2, \tau^2} = \frac{n\nu}{2(\nu + 3)\tau^4} , \quad \kappa_{\nu, \tau^2} = \kappa_{\tau^2, \nu} = -\frac{n}{(\nu + 1)(\nu + 3)\tau^2}$$

e

$$\kappa_{\nu, \nu} = -\frac{n}{2} \left\{ \frac{1}{2} \psi' \left[\frac{1}{2}(\nu + 1) \right] - \frac{1}{2} \psi' \left(\frac{1}{2}\nu \right) + \frac{5}{\nu(\nu + 1)(\nu + 3)} \right\} .$$

Aqui ψ' denota a derivada da função digama ψ .

a. O algoritmo de scoring

Seja $\theta^{(l)} = (\beta^{(l)\top}, \tau^{2(l)}, \nu^{(l)})^\top$ a estimativa de θ na l -ésima iteração. Se $U = U(\theta) = \partial L / \partial \theta$ denota o vetor escore, então para o algoritmo de scoring tem-se que

$$\theta^{(l+1)} = \theta^{(l)} + \lambda_l (K^{(l)})^{-1} U^{(l)},$$

onde $K^{(l)} = K(\theta^{(l)})$, $U^{(l)} = U(\theta^{(l)})$ e $\lambda_l > 0$ é o comprimento ao passo l (usualmente escolhe-se $\lambda_l = 1$). Escrevendo $\theta = (\beta^\top, \theta_2^\top)$ e $U = (U_1^\top, U_2^\top)$, onde $\theta_2 = (\tau^2, \nu)^\top$, $U_1 = \partial L / \partial \beta$ e $U_2 = \partial L / \partial \theta_2$, então do fato de que $K = \text{Diag}\{K_1, K_2\}$ segue que a equação dada acima pode ser decomposta em

$$\begin{aligned} \beta^{(l+1)} &= \beta^{(l)} + \lambda_l (K_1^{(l)})^{-1} U_1^{(l)} \\ &= \beta^{(l)} + \lambda_l \frac{\nu^{(l)} + 3}{\nu^{(l)} + 1} (X^\top X)^{-1} X^\top D(\omega^{(l)}) \epsilon^{(l)} \end{aligned}$$

e

$$\theta_2^{(l+1)} = \theta_2^{(l)} + \lambda_l (K_2^{(l)})^{-1} U_2^{(l)},$$

onde $\epsilon^{(l)} = Y - X\beta^{(l)}$ é o vetor de resíduos da l -ésima iteração. Para ν conhecido tem-se que $\theta_2 = \tau^2$ e a última equação se reduz a

$$\tau^{2(l+1)} = \tau^{2(l)} + \frac{\lambda_l}{4} \frac{\nu}{\nu + 3} \left(\frac{1}{n} \epsilon^{(l)\top} D(\omega^{(l)}) \epsilon^{(l)} - \tau^{2(l)} \right).$$

O valor inicial $\theta^{(0)} = (\beta^{(0)\top}, \tau^{2(0)}, \nu^{(0)})^\top$ pode ser escolhido tomando, por exemplo, $\beta^{(0)} = \hat{\beta}_{MQ}$, $\tau^{2(0)} = \{(\nu^{(0)} - 2/\nu^{(0)})\} \hat{\sigma}_{MQ}^2$ e $\nu^{(0)} = \hat{\nu}_{MQ}$, que é uma estimativa consistente para $\nu > 4$. Uma outra alternativa é tomar $\beta^{(0)} = \hat{\beta}_{MQ}$, $\tau^{2(0)} = \hat{\sigma}_{MQ}^2$ e escolher $\nu^{(0)}$ de forma arbitrária em seu conjunto de valores admissíveis.

b. O algoritmo EM

O algoritmo EM (Dempster, Laird e Rubin (1977); veja também Little e Rubin (1987) e Little (1988)) aumenta os dados $Y = (Y_1, Y_2, \dots, Y_n)^\top$ considerando dados hipotéticos adicionais, digamos $V = (V_1, V_2, \dots, V_n)^\top$, de tal forma que o EMV de θ baseado na função de log-verossimilhança, $L_T(\theta)$, dos dados aumentados, $(Y^\top, V^\top)^\top$, seja fácil de obter. Se $\theta^{(l)}$ é o valor de θ obtido na l -ésima iteração, então a $(l+1)$ -ésima iteração do algoritmo EM consiste em dois passos: E e M . O passo E determina o valor esperado de $L_T(\theta)$ com respeito à distribuição condicional de V , dados $Y = y$ e $\theta^{(l)}$. O passo M maximiza a função resultante com respeito a θ e obtém a nova estimativa $\theta^{(l+1)}$ de θ . A implementação do algoritmo EM para obter os EMV dos parâmetros do modelo (4.29) é facilitada pelo fato de que este modelo pode ser especificado a partir de (4.33) (ou (4.32)). Assim, de (4.33) segue que a função de verossimilhança de $(Y^\top, V^\top)^\top$ é dada por

$$\begin{aligned} p_{Y,V}(y, v | \beta, \tau^2, \nu) &= p_Y(y | v; \beta, \tau^2) p_V(v | \nu) \\ &= \left\{ \frac{(\frac{1}{2}\nu)^{\frac{1}{2}\nu}}{(2\pi\tau^2)^{\frac{1}{2}}\Gamma(\frac{1}{2}\nu)} \right\}^n |D(v)|^{-\frac{1}{2}(\nu+3)} e^{\text{tr}\{-\frac{1}{2}D^{-1}(v)(\nu + tt^\top)\}}, \end{aligned}$$

de onde segue que

$$\begin{aligned} L_T(\theta) &= \log p_{Y,V}(y, v | \theta) \\ &= n \left\{ \frac{\nu}{2} \log\left(\frac{\nu}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log(2\pi\tau^2) \right\} - \frac{(\nu+3)}{2} \log |D(v)| \\ &\quad + \text{tr} \left\{ -\frac{1}{2} D^{-1}(v)(\nu + tt^\top) \right\}, \end{aligned}$$

onde $\theta = (\beta^\top, \tau^2, \nu)^\top$, $t = t(\theta) = (y - X\beta)/\tau$ e $D(v) = \text{Diag}\{v_1, v_2, \dots, v_n\}$. Consideremos agora $L_T(\theta)$ como função do vetor aleatório $(Y^\top, V^\top)^\top$. Usando o fato de que (Y_i, V_i) , $i = 1, 2, \dots, n$, são independentes, é fácil ver que (veja (3.25))

$$\begin{aligned} E(D^{-1}(V) | Y=y) &= \text{Diag}\{E(V_1^{-1} | Y_1=y_1), E(V_2^{-1} | Y_2=y_2), \dots, E(V_n^{-1} | Y_n=y_n)\} \\ &= \text{Diag}\{\omega_1, \omega_2, \dots, \omega_n\} \\ &= D(\omega), \end{aligned}$$

onde $\omega_i = \omega_i(\theta)$, $i = 1, 2, \dots, n$, estão definidos em (4.37). Similarmente,

$$\begin{aligned} E(\log |D(V)| \mid Y = y) &= \sum_{i=1}^n E(\log V_i \mid Y_i = y_i) \\ &= - \sum_{i=1}^n E(\log V_i^{-1} \mid Y_i = y_i) \\ &= - \sum_{i=1}^n \log \omega_i - nE(\log Z) + n \log(\nu + 1) , \end{aligned}$$

uma vez que $(V_i^{-1} \mid Y = y) =^d \omega_i Z / (\nu + 1)$, $i = 1, 2, \dots, n$, onde $Z \sim \chi_{\nu+1}^2$ (veja (3.24)).

Usando que $E(\log(Z/2)) = \psi[\frac{1}{2}(\nu + 1)]$, tem-se que

$$E(\log |D(V)| \mid Y = y) = - \log |D(\omega)| - n\psi[\frac{1}{2}(\nu + 1)] + n \log(\frac{\nu + 1}{2}) .$$

Assim, no passo E da iteração $(l + 1)$ do algoritmo EM tem-se que

$$\begin{aligned} L_E(\theta) &= E(L_T(\theta; Y, V) \mid Y = y, \theta^{(l)}) \\ &= n \left\{ \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log(2\pi\tau^2) \right\} \\ &\quad + \frac{(\nu + 3)}{2} \left\{ \log |D(\omega^{(l)})| + n\psi\left(\frac{\nu^{(l)} + 1}{2}\right) - n \log\left(\frac{\nu^{(l)} + 1}{2}\right) \right\} \\ &\quad + \text{tr} \left\{ -\frac{1}{2} D(\omega^{(l)}) (\nu + tt^\top) \right\} . \end{aligned}$$

Note-se que

$$\begin{aligned} L_E(\theta) &\propto n \left\{ \frac{\nu}{2} \left(\log \frac{\nu}{2} + \log \bar{\omega}_g^{(l)} - \bar{\omega}_a^{(l)} \right) - \log \Gamma\left(\frac{\nu}{2}\right) \right\} \\ &\quad - \frac{n}{2} \log \tau^2 - \frac{1}{2\tau^2} (y - X\beta)^\top D(\omega^{(l)}) (y - X\beta) , \end{aligned} \quad (4.40)$$

onde $\log \bar{\omega}_g = (1/n) \log |D(\omega)| = (1/n) \sum_{i=1}^n \log \omega_i$, $\bar{\omega}_a = (1/n) \text{tr}(D(\omega)) = (1/n) \sum_{i=1}^n \omega_i$ e $\omega^{(l)} = \omega(\theta^{(l)})$. Agora o algoritmo EM procede com o passo M , no qual a função $L_E(\theta)$ é maximizada com respeito a θ . De (4.40) é fácil ver que, neste passo, $\beta^{(l+1)}$ e $\tau^{2(l+1)}$ são dados por

$$\beta^{(l+1)} = (X^\top D(\omega^{(l)}) X)^{-1} X^\top D(\omega^{(l)}) y \quad \text{e} \quad \tau^{2(l+1)} = \frac{1}{n} \epsilon^{(l+1)} D(\omega^{(l)}) \epsilon^{(l+1)} , \quad (4.41)$$

onde $\epsilon^{(l+1)} = y - X\beta^{(l+1)}$ é o vetor de resíduos da iteração $l + 1$, enquanto que $\nu^{(l+1)}$ é a solução da equação dada por

$$\psi\left(\frac{\nu^{(l+1)}}{2}\right) - \log\left(\frac{\nu^{(l+1)}}{2}\right) = \bar{\omega}_a^{(l)} - \log \bar{\omega}_g^{(l)}. \quad (4.42)$$

Observe-se que os resultados dados em (4.41) são sugeridos pelas equações estabelecidas em (4.38). Assim, a conveniência do algoritmo EM é particularmente clara quando ν é conhecido, uma vez que o algoritmo se reduz a avaliar iterativamente as estimativas de mínimos quadrados ponderados (EMQP) dadas em (4.41). O valor inicial $\theta^{(0)}$ pode ser escolhido tomando $D(\omega^{(0)}) = I_n$, o que equivale a tomar $\beta^{(0)} = \hat{\beta}_{MQ}$ e $\tau^{2(0)} = \hat{\sigma}_{MQ}^2$. Para ν desconhecido, em lugar de resolver (4.42), um procedimento alternativo é considerar uma rede de valores para ν , aplicar o algoritmo EM para cada um destes valores e escolher aquela estimativa de θ que maximiza a verossimilhança de Y dada em (4.27). De outra forma, algoritmos auxiliares (por exemplo, o algoritmo de scoring) podem ser aplicados para resolver a equação (4.42), o que equivale a maximizar com respeito a ν a primeira parcela da relação dada em (4.40).

4.3.2 Robustez e outras propriedades do estimador de máxima verossimilhança

Para estudar as propriedades do EMV de θ relativo ao modelo (4.29), note-se em primeiro lugar que as contribuições de cada observação nas equações de estimação (4.34) e (4.35) de β e τ^2 , respectivamente, envolvem o fator ω_i definido em (4.37). Este fator pode ser interpretado como o peso da i -ésima observação na estimação de β e τ^2 . Assim, para $\nu = \infty$ (modelo normal), tem-se que $\omega_i = 1$, $i = 1, 2, \dots, n$, de modo que todas as observações têm o mesmo peso na estimação de β e τ^2 . No entanto, para $\nu < \infty$, tem-se que o peso $\omega_i = \omega(\nu, |t_i|)$ é uma função decrescente na distância $|t_i|^2 = (y_i - x_i^T \beta)^2 / \tau^2$. Além disto, de (4.37) é fácil ver também que $\omega_i < 1$, se $|t_i| > 1$; $\omega_i = 1$, se $|t_i| = 1$ e $\omega_i > 1$, se $|t_i| < 1$, qualquer que seja o valor de $\nu > 0$. Consequentemente, quando $\nu < \infty$ os

EMV de β e τ^2 são robustos no sentido de que as observações com um valor de t_i^2 grande, digamos $|t_i| > 1$, têm um peso $\omega_i < 1$, o qual decresce quando t_i^2 cresce. Observe-se também que para $|t_i| > 1$, $\omega_i = \omega(\nu, |t_i|)$ é uma função decrescente em ν , de modo que valores menores de ν implicam em pesos menores para observações com valores grandes em t_i^2 . Em outras palavras, ν é um parâmetro de robustificação, de modo que para cada $\nu < \infty$ fixado, os EMV de β e τ^2 pertencem à classe dos M -estimadores B -robustos (veja Hampell et al. (1986) ou Huber (1981)), uma vez que para cada $i = 1, 2, \dots, n$, as contribuições $c_\beta(t_i, x_{ij}) = x_{ij}\omega_i t_i$, $j = 1, 2, \dots, k$ e $c_{\tau^2}(t_i) = \omega_i t_i^2 - 1$ das equações de estimação de β e τ^2 , respectivamente, são funções de t_i limitadas. Em particular, $c_\beta(|t_i|, x_{ij}) = x_{ij}\omega_i |t_i| \rightarrow 0$, quando $|t_i| \rightarrow \infty$.

Note-se que os valores t_i , $i = 1, 2, \dots, n$, podem ser considerados como realizações independentes de uma variável aleatória $T \sim t_1(0, 1; \nu)$. Assim, ω_i , $i = 1, 2, \dots, n$, serão também realizações independentes da variável aleatória $W = (\nu + 1)/(\nu + T^2)$. É claro que a distribuição de W depende somente de ν e tem momentos finitos para todo $\nu > 0$. De fato, de (3.27) segue que

$$E(W^m) = \left(\frac{\nu + 1}{\nu}\right)^m \frac{\nu(\nu + 2) \cdots (\nu + 2(m - 1))}{(\nu + 1)(\nu + 3) \cdots (\nu + 1 + 2(m - 1))}, \quad m = 1, 2, \dots$$

Em particular, $E(W) = 1$. Desta forma, os valores $z_i = \omega_i t_i$, $i = 1, 2, \dots, n$, podem ser interpretados como n realizações independentes da variável aleatória definida por $Z = WT$, cuja distribuição também depende somente de ν , é simétrica com respeito à origem e tem momentos finitos para todo ν . Assim, Z tem momentos de ordem ímpar nulos e momentos de ordem par dados por

$$E(Z^{2m}) = \frac{(\nu + 1)^{2m}}{\nu^m} \frac{\nu(\nu + 2) \cdots (\nu + 2(m - 1))}{(\nu + 1)(\nu + 3) \cdots (\nu + 1 + 2(m - 1))}, \quad m = 1, 2, \dots$$

Em particular, $E(Z) = 0$ e $E(Z^2) = (\nu + 1)/(\nu + 3)$ de modo que $\{(\nu + 3)/(\nu + 1)\}^{\frac{1}{2}} z_i$, $i = 1, 2, \dots, n$, constituem resíduos padronizados.

Por outro lado, o fato de que β e $(\tau^2, \nu)^\top$ sejam globalmente ortogonais, no sentido de que a matriz de informação $K(\beta, \tau^2, \nu)$ do modelo (4.29) é bloco diagonal nas correspondentes submatrizes K_1 e K_2 relativas a β e $(\tau^2, \nu)^\top$, respectivamente, implica que os EMV $\hat{\beta}$ e $(\hat{\tau}^2, \hat{\nu})^\top$ são assintoticamente não correlacionados. Além disto, como o modelo (4.29) verifica as condições de regularidade sobre as quais o EMV do vetor de parâmetros é assintoticamente normal, então, sob a condição dada em (4.4), tem-se também que

$$(\hat{\beta}^\top, \hat{\tau}^2, \hat{\nu}) \sim AN_{k+2}((\beta^\top, \tau^\top, \nu)^\top, \text{Diag}\{K_1^{-1}, K_2^{-1}\}),$$

onde K_1 e K_2 estão definidas em (4.39). Em outras palavras, tem-se que

$$\hat{\beta} \sim AN_k(\beta, K_1^{-1}) \quad \text{e} \quad (\hat{\tau}^2, \hat{\nu})^\top \sim AN_2((\tau^2, \nu)^\top, K_2^{-1}), \quad (4.43)$$

e são assintoticamente independentes.

4.3.3 Parametrização ortogonal de $(\tau^2, \nu)^\top$

Será demonstrado aqui que a submatriz de informação, K_2 , de $(\tau^2, \nu)^\top$ pode ser “diagonalizada” via parametrização ortogonal de Cox e Reid (1987), sem alterar a submatriz de informação, K_1 , de β . De fato, como β e $(\tau^2, \nu)^\top$ são globalmente ortogonais segundo a definição de Cox e Reid (1987), então é suficiente aplicar o procedimento desenvolvido por estes autores para obter uma parametrização ortogonal de $(\tau^2, \nu)^\top$, assumindo que β é conhecido. Em tal caso, a matriz de informação do modelo é $K_2 = K_2(\tau^2, \nu)$. Suponha (por conveniência) que ν é o parâmetro de interesse. Sejam L^* e K_2^* a função verossimilhança e a matriz de informação da nova parametrização, digamos, $(\lambda, \nu)^\top$ ($\lambda > 0$). Então, como $(\tau^2, \nu)^\top = (\tau^2(\lambda, \nu), \nu)^\top$ e $L^*(\beta, \lambda, \nu) = L(\beta, \tau^2(\lambda, \nu), \nu)$, tem-se que

$$\frac{\partial L^*}{\partial \nu} = \frac{\partial L}{\partial \nu} + \frac{\partial L}{\partial \tau^2} \frac{\partial \tau^2}{\partial \nu},$$

de onde segue que

$$\frac{\partial^2 L^*}{\partial \nu \partial \lambda} = \frac{\partial^2 L}{\partial \nu \partial \tau^2} \frac{\partial \tau^2}{\partial \lambda} + \frac{\partial^2 L}{\partial (\tau^2)^2} \frac{\partial \tau^2}{\partial \lambda} \frac{\partial \tau^2}{\partial \nu} + \frac{\partial L}{\partial \tau^2} \frac{\partial^2 \tau^2}{\partial \nu \partial \lambda}.$$

Tomando a esperança nesta última equação, tem-se que

$$\kappa_{\nu,\lambda}^* = \kappa_{\nu,\tau^2} \frac{\partial \tau^2}{\partial \lambda} + \kappa_{\tau^2,\tau^2} \frac{\partial \tau^2}{\partial \lambda} \frac{\partial \tau^2}{\partial \nu} .$$

Como desejamos uma transformação de $(\tau^2, \nu)^\top$ para $(\lambda, \nu)^\top$ tal que $\kappa_{\nu,\lambda}^* = 0$ e $\partial \tau^2 / \partial \lambda \neq 0$, desta última equação juntamente com (4.39) segue que $(\lambda, \nu)^\top$ é a solução da equação diferencial

$$\frac{1}{\tau^2} \frac{\partial \tau^2}{\partial \nu} = 2 \left\{ \frac{1}{\nu} - \frac{1}{\nu + 1} \right\} ,$$

com solução $\tau^2 \left(\frac{\nu+1}{\nu} \right)^2 = \alpha(\lambda)$, onde α é uma função arbitrária de λ . Uma escolha conveniente é dada por $\alpha(\lambda) = \lambda$ e, portanto, tem-se que

$$\tau^2 = \lambda \left(\frac{\nu}{\nu + 1} \right)^2 \quad (\lambda > 0) . \quad (4.44)$$

Além disto, é fácil ver que

$$\begin{aligned} K_1^*(\beta, \lambda, \nu) &= -E \left(\frac{\partial^2 L^*}{\partial \beta \partial \beta^\top} \right) = -E \left(\frac{\partial^2 L}{\partial \beta \partial \beta^\top} \right) = K_1(\beta, \tau^2(\lambda, \nu), \nu) , \\ E \left(\frac{\partial^2 L^*}{\partial \nu \partial \beta} \right) &= E \left(\frac{\partial^2 L}{\partial \nu \partial \beta} + \frac{\partial^2 L}{\partial \tau^2 \partial \beta} \frac{\partial \tau^2}{\partial \nu} \right) = 0 , \\ E \left(\frac{\partial^2 L^*}{\partial \lambda \partial \beta} \right) &= E \left(\frac{\partial^2 L}{\partial \tau^2 \partial \beta} \frac{\partial \tau^2}{\partial \lambda} \right) = 0 , \\ \kappa_{\lambda,\lambda}^* &= -E \left(\frac{\partial^2 L^*}{\partial \lambda^2} \right) \\ &= -E \left(\frac{\partial^2 L}{\partial (\tau^2)^2} \left(\frac{\partial \tau^2}{\partial \lambda} \right)^2 + \frac{\partial L}{\partial \tau^2} \frac{\partial^2 \tau^2}{\partial \lambda^2} \right) = -\kappa_{\tau^2,\tau^2} \left(\frac{\partial \tau^2}{\partial \lambda} \right)^2 , \\ \kappa_{\nu,\nu}^* &= -E \left(\frac{\partial^2 L^*}{\partial \nu^2} \right) \\ &= -E \left(\frac{\partial^2 L}{\partial \nu^2} + 2 \frac{\partial^2 L}{\partial \nu \partial \tau^2} \frac{\partial \tau^2}{\partial \nu} + \frac{\partial^2 L}{\partial (\tau^2)^2} \left(\frac{\partial \tau^2}{\partial \nu} \right)^2 + \frac{\partial L}{\partial \tau^2} \frac{\partial^2 \tau^2}{\partial \nu^2} \right) \\ &= \kappa_{\nu,\nu} + 2\kappa_{\nu,\tau^2} \left(\frac{\partial \tau^2}{\partial \nu} \right) + \kappa_{\tau^2,\tau^2} \left(\frac{\partial \tau^2}{\partial \nu} \right)^2 \\ &= \kappa_{\nu,\nu} + \kappa_{\nu,\tau^2} \left(\frac{\partial \tau^2}{\partial \nu^2} \right) , \end{aligned}$$

onde os $\kappa_{i,j}$ são avaliados em $(\beta, \tau^2(\lambda, \nu), \nu)^\top = (\beta, \lambda(\frac{\nu}{\nu+1})^2, \nu)^\top$. Logo, para a nova matriz de informação $K^* = K^*(\beta, \lambda, \nu)$, tem-se que $K^* = \text{Diag} \{K_1^*, K_2^*\}$, onde (veja (4.39))

$$K_1^* = \frac{1}{\lambda} \left(\frac{\nu+1}{\nu} \right)^2 \left(\frac{\nu+1}{\nu+3} \right) X^\top X \quad \text{e} \quad K_2^* = \begin{pmatrix} \kappa_{\lambda,\lambda}^* & 0 \\ 0 & \kappa_{\nu,\nu}^* \end{pmatrix}, \quad (4.45)$$

com

$$\begin{aligned} \kappa_{\lambda,\lambda}^* &= \frac{n\nu}{2(\nu+3)\lambda^2}, \\ \kappa_{\nu,\nu}^* &= -\frac{n}{2} \left\{ \frac{1}{2} \psi' \left[\frac{1}{2}(\nu+1) \right] - \frac{1}{2} \psi' \left(\frac{1}{2}\nu \right) + \frac{\nu+5}{\nu(\nu+1)(\nu+3)} \right\} - \frac{2n}{\nu(\nu+1)^2(\nu+3)}. \end{aligned}$$

Por outro lado, sendo $\theta = (\beta^\top, \lambda, \nu)^\top$, em (4.29) tem-se que

$$Y = (Y_1, Y_2, \dots, Y_n)^\top \sim t_n^{(I)} \left(X\beta, \lambda \left(\frac{\nu}{\nu+1} \right)^2 I_n; \nu \right), \quad (4.46)$$

com função de log-verossimilhança dada por

$$L(\theta) = n \left\{ \log k(1, \nu) + \frac{1}{2} \nu \log \nu - \log \nu + \log(\nu+1) - \frac{1}{2} \log \lambda \right\} - \frac{1}{2}(\nu+1) \sum_{i=1}^n \log \{ \nu + t_i^2 \},$$

onde

$$\log k(1, \nu) = \log \Gamma \left[\frac{1}{2}(\nu+1) \right] - \log \Gamma \left(\frac{1}{2}\nu \right) - \frac{1}{2} \log \pi$$

e

$$t_i^2 = t_i(\theta) = (\nu+1)(y_i - x_i^\top \beta) / \nu \sqrt{\lambda}, \quad i = 1, 2, \dots, n. \quad (4.47)$$

Assim, o EMV $\hat{\theta}$ de θ é a solução das equações dadas por

$$\begin{aligned} \frac{\partial L}{\partial \beta_j} \Big|_{\theta=\hat{\theta}} &= \frac{\hat{\nu}+1}{\hat{\nu}\sqrt{\hat{\lambda}}} \sum_{i=1}^n \hat{\omega}_i \hat{t}_i x_{ij} = 0, \quad j = 1, 2, \dots, k, \\ \frac{\partial L}{\partial \lambda} \Big|_{\theta=\hat{\theta}} &= -\frac{n}{2\hat{\lambda}} + \frac{1}{2\hat{\lambda}} \sum_{i=1}^n \omega_i t_i^2 = 0 \end{aligned}$$

e

$$\begin{aligned} \frac{\partial L}{\partial \nu} \Big|_{\theta=\hat{\theta}} &= \frac{n}{2} \left\{ \psi \left[\frac{1}{2}(\hat{\nu}+1) \right] - \psi \left(\frac{1}{2}\hat{\nu} \right) - \log(\hat{\nu}+1) + \log \hat{\nu} + \log \hat{\omega}_g - \hat{\omega}_a + 1 \right\} \\ &+ \left\{ \frac{1}{\hat{\nu}} - \frac{1}{\hat{\nu}+1} \right\} \left\{ \sum_{i=1}^n \hat{\omega}_i \hat{t}_i^2 - n \right\} = 0, \end{aligned}$$

onde $\omega_i = \omega_i(\theta)$ é definido como em (4.37), com $t_i^2 = t_i^2(\theta)$ dado por (4.47) e $\bar{\omega}_g$ e $\bar{\omega}_a$ são as médias geométrica e aritmética de $\omega_1, \omega_2, \dots, \omega_n$. De (4.47) segue que a primeira e segunda destas equações são equivalentes a

$$\hat{\beta} = (X^T D(\hat{\omega}) X)^{-1} X^T D(\hat{\omega}) y \quad \text{e} \quad \hat{\lambda} \left(\frac{\hat{\nu}}{\hat{\nu} + 1} \right)^2 = \frac{1}{n} \hat{\epsilon}^T D(\hat{\omega}) \hat{\epsilon} ,$$

respectivamente, onde $D(\omega) = \text{Diag} \{ \omega_1, \omega_2, \dots, \omega_n \}$ e $\hat{\epsilon} = y - X\hat{\beta}$. A segunda equação implica também que $\sum_{i=1}^n \hat{\omega}_i \hat{t}_i^2 = n$, de modo que a terceira equação se reduz a

$$\psi \left[\frac{1}{2}(\hat{\nu} + 1) \right] - \psi \left(\frac{1}{2}\hat{\nu} \right) - \log(\hat{\nu} + 1) - \log \hat{\nu} + \log \hat{\omega}_g - \hat{\omega}_a + 1 = 0 ,$$

que é equivalente à equação de estimação de ν na parametrização original (veja a equação (4.36)). Assim, é claro que as inferências sobre β e ν não são alteradas com a nova parametrização. Além disso, dado o EMV $(\hat{\lambda}, \hat{\nu})^T$ de (λ, ν) , o EMV dos parâmetros de escala original τ^2 pode ser obtido de (4.45), isto é, como

$$\hat{\tau}^2 = \hat{\lambda} \left(\frac{\hat{\nu}}{\hat{\nu} + 1} \right)^2 .$$

Observe-se também que a nova parametrização não altera a implementação do algoritmo EM e simplifica o algoritmo de scoring, uma vez que, agora, a submatriz K_2^* é diagonal. Por último, para o EMV $(\hat{\beta}^T, \hat{\lambda}, \hat{\nu})^T$ de $(\beta^T, \lambda, \nu)^T$ tem-se que

$$\hat{\beta} \sim AN_k(\beta, K_1^{*-1}) , \quad \hat{\lambda} \sim N_1(\lambda, \kappa_{\lambda, \lambda}^{*-1}) \quad \text{e} \quad \hat{\nu} \sim N_1(\nu, \kappa_{\nu, \nu}^{*-1})$$

e são assintoticamente independentes, onde K_1^* , $\kappa_{\lambda, \lambda}^*$ e $\kappa_{\nu, \nu}^*$ estão dadas em (4.45).

Uma utilidade adicional da parametrização ortogonal é discutida na próxima seção.

4.3.4 Testes assintóticos da hipótese linear $H_0 : A\beta = \gamma_0$

Seja $Y \sim t_n^{(I)}(X\beta, \tau^2 I_n; \nu)$ o modelo de regressão t independente cuja estimação foi estudada na seção anterior. Suponha agora que deseja-se derivar um teste da hipótese linear $H_0 : A\beta = \gamma_0$, onde $A(q \times k)$ tem posto $r(A) = q \leq k$ e $\gamma_0(q \times 1)$ é conhecido. É claro

que sob este modelo qualquer teste de H_0 deve se basear na teoria assintótica. Assim, por exemplo, uma possibilidade é assumir que $\nu > 2$ e τ^2 são conhecidos e usar a distribuição assintótica de $\hat{\beta}_{MQ}$, dada em (4.5), para derivar a estatística dada por

$$W_{MQ} = (A\hat{\beta}_{MQ} - \gamma_0)^\top [\sigma^2 A(X^\top X)^{-1} A^\top]^{-1} (A\hat{\beta}_{MQ} - \gamma_0),$$

onde $\sigma^2 = \{\nu/(\nu - 2)\}\tau^2$. É fácil ver que sob H_0 , W_{MQ} converge em distribuição para uma variável aleatória tendo distribuição qui-quadrado com q graus de liberdade (χ_q^2), isto é, sob H_0 W_{MQ} tem distribuição assintótica χ_q^2 , o que será denotado por $W_{MQ} \sim A\chi_q^2$. Então o teste que rejeita H_0 para $W_{MQ}(y) > \chi_{q;1-\alpha}^2$ tem nível aproximadamente α para amostras suficientemente grandes. Se ainda, o parâmetro σ^2 é desconhecido, então ele pode ser substituído por seu estimador consistente $\hat{\sigma}_{MQ}^2$, de modo que este teste continua sendo válido assintoticamente. Contudo, este teste é baseado num estimador não robusto de β e não considera o fato de que o modelo assumido seja (4.29).

Três testes alternativos e assintoticamente equivalentes ao anterior, mas que incorporam este último fato, uma vez que baseiam-se no comportamento assintótico do EMV do modelo (4.29) e, portanto, numa estimativa robusta de β , são dados pelas estatísticas da razão de verossimilhança (G), escore (S) e de Wald (W). Estas estatísticas são derivadas no que segue para três suposições sobre o vetor de parâmetros θ do modelo (4.29). Sejam $\tilde{\theta}$ e $\hat{\theta}$ os EMV de θ sob $H_0 : A\beta = \gamma_0$ e sob $H_1 : A\beta \neq \gamma_0$, respectivamente. Então, $\tilde{\theta}$ é tal que $A\tilde{\beta} = \gamma_0$ e as estatísticas G , S e W podem ser obtidas como

$$G = 2(L(\hat{\theta}) - L(\tilde{\theta})), \quad S = U_1(\tilde{\theta})^\top K_1^{-1}(\tilde{\theta}) U_1(\tilde{\theta}) \quad \text{e} \quad W = (A\hat{\beta} - \gamma_0)^\top [AK_1^{-1}(\hat{\theta})A^\top]^{-1} (A\hat{\beta} - \gamma_0), \quad (4.48)$$

onde L é a função de log-verossimilhança, $U_1 = \partial L / \partial \beta$ e $K_1 = E\{U_1^\top U_1\} = -E\{\partial^2 L / \partial \beta \partial \beta^\top\}$ são o vetor escore e a matriz de informação relativos a β , respectivamente. Considerando (4.48) juntamente com (4.27), (4.34) e (4.39), os resultados que seguem são imediatos:

(a) Se τ^2 e ν são desconhecidos, então $\theta = (\beta^\top, \tau^2, \nu)^\top$ e

$$\begin{aligned} G = & n \left\{ 2 \left(\log \frac{\Gamma[\frac{1}{2}(\hat{\nu} + 1)]}{\Gamma[\frac{1}{2}(\tilde{\nu} + 1)]} + \log \frac{\Gamma(\frac{1}{2}\hat{\nu})}{\Gamma(\frac{1}{2}\tilde{\nu})} \right) + \hat{\nu} \log \hat{\nu} - \tilde{\nu} \log \tilde{\nu} \right. \\ & + (\tilde{\nu} + 1) \log(\tilde{\nu} + 1) - (\hat{\nu} + 1) \log(\hat{\nu} + 1) \} \\ & - n \left\{ (\tilde{\nu} + 1) \log \tilde{\omega}_g - (\hat{\nu} + 1) \log \hat{\omega}_g \right\} + n \left\{ \log \tilde{\tau}^2 - \log \hat{\tau}^2 \right\} , \end{aligned}$$

$$S = \frac{\tilde{\nu} + 3}{\tilde{\nu} + 1} \tilde{t}^\top D(\tilde{\omega}) P D(\tilde{\omega}) \tilde{t}$$

e

$$W = \frac{\hat{\nu} + 1}{\hat{\nu} + 3} \frac{1}{\tau^2} (A\hat{\beta} - \gamma_0)^\top [A(X^\top X)^{-1} A^\top] (A\hat{\beta} - \gamma_0) ,$$

onde t , ω , $\bar{\omega}_g$ e $D(\omega)$ estão definidos na seção anterior e $P = X(X^\top X)^{-1} X^\top$;

(b) Se ν é conhecido, então $\theta = (\beta^\top, \tau^2)^\top$, de modo que S e W são como em (a) com $\tilde{\nu} = \hat{\nu} = \nu$ e G se reduz a

$$G = n(\nu + 1) \left\{ \log \hat{\omega}_g - \log \tilde{\omega}_g \right\} + n \left\{ \log \tilde{\tau}^2 - \log \hat{\tau}^2 \right\} ;$$

(c) Se ν e τ^2 são conhecidos, então $\theta = \beta$; S e W são como em (a) com $\tilde{\nu} = \hat{\nu} = \nu$ e $\tilde{\tau}^2 = \hat{\tau}^2 = \tau^2$ e G se reduz a

$$G = n(\nu + 1) \left\{ \log \hat{\omega}_g - \log \tilde{\omega}_g \right\} .$$

Em cada uma destas situações pode ser demonstrado que sob a hipótese nula $H_0 : A\beta = \gamma_0$, as três estatísticas têm distribuição assintótica χ_q^2 . Em outras palavras, se T_n representa qualquer uma destas estatísticas, então sob $H_0 : A\beta = \gamma_0$ tem-se que

$$T_n =^d Z + Z_n ,$$

onde $Z \sim \chi_q^2$ e $\text{plim } Z_n = 0$ ou $Z_n = o_p(1)$. Assim, sob H_0 , $T_n \sim A\chi_q^2$ e o teste que rejeita H_0 para $T_n(y) > \chi_{q;1-\alpha}^2$ tem nível aproximadamente α quando n é suficientemente grande.

Nas seções que seguem o interesse está centrado nas estatísticas G e S para testar hipóteses da forma $H_0 : \beta_2 = \gamma_0$, onde β_2 é um subvetor de β . Por esta razão, se discutem a seguir alguns resultados adicionais relativos ao modelo restrito pela hipótese $H_0 : A\beta = \gamma_0$. Por simplicidade suponha que $\gamma_0 = 0$ e que as colunas linearmente independentes da matriz $A(q \times k)$ correspondem às q últimas (sem perda de generalidade nesta suposição, uma vez que os componentes do vetor β podem ser reordenados se necessário). Então, $A = [A_1, A_2]$, onde A_2 é uma matriz de dimensão $q \times q$ e, portanto, não singular. Logo, se $\beta = (\beta_1^\top, \beta_2^\top)^\top$ e $X = [X_1, X_2]$ são as partições induzidas no vetor β e na matriz X , então sob $H_0 : A\beta = 0$ tem-se que

$$\beta_2 = -A_2^{-1}A_1\beta_1 ,$$

de onde segue que

$$X\beta = X_1\beta_1 + X_2\beta_2 = X_A\beta_A ,$$

com $\beta_A = \beta_1$ e $X_A = X_1 - X_2A_2^{-1}A_1$. Note-se que a matriz X_A tem dimensão $n \times (k - q)$ e é de posto completo, isto é, $r(X_A) = k - q$ (veja Seber (1977, p.115)). Esta representação do modelo de regressão é denominado de forma “canônica”. Com estes resultados, o modelo (4.29) sujeito à restrição $A\beta = 0$ (o modelo restrito) pode ser expresso como

$$Y \sim t_n^{(I)}(X_A\beta_A, \tau^2 I_n; \nu) . \quad (4.49)$$

Assim, se θ_A denota o vetor de parâmetros deste modelo, então o EMV $\tilde{\theta}_A$ de θ_A pode ser obtido aplicando os métodos descritos na seção anterior para maximizar a função de log-verossimilhança $L_A = L_A(\theta_A)$ de (4.49) (uma forma alternativa é dada por maximizar a log-verossimilhança $L(\theta)$ de (4.49) sujeita à restrição $A\beta = 0$ usando o método de Lagrange). Com esta notação, é claro que G e S dependem da matriz A , já que sob $H_0 : A\beta = 0$ tem-se que $\eta = X\beta = X_A\beta_A = \eta_A$ o que implica que $t = (y - \eta)/\tau = (y - \eta_A)/\tau = t_A$ e $\omega = \omega(t) = \omega(t_A) = \omega_A$. Logo, $\tilde{t} = \tilde{t}_A$ e $\tilde{\omega} = \tilde{\omega}_A$, uma vez que $\tilde{\eta} = \tilde{\eta}_A$.

Note-se que para $\gamma_0 \neq 0$ é suficiente redefinir Y em (4.49) como $Y_A = Y - X_2 A_2^{-1} \gamma_0$. Em tal caso, t_A é dado por $t_A = (y_A - \eta_A)/\tau$. Finalmente, para o caso particular em que $A_1 = 0$ e $A_2 = I_q$, isto é, para $H_0 : \beta_2 = \gamma_0$ ($\beta_2(q \times 1)$ é um subvetor de β) tem-se que $X_A = X_1$ e $t_A = (y - X_1 \beta_1 - X_2 \gamma_0)/\tau$. Além disso, usando o fato de que $X_1^\top D(\tilde{\omega}) \tilde{t} = X_A^\top D(\tilde{\omega}_A) \tilde{t}_A = 0$, tem-se também que $X^\top D(\tilde{\omega}) \tilde{t} = (\tilde{t}^\top D(\omega) X_1, \tilde{t}^\top D(\tilde{\omega}) X_2)^\top = (0^\top, \tilde{t}^\top D(\tilde{\omega}) X_2)^\top$, de onde segue que a expressão da estatística S dada em (a) se reduz a

$$S = \frac{\tilde{\nu} + 3}{\tilde{\nu} + 1} \tilde{t}^\top D(\tilde{\omega}) X_2 [X_2^\top M_1 X_2]^{-1} X_2^\top D(\tilde{\omega}) \tilde{t},$$

onde $M_1 = I_n - P_1$ e $P_1 = X_1(X_1^\top X_1)^{-1} X_1^\top$ são as matrizes de projeção do modelo reduzido $Y_A = Y - X_2 \gamma_0 \sim t_n(X_1 \beta_1, \tau^2 I_n; \nu)$.

4.4 Testes assintóticos modificados por fatores tipo Bartlett no modelo de regressão t independente

4.4.1 Resultados teóricos gerais

A utilidade dos testes assintóticos como os descritos na seção anterior pode ser limitada em algumas situações práticas, uma vez que eles requerem amostras suficientemente grandes para que a distribuição da estatística do teste seja próxima da distribuição de referência. Contudo, em problemas que envolvem uma função de verossimilhança regular, isto é, quando é permitida a troca de derivadas da função de verossimilhança com respeito aos parâmetros com integrais sobre o espaço amostral, a aproximação das distribuições das estatísticas G e S com a distribuição de referência qui-quadrado pode ser melhorada, em modelos com distribuição amostral contínua, multiplicando estas estatísticas por fatores de correção do tipo Bartlett. Este procedimento foi originalmente proposto por Bartlett (1937) para o teste de homogeneidade de variâncias em populações normais. Posteriormente, Lawley (1956) desenvolve o fator de Bartlett da estatística G num contexto bem geral (veja também Hayakawa (1977), Barndorff-Nielsen e Cox (1984), Cordeiro (1987)

e McCullagh e Cox (1986)). Recentemente, Cordeiro e Ferrari (1991) mostram que o procedimento também pode ser aplicado para a estatística S , mas neste caso o fator de correção obtido é uma função polinomial de segundo grau na própria estatística. Deste modo, estes autores argumentam que em amostras finitas, em particular em amostras de tamanho moderado, as estatísticas modificadas deverão ter um melhor desempenho que as respectivas estatísticas originais, uma vez que as modificadas têm distribuições mais próximas da distribuição qui-quadrado de referência que as não corrigidas. Para as estatísticas corrigidas tal aproximação é garantida pelo menos até ordem n^{-1} (com n sendo o tamanho da amostra), no sentido de que sob H_0 as funções de distribuição acumulada (FDA) destas estatísticas coincidem com a FDA qui-quadrado de referência até ordem n^{-1} (ignorando os termos de ordem inferior a n^{-1}). Como será visto mais adiante, os fatores de correção de ambas as estatísticas são funções de cumulantes conjuntos de derivadas (com respeito aos parâmetros) até quarta ordem da log-verossimilhança, de modo que suas derivações podem envolver desenvolvimentos algébricos muito trabalhosos em alguns problemas particulares. Este fato tem motivado diferentes trabalhos com o propósito de obter fórmulas para estes fatores que sejam de fácil implementação computacional e que englobem uma ampla gama de modelos probabilísticos. Neste sentido, Cordeiro (1983), Cordeiro e Paula (1989) e Cordeiro, Paula e Botter (1993) desenvolvem uma formulação matricial do fator de Bartlett da estatística G no contexto dos modelos lineares generalizados e suas extensões. Da mesma forma, Cordeiro (1993) deriva este fator no modelo de regressão normal heterocedástico. Similarmente, expressões matriciais dos coeficientes envolvidos no fator de correção da estatística S são obtidas por Cordeiro, Ferrari e Paula (1993) e Cribari-Neto e Ferrari (1995b) para os modelos lineares generalizados e em Cribari-Neto e Ferrari (1995a) para o teste de homogeneidade de variâncias no modelo de regressão normal. Estes autores também mostraram via simulação que o desempenho do teste escore modificado é melhor que o do teste escore usual em diferentes situações

particulares. As estatísticas G e S corrigidas do modelo em estudo (dado em (4.29)) são derivadas na próxima seção. A análise será restrita ao teste de $H_0 : \beta_2 = \gamma_0$ versus $H_1 : \beta_2 \neq \gamma_0$, com β_2 sendo um subvetor com q componentes arbitrários (mas diferentes) do vetor de parâmetros de regressão β . Contudo, dos resultados obtidos para este caso particular, pode-se deduzir que no caso geral da hipótese linear $H_0 : A\beta = \gamma_0$, é suficiente substituir a matriz X_1 pela matriz $X_A = X_1 - X_2 A_2^{-1} A_1$ na derivação dos fatores de correção das estatísticas.

Para derivar as estatísticas G e S modificadas no modelo (4.29), a seguinte notação é requerida. Seja $Y(n \times 1)$ um vetor aleatório com função log-verossimilhança $L(\theta) = \log p(Y|\theta)$ e denote por

$$L_i = \frac{\partial L}{\partial \theta_i}, \quad L_{ij} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j}, \quad L_{ijk} = \frac{\partial^3 L}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad \text{etc.},$$

suas derivadas parciais relativas aos componentes de $\theta = (\theta_1, \theta_2, \dots, \theta_p)^\top$. Para estas derivadas e seus produtos definem-se os cumulantes

$$\begin{aligned} \kappa_{ij} &= E(L_{ij}), \quad \kappa_{ijk} = E(L_{ijk}), \quad \text{etc.}, \\ \kappa_{i,j} &= E(L_i L_j), \quad \kappa_{i,j,k} = E(L_i L_j L_k), \quad \kappa_{i,jk} = E(L_i L_{jk}), \\ \kappa_{ij,kl} &= E(L_{ij} L_{kl}) - \kappa_{ij} \kappa_{kl}, \quad \kappa_{i,j,kl} = E(L_i L_j L_{kl}) - \kappa_{i,j} \kappa_{kl}, \\ \kappa_{i,j,k,l} &= E(L_i L_j L_k L_l) - \kappa_{i,j} \kappa_{k,l} - \kappa_{i,k} \kappa_{j,l} - \kappa_{i,l} \kappa_{j,k}, \quad \text{etc.} \end{aligned}$$

Similarmente, as derivadas destes cumulantes com respeito aos parâmetros serão denotadas por

$$\kappa_{ij}^{(k)} = \frac{\partial \kappa_{ij}}{\partial \theta_k}, \quad \kappa_{ij}^{(kl)} = \frac{\partial^2 \kappa_{ij}}{\partial \theta_k \partial \theta_l}, \quad \text{etc.}$$

Se $L = L(\theta)$ é a log-verossimilhança de um modelo regular, de modo que suas derivadas L_i , L_{ij} , etc. são funções tais que é admissível a troca de derivadas por integrais nas expressões $\partial E(L_i)/\partial \theta_j$, $\partial E(L_{ij})/\partial \theta_k$, etc. (isto é possível, por exemplo, se L_i , L_{ij} , etc. são dominadas por funções integráveis sob o espaço amostral), então os cumulantes verificam relações

que podem facilitar seus cálculos. Relações deste tipo são, por exemplo, $\kappa_{i,j} = -\kappa_{ij}$, $\kappa_{ij}^{(k)} = \kappa_{ijk} + \kappa_{ij,k}$, $\kappa_{ij}^{(k)} + \kappa_{i,j,k} + \kappa_{i,jk} + \kappa_{j,ik} = 0$, $\kappa_{ijk}^{(l)} = \kappa_{ijkl} + \kappa_{l,ijk}$, $\kappa_{i,j,k} = -\kappa_{ijk} - \sum_{(3)} \kappa_{i,jk}$, $\kappa_{i,j,k} = 2\kappa_{ijk} - \sum_{(3)} \kappa_{ij}^{(k)}$, $\kappa_{i,j,k,l} = -3\kappa_{ijkl} + 2 \sum_{(4)} \kappa_{ijk}^{(l)} - \sum_{(6)} \kappa_{ij}^{(kl)} + \sum_{(3)} \kappa_{ij,k,l}$ e $\kappa_{i,j,kl} = \kappa_{ijkl} - 2\kappa_{ikl}^{(j)} + \kappa_{kl}^{(ij)} - \kappa_{ij,kl}$, onde $\sum_{(k)}$ denota o somatório que varre todas as k permutações de índices; por exemplo, $\sum_{(2)} \kappa_{i,jk} = \kappa_{i,jk} + \kappa_{j,ik} + \kappa_{k,ij}$. Assim, se K e K^{-1} são a matriz de informação e sua inversa, então tem-se que $K = [\kappa_{i,j}] = [-\kappa_{ij}]$ e $K^{-1} = [\kappa^{i,j}] = [-\kappa^{ij}]$. Seja $\hat{\theta}$ o EMV de θ . Com esta notação, Lawley (1956) expande (em torno de θ) a função $2(L(\hat{\theta}) - L(\theta))$ e mostra que $E\{2(L(\hat{\theta}) - L(\theta))\} = p + e_p + O(n^{-2})$, onde o termo e_p é de ordem n^{-1} e é dado por

$$e_p = \sum' \kappa^{ij} \kappa^{kr} \left\{ \frac{1}{4} \kappa_{ijk} \kappa_{jkr} + \kappa_{ijk}^{(r)} + \kappa_{ik}^{(jr)} \right\} - \sum' \kappa^{ij} \kappa^{kr} \kappa^{st} \left\{ \frac{1}{6} \kappa_{iks} \kappa_{jrt} + \frac{1}{4} \kappa_{ikr} \kappa_{jst} - \kappa_{iks} \kappa_{jt}^{(r)} - \kappa_{ikr} \kappa_{jt}^{(s)} + \kappa_{ik}^{(s)} \kappa_{jt}^{(r)} + \kappa_{ik}^{(r)} \kappa_{jt}^{(s)} \right\}, \quad (4.50)$$

onde $\sum'$ denota a soma sobre os parâmetros envolvidos (neste caso geral sobre $\theta_1, \theta_2, \dots, \theta_p$).

Suponha agora que θ é particionado como $\theta = (\theta_1^\top, \theta_2^\top)^\top$, onde θ_2 é um vetor de dimensão $q \times 1$, e considere o teste da razão de verossimilhança da hipótese composta $H_0 : \theta_2 = \gamma_0$ (contra $H_1 : \theta_2 \neq \gamma_0$). Sob H_0 , tem-se que $\theta = (\theta_1^\top, \gamma_0^\top)^\top$ e seja então $\tilde{\theta} = (\tilde{\theta}_1^\top, \gamma_0^\top)^\top$ o respectivo EMV. Procedendo de forma similar para $2(L(\tilde{\theta}) - L(\theta))$, tem-se que $E\{2(L(\tilde{\theta}) - L(\theta))\} = p - q + e_{p-q} + O(n^{-2})$, onde e_{p-q} é derivado da mesma forma que e_p , isto é, a partir de (4.50), mas somando somente sobre os $p - q$ componentes de θ_1 . Seja, finalmente, G a estatística da razão de verossimilhança para o teste de $H_0 : \theta_2 = \gamma_0$ contra $H_1 : \theta_2 \neq \gamma_0$. Então, notando que $G = 2(L(\hat{\theta}) - L(\theta)) - 2(L(\tilde{\theta}) - L(\theta))$, sob $H_0 : \theta_2 = \gamma_0$, tem-se que

$$E(G) = q\{1 + d\} + O(n^{-2}),$$

onde

$$d = \frac{e_p - e_{p-q}}{q} \quad (4.51)$$

é um termo de ordem n^{-1} ($d = d_n = O(n^{-1})$). Daqui segue que a estatística G multiplicada pelo fator de Bartlett $\{1 + d\}^{-1}$ (ou $\{1 - d\}$), digamos

$$G^* = \{1 + d\}^{-1}G \quad (4.52)$$

(ou $G^* = \{1 - d\}G$) é tal que $E(G^*) = q + O(n^{-2})$, sob a hipótese nula. Note-se que $E(G) = q + O(n^{-1})$, sob H_0 . Obviamente que, sob H_0 , G^* tem distribuição assintótica χ_q^2 , isto é, $G^* \sim A\chi_q^2$, uma vez que $G \sim A\chi_q^2$ e $d = d_n = O(n^{-1}) = o(1)$. Lawley (1956) também mostra que, sob H_0 , $E(\{G^*\}^m) = \mu_m + O(n^{-2})$, $m = 1, 2, \dots$, onde $\mu_m = q(q+2) \cdots (q+2(m-1))$ é o momento de ordem m da distribuição χ_q^2 , enquanto que $E(G^m) = \mu_m + O(n^{-1})$, $m = 1, 2, \dots$. Isto prova que, sob H_0 , a estatística G^* tem todos seus momentos mais próximos dos respectivos momentos da distribuição de referência χ_q^2 que a estatística original G .

Por outro lado, resultados similares aos obtidos em Lawley (1956) são desenvolvidos recentemente por Cordeiro e Ferrari (1991) para a estatística escore S . Sob as mesmas condições descritas acima, estes autores derivam um fator tipo Bartlett para corrigir a estatística S , de modo que, uma vez corrigida, esta estatística tenha distribuição χ_q^2 até ordem n^{-1} sob a hipótese nula. Neste caso, o fator obtido envolve uma função polinomial na estatística original. Assim, a estatística escore modificada é dada por

$$S^* = \{1 + c + bS + aS^2\}^{-1}S \quad (4.53)$$

(ou $S^* = \{1 - c - bS - aS^2\}S$), onde os coeficientes c, b e a são de ordem n^{-1} e são dados por

$$c = \frac{a_1 - a_2 + a_3}{12q}, \quad b = \frac{a_2 - 2a_3}{12q(q+2)} \quad \text{e} \quad a = \frac{a_3}{12q(q+2)(q+4)}. \quad (4.54)$$

Para definir a_1, a_2 e a_3 , a notação que segue é requerida. A partição de θ em θ_1 e θ_2 induz as seguintes partições da matriz de informação K e sua inversa K^{-1} ,

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \quad \text{e} \quad K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}.$$

Agora, sob $H_0 : \theta_2 = \gamma_0$, tem-se que a matriz de informação se reduz a K_{11} . Sejam $A_0 = [\alpha_{ij}]$ e $C = [\gamma_{ij}]$ as matrizes de dimensão $p \times p$ definidas como

$$A_0 = \begin{pmatrix} K_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{e} \quad C = K^{-1} - A_0 .$$

Assim, em termos desta notação, os coeficientes a_1 , a_2 e a_3 são dados por (Harris (1985); veja também Cordeiro, Ferrari e Paula (1993))

$$\begin{aligned} a_1 = & 3\sum'(\kappa_{ijk} + 2\kappa_{i,j,k})\alpha_{ij}\alpha_{st}\gamma_{kr} - 6\sum'(\kappa_{ijk} + 2\kappa_{i,j,k})\kappa_{r,s,t}\alpha_{ij}\alpha_{kr}\gamma_{st} \\ & + 6\sum'(\kappa_{i,j,k} - \kappa_{i,j,k})(\kappa_{rst} + 2\kappa_{r,s,t})\alpha_{js}\alpha_{kt}\gamma_{ir} - 6\sum'(\kappa_{i,j,k,r} + \kappa_{i,j,k,r})\alpha_{kr}\gamma_{ij} \end{aligned} \quad (4.55)$$

$$\begin{aligned} a_2 = & -3\sum'\kappa_{i,j,k}\alpha_{kr}\gamma_{ij}\gamma_{st} + 6\sum'(\kappa_{ijk} + 2\kappa_{i,j,k})\kappa_{r,s,t}\alpha_{ij}\gamma_{kr}\gamma_{st} \\ & - 6\sum'\kappa_{i,j,k}\kappa_{r,s,t}\alpha_{kt}\gamma_{ir}\gamma_{js} + 3\sum'\kappa_{i,j,k,r}\gamma_{ij}\gamma_{kr} \end{aligned} \quad (4.56)$$

e

$$a_3 = 3\sum'\kappa_{i,j,k}\kappa_{r,s,t}\gamma_{ij}\gamma_{kr}\gamma_{st} + 2\sum'\kappa_{i,j,k}\kappa_{r,s,t}\gamma_{ir}\gamma_{js}\gamma_{kt} . \quad (4.57)$$

Note-se, finalmente, que os coeficientes a , b , c e d podem depender de θ . Em tal caso, estes coeficientes podem ser estimados consistentemente usando o EMV de θ sob a hipótese nula sem alterar a ordem da aproximação entre as distribuições das estatísticas modificadas e distribuição de referência χ_q^2 (veja Lawley (1956) e Cordeiro e Ferrari (1991)).

4.4.2 Fatores de correção supondo τ^2 e ν conhecidos

Considere-se o modelo de regressão t independente definido em (4.29). Tal modelo verifica as condições de regularidade sobre as quais os resultados apresentados na Seção 4.4.1 são válidos. Assim, as estatísticas G^* e S^* podem ser derivadas para diferentes hipóteses sobre os parâmetros deste modelo. Como já foi mencionado, aqui o interesse está centrado no teste de $H_0 : \beta_2 = \gamma_0$ contra $H_1 : \beta_2 \neq \gamma_0$, onde β_2 é um subvetor com q componentes do vetor de parâmetros de regressão β . Começamos supondo que τ^2 e ν são conhecidos,

de modo que $\theta = \beta$ ($p = k$). Assim, de acordo com a notação desenvolvida na Seção 4.4.1, tem-se que $\theta_1 = \beta_1$ e $\theta_2 = \beta_2$. Considerando a correspondente partição da matriz do modelo $X = [X_1, X_2]$, então para a partição da matriz de informação estabelecida anteriormente tem-se que

$$K_{11} = \frac{\nu+1}{\nu+3} \frac{1}{\tau^2} X_1^\top X_1, \quad K_{12} = \frac{\nu+1}{\nu+3} \frac{1}{\tau^2} X_1^\top X_2 = K_{21}^\top \quad \text{e} \quad K_{22} = \frac{\nu+1}{\nu+3} \frac{1}{\tau^2} X_2^\top X_2.$$

Destes fatos podem ser obtidas também as matrizes $A_0 = \text{Diag}\{K_{11}^{-1}, 0\}$ e $C = K^{-1} - A_0$. Assim, para derivar os coeficientes a , b , c e d considerando τ^2 e ν conhecidos, resta somente calcular os cumulantes que envolvem os parâmetros β 's e então substituí-los nas expressões dadas para a , b , c e d . A derivação destes cumulantes é resumida no Apêndice A.2. A expressão (3.27) é de grande utilidade neste processo. Duas propriedades importantes dos cumulantes relativos ao modelo (4.29) são as seguintes:

- (a) Todos os cumulantes dependem somente de τ^2 e ν (independente de β);
- (b) Qualquer cumulante que envolva um número ímpar de parâmetros β é nulo.

Destes fatos segue que e_k , a_1 , a_2 e a_3 (veja (4.50) e (4.55) a (4.56)) se reduzem a

$$\begin{aligned} e_k &= \frac{1}{4} \sum' \kappa^{ij} \kappa^{kr} \kappa_{ijkr}, \\ a_1 &= -6 \sum' (\kappa_{i,j,k,r} + \kappa_{i,j,kr}) \alpha_{kr} \gamma_{ij}, \\ a_2 &= 3 \sum' \kappa_{i,j,k,r} \gamma_{ij} \gamma_{kr}, \\ a_3 &= 0. \end{aligned}$$

Note-se que o fato de que $a_3 = 0$ implica que em (4.54) tem-se que $a = 0$. Ao substituir os cumulantes envolvidos nestas últimas expressões de e_k , a_1 e a_2 (veja o Apêndice A.3), segue-se, após longa álgebra, que

$$e_k = \frac{3}{2} \frac{(\nu+2)(\nu+3)^2}{\nu(\nu+1)(\nu+3)(\nu+7)} \sum_{l=1}^n \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{i,j} x_{li} \kappa^{ij} x_{lj} \right) \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{k,r} x_{lr} \kappa^{kr} x_{lr} \right),$$

$$a_1 = 72 \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \sum_{l=1}^n \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{r,s} x_{lr} \alpha_{rs} x_{ls} \right) \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{i,j} x_{li} \gamma^{ij} x_{lj} \right)$$

e

$$a_2 = -54 \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \sum_{l=1}^n \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{i,j} x_{li} \gamma_{ij} x_{lj} \right) \left(\frac{\nu+1}{\nu+3} \frac{1}{\tau^2} \sum_{k,r} \gamma_{kr} x_{lr} \right) .$$

Considerem-se agora as matrizes de projeção $P = [p_{ij}] = X(X^\top X)^{-1}X^\top$ e $P_1 = [p_{1ij}] = X_1(X_1^\top X_1)^{-1}X_1^\top$, e note que $\hat{\eta} = X\hat{\beta}$ e $\tilde{\eta} = X\tilde{\beta} = X_1\tilde{\beta}_1 + X_2\gamma_0$ (os EMV do vetor de locação $\eta = X\beta$ sob H_1 e H_0 , respectivamente), têm covariâncias assintóticas dadas por

$$\text{Cov } A(\hat{\eta}) = XK^{-1}X^\top = \frac{\nu+3}{\nu+1}\tau^2P \quad \text{e} \quad \text{Cov } A(\tilde{\eta}) = XA_0X^\top = X_1K_{11}^{-1}X_1^\top = \frac{\nu+3}{\nu+1}\tau^2P_1 .$$

Observe também que

$$\text{Cov } A(\hat{\eta}) - \text{Cov } A(\tilde{\eta}) = XCX^\top = \frac{\nu+3}{\nu+1}\tau^2(P - P_1) .$$

Com estes últimos fatos é fácil ver que as expressões dadas acima para ε_k , a_1 e a_2 ficam reduzidas a

$$e_k = \frac{3(\nu+2)(\nu+3)^2}{\nu(\nu+1)(\nu+5)(\nu+7)} \sum_{l=1}^n p_{ll}^2 ,$$

$$a_1 = 72 \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \sum_{l=1}^n p_{1ll}(p_{ll} - p_{1ll}) ,$$

$$a_2 = -54 \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \sum_{l=1}^n (p_{ll} - p_{1ll})^2 .$$

O resultado para e_{k-q} segue de substituir $p_{ll}(P)$ por $p_{1ll}(P_1)$ na expressão obtida para e_k . Substituindo agora estes resultados em (4.51) e (4.54), tem-se para τ^2 e ν conhecidos ($\theta = \beta$) que d , c , b e a são dados por

$$d = \frac{3}{2nq} \frac{(\nu+2)(\nu+3)^2}{\nu(\nu+1)(\nu+5)(\nu+7)} \{\rho_{XX} - \rho_{X_1X_1}\} , \quad (4.58)$$

$$c = \frac{3}{2nq} \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \{3\rho_{XX} - 2\rho_{XX_1} + \rho_{X_1X_1}\} , \quad (4.59)$$

$$b = -\frac{9}{2nq(q+2)} \frac{\nu(\nu+4)-1}{\nu(\nu+5)(\nu+7)} \{\rho_{XX} - 2\rho_{XX_1} + \rho_{X_1X_1}\} \quad (4.60)$$

e

$$a = 0, \quad (4.61)$$

onde

$$\rho_{XX} = n \sum_{i=1}^n p_{ii}^2, \quad \rho_{XX_1} = n \sum_{i=1}^n p_{ii} p_{1ii}, \quad \rho_{X_1X_1} = n \sum_{i=1}^n p_{1ii}^2.$$

Para dar uma interpretação destes últimos termos, considere-se o modelo de regressão com um termo constante. Isto é, se $X = [x_{ij}]$ é a matriz do modelo, então se assume que $x_{i1} = 1, i = 1, 2, \dots, n$. Seja X^* a matriz de dimensão $n \times (k-1)$ cuja i -ésima linha é dada por $x_i^{*\top} = (x_{i2}, \dots, x_{ik}), i = 1, 2, \dots, n$. Então, X pode ser escrito como $X = [1_n, X^*]$ e sua i -ésima linha como $x_i^\top = (1, x_i^{*\top}), i = 1, 2, \dots, n$. Sejam $\bar{x}^* = n^{-1} \sum_{i=1}^n x_i^*$ e $S_{X^*X^*} = n^{-1} X^{*\top} X^* - \bar{x}^* \bar{x}^{*\top}$, o vetor de médias e a matriz de dispersão amostrais dos vetores $x_1^*, x_2^*, \dots, x_n^*$. Então, para a matriz $P = [p_{ij}] = X(X^\top X)^{-1} X^\top$ tem-se que $p_{ii} = n^{-1} \{1 + \delta_i\}$, onde $\delta_i = (x_i^* - \bar{x}^*)^\top S_{X^*X^*}^{-1} (x_i^* - \bar{x}^*)$. Note-se que $\bar{\delta} = n^{-1} \sum_{i=1}^n \delta_i = k-1$. Com esta notação é fácil ver que

$$\rho_{XX} = 1 + 2(k-1) + \frac{1}{n} \sum_{i=1}^n \delta_i^2.$$

Similarmente, se a coluna constante é incorporada à matriz X_1 , então para esta matriz tem-se também que

$$\rho_{X_1X_1} = 1 + 2(k-q-1) + \frac{1}{n} \sum_{i=1}^n \delta_{1i}^2$$

e

$$\rho_{XX_1} = 1 + (k-1) + (k-q-1) + \frac{1}{n} \sum_{i=1}^n \delta_i \delta_{1i},$$

onde δ_{1i} se define de forma análoga a δ_i , mas considerando a matriz X_1 . Note-se que $n^{-1} \sum_{i=1}^n \delta_i^2$ e $n^{-1} \sum_{i=1}^n \delta_{1i}^2$ são medidas de curtose multivariadas relativas às colunas não constantes em X e X_1 , respectivamente. Para ilustrar considerem-se os seguintes exemplos:

Exemplo 4.1 Sejam $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, $i = 1, 2, \dots, n$, a regressão linear simples. Suponha que $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são iid $t_1(0, \tau^2; \nu)$, onde τ^2 e ν são conhecidos. Considere-se o teste de $H_0 : \beta_2 = \gamma_0$ contra $H_1 : \beta_2 \neq \gamma_0$. Para este modelo tem-se que $k = 2$ e $q = 1$ com $X = [X_1, X_2]$, onde $X_1 = 1_n$ e $X_2 = (x_1, x_2, \dots, x_n)^\top$, de modo que $X^* = X_2$. Sejam $\bar{x} = n^{-1} \sum_{i=1}^n x_i$, $S_x^{(m)} = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^m$, $m = 2, 4$, e $\delta_x = S_x^{(4)} / (S_x^{(2)})^2$. Note-se que $\delta_x - 3$ é o coeficiente de curtose amostral de $x_1, x_2, \dots, x_n$. Então, neste modelo tem-se que $\rho_{XX} = 3 + \delta_x$, $\rho_{X_1 X_1} = 1$ e $\rho_{X X_1} = 2$. Colocando estes resultados em (4.58) a (4.60), tem-se que

$$d = \frac{3}{2n} \frac{(\nu + 2)(\nu + 3)^2}{\nu(\nu + 1)(\nu + 5)(\nu + 7)} (\delta_x + 2), \quad c = \frac{9}{2n} \frac{\nu(\nu + 4) - 1}{\nu(\nu + 5)(\nu + 7)} (\delta_x + 2),$$

$$b = -\frac{3}{2n} \frac{\nu(\nu + 4) - 1}{\nu(\nu + 5)(\nu + 7)} \delta_x.$$

Assim, para $\nu = 1$ (o modelo Cauchy) tem-se que $d = 3(\delta_x + 2)/4n$, $c = 3(\delta_x + 2)/8n$ e $b = -\delta_x/8n$.

Exemplo 4.2 Sejam $\{Y_{ij}; j = 1, 2, \dots, n_i\}$, $i = 1, 2, \dots, k$, amostras independentes de $k \geq 2$ populações distintas. Suponha que $Y_{ij} \sim t_1(\beta + \beta_i, \tau^2; \nu)$, onde $\beta_k = 0$ e τ^2, ν são assumidos conhecidos. Considere a hipótese de homogeneidade das locações das k populações, isto é, $H_0 : \beta_1 = \beta_2 = \dots = \beta_{k-1} = 0$. Para o teste de H_0 tem-se que $q = k - 1$, $\rho_{XX} = \rho = n \sum_{i=1}^k n_i^{-1}$, $\rho_{X X_1} = k$ e $\rho_{X_1 X_1} = 1$, de modo que

$$d = \frac{3}{2n(k-1)} \frac{(\nu + 2)(\nu + 3)^2}{\nu(\nu + 1)(\nu + 5)(\nu + 7)} (\rho - 1),$$

$$c = \frac{3}{2n(k-1)} \frac{\nu(\nu + 4) - 1}{\nu(\nu + 5)(\nu + 7)} (3\rho - 2k + 1),$$

$$b = -\frac{9}{2n(k^2 - 1)} \frac{\nu(\nu + 4) - 1}{\nu(\nu + 5)(\nu + 7)} (\rho - 2k + 1).$$

Em particular, para o modelo Cauchy ($\nu = 1$) tem-se que $d = 3(\rho - 1)/(4n(k - 1))$, $c = 3(3\rho - 2k + 1)/(8n(k - 1))$ e $b = -3(\rho - 2k + 1)/(8n(k^2 - 1))$.

Algumas propriedades interessantes dos coeficientes d , c , b e a obtidos em (4.58) a (4.61) são discutidas no que segue:

- (i) Em primeiro lugar, o fato de que $a = 0$ implica que para a estatística de escore, relativa ao modelo (4.29), o fator de correção se reduz a $\{1 + c + bS\}^{-1}$ (ou $\{1 - c - bS\}$), que é uma função crescente em S , uma vez que $b < 0$. Note-se que $a = 0$ para qualquer hipótese nula que envolva os parâmetros de locação do modelo t (ou um modelo elíptico), já que a propriedade (b) dos cumulantes implica que em (4.57) $a_3 = 0$. Portanto, em (4.54) tem-se que $a = 0$, $b = a_2/12q(q+2)$ e $c = (q+2)b + a_1/12q$.
- (ii) Em segundo lugar, note-se que estes coeficientes não dependem de β e τ^2 , e dependem de ν através de funções que são definidas para todo $\nu > 0$ e que se anulam quando $\nu = \infty$ (o modelo normal). De fato, para o modelo de regressão normal as estatísticas da razão de verossimilhança e escore, digamos $G_{(\infty)}$ e $S_{(\infty)}$, têm distribuição χ_q^2 exata sob a hipótese nula.
- (iii) Em terceiro lugar, observe-se que estes coeficientes dependem da matriz modelo $X = [X_1, X_2]$ somente através das matrizes de projeções $P = X(X^\top X)^{-1}X^\top$ e $P_1 = X_1(X_1^\top X_1)^{-1}X_1^\top$. Sabe-se que se $X_*(n \times k)$ é uma outra matriz cujas colunas são também linearmente independentes (matriz base), então $P_* = X_*(X_*^\top X_*)^{-1}X_*^\top = X(X^\top X)^{-1}X^\top = P$. Isto é, a matriz de projeção P independe da escolha da matriz base X . Além disto, existe $B(k \times k)$ inversível tal que $X = X_*B$ (veja Arnold (1981, p.38)). Deste fato, juntamente ao fato de que os coeficientes d , c , b e a não dependem de β , tem-se que estes coeficientes podem ser obtidos supondo, por exemplo, que a matriz base X multiplicada por $n^{-1/2}$ seja ortonormal, isto é, assumindo que $n^{-1}X^\top X = I_k$. Se $X = [x_{ij}]$, isto significa que $n^{-1} \sum_{i=1}^n x_{ij}^2 = 1$ e $\sum_{i=1}^n x_{ij}x_{ij'} = 0$, $j \neq j'$, $j, j' = 1, 2, \dots, k$. Suponha que $X = [X_1, X_2]$, com X_2 sendo de dimensão $n \times q$. Então $P = n^{-1}XX^\top = n^{-1}X_1X_1^\top + n^{-1}X_2X_2^\top = P_1 + P_2$, com $P_1P_2 = 0$. Em outras palavras, se $P = [p_{rs}]$ e $P_i = [p_{irs}]$, $i = 1, 2$, então

$p_{rs} = n^{-1} x_r^\top x_s^\top = n^{-1} x_r^{(1)\top} x_s^{(1)\top} + n^{-1} x_r^{(2)\top} x_s^{(2)\top} = p_{1rs} + p_{2rs}$, onde $x_r^\top$ e $x_r^{(i)\top}$, $i = 1, 2$, denotam a r -ésima linha de X e X_i , $i = 1, 2$, respectivamente, de modo que $x_r^\top = (x_r^{(1)\top}, x_r^{(2)\top})$. Desta forma, tem-se que

$$\rho_{XX} = n^{-1} \sum_{l=1}^n (x_l^\top x_l)^2 = \sum_j \rho_{jj} + 2 \sum_{j,j'} \rho_{jj'},$$

onde

$$\rho_{jj'} = n^{-1} \sum_{l=1}^n x_{lj}^2 x_{lj'}^2 = \rho_{j'j}, \quad j, j' = 1, 2, \dots, k.$$

Note-se que $\rho_{jj} - 3$ é o coeficiente de curtose amostral da j -ésima variável explicativa (coluna de X) do modelo. Similarmente,

$$\rho_{X_i X_i} = n^{-1} \sum_{l=1}^n (x_l^{(i)\top} x_l^{(i)})^2 = \sum_{j_i} \rho_{j_i j_i} + 2 \sum_{j_i j'_i} \rho_{j_i j'_i},$$

onde j_i e j'_i variam sobre as colunas da matriz X_i , $i = 1, 2$. Além disso, é fácil ver que $\rho_{XX} = \rho_{X_1 X_1} + \rho_{X_2 X_2} + 2\rho_{X_1 X_2}$ e $\rho_{X_2 X_2} = \rho_{XX} - 2\rho_{X X_1} + \rho_{X_1 X_1}$. Estes resultados se ilustram no exemplo que segue.

Exemplo 4.3 Seja $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, $i = 1, 2, \dots, n$, onde $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são iid $t_1(0, 1; 1)$. Suponha que $x_1, x_2, \dots, x_n$ estão medidos de modo que $\sum_{i=1}^n x_i = 0$ e $n^{-1} \sum_{i=1}^n x_i^2 = 1$. Então, para o teste de $H_0 : \beta_2 = \gamma_0$ contra $H_1 : \beta_2 \neq \gamma_0$, tem-se que $X_1 = 1_n$, $X_2 = (x_1, x_2, \dots, x_n)^\top$ com $k = 2$ e $q = 1$. Logo, $\rho_{11} = n^{-1} \sum_{i=1}^n 1^4 = 1$, $\rho_{12} = n^{-1} \sum_{i=1}^n 1^2 x_i^2 = 1$ e $\rho_{22} = n^{-1} \sum_{i=1}^n x_i^4 = \delta_x$, de modo que $\rho_{XX} = 3 + \delta_x$, $\rho_{X_1 X_1} = 1$, $\rho_{X_2 X_2} = \delta_x$ e $\rho_{X X_1} = (1/2)(\rho_{XX} + \rho_{X_1 X_1} - \rho_{X_2 X_2}) = 2$. Assim, para $\nu = 1$ em (4.58) a (4.60) tem-se que $d = 3(\delta_x + 2)/4n$, $c = 9(\delta_x + 8)/8n$ e $b = -3\delta_x/8n$ ($a = 0$), que são os mesmos resultados obtidos no Exemplo 4.1, uma vez que $\delta_x = S_x^{(4)}/(S_x^{(2)})^2$ quando $x_1, x_2, \dots, x_n$ não estão padronizados.

Finalmente, note-se que para $q = k$, isto é, quando $H_0 : \beta = \gamma_0$ é simples, tem-se que $X_1 = 0$ e $X_2 = X$, de modo que $\rho_{X_1 X_1} = \rho_{X X_1} = 0$ e $\rho_{X_2 X_2} = \rho_{XX}$. Além disso, se

$\mu_m = E(S^m)$, $m = 1, 2, \dots$, onde S é a estatística escore, então sob $H_0 : \beta = \gamma_0$ pode-se mostrar que

$$\begin{aligned}\mu_1 &= k, \quad \mu_2 = k(k+2) - 18h_1 \sum_{l=1}^n p_{ll}^2 \quad \text{e} \\ \mu_3 &= k(k+2)(k+4) - 54(k+4)h_1 \sum_{l=1}^n p_{ll}^2 + 15(h_2 - 6h_1h_2 + 2) \sum_{l=1}^n p_{ll}^3,\end{aligned}$$

onde $h_1 = \{\nu(\nu+4)-1\}/\{\nu(\nu+5)(\nu+7)\}$ e $h_2 = \{(\nu+1)(\nu+3)(\nu+4)\}/\{\nu(\nu+9)(\nu+11)\}$.

Note-se que a última parcela em μ_3 é de ordem n^{-2} (e deve ser ignorada). Dados μ_1, μ_2 e μ_3 (os três primeiros momentos de S sob a hipótese nula), os coeficientes a, b e c podem ser derivados a partir das seguintes relações (Harris (1985)):

$$\begin{aligned}\mu_1 &= q + \frac{a_1}{12} + O(n^{-2}), \quad \mu_2 = q(q+2) + \frac{(q+2)a_1 + 2a_2}{6} + O(n^{-2}) \quad \text{e} \\ \mu_3 &= q(q+2)(q+4) + \frac{(q+2)(q+4)a_1 + 4(q+4)a_2 + 8a_3}{4} + O(n^{-2}).\end{aligned}$$

Isto também é uma forma de conferir os resultados obtidos para a, b e c , uma vez que, sob a hipótese nula simples, μ_1, μ_2 e μ_3 são fáceis de serem calculados.

4.4.3 Fatores de correção supondo τ^2 desconhecido e ν conhecido

Denotemos por $d_\beta, c_\beta, b_\beta$ e a_β os coeficientes d, c, b e a obtidos em (4.58) a (4.61), isto é, supondo que τ^2 e ν são conhecidos. Suponha agora que τ^2 também é parâmetro desconhecido, de modo que somente ν está fixado num valor assumido conhecido. Considere o teste de $H_0 : \beta_2 = \gamma_0$ contra $H_1 : \beta_2 \neq \gamma_0$. Então, no que segue será demonstrado que

$$d = d_\beta + d_{\beta, \tau^2}, \quad c = c_\beta + c_{\beta, \tau^2}, \quad b = b_\beta + b_{\beta, \tau^2} \quad \text{e} \quad a = 0, \quad (4.62)$$

onde $d_{\beta, \tau^2}, b_{\beta, \tau^2}$ e c_{β, τ^2} representam as contribuições adicionais devidas ao fato de que τ^2 é também desconhecido e são dadas por

$$d_{\beta, \tau^2} = \frac{(2k-q)(\nu+2)^2(\nu+3)}{2n} + \frac{1}{n} \frac{(\nu+2)(\nu+3)(\nu^2+9\nu+2)}{\nu(\nu+5)^2(\nu+7)}, \quad (4.63)$$

$$c_{\beta, \tau^2} = \frac{q+2}{2n} \frac{(\nu-1)^2(\nu+3)}{\nu(\nu+5)^2} + \frac{k-q}{n} \frac{(\nu-1)(\nu+2)(\nu+3)}{\nu(\nu+5)^2} + \frac{6}{n} \frac{(\nu+1)(\nu+2)(\nu+3)}{\nu(\nu+5)^2(\nu+7)}, \quad (4.64)$$

e

$$b_{\beta, \tau^2} = -\frac{1}{2n} \frac{(\nu-1)^2(\nu+3)}{\nu(\nu+5)^2}. \quad (4.65)$$

Para provar estes resultados considere-se (por simplicidade) o parâmetro $\phi = 1/\tau^2$ em lugar de τ^2 . Isto não altera os resultados, uma vez que no modelo t que se estuda os coeficientes d , c , b e a , além de não dependerem da parametrização escolhida para o vetor de locação η , também independem da estrutura paramétrica da matriz de dispersão (aqui $\tau^2 I_n$).

Seja $\theta = (\beta^\top, \phi)^\top$ o vetor de parâmetros, cuja dimensão é $p = k+1$. Então, na partição $\theta = (\theta_1^\top, \theta_2^\top)^\top$ tem-se que $\theta_1 = (\beta_1^\top, \phi)^\top$ e $\theta_2 = \beta_2$, já que a hipótese nula considerada é $H_0 : \beta_2 = \gamma_0$. Assim, a correspondente partição da matriz de informação $K = K(\beta, \phi)$ é dada por

$$\begin{aligned} K_{11} &= \text{Diag} \left\{ \frac{\nu+1}{\nu+3} \phi X_1^\top X_1, \frac{n\nu}{2(\nu+3)\phi^2} \right\}, \quad K_{12} = \left[\frac{\nu+1}{\nu+3} \phi X_1^\top X_2, 0 \right] = K_{21}^\top, \\ K_{22} &= \frac{\nu+1}{\nu+3} \phi X_2^\top X_2, \end{aligned}$$

uma vez que $\kappa_{\phi, \phi} = (\partial \tau^2 / \partial \phi)^2 \kappa_{\tau^2, \tau^2} |_{\tau^2 = \phi} = n\nu / (2(\nu+3)\phi^2)$ (veja (4.39)).

Notando agora que e_{k+1} pode ser escrita como (veja (4.50))

$$e_{k+1} = e_{k;\beta} + e_{1;\phi} + e_{k+1;\beta;\phi},$$

onde $e_{k;\beta}$ é a contribuição de β supondo ϕ conhecido e foi derivada na Seção 4.4.2, $e_{1;\phi}$ é a contribuição de ϕ que resulta de supor β conhecido e $e_{k+1;\beta;\phi}$ é a contribuição conjunta de β e ϕ e é dada por aquelas parcelas de (4.50) que envolvem cumulantes conjuntos de ϕ com algum parâmetro β_j . Lembrando agora que todos os cumulantes independem de β e notando que os cumulantes relativos a ϕ não dependem da matriz X (veja o Apêndice

A.3), então sob $H_0 : \beta_2 = \gamma_0$ tem-se também que

$$e_{k-q+1} = e_{k-q;\beta_1} + e_{1;\phi} + e_{k-q+1;\beta_1,\phi} .$$

Usando (4.51) segue que $d = d_\beta + d_{\beta,\phi}$, onde

$$d_{\beta,\phi} = \frac{e_{k+1;\beta,\phi} - e_{k-q+1;\beta_1,\phi}}{q} . \quad (4.66)$$

Isto mostra a primeira identidade em (4.62), já que $d_{\beta,\phi} = d_{\beta,\tau^2}$.

Por outro lado, usando que β e ϕ são globalmente ortogonais e considerando as propriedades dos cumulantes de (4.29), em (4.50) tem-se após alguma álgebra que

$$\begin{aligned} e_{k+1;\beta,\phi} = & \frac{1}{2} \kappa^{\phi\phi} \sum_{\beta} \kappa^{ij} (\kappa_{ij\phi\phi} - 2\kappa_{ij\phi}^{(\phi)} - \frac{1}{2} (\kappa^{\phi\phi})^2 (\kappa_{\phi\phi\phi} - 2\kappa_{\phi\phi}^{(\phi)}) \sum_{\beta} \kappa^{ij} \kappa_{ij\phi} \\ & - \frac{1}{2} \kappa^{\phi\phi} \sum_{\beta} \kappa^{ij} \kappa^{rs} \kappa_{ir\phi} (\kappa_{js\phi} - 2\kappa_{js}^{(\phi)}) - \frac{1}{4} \kappa^{\phi\phi} \left(\sum_{ij} \kappa^{ij} \kappa_{ij\phi} \right)^2 , \end{aligned} \quad (4.67)$$

onde $\sum_{\beta}$ denota a soma sobre os parâmetros β 's. Substituindo os cumulantes envolvidos nesta última expressão (veja o Apêndice A.2), segue que (após uma certa álgebra)

$$e_{k+1;\beta,\phi} = \frac{k}{n} \frac{(\nu+2)(\nu+3)(\nu^2+9\nu+2)}{\nu(\nu+5)^2(\nu+7)} + \frac{k^2}{2n} \frac{(\nu+2)^2(\nu+3)}{\nu(\nu+5)^2} .$$

Para $e_{k-q+1;\beta_1,\phi}$ é suficiente substituir k por $(k-q)$ nesta última expressão. Colocando estes resultados em (4.66) tem-se (4.63).

Para os coeficientes a , b e c se procede de forma similar. Sejam $a_{1;\beta}$, $a_{2;\beta}$ e $a_{3;\beta}$ os valores de a_1 , a_2 e a_3 que resultam de (4.55), (4.56) e (4.57), respectivamente, quando ϕ é suposto conhecido (veja a Seção 4.4.2). Então, para ϕ desconhecido tem-se que

$$a_i = a_{i;\beta} + a_{i;\beta,\phi} , \quad i = 1, 2, 3 ,$$

onde $a_{i;\beta,\phi}$ é dado por aquelas parcelas que envolvem simultaneamente o parâmetro ϕ com algum parâmetro de β na respectiva expressão de a_i . Para derivar estes últimos termos, note-se que as definições das matrizes $A_0 = [\alpha_{ij}] = \text{Diag} \{K_{11}^{-1}, 0\}$ e $C = [\gamma_{ij}] = K^{-1} - A_0$ implicam que $\gamma_{\phi\phi} = 0$. Da mesma forma, $\alpha_{j\phi} = \gamma_{j\phi} = 0$, para todo $j = 1, 2, \dots, k$, uma

vez que β e ϕ são globalmente ortogonais. Se adicionalmente consideram-se os cumulantes nulos (veja Apêndice A.2), então de (4.55) a (4.57) segue que

$$\begin{aligned}
 a_{1;\beta,\phi} = & -6\alpha_{\phi\phi}^2(\kappa_{\phi\phi\phi} + 2\kappa_{\phi,\phi\phi})\sum_{\beta}\kappa_{\phi,i,j}\gamma_{ij} \\
 & +6\alpha_{\phi\phi}\sum_{\beta}(\kappa_{i,\phi j} - \kappa_{i,\phi,j})(\kappa_{k\phi r} + \kappa_{k\phi,r})\alpha_{jr}\gamma_{ik} \\
 & -6\alpha_{\phi\phi}\sum_{\beta}(\kappa_{i,j\phi} + 2\kappa_{i,j,\phi})\kappa_{\phi,k,r}\alpha_{ij}\gamma_{kr} \\
 & +6\alpha_{\phi\phi}\sum_{\beta}(\kappa_{i,j\phi} - \kappa_{i,j,\phi})(\kappa_{kr\phi} + 2\kappa_{kr,\phi})\alpha_{jr}\gamma_{ik} \\
 & -6\alpha_{\phi\phi}\sum_{\beta}(\kappa_{i,j,\phi\phi} + \kappa_{i,j,\phi\phi})\gamma_{ij} ,
 \end{aligned} \tag{4.68}$$

$$a_{2;\beta,\phi} = -3\alpha_{\phi\phi}\sum_{\beta}\kappa_{i,j,\phi}\kappa_{k,r,\phi}\gamma_{ij}\gamma_{kr} - 6\alpha_{\phi\phi}\sum_{\beta}\kappa_{i,j,\phi}\kappa_{k,r,\phi}\gamma_{ik}\gamma_{jr} , \tag{4.69}$$

$$a_{3;\beta,\phi} = 0 ,$$

onde $\sum_{\beta}$ denota a somatória sobre os índices dos parâmetros em β . Substituindo os cumulantes aqui envolvidos, estas expressões se reduzem a

$$\begin{aligned}
 a_{1;\beta,\phi} &= \frac{72q}{n} \frac{(\nu+1)(\nu+2)(\nu+3)}{\nu(\nu+5)^2(\nu+7)} + \frac{12(k-q)}{n} \frac{(\nu-1)(\nu+2)(\nu+3)}{\nu(\nu+5)^2} , \\
 a_{2;\beta,\phi} &= \frac{6q(q+2)}{n} \frac{(\nu-1)^2(\nu+3)}{\nu(\nu+5)^2} .
 \end{aligned}$$

Assim, os resultados dados para a , b e c em (4.62), (4.64) e (4.65) seguem de substituir $a_i = a_{i;\beta} + a_{i;\beta,\phi}$, $i = 1, 2, 3$, em (4.54).

Exemplo 4.4 Suponha que no Exemplo 4.1 se tenha $\nu = 1$. Então, $d_{\beta,\tau^2} = 2/n$, $c_{\beta,\tau^2} = 1/2n$ e $b_{\beta,\tau^2} = 0$. Logo, para testes sobre o coeficiente β_2 da regressão linear simples com erros Cauchy, tem-se que $d = \{3(\delta_x + 2)/4n\} + 2/n$, $c = \{3(\delta_x + 2)/8n\} + 1/2n$ e $b = b_{\beta} = -\delta_x/8n$. Assim, se $G_{(1)}$ e $S_{(1)}$ denotam as estatísticas G e S do modelo Cauchy, então as respectivas estatísticas modificadas são dadas por

$$G_{(1)}^* = \left\{ 1 + \frac{3(\delta_x + 2)}{4n} + \frac{2}{n} \right\}^{-1} G_{(1)} \quad \text{e} \quad S_{(1)}^* = \left\{ 1 + \frac{3(\delta_x + 2)}{8n} + \frac{1}{2n} - \frac{\delta_x}{8n} S_{(1)} \right\}^{-1} S_{(1)} .$$

Exemplo 4.5 Suponha que no Exemplo 4.2 se tenha que $k = 2$ com $\nu = 1$ (duas populações Cauchy). Então, para o teste de $H_0 : \beta_1 = 0$, tem-se que $d = \{3(\rho-1)/4n\} + 1/2n$,

$c = \{3(3\rho - 5)/8n\} + 1/2n$ e $b = b_\beta = -(\rho - 3)/8n$. Para $n_1 = n_2 = n/2$, $\rho = 4$ e as estatísticas modificadas são dadas por

$$G_{(1)}^* = \left\{1 + \frac{11}{4n}\right\}^{-1} G_{(1)} \quad \text{e} \quad S_{(1)}^* = \left\{1 + \frac{25}{8n} - \frac{1}{8n}S_{(1)}\right\}^{-1} S_{(1)},$$

onde $G_{(1)}$ e $S_{(1)}$ são as correspondentes estatísticas originais.

Exemplo 4.6 Seja $Y_n \sim N_n(X_1\beta_1 + X_2\beta_2, \tau^2 I_n)$, onde τ^2 é desconhecido e considere o teste de $H_0 : \beta_2 = 0$ contra $H_1 : \beta_2 \neq 0$. Como supor normalidade equivale a supor $\nu = \infty$ (ν muito grande) no modelo $t_n^{(I)}(X_1\beta_1 + X_2\beta_2, \tau^2 I_n; \nu)$, então, lembrando que $d_\beta = b_\beta = c_\beta = 0$ para $\nu = \infty$ (veja (4.58) a (4.60)), de (4.63) a (4.65) segue que $d = c = (2k - q + 2)/2n$ e $b = -1/2n$. Assim, para o modelo normal as estatísticas da razão de verossimilhança e escore modificadas, digamos $G_{(\infty)}^*$ e $S_{(\infty)}^*$ são dadas por

$$G_{(\infty)}^* = \left\{1 + \frac{2k - q + 2}{2n}\right\}^{-1} G_{(\infty)} \quad \text{e} \quad S_{(\infty)}^* = \left\{1 + \frac{2k - q + 2}{2n} - \frac{1}{2n}S_{(\infty)}\right\}^{-1} S_{(\infty)},$$

(veja Cordeiro (1987) e Cribari-Neto e Ferrari (1995b), respectivamente), onde $G_{(\infty)}$ e $S_{(\infty)}$ são as estatísticas originais. De (4.23) segue que

$$G_{(\infty)} = n \log \left\{1 + \frac{q}{n - k} F_{(\infty)}\right\} \quad \text{e} \quad S_{(\infty)} = n \left\{\frac{q}{n - k} F_{(\infty)}\right\} \left\{1 - \frac{q}{n - k} F_{(\infty)}\right\}^{-1},$$

onde a estatística $F_{(\infty)}$ é da forma dada em (4.21) (com $A = [0, I_q]$ e $\gamma_0 = 0$), mas considerando o modelo normal, e tem distribuição $F_{q, n-k}$ sob a hipótese nula. Deste modo, a correção não é necessária uma vez que os testes baseados em $G_{(\infty)}$ e $S_{(\infty)}$ podem se reduzir a um teste exato (veja (4.22)). Contudo, isto permite avaliar o comportamento das estatísticas modificadas para ν grande.

4.4.4 Fatores de Bartlett supondo τ^2 e ν desconhecidos

Finalmente estuda-se o caso em que ν também é assumido desconhecido. Para tal efeito define-se novamente o parâmetro $\phi = 1/\tau^2$, de modo que agora $\theta = (\beta^\top, \phi, \nu)^\top$, cuja

dimensão é $p = k + 2$. Como o estudo é restrito a hipóteses sobre β , então $a = 0$ (veja a propriedade (i) na Seção 4.4.2) e para $z = d$, c e b tem-se que

$$z = z_\beta + z_{\beta,\phi,\nu} , \quad (4.70)$$

onde $z_\beta = d_\beta, c_\beta$ e b_β estão definidos em (4.58) a (4.60) e $z_{\beta,\phi,\nu}$ representa a respectiva contribuição adicional que resulta de incorporar os parâmetros ϕ e ν ao modelo. Note-se que $z_{\beta,\phi,\nu} = z_{\beta,\phi} = d_{\beta,\phi}, c_{\beta,\phi}$ e $b_{\beta,\phi}$ (definidos em (4.63) a (4.65)) quando ν é conhecido. Similarmente, $z_{\beta,\phi,\nu} = z_{\beta,\nu}$ quando ϕ é conhecido, onde $z_{\beta,\nu}$ se deriva da mesma forma que $z_{\beta,\phi}$, mas substituindo ϕ por ν em (4.67) a (4.69). Assim, se tanto ϕ quanto ν são desconhecidos, então $z_{\beta,\phi,\nu}$ pode se decompor em

$$z_{\beta,\phi,\nu} = z_{\beta,\phi,\nu}^\phi + z_{\beta,\phi,\nu}^\nu + z_{\beta,\phi,\nu}^{\phi\nu} , \quad (4.71)$$

onde $z_{\beta,\phi,\nu}^x$, $x = \phi, \nu$, se derivam como $z_{\beta,x}$, $x = \phi, \nu$, mas considerando agora que em (4.67) a (4.69) tem-se que $\kappa^{xx} \neq \kappa_{xx}^{-1}$ e $\alpha_{xx} = \kappa^{x,x} \neq -\kappa_{xx}^{-1}$, $x = \phi, \nu$, uma vez que $\kappa_{\phi\nu} \neq 0$ (veja o Apêndice A.2); isto é,

$$\begin{aligned} z_{\beta,\phi,\nu}^x &= \frac{e_{k+1;\beta,x} - e_{k-q+1;\beta,x}}{q} , \quad \text{se } z = d, x = \phi, \nu , \\ &= \frac{a_{1;\beta,x} - a_{2;\beta,x}}{12q} , \quad \text{se } z = c, x = \phi, \nu , \\ &= \frac{a_{2;\beta,x}}{12q(q+2)} , \quad \text{se } z = b, x = \phi, \nu , \end{aligned} \quad (4.72)$$

onde $e_{m;\beta,x}$, $a_{1;\beta,x}$ e $a_{2;\beta,x}$ são como em (4.67), (4.68) e (4.69), respectivamente (com ϕ substituído por ν para $x = \nu$), mas lembrando o fato de que $[\kappa^{ij}] = [\kappa_{ij}]^{-1}$ implica que, para ϕ e ν desconhecidos, κ^{xx} e $\alpha_{xx} = -\kappa^{xx}$, $x = \phi, \nu$, são dados por

$$\begin{pmatrix} \kappa^{\phi\phi} & \kappa^{\phi\nu} \\ \kappa^{\nu\phi} & \kappa^{\nu\nu} \end{pmatrix} = \begin{pmatrix} \kappa_{\phi\phi} & \kappa_{\phi\nu} \\ \kappa_{\nu\phi} & \kappa_{\nu\nu} \end{pmatrix}^{-1} = (\kappa_{\phi\phi}\kappa_{\nu\nu} - \kappa_{\phi\nu}\kappa_{\nu\phi})^{-1} \begin{pmatrix} \kappa_{\nu\nu} & -\kappa_{\phi\nu} \\ -\kappa_{\nu\phi} & \kappa_{\phi\phi} \end{pmatrix} ,$$

de onde é fácil ver que

$$\kappa^{\phi\phi} = \kappa_{\nu\nu}(\kappa_{\phi\phi}\kappa_{\nu\nu}^*)^{-1} , \quad \kappa^{\phi\nu} = \kappa^{\nu\phi} = \frac{2\phi}{\nu(\nu+1)}\kappa_{\nu\nu}^* \quad \text{e} \quad \kappa^{\nu\nu} = (\kappa_{\nu\nu}^*)^{-1} , \quad (4.73)$$

com (veja o Apêndice A.2; veja também (4.39) com $\phi = 1/\tau^2$)

$$\begin{aligned}\kappa_{\phi\phi} &= -\frac{n}{2\phi^2} \frac{\nu}{\nu+3}, \quad \kappa_{\phi\nu} = \kappa_{\nu\phi} = \frac{2\phi}{\nu(\nu+1)} \kappa_{\phi\phi}, \\ \kappa_{\nu\nu} &= \frac{n}{2} \left\{ \psi'[\frac{1}{2}(\nu+1)] - \psi'[\frac{1}{2}\nu] + \frac{\nu+5}{\nu(\nu+1)(\nu+3)} \right\}\end{aligned}\quad (4.74)$$

e (veja (4.45))

$$\kappa_{\nu\nu}^* = \kappa_{\nu\nu} + \frac{2n}{\nu(\nu+1)^2(\nu+3)} = \frac{n}{2} \left\{ \psi'[\frac{1}{2}(\nu+1)] - \psi'[\frac{1}{2}\nu] + \frac{\nu^2+6\nu+1}{\nu(\nu+1)^2(\nu+3)} \right\}. \quad (4.75)$$

(Note-se novamente aqui que $z_{\beta,\phi,\nu}^\phi = z_{\beta,\phi}$ para ν conhecido e $z_{\beta,\phi,\nu} = z_{\beta,\nu}$ para ϕ conhecido, já que, em tais casos, $\kappa^{xx} = \kappa_{xx}^{-1}$ e $\alpha_{xx} = -\kappa_{xx}^{-1}$ para $x = \phi$ e ν , respectivamente).

De forma similar, para o componente $z_{\beta,\phi,\nu}^{\phi\nu}$ de (4.71) tem-se que

$$\begin{aligned}z_{\beta,\phi,\nu}^{\phi,\nu} &= \frac{e_{k+2;\beta,\phi,\nu} - e_{k-q+2;\beta,\phi,\nu}}{q}, \quad \text{se } z = d, x = \phi, \nu, \\ &= \frac{a_{1;\beta,\phi,\nu} - a_{2;\beta,\phi,\nu}}{12q}, \quad \text{se } z = c, x = \phi, \nu, \\ &= \frac{a_{2;\beta,\phi,\nu}}{12q(q+2)}, \quad \text{se } z = b, x = \phi, \nu,\end{aligned}\quad (4.76)$$

onde

$$\begin{aligned}e_{m;\beta,\phi,\nu} &= \frac{1}{2} \kappa^{\nu\phi} \{ 2 \sum_{\beta} \kappa^{ij} (\kappa_{ij\phi\nu} - \kappa_{ij\phi}^{(\nu)} - \kappa_{ij\nu}^{(\phi)}) - \sum_{\beta} \kappa^{ij} \kappa^{st} \kappa_{is\phi} (\kappa_{jt\nu} - 2\kappa_{jt}^{(\nu)}) \\ &\quad - \sum_{\beta} \kappa^{ij} \kappa^{st} \kappa_{is\nu} (\kappa_{jt\phi} - 2\kappa_{jt}^{(\phi)}) - \sum_{\beta} \kappa^{ij} \kappa_{ij\phi} (\sum_{\beta} \kappa^{ij} \kappa_{ij\nu}) \} \\ &\quad - \frac{1}{2} \kappa^{\phi\phi} \kappa^{\nu\nu} \{ (\kappa_{\phi\nu\nu} - 2\kappa_{\phi\nu}^{(\nu)}) \sum_{\beta} \kappa^{ij} \kappa_{ij\phi} + (\kappa_{\nu\phi\phi} - 2\kappa_{\nu\phi}^{(\phi)}) \sum_{\beta} \kappa^{ij} \kappa_{ij\nu} \},\end{aligned}\quad (4.77)$$

$$\begin{aligned}a_{1;\beta,\phi,\nu} &= -12\alpha_{\phi\nu}^2 \{ (\kappa_{\phi\phi\nu} + \kappa_{\phi,\phi\nu} + \kappa_{\nu,\phi\phi}) \sum_{\beta} \kappa_{\nu,s,t} \gamma_{st} + (\kappa_{\phi\nu\nu} + \kappa_{\nu,\phi\nu} + \kappa_{\phi,\nu\nu}) \sum_{\beta} \kappa_{\phi,s,t} \gamma_{st} \} \\ &\quad - 6\alpha_{\phi\phi} \alpha_{\nu\nu} \{ (\kappa_{\phi\phi\nu} + 2\kappa_{\phi,\phi\nu}) \sum_{\beta} \kappa_{r,s,t} \gamma_{st} + (\kappa_{\nu,\phi} + 2\kappa_{\nu,\nu\phi}) \sum_{\beta} \kappa_{\phi,s,t} \gamma_{st} \} \\ &\quad - 6\alpha_{\phi\nu} \{ \sum_{\beta} (\kappa_{ij\phi} + 2\kappa_{i,j\phi}) \kappa_{\nu,s,t} \gamma_{st} + \sum_{\beta} (\kappa_{ij\nu} + 2\kappa_{i,j\nu}) \kappa_{\phi,s,t} \gamma_{st} \} \\ &\quad + 6\alpha_{\phi\nu} \{ \sum_{\beta} (\kappa_{i,\phi k} - \kappa_{i,\phi,k}) (\kappa_{r\nu t} + 2\kappa_{r,s,\nu}) \alpha_{ij} \gamma_{ir} \\ &\quad + \sum_{\beta} (\kappa_{i,j\phi} - \kappa_{i,j,\phi}) (\kappa_{rs\nu} + 2\kappa_{r,s,\nu}) \alpha_{js} \gamma_{ir} \}\end{aligned}$$

$$\begin{aligned}
& + \sum_{\beta} (\kappa_{i,\nu k} - \kappa_{i,\nu,k}) (\kappa_{r\phi t} + 2\kappa_{r\phi,t}) \alpha_{kt} \gamma_{ir} \\
& + \sum_{\beta} (\kappa_{i,j\nu} - \kappa_{i,j,\nu}) (\kappa_{rs\phi} + 2\kappa_{rs,\phi}) \alpha_{ij} \gamma_{ir} \} \\
& - 12\alpha_{\phi\nu} \sum_{\beta} (\kappa_{i,j,\phi\nu} + \kappa_{i,j,\phi\nu}) \gamma_{ij}
\end{aligned} \tag{4.78}$$

e

$$\begin{aligned}
a_{2;\beta,\phi,\nu} &= -3\alpha_{\phi\nu} \{ (\sum_{\beta} \kappa_{i,j,\phi} \gamma_{ij})^2 + (\sum_{\beta} \kappa_{i,j,\nu} \gamma_{ij})^2 \} \\
&+ 6\alpha_{\phi\nu} \{ \sum_{\beta} (\kappa_{i,j,\phi} \kappa_{r,s,\nu} + \kappa_{i,j,\nu} \kappa_{r,s,\phi}) \gamma_{ir} \gamma_{js} \} ,
\end{aligned} \tag{4.79}$$

com $\alpha_{\phi\phi} = -\kappa^{\phi\phi}$, $\alpha_{\nu\nu} = -\kappa^{\nu\nu}$, $\alpha_{\phi\nu} = -\kappa^{\phi\nu}$ e com $\kappa^{\phi\phi}$, $\kappa^{\nu\nu}$ e $\kappa^{\phi\nu}$ sendo como em (4.73).

Para ilustrar deriva-se no que se segue o componente $d_{\beta,\phi,\nu}$ do coeficiente d relativo à estatística da razão de verossimilhança modificada. Os cumulantes necessários são dados no Apêndice A.2. Assim, de (4.67) juntamente com a primeira expressão de (4.72) segue (após uma certa álgebra) que

$$\begin{aligned}
d_{\beta,\phi,\nu}^{\phi} &= -\frac{1}{n} \kappa^{\phi\phi} \kappa_{\phi\phi} \left\{ \frac{(\nu+2)(\nu+3)}{\nu(\nu+5)} \right\} \left\{ \frac{\nu^2+9\nu+26}{(\nu+5)(\nu+7)} \right. \\
&\quad \left. - 2\kappa^{\phi\phi} \kappa_{\phi\phi} \left(\frac{\nu+2}{\nu+5} \right) - \frac{2k-q}{2} \left(\frac{\nu+2}{\nu+5} \right) \right\}
\end{aligned}$$

e

$$\begin{aligned}
d_{\beta,\phi,\nu}^{\nu} &= 2\kappa^{\nu\nu} \left\{ \frac{\nu^5+13\nu^4+37\nu^3-19\nu^2-206\nu-66}{\nu^2(\nu+1)^2(\nu+3)(\nu+5)^2(\nu+7)} \right\} \\
&+ n \{ \kappa^{\nu\nu} \}^2 \left\{ \frac{\nu-1}{\nu(\nu+1)(\nu+5)} \right\} \left\{ \frac{1}{4} \psi'' \left[\frac{1}{2}(\nu+1) \right] \right. \\
&\quad \left. - \frac{1}{4} \psi'' \left(\frac{1}{2}\nu \right) + \frac{2\nu^5+20\nu^4+53\nu^3-27\nu^2-183\nu-57}{\nu^2(\nu+1)^2(\nu+3)^2(\nu+5)} \right\} \\
&- (2k-q) \kappa^{\nu\nu} \left\{ \frac{\nu-1}{\nu(\nu+1)(\nu+5)} \right\}^2 ,
\end{aligned}$$

onde ψ'' denota a segunda derivada da função digama ψ , e de acordo com (4.73), $\kappa^{\phi\phi} \kappa_{\phi\phi} = \kappa_{\nu\nu} (\kappa_{\nu\nu}^*)^{-1}$ e $\kappa^{\nu\nu} = (\kappa_{\nu\nu}^*)^{-1}$, onde $\kappa_{\nu\nu}$ e $\kappa_{\nu\nu}^*$ são dados em (4.74) e (4.75), respectivamente. Daqui é fácil ver que $d_{\beta,\phi,\nu}^{\phi} = d_{\beta,\phi}$ para ν conhecido, uma vez que em tal caso tem-se que $\kappa^{\phi\phi} \kappa_{\phi\phi} = 1$. Similarmente, $\kappa^{\nu\nu} = \kappa_{\nu\nu}^{-1}$ e $d_{\beta,\phi,\nu}^{\nu} = d_{\beta,\nu}$ para ϕ conhecido.

Para obter $d_{\beta,\phi,\nu}^{\phi\nu}$ deve-se usar (4.77) juntamente com a primeira expressão de (4.76).

Após um certo trabalho algébrico e aplicando (4.73) chega-se a

$$d_{\beta,\phi,\nu}^{\phi\nu} = -\frac{2}{\kappa_{\nu\nu}^*} \left\{ \frac{\nu^4 + 56\nu^3 + 254\nu^2 + 160\nu - 279}{\nu^2(\nu+1)^3(\nu+3)^3(\nu+5)^2(\nu+7)} - (2k-q) \frac{(\nu-1)(\nu+2)}{\nu^2(\nu+1)^3(\nu+3)(\nu+5)^2} \right\} \\ + \frac{\kappa_{\nu\nu}}{(\kappa_{\nu\nu}^*)^2} \left\{ \frac{2\nu^4 + 7\nu^3 + 13\nu^2 + 32\nu + 21}{\nu^2(\nu+1)^2(\nu+3)(\nu+5)^2} \right\},$$

onde $\kappa_{\nu\nu}$ e $\kappa_{\nu\nu}^*$ estão definidos em (4.74) e (4.75), respectivamente. Assim colocando $z = d$ em (4.70), tem-se que $d = d_\beta + d_{\beta,\phi,\nu}$, onde d_β está definido em (4.58) e, de acordo com (4.71),

$$d_{\beta,\phi,\nu} = d_{\beta,\phi,\nu}^\phi + d_{\beta,\phi,\nu}^\nu + d_{\beta,\phi,\nu}^{\phi\nu}.$$

De forma similar pode-se obter os componentes $c_{\beta,\phi,\nu}$ e $b_{\beta,\phi,\nu}$ dos coeficientes c e b , respectivamente, para corrigir a estatística escore no caso em que tanto ϕ como ν são parâmetros desconhecidos. A derivação adicional dos cumulantes “com vírgula” que envolvem o parâmetro ν é requerida para tais efeitos. Este processo pode ser simplificado considerando a parametrização ortogonal de $(\phi, \nu)^\top$, digamos $(\lambda, \nu)^\top$, dada por (veja a Seção 4.3.3 com $\phi = 1/\tau^2$)

$$\phi = \lambda \left(\frac{\nu+1}{\nu} \right)^2, \quad (\lambda > 0). \quad (4.80)$$

Para esta parametrização a matriz de informação, digamos, $K^* = [-\kappa_{ij}^*]$, é bloco-diagonal em β , λ e ν , isto é, $\kappa_{i\lambda}^* = \kappa_{i\nu}^* = \kappa_{\lambda\nu}^* = 0$, $i = 1, 2, \dots, k$. Além disso, aqui também se tem que $a = 0$ para hipóteses sobre β e que os coeficientes $z = d$, c e b podem se decompor em

$$z = z_\beta + z_{\beta,\lambda,\nu}^\lambda + z_{\beta,\lambda,\nu}^\nu + z_{\beta,\lambda,\nu}^{\lambda\nu}.$$

A diferença em relação à parametrização original é que aqui $z_{\beta,\lambda,\nu}^x = z_{\beta,x}$, uma vez que $\kappa_{\lambda\nu}^* = 0$ e, portanto, $\kappa^{*xx} = (\kappa_{xx}^*)^{-1}$, $x = \lambda, \nu$, e $z_{\beta,\lambda,\nu}^{\lambda\nu}$ é mais simples de avaliar já que em (4.77) a (4.79) (substituídos os cumulantes κ 's da parametrização original $(\beta^\top, \phi, \nu)^2$ pelos respectivos cumulantes κ^* 's da parametrização ortogonal $(\beta^\top, \lambda, \nu)^\top$) tem-se que

$\kappa^{*\lambda\nu} = -\alpha_{\lambda\nu}^* = 0$. Em particular, em (4.79) tem-se que $a_{2;\beta,\lambda,\nu} = 0$, o que implica que $b_{\beta,\lambda,\nu} = 0$ e $c_{\beta,\lambda,\nu} = (12q)^{-1}a_{1;\beta,\lambda,\nu}$. Observe-se que z_β é o mesmo que na parametrização original e $z_{\beta,\lambda} = z_{\beta,\phi}$, uma vez que estes coeficientes independem dos parâmetros de locação de e de escala do modelo. Finalmente, note-se que os cumulantes κ^* 's podem ser obtidos a partir dos cumulantes κ 's, usando que $L^*(\beta, \lambda, \nu) = L(\beta, \phi(\lambda, \nu), \nu)$. Assim, para fixar esta idéia, considere-se a notação dada no começo da Seção 4.4.1 para as derivadas da log-verossimilhança com respeito aos parâmetros, e denote-se por

$$\phi_x = \frac{\partial \phi}{\partial x}, \quad \phi_{xy} = \frac{\partial^2 \phi}{\partial x \partial y}, \quad \text{etc.},$$

as derivadas de $\phi = \phi(\lambda, \nu)$ (dado em (4.80)) com respeito a $x, y = \lambda, \nu$. Com esta notação tem-se, por exemplo, que

- (1) $L_\lambda^* = \phi_\lambda L_\phi$, $L_{\lambda\lambda}^* = \phi_{\lambda\lambda} L_\phi + \phi_\lambda^2 L_{\phi\phi}$, $L_{\lambda\lambda\lambda}^* = \phi_{\lambda\lambda\lambda} L_\phi + 3\phi_\lambda \phi_{\lambda\lambda} L_{\phi\phi} + \phi_\lambda^3 L_{\phi\phi\phi}$, etc.
- (2) $L_\nu^* = \phi_\nu L_\phi + L_\nu$, $L_{\nu\nu}^* = \phi_{\nu\nu} L_\phi + \phi_\nu^2 L_{\phi\phi} + 2\phi_\nu L_{\phi\nu} + L_{\nu\nu}$, $L_{\nu\nu\nu}^* = \phi_{\nu\nu\nu} L_\phi + 3\phi_\nu \phi_{\nu\nu} L_{\phi\phi} + 2\phi_{\nu\nu} L_{\phi\nu} + \phi_\nu^3 L_{\phi\phi\phi} + 3\phi_\nu^2 L_{\phi\phi\nu} + 3\phi_\nu L_{\phi\nu\nu} + L_{\nu\nu\nu}$, etc.
- (3) $L_{\lambda j}^* = \phi_\lambda L_{\phi j}$, $L_{\lambda jr}^* = \phi_\lambda L_{\phi jr}$, $L_{\nu j}^* = \phi_\nu L_{\phi j} + L_{\nu j}$, $L_{\nu jr} = \phi_\nu L_{\phi jr} + L_{\nu jr}$, $j, r = 1, 2, \dots, k$; etc.

Destas relações, juntamente ao fato de que $\phi_{\lambda\lambda} = \phi_{\lambda\lambda\lambda} = \dots = 0$, tem-se, por exemplo, que

- (1)' $\kappa_{\lambda\lambda}^* = \phi_\lambda^2 \kappa_{\phi\phi}$, $\kappa_{\lambda\lambda\lambda}^* = \phi_\lambda^3 \kappa_{\phi\phi\phi}$, etc.
- (2)' $\kappa_{\nu\nu}^* = \phi_\nu^2 \kappa_{\phi\phi} + 2\phi_\nu \kappa_{\phi\nu} + \kappa_{\nu\nu} = \kappa_{\nu\nu} - \phi_\nu^2 \kappa_{\phi\phi}$, uma vez que $\kappa_{\lambda\nu}^* = 0$ implica que $\phi_\nu \kappa_{\phi\phi} = -\kappa_{\phi\nu}$, etc.,
- (3)' $\kappa_{\lambda j}^* = \phi_\lambda \kappa_{\phi j} = 0$, $\kappa_{\lambda jr}^* = \phi_\lambda L_{\phi jr}$, $\kappa_{\nu j}^* = 0$, $\kappa_{\nu jr}^* = \phi_\nu \kappa_{\phi jr} + \kappa_{\nu jr}$, $j, r = 1, 2, \dots, k$, etc.,

onde os cumulantes κ 's estão avaliados em $\phi = \lambda(\nu + 1)^2/\nu^2$.

4.4.5 Alguns estudos de simulação

Para ilustrar o desempenho das estatísticas G e S e suas versões modificadas, nesta seção são apresentados alguns resultados numéricos obtidos via simulação dos tamanhos e os poderes destes testes. Estes resultados são obtidos considerando 10.000 réplicas simuladas de uma amostra de tamanho n e são dados para diferentes valores de n e distintos níveis nominais α (ou níveis de referências), digamos, $n = 10, 20, 30$ e 40 e $\alpha = 10\%, 5\%, 2, 5\%$ e 1% . Em cada caso o vetor de respostas $Y(n \times 1)$ é gerado supondo que $Y \sim t_n^{(I)}(\eta, 4I_n; \nu)$ com ν variando em $\{1, 3\}$ e utilizando somente alguns casos simples na especificação de $\eta = X\beta$. Na primeira destas especificações se considera o problema descrito no Exemplo 4.5, onde se comparam as locações de duas populações independentes supondo amostras de igual tamanho. Assim, com a notação do Exemplo 4.5, para este experimento as respostas são geradas fixando $\beta = 1$. O poder dos testes não é alterado. Os resultados destas simulações são dados nas Tabelas 4.1 e 4.2.

Das Tabelas 4.1 e 4.2 se observa que o teste da razão de verossimilhança rejeita a hipótese nula ($H_0 : \beta_1 = 0$) mais frequentemente do que o esperado com o tamanho nominal. Este resultado é particularmente claro quando $n = 10$ e $\nu = 1$ (Tabela 4.1). No entanto, os tamanhos relativos ao mesmo teste, mas corrigido pelo fator de Bartlett, estão efetivamente mais próximos dos respectivos níveis nominais em qualquer um dos tamanhos da amostra aqui considerados. Este fornece uma forte evidência sobre a efetividade do fator de Bartlett, a qual é obviamente mais importante nas amostras de tamanhos menores. Os resultados dados nestas tabelas sugerem também que, sob H_0 , a estatística escore S tem distribuição mais próxima da distribuição qui-quadrado de referência do que a estatística da razão de verossimilhança G , uma vez que o teste escore tem níveis simulados relativamente mais próximos dos respectivos níveis de referência que o teste da razão de verossimilhança, qualquer que seja o tamanho da amostra. De forma consistente com esse fato, se observa nestas tabelas que a correção da estatística S não é tão importante quanto a devida correção da estatística G . De fato, os fatores de correção podem ser

Tabela 4.1. Tamanhos do teste de homogeneidade das locações de duas populações independentes com $\nu = 1$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	21.1	9.3	12.6	10.6
	5.0	13.1	4.2	6.5	5.4
	2.5	8.0	1.9	3.4	2.8
	1.0	4.2	0.8	1.3	1.1
20	10.0	13.9	9.5	10.8	9.6
	5.0	7.7	4.7	5.3	4.8
	2.5	4.4	2.4	2.7	2.4
	1.0	2.1	0.9	1.1	1.0
30	10.0	13.2	10.3	11.0	10.5
	5.0	7.1	5.2	5.5	5.2
	2.5	4.0	2.6	2.8	2.7
	1.0	1.7	1.0	1.2	1.1
40	10.0	11.6	9.8	10.1	9.7
	5.0	6.2	4.6	5.1	4.7
	2.5	3.0	2.3	2.3	2.3
	1.0	1.3	0.9	0.9	0.8

Tabela 4.2. Tamanhos do teste de homogeneidade das locações de duas populações independentes com $\nu = 3$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	15.3	9.5	12.3	9.7
	5.0	8.6	4.4	6.1	4.7
	2.5	4.6	2.2	2.9	2.3
	1.0	2.4	0.9	1.2	1.1
20	10.0	12.4	10.0	11.2	10.1
	5.0	6.8	4.9	5.6	4.9
	2.5	3.6	2.4	2.5	2.4
	1.0	1.7	1.1	1.1	1.0
30	10.0	11.5	9.7	10.5	9.8
	5.0	6.0	5.3	5.5	5.1
	2.5	3.4	2.6	2.9	2.6
	1.0	1.5	1.1	1.4	1.1
40	10.0	12.0	10.7	11.4	10.8
	5.0	6.1	5.2	5.7	5.4
	2.5	3.4	2.9	3.1	2.9
	1.0	1.6	1.2	1.2	1.1

utilizados como indicadores da proximidade entre as distribuições da estatística do teste e de referência (sob a hipótese nula), de modo que valores pequenos nestes fatores são evidência de que estas distribuições estão próximas.

Para o segundo experimento considera-se o modelo de regressão linear simples $\eta_i = \beta_1 + \beta_2 x_i$, $i = 1, 2, \dots, n$, definido no Exemplo 4.4. Sob este modelo se estudará o desempenho dos testes de $H_0 : \beta_2 = 0$ contra $H_1 : \beta_2 \neq 0$. As respostas $Y_1, Y_2, \dots, Y_n$ são geradas fixando $\beta_1 = 1$. Do Exemplo 4.4, segue que nesse caso os fatores de correção dependem da curtose $\delta_x - 3$ dos valores $x_1, x_2, \dots, x_n$, de modo que os testes modificados são afetados por esta curtose. Para controlar este efeito as quantidades $x_1, x_2, \dots, x_n$ são escolhidas de acordo com o procedimento estabelecido por Cribari-Neto e Ferrari (1995). Dada uma certa distribuição da covariável x e começando com $n = 10$, estes autores fixam $x_1, x_2, \dots, x_{10}$, de modo que

$$x_i = \text{quantil} \left\{ \frac{i - 0.5}{10} \right\}, \quad i = 1, 2, \dots, 10.$$

Assim, para $n = 20, 30$ e 40 os valores restantes são obtidos replicando $x_1, x_2, \dots, x_{10}$. Por exemplo, para $n = 20$, se escolhem os valores $x_1, x_1, x_2, x_2, \dots, x_{10}, x_{10}$. A conveniência deste procedimento é que o valor da curtose (ou δ_x) não varia com n . Para gerar $x_1, x_2, \dots, x_n$, aqui se usam as distribuições uniforme em $(0, 1)$ ($U(0, 1)$) e log-normal padrão ($LN(0, 1)$), cujas respectivas curtoses são -1.22 ($\delta_x = 1.78$) e 1.42 ($\delta_x = 4.42$). Os resultados em 10.000 réplicas de cada tamanho amostral são resumidos nas Tabelas 4.3 a 4.6.

Tabela 4.3. Tamanhos quando a covariável é gerada com a distribuição $U(0,1)$ e $\nu = 1$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	21.5	7.5	11.5	9.7
	5.0	13.5	3.3	5.8	4.8
	2.5	8.2	1.3	2.4	2.5
	1.0	4.5	0.3	1.0	0.9
20	10.0	15.0	10.1	11.2	10.2
	5.0	8.9	4.9	5.7	5.3
	2.5	5.0	2.5	2.7	2.5
	1.0	2.4	1.0	0.9	0.9
30	10.0	13.1	10.0	10.3	10.1
	5.0	7.1	4.8	5.3	5.0
	2.5	3.7	2.2	2.5	2.4
	1.0	1.6	1.0	1.0	1.0
40	10.0	13.0	10.4	11.0	10.3
	5.0	6.7	5.0	5.6	5.4
	2.5	3.7	2.6	2.8	2.7
	1.0	1.8	1.3	1.4	1.2

Tabela 4.4. Tamanhos quando a covariável é gerada com a distribuição $U(0,1)$ e $\nu = 3$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	15.7	9.1	12.2	9.7
	5.0	8.8	4.4	5.8	4.6
	2.5	5.0	2.0	2.6	2.1
	1.0	2.4	0.9	0.9	0.8
20	10.0	12.3	9.7	10.9	9.8
	5.0	6.7	5.0	5.7	4.8
	2.5	3.9	2.5	3.1	2.6
	1.0	1.7	0.9	1.3	1.1
30	10.0	11.9	10.2	10.9	10.1
	5.0	6.4	5.1	5.5	5.0
	2.5	3.6	2.6	2.6	2.4
	1.0	1.5	1.3	1.3	1.2
40	10.0	11.2	9.8	10.5	9.9
	5.0	6.0	5.2	5.3	5.0
	2.5	3.3	2.5	2.8	2.6
	1.0	1.4	1.1	1.2	1.2

Tabela 4.5. Tamanhos quando a covariável é gerada com a distribuição $LN(0, 1)$ e $\nu = 1$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	26.9	3.0	11.2	8.8
	5.0	17.3	0.7	4.3	3.8
	2.5	11.4	0.2	1.2	1.2
	1.0	6.0	0.1	0.1	0.2
20	10.0	16.7	8.6	11.3	10.2
	5.0	9.8	3.4	5.3	4.9
	2.5	5.7	1.6	2.1	2.1
	1.0	2.6	0.6	0.6	0.8
30	10.0	14.6	9.4	10.3	9.7
	5.0	7.9	4.4	5.0	4.7
	2.5	4.2	2.2	2.2	2.2
	1.0	2.0	0.6	0.8	0.9
40	10.0	13.2	9.9	10.3	9.7
	5.0	7.1	4.9	4.9	4.8
	2.5	4.0	2.2	2.4	2.4
	1.0	1.7	0.8	0.7	0.8

Tabela 4.6. Tamanhos quando a covariável é gerada com a distribuição $LN(0, 1)$ e $\nu = 3$ (porcentagem de rejeição em 10.000 réplicas).

n	Tamanho nominal	G	Tamanho do teste		
			G^*	S	S^*
10	10.0	18.7	9.2	11.6	8.6
	5.0	11.0	4.8	4.2	3.5
	2.5	6.7	2.2	1.1	1.1
	1.0	3.7	0.8	0.1	0.2
20	10.0	12.8	9.5	11.3	10.0
	5.0	7.2	4.8	5.3	4.7
	2.5	4.1	2.4	2.2	2.2
	1.0	1.9	1.0	0.7	0.8
30	10.0	12.8	10.4	11.4	10.6
	5.0	6.9	5.2	5.5	5.3
	2.5	3.6	2.5	2.5	2.6
	1.0	1.5	0.9	0.9	1.1
40	10.0	11.4	9.7	10.4	9.9
	5.0	6.3	5.2	5.2	5.0
	2.5	3.4	2.5	2.6	2.5
	1.0	1.5	1.0	0.9	1.0

Das Tabelas 4.3 a 4.6 se observa que, excluindo o caso com $n = 10$ e $\nu = 1$, os resultados deste último experimento são similares aos do primeiro experimento relativo aos tamanhos dos diferentes testes (convém notar aqui que o primeiro experimento pode se parametrizar como o modelo de regressão $\eta_i = \beta_1 + \beta_2 x_i$, $i = 1, 2, \dots, n$, colocando $x_i = 1$ para $1 \leq i \leq n/2$ e $x_i = 0$ para $n/2 + 1 \leq i \leq n$, de modo que $x_1, x_2, \dots, x_n$ têm curtose $\delta_x - 3 = -2$ ou $\delta_x = 1$). No caso em que $n = 10$ e $\nu = 1$, o fator de Bartlett tende a diminuir demasiadamente o tamanho do teste baseado na estatística G . Este fato é notadamente mais importante quando se usa a distribuição $LN(0, 1)$ para gerar os valores amostrais de x , e poder-se-ia esperar, já que, neste caso, o valor do fator de Bartlett (0.318) indica que o ajuste da estatística deve ser forte. Contudo, Cox (1988, p.330) nota que “a large adjustment (to the likelihood ratio statistic) would have to be interpreted partly as a warning, in particular to investigate the distribution more carefully, perhaps by simulation”.

Para finalizar esta seção, convém mencionar que neste estudo de simulação se aplica o algoritmo EM (descrito na Seção 4.3.2) para calcular as estimativas de máxima verossimilhança de β e τ^2 (sob H_0 e H_1) assumindo ν conhecido. Na maioria dos casos, o algoritmo converge em poucas iterações, embora em algumas amostras com $\nu = 1$ e $n = 10$ foi necessário entre 50 a 100 iterações para obter convergência do algoritmo.

Capítulo 5

Inferência no modelo de regressão linear simples com erros nas variáveis usando a distribuição t

5.1 Introdução

Neste capítulo estuda-se a regressão linear simples

$$Y_i = \alpha + \beta x_i + e_i, \quad i = 1, 2, \dots, n, \quad (5.1)$$

supondo que as quantidades $x_1, x_2, \dots, x_n$ estão medidas com erro. Isto é, assumindo que em lugar do verdadeiro valor x_i da variável “independente” se observa um valor X_i tal que

$$X_i = x_i + u_i, \quad i = 1, 2, \dots, n, \quad (5.2)$$

onde u_i representa o correspondente erro na medição de x_i . Como é habitual, aqui o principal objetivo é realizar inferências sobre os parâmetros estruturais α e β a partir das observações (Y_i, X_i) , $i = 1, 2, \dots, n$. Com este propósito, os seguintes postulados sobre os erros aleatórios e_i e u_i são comumente aceitos:

- (i) $E(e_i) = 0$, $\text{Var}(e_i) = \sigma_{ee}$ e $\text{Cov}(e_i, e_j) = 0, i \neq j$;
- (ii) $E(u_i) = 0$, $\text{Var}(u_i) = \sigma_{uu}$ e $\text{Cov}(u_i, u_j) = 0, i \neq j$;

(iii) $\text{Cov}(e_i, u_i) = 0$;

(iv) $\text{Cov}(x_i, e_i) = \text{Cov}(x_i, u_i) = 0$.

Com tais hipóteses, o modelo definido pelas equações (5.1) e (5.2) é denominado modelo de regressão linear simples com erros nas variáveis. Note-se que a condição (iv) é irrelevante se as quantidades não observáveis $x_1, x_2, \dots, x_n$ são assumidas não aleatórias. A consideração ou não desta suposição envolve duas formas de tratar e denominar este modelo (cujas diferenças são mais conceituais que operacionais):

- (a) **O Modelo Estrutural:** Suponha que $x_1, x_2, \dots, x_n$ são variáveis aleatórias não correlacionadas ou independentes. Mesmo que a distribuição de x_i possa ser variável com i , é habitual assumir que $x_1, x_2, \dots, x_n$ têm distribuição comum, com média μ_x e variância σ_{xx} , ambas desconhecidas;
- (b) **O Modelo Funcional:** Suponha que $x_1, x_2, \dots, x_n$ são valores fixos (não aleatórios), de modo que são tratados como parâmetros desconhecidos (denominados parâmetros incidentais) que devem ser estimados. Neste caso, os parâmetros μ_x e σ_{xx} podem ser interpretados como os limites quando $n \rightarrow \infty$ das seqüências $\bar{x}_n = n^{-1} \sum_{i=1}^n x_i$ e $S_{x,n}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$, respectivamente. De qualquer forma, supondo que $x_1, x_2, \dots, x_n$ são variáveis aleatórias degeneradas (ou por condicionamento), este modelo pode ser considerado como um caso particular de (a).

Ambos os enfoques apresentam certos inconvenientes que dificultam a estimação consistente dos parâmetros de interesse (ou estruturais) do modelo. Assim, em (a) tem-se que o modelo pode ser não identificável, no sentido de que valores diferentes do vetor de parâmetros podem dar origem à mesma distribuição das observações. Isto acontece, por exemplo, com o modelo normal (veja Fuller (1987, pp.9-10)), a menos de uma suposição adicional (além das dadas em (i) a (iv)) que permita reduzir o número de parâmetros

envolvidos no modelo. Considerando a normalidade, tal suposição é habitualmente feita sobre as variâncias de x_i , e_i e u_i (veja a Seção 5.1.2). Em (b), tem-se que o número de parâmetros incidentais (aqui, $x_1, x_2, \dots, x_n$) cresce com n (o tamanho da amostra), o que em geral impede obter estimações consistentes dos parâmetros de real interesse (Neyman e Scott (1948)). Além disso, para o modelo funcional normal também deve ser feita uma suposição adicional sobre o vetor de parâmetros para que exista (e seja admissível) uma solução de máxima verossimilhança (veja Solari (1969) e Zellner (1971)).

Seria demasiadamente extenso analisar os dois enfoques mencionados acima do modelo dado em (5.1) e (5.2), principalmente nesta seção que pretende dar uma idéia geral sobre o problema. Assim, no que segue desta seção, estuda-se este modelo usando essencialmente seu enfoque estrutural. Aspectos adicionais aos aqui mencionados para o enfoque funcional e exemplos de situações práticas em que aplicam-se estes modelos podem ser encontrados, por exemplo, em Kendall e Stuart (1979, cap. 29), Fuller (1987) e Bolfarine, Rodriguez e Cordani (1992).

5.1.1 Estimação de mínimos quadrados ordinários e generalizados de β

Voltando ao modelo de regressão definido em (5.1) e (5.2), no que segue se discutem algumas conseqüências imediatas do fato estabelecido em (5.2) sobre a estimação dos parâmetros estruturais α e β , em particular sobre β . Assim, substituindo (5.2) em (5.1), obtém-se a regressão linear em termos de ambas as variáveis observadas, isto é,

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (5.3)$$

onde $\epsilon_i = e_i - \beta u_i$. É fácil ver que se (i) e (iv) são satisfeitas, então os erros $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são tais que $E(\epsilon_i) = 0$, $\text{Var}(\epsilon_i) = \sigma_{ee} + \beta^2 \sigma_{uu}$ e $\text{Cov}(\epsilon_i, \epsilon_j) = 0$, $i \neq j$. Contudo, o problema aqui é que a variável observada X_i , além de ser aleatória, está correlacionada com ϵ_i , uma vez que $\text{Cov}(X_i, \epsilon_i) = \text{Cov}(u_i, \epsilon_i) = -\beta \sigma_{uu} \neq 0$. A implicação deste fato é

que $(\hat{\alpha}_{MQ}, \hat{\beta}_{MQ})^T$ é um estimador inconsistente de $(\alpha, \beta)^T$. Em particular, para

$$\hat{\beta}_{MQ} = \frac{S_{XY}}{S_{XX}}, \quad (5.4)$$

tem-se que

$$\hat{\beta}_{MQ} = \beta + \frac{S_{x\epsilon} + S_{u\epsilon}}{S_{xx} + 2S_{xu} + S_{uu}},$$

onde, para uma seqüência de valores bivariados (a_i, b_i) , $i = 1, 2, \dots, n$, tem-se que $S_{aa} = (1/n) \sum_{i=1}^n (a_i - \bar{a})^2$, $S_{ab} = (1/n) \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})$, com $\bar{a} = (1/n) \sum_{i=1}^n a_i$ e $\bar{b} = (1/n) \sum_{i=1}^n b_i$. Daqui pode-se mostrar que (usando as condições (i) a (iv))

$$\text{plim} \hat{\beta}_{MQ} = \beta - \beta \frac{\sigma_{uu}}{\sigma_{xx} + \sigma_{uu}} = \beta \frac{\sigma_{xx}}{\sigma_{xx} + \sigma_{uu}},$$

de onde segue que

$$\text{plim} \left\{ \left[\frac{\sigma_{xx}}{\sigma_{xx} + \sigma_{uu}} \right]^{-1} \hat{\beta}_{MQ} \right\} = \beta.$$

Em certas situações (particularmente em economia, psicologia, sociologia, biologia e amostragem; veja, por exemplo, Fuller (1987)) é possível obter informação que permita conhecer o fator de atenuação k_x definido como

$$k_x = \frac{\sigma_{xx}}{\sigma_{xx} + \sigma_{uu}} = \frac{\lambda_x}{\lambda_x + 1}, \quad (5.5)$$

onde $\lambda_x = \sigma_{xx}/\sigma_{uu}$. Neste caso, do último resultado dado acima, segue que um estimador consistente de β é dado por

$$\hat{\hat{\beta}}_{MQ} = k_x^{-1} \hat{\beta}_{MQ}. \quad (5.6)$$

Além disso, o fato de supor k_x conhecido permite também identificar o modelo estrutural quando se assume normalidade na distribuição conjunta dos vetores $(x_i, e_i, u_i)^T$, $i = 1, 2, \dots, n$. Neste caso, $\hat{\hat{\beta}}_{MQ}$ coincide com o EMV de β . Isto pode ser demonstrado usando o método de momentos, uma vez que, sob normalidade, os EMV em geral coincidem com os respectivos estimadores de momentos.

Uma outra forma de obter um estimador consistente para β a partir de (5.3) é desenvolvida em Sprent (1966) (para o modelo funcional) supondo que a razão de variâncias

$$\lambda_e = \frac{\sigma_{ee}}{\sigma_{uu}}, \quad (5.7)$$

é conhecida (note-se que, sem perda de generalidade, pode-se assumir $\lambda_e = 1$, uma vez que para $\lambda_e \neq 1$ as observações $Y_1, Y_2, \dots, Y_n$ podem ser redefinidas dividindo-se por $\sqrt{\lambda_e}$). Usando (5.7), em (5.3) tem-se que (sob a condição (iii))

$$\begin{aligned} \sigma_{\epsilon\epsilon} = \text{Var}(\epsilon_i) &= \sigma_{ee} + \beta^2 \sigma_{uu} \\ &= \sigma_{uu}(\lambda_e + \beta^2) \end{aligned}$$

é uma função de β . Considerando este resultado, um estimador consistente de $(\alpha, \beta)^\top$, digamos $(\hat{\alpha}_S, \hat{\beta}_S)^\top$, segue de minimizar

$$Q(\alpha, \beta) = \sum_{i=1}^n \frac{\epsilon_i^2}{\sigma_{\epsilon\epsilon}} = \sum_{i=1}^n \frac{(Y_i - \alpha - \beta X_i)^2}{\sigma_{uu}(\lambda_e + \beta^2)},$$

de onde tem-se que (veja Sprent (1966))

$$\hat{\beta}_S = \frac{S_{YY} - \lambda_e S_{XX} + \{(S_{YY} - \lambda_e S_{XX})^2 + 4\lambda_e S_{YX}^2\}^{\frac{1}{2}}}{2S_{YX}} \quad (5.8)$$

Denotando por $x = (x_1, x_2, \dots, x_n)^\top$, $\epsilon_i = (e_i, u_i)^\top$ e $\Sigma_\epsilon = \text{Cov}(\epsilon_i) = \text{Diag}\{\sigma_{ee}, \sigma_{uu}\}$ (veja (5.9), abaixo), é fácil ver que no modelo funcional, o procedimento anterior equivale a derivar o estimador de mínimos quadrados generalizados de $(\alpha, \beta, x^\top)^\top$, minimizando

$$Q_G(\alpha, \beta, x) = \sum_{i=1}^n \epsilon_i^\top \Sigma_\epsilon^{-1} \epsilon_i = \sum_{i=1}^n \frac{(Y_i - \alpha - \beta x_i)^2}{\lambda_e \sigma_{uu}} + \sum_{i=1}^n \frac{(X_i - x_i)^2}{\sigma_{uu}}.$$

Assim, para o modelo funcional normal com λ_e conhecido o resultado é análogo. Similarmente, o modelo estrutural normal fica identificado para λ_e sendo conhecido e, em tal caso, $\hat{\beta}_S$ também coincide com o EMV de β .

Os resultados anteriores são formalizados na próxima seção para o modelo normal estrutural. Como o parâmetro de maior interesse é β , as inferências serão (em geral)

restritas a este parâmetro. Tais inferências podem ser simplificadas via reparametrizações ortogonais de Cox e Reid (1987), tanto no modelo funcional quanto no modelo estrutural. Alguns fatos deste tipo serão também apresentados nas próximas seções.

Para incorporar a hipótese de normalidade, a seguinte notação será usada. Sejam

$$Z_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \quad g_i = g(x_i) = \begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix} \quad \text{e} \quad \varepsilon_i = \begin{pmatrix} e_i \\ u_i \end{pmatrix},$$

$i = 1, 2, \dots, n$. Com esta notação, o modelo definido em (5.1) e (5.2) pode ser expresso como

$$Z_i = g_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (5.9)$$

e as hipóteses (i) a (iv) como:

$$(i)' \quad E(\varepsilon_i) = 0;$$

$$(ii)' \quad \text{Cov}(\varepsilon_i) = \text{Diag} \{ \sigma_{ee}, \sigma_{uu} \};$$

$$(iii)' \quad \text{Cov}(g_i, \varepsilon_i) = 0.$$

Assim, para o vetor de médias e a matriz de covariâncias de Z_i , tem-se que:

(a) no modelo estrutural,

$$E(Z_i) = E(g_i) + E(\varepsilon_i) = \begin{pmatrix} \alpha + \beta \mu_x \\ \mu_x \end{pmatrix} \quad (5.10)$$

e

$$\begin{aligned} \text{Cov}(Z_i) &= \text{Cov}(g_i) + \text{Cov}(\varepsilon_i) \\ &= \sigma_{xx} \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix} + \begin{pmatrix} \sigma_{ee} & 0 \\ 0 & \sigma_{uu}^2 \end{pmatrix} \\ &= \begin{pmatrix} \beta^2 \sigma_{xx} + \sigma_{ee} & \beta \sigma_{xx} \\ \beta \sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{pmatrix}; \end{aligned} \quad (5.11)$$

(b) no modelo funcional,

$$E(Z_i) = \mu_i = \begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}$$

e

$$\text{Cov}(Z_i) = \text{Cov}(\varepsilon_i) = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \sigma_{uu} \end{pmatrix}.$$

Sejam $\Sigma = \text{Cov}(Z_i)$ e $\Sigma_\varepsilon = \text{Cov}(\varepsilon_i)$. Então, tem-se que

$$\Sigma = \begin{cases} \Sigma_x + \Sigma_\varepsilon, & \text{no caso (a),} \\ \Sigma_\varepsilon, & \text{no caso (b),} \end{cases}$$

onde $\Sigma_x = \text{Cov}(g_i) = \sigma_{xx}\mathbf{b}\mathbf{b}^\top$ com $\mathbf{b} = (\beta, 1)^\top$. Note-se que em (b) tem-se que $\text{Cov}(Y_i, X_i) = 0$, enquanto que em (a), $\text{Cov}(Y_i, X_i) = \beta\sigma_{xx} \neq 0$.

5.1.2 Inferência no modelo estrutural normal supondo razões de variâncias conhecidas

Lembrando que no enfoque estrutural as quantidades (não observadas) x_i , $i = 1, 2, \dots, n$, são assumidas aleatórias, em (5.9) tem-se que os vetores $g_i = g(x_i)$, $i = 1, 2, \dots, n$, são também aleatórios. Além disso, notando que $g_i = g(x_i)$ é linear em x_i e $Z_i = g_i + \varepsilon_i$, tem-se que a suposição de normalidade deve ser feita sobre a distribuição conjunta dos vetores $(x_i, \varepsilon_i^\top)^\top = (x_i, e_i, u_i)^\top$, $i = 1, 2, \dots, n$, para que a distribuição conjunta dos vetores (observados) Z_i , $i = 1, 2, \dots, n$, seja normal. Suponha, então, que

$$\begin{pmatrix} x_i \\ \varepsilon_i^\top \end{pmatrix} \stackrel{\text{iid}}{\sim} N_3 \left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & 0 \\ 0 & \Sigma_\varepsilon \end{pmatrix} \right), \quad (5.12)$$

para $i = 1, 2, \dots, n$, onde $\mu_x = E(x_i)$, $\sigma_{xx} = \text{Var}(x_i)$ e $\Sigma_\varepsilon = \text{Cov}(\varepsilon_i) = \text{Diag}\{\sigma_{ee}, \sigma_{uu}\}$. É fácil ver que (5.12) implica que

$$Z_i \stackrel{\text{iid}}{\sim} N_2(\mu, \Sigma), \quad (5.13)$$

para $i = 1, 2, \dots, n$, onde $\mu = \mu(\theta) = Z(Z_i)$ e $\Sigma = \Sigma(\theta) = \text{Cov}(Z_i)$ estão dados em (5.10) e (5.11), respectivamente, e

$$\theta = (\alpha, \mu_x, \sigma_{xx}, \sigma_{ee}, \sigma_{uu}, \beta)^\top. \quad (5.14)$$

Assim, $Z_1, Z_2, \dots, Z_n$ têm densidade conjunta dada por

$$p_{z_1, z_2, \dots, z_n}(z_1, z_2, \dots, z_n | \theta) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (z_i - \mu)^\top \Sigma^{-1} (z_i - \mu)}$$

e sua função log-verossimilhança, $L = L(\mu(\theta), \Sigma(\theta))$, é tal que

$$L \propto -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n (z_i - \mu)^\top \Sigma^{-1} (z_i - \mu) . \quad (5.15)$$

Como é bem sabido, os EMV de μ e Σ são dados por

$$\hat{\mu} = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i \quad \text{e} \quad \hat{\Sigma} = S_{ZZ} = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^\top ,$$

de modo que o EMV de θ , digamos $\hat{\theta}$, pode ser obtido como a solução das equações de momentos

$$\mu(\hat{\theta}) = \bar{Z} \quad \text{e} \quad \Sigma(\hat{\theta}) = S_{ZZ} ,$$

ou, equivalentemente,

$$\begin{pmatrix} \hat{\alpha} + \hat{\beta} \hat{\mu}_x \\ \hat{\mu}_x \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} \hat{\beta}^2 \hat{\sigma}_{xx} + \hat{\sigma}_{ee} \\ \hat{\beta} \hat{\sigma}_{xx} \\ \hat{\sigma}_{xx} + \hat{\sigma}_{uu} \end{pmatrix} = \begin{pmatrix} S_{YY} \\ S_{YX} \\ S_{XX} \end{pmatrix} \quad (5.16)$$

onde $\hat{\sigma}_{xx}$, $\hat{\sigma}_{ee}$ e $\hat{\sigma}_{uu}$ devem ser não-negativos. Daqui é claro que não é possível obter o EMV de θ sem informação adicional (veja Kendal e Stuart (1979, p.404)), uma vez que θ não é identificável pelo modelo (5.13). Em particular, a estimação de β não é possível. Contudo, é fácil ver que as equações acima implicam que

$$\frac{|S_{YX}|}{S_{XX}} \leq |\hat{\beta}| \leq \frac{S_{YY}}{|S_{YX}|} , \quad \text{se } \hat{\beta}, S_{YX} \neq 0 ,$$

de modo que $|\hat{\beta}|$ é limitado inferiormente pelo valor absoluto do coeficiente da regressão de Y em X (ou $\hat{\beta}_{MQ}$) e superiormente pelo recíproco do valor absoluto do coeficiente da regressão de X em Y . Por outro lado, note-se que dos seis parâmetros em θ , somente μ_x está identificado. Assim, em termos de identificar os parâmetros restantes, a maioria dos autores concordam em que não é desejável impor restrições sobre α e β (já que são

os parâmetros de maior interesse) ou sobre σ_{xx} (já que x não é observável). Consequentemente, as condições que mais comumente são aceitas vêm dadas sobre as variâncias dos erros (σ_{ee} e σ_{uu}). Várias destas condições são discutidas, por exemplo, em Kendall e Stuart (1979, cap. 29). Em vista dos resultados obtidos em (5.6) e (5.8), aqui estudam-se somente as condições de identificabilidade estabelecidas em (5.6) e (5.7), notando que a primeira delas envolve o parâmetro σ_{xx} . No que segue, estas restrições serão referidas como λ_x e λ_e conhecidos, entendendo-se que $\lambda_x = \sigma_{xx}/\sigma_{uu}$ e $\lambda_e = \sigma_{ee}/\sigma_{uu}$, e observando que, no primeiro caso, conhecer λ_x equivale a conhecer k_x (veja (5.5)). No entanto, no segundo caso, não se perde generalidade em supor que $\lambda_e = 1$ ($\sigma_{ee} = \sigma_{uu}$). Para cada uma destas suposições, o vetor de parâmetros θ definido em (5.14) se reduz a

$$\theta = \begin{cases} (\alpha, \mu_x, \sigma_{ee}, \sigma_{uu}, \beta)^\top, & \text{se } \lambda_x \text{ é conhecido,} \\ (\alpha, \mu_x, \sigma_{xx}, \sigma_{uu}, \beta)^\top, & \text{se } \lambda_e \text{ é conhecido.} \end{cases} \quad (5.17)$$

Considerando, então, as condições de identificabilidade λ_x ou λ_e conhecidos, tem-se que a matriz $\Sigma = \Sigma(\theta) = \text{Cov}(Z_i)$ (veja (5.11)) se reduz a

$$\Sigma = \begin{cases} \begin{pmatrix} \lambda_e \beta^2 \sigma_{uu} + \sigma_{ee} & \lambda_x \beta \sigma_{uu} \\ \lambda_x \beta \sigma_{uu} & (\lambda_x + 1) \sigma_{uu} \end{pmatrix}, & \text{se } \lambda_x \text{ é conhecido,} \\ \begin{pmatrix} \beta^2 \sigma_{xx} + \lambda_e \sigma_{uu} & \beta \sigma_{xx} \\ \beta \sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{pmatrix}, & \text{se } \lambda_e \text{ é conhecido,} \end{cases} \quad (5.18)$$

cujo determinante é dado por

$$|\Sigma| = \begin{cases} [\lambda_x \beta^2 \sigma_{uu} + (\lambda_x + 1) \sigma_{ee}] \sigma_{uu}, & \text{se } \lambda_x \text{ é conhecido,} \\ [\lambda_e \sigma_{uu} + (\beta^2 + \lambda_e) \sigma_{xx}] \sigma_{uu}, & \text{se } \lambda_e \text{ é conhecido.} \end{cases} \quad (5.19)$$

Assim, se $\hat{\beta}$ denota o EMV de β que segue de resolver (5.16) sob uma determinada condição de identificabilidade, então tem-se que

$$\hat{\beta} = \begin{cases} \hat{\beta}_{MQ}, & \text{se } \lambda_x \text{ é conhecido,} \\ \hat{\beta}_S, & \text{se } \lambda_e \text{ é conhecido,} \end{cases}$$

onde $\hat{\beta}_{MQ}$ e $\hat{\beta}_S$ estão definidos em (5.6) e (5.8), respectivamente. A distribuição exata de $\hat{\beta}$ é particularmente complicada para o caso com λ_e conhecido (veja Anderson e Sawa (1982)

e Kendall e Stuart (1979, p.412-414)), sendo comum usar sua distribuição assintótica na obtenção de intervalos de confiança ou testes de hipóteses para β . Alguns resultados desta natureza são dados no que segue. Em particular, deriva-se a distribuição exata de $\hat{\beta}$ quando λ_x é suposto conhecido (Seção 5.1.2.c) e a estatística escore sob ambas as condições de identificabilidade (Seção 5.1.2.d).

(a) Matriz de informação e parametrização ortogonal

Seja $K = K(\theta) = [\kappa_{i,j}]$ a matriz de informação ou modelo normal estrutural definido em (5.13), onde θ tem uma das especificações dadas em (5.17). Para derivar esta matriz, em (5.19) convém escrever $L = L(\theta)$ como $L = \sum_{i=1}^n \ell_i$, onde (omitindo o índice i) $\ell = \ell(\mu(\theta), \Sigma(\theta))$ é tal que

$$\ell \propto -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \|T\|^2, \quad (5.20)$$

com $T = \Sigma^{-1/2}(Z - \mu)^\top \sim N_2(0, I_2)$, $\mu = \mu(\theta) = (\alpha + \beta\mu_x, \mu_x)^\top$ e $\Sigma = \Sigma(\theta)$ tendo uma das expressões dadas em (5.18). Seja θ_i o i -ésimo componente do vetor θ (definido em (5.17)). Então,

$$\begin{aligned} \frac{\partial \|T\|^2}{\partial \theta_i} &= -2 \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1} (Z - \mu) + (Z - \mu)^\top \frac{\partial \Sigma^{-1}}{\partial \theta_i} (Z - \mu) \\ &= -2 \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1/2} T - T^\top \Sigma^{-1/2} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1/2} T, \end{aligned} \quad (5.21)$$

uma vez que

$$\frac{\partial \Sigma^{-1}}{\partial \theta_i} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1}.$$

Usando agora que

$$\frac{\partial \log |\Sigma|}{\partial \theta_i} = \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \right),$$

de (5.20) e (5.21) tem-se que

$$\frac{\partial \ell}{\partial \theta_i} = -\frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \right) + \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1/2} T + \frac{1}{2} T^\top \Sigma^{-1/2} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1/2} T. \quad (5.22)$$

Considerando este último resultado juntamente com (2.77) e (2.78) (lembrando também que aqueles termos que envolvem funções ímpares em $T \sim N_2(0, I_2)$, por exemplo, $(T^\top AT)BT$, têm esperança nula), é fácil ver que

$$\begin{aligned}\kappa_{i,j} &= nE \left(\frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j} \right) \\ &= n \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} + \frac{n}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right) .\end{aligned}$$

Daqui segue que

$$\kappa_{i,j} = \begin{cases} n \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} , & \text{se } i = 1, 2 \text{ e } j = 1, 2, 5, \\ \frac{n}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right) , & \text{se } i = 3, 4 \text{ e } j = 3, 4, 5, \\ 0 , & \text{se } i = 1, 2 \text{ e } j = 3, 4, \end{cases} \quad (5.23)$$

e

$$\kappa_{5,5} = \kappa_{\beta,\beta} = n \frac{\partial \mu^\top}{\partial \beta} \Sigma^{-1} \frac{\partial \mu}{\partial \beta} + \frac{n}{2} \text{tr} \left\{ \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right)^2 \right\} . \quad (5.24)$$

De (5.23) segue que, em particular, $\kappa_{\beta,j} \neq 0$ qualquer que seja o parâmetro θ_j , uma vez que β é parâmetro de locação e escala. Este fato dificulta, por exemplo, a determinação dos fatores de correção dos testes assintóticos da razão de verossimilhança e escore para hipóteses sobre β . Contudo, se o interesse é centrado neste parâmetro, tal dificuldade pode ser removida aplicando o procedimento de Cox e Reid (1977) para obter uma reparametrização do modelo na qual o parâmetro de interesse β seja ortogonal aos parâmetros restantes. Isto equivale a considerar a transformação do vetor θ (definido em (5.18)),

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \beta)^\top \rightarrow \phi = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)^\top ,$$

de modo que $\theta_i = \theta_i(\phi)$, $i = 1, 2, 3, 4$, sejam uma solução das equações diferenciais

$$\sum_{i=1}^4 \kappa_{i,j} \frac{\partial \theta_i}{\partial \beta} = -\kappa_{\beta,j} , \quad j = 1, 2, 3, 4 . \quad (5.25)$$

Resolver sistemas deste tipo em geral pode ser complicado. Além disso, quando o sistema tem solução, ela é arbitrária, sendo conveniente escolher uma na qual ϕ tenha uma interpretação adequada para o modelo estrutural em estudo. Uma solução de (5.25) é dada em

Wong (1989) no caso em que λ_e é suposto conhecido (e colocando $\lambda_e = 1$) e em Bolfarine e Cordani (1990) no caso em que λ_x (ou k_x) é assumido conhecido. Convém notar que em ambos os casos, (5.25) pode ser resolvida de forma independente para os parâmetros de locação (θ_1 e θ_2) e de escala (θ_3 e θ_4), uma vez que no modelo normal, em particular, ou elíptica, em geral, estes parâmetros são globalmente ortogonais. Assim, uma solução de (5.25) segue de escolher $\phi = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)^T$ como

$$\begin{aligned} \phi_1 &= \alpha + \beta\mu_x, \quad \phi_2 = \mu_x, \quad \phi_4 = \sigma_{uu} \quad \text{e} \\ \phi_3 &= \begin{cases} \lambda_x\beta^2\sigma_{uu} + (\lambda_x + 1)\sigma_{ee}, & \text{para } \lambda_x \text{ conhecido,} \\ (\beta^2 + \lambda_e)\sigma_{xx} + \lambda_e\sigma_{uu}, & \text{para } \lambda_e \text{ conhecido.} \end{cases} \end{aligned} \quad (5.26)$$

Note-se que a escolha dos novos parâmetros de locação, ϕ_1 e ϕ_2 , vem diretamente do fato que no modelo normal (ou elíptico), tem-se que o vetor de locação μ e a matriz de dispersão Σ são globalmente ortogonais (veja Mitchell (1989)). Assim, em (5.23), o fato de que $\kappa_{\beta,j} \neq 0$ se deve ao fato de que β está envolvido em μ e em Σ . Logo, se β é removido de μ ou de Σ se consegue que ele seja somente parâmetro de escala ou de locação, respectivamente. Como remover β de Σ não é simples, então escolhendo $\phi_1 = E(Y_i)$ e $\phi_2 = E(X_i)$ tem-se que em (5.26) β é removido de μ . A escolha dos novos parâmetros de escala não é tão óbvia quanto ϕ_1 e ϕ_2 , mas é motivada por (5.18), já que ϕ_3 e ϕ_4 são escolhidos de modo que $|\Sigma| = \phi_3\phi_4$ na nova parametrização (veja (5.29), abaixo).

Considerando (5.26), tem-se então que $\mu = E(Z_i)$ e $\Sigma = \text{Cov}(Z_i)$ são dados por

$$\mu = \mu(\phi_L) = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad (5.27)$$

e

$$\Sigma = \Sigma(\phi_E) = \begin{cases} (\lambda_x + 1)^{-1} \begin{pmatrix} \phi_3 + (\lambda_x\beta)^2\phi_4 & (\lambda_x + 1)\lambda_x\beta\phi_4 \\ (\lambda_x + 1)\lambda_x\beta\phi_4 & (\lambda_x + 1)^2\phi_4 \end{pmatrix}, & \text{para } \lambda_x \text{ conhecido,} \\ (\beta^2 + \lambda_e)^{-1} \begin{pmatrix} \beta^2\phi_3 + \lambda_e^2\phi_4 & \beta(\phi_3 - \lambda_e\phi_4) \\ \beta(\phi_3 - \lambda_e\phi_4) & \phi_3 + \beta^2\phi_4 \end{pmatrix}, & \text{para } \lambda_e \text{ conhecido,} \end{cases} \quad (5.28)$$

onde $\phi_L = (\phi_1, \phi_2)^\top$ e $\phi_E = (\phi_3, \phi_4, \beta)^\top$. Para ambos os casos, o determinante de Σ é dado por

$$|\Sigma| = \phi_3 \phi_4 . \quad (5.29)$$

Em particular, se no segundo caso, $\lambda_e = 1$, então $\text{tr}(\Sigma) = \phi_3 + \phi_4$ e, portanto, ϕ_3 e ϕ_4 são os valores próprios de Σ (com $\phi_3 > \phi_4 > 0$).

No que segue são dadas algumas propriedades de Σ e suas derivadas que facilitam a derivação dos cumulantes das derivadas de $L = L(\phi)$. Sejam $\{\alpha_3, \alpha_4\}$ os vetores definidos por

$$\{\alpha_3, \alpha_4\} = \begin{cases} \left\{ (\lambda_x + 1)^{-\frac{1}{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, (\lambda_x + 1)^{-\frac{1}{2}} \begin{pmatrix} \lambda_x \beta \\ \lambda_x + 1 \end{pmatrix} \right\}, & \text{para } \lambda_x \text{ conhecido,} \\ \left\{ (\beta^2 + \lambda_e)^{-\frac{1}{2}} \begin{pmatrix} \beta \\ 1 \end{pmatrix}, (\beta^2 + \lambda_e)^{-\frac{1}{2}} \begin{pmatrix} \lambda_e \\ -\beta \end{pmatrix} \right\}, & \text{para } \lambda_e \text{ conhecido.} \end{cases} \quad (5.30)$$

Então, com esta notação, tem-se os seguintes resultados:

- (i) $\frac{\partial \Sigma}{\partial \phi_i} = \alpha_i \alpha_i^\top$, $i = 3, 4$;
- (ii) $\Sigma = \phi_3 \frac{\partial \Sigma}{\partial \phi_3} + \phi_4 \frac{\partial \Sigma}{\partial \phi_4} = \phi_3 \alpha_3 \alpha_3^\top + \phi_4 \alpha_4 \alpha_4^\top$;
- (iii) $I_2 = \phi_3 \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_3} + \phi_4 \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_4} = \bar{\alpha}_3 \alpha_3^\top + \bar{\alpha}_4 \alpha_4^\top$,

onde I_2 denota a matriz identidade de dimensão 2×2 e $\bar{\alpha}_i = \phi_i \Sigma^{-1} \alpha_i$, $i = 3, 4$ são vetores tais que

$$\bar{\alpha}_i^\top \alpha_i = 1, \quad i = 3, 4, \quad \bar{\alpha}_i^\top \alpha_j = 0 \quad \text{e} \quad \|\bar{\alpha}_i\| = \|\alpha_j\|, \quad i, j = 3, 4, \quad i \neq j,$$

uma vez que $\Lambda_i = \bar{\alpha}_i \alpha_i^\top$, $i = 3, 4$, são matrizes idempotentes com $\text{tr}(\Lambda_i) = 1$, $i = 3, 4$, e $\Lambda_3 \Lambda_4 = 0$. É fácil ver que os vetores $\{\bar{\alpha}_3, \bar{\alpha}_4\}$ são dados por

$$\{\bar{\alpha}_3, \bar{\alpha}_4\} = \begin{cases} \left\{ (\lambda_x + 1)^{-\frac{1}{2}} \begin{pmatrix} \lambda_x + 1 \\ -\lambda_x \beta \end{pmatrix}, (\lambda_x + 1)^{-\frac{1}{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, & \text{para } \lambda_x \text{ conhecido,} \\ \left\{ (\beta^2 + \lambda_e)^{-\frac{1}{2}} \begin{pmatrix} \beta \\ \lambda_e \end{pmatrix}, (\beta^2 + \lambda_e)^{-\frac{1}{2}} \begin{pmatrix} 1 \\ -\beta \end{pmatrix} \right\}, & \text{para } \lambda_e \text{ conhecido.} \end{cases} \quad (5.31)$$

Note-se que no segundo caso tem-se que $\bar{\alpha}_i = \alpha_i$, $i = 3, 4$, se $\lambda_e = 1$. De qualquer modo, definindo as matrizes $A^\top = [\alpha_3, \alpha_4]$, $\bar{A}^\top = [\bar{\alpha}_3, \bar{\alpha}_4]$ e $\Phi = \text{Diag}\{\phi_3, \phi_4\}$, tem-se

que $A\Sigma^{-1}A^\top = \Phi^{-1}$ e $\bar{A}\Sigma\bar{A}^\top = \Phi$. Assim, dos resultados dados em (i) – (iii) (ou usando (5.29) segue que

$$\begin{aligned} \text{tr} \left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \right\} &= \phi_i^{-1}, \quad i = 3, 4, \\ \text{tr} \left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_j} \right\} &= \begin{cases} \phi_i^{-2}, & i = j, \\ 0, & i \neq j. \end{cases} \end{aligned} \quad (5.32)$$

De (iii) segue também que

$$(iv) \quad \Sigma^{-1} = \phi_3 \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_3} \Sigma^{-1} + \phi_4 \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_4} \Sigma^{-1} = \phi_3^{-1} \bar{\alpha}_3 \bar{\alpha}_3^\top + \phi_4^{-1} \bar{\alpha}_4 \bar{\alpha}_4^\top.$$

Usando agora que

$$\left\{ \frac{\partial \alpha_3}{\partial \beta}, \frac{\partial \alpha_4}{\partial \beta} \right\} = \begin{cases} \{0, \lambda_x \alpha_3\}, & \text{para } \lambda_x \text{ conhecido,} \\ \{(\beta^2 + \lambda_e)^{-1} \alpha_4, (\beta^2 + \lambda_e)^{-1} (-\lambda_e \alpha_3)\}, & \text{para } \lambda_e \text{ conhecido,} \end{cases}$$

tem-se que:

$$(v) \quad \frac{\partial \Sigma}{\partial \beta} = \phi_3 \frac{\partial^2 \Sigma}{\partial \phi_3 \partial \beta} + \phi_4 \frac{\partial^2 \Sigma}{\partial \phi_4 \partial \beta} = (\phi_3 \phi_4 / \sigma_\beta^2)^{\frac{1}{2}} (\alpha_3 \alpha_4^\top + \alpha_4 \alpha_3^\top),$$

onde $\sigma_\beta^2 = n \kappa_{\beta, \beta}^{-1}$ e é dada por

$$\sigma_\beta^2 = \begin{cases} \frac{\phi_3}{\lambda_x^2 \phi_4}, & \text{para } \lambda_x \text{ conhecido,} \\ \left(\frac{\beta^2 + \lambda_e}{\phi_3 - \lambda_e \phi_4} \right)^2 \phi_3 \phi_4, & \text{para } \lambda_e \text{ conhecido;} \end{cases} \quad (5.33)$$

$$(vi) \quad \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} = (\phi_3 \phi_4 / \sigma_\beta^2)^{\frac{1}{2}} (\phi_3^{-1} \bar{\alpha}_3 \alpha_4^\top + \phi_4^{-1} \bar{\alpha}_4 \alpha_3^\top),$$

$$\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right)^2 = (1 / \sigma_\beta^2) I_2 \quad \text{e} \quad \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-1} = (1 / \sigma_\beta^2 \phi_3 \phi_4)^{\frac{1}{2}} (\bar{\alpha}_3 \bar{\alpha}_4^\top + \bar{\alpha}_4 \bar{\alpha}_3^\top).$$

De (vi) segue que (ou diretamente de (5.29))

$$\text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right) = 0 \quad \text{e} \quad \text{tr} \left\{ \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right)^2 \right\} = \text{tr} \left\{ \Sigma^{-1} \frac{\partial^2 \Sigma}{\partial \beta^2} \right\} = 2 \sigma_\beta^{-2}. \quad (5.34)$$

Finalmente, seja $W = (W_3, W_4)^\top$ a transformação (ortogonal) de $T = \Sigma^{-\frac{1}{2}}(Z - \mu) \sim$

$N_2(0, I_2)$, definida por

$$W_i = \phi_i^{\frac{1}{2}} \alpha_i^\top \Sigma^{\frac{1}{2}} T = \phi_i^{-\frac{1}{2}} \bar{\alpha}_i^\top (Z - \mu), \quad i = 3, 4. \quad (5.35)$$

Então, (5.32) e (5.34) também podem ser obtidas a partir desta transformação, observando que

(vii) $W = {}^d T \sim N_2(0, I_2)$, isto é, W_3 e W_4 são iid $N(0, 1)$;

(viii) $(Z - \mu)^\top \Sigma^{-1} (Z - \mu) = \|T\|^2 = \|W\|^2$;

(ix) $-\frac{\partial \|T\|^2}{\partial \phi_i} = T^\top \Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \phi_i} \Sigma^{-\frac{1}{2}} T = \phi_i^{-1} W_i^2, i = 3, 4$;

(x) $-\frac{\partial \|T\|^2}{\partial \beta} = T^\top \Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-\frac{1}{2}} T = 2\sigma_\beta^{-1} W_3 W_4$.

Para provar estes últimos resultados é conveniente considerar a matriz $\bar{A}^\top = \bar{A}^\top(\beta) = [\bar{\alpha}_3, \bar{\alpha}_4]$ e notar que:

(xi) $\bar{A}(Z - \mu) \sim N_2(0, \Phi)$, onde $\Phi = [\bar{\alpha}_i^\top \Sigma \bar{\alpha}_j] = \text{Diag}\{\phi_3, \phi_4\}$.

Como foi mencionado, os resultados dados acima são úteis na derivação dos cumulantes das derivadas de L (veja o Apêndice A.1). Em particular, para obter a matriz de informação. Assim, denotando esta matriz por $K = K(\phi) = [\kappa_{i,j}]$ (com abuso de notação), de (5.23) e (5.24) (com θ substituído por ϕ) segue que

$$\kappa_{i,j} = \begin{cases} n \frac{\partial \mu^\top}{\partial \phi_i} \Sigma^{-1} \frac{\partial \mu}{\partial \phi_j}, & i, j = 1, 2 \text{ (localização)} \\ \frac{n}{2} \text{tr} \left\{ \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \right)^2 \right\}, & i = 3, 4, 5 \text{ (escala; } \phi_5 = \beta) \\ 0, & i = 1, 2, j = 3, 4, 5, \end{cases}$$

onde $\mu = \mu(\phi_L)$ e $\Sigma = \Sigma(\phi_E)$ estão dadas em (5.27) e (5.28), respectivamente. Considerando agora (5.27), (5.30) e (5.32), tem-se que

$$K = \text{Diag}\{K_L, K_E\}, \quad (5.36)$$

onde $K_L = n\Sigma^{-1}$ e $K_E = \text{Diag}\{\kappa_{3,3}, \kappa_{4,4}, \kappa_{\beta,\beta}\} = \text{Diag}\{n/2\phi_3^2, n/2\phi_4^2, n/\sigma_\beta^2\}$ são as submatrizes de informação dos vetores de localização $\phi_L = (\phi_1, \phi_2)^\top$ e escala $\phi_E = (\phi_3, \phi_4, \beta)^\top$,

respectivamente, e $\sigma_\beta^2 = \sigma_\beta^2(\phi_E)$ sendo definido como em (5.33). Convém notar aqui que σ_β^2 pode ser expressa como

$$\sigma_\beta^2 = \beta^2 \frac{\sigma_{XX} \sigma_{YY.X}}{\sigma_{YX}^2} = \beta^2 \left(\frac{1 - \rho_{YX}^2}{\rho_{YX}^2} \right), \quad (5.37)$$

onde $\sigma_{YX} = \text{Cov}(Y_i, X_i)$, $\sigma_{XX} = \text{Var}(X_i)$, $\sigma_{YY.X} = \text{Var}(Y_i|X_i)$ e $\rho_{YX} = \text{Correlação}(Y_i, X_i) = \text{Cov}(Y_i, X_i)/(\text{Var}(Y_i)\text{Var}(X_i))^{1/2}$.

(b) Estimação de máxima verossimilhança usando a parametrização ortogonal

Seja $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4, \hat{\beta})^\top$ o EMV de $\phi = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)^\top$, onde ϕ é parametrização ortogonal definida em (5.26) do modelo (5.13). Então, usando (5.22) (com θ substituído por ϕ) tem-se que $\hat{\phi}$ é a solução das equações dadas por

$$\begin{aligned} \frac{\partial L}{\partial \phi_i} \Big|_{\phi=\hat{\phi}} &= \left[\sum_{l=1}^n \frac{\partial \mu^\top}{\partial \phi_i} \Sigma^{-1} (Z_l - \mu) \right]_{\phi=\hat{\phi}} = 0, \quad i = 1, 2, \\ \frac{\partial L}{\partial \phi_i} \Big|_{\phi=\hat{\phi}} &= \left[-\frac{n}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \right) + \frac{1}{2} \sum_{l=1}^n (Z_l - \mu)^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \Sigma^{-1} (Z_l - \mu) \right]_{\phi=\hat{\phi}} = 0, \quad i = 3, 4, \\ \frac{\partial L}{\partial \beta} \Big|_{\phi=\hat{\phi}} &= \left[\frac{1}{2} \sum_{l=1}^n (Z_l - \mu)^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-1} (Z_l - \mu) \right]_{\phi=\hat{\phi}} = 0. \end{aligned} \quad (5.38)$$

Da primeira equação é claro que $\hat{\mu} = \mu(\hat{\phi}_L) = \bar{Z}$, de onde segue que para ambas as condições de identificabilidade, os EMV dos parâmetros de locação ϕ_1 e ϕ_2 são dados por

$$\hat{\phi}_1 = \bar{Y} \quad \text{e} \quad \hat{\phi}_2 = \bar{X}. \quad (5.39)$$

Considerando (5.39), em (5.30) tem-se que a segunda e terceira equações se reduzem a (veja também (iv) e (vi) em 5.1.2a)

$$\hat{\phi}_i = \bar{\alpha}_i^\top(\hat{\beta}) S_{ZZ} \bar{\alpha}_i(\hat{\beta}), \quad i = 3, 4 \quad \text{e} \quad \bar{\alpha}_3^\top(\hat{\beta}) S_{ZZ} \bar{\alpha}_4(\hat{\beta}) = 0,$$

respectivamente, onde $\bar{\alpha}_i(\hat{\beta})$, $i = 3, 4$, são os vetores definidos em (5.31) mas avaliados em $\hat{\beta}$. Aplicando (5.31) neste último sistema, segue que os EMV dos parâmetros de escala ϕ_3, ϕ_4 e β vem dados por:

(a) Para λ_x conhecido,

$$\left. \begin{aligned} \hat{\phi}_3 &= (\lambda_x + 1)S_{YY} - 2(\lambda_x \hat{\beta})S_{YX} + (\lambda_x \hat{\beta})^2 \hat{\phi}_4, \\ \hat{\phi}_4 &= \frac{S_{XX}}{\lambda_x + 1}, \\ \hat{\beta} &= \left(\frac{\lambda_x + 1}{\lambda_x} \right) \frac{S_{YX}}{S_{XX}}. \end{aligned} \right\} \quad (5.40)$$

Substituindo $\hat{\phi}_4$ e $\hat{\beta}$ na expressão de $\hat{\phi}_3$, tem-se que

$$\hat{\phi}_3 = (\lambda_x + 1)S_{YY.X},$$

onde $S_{YY.X} = S_{YY} - S_{YX}^2 S_{XX}^{-1} = S_{YY}(1 - r_{YX}^2)$ e $r_{YX} = S_{YX}/(S_{YY}S_{XX})^{\frac{1}{2}}$.

(b) Para λ_e conhecido,

$$\left. \begin{aligned} \hat{\phi}_3 &= \frac{\hat{\beta}^2 S_{YY} + 2\lambda_e \hat{\beta} S_{YX} + \lambda_e S_{XX}}{\hat{\beta}^2 + \lambda_e}, \\ \hat{\phi}_4 &= \frac{S_{YY} - 2\hat{\beta} S_{YX} + \hat{\beta}^2 S_{XX}}{\hat{\beta}^2 + \lambda_e}, \\ \hat{\beta} &= \frac{S_{YY} - \lambda_e S_{XX} + \{(S_{YY} - \lambda_e S_{XX})^2 + 4\lambda_e S_{YX}^2\}^{\frac{1}{2}}}{2S_{YX}}. \end{aligned} \right\} \quad (5.41)$$

Em ambos os casos todos os estimadores podem ser expressos como funções explícitas dos elementos da matriz S_{ZZ} . Também pode-se mostrar que $\hat{\phi}_E = (\hat{\phi}_3, \hat{\phi}_4, \hat{\beta})^T$ é dado pela solução da equação de momentos

$$\Sigma(\hat{\phi}_E) = S_{ZZ}.$$

Contudo, formas diferentes (mas equivalentes) de alguns estimadores podem ser obtidas com este procedimento. Por exemplo, para $\hat{\phi}_3$ tem-se que

$$\hat{\phi}_3 = \begin{cases} (\lambda_x + 1)S_{YY} - (\lambda_x \hat{\beta})^2 \hat{\phi}_4, & \text{para } \lambda_x \text{ conhecido,} \\ \lambda_e S_{YY} - \hat{\beta} S_{YX}, & \text{para } \lambda_e \text{ conhecido.} \end{cases}$$

Note-se também que para ambos os casos $\hat{\phi}_3\hat{\phi}_4 = |\Sigma(\hat{\phi}_E)| = |S_{ZZ}| = S_{YY.X}S_{XX}$.

(c) Propriedades dos EMV

Para estudar as propriedades de $\hat{\phi} = (\hat{\phi}_L^\top, \hat{\phi}_E^\top)^\top$, lembre-se em primeiro lugar que (5.13) (o modelo normal) implica que

$$\hat{\mu} = \bar{Z} \sim N_2(\mu, \frac{1}{n}\Sigma) \quad \text{e} \quad \hat{\Sigma} = S_{ZZ} \sim W_2(\frac{1}{n}\Sigma, n-1), \quad (5.42)$$

são independentes, onde

$$\mu = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \quad \bar{Z} = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{YY} & \sigma_{YX} \\ \sigma_{YX} & \sigma_{XX} \end{pmatrix} \quad \text{e} \quad S_{ZZ} = \begin{pmatrix} S_{YY} & S_{YX} \\ S_{YX} & S_{XX} \end{pmatrix},$$

e $W_k(A, m)$ denota a distribuição Wishart k -variada com matriz de dispersão A e m graus de liberdade (veja, por exemplo, Muirhead (1982, cap. 3)). Usando agora o fato de que, na parametrização ortogonal, tem-se $\hat{\phi}_L = \phi_L(\hat{\mu})$ e $\hat{\phi}_E = \phi_E(\hat{\Sigma})$, então (5.42) implica que:

(1) $\hat{\phi}_L = (\hat{\phi}_1, \hat{\phi}_2)^\top$ e $\hat{\phi}_E = (\hat{\phi}_3, \hat{\phi}_4, \hat{\beta})^\top$ são independentes;

(2) $\hat{\phi}_L = \bar{Z} \sim N_2(\phi_L, \frac{1}{n}\Sigma)$.

Assim, para obter regiões de confiança de ϕ_L ou para testar a hipótese $H_0 : \phi_L = \gamma_0$ pode-se usar o seguinte fato (veja Muirhead (1982, p.98)):

$$(3) \quad F = \left(\frac{n-2}{2} \right) (\hat{\phi}_L - \phi_L)^\top (\Sigma(\hat{\phi}_E))^{-1} (\hat{\phi}_L - \phi_L) \\ = \left(\frac{n-2}{2} \right) (\bar{Z} - \mu)^\top S_{ZZ}^{-1} (\bar{Z} - \mu) \sim F_{2, n-2}.$$

Similarmente, intervalos de confiança ou testes de hipóteses para funções de ϕ_L da forma $a^\top \phi_L$ (com a sendo um vetor desconhecido) podem ser obtidos através do seguinte resultado:

$$(4) \quad \sqrt{n}(a^\top \hat{\phi}_L - a^\top \phi_L) / (a^\top S_{ZZ} a)^{\frac{1}{2}} \sim t(0, 1; n-1).$$

A distribuição exata de $\hat{\phi}_E = \phi_E(S_{ZZ})$ é determinada pela distribuição de S_{ZZ} . As distribuições marginais de $\hat{\phi}_E$ são particularmente difíceis de derivar no caso com λ_e conhecido, uma vez que neste caso as componentes de $\hat{\phi}_E = \phi_E(S_{ZZ})$ são funções complicadas de S_{ZZ} (veja (5.41)). Por exemplo, para $\lambda_e = 1$, $\hat{\phi}_3$ e $\hat{\phi}_4$ são os valores próprios da matriz Wishart $\Sigma(\hat{\phi}_E) = S_{ZZ} \sim W_2(n^{-1}\Sigma; n-1)$. Nesta situação, a distribuição de $(\hat{\phi}_3, \hat{\phi}_4)^\top$ pode ser representada em termos de séries infinitas (veja Muirhead (1982, p.106)). Similarmente, neste modelo com λ_e conhecido a distribuição exata de $\hat{\beta}$ também deve ser representada em termos de séries infinitas (Anderson e Sawa (1982) derivam a distribuição de $\hat{\beta}$ no modelo funcional supondo $\sigma_{ee} = \sigma_{uu} = 1$). Por esta razão, para λ_e conhecido, as inferências sobre β são comumente baseadas na distribuição assintótica de $\hat{\beta}$, a saber,

$$\hat{\beta} \sim N(\beta, \sigma_\beta^2/n),$$

onde σ_β^2 está definido em (5.33) (veja também (5.37)). Contudo, usando as propriedades da distribuição Wishart, pode-se estudar as distribuições de certas funções de $\hat{\phi}_E$ em ambos os modelos (λ_x e λ_e conhecidos). No que segue são desenvolvidos alguns resultados ilustrativos. Em particular, são derivadas as distribuições marginais de $\hat{\phi}_3$, $\hat{\phi}_4$ e $\hat{\beta}$ para o modelo que supõe λ_x conhecido.

Considere-se a distribuição Wishart dada em (5.42). Então, como é bem sabido, nesta distribuição tem-se que (veja, por exemplo, Muirhead (1982, p.93)):

- (5) (i) $S_{YY.X} = S_{YY} - \frac{S_{YX}^2}{S_{XX}} = S_{YY}(1 - r_{YX}^2)$ é independente de (S_{YX}, S_{XX}) ;
(ii) $n \frac{S_{YY.X}}{\sigma_{YY.X}} \sim \chi_{n-2}^2$, onde $\sigma_{YY.X} = \sigma_{YY} - \frac{\sigma_{YX}^2}{\sigma_{XX}} = \sigma_{YY}(1 - \rho_{YX}^2)$;
(iii) $(S_{YX}|S_{XX}) \sim N\left(\frac{\sigma_{YX}}{\sigma_{XX}}S_{XX}, \frac{\sigma_{YY.X}}{n}S_{XX}\right)$;
(iv) $n \frac{S_{YY}}{\sigma_{YY}} \sim \chi_{n-1}^2$ e $n \frac{S_{XX}}{\sigma_{XX}} \sim \chi_{n-1}^2$.

Note-se que (5.iv) vem diretamente do fato que (5.16) implica que $Y = (Y_1, Y_2, \dots, Y_n)^\top \sim N_n(\mu_Y \mathbf{1}_n, \sigma_{YY} I_n)$ e $X = (X_1, X_2, \dots, X_n)^\top \sim N_n(\mu_x \mathbf{1}_n, \sigma_{xx} I_n)$, enquanto que (5.i) a (5.iii)

podem ser demonstrados considerando o modelo de regressão condicional definido por

$$(Y_i|X_i) \stackrel{\text{ind.}}{\sim} N(\mu_Y + \sigma_{YX}\sigma_{XX}^{-1}(X_i - \mu_X), \sigma_{YY.X}) .$$

Lembre-se também que $\mu = (\mu_Y, \mu_X)^\top = (\phi_1, \phi_2)^\top$. As representações em termos de $\phi_E = (\phi_3, \phi_4, \beta)^\top$ das funções de $\Sigma = [\sigma_{ij}]$ envolvidas acima vêm diretamente de (5.28). De forma análoga, obtém-se as funções de $S_{ZZ} = [S_{ij}]$ em termos de $\hat{\phi}_E$.

Observe-se agora que a propriedade (5.iii) pode ser expressa como

$$\left\{ \left(\frac{S_{YX}}{S_{XX}} - \frac{\sigma_{YX}}{\sigma_{XX}} \right) | S_{XX} \right\} \sim N \left(0, \frac{\sigma_{YY.X}}{nS_{XX}} \right) . \quad (5.43)$$

Sejam

$$U_{XX} = \frac{S_{YX}}{S_{XX}} - \frac{\sigma_{YX}}{\sigma_{XX}} \quad \text{e} \quad V_{XX} = \frac{(n-1)\sigma_{XX}}{nS_{XX}} .$$

Então, de (5.43), segue que

$$(U_{XX}|S_{XX}) \stackrel{d}{=} (U_{XX}|V_{XX}) \sim N \left(0, \frac{\sigma_{YY.X}}{(n-1)\sigma_{XX}} V_{XX} \right) ,$$

onde, de acordo com (5.iv),

$$V_{XX} \sim \frac{n-1}{\chi_{n-1}^2} ,$$

de modo que a distribuição incondicional de U_{XX} é dada por (veja o Capítulo 3)

$$U_{XX} \sim t \left(0, \frac{\sigma_{YY.X}}{(n-1)\sigma_{XX}}; n-1 \right) .$$

Daqui obtem-se a seguinte propriedade:

$$(6) \quad \frac{S_{YX}}{S_{XX}} \sim t \left(\frac{\sigma_{YX}}{\sigma_{XX}}, \frac{\sigma_{YY.X}}{(n-1)\sigma_{XX}}; n-1 \right)$$

Usando novamente (5.43) tem-se também que

$$\left\{ S_{XX}^{\frac{1}{2}} U_{XX} | S_{XX} \right\} \sim N \left(0, \frac{\sigma_{YY.X}}{n} \right) ,$$

para todo S_{XX} . Isto implica que $S_{XX}^{\frac{1}{2}} U_{XX}$ é independente de S_{XX} , de modo que sua distribuição incondicional é dada por

$$(7) \quad S_{XX}^{\frac{1}{2}} U_{XX} = S_{XX}^{\frac{1}{2}} \left(\frac{S_{YX}}{S_{XX}} - \frac{\sigma_{YX}}{\sigma_{XX}} \right) \sim N \left(0, \frac{\sigma_{YY.X}}{n} \right).$$

Utilizando agora que $\hat{\sigma}_{YY.X} = S_{YY.X}$, então as propriedades (7), (5.i) e (5.ii) implicam que

$$(8) \quad \left\{ \frac{(n-2)S_{XX}}{S_{YY.X}} \right\}^{\frac{1}{2}} \left(\frac{S_{YX}}{S_{XX}} - \frac{\sigma_{YX}}{\sigma_{XX}} \right) \sim t(0, 1; n-2).$$

Este resultado permite inferir sobre a função de ϕ_E dada por

$$\frac{\sigma_{YX}}{\sigma_{XX}} = \begin{cases} \left(\frac{\lambda_x}{\lambda_x + 1} \right) \beta, & \text{para } \lambda_x \text{ conhecido,} \\ \left(\frac{\phi_3 - \lambda_e \phi_4}{\phi_3 + \beta^2 \phi_4} \right) \beta, & \text{para } \lambda_e \text{ conhecido.} \end{cases}$$

Em particular, usando (8) juntamente com (5.40), é fácil ver que sob a hipótese de independência entre Y e X , isto é, sob $H_0 : \beta = 0$, tem-se (veja também o Exemplo 2.12)

$$\frac{\sqrt{n-2}\hat{\beta}}{\hat{\sigma}_\beta} = \frac{\sqrt{n-2}r_{YX}}{\sqrt{1-r_{YX}^2}} = \frac{\sqrt{n-2}S_{YX}}{\sqrt{S_{XX}S_{YY.X}}} \sim t(0, 1; n-2),$$

onde $\sigma_\beta = \sigma_\beta(\phi_E)$ está definida em (5.33).

Usando os resultados dados em (5) também pode-se determinar a distribuição da função de $\hat{\phi}_E$ dada por

$$\hat{\phi}_3 \hat{\phi}_4 = |\Sigma(\hat{\phi}_E)| = |S_{ZZ}| = S_{YY.X} S_{XX}.$$

Daqui segue por exemplo que:

$$(9) \quad (i) \quad E(\hat{\phi}_3 \hat{\phi}_4) = E(S_{YY.X})E(S_{XX}) = \left(\frac{n-2}{n} \right) \left(\frac{n-1}{n} \right) \phi_3 \phi_4, \text{ uma vez que}$$

$$\sigma_{YY.X} \sigma_{XX} = |\Sigma| = \phi_3 \phi_4;$$

$$\begin{aligned} (ii) \quad E \left\{ \log \left(\frac{\hat{\phi}_3 \hat{\phi}_4}{\phi_3 \phi_4} \right) \right\} &= E \left\{ \log \left(\frac{S_{YY.X} S_{XX}}{\sigma_{YY.X} \sigma_{XX}} \right) \right\} \\ &= E \left\{ \log \left(\frac{S_{YY.X}}{\sigma_{YY.X}} \right) \right\} + E \left\{ \log \left(\frac{S_{XX}}{\sigma_{XX}} \right) \right\} \\ &= \psi \left(\frac{n-2}{2} \right) + \psi \left(\frac{n-1}{2} \right) - 2 \log \frac{n}{2}, \end{aligned}$$

onde se usa o fato de que se $V \sim Ga(\nu, \delta)$, então $E(\log V) = \psi(\nu) - \log \delta$, com ψ sendo a função digama. No modelo com λ_e conhecido, se $\lambda_e = 1$ tem-se também que $\hat{\phi}_3 + \hat{\phi}_4 = \text{tr}(\Sigma(\hat{\phi}_e)) = \text{tr}(S_{ZZ}) = S_{YY} + S_{XX}$, de onde segue que $E(\hat{\phi}_3 + \hat{\phi}_4) = (\frac{n-1}{n})(\phi_3 + \phi_4)$.

Considere-se, finalmente, o modelo com λ_x conhecido. De (5.28) é fácil ver que neste modelo

$$\phi_3 = (\lambda_x + 1)\sigma_{YY.X}, \quad \phi_4 = \frac{\sigma_{XX}}{\lambda_x + 1} \quad \text{e} \quad \beta = \left(\frac{\lambda_x + 1}{\lambda_x} \right) \frac{\sigma_{YX}}{\sigma_{XX}},$$

cujos respectivos EMV são dados por (veja (5.40))

$$\hat{\phi}_3 = (\lambda_x + 1)S_{YY.X}, \quad \hat{\phi}_4 = \frac{S_{XX}}{\lambda_x + 1} \quad \text{e} \quad \hat{\beta} = \left(\frac{\lambda_x + 1}{\lambda_x} \right) \frac{S_{YX}}{S_{XX}} = \frac{S_{YX}}{\lambda_x \hat{\phi}_4}.$$

Considerando estas relações, das propriedades estabelecidas em (5), (6), (7) e (8) segue que

(10) se λ_x é conhecido, então:

- (i) $\hat{\phi}_3, \hat{\phi}_4$ e $\hat{\phi}_4^{\frac{1}{2}}(\hat{\beta} - \beta)$ são independentes;
- (ii) $\frac{n\hat{\phi}_3}{\phi_3} \sim \chi_{n-2}^2$ e $\frac{n\hat{\phi}_4}{\phi_4} \sim \chi_{n-1}^2$;
- (iii) $\hat{\phi}_4^{\frac{1}{2}}(\hat{\beta} - \beta) \sim N(0, \frac{\sigma_{\beta}^2 \phi_4}{n})$, onde $\sigma_{\beta}^2 = \frac{\phi_3}{\lambda_x^2 \phi_4}$ (veja (5.33));
- (iv) $\hat{\beta} \sim t\left(\beta, \frac{\sigma_{\beta}^2}{n-1}; n-1\right)$,
- (v) $\frac{\sqrt{n-2}(\hat{\beta} - \beta)}{\hat{\sigma}_{\beta}} \sim t(0, 1; n-2)$.

A prova de (10.i) vem diretamente de (5.i) e (7). As propriedades (5.ii) e (5.iv) implicam (10.ii). Usando (7) tem-se (10.iii). Enquanto que (10.iv) e (10.v) seguem de (6) e (8), respectivamente.

Note-se que (10.iv) implica que

$$E(\hat{\beta}) = \beta, \quad (n-1 > 1) \quad \text{e} \quad \text{Var}(\hat{\beta}) = \frac{\sigma_{\beta}^2}{n-3}, \quad (n-1 > 2).$$

De (10.iv) também é claro que $\sqrt{n}(\hat{\beta} - \beta)$ converge em lei para uma variável aleatória com distribuição $N(0, \sigma_\beta^2)$, uma vez que $\sqrt{n-1}(\hat{\beta} - \beta) \sim t(0, \sigma_\beta^2; n-1)$ tende para a lei $N(0, \sigma_\beta^2)$ quando $n \rightarrow \infty$ e $\sqrt{n}(\hat{\beta} - \beta) = \sqrt{\frac{n}{n-1}} \sqrt{n-1}(\hat{\beta} - \beta)$. Além disso, de (10.v) decorre que um intervalo de confiança exato para β , com nível $1 - \gamma$, é dado por

$$\beta = \hat{\beta} \pm t_{n-1; 1-\gamma/2} \frac{\hat{\sigma}_\beta}{\sqrt{n-2}},$$

onde $t_{n-1; 1-\gamma/2}$ denota o percentil $(1 - \gamma/2)100\%$ da distribuição t padrão (ou Student) com $n - 1$ graus de liberdade.

(d) Testes da razão de verossimilhança e escore da hipótese $H_0 : \beta = \beta_0$

Seja $\tilde{\phi} = (\tilde{\phi}_L^\top, \tilde{\phi}_E^\top)^\top$ o EMV de $\phi = (\phi_L^\top, \phi_E^\top)^\top$ para o modelo (5.13) sob a hipótese nula $H_0 : \beta = \beta_0$. Então, é fácil ver que $\hat{\phi}_L = \tilde{\phi}_L = \bar{Z}$ e $\tilde{\phi}_E = (\tilde{\phi}_3, \tilde{\phi}_4, \tilde{\beta})^\top$ é dado por

$$\tilde{\beta} = \beta_0 \quad \text{e} \quad \tilde{\phi}_i = \bar{\alpha}_i^\top(\beta_0) S_{ZZ} \bar{\alpha}_i(\beta_0), \quad i = 3, 4,$$

onde $\bar{\alpha}_i(\beta_0)$, $i = 3, 4$, são os vetores definidos em (5.31) mas avaliados em $\beta = \beta_0$, isto é, $\tilde{\phi}_3$ e $\tilde{\phi}_4$ podem ser obtidos a partir das expressões de $\hat{\phi}_3$ e $\hat{\phi}_4$ (dadas em (5.40) ou (5.41)), mas substituindo $\hat{\beta}$ por $\tilde{\beta} = \beta_0$. Note-se também que no modelo com λ_x conhecido, tem-se que $\tilde{\phi}_4 = \hat{\phi}_4$.

Por outro lado, sob $H_0 : \beta = \beta_0$ tem-se que $nS_{XX} \sim W_2(\Sigma_0, n-1)$, onde $\Sigma_0 = \Sigma(\phi_3, \phi_4, \beta_0)$. Seja $\bar{A}_0 = \bar{A}(\beta_0)$, onde $\bar{A}^\top = [\bar{\alpha}_3, \bar{\alpha}_4]$. Então, do fato de que $\bar{A} S_{ZZ} \bar{A}^\top \sim W_2(n^{-1}\Phi; n-1)$, onde $\Phi = \bar{A} \Sigma \bar{A}^\top = [\bar{\alpha}_i^\top \Sigma \bar{\alpha}_j] = \text{Diag}\{\phi_3, \phi_4\}$, segue que, sob $H_0 : \beta = \beta_0$, a distribuição de $\tilde{\phi}_i$ é dada por

$$\frac{n\tilde{\phi}_i}{\phi_i} \sim \chi_{n-1}^2, \quad i = 3, 4. \quad (5.44)$$

Contudo, a independência entre $\tilde{\phi}_3$ e $\tilde{\phi}_4$ é válida somente no modelo com λ_x conhecido. Com estes resultados pode-se estudar as estatísticas da razão de verossimilhança (G) e escore (S), para testar a hipótese nula $H_0 : \beta = \beta_0$ (contra a alternativa $H_1 : \beta \neq \beta_0$) no

modelo (5.13). Assim, usando (5.15) tem-se que G é dada por

$$G = 2(L(\hat{\phi}) - L(\tilde{\phi})) = n \log \left\{ \frac{|\Sigma(\tilde{\phi}_E)|}{|\Sigma(\hat{\phi}_E)|} \right\} = n \log \left\{ \frac{\tilde{\phi}_3 \tilde{\phi}_4}{\hat{\phi}_3 \hat{\phi}_4} \right\}.$$

Sob $H_0 : \beta = \beta_0$, a estatística G tem distribuição assintótica qui-quadrado com um grau de liberdade ($G \sim A\chi_1^2$), e pode ser expressa em termos de β_0 e dos elementos de S_{ZZ} como

$$G(\beta_0) = \begin{cases} n \log \left\{ \frac{(\lambda_x + 1)S_{YY} - 2\beta_0\lambda_x S_{YX} + (\beta_0\lambda_x)^2 S_{XX}}{(\lambda_x + 1)S_{YY.X}} \right\}, \\ \text{para } \lambda_x \text{ conhecido,} \\ n \log \left\{ \frac{(\beta_0^2 S_{YY} + 2\lambda_e \beta_0 S_{YX} + \lambda_e^2 S_{XX})(S_{YY} - 2\beta_0 S_{YX} + \beta_0^2 S_{XX})}{(\beta_0^2 + \lambda_e)^2 S_{YY.X} S_{XX}} \right\}, \\ \text{para } \lambda_e \text{ conhecido,} \end{cases} \quad (5.45)$$

onde $S_{YY.X} = S_{YY}(1 - r_{YX}^2)$. Note-se que no caso com λ_x conhecido,

$$G = n \log(\tilde{\phi}_3 / \hat{\phi}_3).$$

Como foi visto no Capítulo 4 (veja a Seção 4.3.3), a aproximação da distribuição de G (sob a hipótese nula) para a distribuição qui-quadrado (χ_1^2 , neste caso) pode ser melhorada utilizando-se o fator de Bartlett (veja (4.52)). Este fator é derivado por Wong (1989) para o modelo com λ_e conhecido (com $\lambda_e = 1$) e por Bolfarine e Nascimento (1992a) para λ_x conhecido. Em ambos os trabalhos o fator é determinado utilizando o procedimento de Lawley (1956) descrito na Seção 4.3.3 do Capítulo 4. Tal procedimento requer um processo algébrico extenso e muito trabalhoso, uma vez que envolve cumulantes conjuntos das derivadas da função de log-verossimilhança de até quarta ordem (veja a Seção 5.4.2). No que segue se mostra que este fator pode ser obtido avaliando diretamente o valor esperado de G sob $H_0 : \beta = \beta_0$. De fato, se $E_0(G)$ denota o valor esperado de G sob H_0 , então considerando (5.44) e procedendo como em (9.ii), tem-se que

$$E_0(G) = n \left\{ E_0 \left(\log \left\{ \frac{\tilde{\phi}_3 \tilde{\phi}_4}{\phi_3 \phi_4} \right\} \right) - E_0 \left(\log \left\{ \frac{\hat{\phi}_3 \hat{\phi}_4}{\phi_3 \phi_4} \right\} \right) \right\}$$

$$\begin{aligned}
&= n \left\{ \left(\psi \left(\frac{n-1}{2} \right) + \psi \left(\frac{n-1}{2} \right) - 2 \log \frac{n}{2} \right) - \left(\psi \left(\frac{n-2}{2} \right) - \psi \left(\frac{n-1}{2} \right) - 2 \log \frac{n}{2} \right) \right\} \\
&= n \left(\psi \left(\frac{n-1}{2} \right) - \psi \left(\frac{n-2}{2} \right) \right),
\end{aligned}$$

onde $\psi(m)$ é a função digama avaliada em m . Usando o fato de que (veja Abramowitz e Stegun (1965))

$$\psi(m) = \psi(m-1) + \frac{1}{m-1}$$

e

$$\begin{aligned}
\psi(m) &= \log m - \frac{1}{2m} - \frac{1}{12m^2} + \frac{1}{120m^4} - \frac{1}{252m^6} + \dots \\
&= \log m - \frac{1}{2m} - \frac{1}{12m^2} + O(m^{-4}),
\end{aligned}$$

tem-se que

$$\begin{aligned}
E_0(G) &= n \left\{ \left(\psi \left(\frac{n+1}{2} \right) - \psi \left(\frac{n}{2} \right) \right) - 2 \left(\frac{1}{n-1} - \frac{1}{n-2} \right) \right\} \\
&= n \left\{ \left[\left(\log \left(\frac{n+1}{2} \right) - \log \frac{n}{2} \right) - \left(\frac{1}{n+1} - \frac{1}{n} \right) - \frac{1}{3} \left(\frac{1}{(n+1)^2} - \frac{1}{n^2} \right) \right. \right. \\
&\quad \left. \left. + O(n^{-4}) \right] - 2 \left(\frac{1}{n-1} - \frac{1}{n-2} \right) \right\} \\
&= n \left\{ \log \left(1 + \frac{1}{n} \right) - \frac{1}{n} \left[\left(1 + \frac{1}{n} \right)^{-1} - 1 \right] - \frac{1}{3n^2} \left[\left(1 + \frac{1}{n} \right)^{-2} - 1 \right] \right. \\
&\quad \left. + O(n^{-4}) - \frac{2}{n} \left[\left(1 - \frac{1}{n} \right)^{-1} - \frac{1}{2} \left(\frac{1}{2} - \frac{1}{n} \right)^{-1} \right] \right\}.
\end{aligned}$$

Substituindo nesta última expressão as expansões dadas por

$$\begin{aligned}
\log \left(1 + \frac{1}{n} \right) &= \frac{1}{n} + \frac{1}{2n^2} + O(n^{-3}), \quad \left(1 + \frac{1}{n} \right)^{-k} = 1 - \frac{k}{n} + O(n^{-2}), \quad k = 1, 2, \\
\left(1 - \frac{1}{n} \right)^{-1} &= 1 + \frac{1}{n} + O(n^{-2}), \quad \left(\frac{1}{2} - \frac{1}{n} \right)^{-1} = 2 + \frac{4}{n} + O(n^{-2}),
\end{aligned}$$

segue que

$$\begin{aligned}
E_0(G) &= n \left\{ \frac{1}{n} \left[1 + \frac{1}{2n} + O(n^{-2}) \right] - \frac{1}{n} \left[\left(1 - \frac{1}{n} + O(n^{-2}) \right) - 1 \right] \right. \\
&\quad \left. - \frac{1}{3n^2} \left[\left(1 - \frac{2}{n} + O(n^{-2}) \right) - 1 \right] + O(n^{-4}) \right\}
\end{aligned}$$

$$\begin{aligned}
& -\frac{2}{n} \left[\left(1 + \frac{1}{n} + O(n^{-2}) \right) - \frac{1}{2} \left(2 + \frac{4}{n} + O(n^{-2}) \right) \right] \} \\
& = 1 + \frac{5}{2n} + O(n^{-2}) .
\end{aligned}$$

Isto é, para ambos os modelos (λ_x ou λ_e conhecido) a estatística G modificada do teste de $H_0 : \beta = \beta_0$ contra $H_1 : \beta \neq \beta_0$ é dada por

$$G^* = \left\{ 1 + \frac{5}{2n} \right\}^{-1} G , \quad (5.46)$$

(ou $G^* = \{1 - d\}G$) com G sendo como em (5.45).

Um teste alternativo para $H_0 : \beta = \beta_0$ é dado pela estatística escore S . Esta estatística não tem sido mencionada na literatura relativa aos modelos com erros. Contudo, o teste baseado em S apresentou um bom desempenho no estudo de simulação realizado no Capítulo 4 (veja Seção 4.4.5). Devido a estes fatos, no que segue deriva-se a estatística S para o modelo (5.13) sob parametrização ortogonal.

De (5.26) (com θ substituído por ϕ) segue que o escore relativo a β , avaliado em $\tilde{\phi}$, é dado por

$$\frac{\partial L(\phi)}{\partial \beta} \Big|_{\phi=\tilde{\phi}} = \frac{n}{2} \text{tr} (B(\tilde{\phi}) S_{ZZ}) ,$$

onde (veja o Fato (vi) dado em a)

$$B(\phi) = \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-1} = \frac{1}{\sigma_\beta \sqrt{\phi_3 \phi_4}} (\bar{\alpha}_3 \bar{\alpha}_4^\top + \bar{\alpha}_4 \bar{\alpha}_3^\top) ,$$

onde σ_β e $\bar{\alpha}_i$, $i = 3, 4$, são definidos como em (5.33) e (5.31), respectivamente. Então,

$$\frac{\partial L(\phi)}{\partial \beta} \Big|_{\phi=\tilde{\phi}} = n \frac{\bar{\alpha}_3^\top(\beta_0) S_{ZZ} \bar{\alpha}_4(\beta_0)}{\tilde{\sigma}_\beta \sqrt{\tilde{\phi}_3 \tilde{\phi}_4}}$$

onde $\tilde{\sigma}_\beta = \sigma_\beta(\tilde{\phi}_E)$. Lembrando que $\sigma_\beta^2 = n \kappa_{\beta,\beta}^{-1}$ (veja (5.33)), tem-se que

$$S = \frac{1}{\tilde{\kappa}_{\beta,\beta}} \left(\frac{\partial L(\phi)}{\partial \beta} \Big|_{\phi=\tilde{\phi}} \right)^2 = n \frac{(\bar{\alpha}_3^\top(\beta_0) S_{ZZ} \bar{\alpha}_4(\beta_0))^2}{\tilde{\phi}_3 \tilde{\phi}_4} ,$$

onde $\tilde{\phi}_i = \bar{\alpha}_i^\top(\beta_0)S_{ZZ}\bar{\alpha}_i(\beta_0)$, $i = 3, 4$. Sob $H_0 : \beta = \beta_0$, $S \sim A\chi_1^2$ e pode ser expressa em termos de β_0 e dos elementos da matriz S_{ZZ} aplicando (5.31). Isto é,

$$S(\beta_0) = \begin{cases} \frac{n((\lambda_x + 1)S_{YX} - \lambda_x\beta_0S_{XX})^2}{((\lambda_x + 1)^2S_{YY} - 2\lambda_x(\lambda_x + 1)\beta_0S_{YX} + \lambda_x^2\beta_0^2S_{XX})S_{XX}} , & \text{para } \lambda_x \text{ conhecido,} \\ \frac{n(\beta_0(S_{YY} - \lambda_eS_{XX}) - (\beta_0^2 - \lambda_e)S_{YX})^2}{(\beta_0^2S_{YY} + 2\lambda_e\beta_0S_{YX} + \lambda_e^2S_{XX})(S_{YY} - 2\beta_0S_{YX} + \beta_0^2S_{XX})} , & \text{para } \lambda_e \text{ conhecido.} \end{cases} \quad (5.47)$$

É fácil ver que no modelo com λ_x conhecido, tem-se que

$$S = n \left(\frac{\hat{\beta} - \beta_0}{\tilde{\sigma}_\beta} \right)^2 ,$$

uma vez que $n\kappa_{\beta,\beta}^{-1} = \tilde{\sigma}_\beta^2 = \tilde{\phi}_3/\lambda_x^2\tilde{\phi}_4$.

Para ambos os modelos, a estatística S tem distribuição exata. De fato, seja $\tilde{\Phi} = [\tilde{\phi}_{ij}]$ a matriz definida por

$$\tilde{\Phi} = \bar{A}_0 S_{ZZ} \bar{A}_0^\top ,$$

onde $\bar{A}_0^\top = \bar{A}^\top(\beta_0) = [\bar{\alpha}_3(\beta_0), \bar{\alpha}_4(\beta_0)]$, isto é,

$$\tilde{\phi}_{ij} = \bar{\alpha}_i^\top(\beta_0)S_{ZZ}\bar{\alpha}_j(\beta_0) , \quad i, j = 3, 4,$$

e $\tilde{\phi}_{ii} = \tilde{\phi}_i$. É claro que S pode ser escrita como

$$S = n \frac{\tilde{\phi}_{34}^2}{\tilde{\phi}_3\tilde{\phi}_4} ,$$

de modo que sua distribuição é determinada pela distribuição da matriz $\tilde{\Phi}$. Mas, sob $H_0 : \beta = \beta_0$,

$$\tilde{\Phi} \sim W_2(n^{-1}\Phi_0; n-1) ,$$

onde $\Phi_0 = \bar{A}_0\Sigma_0\bar{A}_0^\top = \text{Diag}\{\phi_3, \phi_4\}$. Então, aplicando a propriedade (8) (dada em c) para a matriz $\tilde{\Phi}$, tem-se que, sob H_0 ,

$$t^2 = \left\{ \frac{(n-2)\tilde{\phi}_4}{\tilde{\phi}_{3.4}} \right\}^{1/2} \left(\frac{\tilde{\phi}_{34}}{\tilde{\phi}_4} - \frac{\phi_{34}}{\phi_4} \right) \sim t(0, 1; n-2) ,$$

onde $\tilde{\phi}_{3.4} = \tilde{\phi}_3 - \tilde{\phi}_{34}^2 \tilde{\phi}_4^{-1} = \tilde{\phi}_3 \tilde{\phi}_4 (1 - S/n)$ e $\phi_{34} = 0$, já que Φ_0 é uma matriz diagonal.

Assim, desta última expressão segue que

$$S = n \left(\frac{t^2}{n-2} \right) \left(1 + \frac{t^2}{n-2} \right)^{-1},$$

de modo que, sob H_0 , $n^{-1}S \sim \text{Beta}(1/2, (n-2)/2)$.

No modelo com λ_x conhecido, tem-se também que um teste exato de $H_0 : \beta = \beta_0$ pode ser obtido usando o fato de que, sob H_0 , (veja (10.v) em c),

$$\frac{\sqrt{n-2}(\hat{\beta} - \beta_0)}{\hat{\sigma}_\beta} \sim t(0, 1; n-2).$$

Na seção que segue, são estudadas duas extensões baseadas nas distribuições elípticas do modelo normal estrutural desenvolvido acima. Na primeira destas extensões supõe-se que a distribuição conjunta dos vetores $(x_i, e_i, u_i)^T$, $i = 1, 2, \dots, n$ é elíptica, de modo que estes vetores terão distribuições marginais elípticas, mas não serão independentes. Na segunda extensão assume-se que estes vetores são independentes com distribuições marginais elípticas. Para ambos os modelos, tem-se que x_i , e_i e u_i podem estar não correlacionados sem serem independentes e os vetores observados $Z_i = (Y_i, X_i)^T$, $i = 1, 2, \dots, n$, terão distribuições marginais elípticas sendo dependentes no primeiro modelo e independentes no segundo. Tais modelos serão denominados de modelo estrutural elíptico dependente e independente, respectivamente, e são ilustrados usando a distribuição t nas seções finais deste capítulo (veja, também, as seções 4.2 e 4.3 do Capítulo 4). Mostra-se, em particular, que o modelo t dependente produz um EMV de β análogo ao do modelo normal e, portanto, não robusto, enquanto que o modelo t independente proporciona um EMV robusto para β . O interesse de robustificar a estimação de β tem sido explorada em diversos trabalhos tais como Abdullah (1989), Brown (1982), Carroll e Gallo (1982), Kelly (1984) e Loung (1991). Como resultado adicional, neste último modelo, derivam-se os fatores de (tipo) Bartlett dos testes de razão de verossimilhança e escore da hipótese $H_0 : \beta = \beta_0$.

5.2 Modelos estruturais elípticos

5.2.1 Introdução

Os modelos elípticos, dos quais o modelo normal é um membro particular, têm aparecido com muita frequência na literatura estatística mais recente. Por exemplo, no contexto dos modelos lineares, vários autores têm usado estes modelos para explorar as consequências de ter erros não-normais ou com distribuição tendo caudas mais pesadas (ou mais leves) que a distribuição normal. Contudo, na literatura aparecem duas formas de especificar um modelo de regressão elíptico, as quais foram ilustradas no Capítulo 4 (Seções 4.2 e 4.3) usando a distribuição t . Como no referido capítulo, tais modelos podem ser denominados:

- (i) *Modelo elíptico dependente*: assume que os erros (ou as observações) têm distribuição conjunta elíptica, de modo que eles podem ter covariância nula sem serem independentes (veja, por exemplo, Zellner (1976); Ghosh e Sinha (1980); Ullah e Walsh (1984); Chib, Tiwari e Jammaladaka (1988) e Anderson, Fang e Hsu (1986));
- (ii) *Modelo elíptico independente*: supõe que os erros (ou as observações) são independentes e têm distribuições marginais elípticas (veja, por exemplo, Lange, Little e Taylor (1989); Taylor (1992) e Kano, Berkane e Bentler (1993)).

Note-se que em (ii) a distribuição conjunta é o produto de distribuições elípticas, mas ela não é elíptica. Convém lembrar aqui que supor que a distribuição conjunta é elíptica juntamente com independência equivale a assumir normalidade (veja o Fato (b) na Seção 2.6).

Os autores que seguem o enfoque (i) mostram que muitas das propriedades estatísticas do modelo normal permanecem válidas nesta classe geral de modelos elípticos dependentes (veja os exemplos 2.11 – 2.13, no Capítulo 2). Isto fornece certa robustez de alguns procedimentos usados (como os mínimos quadrados; veja Ghosh e Sinha (1988) e Ullah e Walsh (1984)) com respeito a este tipo de não-normalidade. Resultados gerais sobre a

existência e unicidade dos EMV dos parâmetros de locação (μ) e escala ($\Sigma > 0$) de um modelo elíptico dependente podem ser encontrados em Anderson, Fang e Hsu (1986). Em particular, tais autores mostram que estes EMV coincidem (a menos de uma constante positiva no caso de Σ) com os respectivos EMV do modelo normal. Uma consequência direta deste fato é que tais estimadores serão altamente sensíveis (não-robustos) com respeito a observações extremas. Esta limitação pode ser superada escolhendo-se uma família elíptica apropriada, como a família t , e especificando o modelo de acordo com o enfoque (ii) (veja Lange, Little e Taylor (1989)). Neste caso tem-se que, em geral, os EMV não têm uma forma fechada, de modo que devem ser avaliados com algoritmos de solução numérica. Além disso, neste modelo as inferências devem basear-se na teoria assintótica. Resultados gerais sobre este último aspecto podem ser encontrados em Kano, Berkane e Bentler (1993).

Nesta seção são dadas apenas algumas idéias gerais sobre a especificação dos modelo estrutural elíptico seguindo os enfoques (i) e (ii). Em particular, mostra-se que em ambos os modelos prevalecem tanto o problema de identificabilidade quanto a parametrização ortogonal do modelo normal. Considerando esta parametrização ortogonal, estes modelos são desenvolvidos detalhadamente no contexto da distribuição t , nas Seções 5.3 e 5.4.

5.2.2 Especificação de um modelo estrutural elíptico

Sejam

$$Z_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \quad r_i = \begin{pmatrix} x_i \\ e_i \\ u_i \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \quad B = [\beta, I_2] \quad \text{e} \quad \beta = \begin{pmatrix} \beta \\ 1 \end{pmatrix}$$

e note-se que a relação em (5.9) pode ser reescrita como

$$Z_i = \alpha + Br_i, \quad i = 1, 2, \dots, n. \quad (5.48)$$

Com esta relação linear não é difícil obter a distribuição de Z_i , $i = 1, 2, \dots, n$, uma vez postulada a distribuição de r_i , $i = 1, 2, \dots, n$. Assim, se assume-se que $r_1, r_2, \dots, r_n$ são

não-correlacionados e

$$r_i \sim El_3(\eta_i, \Omega_i; \phi) ,$$

então $Z_1, Z_2, \dots, Z_n$ serão não-correlacionados e (veja (2.47))

$$Z_i \sim El_2(\mu_i, \Sigma_i; \phi) ,$$

onde $\mu_i = \alpha + \beta\eta_i$ e $\Sigma_i = B\Omega_i B^\top$. Contudo, estes resultados são coerentes com duas especificações diferentes da distribuição conjunta de $r_1, r_2, \dots, r_n$, isto é, supondo que:

$$(i) \mathbf{r} \sim El_{3n}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi), \text{ ou}$$

$$(ii) \mathbf{r} \sim El_{3n}^{(I)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \phi) \Leftrightarrow r_i \overset{\text{ind.}}{\sim} El_3(\eta_i, \Omega_i; \phi), i = 1, 2, \dots, n, \text{ onde}$$

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Omega} = \begin{pmatrix} \Omega_1 & & 0 \\ & \Omega_2 & \\ 0 & & \Omega_n \end{pmatrix}.$$

Para ambos os casos tem-se que (desde que existem; veja a Seção 2.5.1)

$$E(\mathbf{r}) = \boldsymbol{\eta} \quad \text{e} \quad \text{Cov}(\mathbf{r}) = \alpha_\phi \boldsymbol{\Omega} ,$$

isto é,

$$E(r_i) = \eta_i, \quad \text{Cov}(r_i) = \alpha_\phi \Omega_i \quad \text{e} \quad \text{Cov}(r_i, r_j) = 0, \quad i \neq j ,$$

onde $\alpha_\phi = -2\phi'(0)$. A diferença entre (i) e (ii) é que em (i) os vetores $r_1, r_2, \dots, r_n$ não são independentes, a menos que se tenha normalidade. Em tal caso (i) e (ii) são equivalentes. Similarmente, excetuando o caso normal, para ambos os modelos a escolha de Ω_i como sendo diagonal não implica independência entre as componentes de $r_i = (x_i, e_i, u_i)^\top$. Aplicando (2.47) em (5.48), para (i) e (ii) segue que:

$$(i) \mathbf{Z} \sim El_{2n}(\mathbf{u}, \boldsymbol{\Omega}; \phi),$$

$$(ii) \mathbf{Z} \sim El_{2n}^{(I)}(\mathbf{u}, \Sigma; \phi) \Leftrightarrow Z_i \stackrel{\text{ind}}{\sim} El_2(u_i, \Sigma_i; \phi), i = 1, 2, \dots, n,$$

onde

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_1 & & & 0 \\ & \Sigma_2 & & \\ & & \ddots & \\ 0 & & & \Sigma_n \end{pmatrix},$$

com $u_i = \alpha + B\eta_i$ e $\Sigma_i = B\Omega_i B^\top$. Similarmente (e desde que existam),

$$E(\mathbf{Z}) = \mathbf{u} \quad \text{e} \quad \text{Cov}(\mathbf{Z}) = \alpha_\phi \Sigma$$

ou

$$E(Z_i) = u_i, \quad \text{Cov}(Z_i) = \alpha_\phi \Sigma_i \quad \text{e} \quad \text{Cov}(Z_i, Z_j) = 0, \quad i \neq j.$$

Se consideramos somente as distribuições normais compostas (veja os Exemplos 2.6 e 2.10), então (i) e (ii) podem ser formulados como:

$$(i) (r_i|V = v) \stackrel{\text{ind}}{\sim} N_3(\eta_i, v\Omega_i) \Rightarrow (Z_i|V = v) \stackrel{\text{ind}}{\sim} N_2(u_i, v\Sigma_i), i = 1, 2, \dots, n, \text{ onde } V \sim F_V \text{ (sobre } [0, \infty)),$$

$$(ii) (r_i|V_i = v_i) \stackrel{\text{ind}}{\sim} N_3(\eta_i, v_i\Omega_i) \Rightarrow (Z_i|V_i = v_i) \stackrel{\text{ind}}{\sim} N_2(u_i, v_i\Sigma_i), i = 1, 2, \dots, n, \text{ onde } V_1, V_2, \dots, V_n \text{ são iid } F_V \text{ (sobre } [0, \infty)).$$

Por outro lado, se $\eta_i = \eta$ e $\Omega_i = \Omega$ ($\Omega = I_n \otimes \Omega$), $i = 1, 2, \dots, n$, tem-se que $\mu_i = \mu = \alpha + B\eta$ e $\Sigma_i = \Sigma = B\Omega B^\top$ ($\Sigma = I_n \otimes \Sigma$), $i = 1, 2, \dots, n$. Então, especificando η e Ω (a locação e dispersão de $r_i = (x_i, e_i, u_i)^\top$) como

$$\eta = \begin{pmatrix} u_x \\ 0 \\ 0 \end{pmatrix} \quad \text{e} \quad \Omega = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{ee} & 0 \\ 0 & 0 & \sigma_{uu} \end{pmatrix},$$

para μ e Σ (a locação e dispersão de $Z_i = (Y_i, X_i)^\top$) tem-se que (veja (5.10) e (5.11))

$$\mu = \mu(\theta) = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \Sigma = \Sigma(\theta) = \begin{pmatrix} \beta^2\sigma_{xx} + \sigma_{ee} & \beta\sigma_{xx} \\ \beta\sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{pmatrix},$$

onde (veja (5.14))

$$\theta = (\alpha, u_x, \sigma_{xx}, \sigma_{ee}, \sigma_{uu}, \beta)^\top .$$

Desta forma o modelo estrutural normal estende-se para um modelo estrutural elíptico. Assim, sob a hipótese adicional de existência de densidade, este último modelo pode ser definido em termos da densidade de $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$, digamos,

$$p_{\mathbf{Z}}(\mathbf{z}|\theta) = p_{z_1, z_2, \dots, z_n}(z_1, z_2, \dots, z_n|\theta) ,$$

como segue:

(i) o modelo estrutural elíptico dependente:

$$p_{\mathbf{Z}}(\mathbf{z}|\theta) = |\Sigma|^{n/2} f \left(\sum_{i=1}^n (z_i - \mu)^\top \Sigma^{-1} (z_i - \mu) \right) , \quad (5.49)$$

e será denotado por

$$\mathbf{Z} \sim El_{2n}(\mu, I_n \otimes \Sigma; f) ,$$

onde $f(\cdot)$ é uma densidade esférica em $\mathbb{R}^{2n}$ (a densidade conjunta de $T_i = \Sigma^{-1/2}(Z_i - \mu)$, $i = 1, 2, \dots, n$),

(ii) o modelo estrutural elíptico independente:

$$p_{\mathbf{Z}}(\mathbf{z}|\theta) = |\Sigma|^{-n/2} \prod_{i=1}^n g((z_i - \mu)^\top \Sigma^{-1} (z_i - \mu)) , \quad (5.50)$$

e será denotado por

$$\mathbf{Z} = El_{2n}^{(I)}(\mu, I_n \otimes \Sigma; g) ,$$

onde $g(\cdot)$ é uma densidade esférica em $\mathbb{R}^2$ (a densidade marginal de $T_i = \Sigma^{-1/2}(Z_i - \mu)$).

Convém notar aqui que tanto f em (5.49) como g em (5.50) não dependem de $(\mu, \Sigma) = (\mu(\theta), \Sigma(\theta))$.

5.2.3 Identificabilidade e estimação consistente dos parâmetros estruturais

Em primeiro lugar, note-se que tanto em (5.49) quanto em (5.50) tem-se que

$$p_{\mathbf{Z}}(\mathbf{z}|\theta) = p_{\mathbf{Z}}(\mathbf{z}|\mu(\theta); \Sigma(\theta)) ,$$

de modo que a distribuição de $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$ depende de θ somente através de $(\mu, \Sigma) = (\mu(\theta), \Sigma(\theta))$. Então, dado f em (5.49) ou g em (5.50), a distribuição conjunta de $Z_1, Z_2, \dots, Z_n$ é completamente caracterizada pela locação e dispersão de Z_i , isto é, por

$$\mu = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \sigma_{YY} & \sigma_{YX} \\ \sigma_{YX} & \sigma_{XX} \end{pmatrix} ,$$

com um total de cinco parâmetros. Como $(\mu, \Sigma) = (\mu(\theta), \Sigma(\theta))$, onde $\theta = (\alpha, \mu_x, \sigma_{xx}, \sigma_{ee}, \sigma_{uu}, \beta)^\top$ e envolve um total de seis parâmetros, então existem diferentes valores de θ sobre os quais $Z_1, Z_2, \dots, Z_n$ terão a mesma locação $\mu = \mu(\theta)$ e dispersão $\Sigma = \Sigma(\theta)$. Em outras palavras, existem θ_1 e θ_2 tais que

$$\begin{aligned} \theta_1 \neq \theta_2 &\Rightarrow (\mu(\theta_1), \Sigma(\theta_1)) = (\mu(\theta_2), \Sigma(\theta_2)) \\ &\Rightarrow p_{\mathbf{Z}}(\mathbf{z}|\mu(\theta_1), \Sigma(\theta_1)) = p_{\mathbf{Z}}(\mathbf{z}|\mu(\theta_2), \Sigma(\theta_2)) \\ &\Rightarrow p_{\mathbf{Z}}(\mathbf{z}|\theta_1) = p_{\mathbf{Z}}(\mathbf{z}|\theta_2) , \end{aligned}$$

para todas as configurações de observações $\mathbf{z} = (z_1^\top, z_2^\top, \dots, z_n^\top)^\top \in \mathbb{R}^{(3n)}$. Por exemplo, (Fuller (1987, p.10)), para $\theta_1 = (1, 1, 1, 1, 1, 1)^\top$ e $\theta_2 = (3/2, 1, 2, 3/2, 0, 1/2)^\top$ tem-se que $\theta_1 \neq \theta_2$ e

$$\mu(\theta_1) = \mu(\theta_2) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} , \quad \Sigma(\theta_1) = \Sigma(\theta_2) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

Isto mostra que todo modelo estrutural elípticos definido como em (5.49) ou (5.50), em particular, o modelo estrutural normal, é não identificável.

Contudo, como no modelo normal, para $\lambda_x = \sigma_{xx}/\sigma_{uu}$ ou $\lambda_e = \sigma_{ee}/\sigma_{uu}$ conhecido, ambos os modelos tornam-se identificáveis, uma vez que θ se reduz a (5.17) com um total

de cinco parâmetros, ou seja, a relação

$$\theta \rightarrow (\mu, \Sigma) = (\mu(\theta), \Sigma(\theta)) ,$$

é agora um a um. Em tal caso, θ pode ser estimado consistentemente usando o procedimento de máxima verossimilhança. Por outro lado, se $E(Z_i Z_j^\top) < \infty$ (ou $E(r_i r_j^\top) < \infty$), para ambos os modelos tem-se que $E(Z_i) = \mu(\theta)$ e $\text{Cov}(Z_i) = \alpha_g \Sigma(\theta)$, onde α_g ($\alpha_g = \alpha_\phi = -2\phi'(0)$) não depende de θ . Sejam

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix} \quad \text{e} \quad S_{ZZ} = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^\top = \begin{pmatrix} S_{YY} & S_{YX} \\ S_{YX} & S_{XX} \end{pmatrix} ,$$

então, pela lei fraca dos grandes números, no modelo estrutural elíptico independente tem-se que

$$\text{plim} \bar{Z} = \mu(\theta) \quad \text{e} \quad \text{plim} S_{ZZ} = \alpha_g \Sigma(\theta) .$$

Contudo, o primeiro destes resultados também vale no modelo estrutural elíptico dependente. De fato, como $\text{Cov}(Z_i, Z_j) = 0$, $i \neq j$, então $\text{Cov}(\bar{Z}) = \alpha_g \Sigma/n$. Logo, da desigualdade de Chebychev segue que, para todo $\varepsilon > 0$,

$$P(\|\bar{Z} - \mu\|^2 \geq \varepsilon) \leq \frac{\text{tr}\{\text{Cov}(\bar{Z})\}}{\varepsilon^2} = \frac{\alpha_g \text{tr}(\Sigma)}{n\varepsilon^2} \rightarrow 0 ,$$

quando $n \rightarrow \infty$. O segundo resultado não é garantido neste último modelo, uma vez que $\text{plim} S_{ZZ}$ pode ser aleatório. Isto sucede, por exemplo, no modelo normal composto:

$$\begin{cases} (Z_i | V = v) \stackrel{\text{iid}}{\sim} N_2(\mu(\theta), v\Sigma(\theta)), & i = 1, 2, \dots, n, \\ V \sim F_V , \end{cases}$$

onde é fácil ver que

$$\text{plim} \bar{Z} = \mu(\theta) \quad \text{e} \quad \text{plim} S_{ZZ} = V\Sigma(\theta) ,$$

sem a condição de que $E(Z_i Z_j^\top) < \infty$. De qualquer modo, se em ambos os modelos $\text{plim} S_{ZZ}$ é proporcional a $\Sigma(\theta)$, então um estimador consistente de β é dado por seu

estimador de “momentos”, digamos

$$\hat{\beta}_M = \begin{cases} \hat{\beta} = \left(\frac{\lambda_x + 1}{\lambda_x} \right) \frac{S_{YX}}{S_{XX}}, & \text{se } \lambda_x \text{ é conhecido,} \\ \hat{\beta}_S = \frac{S_{YY} - \lambda_e S_{XX} + \{(S_{YY} - \lambda_e S_{XX})^2 + 4\lambda_e S_{YX}^2\}^{\frac{1}{2}}}{2S_{YX}}, & \text{se } \lambda_e \text{ é conhecido.} \end{cases}$$

Além disto, $\hat{\beta}_M$ pode ser o EMV de β no modelo estrutural elíptico dependente. Este resultado despreende-se do Teorema 1 dado em Anderson, Fang e Hsu (1986). Nesse teorema os autores estabelecem que num modelo elíptico como (5.49), os EMV de μ e Σ são dados por

$$\hat{\mu} = \hat{\mu}_N \quad \text{e} \quad \hat{\Sigma} = \frac{np}{u_f} \hat{\Sigma}_N, \quad (5.51)$$

onde $\hat{\mu}_N$ e $\hat{\Sigma}_N$ são os respectivos EMV sob normalidade, $p = 2$ é a dimensão de Z_i ($p < n$, de modo que $\hat{\Sigma}_N > 0$ com probabilidade um) e u_f (assumido positivo e finito) é o máximo de $u^{np/2} f(u)$. Por exemplo, se $f(u) = k(p, \nu) \{\nu + u\}^{-(\nu+p)/2}$ (o modelo t com ν graus de liberdade) é fácil ver que $u_f = np$ para todo $\nu > 0$. Em geral, se $f(u)$, $u \geq 0$, é uma função contínua e decrescente para u suficientemente grande, então o máximo u_f de $u^{np/2} f(u)$ existe e é finito. Este fato vale em particular quando a densidade $f(z^\top z)$, $z \in \mathbb{R}^{np}$, tem segundos momentos finitos (veja Anderson et al. (1986)).

Note-se que o resultado de Anderson et al. (1986) mencionado acima tem implicações adicionais de grande importância na teoria de inferência baseada num modelo elíptico. De fato, se $X \sim El_p(0, I_p; \phi)$, $P(X = 0) = 0$ e $\delta(X)$ é uma estatística tal que $\delta(aX) =^d \delta(X)$ para cada $a > 0$, então tem-se que (veja Fang et al. (1990, p.51))

$$\delta(X) =^d \delta(Z), \quad Z \sim N_p(0, I_p).$$

Esta propriedade segue do fato de que (veja o Capítulo 2)

$$\frac{X}{\|X\|} =^d \frac{Z}{\|Z\|} = U^{(P)} \sim U(S_p).$$

No modelo estrutural elíptico dependente $\mathbf{Z} \sim El_{2n}(\boldsymbol{\mu}, I_n \otimes \Sigma; f)$, algumas estatísticas que verificam esta propriedade de robustez (numérica e em distribuição) são, por exemplo:

$$(1) F = \left(\frac{n-2}{2}\right)(\bar{Z} - \mu)^\top S_{ZZ}^{-1}(\bar{Z} - \mu) \sim F_{2,n-2};$$

$$(2) t = \sqrt{n}(a^\top \bar{Z} - a^\top \mu)/(a^\top S_{ZZ} a)^{1/2} \sim t(0, 1; n-1), a \in \mathbb{R}^2;$$

(3) O EMV de β . Em particular, para λ_x conhecido tem-se que

$$\hat{\beta} = \hat{\beta}_{MQ} = \left(\frac{\lambda_x + 1}{\lambda_x}\right) \frac{S_{YX}}{S_{XX}} \sim t\left(\beta, \frac{\sigma_\beta^2}{n-1}; n-1\right),$$

onde $\sigma_\beta^2 = \phi_3/\lambda_x^2 \phi_4$ (veja (5.33)).

(4) O coeficiente de correlação amostral $r_{YX} = S_{YX}/(S_{YY}S_{XX})^{1/2}$.

A prova direta destes resultados é simples para aqueles modelos elípticos dependentes que são mistura de escala do modelo normal. De fato, para estes modelos tem-se que

$$(Z_i|V=v) \stackrel{\text{iid}}{\sim} N_2(\mu, v\Sigma), \quad i = 1, 2, \dots, n,$$

de onde segue que

$$(\bar{Z}|V=v) \sim N_2\left(\mu, \frac{v}{n}\Sigma\right) \quad \text{e} \quad (S_{ZZ}|V=v) \sim W_2\left(\frac{v}{n}\Sigma, n-1\right)$$

são independentes. Assim, a prova decorre de observar que para cada uma das estatísticas dadas acima tem-se que suas distribuições condicionais dado $V=v$ não dependem de v .

Note-se que em (3) e em (4) a distribuição das estatísticas envolvidas depende somente da distribuição da matriz S_{ZZ} . A distribuição desta matriz está derivada em Sutradhar e Ali (1989) e constitui uma generalização da matriz Wishart. A partir deste resultado, Ali e Joarder (1994) derivam a distribuição de r_{YX} para o caso geral de um modelo elíptico dependente.

5.2.4 Matriz de informação e parametrização ortogonal

Seja $L = L(\theta)$ a função de log-verossimilhança do modelo estrutural elíptico (5.49) ou (5.50). Então,

$$L = \begin{cases} -\frac{n}{2} \log |\Sigma| + \log f \left(\sum_{i=1}^n \|T_k\|^2 \right), & \text{no caso (i)} \\ -\frac{n}{2} \log |\Sigma| + \sum_{k=1}^n \log g(\|T_k\|^2), & \text{no caso (ii)}, \end{cases}$$

onde $T_k = \Sigma^{-1/2}(Z_k - \mu)$ e $\sum_{k=1}^n \|T_k\|^2 = \|\mathbf{T}\|^2$, onde $\mathbf{T} = (T_1^\top, T_2^\top, \dots, T_n^\top)^\top$, de modo que:

(i) no modelo elíptico dependente:

$$\mathbf{T} \sim El_{2n}(0, I_{2n}; f);$$

(ii) no modelo elíptico independente:

$$\mathbf{T} \sim El_{2n}^{(I)}(0, I_{2n}; g) \Leftrightarrow T_i \stackrel{\text{iid}}{\sim} El_2(0, I_2; g).$$

Note-se que $\mathbf{T}$ pode ser escrito como

$$\mathbf{T} = \Sigma^{-1/2}(\mathbf{Z} - \boldsymbol{\mu}),$$

onde $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$, $\boldsymbol{\mu} = (\mu^\top, \mu^\top, \dots, \mu^\top)^\top$ e $\Sigma = I_n \otimes \Sigma = \text{Diag}\{\Sigma, \Sigma, \dots, \Sigma\}$.

Usando esta notação tem-se

$$\frac{\partial L}{\partial \theta_i} = \begin{cases} -\frac{n}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \right) + W_f \left(\sum_{k=1}^n \|T_k\|^2 \right) \sum_{k=1}^n \frac{\partial \|T_k\|^2}{\partial \theta_i}, & \text{no caso (i)}, \\ -\frac{n}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \right) + \sum_{k=1}^n W_g(\|T_k\|^2) \frac{\partial \|T_k\|^2}{\partial \theta_i}, & \text{no caso (ii)}, \end{cases}$$

onde (veja (2.83) e (5.21))

$$\begin{aligned} W_h(u) &= \frac{\partial \log k(u)}{\partial u} = \frac{h'(u)}{h(u)}, \quad k = g, f, \\ \frac{\partial \|T_k\|^2}{\partial \theta_i} &= -2 \frac{\partial \mu^\top}{\partial \theta_i} \Sigma^{-1/2} T_k - T_k^\top \Sigma^{-1/2} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1/2} T_k \end{aligned}$$

e

$$\begin{aligned}\sum_{k=1}^n \frac{\|T_k\|^2}{\partial\theta_i} &= \frac{\partial\|\mathbf{T}\|^2}{\partial\theta_i} \\ &= -2\frac{\partial\boldsymbol{\mu}^\top}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_j} - \mathbf{T}^\top\boldsymbol{\Sigma}^{-1/2}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1/2}\mathbf{T}.\end{aligned}$$

Seja

$$K = K(\theta) = [\kappa_{i,j}] = \left[E \left(\frac{\partial L}{\partial\theta_i} \frac{\partial L}{\partial\theta_j} \right) \right]$$

a matriz de informação do modelo elíptico, então, aplicando os resultados anteriores e usando a simetria das distribuições de $\mathbf{T}$ e T_k juntamente com (2.84) e (2.85), segue (após uma certa álgebra) que:

(i) para o modelo elíptico dependente:

$$\begin{aligned}\kappa_{i,j} &= \frac{4n}{p}a_f(1,2)\frac{\partial\boldsymbol{\mu}^\top}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_j} \\ &+ n\left(\frac{n}{4} + \frac{n}{p}a_f(1,1) + \frac{1}{p(p+2)}a_f(2,2)\right)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\right)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right) \\ &+ \frac{2n}{p(p+2)}a_f(2,2)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right),\end{aligned}\quad (5.52)$$

onde $p = 2n$ e denota a dimensão de $\mathbf{T}$,

(ii) para o modelo elíptico independente:

$$\begin{aligned}\kappa_{i,j} &= \frac{4n}{p}a_g(1,2)\frac{\partial\boldsymbol{\mu}^\top}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_j} \\ &+ n\left(\frac{1}{4} + \frac{1}{p}a_g(1,1) + \frac{1}{p(p+2)}a_g(2,2)\right)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\right)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right) \\ &+ \frac{2n}{p(p+2)}a_g(2,2)\text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right),\end{aligned}\quad (5.53)$$

onde $p = 2$ e denota a dimensão de T_k .

Para ambos os modelos, os coeficientes $a_k(i,j)$ são definidos por

$$a_k(i,j) = \begin{cases} E \left\{ \left(\sum_{k=1}^n \|T_k\|^2 \right)^i \left(W_f \left(\sum_{k=1}^n \|T_k\|^2 \right) \right)^j \right\}, & \text{no caso (i) } (h = f), \\ E \left\{ \|T_k\|^{2i} (W_g(\|T_k\|^2))^j \right\}, & \text{no caso (i) } (h = g), \end{cases}\quad (5.54)$$

de modo que não dependem de $(\mu(\theta), \Sigma(\theta))$. Em particular, para o modelo normal tem-se que

$$W_h(u) = -\frac{1}{2}, \quad h = f \text{ ou } g; \quad \|\mathbf{T}\|^2 = \sum_{k=1}^n \|T_k\|^2 \sim \chi_{2n}^2 \quad \text{e} \quad \|T_k\|^2 \sim \chi_2^2,$$

de modo que

$$a_h(i, j) = \left(-\frac{1}{2}\right)^j \mu_i(\chi_p^2),$$

onde $\mu_k(\chi_p^2) = E(\chi_p^{2k}) = p(p+2) \cdots (p+2(k-1))$ é o k -ésimo momento da distribuição χ_p^2 . É fácil ver que neste caso tanto (5.52) quanto (5.53) se reduzem a (5.23) – (5.24).

Convém notar que se θ_i é tal que $\text{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i}) \neq 0$, então tanto em (5.52) quanto em (5.53) tem-se que (veja o Apêndice A.1)

$$E \left\{ \frac{\partial L}{\partial \theta_i} \right\} = 0 \Rightarrow a_h(1, 1) = -\frac{p}{2}. \quad (5.55)$$

De (5.51) e (5.53) é claro que $\kappa_{i,j} = 0$ quando θ_i é parâmetro de locação e θ_j de escala, de modo que para ambos os modelos elípticos tem-se que $\mu(\theta)$ e $\Sigma(\theta)$ são (globalmente) ortogonais. Consequentemente, como na parametrização $\phi = \phi(\theta) = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)^\top$ definida em (5.26) tem-se que $\mu(\phi) = \mu(\phi_L)$ e $\Sigma(\phi) = \Sigma(\phi_E)$, então a ortogonalidade de $\phi_L = (\phi_1, \phi_2)^\top$ com $\phi_E = (\phi_3, \phi_4, \beta)^\top$ sob o modelo normal é preservada nos modelos elípticos. Para mostrar que também prevalece a ortogonalidade de β com $(\phi_3, \phi_4)^\top$ é suficiente lembrar que o fato de que $|\Sigma| = \phi_3 \phi_4$ implica que

$$\text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right) = \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \right) = 0, \quad i = 3, 4.$$

Contudo, nestes modelos elípticos a ortogonalidade de ϕ_3 com ϕ_4 não é necessariamente conservada. Assim, se $K = K(\phi)$ denota (com abuso da notação) a matriz de informação sob a parametrização ortogonal, então,

$$K = \text{Diag} \{K_L, K_E\},$$

onde K_L e K_E são as submatrizes relativas a $\phi_L = (\phi_1, \phi_2)^\top$ e $\phi_E = (\phi_3, \phi_4, \beta)^\top$, respectivamente. Substituindo θ por ϕ em (5.52) e (5.53) e considerando (5.54), para K_L tem-se

que

$$K_L = \frac{4a_h(1, 2)}{p} \left(\frac{1}{n} \Sigma \right)^{-1}. \quad (5.56)$$

Procedendo da mesma forma e usando (5.32), (5.34) e (5.55), tem-se que $K_E = [\kappa_{i,j}]$ é dada por

$$\kappa_{i,j} = \begin{cases} b_{ij} \left(\frac{2\phi_i\phi_j}{n} \right)^{-1}, & i, j = 3, 4, \\ \frac{4a_k(2, 2)}{p(p+2)} \left(\frac{\sigma_\beta^2}{n} \right)^{-1}, & i, j = \beta, \\ 0, & i = 3, 4, j = \beta, \end{cases} \quad (5.57)$$

onde $\sigma_\beta^2 = \sigma_\beta^2(\phi_E)$ está definida em (5.33),

$$b_{ii} = \begin{cases} \frac{(n+2)a_f(2, 2)}{p(p+2)} - \frac{n}{4}, & \text{no caso (i), } (p = 2n), \\ \frac{3a_g(2, 2)}{p(p+2)} - \frac{1}{4}, & \text{no caso (ii), } (p = 2), \end{cases}$$

para $i = 3, 4$ e

$$b_{34} = \begin{cases} \frac{na_f(2, 2)}{p(p+2)} - \frac{n}{4}, & \text{no caso (i),} \\ \frac{a_g(2, 2)}{p(p+2)} - \frac{1}{4}, & \text{no caso (ii).} \end{cases}$$

Note-se que $\kappa_{3,4} = 0$ para o modelo normal.

Resultados adicionais sobre a matriz de informação de um modelo elíptico p -variado, $El_p(\mu, \Sigma; h)$, são dados em Mitchell (1989) considerando o vetor de parâmetros de dimensão $p + p(p+1)/2$ dado por

$$\theta = (\mu_1, \dots, \mu_p, \sigma_{11}, \sigma_{12}, \dots, \sigma_{1p}, \sigma_{21}, \dots, \sigma_{pp})^\top.$$

Nas próximas seções, estuda-se os modelos estruturais- t dependente (Seção 5.3) e independente (Seção 5.4) supondo os coeficientes de λ_x e λ_e conhecidos. Em vista dos resultados desenvolvidos acima para o contexto geral das distribuições elípticas, estes modelos são formulados diretamente para os vetores observados $Z_1, Z_2, \dots, Z_n$ e são estudados apenas sob a parametrização ortogonal de (μ, Σ) .

5.3 Inferência no modelo estrutural- t dependente

Seja $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$, onde $Z_i = (Y_i, X_i)^\top$. Então, segundo Zellner (1976), o modelo estrutural normal pode estender-se para o modelo estrutural- t dependente supondo que

$$\mathbf{Z} \sim t_{2n}(\boldsymbol{\mu}, I_n \otimes \Sigma; \nu) . \quad (5.58)$$

Isto é, supondo que $\mathbf{Z}$ tem densidade dada por

$$p_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\mu}, \Sigma) = |\Sigma|^{-\frac{n}{2}} k(2n, \nu) \nu^{\frac{n}{2}} \left\{ \nu + \sum_{i=1}^n (Z_i - \boldsymbol{\mu})^\top \Sigma^{-1} (Z_i - \boldsymbol{\mu}) \right\}^{-\frac{2n+\nu}{2}} ,$$

onde $k(2n, \nu) = \Gamma[(2n + \nu)/2] / \Gamma(\nu/2) \pi^n$. É fácil ver (veja o Capítulo 3) que para este modelo tem-se que

$$Z_i \sim t_2(\boldsymbol{\mu}, \Sigma; \nu) , \quad i = 1, 2, \dots, n,$$

com $\text{Cov}(Z_i, Z_j) = 0$, $i \neq j$ e $\nu > 2$, mas $Z_1, Z_2, \dots, Z_n$ não são independentes. Note-se também que

$$E(Z_i) = \boldsymbol{\mu} , \quad \nu > 1 ; \quad \text{Cov}(Z_i) = \frac{\nu}{\nu - 2} \Sigma , \quad \nu > 2 .$$

Uma representação alternativa para (5.57) vem dada por (veja o Capítulo 3)

$$(\mathbf{Z}|\mathbf{V}) \sim N_{2n}(\mathbf{u}, \mathbf{V} I_n \otimes \Sigma) , \quad (5.59)$$

onde $\mathbf{V} \sim \nu/\chi_\nu^2$. Com estas especificações, as propriedades de (5.57) podem ser estudadas usando as respectivas propriedades do modelo normal. Note-se também que (5.59) implica que (veja (3.24))

$$(V|\mathbf{Z}) \sim (V + \sum_{i=1}^n (Z_i - \boldsymbol{\mu})^\top \Sigma^{-1} (Z_i - \boldsymbol{\mu})) / \chi_{\nu+2n}^2 ,$$

de onde segue, por exemplo, que (veja (3.25))

$$E(V^{-1}|\mathbf{Z}) = \frac{\nu + 2n}{\nu + \sum_{i=1}^n (Z_i - \boldsymbol{\mu})^\top \Sigma^{-1} (Z_i - \boldsymbol{\mu})} .$$

No que segue assume-se que $(\boldsymbol{\mu}, \Sigma) = (\boldsymbol{\mu}(\boldsymbol{\phi}_L), \Sigma(\boldsymbol{\phi}_E))$, onde $\boldsymbol{\phi} = (\boldsymbol{\phi}_L^\top, \boldsymbol{\phi}_E^\top)^\top$ é a parametrização ortogonal definida em (5.26).

5.3.1 Matriz de informação e EMV da parametrização ortogonal

Considerando o modelo estrutural t -dependente definido em (5.57), nesta seção são derivadas a matriz de informação $K = K(\phi)$ e o EMV da parametrização ortogonal ϕ .

a) Derivação da matriz de informação

Como $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, I_n \otimes \Sigma; \nu) = El_p(\boldsymbol{\mu}, I_n \otimes \Sigma; h)$ com $p = 2n$ (veja o Capítulo 3) e $h(u) = k(p, \nu) \nu^{\nu/2} \{\nu + u\}^{-(\nu+p)/2}$, $u \geq 0$, então a matriz de informação deste modelo pode ser derivada a partir de (5.56) e (5.57). Assim, para derivar os coeficientes $a_k(i, j)$ definidos em (5.54), note-se em primeiro lugar que

$$W_h(u) = \frac{\partial \log h(u)}{\partial u} = -\frac{1}{2} \frac{\nu + p}{\nu + u},$$

de modo que $a_h(i, j) = (-\frac{\nu+p}{2})^j E\{\|\epsilon\|^{2i} (\nu + \|\epsilon\|^2)^{-j}\}$, onde $\epsilon \sim t_p(0, I_p; \nu)$, e estão dados no Exemplo 2.14(ii) (veja também (3.27)). Destes resultados segue que para a distribuição t os coeficientes $a_h(i, j) = a_\nu(i, j)$ são da forma

$$a_\nu(i, j) = \left(-\frac{1}{2}\right)^j \mu_i(\chi_p^2) \left(\frac{\nu + p}{\nu}\right)^{j-i} \frac{(\nu + p)^i \mu_{j-i}(\chi_\nu^2)}{\mu_j(\chi_{\nu+p}^2)}, \quad j \geq i, \quad (5.60)$$

$p \geq 2$ ($\mu_i(\chi_p^2) \equiv 1, p = 1$), onde $\mu_0(\cdot) = 1$ e $\mu_k(\chi_s^2) = E(\chi_s^{2k}) = s(s+2) \cdots (s+2(k-1))$ é o k -ésimo momento da distribuição χ_s^2 . De (5.60) é fácil ver que $a_\nu(i, j) \rightarrow a_\infty(i, j)$, para todo $j \geq i$, quando $\nu \rightarrow \infty$, onde $a_\infty(i, j) = (-1/2)^j \mu_i(\chi_p^2)$ e é o respectivo coeficiente do modelo normal. Em particular, para $i, j = 1, 2, j \geq i$, tem-se que

$$a_\nu(1, 1) = -\frac{p}{2}, \quad a_\nu(1, 2) = \frac{p}{4} \frac{\nu + p}{\nu + p + 2}, \quad a_\nu(2, 2) = \frac{p(p+2)}{4} \frac{\nu + p}{\nu + p + 2},$$

de modo que na matriz de informação $K(\phi_L, \phi_E) = \text{Diag}\{K_L, K_E\}$ do modelo estrutural t -dependente tem-se que (veja (5.56) e (5.57), com $p = 2n$)

$$K_L = \frac{\nu + 2n}{\nu + 2n + 2} \left(\frac{1}{n} \Sigma\right)^{-1}$$

e

$$K_E = \begin{pmatrix} \kappa_{3,3} & \kappa_{3,4} & 0 \\ \kappa_{3,4} & \kappa_{4,4} & 0 \\ 0 & 0 & \kappa_{\beta,\beta} \end{pmatrix} = \begin{pmatrix} \frac{\nu+n}{\nu+2n+2} \left(\frac{2\phi_3^2}{n} \right)^{-1} & -\frac{n}{\nu+2n+2} \left(\frac{2\phi_3\phi_4}{n} \right)^{-1} & 0 \\ -\frac{n}{\nu+2n+2} \left(\frac{2\phi_3\phi_4}{n} \right)^{-1} & \frac{\nu+n}{\nu+2n+2} \left(\frac{2\phi_4^2}{n} \right)^{-1} & 0 \\ 0 & 0 & \frac{\nu+2n}{\nu+2n+2} \left(\frac{\sigma_\beta^2}{n} \right)^{-1} \end{pmatrix},$$

onde $\sigma_\beta^2 = \sigma_\beta^2(\phi_E)$ está definida em (5.33). Assim, quando $\nu \rightarrow \infty$ verifica-se que $K = K_\nu \rightarrow K_\infty$, onde K_∞ é a matriz de informação do modelo normal.

Convém notar que para os modelos que são mistura de escala do modelo normal, como o caso do modelo t (veja o Capítulo 3), a matriz de informação pode ser derivada de forma alternativa usando (5.59). Isto é (veja Louis (1982)),

$$K = E \left\{ E \left[\frac{\partial \log p(\mathbf{Z}|V; \phi)}{\partial \phi} \middle| \mathbf{Z} \right] E \left[\frac{\partial \log p(\mathbf{Z}|V; \phi)}{\partial \phi^\top} \middle| \mathbf{Z} \right] \right\},$$

onde

$$\log p(\mathbf{Z}|V; \phi) \propto -\frac{n}{2} \log |\Sigma| - \frac{1}{2V} \sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu);$$

é a log-verossimilhança de (5.59).

b) Derivação dos EMV

Seja $L = L(\phi)$ a função de log-verossimilhança do modelo (5.57). Então, para cada $\nu > 0$ fixado, tem-se que

$$\begin{aligned} L &\propto -\frac{n}{2} \log |\Sigma| - \frac{\nu+2n}{2} \log \left\{ \nu + \sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) \right\} \\ &= -\frac{n}{2} \log |\Sigma| - \frac{\nu+2n}{2} \log \left\{ \nu + n \text{tr}(\Sigma^{-1} S_{ZZ}) + n(\bar{Z} - \mu)^\top \Sigma^{-1} (\bar{Z} - \mu) \right\}, \quad (5.61) \end{aligned}$$

onde

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i \quad \text{e} \quad S_{ZZ} = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^\top.$$

Como $\Sigma > 0$ e $n > 2$, da segunda expressão dada acima para L é fácil ver que neste modelo os EMV de μ e Σ são dados por

$$\hat{\mu} = \bar{Z} \quad \text{e} \quad \hat{\Sigma} = S_{ZZ}.$$

Note-se que usando o fato de Anderson et al. (1986) dado em (5.51), este último resultado é direto, uma vez que para a distribuição t com ν graus de liberdade tem-se que $u_f = np$ para cada $\nu > 0$.

Do resultado anterior segue que se λ_x ou λ_e é conhecido, então (como no modelo normal) o EMV de $\phi = (\phi_L^T, \phi_E^T)^T$ pode ser obtido resolvendo as equações (“de momentos”) dadas por

$$\mu(\hat{\phi}_L) = \bar{Z} \quad \text{e} \quad \Sigma(\hat{\phi}_E) = S_{ZZ}.$$

Daqui é claro que $\hat{\phi}_L = \bar{Z}$ e que $\hat{\phi}_E = \phi_E(S_{ZZ})$ é dado por (5.40) para λ_x conhecido e por (5.41) para λ_e conhecido. Tal resultado pode ser conferido resolvendo diretamente as equações de log-verossimilhança, e pode estender-se para todos os modelos estruturais elípticos dependentes que satisfazem as condições da existência e unicidade de EMV (μ, Σ) (veja Anderson et al. (1986)). Desta forma, as propriedades do EMV $\hat{\phi}$ podem ser estudadas sob o contexto geral do modelo elíptico. Contudo, no que segue, são dados apenas alguns resultados no contexto do modelo t que estuda-se nesta seção.

5.3.2 Propriedades dos EMV

Como foi visto no Capítulo 2 (veja o Exemplo 2.15), a independência de $\bar{Z}$ e S_{ZZ} é própria do modelo normal. Isto é, no modelo t (ou em qualquer modelo elíptico) os EMV $\hat{\phi}_L = \phi_L(\bar{Z})$ e $\hat{\phi}_E = \phi_E(S_{ZZ})$ não são independentes. Contudo, as propriedades destes EMV podem ser estudadas usando que (5.59) implica que

$$(\bar{Z}|V) \sim N_2(\mu, \frac{V}{n}\Sigma) \quad \text{e} \quad (S_{ZZ}|V) \sim W_2(\frac{V}{n}\Sigma, n-1)$$

são independentes, onde $V \sim \nu/\chi_\nu^2$. Daqui é fácil ver, por exemplo, que:

$$\hat{\phi}_L = \bar{Z} \sim t_2(\phi_L, \frac{1}{n}\Sigma(\phi_E); \nu),$$

de modo que

$$E(\hat{\phi}_L) = \phi_L \quad \text{e} \quad \text{Cov}(\hat{\phi}_L) = \frac{1}{n} \frac{\nu}{\nu-2} \Sigma,$$

para $\nu > 1$ e $\nu > 2$, respectivamente. De forma similar pode-se derivar a distribuição incondicional de S_{ZZ} (veja Dawid (1977, 1981) e Sutradhar e Ali (1989)) para obter a distribuição de $\hat{\phi}_E$. Contudo, como foi mencionado na Seção 5.2.2, as propriedades obtidas supondo normalidade das estatísticas que são invariantes sob multiplicações de escalares positivos são preservadas neste modelo. Em particular, uma tal destas estatísticas vem dada pelo EMV $\hat{\beta}$ de β . De fato, tanto no modelo normal quanto no modelo dependente o EMV de β é dado por (5.40) para λ_x conhecido e por (5.41) para λ_e conhecido. Usando estas expressões é fácil ver que $\hat{\beta}(az) = \hat{\beta}(z)$, $a \in \mathbb{R}$, $z \in \mathbb{R}^{2n}$. Consequentemente, $\hat{\beta}$ tem a mesma distribuição sob ambos os modelos. Para ilustrar com mais detalhes considere-se o caso com λ_x conhecido. Neste caso, o EMV de β é dado por

$$\hat{\beta} = \left(\frac{\lambda_x + 1}{\lambda_x} \right) \frac{S_{YX}}{S_{XX}}.$$

Considerando o modelo t dependente, então (5.59) implica que $(S_{ZZ}|V = v) \sim W_2(n^{-1}v\Sigma, n-1)$, de onde segue que (veja a propriedade (10) na Seção 5.1.2c)

$$(\hat{\beta}|V = v) \sim t\left(\beta, \frac{\sigma_\beta^2}{n-1}; n-1\right),$$

para todo $v > 0$. Como esta distribuição condicional não depende de v , tem-se que $\hat{\beta}$ é independente V e, portanto, $\hat{\beta} \stackrel{d}{=} (\hat{\beta}|V = v) \sim t(\beta, \sigma_\beta^2/(n-1); n-1)$.

Considerando as propriedades desenvolvidas na Seção 5.1.2c sob normalidade, resultados similares podem ser derivados para outras estatísticas no modelo t dependente, como, por exemplo, para a estatística escore da hipótese $H_0 : \beta = \beta_0$. Contudo, a análise deste modelo será finalizada com o estudo da estatística da razão de verossimilhança da hipótese $H_0 : \beta = \beta_0$.

5.3.3 Testes da razão de verossimilhança e escore de $H_0 : \beta = \beta_0$

Considere-se o modelo t -dependente dado em (5.57) sob a parametrização ortogonal $(\mu(\phi), \Sigma(\phi))$. Suponha que deseja-se testar a hipótese nula $H_0 : \beta = \beta_0$ (contra $H_1 :$

$\beta = \beta_1$) usando a estatística de razão de verossimilhança $G = 2(L(\hat{\phi}) - L(\tilde{\phi}))$, onde $\hat{\phi}$ e $\tilde{\phi}$ são os EMV não restritos (sob H_1) e restrito (sob H_0) de ϕ . É claro que tanto $\hat{\phi}$ quanto $\tilde{\phi}$ coincidem com os respectivo EMV do modelo normal (veja a Seção 5.1.2d). Mas, para um modelo elíptico (em particular, para o modelo t ; veja (5.61)), $L(\phi)$ é da forma

$$\begin{aligned} L(\phi) &= L(\mu(\phi), \Sigma(\phi)) = \log \left\{ |\Sigma|^{-n/2} f \left(\sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) \right) \right\} \\ &= \log \left\{ |\Sigma|^{-n/2} f(n \text{tr}(\Sigma^{-1} S_{ZZ}) + n(\bar{Z} - \mu)^\top \Sigma^{-1} (\bar{Z} - \mu)) \right\}. \end{aligned}$$

Logo, aplicando (5.51) sob H_0 e H_1 , tem-se que $\tilde{\mu} = \hat{\mu} = \bar{Z}$ e $\text{tr}(\tilde{\Sigma}^{-1} S_{ZZ}) = \text{tr}(\hat{\Sigma}^{-1} S_{ZZ}) = np/uf$, de onde segue que

$$G = n \log \left\{ \frac{|\Sigma(\tilde{\phi}_E)|}{|\Sigma(\hat{\phi}_E)|} \right\} = n \log \left\{ \frac{\tilde{\phi}_3 \tilde{\phi}_4}{\hat{\phi}_3 \hat{\phi}_4} \right\},$$

e pode-se escrever em termos dos elementos de $S_{ZZ} = [S_{ij}]$ como em (5.45). É claro então que esta estatística verifica a propriedade de invariância sob multiplicações de escalares, de modo que preserva a distribuição da correspondente estatística sob normalidade. Com tal resultado é válido para cada n , então a distribuição limite de G também deve ser χ_1^2 . Segue-se então que no modelo t , em particular, ou elíptico, em geral, a correspondente estatística modificada vem dada por (5.46).

Um resultado similar pode ser obtido para a estatística escore S . De fato, no caso do modelo t -dependente é fácil ver que $S = S_\nu$, com

$$S_\nu = \frac{\nu + 2n + 2}{\nu + 2n} S_\infty,$$

onde $S_\infty = \lim_{\nu \rightarrow \infty} S_\nu$ é a respectiva estatística para o modelo normal. Considerando que, sob $H_0 : \beta = \beta_0$, $n^{-1} S_\infty \sim \text{Beta}(1/2, (n-2)/2)$, a distribuição de S_ν é imediata.

5.4 Inferência no modelo estrutural t -independente

É bem sabido que sob normalidade, especificar a distribuição de $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$ supondo que os vetores $Z_i = (Y_i, X_i)^\top$, $i = 1, 2, \dots, n$, sejam não correlacionados é equiva-

lente a supor que estes vetores sejam independentes. Em outras palavras, para o modelo estrutural normal tem-se que

$$\mathbf{Z} \sim N_{2n}(\boldsymbol{\mu}, I_n \otimes \Sigma) \Leftrightarrow Z_i \stackrel{\text{iid}}{\sim} N_2(\mu, \Sigma), \quad i = 1, 2, \dots, n.$$

Seguindo então o enfoque de Lange et al. (1989), este modelo pode ser estendido, preservando a independência, para o modelo estrutural t -independente definido por

$$\mathbf{Z} \sim t_{2n}^{(I)}(\boldsymbol{\mu}, I_n \otimes \Sigma; \nu) \Leftrightarrow Z_i \stackrel{\text{iid}}{\sim} t_2(\mu, \Sigma; \nu), \quad i = 1, 2, \dots, n. \quad (5.62)$$

Logo, neste modelo também tem-se que $E(Z_i) = \nu$ para $\nu > 1$ e

$$\text{Cov}(Z_i, Z_j) = \begin{cases} \frac{\nu}{\nu-2} \Sigma, & i = j; \\ 0, & i \neq j, \end{cases}$$

se $\nu > 2$. A diferença do modelo estrutural t -dependente, para (5.62), a densidade de $\mathbf{Z} = (Z_1^\top, Z_2^\top, \dots, Z_n^\top)^\top$ é dada por

$$\begin{aligned} p_{\mathbf{Z}}(\mathbf{z}|\mu, \Sigma) &= \prod_{i=1}^n \left\{ k(2, \nu) |\Sigma|^{-1/2} \nu^{\nu/2} \left\{ \nu + (z_i - \mu)^\top \Sigma^{-1} (z_i - \mu) \right\}^{-(\nu+2)/2} \right\} \\ &= (2\pi)^{-n} |\Sigma|^{-n/2} \nu^{n(\nu+2)/2} \prod_{i=1}^n \left\{ \nu + (z_i - \mu)^\top \Sigma^{-1} (z_i - \mu) \right\}^{-(\nu+2)/2}, \end{aligned}$$

onde a segunda destas expressões segue do fato de que $k(2, \nu) = \Gamma[(\nu+2)/2]/\Gamma(\nu/2)\pi = \nu/2\pi$.

Este modelo também pode ser formulado como mistura de escala do modelo estrutural normal supondo que

$$(\mathbf{Z}|\mathbf{V}) \sim N_{2n}(\boldsymbol{\mu}, D(\mathbf{V}) \otimes \Sigma) \Leftrightarrow (Z_i|V_i) \stackrel{\text{iid}}{\sim} N_2(\mu, V_i \Sigma), \quad i = 1, 2, \dots, n,$$

onde $\mathbf{V} = (V_1, V_2, \dots, V_n)^\top$, $D(\mathbf{V}) = \text{Diag}\{V_1, V_2, \dots, V_n\}$ e $V_i \stackrel{\text{iid}}{\sim} \nu/\chi_\nu^2$. Daqui é fácil ver que (veja (3.24) e (3.25))

$$E(V_i|Z_i) = \frac{\nu+2}{\nu + \|T_i\|^2} = \omega_i, \quad i = 1, 2, \dots, n, \quad (5.63)$$

onde $\|T_i\|^2 = (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu)$. Note-se que

$$\|T_i\|^2 \stackrel{\text{iid}}{\sim} n F_{2,\nu}, \quad i = 1, 2, \dots, n.$$

Tal como é ilustrado em Lange et al. (1989) em diferentes contextos, a conveniência do modelo t -independente vem dada pelo fato de que se ν é fixado num valor adequado, então seus EMV são robustos com respeito a observações aberrantes. Este resultado é devido a que nas equações de log-verossimilhança, a contribuição da i -ésima observação está ponderada pelo fator $\omega_i = \omega(\|T_i\|^2)$ definido em (5.63), o qual é uma função decrescente na distância $\|T_i\|^2 = (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu)$.

No que segue estuda-se o modelo estrutural t -independente seguindo os desenvolvimentos feitos nas Seções 4.3 e 4.4, de modo que muitos detalhes serão omitidos. Começa-se incorporando a parametrização ortogonal ϕ (definida em (5.26)) de (μ, Σ) , para derivar o EMV de ϕ e a matriz de informação $K(\phi)$. Finaliza-se com a derivação dos testes da razão de verossimilhança (G) e escore (S) modificados para a hipótese $H_0 : \beta = \beta_0$. Um experimento de simulação também é feito com o propósito de ilustrar o comportamento dos testes modificados. Para desenvolver alguns destes resultados a seguinte notação será requerida.

Sejam

$$\bar{Z}(\omega) = \frac{\sum_{i=1}^n \omega_i Z_i}{\sum_{i=1}^n \omega_i} \quad \text{e} \quad S_{ZZ}(\omega) = \frac{1}{n} \sum_{i=1}^n \omega_i (Z_i - \bar{Z}(\omega))(Z_i - \bar{Z}(\omega))^\top, \quad (5.64)$$

a média e a covariância de $Z_1, Z_2, \dots, Z_n$ ponderadas pelos fatores $\omega_1, \omega_2, \dots, \omega_n$ definidos em (5.63). Os elementos $(\bar{Y}(\omega), \bar{X}(\omega))$ de $\bar{Z}(\omega)$ e $(S_{YY}(\omega), S_{YX}(\omega), S_{XX}(\omega))$ de $S_{ZZ}(\omega)$ são definidos da mesma forma. Note-se que $\omega = (\omega_1, \omega_2, \dots, \omega_n)^\top$ é desconhecido, uma vez que depende dos parâmetros $\mu = \mu(\phi_L)$, $\Sigma = \Sigma(\phi_E)$ e ν , onde $\phi = (\phi_L^\top, \phi_E^\top)^\top$ é a parametrização ortogonal definida em (5.26).

5.4.1 EMV e matriz de informação da parametrização ortogonal

(a) Derivação dos EMV

Seja $L = L(\phi, \nu)$ a função de log-verossimilhança de (5.62). Então, $L = \sum_{i=1}^n \ell_i$, onde

$$\ell_i = \log(2\pi) - \frac{1}{2} \log |\Sigma| + \frac{\nu + 2}{2} \log \nu - \frac{\nu + 2}{\nu} \log \{ \nu + \|T_i\|^2 \}, \quad (5.65)$$

e $T_i = \Sigma^{-1/2}(Z_i - \mu)$. Considerando a parametrização ortogonal $(\mu(\phi_L), \Sigma(\phi_E))$ tem-se que $|\Sigma| = \phi_3 \phi_4$ (veja (5.27) - (5.29)). Supondo ν conhecido, é fácil ver que o EMV de $\phi = (\phi_L^T, \phi_E^T)^T$ vem dado pela solução das equações de “momentos ponderados” definidas por

$$\mu(\hat{\phi}_L) = \bar{Z}(\hat{\omega}) \quad \text{e} \quad \Sigma(\hat{\phi}_E) = S_{ZZ}(\hat{\omega}), \quad (5.66)$$

onde $\hat{\omega} = \omega(\hat{\phi})$. Para um dado $\hat{\omega}$ é claro que a primeira equação de (5.66) implica em (5.39) com $(\bar{Y}, \bar{X})$ substituído por $(\bar{Y}(\hat{\omega}), \bar{X}(\hat{\omega}))$. Similarmente, se λ_x ou λ_e é suposto conhecido, a segunda equação de (5.66) implica em (5.40) ou (5.41) com (S_{YY}, S_{YX}, S_{XX}) substituído por $(S_{YY}(\hat{\omega}), S_{YX}(\hat{\omega}), S_{XX}(\hat{\omega}))$. Contudo, para obter a solução destas equações deve-se usar algoritmos numéricos. Novamente, a conveniência do algoritmo EM é clara, uma vez que este algoritmo se reduz a reavaliar iterativamente (5.66), passando previamente pelas equações (5.63) e (5.64). O valor inicial $\phi^{(0)}$ pode ser escolhido como a solução clássica que segue de resolver (5.66) para $(\bar{Z}(\hat{\omega}), S_{ZZ}(\hat{\omega})) = (\bar{Z}, S_{ZZ})$. Assim, se $\phi^{(\ell)}$ denota solução da ℓ -ésima iteração, então na $(\ell + 1)$ -iteração o algoritmo procede como segue:

Etapla E: compute $\omega^{(\ell)} = \omega(\phi^{(\ell)})$ e $(Z(\omega^{(\ell)}), S_{ZZ}(\omega^{(\ell)}))$ usando (5.63) e (5.66), respectivamente;

Etapla M: compute $\phi^{(\ell+1)}$ a partir de (5.66) (ou (5.39) - (5.41)) usando $(\bar{Z}(\omega^{(\ell)}), S_{ZZ}(\omega^{(\ell)}))$.

Se ν também é desconhecido, então $\nu^{(\ell+1)}$ pode ser obtido a partir da equação dada por (veja (4.42))

$$\psi\left(\frac{\nu^{(\ell+1)}}{2}\right) - \log\left(\frac{\nu^{(\ell+1)}}{2}\right) = \bar{\omega}_a^{(\ell)} - \log \bar{\omega}_g^{(\ell)},$$

onde $\psi(\cdot)$ denota a função digama e $\bar{\omega}_a$ e $\bar{\omega}_g$ são as médias aritmética e geométrica de $\omega_1, \omega_2, \dots, \omega_n$, respectivamente. Uma forma alternativa é dada por aplicar o algoritmo sucessivamente para uma rede de valores de ν , e escolher aquele valor que maximiza a verossimilhança.

De qualquer forma, quando $\nu > 4$, um estimador consistente deste parâmetro pode ser construído a partir dos resíduos padronizados $S_{ZZ}^{-1/2}(Z_i - \bar{Z})$, $i = 1, 2, \dots, n$. De fato, para $\nu > 2$ tem-se que $\mu_Z = E(Z_i) = \mu_Z$, e $\Sigma_Z = \text{Cov}(Z_i) = \alpha_\nu \Sigma$, onde $\alpha_\nu = \nu/(\nu - 2)$, são finitos. Em tal caso, tem-se que

$$\text{plim}(\bar{Z}, S_{ZZ}) = (\mu_Z, \Sigma_Z) = (\mu, \alpha_\nu \Sigma).$$

Sejam (veja (2.45)),

$$\beta_{2,2} = E\{[(Z_1 - \mu_Z)^\top \Sigma_Z^{-1}(Z_1 - \mu_Z)]^2\} \quad \text{e} \quad \hat{\beta}_{2,2} = \frac{1}{n} \sum_{i=1}^n [(Z_i - \bar{Z})^\top S_{ZZ}^{-1}(Z_i - \bar{Z})]^2.$$

Então, para $\nu > 4$, é fácil ver que

$$\text{plim} \hat{\beta}_{2,2} = \beta_{2,2} = 4 \frac{\nu - 2}{\nu - 4},$$

de onde segue que um estimador consistente para ν é dado por

$$\hat{\nu}_M = 4 \frac{\hat{\beta}_{2,2} - 2}{\hat{\beta}_{2,2} - 4}.$$

Como é indicado pela notação, $\hat{\nu}_M$ é denominado o estimador de momentos de ν .

Um outro método para obter a solução de máxima verossimilhança vem dado pelo algoritmo de scoring. As derivadas da log-verossimilhança e a matriz de informação são requeridas para a implementação deste algoritmo. Usando (5.65) (veja também a Seção

5.2.3) tem-se que

$$\frac{\partial L}{\partial \phi_i} = \sum_{k=1}^n \frac{\partial \ell_k}{\partial \phi_i} = \begin{cases} \sum_{k=1}^n \omega_k \frac{\partial \mu^\top}{\partial \phi_i} \Sigma^{-1/2} T_k, & i = 1, 2, \\ \sum_{k=1}^n \left(-\frac{1}{2} \phi_i^{-1} + \frac{1}{2} \omega_k T_k^\top \Sigma^{-1/2} \frac{\partial \Sigma}{\partial \phi_i} \Sigma^{-1/2} T_k \right), & i = 3, 4, \\ \sum_{k=1}^n \frac{1}{2} \omega_k T_k^\top \Sigma^{-1/2} \frac{\partial \Sigma}{\partial \beta} \Sigma^{-1/2} T_k, & \phi_i = \beta, \end{cases}$$

onde ω_k e T_k estão definidos em (5.63). Se ν é desconhecido, tem-se também que

$$\frac{\partial L}{\partial \nu} = \sum_{k=1}^n \frac{1}{2} \left\{ \frac{\nu+2}{\nu} - \log \left(\frac{\nu+2}{\nu} \right) + \log \omega_k - \omega_k \right\}.$$

Formas alternativas e mais convenientes destas equações podem ser obtidas com uma simples álgebra. Por exemplo (veja (i) - (iv) na Seção 5.1.2a e (5.38)), estas equações podem ser expressas como

$$\begin{aligned} \frac{\partial L}{\partial \phi_L} &= n \bar{\omega}_a \Sigma^{-1} (\bar{Z}(\omega) - \mu), \\ \frac{\partial L}{\partial \phi_i} &= \frac{n}{2 \phi_i^2} \{ \bar{\alpha}_i^\top (S_{ZZ}(\omega) + \bar{\omega}_a (\bar{Z}(\omega) - \mu) (\bar{Z}(\omega) - \mu)^\top) \bar{\alpha}_i - \phi_i \}, \quad i = 3, 4, \\ \frac{\partial L}{\partial \beta} &= \frac{n}{\sigma_\beta \sqrt{\phi_3 \phi_4}} \bar{\alpha}_3^\top (S_{ZZ}(\omega) - \bar{\omega}_a (\bar{Z}(\omega) - \mu) (\bar{Z}(\omega) - \mu)^\top) \bar{\alpha}_4, \\ \frac{\partial L}{\partial \nu} &= \frac{n}{2} \left\{ \frac{\nu+2}{\nu} - \log \left(\frac{\nu+2}{2} \right) + \log \bar{\omega}_g - \bar{\omega}_a \right\}, \end{aligned} \quad (5.67)$$

onde $\bar{Z}(\omega)$ e $S_{ZZ}(\omega)$ estão definidos em (5.64); $\bar{\omega}_a$ e $\bar{\omega}_g$ são a média aritmética e geométrica das ponderações $\omega_1, \omega_2, \dots, \omega_n$ dadas em (5.63) e $\bar{\alpha}_j$, $j = 3, 4$, são os vetores definidos em (5.31). Note-se que as três primeiras equações se reduzem a (5.38) quando $\omega_i = 1$, $i = 1, 2, \dots, n$. Além disso, destas equações tem-se (5.66) de forma clara.

(b) Derivação da matriz de informação

Para ν conhecido, a matriz de informação pode ser obtida diretamente de (5.56), (5.57) e (5.60) colocando $p = 2$. Daqui segue que $K(\phi_L, \phi_E) = \text{Diag} \{K_L, K_E\}$, onde

$$K_L = \left(\frac{\nu+2}{\nu+4} \right) \left(\frac{1}{n} \Sigma \right)^{-1}$$

e

$$K_E = \begin{pmatrix} \kappa_{3,3} & \kappa_{3,4} & 0 \\ \kappa_{3,4} & \kappa_{4,4} & 0 \\ 0 & 0 & \kappa_{\beta,\beta} \end{pmatrix} = \begin{pmatrix} \frac{\nu+2}{\nu+4} \left(\frac{2\phi_3^2}{n} \right)^{-1} & -\frac{1}{\nu+4} \left(\frac{2\phi_3\phi_4}{n} \right)^{-1} & 0 \\ -\frac{1}{\nu+4} \left(\frac{2\phi_3\phi_4}{n} \right)^{-1} & \frac{\nu+2}{\nu+4} \left(\frac{2\phi_4^2}{n} \right)^{-1} & 0 \\ 0 & 0 & \frac{\nu+2}{\nu+4} \left(\frac{\sigma_\beta^2}{n} \right)^{-1} \end{pmatrix}, \quad (5.68)$$

onde $\sigma_\beta^2 = \sigma_\beta^2(\phi_E)$ está definida em (5.33). Para ν desconhecido, usando a simetria da distribuição de $T_k = \Sigma^{-1/2}(Z_k - \mu) \sim t_2(0, I_2; \nu)$, é fácil ver que $\kappa_{\nu,i} = 0$, $i = 1, 2$, de modo que ν é (globalmente) ortogonal com o vetor de locação $\mu = \phi_L = (\phi_1, \phi_2)^\top$. Para obter $\kappa_{\nu,i}$, $i = 3, 4$, $\kappa_{\nu,\beta}$ e $\kappa_{\nu,\nu}$ é mais simples usar as respectivas segundas derivadas da log-verossimilhança. Procedendo desta forma e usando (5.60) para $p = 2$, tem-se que

$$\begin{aligned} \kappa_{\nu,i} &= \frac{n}{(\nu+2)(\nu+4)} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \phi_i} \right) = \begin{cases} -\frac{n}{(\nu+2)(\nu+4)} \phi_i^{-1}, & i = 3, 4, \\ 0, & i = 5, \end{cases} \\ \kappa_{\nu,\nu} &= \frac{8n}{\nu^2(\nu+2)(\nu+4)}. \end{aligned}$$

Note-se que $\kappa_{\nu,5} = \kappa_{\nu,\beta} = 0$ segue do fato de que na parametrização ortogonal tem-se que $\text{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta}) = 0$ (veja (5.34)), o qual não sucede na parametrização original do modelo estrutural que estuda-se. Por outro lado, o fato de que $\kappa_{\nu,i} \neq 0$, $i = 3, 4$, altera a estrutura diagonal em blocos da matriz de informação total $K(\phi_L, \phi_E, \nu) = K(\phi_1, \phi_2, \phi_3, \phi_4, \nu)$, que complica, entre outras coisas, a inversão desta matriz. Contudo, segundo o procedimento feito na Seção 4.3.3, pode-se omitir esta dificuldade derivando uma parametrização ortogonal tipo Cox e Reid (1987) para o subvetor $(\phi_3, \phi_4, \nu)^\top$, supondo ν como o parâmetro de interesse e fixando os parâmetros restantes ϕ_1 , ϕ_2 e β . Assim, a matriz de informação se reduz a

$$K(\phi_3, \phi_4, \nu) = \begin{pmatrix} \kappa_{3,3} & \kappa_{3,4} & \kappa_{3,\nu} \\ \kappa_{4,3} & \kappa_{4,4} & \kappa_{4,\nu} \\ \kappa_{\nu,3} & \kappa_{\nu,4} & \kappa_{\nu,\nu} \end{pmatrix},$$

onde os $\kappa_{i,j}$'s estão definidos acima. Se $(\varphi_3, \varphi_4, \nu)^\top$ é o novo vetor de parâmetros tal que ν seja ortogonal com (φ_3, φ_4) , então como é estabelecido em Cox e Reid (1987), $\phi_i = \phi_i(\varphi_3, \varphi_4, \nu)$ é a solução do sistema de equações diferenciais dado por

$$\sum_{i=3}^4 \kappa_{i,j} \frac{\partial \phi_i}{\partial \nu} = -\kappa_{\nu,j}, \quad j = 3, 4.$$

Dos resultados dados acima, é fácil ver que este sistema pode ser expresso matricialmente como

$$\begin{bmatrix} \nu+2 & -1 \\ -1 & \nu+2 \end{bmatrix} \begin{bmatrix} \frac{1}{\phi_3} \frac{\partial \phi_3}{\partial \nu} \\ \frac{1}{\phi_4} \frac{\partial \phi_4}{\partial \nu} \end{bmatrix} = \frac{2}{\nu+2} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

de onde segue que

$$\frac{1}{\phi_i} \frac{\partial \phi_i}{\partial \nu} = 2 \left(\frac{1}{\nu+1} - \frac{1}{\nu+2} \right), \quad i = 3, 4.$$

Estas equações têm solução da forma $\phi_i = c_i(\varphi_3, \varphi_4) \left(\frac{\nu+1}{\nu+2} \right)^2$, onde $c_i(\varphi_3, \varphi_4)$ é uma função arbitrária de $(\varphi_3, \varphi_4)^\top$. Uma escolha apropriada vem dada por $c_i(\varphi_3, \varphi_4) = \varphi_i$, $i = 3, 4$. Neste caso, tem-se que

$$\phi_i = \varphi_i \left(\frac{\nu+1}{\nu+2} \right)^2, \quad i = 3, 4.$$

Sejam $L^* = L^*(\phi_1, \phi_2, \varphi_3, \varphi_4, \beta, \nu)$ e $K^* = K^*(\phi_1, \phi_2, \varphi_3, \varphi_4, \beta, \nu)$ a log-verossimilhança e a matriz de informação da nova parametrização. Então, usando que $L^* = L(\phi_1, \phi_2, \phi_3(\phi_3, \nu), \phi_4(\varphi_4, \nu), \beta, \nu)$ e procedendo como na Seção 4.3.3, é fácil ver que $K^* = [\kappa_{i,j}^*]$ é da forma

$$K^* = \text{Diag} \{ K_1^*, K_2^*, \kappa_{\beta,\beta}^*, \kappa_{\nu,\nu}^* \},$$

com

$$\begin{aligned} K_1^* &= [\kappa_{i,j}(\phi^*)] = K_L(\phi^*), \quad i, j = 1, 2, \\ K_2^* &= \left[\frac{\partial \phi_i}{\partial \varphi_i} \frac{\partial \phi_j}{\partial \varphi_j} \kappa_{i,j}(\phi^*) \right], \quad i, j = 3, 4, \\ \kappa_{\beta,\beta}^* &= \kappa_{\beta,\beta}(\phi^*), \\ \kappa_{\nu,\nu}^* &= \kappa_{\nu,\nu}(\phi^*) + \sum_{i=3}^4 \frac{\partial \phi_i}{\partial \nu} \kappa_{\nu,i}(\phi^*), \end{aligned}$$

onde os $\kappa_{i,j}(\phi^*)$'s são os elementos da matriz original $K = [\kappa_{i,j}]$ definida mais acima, mas avaliados em $\phi^* = (\phi_1, \phi_2, \varphi_3 \left(\frac{\nu+1}{\nu+2} \right)^2, \varphi_4 \left(\frac{\nu+1}{\nu+2} \right)^2, \beta, \nu)^\top$. Isto é,

$$\begin{aligned} K_1^* &= \left(\frac{\nu+2}{\nu+1} \right)^2 \left(\frac{\nu+2}{\nu+4} \right) \left(\frac{1}{n} \Sigma(\varphi_E) \right)^{-1}, \\ K_2^* &= \left(\frac{\nu+2}{\nu+1} \right)^2 \begin{pmatrix} \frac{\nu+2}{\nu+4} \left(\frac{2\varphi_3^2}{n} \right)^{-1} & -\frac{1}{\nu+4} \left(\frac{2\varphi_3\varphi_4}{n} \right)^{-1} \\ -\frac{1}{\nu+4} \left(\frac{2\varphi_3\varphi_4}{n} \right)^{-1} & \frac{\nu+2}{\nu+4} \left(\frac{2\varphi_4^2}{n} \right)^{-1} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned}\kappa_{\beta,\beta}^* &= \frac{\nu + 2}{\nu + 4} \left(\frac{\sigma_{\beta}^2(\varphi_E)}{n} \right)^{-1}, \\ \kappa_{\nu,\nu}^* &= \frac{4n(\nu^2 + 6\nu + 4)}{\nu^2(\nu + 1)(\nu + 2)(\nu + 4)},\end{aligned}$$

onde $\Sigma(\varphi_E)$ e $\sigma_{\beta}^2(\varphi_E)$ resultam de substituir $\phi_E = (\phi_3, \phi_4, \beta)^\top$ por $\varphi_E = (\varphi_3, \varphi_4, \beta)^\top$ em (5.28) e (5.33), respectivamente. Note-se que $\kappa_{\beta,\beta}^* = \kappa_{\beta,\beta}(\varphi_E, \nu)$, de modo que sua forma não é alterada pela reparametrização de (ϕ_3, ϕ_4) .

Finalmente, convém notar que esta parametrização não altera a implementação do algoritmo EM e simplifica a do algoritmo de scoring. Além disso, as inferências sobre os parâmetros $(\phi_1, \phi_2)^\top$, $(\varphi_3, \varphi_4)^\top$, β e ν é simplificada, uma vez que seus respectivos EMV são assintoticamente independentes com $(\hat{\phi}_1, \hat{\phi}_2)^\top \sim AN_2((\phi_1, \phi_2)^\top, (K_1^*)^{-1})$, $(\hat{\varphi}_3, \hat{\varphi}_4)^\top \sim N_2((\varphi_3, \varphi_4)^\top, (K_2^*)^{-1})$, $\hat{\beta} \sim AN(\beta, 1/\kappa_{\beta,\beta}^*)$ e $\hat{\nu} \sim AN(\nu, 1/\kappa_{\nu,\nu}^*)$.

5.4.2 Testes da razão de verossimilhança e escore modificados da hipótese $H_0 : \beta = \beta_0$

Tanto no enfoque estrutural quanto no funcional do modelo com erros nas variáveis definido em (5.1) e (5.2), o objetivo principal é a estimação da relação linear $y = \alpha + \beta x$ entre as variáveis não observáveis (y, x) . Neste sentido, dadas n observações medidas com erro de (y, x) , digamos, $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$, as inferências centram-se usualmente no coeficiente β da relação linear. Assim, o teste da hipótese nula $H_0 : \beta = \beta_0$ tem grande relevância nas diferentes áreas de aplicação destes modelos. Contudo, como já foi mencionado, no modelo estrutural (ou funcional) as inferências sobre β em geral devem apoiar-se na teoria assintótica. Assim, por exemplo, os testes de $H_0 : \beta = \beta_0$ estão usualmente baseados na distribuição assintótica da estatística do teste, uma vez que a distribuição exata desta estatística é desconhecida ou é difícil de derivar. Isto acontece, por exemplo, com os testes da razão de verossimilhança (G) e escore (S) num determinado modelo estrutural, como o modelo t -independente que estuda-se nesta seção. À apro-

ximação (sob $H_0 : \beta = \beta_0$) para a distribuição qui-quadrado com um grau de liberdade (χ_1^2) é considerada neste caso. Como foi estabelecido na Seção 4.4 do Capítulo 4, a inconveniência destes procedimentos assintóticos é que a aproximação entre as distribuições da estatística do teste e de referência está garantida somente para valores “suficientemente grandes” do tamanho da amostra (n). Por esta razão, tais procedimentos podem ser de pouca utilidade prática em situações com tamanhos moderados a pequenos de n . Daqui segue que o interesse por considerar os testes assintóticos modificados G^* e S^* definidos em (4.52) e (4.53), respectivamente, para os quais a aproximação com a distribuição de referência χ_1^2 está garantida em amostras finitas pelo menos até ordem n^{-1} (veja a Seção 4.4).

Nesta seção são derivados os fatores de correção definidos na Seção 4.4 do Capítulo 4, para obter as estatísticas modificadas G^* e S^* dos testes da razão de verossimilhança e escore, respectivamente, da hipótese $H_0 : \beta = \beta_0$ no modelo estrutural t -independente definido em (5.62). A derivação destes fatores é feita supondo ν conhecido. O caso com ν desconhecido requer a avaliação de parcelas adicionais (veja a Seção 4.4.4), que podem ser derivadas de forma mais simples usando a parametrização ortogonal obtida na seção anterior para $(\phi_3, \phi_4, \nu)^T$. De qualquer modo, a conjectura aqui feita é que a contribuição de tais parcelas é desprezível. Como ν é um parâmetro de incômodo (“nuisance parameter”), um enfoque alternativo, que pode ser estudado no contexto bem geral, está dado por desenvolver os fatores de correção a partir da pseudo-log-verossimilhança $L(\phi|\tilde{\nu})$, onde $\tilde{\nu}$ é uma estimativa consistente de ν . Resultados relativos às propriedades assintóticas do pseudo-EMV $\hat{\phi}(\tilde{\nu})$ de uma pseudo-log-verossimilhança arbitrária podem ser encontrados em Gong e Samaniego (1981) e Parke (1986). Resultados similares, mas aplicados às distribuições elípticas, podem ser vistos em Kano et al. (1993).

Considere-se então o modelo estrutural t -independente definido em (5.62) com $\mu = \mu(\phi_L)$ e $\Sigma = \Sigma(\phi_E)$ dados por (5.27) e (5.28), respectivamente, onde $\phi_L = (\phi_1, \phi_2)^T$ e

$\phi_E = (\phi_3, \phi_4, \beta)^\top$ é parametrização ortogonal estabelecida em (5.26). Suponha que se deseja testar a hipótese $H_0 : \beta = \beta_0$. Então, para valores “grandes” de n uma alternativa é dada pelo fato de que $\hat{\beta} \sim AN(\beta, (\nu + 4)\sigma_\beta^2/n(\nu + 2))$, onde $\sigma_\beta^2 = \sigma_\beta^2(\phi_E)$ está definida por (veja (5.33) e (5.37))

$$\sigma_\beta^2(\phi_E) = \beta \left\{ \frac{1 - \rho_{YX}^2(\phi_E)}{\rho_{YX}^2(\phi_E)} \right\} = \begin{cases} \frac{\phi_3}{\lambda_x^2 \phi_4}, & \text{se } \lambda_x \text{ é conhecido,} \\ \left(\frac{\beta^2 + \lambda_e}{\phi_3 - \lambda_e \phi_4} \right)^2 \phi_3 \phi_4, & \text{se } \lambda_e \text{ é conhecido,} \end{cases}$$

de modo que pode ser estimada como $\hat{\sigma}_\beta^2 = \sigma_\beta^2(\hat{\phi}_E) = \hat{\beta}(1 - r_{YX}^2(\hat{\omega}))/r_{YX}^2(\hat{\omega})$, onde $r_{YX}(\hat{\omega}) = S_{YX}(\hat{\omega})/(S_{YY}(\hat{\omega})S_{XX}(\hat{\omega}))^{1/2}$ e $S_{ij}(\omega)$ sendo definidos como em (5.64). Se ν também é desconhecido, então pode-se estimá-lo usando seu respectivo EMV. Uma outra alternativa é basear o teste na estatística da razão de verossimilhança G ou escore S . Para ν conhecido, estas estatísticas são definidas por (veja (5.64))

$$G = 2(L(\hat{\phi}) - L(\tilde{\phi})) = n \left\{ \log \left(\frac{\tilde{\phi}_3 \tilde{\phi}_4}{\hat{\phi}_3 \hat{\phi}_4} \right) - (\nu + 2) \log \left(\frac{\tilde{\omega}_g}{\hat{\omega}_g} \right) \right\} \quad (5.69)$$

e (veja (5.66) e (5.67))

$$S = \frac{1}{\tilde{K}_{\beta, \beta}} \left(\frac{\partial L(\phi)}{\partial \beta} \Big|_{\phi=\tilde{\phi}} \right)^2 = n \left(\frac{\nu + 4}{\nu + 2} \right) \frac{(\bar{\alpha}_3^\top(\beta_0) S_{ZZ}(\tilde{\omega}) \bar{\alpha}_4(\beta_0))^2}{\tilde{\phi}_3 \tilde{\phi}_4}, \quad (5.70)$$

onde $\hat{\phi}$ e $\tilde{\phi}$ são os EMV sob $H_1 : \beta \neq \beta_0$ e $H_0 : \beta = \beta_0$, respectivamente; $\bar{\omega}_g = (\prod_{i=1}^n \omega_i)^{1/n}$ com ω_i definido em (5.63); $S_{ZZ}(\omega)$ é matriz de covariância ponderada dada em (5.64) e $\bar{\alpha}_j(\beta_0)$, $j = 3, 4$, são vetores definidos em (5.31) avaliados em $\beta = \beta_0$. Note-se que tanto a primeira parcela de G em (5.69) quanto S em (5.70) podem ser expressos em termos de β_0 e dos elementos da matriz $S_{ZZ}(\hat{\omega}) = [S_{ij}(\hat{\omega})]$ de forma análoga a (5.45) e (5.47), respectivamente. Como o modelo que estuda-se verifica as condições de regularidade usuais, sob as quais as estatísticas G e S convergem em distribuição (sob $H_0 : \beta = \beta_0$) para a distribuição χ_1^2 , então, como é bem sabido, o teste que rejeita H_0 para valores de G ou S maiores que o percentil $(1 - \gamma)100\%$ da distribuição χ_1^2 , tem nível aproximadamente γ em amostras suficientemente grandes.

Como foi visto no Capítulo 4 (veja a Seção 4.4), em amostras finitas o desempenho destes testes pode melhorar usando-se as respectivas estatísticas modificadas $G^* = \{1 + d\}^{-1}G$ e $S^* = \{1 + c + bS + aS^2\}^{-1}S$, onde os coeficientes d e a, b, c estão definidos em (4.51) e (5.54), respectivamente. A derivação destes coeficientes é particularmente trabalhosa no modelo estrutural que estuda-se e será feita no que segue desta seção, usando a parametrização ortogonal. Com este propósito, utiliza-se tanto a notação quanto os procedimentos desenvolvidos na Seção 4.4, de modo que muitos detalhes serão omitidos. Começa-se com a derivação do coeficiente d envolvido no fator de Bartlett $\{1 + d\}^{-1}$ (ou $\{1 - d\}$) que modifica o teste da razão de verossimilhança. Então, derivam-se os coeficientes a, b e c para corrigir o teste de escore. Nota-se que em ambos os testes estes coeficientes independem de $(\mu(\phi), \Sigma(\phi))$ e das condições de identificabilidade consideradas neste trabalho (λ_x ou λ_e conhecido).

a) Teste da razão de verossimilhança modificado

Sejam $Z_1, Z_2, \dots, Z_n$ iid $t_2(\mu(\phi_L), \Sigma(\phi_E); \nu)$, onde $\mu(\phi_L) = \phi_L = (\phi_1, \phi_2)$ e $\Sigma(\phi_E) = \Sigma(\phi_3, \phi_4, \beta)$ está definida em (5.28). Suponha que ν é conhecido e considere a estatística da razão de verossimilhança modificada do teste de $H_0 : \beta = \beta_0$, isto é,

$$G^* = \{1 + d\}^{-1}G ,$$

onde G está definida em (5.69). Então, para ambas as condições de identificabilidade, λ_x ou λ_e conhecido, tem-se que

$$\begin{aligned} d &= \frac{1}{2n} \left(\frac{\nu + 4}{\nu + 2} \right) \left\{ \frac{5\nu^3 + 86\nu^2 + 416\nu + 744}{(\nu + 6)^2(\nu + 8)} \right\} \\ &= \frac{5}{2n} \left\{ 1 - \frac{4}{5} \frac{\nu^3 + 25\nu^2 + 88\nu - 24}{(\nu + 2)(\nu + 6)^2(\nu + 8)} \right\} , \end{aligned} \quad (5.71)$$

É claro que $d = d_\nu \rightarrow d_\infty = 5/2n$, quando $\nu \rightarrow \infty$, onde d_∞ é o respectivo coeficiente sob o modelo normal (veja (5.46)). Considerando este fato, da segunda expressão dada

em (5.71) segue que d_ν pode ser escrito como

$$d_\nu = d_\infty \{1 + v_\nu\} ,$$

onde

$$v_\nu = -\frac{4}{5} \frac{\nu^3 + 25\nu^2 + 88\nu - 24}{(\nu + 2)(\nu + 6)^2(\nu + 8)} ,$$

é uma função que assume valores pequenos para todo $\nu > 0$. Por exemplo, para o modelo Cauchy tem-se que $v_1 = -0.054$, de modo que

$$d_1 = 0,946d_\infty \approx d_\infty .$$

É fácil ver que $|v_\nu| \leq 0.088$ e assume seu valor máximo em $\nu \approx 3$. Em outras palavras, o fato de aproximar d_ν por d_∞ pode implicar no máximo num erro de 8,8%.

Para mostrar o resultado dado em (5.71), por conveniência vamos denotar o vetor de parâmetros do modelo como $\theta = (\phi^\top, \beta)^\top$, onde $\phi = (\phi_1, \phi_2, \phi_3, \phi_4)^\top$. Assim, como a hipótese de interesse é $H_0 : \beta = \beta_0$, então de acordo com a notação da Seção 4.4.1, na partição $\theta = (\theta_1^\top, \theta_2^\top)^\top$ tem-se que $\theta_1 = \phi$ e $\theta_2 = \beta$. Note-se que $p = 5$ (a dimensão de θ) e $q = 1$ (a dimensão de θ_2). Convém notar também que o fato de que $\theta_1 = \phi$ e $\theta_2 = \beta$ sejam ortogonais implica que as partições induzidas para as matrizes $K = [\kappa_{i,j}] = [-\kappa_{ij}]$ e $K^{-1} = [\kappa^{ij}] = [-\kappa^{ij}]$ estão dadas por

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} K_{11} & 0 \\ 0 & \kappa_{\beta,\beta} \end{bmatrix} \quad \text{e} \quad K^{-1} = \begin{bmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{bmatrix} = \begin{bmatrix} K_{11}^{-1} & 0 \\ 0 & \kappa_{\beta,\beta}^{-1} \end{bmatrix}$$

onde $K_{11} = [\kappa_{ij}]$, $i, j = 1, 2, 3, 4$ e a submatriz de informação relativa a $\phi = (\phi_1, \phi_2, \phi_3, \phi_4)$. Como (ϕ_1, ϕ_2) e (ϕ_3, ϕ_4) também são ortogonais, então $K^{11} = K_{11}^{-1} = \text{Diag} \{K_1^{-1}, K_2^{-2}\}$, onde $K_1 = [\kappa_{i,j}]$, $i, j = 1, 2$ e $K_2 = [\kappa_{i,j}]$, $i, j = 3, 4$, isto é, (veja a Seção 5.3.1),

$$\begin{aligned} K_1^{-1} &= K_L^{-1} = \frac{1}{n} \frac{\nu + 4}{\nu + 2} \Sigma , \\ K_2^{-1} &= \begin{pmatrix} \kappa^{3,3} & \kappa^{3,4} \\ \kappa^{3,4} & \kappa^{4,4} \end{pmatrix} = \frac{1}{n} \frac{\nu + 4}{\nu + 2} \begin{pmatrix} 2(\frac{\nu+1}{\nu})\phi_3^2 & 2\frac{1}{\nu}\phi_3\phi_4 \\ 2\frac{1}{\nu}\phi_3\phi_4 & 2(\frac{\nu+1}{\nu})\phi_4^2 \end{pmatrix} . \end{aligned}$$

Colocando, então, $q = 1$ em (4.51), tem-se que

$$d = e_p - e_{p-1} , \quad (5.72)$$

onde e_p e e_{p-1} devem ser derivados a partir de (4.50), mas aplicando o somatório Σ' sobre $\{\phi_1, \phi_2, \phi_3, \phi_4, \beta\}$ e sobre $\{\phi_1, \phi_2, \phi_3, \phi_4\}$, respectivamente, e avaliando os cumulantes envolvidos em $\beta = \beta_0$. Decompondo $\Sigma' = \Sigma_\beta + \Sigma_{\beta\phi} + \Sigma_\phi$; onde Σ_β é o somatório com todos os índices fixados em β ; $\Sigma_{\beta\phi}$ é o somatório com alguns índices fixados em β e alguns dos índices restantes em $\{\phi_1, \phi_2, \phi_3, \phi_4\}$; Σ_ϕ é somatório sobre $\{\phi_1, \phi_2, \phi_3, \phi_4\}$; e definindo

$$\delta_{ijkl} = \frac{1}{4} \kappa^{ij} \kappa^{kl} \{ \kappa_{ijkl} - 4\kappa_{ijk}^{(l)} + 4\kappa_{ik}^{(jl)} \} , \quad (5.73)$$

$$\begin{aligned} \delta_{ijklrs} = & \frac{1}{12} \kappa^{ij} \kappa^{kl} \kappa^{rs} \{ 2\kappa_{ikr} \kappa_{jls} + 3\kappa_{ikl} \kappa_{jrs} \\ & - 12(\kappa_{ikr} \kappa_{js}^{(l)} + \kappa_{ikl} \kappa_{js}^{(r)}) + 12(\kappa_{ik}^{(r)} \kappa_{js}^{(l)} + \kappa_{lk}^{(l)} \kappa_{jr}^{(r)}) \} , \end{aligned} \quad (5.74)$$

então de (4.50) segue que

$$\begin{aligned} e_p &= \Sigma' \delta_{ijkl} - \Sigma' \delta_{ijklrs} \\ &= (\Sigma_\beta \delta_{ijkl} - \Sigma_\beta \delta_{ijklrs}) + (\Sigma_{\beta\phi} \delta_{ijkl} - \Sigma_{\beta\phi} \delta_{ijkl}) + (\Sigma_\phi \delta_{ijkl} - \Sigma_\phi \delta_{ijklrs}) \\ &= e_\beta + e_{\beta\phi} + e_\phi . \end{aligned}$$

Usando agora que a hipótese $H_0 : \beta = \beta_0$ implica que

$$e_{p-1} = e_\phi = \Sigma_\phi \delta_{ijkl} - \Sigma_\phi \delta_{ijklrs} ,$$

tem-se que (5.72) se reduz a

$$d = e_\beta + e_{\beta\phi} ,$$

onde

$$e_\beta = \Sigma_\beta \delta_{ijkl} - \Sigma_\beta \delta_{ijklrs} = \delta_{\beta\beta\beta\beta} - \delta_{\beta\beta\beta\beta\beta\beta} \quad \text{e} \quad e_{\beta\phi} = \Sigma_{\beta\phi} \delta_{ijkl} - \Sigma_{\beta\phi} \delta_{ijklrs} .$$

A avaliação de e_β é simples, uma vez que (5.73) e (5.74) implicam que:

$$(a) \quad \delta_{\beta\beta\beta\beta} = \frac{1}{4}(\kappa^{\beta\beta})^2 \{ \kappa_{\beta\beta\beta\beta} - 4\kappa_{\beta\beta\beta}^{(\beta)} + 4\kappa_{\beta\beta}^{(\beta\beta)} \} ,$$

$$(b) \quad \delta_{\beta\beta\beta\beta\beta\beta} = \frac{1}{12}(\kappa^{\beta\beta})^3 \{ 5(\kappa_{\beta\beta\beta})^2 - 24(\kappa_{\beta\beta\beta} - \kappa_{\beta\beta}^{(\beta)})\kappa_{\beta\beta}^{(\beta)} \} ,$$

sendo suficiente substituir os cumulantes envolvidos nestas expressões. A determinação de $e_{\beta\phi}$ não é tão simples quanto e_β , uma vez que devem ser avaliados um grande número de parcelas. Contudo, como foi visto acima, o fato de que β seja ortogonal com ϕ_i , $i = 1, 2, 3, 4$, implica que $\kappa^{\beta i} = 0$, $i = 1, 2, 3, 4$, de modo que em (5.73) e (5.74) as parcelas δ_{ijkl} e δ_{ijklrs} que eventualmente são não nulas estão dadas por:

$$(a1) \quad \delta_{\beta\beta kl} = \frac{1}{4}\kappa^{\beta\beta}\kappa^{kl}\{\kappa_{\beta\beta kl} - 4\kappa_{\beta\beta k}^{(l)}\}, \quad k, l \neq \beta,$$

$$(a2) \quad \delta_{ij\beta\beta} = \frac{1}{4}\kappa^{ij}\kappa^{\beta\beta}\{\kappa_{ij\beta\beta} - 4\kappa_{ij\beta}^{(\beta)}\}, \quad i, j \neq \beta,$$

$$(b1) \quad \delta_{\beta\beta klrs} = \frac{1}{12}\kappa^{\beta\beta}\kappa^{kl}\kappa^{rs}\{2\kappa_{\beta k r}\kappa_{\beta l s} + 3\kappa_{\beta kl}\kappa_{\beta rs}\}, \quad k, l, r, s \neq \beta,$$

$$(b2) \quad \delta_{ij\beta\beta rs} = \frac{1}{12}\kappa^{ij}\kappa^{\beta\beta}\kappa^{rs}\{2\kappa_{i\beta r}\kappa_{j\beta s} + 3\kappa_{i\beta\beta}\kappa_{jrs} - 12(\kappa_{i\beta r}\kappa_{js}^{(\beta)} + \kappa_{i\beta\beta}\kappa_{js}^{(r)})\}, \quad i, j, r, s \neq \beta,$$

$$(b3) \quad \delta_{ijkl\beta\beta} = \frac{1}{12}\kappa^{ij}\kappa^{kl}\kappa^{\beta\beta}\{2\kappa_{ik\beta}\kappa_{jl\beta} + 3\kappa_{ikl}\kappa_{j\beta\beta}\}, \quad i, j, k, l \neq \beta,$$

$$(b4) \quad \delta_{\beta\beta\beta\beta rs} = \frac{1}{12}(\kappa^{\beta\beta})^2\kappa^{rs}\{2\kappa_{\beta\beta r}\kappa_{\beta\beta s} + 3\kappa_{\beta\beta\beta}\kappa_{\beta rs}\}, \quad r, s \neq \beta,$$

$$(b5) \quad \delta_{\beta\beta kl\beta\beta} = \frac{1}{12}(\kappa^{\beta\beta})^2\kappa^{kl}\{2\kappa_{\beta k\beta}\kappa_{\beta l\beta} + 3\kappa_{\beta kl}\kappa_{\beta\beta\beta} - 12(\kappa_{\beta k\beta}\kappa_{\beta\beta}^{(l)} + \kappa_{\beta kl}\kappa_{\beta\beta}^{(\beta)})\}, \quad k, l \neq \beta,$$

$$(b6) \quad \delta_{ij\beta\beta\beta\beta} = \frac{1}{12}(\kappa^{\beta\beta})^2\kappa^{ij}\{5\kappa_{i\beta\beta}\kappa_{j\beta\beta}\}, \quad i, j \neq \beta.$$

Várias destas parcelas são também nulas, já que (ϕ_1, ϕ_2) e (ϕ_3, ϕ_4) são ortogonais e, portanto, $\kappa^{ij} = 0$ para $i \in \{\phi_1, \phi_2\}$, $j \in \{\phi_3, \phi_4\}$.

Sejam $\sum_L$ e $\sum_E$ os somatórios sobre os parâmetros de locação $\{\phi_1, \phi_2\}$ e escala $\{\phi_3, \phi_4\}$, respectivamente. Então, considerando os resultados anteriores, tem-se que d pode ser escrito como

$$\begin{aligned}
d = \{ \delta_{\beta\beta\beta\beta} - \delta_{\beta\beta\beta\beta\beta\beta} \} &+ \sum_L \{ \delta_{\pi(\beta\beta,ij)} - \delta_{\pi(\beta\beta,\beta\beta,ij)} - \delta_{\pi(\beta\beta,ij,ij)} \} \\
&+ \sum_E \{ \delta_{\pi(\beta\beta,ij)} - \delta_{\pi(\beta\beta,\beta\beta,ij)} - \delta_{\pi(\beta\beta,lj,ij)} \} \\
&- \sum_{i,j} \sum_{k,l} \delta_{\pi(\beta\beta,ij,kl)}
\end{aligned}$$

onde $\delta_{\pi(\beta\beta,ij)} = \delta_{\beta\beta ij} + \delta_{ij\beta\beta}$, $\delta_{\pi(\beta\beta,\beta\beta,ij)} = \delta_{\beta\beta\beta\beta ij} + \delta_{\beta\beta ij\beta\beta} + \delta_{lj\beta\beta\beta\beta}$ e assim sucessivamente.

No que segue usam-se algumas propriedades dos cumulantes do modelo que estuda-se. Tais propriedades seguem-se da simetria da distribuição t e da parametrização utilizada em (μ, Σ) , de modo que também valem para a classe geral dos modelos estruturais elípticos independentes definidos na Seção 5.2. Estas propriedades são:

- (i) todos os cumulantes independem do vetor de locação $\phi_L = (\phi_1, \phi_2)^\top$, consequentemente, as derivadas de qualquer cumulante com respeito a ϕ_1 ou ϕ_2 são todas nulas;
- (ii) qualquer cumulante que envolve um número ímpar de índices em $\{\phi_1, \phi_2\}$ é nulo, por exemplo, $\kappa_{\beta\beta i} = \kappa_{\beta\beta, i} = \kappa_{iii} = \kappa_{\beta i 3} = 0$, $i = 1, 2$;
- (iii) $\kappa_{ij}^{(\beta)} = \kappa_{\beta ij} = 0$, $i, j = 3, 4$ e $\kappa_{ii}^{(j)} = 0$, $i, j = 3, 4$, $i \neq j$.

Considerando estes fatos, tem-se, após uma longa álgebra, que

$$\begin{aligned}
(1) &= \sum_L \delta_{\pi(\beta\beta,ij)} = \frac{1}{2} \kappa^{\beta\beta} \left\{ \sum_{i=1}^2 \kappa^{ii} (\kappa_{\beta\beta ii} - 2\kappa_{\beta ii}^{(\beta)}) + 2\kappa^{12} (\kappa_{\beta\beta 12} - 2\kappa_{\beta 12}^{(\beta)}) \right\}; \\
(2) &= \sum_E \delta_{\pi(\beta\beta,ij)} = \frac{1}{2} \kappa^{\beta\beta} \left\{ \sum_{i=3}^4 \kappa^{ii} (\kappa_{\beta\beta ii} - 2\kappa_{\beta ii}^{(i)}) + 2\kappa^{34} (\kappa_{\beta\beta 34} - \kappa_{\beta\beta 3}^{(4)} - \kappa_{\beta\beta 4}^{(3)}) \right\}; \\
(3) &= \sum_L \delta_{\pi(\beta\beta,\beta\beta,ij)} = \frac{1}{2} (\kappa^{\beta\beta})^2 (\kappa_{\beta\beta\beta\beta} - 2\kappa_{\beta\beta\beta}^{(\beta)}) \left\{ \sum_{i=1}^2 \kappa^{ii} \kappa_{\beta\beta ii} + 2\kappa^{12} \kappa_{\beta 12} \right\}; \\
(4) &= \sum_E \delta_{\pi(\beta\beta,\beta\beta,ij)} = \frac{1}{4} (\kappa^{\beta\beta})^2 \left\{ \sum_{i=3}^4 \kappa^{ii} \kappa_{\beta\beta i} (3\kappa_{\beta\beta i} - 4\kappa_{\beta\beta}^{(i)}) \right. \\
&\quad \left. + 2\kappa^{34} (3\kappa_{\beta\beta 3} \kappa_{\beta\beta 4} - 2\kappa_{\beta\beta 3} \kappa_{\beta\beta}^{(4)} - 2\kappa_{\beta\beta 4} \kappa_{\beta\beta}^{(3)}) \right\};
\end{aligned}$$

$$\begin{aligned}
(5) = \sum_L \delta_{\pi(\beta\beta,ij,ij)} &= \frac{1}{4} \kappa^{\beta\beta} \left\{ \sum_{i=1}^2 (\kappa^{ii})^2 \kappa_{\beta ii} (3\kappa_{\beta ii} - 4\kappa_{ii}^{(\beta)}) \right. \\
&+ 4\kappa^{12} \left[(3\kappa_{\beta 12} - 2\kappa_{12}^{(\beta)}) \sum_{i=1}^2 \kappa^{ii} \kappa_{\beta ii} + 2\kappa_{\beta 12} \sum_{i=1}^2 \kappa^{ii} \kappa_{ii}^{(\beta)} \right] \\
&+ 4(\kappa^{12})^2 \left[\kappa_{\beta 11} \kappa_{\beta 22} - 2(\kappa_{\beta 11} \kappa_{22}^{(\beta)} + \kappa_{\beta 22} \kappa_{11}^{(\beta)}) + 2\kappa_{\beta 12} (\kappa_{\beta 12} - \kappa_{12}^{(\beta)}) \right] \\
&\left. + 2\kappa^{11} \kappa^{22} \left[\kappa_{\beta 11} \kappa_{\beta 22} + 2\kappa_{\beta 12} (\kappa_{\beta 12} - 2\kappa_{12}^{(\beta)}) \right] \right\};
\end{aligned}$$

$$\begin{aligned}
(6) = \sum_E \delta_{\pi(\beta\beta,ij,ij)} &= \frac{1}{2} \kappa^{\beta\beta} \left\{ \sum_{i=3}^4 (\kappa^{ii})^2 \kappa_{\beta\beta i} (\kappa_{iii} - 2\kappa_{ii}^{(i)}) \right. \\
&+ \kappa^{33} \kappa^{34} \left[\kappa_{\beta\beta 3} (3\kappa_{334} - 4\kappa_{34}^{(3)}) + \kappa_{\beta\beta 4} (\kappa_{333} - 2\kappa_{33}^{(3)}) \right] \\
&+ \kappa^{34} \kappa^{44} \left[\kappa_{\beta\beta 4} (3\kappa_{344} - 4\kappa_{34}^{(4)}) + \kappa_{\beta\beta 3} (\kappa_{444} - 2\kappa_{44}^{(4)}) \right] \\
&+ \kappa^{33} \kappa^{44} \left[\kappa_{\beta\beta 3} (\kappa_{344} - 2\kappa_{34}^{(4)}) + \kappa_{\beta\beta 4} (\kappa_{334} - 2\kappa_{34}^{(3)}) \right] \\
&\left. + 2(\kappa^{34})^2 \left[\kappa_{\beta\beta 3} (\kappa_{344} - \kappa_{34}^{(4)}) + \kappa_{\beta\beta 4} (\kappa_{334} - \kappa_{34}^{(3)}) \right] \right\};
\end{aligned}$$

$$\begin{aligned}
(7) = \sum_{i,j} \sum_L \delta_{\pi(\beta\beta,ij,kl)} &= \frac{1}{2} \kappa^{\beta\beta} \left\{ 2\kappa^{12} \sum_{i=3}^4 \kappa^{ii} \kappa_{\beta\beta i} \kappa_{12i} \right. \\
&+ \kappa^{34} \sum_{i=1}^2 \kappa^{ii} (\kappa_{\beta\beta 3} \kappa_{ii4} + \kappa_{\beta\beta 4} \kappa_{ii3}) + 2\kappa^{12} \kappa^{34} (\kappa_{\beta\beta 3} \kappa_{124} + 2\kappa_{\beta\beta 4} \kappa_{123}) \\
&\left. + \sum_{j=3}^4 \kappa^{jj} \kappa_{\beta\beta j} \sum_{i=1}^2 \kappa^{ii} \kappa_{ii j} \right\}.
\end{aligned}$$

Os cumulantes envolvidos nestas expressões são dados no Apêndice A.3 tanto para λ_x conhecido quanto para λ_e conhecido. Considerando o caso com λ_x conhecido, após substituir os cumulantes requeridos nas expressões dadas acima e de efetuar uma considerável álgebra, tem-se que:

$$\begin{aligned}
(0) &= \delta_{\beta\beta\beta\beta} - \delta_{\beta\beta\beta\beta\beta} = \frac{9}{2n} \frac{(\nu+4)^2}{(\nu+2)(\nu+6)(\nu+8)} ; \\
(1) &= \frac{1}{n} \frac{(\nu+4)^2(\nu+14)}{(\nu+2)(\nu+6)(\nu+8)} ; \\
(2) &= \frac{2}{n} \frac{(\nu+4)(\nu^2+17\nu+48)}{(\nu+2)(\nu+6)(\nu+8)} ; \quad (3) = 0 ; \\
(4) &= \frac{(\nu+4)(\nu^2+27\nu+90)}{(\nu+2)(\nu+6)^2} ; \quad (5) = \frac{1}{n} \frac{(\nu+4)^2(\nu+8)}{(\nu+2)(\nu+6)^2} ; \\
(6) &= \frac{2}{n} \frac{(\nu+4)(3\nu+4)}{(\nu+2)(\nu+6)^2} ; \quad (7) = -\frac{1}{n} \frac{(\nu+2)(\nu+4)}{(\nu+6)^2} .
\end{aligned}$$

Para λ_e conhecido (colocando sem perda de generalidade $\lambda_e = 1$), (1), (3), (5), (6) e (7) são análogos ao caso anterior, enquanto que para (0), (2) e (4) tem-se que

$$\begin{aligned}(0) &= \frac{9}{2n} \frac{(\nu+4)^2}{(\nu+2)(\nu+6)(\nu+8)} + \frac{1}{n} \frac{\nu+4}{\nu+2} (\phi_3 \phi_4) (\phi_3 - \phi_4)^{-2} ; \\(2) &= \frac{2}{n} \frac{(\nu+4)(\nu^2+17\nu+48)}{(\nu+2)(\nu+6)(\nu+8)} + \frac{4}{n} \frac{\nu+4}{\nu+2} (\phi_3 \phi_4) (\phi_3 - \phi_4)^{-2} ; \\(4) &= \frac{1}{2n} \frac{(\nu+4)(\nu^2+17\nu+90)}{(\nu+2)(\nu+6)^2} + \frac{5}{n} \frac{\nu+4}{\nu+2} (\phi_3 \phi_4) (\phi_3 - \phi_4)^{-2} .\end{aligned}$$

Considerando que $d = (0) + \{(1) - (3) - (5)\} + \{(2) - (4) - (6)\} - (7)$, obtém-se (5.71) após agrupar apropriadamente estas parcelas.

b) Teste do escore modificado

Sejam $Z_1, Z_2, \dots, Z_n$ iid $t_2(\mu(\phi_L), \Sigma(\phi_E); \nu)$, onde $\mu(\phi_L) = \phi_L$, $\Sigma(\phi_E)$ é definida em (5.28) e $\nu > 0$ é assumido conhecido. Considere a estatística escore modificada do teste de $H_0 : \beta = \beta_0$, ou seja,

$$S^* = \{1 + c + bS + aS^2\}^{-1} S ,$$

(ou $S^* = \{1 - c - bS - aS^2\}S$), onde S está dada em (5.70). Então, tanto para λ_x quanto λ_e conhecido, tem-se que

$$\begin{aligned}a &= 0 , \\b &= -\frac{1}{2n} \left\{ 1 - 4 \frac{\nu^2 + 17\nu + 54}{(\nu+6)^2(\nu+8)} \right\} , \\c &= \frac{5}{2n} \left\{ 1 + \frac{4}{5} \frac{\nu^5 - 37\nu^4 - 344\nu^3 - 808\nu^2 - 312\nu + 384}{\nu(\nu+2)^2(\nu+6)^2(\nu+8)} \right\} .\end{aligned} \tag{5.75}$$

Em particular, para o modelo normal ($\nu = \infty$) estes coeficientes se reduzem a

$$a_\infty = 0 , \quad b_\infty = -\frac{1}{2n} , \quad c_\infty = \frac{5}{2n} ,$$

de modo que neste modelo a estatística modificada é da forma

$$S_\infty^* = \left\{ 1 + \frac{5}{2n} - \frac{1}{2n} S_\infty \right\}^{-1} S_\infty$$

com S_∞ sendo dada por (5.47). Contudo, neste caso tanto b_∞ quanto c_∞ são modificados de forma significativa para valores pequenos a moderados de ν . Por exemplo, no modelo Cauchy ($\nu = 1$) tem-se que $b = b_1 = 0.347b_\infty$ e $c = c_1 = 0.775b_\infty$. Finalmente, deve-se notar também que $b = b_\nu < 0$ e $c = c_\nu > 0$, para todo $\nu > 0$.

Para provar (5.75) note-se em primeiro lugar que as matrizes $A_0 = [\alpha_{ij}] = \text{Diag} \{K_{11}^{-1}, 0\}$ e $C = [\gamma_{ij}] = K^{-1} - A_0$, são dadas por

$$\alpha_{ij} = \begin{cases} 0, & i, j = \beta \\ -\kappa^{ij}, & i, j \neq \beta \end{cases} \quad \text{e} \quad \gamma_{ij} = \begin{cases} -\kappa^{\beta\beta}, & i, j = \beta \\ 0, & i, j \neq \beta. \end{cases}$$

Consequentemente, de (4.55) - (4.57) segue que

$$\begin{aligned} a_1 &= -3\kappa^{\beta\beta} \sum_\phi (\kappa_{ij\beta} + 2\kappa_{i,j\beta}) (\kappa_{\beta st} + 2\kappa_{\beta s,t}) \kappa^{ij} \kappa^{st} \\ &\quad + 6\kappa^{\beta\beta} \sum_\phi (\kappa_{ijk} + 2\kappa_{i,j,k}) \kappa_{r,\beta,\beta} \kappa^{ij} \kappa^{kr} \\ &\quad - 6\kappa^{\beta\beta} \sum_\phi (\kappa_{\beta,jk} - \kappa_{\beta,j,k}) (\kappa_{\beta st} + 2\kappa_{\beta s,t}) \kappa^{js} \kappa^{kt} \\ &\quad - 6\kappa^{\beta\beta} \sum_\phi (\kappa_{\beta,\beta,k,r} + \kappa_{\beta,\beta,kr}) \kappa^{kr}; \\ a_2 &= 3(\kappa^{\beta\beta})^2 \sum_\phi \kappa_{\beta,\beta,k} \kappa_{r,\beta,\beta} \kappa^{kr} - 6(\kappa^{\beta\beta})^2 \kappa_{\beta,\beta,\beta} \sum_\phi (\kappa_{ij\beta} + 2\kappa_{i,j\beta}) \kappa^{ij} \\ &\quad + 6(\kappa^{\beta\beta})^2 \sum_\phi \kappa_{\beta,\beta,k} \kappa_{\beta,\beta,t} \kappa^{kt} + 3(\kappa^{\beta\beta})^2 \kappa_{\beta,\beta,\beta,\beta}; \\ a_3 &= -5(\kappa^{\beta\beta})^3 (\kappa_{\beta,\beta,\beta})^2. \end{aligned}$$

Os cumulantes envolvidos nestas expressões estão dados no Apêndice A.3. A simetria do modelo juntamente com as propriedades da parametrização utilizada para (μ, Σ) implicam que muitos destes cumulantes são nulos. Por exemplo,

$$(i) \quad \kappa_{i,\beta,\beta} = \kappa_{i\beta,\beta} = \kappa_{i\beta\beta} = 0, \quad i = 1, 2;$$

$$(ii) \quad \kappa_{i,j,\beta} = \kappa_{i,j\beta} = \kappa_{ij,\beta} = \kappa_{ij\beta} = 0, \quad i, j = 3, 4.$$

Além disto, como o modelo t é regular, então várias relações entre cumulantes se verificam. Por exemplo, $\kappa_{i,j} = -\kappa_{ij}$ e $\kappa_{ij}^{(k)} = \kappa_{ijk} + \kappa_{ij,k}$ implicam que $\kappa_{i,j,k} = 2\kappa_{ijk} - (\kappa_{ij}^{(k)} + \kappa_{ik}^{(j)} + \kappa_{jk}^{(i)})$.

Lembrando também que $\kappa_{ij} = 0$, $i = 1, 2$, $j = 3, 4$ e $\kappa_{\beta i} = 0$, $i = 1, 2, 3, 4$, tem-se, em particular, que

$$(iii) \quad \kappa_{i,j\beta} = -\kappa_{ij\beta}, \quad \kappa_{i,jk} = -\kappa_{ijk}, \quad i, j = 1, 2, \quad k = 3, 4, \quad \kappa_{\beta,jk} = -\kappa_{\beta jk}, \quad j, k = 3, 4,$$

$$(iv) \quad \kappa_{\beta,\beta,\beta} = 2\kappa_{\beta\beta\beta} - 3\kappa_{\beta,\beta}^{(\beta)} = 0.$$

Usando estes resultados e substituindo os cumulantes relevantes, tanto para λ_x quanto λ_e conhecido, segue-se, após muita álgebra, que

$$\begin{aligned} a_1 &= \left\{ -3\kappa^{\beta\beta} (\sum_L \kappa_{ij\beta} \kappa^{ij})^2 + 6\kappa^{\beta\beta} \sum_L (\kappa_{\beta,jk} - \kappa_{\beta,j,k}) \kappa_{\beta st} \kappa^{js} \kappa^{kt} \right. \\ &\quad \left. - 6\kappa^{\beta\beta} \sum_L (\kappa_{\beta,\beta,k,r} + \kappa_{\beta,\beta,kr}) \kappa^{kr} \right\} + \left\{ -6\kappa^{\beta\beta} \sum_E \kappa_{ijk} \kappa_{r,\beta,\beta} \kappa^{ij} \kappa^{kr} \right. \\ &\quad \left. - 6\kappa^{\beta\beta} \sum_E (\kappa_{\beta,\beta,k,r} + \kappa_{\beta,\beta,kr}) \kappa^{kr} \right\} + \left\{ -6\kappa^{\beta\beta} \sum_{ij} \sum_{kr} \kappa_{ijk} \kappa_{r,\beta,\beta} \kappa^{ij} \kappa^{kr} \right\} \\ &= \left\{ 0 + \frac{12}{n} \frac{\nu(\nu+4)^2}{(\nu+2)(\nu+6)^2} - \frac{24}{n} \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} \right\} \\ &\quad + \left\{ -\frac{48}{n} \frac{(\nu+4)(3\nu^2+9\nu+\nu-6)}{\nu(\nu+2)^2(\nu+6)^2} + \frac{24}{n} \frac{(\nu+4)(11\nu+16)}{(\nu+2)(\nu+6)(\nu+8)} \right\} \\ &\quad + \left\{ \frac{24}{n} \frac{(\nu+2)(\nu+4)}{(\nu+6)^2} \right\} \\ &= \frac{12(\nu+4)(\nu^5+28\nu^4+156\nu^3+340\nu^2+408\nu+192)}{n\nu(\nu+2)^2(\nu+6)^2(\nu+8)} ; \end{aligned}$$

$$\begin{aligned} a_2 &= 9(\kappa^{\beta\beta})^2 \sum_E \kappa_{\beta,\beta,k} \kappa_{\beta,\beta,r} \kappa^{k,r} + 3(\kappa^{\beta\beta})^2 \kappa_{\beta,\beta,\beta,\beta} \\ &= -\frac{36}{n} \frac{\nu(\nu+4)}{(\nu+6)^2} + \frac{9}{n} \left(3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right) \\ &= -\frac{9}{n} \left\{ 1 + \left(\frac{\nu+4}{\nu+6} \right) \frac{\nu^2+8\nu-36}{(\nu+6)(\nu+8)} \right\} ; \end{aligned}$$

$$a_3 = 0 .$$

Aplicando estes resultados em (4.54) obtem-se (5.75) após realizar uma certa álgebra.

(c) Um estudo de simulação

Finalmente, para completar o estudo feito nesta seção, são dados no que segue alguns resultados numéricos com o propósito de ilustrar o desempenho das estatísticas G e S e suas versões modificadas. Tais resultados são apresentados na Tabela 5.1 e foram obtidos a partir de 10.000 réplicas simuladas de $Z_1, Z_2, \dots, Z_n$ iid $t_2(\mu, \Sigma; \nu)$ e para quatro tamanhos amostrais diferentes, digamos, $n = 10, 20, 30$ e 40 . Fixando $\mu = \mu(\phi_L)$, $\Sigma = \Sigma(\phi_E)$ e ν , o i -ésimo valor de $Z = (Y, X) \sim t_2(\mu, \Sigma; \nu)$ é gerado usando que $Z =^d \mu + V^{1/2}AW$, onde $V \sim \nu/\chi_\nu^2$ é independente de $W = (W_1, W_2)^T \sim N_2(0, I_2)$ e $AA^T = \Sigma$. Isto é, gerando valores para $V \sim \nu/\chi_\nu^2$, $W_1 \sim N(0, 1)$ e $W_2 \sim N(0, 1)$ em forma independente, os respectivos valores de Y e X são obtidos como

$$Y = \mu_Y + \sqrt{V}(a_{11}W_1 + a_{12}W_2) \quad \text{e} \quad X = \mu_X + \sqrt{V}(a_{21}W_1 + a_{22}W_2),$$

onde $(\mu_Y, \mu_X)^T = \mu$ e $[a_{ij}] = A = \Sigma^{1/2}$. Assim, o procedimento é repetido para diferentes valores de ν , mas para um único valor do vetor de parâmetros do modelo. Por conveniência, tal valor foi escolhido como

$$\phi_L = (\phi_1, \phi_2)^T = (0, 0)^T \quad \text{e} \quad \phi_E = (\phi_3, \phi_4, \beta)^T = (4, 1/2, 1)^T.$$

Além disso, os dados são gerados a partir do modelo que supõe λ_e conhecido e colocando $\lambda_e = 1$ (ou $\sigma_{ee} = \sigma_{uu}$). Contudo, considerando o valor do vetor de parâmetros usado, de (5.26) segue que esta situação implica que $\lambda_x = (\phi_3 - \phi_4)\phi_4^{-1}(\beta^2 + 1)^{-1} = 7/2$. Finalmente, de (5.27) e (5.28) segue que $\mu(\phi_L)$ e $\Sigma(\phi_E)$ são dados por

$$\mu(\phi_L) = \phi_L = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{e} \quad \Sigma(\phi_E) = \begin{pmatrix} 9/4 & 7/4 \\ 7/4 & 9/4 \end{pmatrix}.$$

A Tabela 5.1 mostra que, em geral, os tamanhos das estatísticas modificadas estão efetivamente mais próximas dos respectivos níveis nominais em qualquer uma das situações consideradas. Note-se que a efetividade do fator de Bartlett, para corrigir a estatística da razão de verossimilhança G , é particularmente clara no caso de amostras pequenas.

Finalmente, convém mencionar que o algoritmo EM apresenta um bom desempenho na obtenção das estimativas de máxima verossimilhança dos parâmetros do modelo.

Tabela 5.1 Tamanhos do teste de $H_0 : \beta = 0$ (porcentagem de rejeições em 10.000 réplicas)

ν	n	Tamanho nominal	G	Tamanho do teste G^*	S	S^*
1	10	10.0	14.8	9.4	11.3	8.4
		5.0	8.1	4.7	5.6	4.2
		2.5	4.8	2.4	2.9	2.1
		1.0	2.3	0.9	0.9	0.7
	20	10.0	13.0	10.6	11.3	10.0
		5.0	6.9	5.0	5.6	4.8
		2.5	3.7	2.5	2.7	2.2
		1.0	1.6	1.2	1.2	1.0
	30	10.0	12.1	10.5	11.1	10.1
		5.0	6.4	5.4	5.7	5.1
		2.5	3.3	2.7	2.8	2.6
		1.0	1.3	1.0	1.0	0.9
	40	10.0	10.7	9.7	10.0	9.4
		5.0	5.3	4.7	4.9	4.6
		2.5	2.7	2.3	2.3	2.2
		1.0	1.1	0.9	0.9	0.8
3	10	10.0	14.6	10.0	12.1	10.0
		5.0	8.5	5.2	6.6	5.3
		2.5	5.1	2.5	3.2	2.6
		1.0	2.4	1.1	1.1	1.0
	20	10.0	12.1	10.0	10.9	10.0
		5.0	6.8	5.0	5.7	5.1
		2.5	3.6	2.6	2.9	2.6
		1.0	1.5	1.0	1.0	0.9
	30	10.0	12.2	10.8	11.4	10.6
		5.0	6.2	5.2	5.6	5.2
		2.5	3.2	2.6	2.7	2.5
		1.0	1.5	1.1	1.1	1.0
	40	10.0	10.9	10.0	10.4	9.8
		5.0	5.7	5.0	5.4	4.9
		2.5	2.9	2.5	2.6	2.5
		1.0	1.3	1.0	1.0	1.0
50	10	10.0	15.4	10.0	12.8	10.1
		5.0	8.9	4.9	5.8	4.8
		2.5	5.1	2.5	2.3	2.2
		1.0	2.6	0.9	0.5	0.6
	20	10.0	12.3	10.0	11.2	9.9
		5.0	6.6	4.8	5.3	4.8
		2.5	3.5	2.4	2.5	2.4
		1.0	1.6	1.0	0.9	1.0
	30	10.0	11.4	10.1	10.7	10.1
		5.0	6.0	4.9	5.2	4.9
		2.5	3.2	2.6	2.7	2.6
		1.0	1.3	1.0	0.9	1.0
	40	10.0	11.3	10.1	10.8	10.1
		5.0	5.8	5.0	5.1	4.9
		2.5	2.9	2.4	2.5	2.4
		1.0	1.3	1.0	1.0	1.0

Apêndice A

Neste apêndice são desenvolvidas expressões gerais para determinar os cumulantes conjuntos de derivadas da função log-verossimilhança de um modelo t p -variado com vetor de locação $\eta = \eta(\theta_L)$, matriz de dispersão $\psi = \psi(\theta_E)$ e $\nu > 0$ graus de liberdade, onde $\theta_L = (\theta_1, \theta_2, \dots, \theta_r)^\top$ e $\theta_E = (\theta_{r+1}, \theta_{r+2}, \dots, \theta_s)^\top$ são vetores de parâmetros de locação e escala, respectivamente, de modo que são ortogonais. A partir destes resultados gerais, os cumulantes relativos aos modelos estudados nos Capítulos 4 e 5 podem ser deduzidos como casos particulares.

A.1 Cumulantes conjuntos de derivadas da função de log-verossimilhança de um modelo t -elíptico p -variado com ν graus de liberdade

Seja $T = (T_1, T_2, \dots, T_p)^\top \sim t_p(0, I_p; \nu)$, isto é, $T \sim El_p(0, I_p; G)$, onde $f(\|t\|^2)$, $\|t\|^2 = t^\top t$, $t \in \mathbb{R}^p$, é a função de densidade de T para f dada por

$$f(u) = k(p, \nu) \nu^{\frac{1}{2}\nu} \{\nu + u\}^{-\frac{1}{2}(\nu+p)}, \quad u = \|t\|^2 \geq 0, \quad (\text{A.1})$$

onde

$$k(p, \nu) = \frac{\Gamma[\frac{1}{2}(\nu + p)]}{\Gamma[\frac{1}{2}\nu] \pi^{\frac{1}{2}p}}.$$

Seja

$$g(u) = \log f(u).$$

Então g é infinitamente diferenciável e sua j -ésima derivada é dada por

$$g^{(j)}(u) = (-1)^j [(j-1)!] \left(\frac{\nu+p}{2} \right) \{\nu+u\}^{-j}, \quad j = 1, 2, \dots$$

Note que $g^{(j+1)}$ pode ser expressa em termos de $g^{(i)}$ como

$$g^{(j+1)}(u) = (j)! \left(\frac{\nu+p}{2} \right)^{-j} \{g^{(1)}(u)\}^{j+1}, \quad j = 1, 2, \dots, \quad (\text{A.2})$$

onde

$$g^{(1)}(u) = -\frac{\nu+p}{2} \{\nu+u\}^{-1}.$$

Considere agora a variável aleatória $W = g^{(1)}(\|T\|^2)$ e defina os momentos

$$a(i, j) = E\{\|T\|^{2i} W^j\}.$$

De (3.27) segue que estes momentos estão definidos para todo i e j tais que $\nu+2(j-1) > 0$.

Assim, para $j \geq i$ tem-se que (veja (5.59))

$$a(i, j) = \left(-\frac{1}{2}\right)^j \mu_i(\chi_p^2) \left(\frac{\nu+p}{\nu}\right)^{j-i} \frac{(\nu+p)^i \mu_{j-i}(\chi_\nu^2)}{\mu_j(\chi_{\nu+p}^2)},$$

se $p \geq 2$ ($\mu_i(\chi_p^2) \equiv 1$ se $p = 1$), onde

$$\mu_0(\chi_s^2) = 1 \quad \text{e}$$

$$\mu_k(\chi_s^2) = E\{\chi_s^{2k}\} = s(s+2) \cdots (s+2(k-1)), \quad k \geq 1,$$

é o k -ésimo momento da distribuição qui-quadrado com s graus de liberdade. Em particular, para o modelo normal ($\nu = \infty$) tem-se que

$$a(i, j) = \left(-\frac{1}{2}\right)^j \mu_i(\chi_p^2).$$

Alguns valores especiais de $a(i, j)$ são dados na seguinte tabela.

Tabela A.1. Valores de $a(i, j)$, $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$, $j \geq i$.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	$-\frac{p}{2}$	$\frac{p}{4} \frac{\nu+p}{\nu+p+2}$	$-\frac{p}{8} \left(\frac{\nu+p}{\nu}\right)^2 \frac{\nu(\nu+2)}{(\nu+p+2)(\nu+p+4)}$	$\frac{p}{16} \left(\frac{\nu+p}{\nu}\right)^3 \frac{\nu(\nu+2)(\nu+4)}{(\nu+p+2)(\nu+p+4)(\nu+p+6)}$
$i = 2$	-	$\frac{p(p+2)}{4} \frac{\nu+p}{\nu+p+2}$	$-\frac{p(p+2)}{8} \frac{(\nu+p)^2}{(\nu+p+2)(\nu+p+4)}$	$\frac{p(p+2)}{16} \left(\frac{\nu+p}{\nu}\right)^2 \frac{(\nu+p)\nu(\nu+2)}{(\nu+p+2)(\nu+p+4)(\nu+p+6)}$
$i = 3$	-	-	$-\frac{p(p+2)(p+4)}{8} \frac{(\nu+p)^2}{(\nu+p+2)(\nu+p+4)}$	$\frac{p(p+2)(p+4)}{16} \left(\frac{\nu+p}{\nu}\right) \frac{(\nu+p)^2 \nu}{(\nu+p+2)(\nu+p+4)(\nu+p+6)}$
$i = 4$	-	-	-	$\frac{p(p+2)(p+4)(p+6)}{16} \frac{(\nu+p)^3}{(\nu+p+2)(\nu+p+4)(\nu+p+6)}$

Suponha agora que

$$Y \sim t_p(\eta, \psi; \nu) ,$$

onde $\eta = \eta(\theta_L) \in \mathbb{R}^p$, $\psi = \psi(\theta_E) > 0$, $\theta_L = (\theta_1, \dots, \theta_r)^\top$ e $\theta_E = (\theta_{r+1}, \dots, \theta_s)^\top$. Seja $\theta = (\theta_L^\top, \theta_E^\top)^\top$. Então, Y tem função densidade dada por

$$P_X(y|\theta, \nu) = |\psi|^{-1/2} f((y - \eta)^\top \psi^{-1} (y - \eta)) , \quad y \in \mathbb{R}^p ,$$

onde $f(u)$, $u \geq 0$, está dada em (A.1). Daqui segue que

$$T = \psi^{-1/2} (Y - \eta) \sim t_p(0, I_p; \nu) ,$$

e tem densidade $f(\|t\|^2)$, $t \in \mathbb{R}^p$. Lembre também que

$$\|T\|^2 = (Y - \eta)^\top \psi^{-1} (Y - \eta) \sim pF_{p, \nu}$$

e é independente de (veja os Capítulos 2 e 3)

$$U = \frac{T}{\|T\|} \sim \mathcal{U}(s_p) .$$

Seja $\ell(\theta, \nu) = \log p_Y(y|\theta, \nu)$. Então,

$$\ell(\theta, \nu) = -\frac{1}{2} \log |\psi| + g(\|t\|^2) ,$$

onde $g(u) = \log f(u)$ e $\|t\|^2 = (y - \eta)^\top \psi^{-1}(y - \eta)$. Defina a variável aleatória

$$\ell = -\frac{1}{2} \log |\psi| + g(\|T\|^2), \quad (\text{A.4})$$

e considere os seguintes fatos

$$\frac{\partial \log |\psi|}{\partial \theta_i} = \text{tr} \left(\psi^{-1} \frac{\partial \psi}{\partial \theta_i} \right), \quad \frac{\partial \psi^{-1}}{\partial \theta_i} = -\psi^{-1} \frac{\partial \psi}{\partial \theta_i} \psi^{-1}, \quad \theta_i \in \Theta_E, \quad (\text{A.5})$$

e

$$\begin{aligned} \frac{\partial \|T\|^2}{\partial \theta_i} &= \begin{cases} -2 \frac{\partial \eta^\top}{\partial \theta_i} \psi^{-1} (Y - \eta) \\ -(Y - \eta)^\top \psi^{-1} \frac{\partial \psi}{\partial \theta_i} \psi^{-1} (Y - \eta) \end{cases} \\ &= \begin{cases} -2 \frac{\partial \eta^\top}{\partial \theta_i} \psi^{-1/2} T & , \theta_i \in \theta_L, \\ -T^\top \psi^{-1/2} \frac{\partial \psi}{\partial \theta_i} \psi^{-1/2} T & , \theta_i \in \theta_E. \end{cases} \end{aligned} \quad (\text{A.6})$$

Se φ é uma função de $\theta = (\theta_1, \theta_2, \dots, \theta_s)^\top$ e ν , denote suas derivadas por

$$\varphi_i = \frac{\partial \varphi}{\partial \theta_i}, \quad \varphi_\nu = \frac{\partial \varphi}{\partial \nu}, \quad \varphi_{ij} = \frac{\partial^2 \varphi}{\partial \theta_i \partial \theta_j}, \quad \text{etc.}$$

Assim, $\eta_i = \partial \eta / \partial \theta_i$, $\psi_i = \partial \psi / \partial \theta_i$, etc. Note-se que $\eta_i = 0$ se $\theta_i \in \theta_E$ e $\psi_i = 0$ se $\theta_L \in \theta_L$.

Defina também

$$e_i = \psi^{-1/2} \eta_i, \quad \psi_i = \psi^{-1/2} \psi_i \psi^{-1/2}, \quad E_{ij} = \psi^{-1/2} \psi_{ij} \psi^{-1/2}, \quad \text{etc.}$$

Usando esta notação juntamente com (A.5), então de (A.4) segue que a derivada de ℓ com respeito a θ_i é dada por

$$\begin{aligned} \ell_i &= \begin{cases} -2g^{(1)}(\|T\|^2) \eta_i^\top \psi^{-1} (Y - \eta) & , \theta_i \in \theta_L, \\ -\frac{1}{2} \text{tr}(\psi^{-1} \psi_i) - g^{(1)}(\|T\|^2) (Y - \eta)^\top \psi^{-1} \psi_i \psi^{-1} (Y - \eta) & , \theta_i \in \theta_E, \end{cases} \\ &= \begin{cases} -2g^{(1)}(\|T\|^2) e_i^\top T & , \theta_i \in \theta_L, \\ -\frac{1}{2} \text{tr}(E_i) - g^{(1)}(\|T\|^2) T^\top E_i T & , \theta_i \in \theta_E, \end{cases} \end{aligned} \quad (\text{A.7})$$

onde $g^{(1)}(\|T\|^2) = W = -(\frac{\nu+2}{2})\{\nu + \|T\|^2\}^{-1}$.

Suponha agora que $\eta = \eta(\theta_L)$ é linear em θ_L , de modo que $\eta_{ij} = \eta_i^\top \eta_j = 0$, $i \neq j$. Então, derivando na primeira expressão de (A.7) com respeito a θ_j e considerando (A.5) e (A.6), tem-se que

$$\ell_{ij} = \begin{cases} 4g^{(2)}(\|T\|^2)T^\top e_j e_i^\top T + 2g^{(1)}(\|T\|^2)e_i^\top e_j, & \theta_i, \theta_j \in \theta_L, \\ 2g^{(2)}(\|T\|^2)(T^\top E_j T)e_i^\top T + 2g^{(1)}(\|T\|^2)e_i^\top E_j T, & \theta_i \in \theta_L, \theta_j \in \theta_E \\ -\frac{1}{2}\text{tr}(E_i E_j) - \frac{1}{2}\text{tr}(E_{ij}) + g^{(2)}(\|T\|^2)(T^\top E_i T)(T^\top E_j T) \\ + 2g^{(1)}(\|T\|^2)T^\top E_i E_j T - g^{(1)}(\|T\|^2)T^\top E_{ij} T, & \theta_i, \theta_j \in \theta_E, \end{cases} \quad (\text{A.8})$$

onde $g^{(1)}(\|T\|^2) = W = -(\frac{\nu+p}{2})\{\nu + \|T\|^2\}^{-1}$ e $g^{(2)}(\|T\|^2) = (\frac{\nu+p}{2})^{-1}W_1^2$ (veja (A.2)).

Defina agora a seguinte notação,

$$\begin{aligned} \lambda_i &= E(\ell_i), \quad \lambda_{ij} = E(\ell_{ij}), \quad \text{etc.}, \\ \lambda_{i,j} &= E(\ell_i \ell_j), \quad \lambda_{i,jk} = E(\ell_i \ell_j \ell_k), \quad \text{etc.}, \\ \lambda_i^{(ij)} &= \frac{\partial \lambda_i}{\partial \theta_j}, \quad \lambda_{ij}^{(k)} = \frac{\partial \lambda_{ij}}{\partial \theta_k}, \quad \text{etc.}, \end{aligned}$$

e, lembre que $\|T\|$ e $U = T/\|T\|$ são independentes. Então, usando (A.7), (2.73) e (2.80) (ou (2.81)) é fácil ver que

$$(a) \quad e_i \neq 0, \theta_i \in \theta_L \text{ implica que } \lambda_i = 0 \Leftrightarrow E(\|T\|g^{(1)}(\|T\|^2)) = a(\frac{1}{2}, 1) < \infty,$$

$$(b) \quad \text{tr}(E_i) \neq 0, \theta_i \in \theta_E \text{ implica que } \lambda_i = 0 \Leftrightarrow E(\|T\|^2 g^{(1)}(\|T\|^2)) = a(1, 1) = -p/2.$$

Por exemplo, considerando a segunda expressão de (A.7) e usando (2.80), (b) segue de observar que $\lambda_i = -\frac{1}{2}\text{tr}(E_i) - \frac{1}{p}a(1, 1)\text{tr}(E_i)$. De (A.3) segue que (a) e (b) valem no modelo t .

O fato de que $\lambda_i = 0$ implica que $\lambda_i^{(j)} = 0$, de onde é fácil ver que $\lambda_{i,j} = -\lambda_{i,j}$. Este resultado é bem conhecido e se verifica para todos os modelos regulares, como o modelo que estuda-se. O que importa notar aqui é que $\lambda_{ij} = -\lambda_{i,j}$ pode ser derivado sem requerer a segunda derivada ℓ_{ij} . Similarmente, considerando que também valem outras relações entre os “cumulantes” λ 's (veja a Seção 4.4.1), então λ_{ijk} pode ser obtido a partir de $\lambda_{i,jk}$

e $\lambda_{i,j}^{(k)}$ sem necessitar a terceira derivada ℓ_{ijk} e assim sucessivamente. O total de derivadas de ℓ que são requeridas depende da ordem desejada nestes cumulantes. Para ilustrar, considere a derivação de λ_{ijk} . Note em primeiro lugar que

$$\frac{\partial p_Y(y|\theta, \nu)}{\partial \theta_i} = p_Y(y|\theta, \nu) \ell_i .$$

Com esta relação é fácil ver que

$$\lambda_{ij} = \int_{\mathbb{R}^n} l_{ij} p_Y(y|\theta, \nu) dy \Rightarrow \lambda_{ij}^{(k)} = \lambda_{ijk} + \lambda_{ij,k} \Rightarrow \lambda_{ijk} = \lambda_{ij}^{(k)} - \lambda_{ij,k} ,$$

de modo que somente a primeira e segunda derivada de ℓ são requeridas para obter λ_{ijk} . Com este procedimento derivam-se no que segue alguns cumulantes λ 's de até quarta ordem, a partir das quais podem ser obtidos os cumulantes κ 's utilizados nos Capítulos 4 e 5. As relações (2.84) – (2.90) do Capítulo 2 são de grande utilidade nos resultados que seguem (veja, também, Mitchell (1989)). Convém notar que estes resultados podem ser aplicados a qualquer modelo elíptico p -variado regular avaliando adequadamente (A.3).

A.1.1 Alguns cumulantes para parâmetros de locação

Usando a primeira expressão de (A.7) juntamente com (2.80) – (2.82) e (2.90), tem-se que

$$(1) \quad -\lambda_{ij} = \lambda_{i,j} = \frac{4}{p} a(1, 2) e_i^\top e_j, \text{ onde } e_i = \psi^{-1/2} \frac{\partial \eta}{\partial \theta_i},$$

$$(2) \quad \lambda_{i,j,k} = 0 \text{ (note que este resultado segue da simetria da distribuição de } T),$$

$$(3) \quad \lambda_{i,i,i,i} = \frac{16 \times 105}{p(p+2)(p+4)} (a(1, 1))^4 (e_i^\top e_i)^4,$$

$$(4) \quad \lambda_{ij}^{(k)} = 0, \text{ uma vez que } \eta = \eta(\theta_L) \text{ é linear em } \theta_L.$$

De (A.7), (A.8) e da simetria da distribuição de T segue que

$$(5) \quad \lambda_{ij,k} = 0 \text{ e } \lambda_{ijk} = \lambda_{ij}^{(k)} - \lambda_{ij,k} = 0.$$

Note que (A.8) implica que, em (1),

$$-\lambda_{ij} = \lambda_{i,j} \Leftrightarrow E(\|T\|^2 g^{(2)}(\|T\|^2)) + \frac{p}{2} a(0, 1) = -a(1, 2) = \frac{p}{2} .$$

A.1.2 Alguns cumulantes para parâmetros de escala

Usando a segunda expressão de (A.7) e que $a(1, 1) = -p/2$, segue que

$$\begin{aligned}
 (1) \quad -\lambda_{ij} &= \lambda_{i,j} = \left(\frac{a(2, 2)}{p(p+2)} - \frac{1}{4} \right) \text{tr}(E_i) \text{tr}(E_j) + 2 \frac{a(2, 2)}{p(p+2)} \text{tr}(E_i E_j), \\
 (2) \quad \lambda_{i,j,k} &= - \left(\frac{a(3, 3)}{p(p+2)(p+4)} + \frac{3}{2} \frac{a(2, 2)}{p(p+2)} - \frac{1}{4} \right) \text{tr}(E_i) \text{tr}(E_j) \text{tr}(E_k) \\
 &\quad - \left(2 \frac{a(3, 3)}{p(p+2)(p+4)} + \frac{a(2, 2)}{p(p+2)} \right) (\text{tr}(E_i) \text{tr}(E_j E_k) + \text{tr}(E_j) \text{tr}(E_i E_k) + \text{tr}(E_k) \text{tr}(E_i E_j)) \\
 &\quad - 8 \frac{a(3, 3)}{p(p+2)(p+4)} \text{tr}(E_i E_j E_k), \\
 (3) \quad \lambda_{i,i,i,i} &= \left(\frac{a(4, 4)}{p(p+2)(p+4)(p+6)} + 2 \frac{a(3, 3)}{p(p+2)(p+4)} + \frac{3}{2} \frac{a(2, 2)}{p(p+2)} - \frac{1}{16} \right) (\text{tr}(E_i))^4 \\
 &\quad \left(12 \frac{a(4, 4)}{p(p+2)(p+4)(p+6)} + 12 \frac{a(3, 3)}{p(p+2)(p+4)} + 3 \frac{a(2, 2)}{p(p+2)} \right) \text{tr}(E_i^2) (\text{tr}(E_i))^2 \\
 &\quad + \left(32 \frac{a(4, 4)}{p(p+2)(p+4)(p+6)} + 16 \frac{a(3, 3)}{p(p+2)(p+4)} \right) \text{tr}(E_i^3) \text{tr}(E_i) \\
 &\quad + 12 \frac{a(4, 4)}{p(p+2)(p+4)(p+6)} (\text{tr}(E_i^2))^2 + 48 \frac{a(4, 4)}{p(p+2)(p+4)(p+6)} \text{tr}(E_i^4),
 \end{aligned}$$

onde $\text{tr}(E_i) = \text{tr} \left(\psi^{-1} \frac{\partial \psi}{\partial \theta_i} \right)$ e $\text{tr}(E_i E_j) = \text{tr} \left(\psi^{-1} \frac{\partial \psi}{\partial \theta_i} \psi^{-1} \frac{\partial \psi}{\partial \theta_j} \right)$, etc.

Note que de (1) e da terceira expressão de (A.8) segue que

$$-\lambda_{ij} = \lambda_{i,j} \Leftrightarrow E(\|T\|^4 g^{(2)}(\|T\|^2)) = \frac{p(p+2)}{4} - a(2, 2).$$

Considerando esta relação e combinando a segunda expressão de (A.7) com a terceira de (A.8), tem-se que

$$\begin{aligned}
 (4) \quad \lambda_{ij,k} &= - \frac{E[\|T\|^6 g^{(1)}(\|T\|^2) g^{(2)}(\|T\|^2)]}{p(p+2)(p+4)} (8 \text{tr}(E_i E_j E_k) + 2 \text{tr}(E_i E_j) \text{tr}(E_k) \\
 &\quad + 2 \text{tr}(E_i E_k) \text{tr}(E_j) + 2 \text{tr}(E_j E_k) \text{tr}(E_i) + \text{tr}(E_i) \text{tr}(E_j) \text{tr}(E_k)) \\
 &\quad - \frac{1}{2} \frac{a(2, 2)}{p(p+2)} (8 \text{tr}(E_i E_j E_k) + 4 \text{tr}(E_i E_j) \text{tr}(E_k) - \text{tr}(E_i) \text{tr}(E_j) \text{tr}(E_k)) - \frac{1}{8} \text{tr}(E_i) \text{tr}(E_j) \text{tr}(E_k)
 \end{aligned}$$

De (A.2) e (A.3) segue que no modelo t ,

$$E(\|T\|^6 g^{(1)}(\|T\|^2) g^{(2)}(\|T\|^2)) = \frac{2}{\nu + p} a(3, 3) .$$

$$(5) \quad \lambda_{ijk} = \lambda_{ij}^{(k)} - \lambda_{ij,k} .$$

A.1.3 Alguns cumulantes mistos

De combinar as duas expressões de (A.7) tem-se que

(1) $\lambda_{ij} = -\lambda_{i,j} = 0$, $i \in \theta_L$, $j \in \theta_E$, resultado que é óbvio já que estamos supondo que θ_L e θ_E são ortogonais. Contudo, este fato segue da simetria dos modelos elípticos, a qual implica que $\eta(\theta_L)$ seja ortogonal a $\psi(\theta_E)$ e, portanto, θ_L com θ_E . De forma similar, tem-se que qualquer cumulante que envolva um número ímpar de parâmetros de locação é nulo. Assim, por exemplo,

$$(2) \quad \lambda_{i,j,k} = \lambda_{i,j,k} = \lambda_{i,j,k} = \lambda_{ijk} = 0, \text{ para todo } \theta_i \in \theta_L \text{ e } \theta_j, \theta_k \in \theta_E,$$

$$(3) \quad \lambda_{i,jkt} = 0, \theta_i \in \theta_L \text{ e } \theta_j, \theta_k, \theta_t \in \theta_E, \text{ etc.}$$

De (A.7) segue também que se $\theta_i, \theta_j \in \theta_L$ e $\theta_k \in \theta_E$, então

$$(4) \quad \lambda_{i,j,k} = - \left(2 \frac{a(1,2)}{p} + 4 \frac{a(2,3)}{p(p+2)} \right) e_i^\top e_j \text{tr}(E_k) - 8 \frac{a(2,3)}{p(p+2)} e_i^\top E_k e_j .$$

Combinando agora a primeira expressão de (A.7) com a segunda de (A.8) tem-se que

$$(5) \quad \lambda_{i,j,k} = -4 \frac{E[\|T\|^4 g^{(1)}(\|T\|^2) g^{(2)}(\|T\|^2)]}{p(p+2)} (2e_i^\top E_k e_j + e_i^\top e_j \text{tr}(E_k)) - 4 \frac{a(1,2)}{p} e_i^\top E_k e_j ,$$

$$(6) \quad \lambda_{ijk} = \lambda_{ij}^{(k)} - \lambda_{ij,k} .$$

Note que no modelo t ,

$$E[\|T\|^4 g^{(1)}(\|T\|^2) g^{(2)}(\|T\|^2)] = \frac{2}{\nu + p} a(1, 2) .$$

O material dado aqui é suficiente para obter os cumulantes requeridos nos Capítulos 4 e 5 quando ν é conhecido. Contudo, com estes cumulantes podem ser derivados vários

outros usando as relações entre eles. Para ν desconhecido, damos apenas os elementos de entrada da matriz de informação (veja também Lange et al. (1989)). Para tais efeitos, considere que

$$\begin{aligned}\ell_{\nu\nu} &= \frac{1}{2}\psi' \left[\frac{\nu+p}{2} \right] - \frac{1}{2}\psi' \left[\frac{1}{2}\nu \right] + \frac{1}{2\nu} + \frac{2}{\nu+p}g^{(1)}(\|T\|^2) + g^{(2)}(\|T\|^2) \\ \ell_{\nu i} &= \begin{cases} -\left(\frac{2}{\nu+p}g^{(1)}(\|T\|^2) + g^{(2)}(\|T\|^2)\right)e_i^\top T, & \theta_i \in \theta_L, \\ -\left(\frac{2}{\nu+p}g^{(1)}(\|T\|^2) + g^{(2)}(\|T\|^2)\right)T^\top E_i T, & \theta_i \in \theta_E, \end{cases}\end{aligned}$$

de modo que

$$\begin{aligned}\lambda_{\nu\nu} &= \frac{1}{2}\psi' \left[\frac{\nu+p}{2} \right] - \frac{1}{2}\psi' \left[\frac{1}{2}\nu \right] + \frac{1}{2\nu} + \frac{2}{\nu+p}(a(0,1) + a(0,2)) \\ \lambda_{\nu i} &= \begin{cases} 0, & \theta_i \in \theta_L, \\ -\frac{1}{p}\frac{2}{\nu+p}(a(0,1) + a(0,2))\text{tr}(E_i), & \theta_i \in \theta_E. \end{cases}\end{aligned}$$

Finalmente, se $Y_1, Y_2, \dots, Y_n$ são n réplicas independentes de Y e $Y_m \sim t_p(n_m, \psi_m; \nu)$.

Então, se L é a log-verossimilhança de Y_m , $i = 1, 2, \dots, n$, tem-se que

$$L = \sum_m \ell_{(m)},$$

e suas derivadas são da forma

$$L_i = \sum_m \ell_{(m)i}, \quad L_{ij} = \sum_m \ell_{(m)ij}, \quad \text{etc.}$$

Defina então os cumulantes κ 's como (veja a Seção 4.4.1),

$$\begin{aligned}\kappa_{i,j} &= E(L_i L_j) = \sum_m E(\ell_{(m)i} \ell_{(m)j}) + \sum_{m \neq m'} E(\ell_{(m)i}) E(\ell_{(m')j}) = \sum_m \lambda_{(m)i,j}, \\ \kappa_{i,jk} &= E(L_i L_{jk}) = \sum_m \lambda_{(m)i,jk}, \\ \kappa_{ij,kl} &= E(L_{ij} L_{kl}) - \kappa_{ij} \kappa_{kl} = \sum_m \lambda_{(m)ij,kl} + \sum_{m \neq m'} \lambda_{(m)ij} \lambda_{(m')kl} - \kappa_{ij} \kappa_{kl}, \quad \text{etc.}\end{aligned}$$

Se $\eta_m = \eta$ e $\psi_m = \psi$, $m = 1, 2, \dots, n$, então,

$$\begin{aligned}\kappa_{i,j} &= n\lambda_{i,j}, \quad \kappa_{i,jk} = n\lambda_{i,jk}, \\ \kappa_{ij,kl} &= n\lambda_{ij,kl} + n(n-1)\lambda_{ij}\lambda_{kl} - \kappa_{ij}\kappa_{kl} = n\lambda_{ij,kl} - \frac{1}{n}\kappa_{ij}\kappa_{kl}, \quad \text{etc.}\end{aligned}$$

A.2 Cumulantes para o modelo de regressão linear t -independente

Para este modelo tem-se que

$$Y \sim t_p(\eta(\theta_L), \psi(\theta_E); \nu)$$

com

$$p = 1, \quad \eta(\theta_L) = x^\top \beta = \sum_{j=1}^r x_j \beta_j \quad \text{e} \quad \psi(\theta_E) = \tau^2,$$

de modo que $\theta_L = \beta$ e $\theta_E = \tau^2$. Por comodidade vamos colocar $\tau^2 = \phi^{-1}$. Assim, de acordo com a notação de A.1, tem-se que

$$\begin{aligned} \eta_i &= \frac{\partial \eta}{\partial \beta_i} = x_i = (i) \quad e_i = \psi^{-\frac{1}{2}} \eta_i = \phi^{\frac{1}{2}}(i), \quad i = 1, 2, \dots, k, \\ \psi_\phi &= \frac{\partial \psi}{\partial \phi} = -\phi^{-2}, \quad E_\phi = \psi^{-\frac{1}{2}} \psi_\phi \psi^{-\frac{1}{2}} = -\phi^{-1}, \quad E_{\phi\phi} = \phi^{-2}, \quad \text{etc.} \end{aligned}$$

A função ℓ é dada por

$$\ell = \frac{1}{2} \log \phi + g(T^2),$$

onde

$$g(T^2) = \log k(1, \nu) + \frac{\nu}{2} \log \nu - \frac{\nu+1}{2} \log\{\nu + T^2\}$$

e

$$T^2 = \phi(Y - x^\top \beta)^2.$$

Assim, de (A.7) e (A.8) segue que

$$\begin{aligned} \ell_i &= -2\phi^{\frac{1}{2}}(i)g^{(1)}(T^2)T, \quad \ell_{ij} = 2\phi(ij)(2g^{(2)}(T^2)T^2 + g^{(1)}(T^2)), \\ \ell_\phi &= \frac{1}{2}\phi^{-1} + \phi^{-1}g^{(1)}(T^2)T^2, \quad \ell_{i\phi} = -2\phi^{-\frac{1}{2}}(i)(g^{(2)}(T^2)T^3 + g^{(1)}(T^2)T), \end{aligned}$$

onde $(i) = x_i$, $(ij) = x_i x_j$ e (veja (A.2))

$$g^{(1)}(T^2) = -\frac{\nu+1}{2}(\nu + T^2)^{-1}, \quad g^{(2)}(T^2) = \frac{\nu+1}{2}(\nu + T^2)^{-2}.$$

Com estas ferramentas e usando a notação dada na parte final de A.1, obtem-se os resultados que seguem.

A.2.1 Alguns cumulantes para os parâmetros de regressão

$$\begin{aligned}\kappa_{ij} &= -\kappa_{i,j} = -\phi \frac{\nu+1}{\nu+3} \sum_m (ij)_m, & \kappa_{ijk} &= \kappa_{i,jk} = \kappa_{i,j,k} = \kappa_{ij}^{(k)} = 0, \\ \kappa_{ijkt} &= 6\phi^2 \frac{(\nu+1)(\nu+2)}{\nu(\nu+5)(\nu+7)} \sum_m (ijkt)_m, & \kappa_{ij,kt} &= \frac{3(\nu+3)^2 - 8}{(\nu+2)(\nu+3)^2} \kappa_{ijkt}, \\ \kappa_{i,j,kt} &= \frac{(\nu+1)\{(\nu+2)^2 - 5\}}{(\nu+2)(\nu+3)^3} \kappa_{ijkt}, & \kappa_{i,j,k,t} &= -3\kappa_{i,j,kt}, \quad \kappa_{ij}^{(kt)} = \kappa_{ij}^{(t)} = 0.\end{aligned}$$

A.2.2 Alguns cumulantes para o parâmetro de escala

$$\begin{aligned}\kappa_{\phi\phi} &= \frac{n}{2} \frac{\nu}{\nu+2} \phi^{-2}, & \kappa_{\phi\phi}^{(\phi)} &= -2\phi^{-1} \kappa_{\phi\phi}, & \kappa_{\phi\phi}^{(\phi\phi)} &= 6\phi^{-2} \kappa_{\phi\phi}, \\ \kappa_{\phi,\phi\phi} &= \frac{6}{\nu+5} \phi^{-1} \kappa_{\phi\phi}, & \kappa_{\phi,\phi,\phi} &= \frac{\nu-1}{\nu+5} \phi^{-1} \kappa_{\phi\phi}, & \kappa_{\phi\phi\phi} &= -2\frac{\nu+8}{\nu+5} \phi^{-1} \kappa_{\phi\phi}, \\ \kappa_{\phi\phi\phi}^{(\phi)} &= -6\frac{\nu+8}{\nu+5} \phi^{-2} \kappa_{\phi\phi}, & \kappa_{\phi\phi\phi\phi} &= 6\frac{(\nu+7)(\nu+8)+15}{(\nu+5)(\nu+7)} \phi^{-2} \kappa_{\phi\phi}.\end{aligned}$$

A.2.3 Cumulantes mistos

$$\begin{aligned}\kappa_{ij}^{(\phi)} &= \phi^{-1} \kappa_{ij}, & \kappa_{ij\phi} &= \frac{\nu+2}{\nu+5} \phi^{-1} \kappa_{ij}, \\ \kappa_{i\phi} &= \kappa_{i\phi\phi} = \kappa_{i,\phi\phi} = \kappa_{i,\phi,\phi} = \kappa_{ij\phi}^{(k)} = \kappa_{ij\phi}^{(\phi)} = 0, & \kappa_{i,j,\phi} &= \frac{\nu-1}{\nu+5} \phi^{-1} \kappa_{ij}, \\ \kappa_{i,j\phi} &= -\kappa_{ij\phi}, & \kappa_{ij,\phi} &= \frac{3}{\nu+5} \phi^{-1} \kappa_{ij}, & \kappa_{ij\phi\phi} &= -\frac{6}{(\nu+7)} \phi^{-1} \kappa_{ij\phi}, \\ \kappa_{i\phi,j\phi} &= -\frac{(\nu+2)(\nu+4)}{(\nu+5)(\nu+7)} \phi^{-1} \kappa_{ij}, & \kappa_{i,j,\phi,\phi} &= \frac{4}{n} \frac{(\nu-3)(\nu^2+3\nu-1)}{\nu(\nu+5)(\nu+7)} \kappa_{ij} \kappa_{\phi\phi}, \\ \kappa_{\phi\phi,ij} &= \frac{12}{n} \frac{(\nu+1)^2-6}{\nu(\nu+5)(\nu+7)} \kappa_{ij} \kappa_{\phi\phi}.\end{aligned}$$

Convém mencionar que todos os cumulantes deste modelo foram obtidos diretamente a partir das respectivas derivadas da log-verossimilhança e foram conferidos usando as relações entre eles.

A.2.4 Alguns cumulantes mistos que envolvem o parâmetro ν

$$\kappa_{\nu\nu} = nD_{\nu}^{(2)} + \frac{n}{2} \frac{\nu+5}{\nu(\nu+1)(\nu+3)}, \quad \kappa_{\nu\nu}^{(\nu)} = nD_{\nu}^{(3)} - \frac{n}{2} \frac{2\nu^3+19\nu^2+40\nu+15}{\nu^2(\nu+1)^2(\nu+3)^2}, \quad \kappa_{\nu\nu}^{(\phi)} = 0,$$

$$\begin{aligned}
\kappa_{\nu\nu\nu} &= nD_{\nu}^{(3)} - \frac{n}{2} \frac{2\nu^3+16\nu^2+41\nu+31}{\nu^2(\nu+1)(\nu+3)(\nu+5)} , \quad \kappa_{\nu\nu\nu\nu} = nD_{\nu}^{(4)} + \frac{n}{2} \left(\frac{1}{\nu^3} + \frac{(\nu+2)(\nu+4)(\nu^2+7\nu+18)}{\nu^3(\nu+1)(\nu+3)(\nu+5)(\nu+7)} \right) , \\
\kappa_{\phi\phi}^{(\nu)} &= \frac{3}{\nu(\nu+3)} \kappa_{\phi\phi} , \quad \kappa_{\phi\phi\nu} = \frac{3(\nu-3)}{\nu(\nu+2)(\nu+5)} \kappa_{\phi\phi} , \quad \kappa_{\nu\nu\phi} = -4 \frac{\nu-1}{\nu^2(\nu+1)(\nu+5)} \phi \kappa_{\phi\phi} , \\
\kappa_{\nu\phi} &= -\frac{n}{(\nu+1)(\nu+3)} \phi^{-1} , \quad \kappa_{ij}^{(\nu)} = \frac{2}{(\nu+1)(\nu+3)} \kappa_{ij} , \quad \kappa_{ij\nu}^{(\phi)} = 2 \frac{(\nu-1)}{\nu(\nu+1)(\nu+5)} \phi^{-1} \kappa_{ij} , \\
\kappa_{ij\phi}^{(\nu)} &= \frac{5\nu^2+26\nu+29}{(\nu+3)^2(\nu+5)^2} \phi^{-1} \kappa_{ij} , \quad \kappa_{\nu\phi}^{(\nu)} = \frac{2}{(\nu+1)(\nu+3)} \kappa_{\nu\phi} , \quad \kappa_{\nu\phi}^{(\phi)} = \frac{n}{(\nu+1)(\nu+3)} \phi^{-2} , \\
\kappa_{\nu\nu j} &= 0 , \quad \kappa_{\nu\nu\nu j} = 0 , \quad \kappa_{ij\phi\nu} = -\frac{(\nu+2)(5\nu-1)}{\nu(\nu+1)(\nu+5)(\nu+7)} \phi^{-1} \kappa_{ij} , \quad \kappa_{\nu\nu ij} = -4 \frac{(\nu-2)(\nu+2)}{\nu^2(\nu+1)(\nu+5)(\nu+7)} \kappa_{ij} , \\
\kappa_{\nu ij} &= 2 \frac{\nu-1}{\nu(\nu+1)(\nu+5)} \kappa_{ij} , \quad \kappa_{\nu ij}^{(\nu)} = -2 \frac{2\nu^3+5\nu^2-16\nu-15}{\nu^2(\nu+1)(\nu+3)(\nu+5)^2} \kappa_{ij} .
\end{aligned}$$

$$\text{Aqui, } D_{\nu}^{(j)} = d_{\nu}^{(j)} \frac{\left\{ \log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) \right\}}{d\nu^j} .$$

A.3 Cumulantes para o modelo estrutural t independente

Para este modelo tem-se que

$$Y \sim t_p(\eta(\theta_L), \psi(\theta_E); \nu)$$

com

$$Y = Z , \quad p = 2 , \quad \eta(\theta_L) = \mu(\phi_L) \quad \text{e} \quad \psi(\theta_E) = \Sigma(\phi_E) ,$$

onde $\theta_L = \phi_L$, $\theta_E = \phi_E$, $\mu(\phi_L)$ e $\Sigma(\phi_E)$ estão definidos em (5.26), (5.27) e (5.28), respectivamente. Usando (5.27) e (5.28) segue, por exemplo, que

$$\eta_1 = \frac{\partial \mu}{\partial \phi_1} = (1, 0)^{\top} , \quad \eta_2 = \frac{\partial \mu}{\partial \phi_2} = (0, 1)^{\top} , \quad \psi_i = \frac{\partial \Sigma}{\partial \phi_i} = \alpha_i \alpha_i^{\top} , \quad i = 3, 4, \text{ etc.},$$

onde os vetores α_i , $i = 3, 4$, estão definidos em (5.30). As propriedades da matriz Σ e suas derivadas desenvolvidas em (i) – (x) da Seção 5.1.2 serão de grande utilidade na derivação dos cumulantes deste modelo. O procedimento algébrico envolvido na derivação destes cumulantes é complicado e muito extenso, de modo que será ilustrado apenas com a determinação de alguns cumulantes nos exemplos que seguem. Contudo, convém notar

que todos os cumulantes foram derivados diretamente de suas definições e foram conferidos a partir das relações entre eles.

Exemplo A.1 Derivação de $\kappa_{\beta,\beta,\beta,\beta}$.

Como $Z_1, Z_2, \dots, Z_n$ são n réplicas iid de $Z \sim t_2(\mu(\phi_L), \Sigma(\phi_E); \nu)$, então tem-se que

$$\kappa_{\beta,\beta,\beta,\beta} = n\lambda_{\beta,\beta,\beta,\beta} - \frac{3}{n}\kappa_{\beta,\beta}^2 ,$$

onde $\kappa_{\beta,\beta} = n\lambda_{\beta,\beta}$ com $\lambda_{\beta,\beta} = E(\ell_\beta^2)$ e $\lambda_{\beta,\beta,\beta,\beta} = E(\ell_\beta^4)$. A função ℓ envolvida aqui está definida em (5.64). Lembrando que (veja (5.34))

$$\text{tr}(E_\beta) = \text{tr}\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta}\right) = 0 ,$$

então do fato (3) dado em A.1.2 ou aplicando (2.91) (veja também (2.90)) diretamente na segunda expressão de A.7, tem-se que

$$\begin{aligned} l_{\beta,\beta,\beta,\beta} &= E\{W^4(T^\top E_\beta T)^4\} = E\{\|T\|^8 W^4\} E\{(U^\top E_\beta U)^4\} \\ &= \frac{12a(4,4)}{p(p+2)(p+4)(p+6)} \{4\text{tr}(E_\beta^4) + (\text{tr}(E_\beta^2))^2\} , \end{aligned}$$

onde $W = g^{(1)}(\|T\|^2)$, $U = T/\|T\|$ são independentes e $a(4,4)$ está definida na Tabela A.1 para $p = 2$. Do fato (vi) dado na Seção 5.1.2 (veja também (5.34)), segue que

$$E_\beta^2 = \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \beta}\right)^2 = \frac{1}{\sigma_\beta^2} I_2 \Rightarrow E_\beta^4 = \frac{1}{\sigma_\beta^4} I_2 ,$$

onde σ_β^2 está definida em (5.33). Este fato juntamente com a Tabela A.1 implica que

$$\lambda_{\beta,\beta,\beta,\beta} = \frac{144a(4,4)}{p(p+2)(p+4)(p+6)} \frac{1}{\sigma_\beta^4} = 9 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} \lambda_{\beta,\beta}^2 ,$$

uma vez que $\lambda_{\beta,\beta} = (\nu+2)/(\nu+4)\sigma_\beta^2$. Lembrando que $\kappa_{\beta,\beta} = n\lambda_{\beta,\beta}$, tem-se que

$$\kappa_{\beta,\beta,\beta,\beta} = \frac{3}{n} \left\{ 3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{\beta,\beta}^2 .$$

Uma forma alternativa que evita a avaliação de alguns traços segue-se de aplicar a transformação dada em (5.35) (veja também os fatos (x) e (xi) dados na Seção 5.1.2).

Exemplo A.2 Derivação de $\kappa_{\beta\beta\beta} = n\lambda_{\beta\beta\beta}$.

Considerando a terceira derivada de ℓ , tem-se que

$$\begin{aligned}\ell_{\beta\beta\beta} = & -g^{(3)}(\|T\|^2)(T^\top E_\beta T)^3 + 3g^{(2)}(\|T\|^2)(T^\top E_{\beta\beta} T) - 2g^{(2)}(\|T\|^2)(T^\top E_\beta T)(T^\top E_\beta T) \\ & + 4g^{(1)}(\|T\|^2)(T^\top E_\beta E_{\beta\beta} T) - g^{(1)}(\|T\|^2)(T^\top E_{\beta\beta\beta} T) ,\end{aligned}$$

de modo que $\lambda_{\beta\beta\beta} = E(\ell_{\beta\beta\beta})$. A dificuldade na avaliação deste cumulante já é bastante clara, por exemplo, o valor esperado da última parcela é dado por

$$\begin{aligned}E(g^{(1)}(\|T\|^2)(T^\top E_{\beta\beta\beta} T)) &= E(\|T\|^2 W)E(U^\top E_{\beta\beta\beta} U) \\ &= a(1, 1)\frac{1}{p}\text{tr}(E_{\beta\beta\beta}) ,\end{aligned}$$

onde $\text{tr}(E_{\beta\beta\beta}) = \text{tr}(\Sigma^{-1}\frac{\partial^3 \Sigma}{\partial \beta^3})$. Para o modelo com λ_x conhecido este traço é fácil de obter, uma vez que $\partial^3 \Sigma / \partial \beta^3 = 0$ (veja (5.28)). Para o caso com λ_e conhecido, a derivação deste traço é complicada, já que envolve a terceira derivada da matriz Σ com respeito a β . Contudo, este traço pode ser obtido como segue (veja (5.34))

$$\text{tr}(E_\beta) = \text{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \beta}\right) = 0 \Rightarrow \text{tr}(E_{\beta\beta} - E_\beta^2) = 0 \Rightarrow \text{tr}(E_{\beta\beta\beta} - 3E_\beta E_{\beta\beta} + 2E_\beta^3) = 0 .$$

Um caminho alternativo para derivar $\lambda_{\beta\beta\beta}$ é dado pelo fato de que $\lambda_{\beta\beta\beta} = \lambda_{\beta\beta}^{(\beta)} - \lambda_{\beta\beta\beta}$. Assim, aplicando o resultado (4) dado em A.1.2, tem-se que

$$\lambda_{\beta\beta,\beta} = -\left(\frac{16}{\nu+2}\frac{a(3,3)}{p(p+2)(p+4)} + 4\frac{a(2,2)}{p(p+2)}\right)\text{tr}(E_\beta^3) - 2\frac{a(2,2)}{p(p+2)}\text{tr}(E_\beta E_{\beta\beta}) .$$

Note agora que

$$E_\beta^2 = \sigma_\beta^{-2} I_2 \Rightarrow E_\beta^3 = \sigma_\beta^{-2} E_\beta \Rightarrow \text{tr}(E_\beta^3) = 0 \Rightarrow \text{tr}(E_{\beta\beta\beta}) = -3\text{tr}(E_\beta E_{\beta\beta}) ,$$

de modo que $\lambda_{\beta\beta,\beta} = 0$ para λ_x conhecido. Para este modelo tem-se também que $\lambda_{\beta\beta}^{(\beta)} = 0$, uma vez que $\lambda_{\beta\beta} = -(\nu+2)/(\nu+4)\sigma_\beta^2$ e σ_β^2 não depende de β (veja (5.33)). Assim, $\lambda_{\beta\beta\beta} = 0$, quando λ_x é conhecido. No modelo com λ_e conhecido pode-se usar a propriedade (vi)

dada na Seção 5.1.2 para mostrar que (colocando $\lambda_e = 1$) $\text{tr}(E_\beta E_{\beta\beta}) = -4\beta(\beta^2 + 1)^{-1}\sigma_\beta^{-2}$.

Logo, usando a Tabela A.1, neste modelo com $\lambda_e = 1$ tem-se que

$$\lambda_{\beta,\beta\beta} = -2 \frac{a(2,2)}{p(p+2)} \text{tr}(E_\beta E_{\beta\beta}) = 2\beta(\beta^2 + 1)^{-1} \lambda_{\beta\beta} ,$$

e $\lambda_{\beta\beta}^{(\beta)} = -4\beta(\beta^2 + 1)^{-1} \lambda_{\beta\beta}$. Destes resultados segue que

$$\kappa_{\beta\beta\beta} = \kappa_{\beta\beta}^{(\beta)} - \kappa_{\beta,\beta\beta} = \begin{cases} 0 , & \text{para } \lambda_x \text{ conhecido} \\ -6\beta(\beta^2 + 1)^{-1} \kappa_{\beta\beta} , & \text{para } \lambda_e = 1. \end{cases}$$

Similarmente, $\kappa_{\beta\beta\beta\beta}$ pode ser derivado usando que

$$\kappa_{\beta,\beta,\beta,\beta} = -3\kappa_{\beta\beta\beta\beta} + 8\kappa_{\beta\beta\beta}^{(\beta)} - 6\kappa_{\beta\beta}^{(\beta\beta)} + 3\kappa_{\beta\beta,\beta\beta} ,$$

o que evita derivar quatro vezes a função ℓ e a matriz Σ .

No que segue assume-se os cumulantes que foram utilizados na derivação dos testes modificados de $H_0 : \beta = \beta_0$ do modelo estrutural t -independente.

A.3.1 Alguns cumulantes para os parâmetros de locação

$$\begin{aligned} \kappa_{ij} = -\kappa_{i,j} &= -n \frac{\nu+2}{\nu+4} \mu_i^\top \Sigma^{-1} \mu_j , \quad \kappa^{ij} = -\kappa^{i,j} = -\frac{1}{n} \frac{\nu+4}{\nu+2} \mu_i^\top \Sigma \mu_j , \\ \kappa_{ijk} = \kappa_{i,jk} = \kappa_{ij}^{(k)} &= 0 \quad i, j = 1, 2, \text{ onde } \mu_1 = \partial\mu/\partial\phi_1 = (1, 0)^\top \text{ e } \mu_2 = \partial\mu/\partial\phi_2 = (0, 1)^\top. \end{aligned}$$

A.3.2 Alguns cumulantes para os parâmetros de escala

Cumulantes que envolvem os parâmetros ϕ_3, ϕ_4 :

$$\begin{aligned} \kappa_{ii} &= -\frac{n}{2} \frac{\nu+1}{\nu+4} \phi_i^2 , \quad \kappa_{34} = \frac{n}{2} \frac{1}{\nu+4} \phi_3 \phi_4 , \quad \kappa^{ii} = -\frac{2}{n} \frac{(\nu+1)(\nu+4)}{\nu(\nu+2)} \phi_i^2 , \quad \kappa^{34} = -\frac{2}{n} \frac{\nu+4}{\nu(\nu+2)} \phi_3 \phi_4 , \\ \kappa_{ii}^{(j)} &= 0 , \quad i \neq j , \quad \kappa_{ii}^{(i)} = -\phi_i^{-1} \kappa_{ii} , \quad \kappa_{34}^{(j)} = -\phi_j^{-1} \kappa_{34} , \quad \kappa_{iii} = -2 \frac{2\nu+9}{\nu+6} \phi_i^{-1} \kappa_{ii} , \\ \kappa_{i34} &= -2 \frac{\nu+3}{\nu+6} \phi_i^{-1} \kappa_{34} , \quad \kappa_{i,ii} = 2 \frac{\nu+3}{\nu+6} \phi_i^{-1} \kappa_{ii} , \quad \kappa_{i,34} = \frac{\nu}{\nu+6} \phi_i^{-1} \kappa_{34} . \end{aligned}$$

Cumulantes que envolvem o parâmetro β :

(a) Para λ_x conhecido,

$$\begin{aligned}\kappa_{\beta\beta} &= -n \frac{\nu+2}{\nu+4} \left(\frac{\lambda_x}{\lambda_x+1} \right)^2 \frac{\phi_4}{\phi_3}, \quad \kappa^{\beta\beta} = -\frac{1}{n} \frac{\nu+4}{\nu+2} \left(\frac{\lambda_x+1}{\lambda_x} \right)^2 \frac{\phi_3}{\phi_4}, \\ \kappa_{\beta\beta}^{(\beta)} &= \kappa_{\beta\beta}^{(\beta\beta)} = \kappa_{\beta\beta\beta} = \kappa_{\beta\beta\beta}^{(\beta)} = \kappa_{\beta,\beta\beta} = \kappa_{\beta,\beta,\beta} = 0, \quad \kappa_{\beta,\beta,\beta,\beta} = \frac{3}{n} \left\{ 3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{\beta\beta}^2, \\ \kappa_{\beta\beta,\beta\beta} &= \frac{1}{n} \left\{ 3 \frac{(\nu+4)(\nu^2+10\nu+28)}{(\nu+2)(\nu+6)(\nu+8)} - 1 \right\} \kappa_{\beta\beta}^2, \quad 3\kappa_{\beta,\beta,\beta\beta} + \kappa_{\beta,\beta,\beta,\beta} = 0 \\ \kappa_{\beta\beta\beta\beta} &= \frac{1}{3} (3\kappa_{\beta\beta,\beta\beta} - \kappa_{\beta,\beta,\beta,\beta}) = \frac{18}{n} \frac{(\nu+4)^2}{(\nu+2)(\nu+6)(\nu+8)} \kappa_{\beta\beta}^2.\end{aligned}$$

(b) Para λ_e conhecido com $\lambda_e = 1$,

$$\begin{aligned}\kappa_{\beta\beta} &= -n \frac{\nu+2}{\nu+4} (\phi_3\phi_4)^{-1} (\phi_3 - \phi_4)^2 (\beta^2 + 1)^{-2}, \quad \kappa^{\beta\beta} = -\frac{1}{n} \frac{\nu+4}{\nu+2} (\phi_3\phi_4) (\phi_3 - \phi_4)^{-2} (\beta^2 + 1)^2, \\ \kappa_{\beta\beta}^{(\beta)} &= -4\beta(\beta^2 + 1)^{-1} \kappa_{\beta\beta}, \quad \kappa_{\beta,\beta}^{(\beta\beta)} = 4(5\beta^2 - 1)(\beta^2 + 1)^{-2} \kappa_{\beta\beta}, \\ \kappa_{\beta,\beta\beta} &= -2\beta(\beta^2 + 1)^{-1} \kappa_{\beta\beta}, \quad \kappa_{\beta,\beta\beta} = 0, \quad \kappa_{\beta\beta\beta}^{(\beta)} = 6(5\beta^2 - 1)(\beta^2 + 1)^{-2} \kappa_{\beta\beta}, \\ \kappa_{\beta\beta\beta} &= \kappa_{\beta\beta\beta}^{(\beta)} - \kappa_{\beta,\beta\beta} = -6\beta(\beta^2 + 1)^{-1} \kappa_{\beta\beta}, \quad \kappa_{\beta,\beta,\beta,\beta} = \frac{3}{n} \left\{ 3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{\beta\beta}^2, \\ \kappa_{\beta\beta,\beta\beta} &= \frac{1}{n} \left\{ 3 \frac{(\nu+4)(\nu^2+10\nu+28)}{(\nu+2)(\nu+6)(\nu+8)} - 1 \right\} \kappa_{\beta\beta}^2 - 4(\beta^2 + 1)^{-1} \kappa_{\beta\beta}, \quad 3\kappa_{\beta,\beta,\beta\beta} + \kappa_{\beta,\beta,\beta,\beta} = 0, \\ \kappa_{\beta\beta\beta\beta} &= \frac{8}{n} \frac{(\nu+4)^2}{(\nu+2)(\nu+6)(\nu+8)} \kappa_{\beta\beta}^2 + 12(3\beta^2 - 1)(\beta^2 + 1)^{-2} \kappa_{\beta\beta}.\end{aligned}$$

A.3.3 Alguns cumulantes mistos que envolvem o parâmetro β

(a) β com ϕ_1 e ϕ_2 .

$$\begin{aligned}\kappa_{\beta i} &= \kappa_{\beta\beta i} = \kappa_{\beta\beta,i} = \kappa_{\beta,\beta i} = \kappa_{\beta,\beta,i} = 0, \quad i = 1, 2, \quad \kappa_{\beta ij} = \frac{\nu+4}{\nu+6} \kappa_{ij}^{(\beta)}, \quad \kappa_{ij,\beta} = \frac{2}{\nu+6} \kappa_{ij}^{(\beta)}, \\ \kappa_{ij\beta\beta} &= \frac{6}{n} \frac{(\nu+4)^2}{(\nu+2)(\nu+6)(\nu+8)} \kappa_{ij} \kappa_{\beta\beta} + \frac{\nu+4}{\nu+6} \kappa_{ij}^{(\beta\beta)}, \quad \kappa_{i\beta,j\beta} = \frac{1}{n} \frac{(\nu+4)(\nu^2+10\nu+28)}{(\nu+2)(\nu+6)(\nu+8)} \kappa_{ij} \kappa_{\beta\beta}, \\ \kappa_{ij,\beta\beta} &= \kappa_{i\beta,j\beta} - \kappa_{ij} \kappa_{\beta\beta} + \frac{2}{\nu+6} \kappa_{ij}^{(\beta\beta)}, \quad \kappa_{i,j,\beta,\beta} = \frac{1}{n} \left\{ 3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{ij} \kappa_{\beta\beta}, \\ \kappa_{ij,\beta,\beta} &= -\frac{1}{n} \left\{ \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{ij} \kappa_{\beta\beta}, \quad i, j = 1, 2.\end{aligned}$$

(b) β com ϕ_3 e ϕ_4 .

$$\kappa_{\beta i} = \kappa_{\beta ij} = \kappa_{\beta,i j} = \kappa_{\beta i,j} = \kappa_{\beta,i,j} = 0, \quad i, j = 3, 4, \quad \kappa_{i\beta\beta} = \left(\frac{3}{\nu+6} - \delta_{3i} \right) \phi_i^{-1} \kappa_{\beta\beta},$$

$$\begin{aligned}
\kappa_{i,\beta\beta} &= \left(\frac{\nu+3}{\nu+6} - \delta_{3i} \right) \phi_i^{-1} \kappa_{\beta\beta} , \quad \kappa_{ii,\beta\beta} = \frac{4}{n} \frac{(\nu+3)(\nu+5)}{(\nu+1)(\nu+8)} \kappa_{ii} \kappa_{\beta\beta} - \frac{4}{n} \frac{(\nu+3)(\nu+4)}{(\nu+1)(\nu+6)} \delta_{3i} \kappa_{ii} \kappa_{\beta\beta} , \\
\kappa_{34,\beta\beta} &= \frac{4}{n} \left\{ 3 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{34} \kappa_{\beta\beta} , \quad \kappa_{i,i,\beta,\beta} = \frac{1}{n} \left\{ \frac{(\nu+4)(5\nu^2+7\nu+6)}{(\nu+1)(\nu+6)(\nu+8)} - 1 \right\} \kappa_{ii} \kappa_{\beta\beta} , \\
\kappa_{i\beta,i\beta} &= \frac{30}{n} \frac{\nu+4}{(\nu+1)(\nu+6)(\nu+8)} \kappa_{ii} \kappa_{\beta\beta} + \frac{2}{n} \frac{\nu(\nu+4)}{(\nu+1)(\nu+6)} \delta_{3i} \kappa_{ii} \kappa_{\beta\beta} , \quad \kappa_{3\beta,4\beta} = \frac{6}{n} \frac{(\nu+4)(\nu+5)}{(\nu+6)(\nu+8)} \kappa_{34} \kappa_{\beta\beta} , \\
\kappa_{3,4,\beta,\beta} &= -\frac{1}{n} \left\{ \frac{(\nu+4)(2\nu^2-5\nu-6)}{(\nu+6)(\nu+8)} + 1 \right\} \kappa_{34} \kappa_{\beta\beta} , \quad \kappa_{ii\beta\beta} = \frac{4}{n} \frac{(\nu+3)(\nu+4)}{(\nu+1)(\nu+6)} \left(\frac{3}{\nu+8} - \delta_{3i} \right) \kappa_{ii} \kappa_{\beta\beta} , \\
\kappa_{34\beta\beta} &= -\frac{6}{n} \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} \kappa_{34} \kappa_{\beta\beta} , \quad \kappa_{ii,\beta,\beta} = -\frac{1}{n} \left\{ \frac{(\nu+4)(5\nu^2+31\nu+18)}{(\nu+1)(\nu+6)(\nu+8)} - 1 \right\} \kappa_{ii} \kappa_{\beta\beta} , \\
\kappa_{34,\beta,\beta} &= -\frac{1}{n} \left\{ 9 \frac{(\nu+2)(\nu+4)}{(\nu+6)(\nu+8)} - 1 \right\} \kappa_{34} \kappa_{\beta\beta} , \quad \text{onde } \delta_{33} = 1 \text{ e } \delta_{34} = 0.
\end{aligned}$$

Bibliografia

- ALI, M.M. (1980). Characterization of the normal distribution among the continuous symmetric spherical class. *Journal of the Royal Statistical Society, B*, **42**, 162–164.
- ALI, M.M.; JOARDER, A.H. (1991). Distribution of the correlation coefficient for the class of bivariate elliptical models. *The Canadian Journal of Statistics*, **19**, 447–452.
- ALBERT, J.; DELAMPADY, M.; POLASEK, W. (1991). A class of distributions for robustness studies. *Journal of Statistical Planning and Inference*, **28**, 291–304.
- ANDERSON, T.W. (1984). *An introduction to multivariate statistical analysis*. 2.ed. New York: Wiley.
- ANDERSON, T.W.; STEPHENS, M.A. (1972). Test for randomness of directions against equatorial and bimodal alternative. *Biometrika*, **59**, 613–621.
- ANDERSON, T.W.; SAWA, T. (1982). Exact and approximate distribution of the maximum likelihood estimator of a slope coefficient. *Journal of the Royal Statistical Society B*, **44**, 52–62.
- ANDERSON, T.W.; FANG, K.T.; HSU, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptical contoured distributions. *The Canadian Journal of Statistics*, **44**, 55–59.
- ANDERSON, T.W.; FANG, K.T. (1987). Cochran's theorem for elliptically contoured distributions. *Sankhyā A*, **49**, 305–315.
- ANDO, A.; KAUFMAN, G.M. (1965). Bayesian analysis of the independent multinormal process—neither mean nor precision knows. *Journal of the American Statistical Association*, **60**, 347–358.

- ANDREWS, D.F.; MALLOWS, C.L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, **36**, 99–102.
- ANTILLE, G.; RITSCHARD, G. (1992). Robust versus classical detection of atypical data. *Computational Statistical and Data Analysis*, **14**, 75–88.
- ARNOLD, S.F.; LYNCH, J. (1982). On Ali's characterization of the spherical normal distribution. *Journal of the Royal Statistical Society B*, **44**, 49–51.
- ARMSTRONG, B. (1985). Measurement error in generalized linear model. *Communications in Statistics B*, **14**, 529–544.
- ABRAMOWITZ, M.; STEGUN, I.A. (1972). *Handbook of mathematical functions*. New York: Dover Publications, Inc.
- ALI, M.M.; SHARMA, S.C. (1993). Robustness to nonnormality of the Durbin-Watson test for autocorrelation. *Journal of Econometrics*, **57**, 117–136.
- ARNOLD, S.F. (1981). *The theory of linear models and multivariate analysis*. New York: Wiley.
- ARNOLD, B.C.; CASTILLO, E.; SARABIA, J.M. (1994). A conditional characterization of the multivariate normal distribution. *Statistics and Probability Letters*, **19**, 313–315.
- BANSAL, N.K. (1990). Testing linear hypotheses in errors in variables model. *Annals of the Institute of Statistical Mathematics*, **42**, 581–596.
- BARNDORFF-NIELSEN, O.E.; COX, D.R. (1984). Bartlett adjustments to the likelihood ratio statistics and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society B*, **46**, 483–495.
- BARNDORFF-NIELSEN, O.E.; BLAESILD, P. (1986). A note on the calculation of Bartlett adjustments. *Journal of the Royal Statistical Society B*, **48**, 353–358.
- BARNDORFF-NIELSEN, O.E.; HALL, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374–778.
- BARNETT, V.D. (1966). Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika*, **53**, 151–165.
- BARNETT, V.D. (1967). A note on linear structural relationships when both residual variance are known. *Biometrika*, **54**, 670–672.

- BENNETT, B.M. (1961). On a certain multivariate non-normal distributions. *Biometrika*, **51**, 434–437.
- BENTLER, P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures. *Psychometrika*, **48**, 493–517.
- BERKANE, M.E.; BENTLER, P.M. (1987). Characterizing parameters of multivariate elliptical distributions. *Communication Statistic – Simulation*, **16**, 193–198.
- BERKANE, M.E.; BENTLER, P.M. (1988). Estimation of contamination parameters and identification of outliers in multivariate data. *Sociological Methods and Research*, **17**, 55–64.
- BHANDARY, M. (1991). Robust M -estimation of a dispersion matrix with a structure. *Annals of the Institute of Statistical Mathematics*, **43**, 689–705.
- BICKEL, P.J.; GHOSH, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction - a Bayesian argument. *The Annals of Statistics*, **18**, 1070–1090.
- BOLFARINE, H.; RODRIGUES, J.; CORDANI, L.K. (1992). Modelos com erros nas variáveis. *ABE*.
- BOLFARINE, H.; CORDANI, L.K. (1993). Estimation of a structural regression model with known reliability ratio. *Annals of the Institute of Statistical Mathematics*, **45**, 3, 531–540.
- BOLFARINE, H.; NASCIMENTO, J.A. (1992). *Bartlett correction factors for the structural regression model with known reliability ratio*. São Paulo, IME-UPS, 1992 (RT-MAE-9206).
- BOX, G.E.P.; TIAO, G.C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley: Menlo Park, CA.
- BROWN, M.L. (1982). Robust line estimation with errors in both variables. *Journal of the American Statistical Association*, **77**, 71–79.
- BROWNE, M.W. (1982). *Covariance structures*. In topics in Applied Multivariate Analysis, ed. D.M. Hawkins, Cambridge, U.K.: Cambridge University Press, 72–141.

- BROWNE, M.W. (1984). Asymptotically distribution-free methods for analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.
- BUTLER, R.J.; McDONALD, N.D.R.; STEVEN, W.B. (1990). Robust and partially adaptive estimation of regression models. *The Review of Economics and Statistics*, **73**, 321–326.
- CACOULLOS, T. (1992). Two LDF characterizations of the normal as a spherical distribution. *Journal of Multivariate Analysis*, **40**, 205–212.
- CACOULLOS, T.; KHATRI, C.G. (1991). Correcting remarks on “characterization of normality withing the class of elliptical contoured distributions”. *Statistics and Probability Letters*, **11**, 551–552.
- CAMBANIS, S.; HUANG, S.; SIMONS, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368–385.
- CARROLL, R.J.; GALLO, P.P. (1982). Some aspects of robustness in the functional errors in variables regression model. *Communication in Statistic – Theory and Methods*, **11**, 2573–2585.
- CARROLL, R.J.; ELTINGE, J.L.; RUPPERT, D. (1993). Robust linear regression in replicated measurement error models. *Statistics and Probability Letters*, **16**, 169–175.
- CASELLA, G.; BERGER, R.L. (1990). *Statistical inference*. Pacific Grove, California: Wadsworth and Brooks/Cole.
- CHANDRA, T.K. (1985). Asymptotic expansions of perturbed chi-squared variables. *Sankhyā A*, **47**, 100–110.
- CHANDRA, T.K.; GHOSH, J.K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-squared variables. *Sankhyā A*, **41**, 22–47.
- CHANDRA, T.K.; JOSHI, S.N. (1983). Comparison of the likelihood ratio, Rao’s and Wald’s tests and a conjecture of C.R. Rao. *Sankhyā A*, **45**, 226–246.
- CHANDRA, T.K.; MUKERJEE, R. (1984). On the optimality of Rao’s statistic. *Communications in Statistics A*, **13**, 1507–1515.

- CHANDRA, T.K.; MUKERJEE, R. (1985). Comparison of the likelihood ratio, Wald's and Rao's tests. *Sankhyā A*, **47**, 271–284.
- CHANDRA, T.K.; MUKERJEE, R. (1991). Bartlett-type modification for Rao's efficient score statistic. *Journal of Multivariate Analysis*, **36**, 103–112.
- CHANG, K.C.; HAN, C.P.; HAWKINS, D.L. (1994). Improved estimators of regression coefficients in measurement error models. *Journal of Statistics in Computation and Simulation*, **39**, 21–35.
- CHIB, S.; TIWARI, R.C.; JAMMALAMADAKA, S.R. (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics*, **38**, 349–360.
- CHIKUSE, Y. (1980). Invariant polynomials with matrix arguments and their applications. *Multivariate Statistical Analysis*, R.P. Gupta (ed.), 53–68.
- CHMIELEWSKI, M.A. (1980). Invariant scale matrix hypotheses tests under elliptical symmetry. *Journal of Multivariate Analysis*, **10**, 343–350.
- CHMIELEWSKI, M.A. (1981). Elliptically symmetric distributions: a review and bibliography. *International Statistical Review*, **49**, 67–74.
- CHOW, Y.S.; TEICHER, H. (1978). *Probability theory*. Springer: New York.
- CHUNSHENG, M. (1992). The order of a univariate elliptical distribution. *Statistic and Probability Letters*, **14**, 307–310.
- COHEN, A.C.; WHITTEN, B.J. (1982). Modified moments and maximum likelihood estimators for parameters of the three-parameter gamma distribution. *Commun. Statist. - Simula. Computa.*, **11**, 197–216.
- CONSONNI, G.; VERONESE, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *Journal of the American Statistical Association*, **87**, 1123–1127.
- CONSTANTINE, A.G. (1963). Some non-central distribution problems in multivariate analysis. *Annals of Mathematics Statistics*, **34**, 1270–1285.
- CORDEIRO, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B*, **45**, 404–413.
- CORDEIRO, G.M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, **74**, 265–274.

- CORDEIRO, G.M. (1993). General matrix formulae for computing Bartlett corrections. *Statistic and Probability Letters*, **16**, 11–18.
- CORDEIRO, G.M.; FERRARI, S.L.P. (1991). A modified score statistic having chi-squared distribution to order n^{-1} . *Biometrika*, **78**, 573–582.
- CORDEIRO, G.M.; FERRARI, S.L.P.; PAULA, G.A. (1993). Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B*, **55**, 661–674.
- CORDEIRO, G.M.; PAULA, G.A. (1989). Improved likelihood statistics for exponential linear models. *Biometrika*, **76**, 93–100.
- CORDEIRO, G.M.; PAULA, G.A.; BOTTER, D.A. (1993). Improved likelihood ratio tests for dispersion models. *International Statistical Review*. To appear.
- COX, D.R. (1988). Some aspects of conditional and asymptotic inference: a review. *Sankhyā A*, **50**, 314–337.
- COX, D.R.; REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, **49**, 1–39.
- CRIBARI-NETO, F.; FERRARI, S.L.P. (1995a). An improved Lagrange multiplier statistic for the test of heteroscedasticity. *Communication in Statistics - Theory and Methods*, a aparecer.
- CRIBARI-NETO, F.; FERRARI, S.L.P. (1995b). Second order asymptotics for score tests in generalized linear models. *Biometrika*, **82**, a aparecer.
- CORNISH, E.A. (1954). The multivariate t -distribution. *Australian Journal of Physics*, **7**, 531–542.
- DALAL, S.R.; HALL, W.J. (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society B*, **45**, n.2, 278–286.
- DATTA, G.S. (1992). A unified Bayesian prediction theory for mixed linear models with application. *Statistics & Decisions*, **10**, 337–365.
- DAVIS, A.W. (1979). Invariant polynomials with two matrix arguments extending the zonal polynomials: applications to multivariate distribution theory. *Ann. Inst. Statist. Math.*, **31**, Part A, 465–485.

- DAVIS, A.W. (1981). On the construction of a class of invariant polynomials in several matrices, extending the zonal polynomials. *Ann. Inst. Statist. Math.*, **33**, Part A, 297–313.
- DAWID, A.P. (1977). Spherical matrix distributions and a multivariate model. *Journal of the Royal Statistical Society B*, **39**, 254–261.
- DAWID, A.P. (1978). Extendibility of spherical matrix distributions. *Journal of Multivariate Analysis*, **8**, 559–566.
- DAWID, A.P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, **68**, **1**, 265–274.
- DELAMPADY, M. (1992). Bayesian robustness for elliptical distributions. *Revista Brasileira de Probabilidade e Estatística*, **6**, 97–119.
- DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1–38.
- DEVLIN, S.J.; GNANADESIKAN, R.; KETTERING, J.R. (1976). Some multivariate applications of elliptical distributions. In: *Essays in Probability and Statistics* (Eds. S.I. Keda et al.), Chapter 24, Tokyo: Shinko Tsusho Co., 365–393.
- DIACONIS, E.; EATON, M.; LAURITZEN, S. (1992). Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.*, **19**, 289–315.
- DIACONIS, E.; ILVISAKER, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics*, v.7, n.2, 269–281.
- DIACONIS, E.; ILVISAKER, D. (1985). Quantifying prior opinion. *Bayesian Statistics*, **2**, 133–156.
- DICKEY, J.M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematics and Statistics*, **38**, 511–518.
- DICKEY, J.M. (1976). A new representation of Student's t as a function of independent t 's, with a generalization to the matrix t . *Journal of Multivariate Analysis*, **6**, 343–346.

- DUNNETT, C.H.W.; SOBEL, M. (1954). A bivariate generalization of Student's t distribution, with tables for certain special cases. *Biometrika*, **41**, 153-169.
- DUNNETT, C.H.W.; SOBEL, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t distribution. *Biometrika*, **42**, 258-260.
- EFRON, B. (1969). Student's t -test under symmetric conditions. *Journal of the American Statistical Association*, 1278-1302.
- EFRON, B. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, **65**, 3, 457-487.
- EATON, M.L. (1977). *N-dimensional versions of some symmetric univariate distributions*. Tech. Report, n. 288, University of Minnesota.
- EATON, M.L. (1986). A characterization of spherical distributions. *Journal of Multivariate Analysis*, **20**, 272-276.
- FAN, T.H.; BERGER, J.O. (1992). Behaviour of the posterior distribution and inferences for a normal mean with t prior distributions. *Statistics & Decisions*, **10**, 99-120.
- FANG, K.T.; BENTLER, P.M. (1991). A largest characterization of spherical and related distributions. *Statistical and Probability Letters*, **11**, 107-110.
- FANG, K.T.; KOTZ, S.; NG, K.W. (1990). *Symmetric multivariate and related distributions*. London-New York: Chapman and Hall.
- FANG, K.T.; ANDERSON, T.W. (1990). *Statistical inference in elliptically contoured and related distributions*. New York: Allerton Press Inc.
- FANG, K.T.; ZHU, L.X.; BENTLER, P.M. (1993). A necessary test of goodness of fit for sphericity. *Journal of Multivariate Analysis*, **45**, 34-55.
- FELLER, W. (1971). *An introduction to probability theory and applications*. v.II. 2.ed. New York: Wiley.
- FERRARI, S.L.P. (1991). *Aperfeiçoamento de testes escore, aplicações e extensões*. São Paulo. 185p. Tese (Doutorado). Instituto de Matemática e Estatística - Universidade de São Paulo.

- FERRARI, S.L.P.; CRIBARI-NETO, F. (1993). On the corrections to the Wald test of non-linear restrictions. *Economic Letters*, **42**, 321–326.
- FRASER, D.A.S (1976). Necessary analysis and adaptive inference. *Journal of the American Statistical Association*, **71**, n.353, 99–113.
- FRIEDMANN, R. (1990). Bounds for exact moments of estimators in the errors-in-variables model and simultaneous equations. *Journal of Econometrics*, **45**, 367–384. North Holland.
- FUJIKOSHI, Y. (1989). Error bounds for asymptotic expansion of the maximum of the multivariate t and F variables with common denominator. *Hiroshima Math. J.*, **19**, 319–327.
- FULLER, W. (1987). *Measurement error models*. New York: Wiley.
- GILES, J.A. (1992). Estimation of the error variance after a preliminary-test of homogeneity in a regression model with spherically symmetric disturbances. *Journal of Econometrics*, **53**, 345–361. North Holland.
- GHOSH, M.; SINHA, K.B. (1980). On the robustness of least squares procedures in regression models. *Journal of Multivariate Analysis*, **10**, 332–342.
- GHOSH, J.K.; MUKERJEE, R. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *Journal of the Royal Statistical Society B*, **54**, n.3, 867–875.
- GOEL, P.K.; DeGROOT, M.H. (1980). Only normal distribution have linear posterior expectations in linear regression. *Journal of the American Statistical Association*, **75**, n.373, 895–900.
- GONG, G.; SAMANIEGO, F.J. (1981). Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, **9**, n.4, 861–869.
- GRAYBILL, F.A. (1983). *Matrices with applications in statistics*. 2.ed. Belmont, California: Wadsworth.
- GREEN, R.F. (1976). Outlier-prone and outlier-resistant distributions. *Journal of the American Statistical Association*, **71**, n.354, 502–505.
- GROSS, A.M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *Journal of the American Statistical Association*, **71**, n.354, 409–416.

- HAMPEL, F.R.; RONCHETTI, E.M.; ROUSSEEUW; STAHEL, W.A. (1986). *Robust statistics: the approach based on influence functions*. New York: Wiley.
- HABILULLAH, M.; KATTI, S.K. (1991). A modified steepest descent method with applications to maximizing likelihood functions. *Ann. Inst. Statist. Math.*, **43**, n.2, 391–404.
- HADI, A.S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistics & Data Analysis*, **14**, 1–27. North-Holland.
- HARRIS, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika*, **72**, 653–659.
- HAUSMAN, J.A.; NEWEY, W.; ICHIMURA, H.; POWEL, J.L. (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics*, **50**, 273–295.
- HAYAKAWA, T. (1966). On the distribution of a quadratic form in multivariate normal sample. *Ann. Inst. Statist. Math.*, **18**, 191–201.
- HAYAKAWA, T. (1969). On the distribution of the latent roots of a positive definite random symmetric matrix I. *Ann. Inst. Statist. Math.*, **21**, 1–21.
- HAYAKAWA, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika*, **62**, 2, 451–460.
- HAYAKAWA, T. (1972). On the distribution of the multivariate quadratic form in multivariate normal sample. *Ann. Inst. Math. Statist.*, **24**, 205–230.
- HAYAKAWA, T. (1972). The asymptotic distributions of the statistics based on the complex Gaussian distribution. *Ann. Inst. Statist. Math.*, **24**, 231–244.
- HAYAKAWA, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Ann. Inst. Statist. Math.*, **29**, Part A, 359–378.
- HAYAKAWA, T.; PURI, M.L. (1985). Asymptotic distributions of likelihood ratio criteria for testing latent roots and latent vectors of a covariance matrix under an elliptical population. *Biometrika*, **72**, 2, 331–338.
- HAYAKAWA, T.; PURI, M.L. (1985). Asymptotic expansions of the distributions of some test statistics. *Ann. Inst. Statist. Math.*, **37**, Part A, 95–108.

- HAYAKAWA, T. (1973). An asymptotic expansion for the distribution of the determinant of a multivariate quadratic form in a normal sample. *Ann. Inst. Statist. Math.*, **25**, 395–421.
- HEYDE, C.C. (1970). Characterization of the normal law by the symmetry of a certain conditional distribution. *Sankhyā : The Indian Journal of Statistics: Series A*, 115–118.
- HOLLAS, E.A., (1991). Performance of Bartlett adjustment for certain likelihood ratio tests. *Comm. Statist. - Simula.*, **20**, **2**, **3**, 449–462.
- HOGG, R.V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, **69**, n.348, 909–927.
- HSIAO, C.H. (1989). Consistent estimation for some nonlinear errors-in-variables models. *Journal of Econometrics*, **41**, 159–185. North-Holland.
- HUBER, P.Y. (1981). *Robust statistics*. New York: Wiley.
- IFRAM, A. (1970). On the characteristic functions of the F and t distributions. *Sankhyā: The Indian Journal of Statistics: Series A*, 350–352.
- IGLESIAS-ZUAZOLA, P.L. (1993). *Finite forms of de Finetti's theorem: a predictivistic approach to statistical inference in finite populations* (in Portuguese). Doctoral Thesis. Instituto de Matemática e Estatística da Universidade de São Paulo.
- IWASHITA, T.; SIOTANI, M. (1994). Asymptotic distribution of functions of a sample covariance matrix under the elliptical distribution. *The Canadian Journal of Statistics*, **22**, 273–283.
- IWATA, S. (1992). Instrumental variables estimation in errors-in-variables models when instruments are correlated with errors. *Journal of Econometrics*, **53**, 297–322. North-Holland.
- JAMES, A.T. (1961). Zonal polynomials of the real positive definite symmetric matrices. *Annals of Mathematics*, **71**, n.3, 456–469.
- JAMES, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical and Statistics*, **35**, 475–501.
- JOHN, S. (1971). Some optimal multivariate tests. *Biometrika*, **58**, **1**, 123–127.

- JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, **59**, 1, 169–173.
- JOHNSON, N.L.; KOTZ, S. (1972). *Distributions in statistics: continuous multivariate distributions*. New York: Wiley.
- KANO, Y.; BERKANE, M.; BENTLER, P. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, **88**, 421, 135–143.
- KARIYA, T. (1981). A robustness property of hotelling's T^2 -test. *The Annals of Statistics*, **9**, 211–214.
- KARIYA, T.; EATON, M.L. (1977). Robust tests for spherical symmetry. *The Annals of Statistics*, **5**(1), 206–215.
- KEILSON, J.; STEUTEL, F.W. (1974). Mixtures of distributions, moment inequalities and measures of exponentiality and normality. *The Annals of Probability*, **2**(1), 112–130.
- KEILSON, J.; STEUTEL, F.W. (1972). Families of infinitely divisible distributions closed under mixing and convolution. *The Annals of Mathematical Statistics*, **43**(1), 242–250.
- KELKER, D. (1970). Distribution theory of special distributions and a location-scale parameter. *Sankhyā*, **A**, **32**, 319–430. spherical
- KELKER, D. (1971). Infinite divisibility and variance mixtures of the normal distribution. *The Annals of Mathematical Statistics*, **42**(2), 802–808.
- KELLY, G.E. (1992). Robust regression estimators - the choice of tuning constants. *The Statistician*, **41**, 303–314.
- KELLY, G.E. (1984). The influence function in the errors in variables problem. *The Annals of Statistics*, **12**(1), 87–100.
- KENDALL, M.G.; STUART, A. (1979). *The advanced theory of statistics*. 4.ed. New York: Hafner.
- KHATRI, C.G.; MUKERJEE, R. (1987). Characterization of normality within the class of elliptical contoured distributions. *Statistics and Probability Letters*, **5**, 187–190.

- KING, M.L. (1980). Robust tests for spherical symmetry and their application to least squares regression. *The Annals of Statistics*, **8**(6), 1265–1271.
- KINGMAN, J.F.C. (1972). On random sequences with spherical symmetry. *Biometrika*, **59**, 493–495.
- KONNO, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F distribution. *Ann. Inst. Statist. Math.*, **43**(1), 157–165.
- KOTLARSKI, I. (1966). On characterizing the chi-squared distribution by the student law. *Journal of the American Statistical Association*, December, 976–981.
- KSHIRSAGAR, A.M. (1960). Some extensions of the multivariate t distribution and the multivariate generalization of the distribution of the regression coefficient. *Proc. Cambridge Phil. Soc.*, **57**, 80–85.
- KUBOKAWA, T.; ROBERT, C.; SALEH, E. (1991). Robust estimation of common regression coefficients under spherical symmetry. *Ann. Inst. Statist. Math.*, **43**(4), 677–688.
- KURITSYN, Y.G. (1988). On the least-squares method for elliptically contoured distributions. *Theory Probab. Appl.*, **31**, 738–740.
- LANGE, K.L.; LITTLE, J.A.; TAYLOR, M.G. (1989). Robust statistical modeling using the t -distribution. *Journal of the American Statistical Association*, **84**(408), 881–896.
- LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.
- LAWRANCE, A.J. (1987). The score statistic for regression transformation. **72**(2), 275–279.
- LE CAM, L. (1990). Maximum likelihood: an introduction. *International Statistical Review*. **58**(2), 153–171.
- LINDSAY, B.G. (1986). Exponential family mixture models (with least-squares estimators). *The Annals of Statistics*, **14**(1), 124–137.
- LITTLE, R.J.A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Appl. Statist.*, **37**(1), 23–38.

- LITTLE, R.J.A.; RUBIN, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley.
- LORD, R.D. (1974). The use of the Hankel transform in statistics. I. General theory and examples. *Biometrika*, **41**, 44–55.
- LOUIS, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *J.R. Statist. Soc., B*, **44**(2), 226–233.
- LUONG, A.; MAK, T. (1991). Robust estimation in a linear functional relationship model. *Comm. Statist. Theory Meth.*, **20**(2), 721–733.
- LYE, J.N.; MARTIN, V.L. (1993). Robust estimation nonnormalities, and generalized exponential distributions. *Journal of the American Statistical Association*, **88**(421), 261–267.
- MAK, T.K. (1983). On Sprent's generalized least-squares estimates. *J.R. Statist.*, **45**(3), 380–381.
- McCULLAGH, P.; COX, D.R. (1986). Invariants and likelihood ratio statistics. *The Annals of Statistics*, **14**, 1419–1480.
- MAC EACHERN, S.N. (1993). A characterization of some conjugate prior distributions for exponential families. *Seand J. Statist.*, **20**, 77–82.
- MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, March, 173–205.
- MAGNUS, J.R.; NEUDECKER, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, **7**(2), 381–394.
- MARDIA, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**(3), 519–530.
- MARONNA, R.A. (1976). Robust M -estimators of multivariate location and scatter. *The Annals of Statistics*, **4**(1), 51–67.
- McDONALD, J.; NEWHEY, W.K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory*, **4**, 428–457.
- MITCHELL, A.F.S.; KRZANOWSKI, W. (1985). The Mahalanobis distance and elliptic distributions. *Biometrika*, **72**(2), 464–467.

- MITCHELL, A.F.S. (1988). Statistical manifolds of univariate elliptic distributions. *International Statistical Review*, **56**(1), 1–16.
- MITCHELL, A.F.S. (1989). The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.*, **41**(2), 289–304.
- MORAN, P.A.P. (1971). Estimating structural and functional relationships. *Journal of Multivariate Analysis*, **1**, 232–255.
- MORRIS, C.N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *The Annals of Statistics*, **11**(2), 515–529.
- MORRIS, C.N. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics*, **10**(1), 65–88.
- MORTAZA, J.; JENNRICH, R.I. (1993). Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, **88**(421), Theory and Methods, 221–228.
- MOUTON, L.H.; WEISSFELD, L.A.; LAURENT, R.T. (1993). Bartlett correction factors in logistic regression models. *Computational Statistics & Data Analysis*, **15**, 1–11.
- MUIRHEAD, R.J. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients: a review. *Multivariate Statistical Analysis*, 143–159.
- MUIRHEAD, R.J. (1982). *Aspects of multivariate statistical theory*. New York: Wiley.
- MUIRHEAD, R.J.; WATERNAUX, C.M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, **67**(1), 31–43.
- NEUDECKER, H.; WANSBEEK, T. (1983). Some results on commutation matrices, with statistical applications. *The Canadian Journal of Statistics*, **11**(3), 221–231.
- NEUDECKER, H.; WANSBEEK, T. (1987). Fourth-order properties of normally distributed random matrices. *Linear Algebra and Its Applications*, **97**, 13–21.
- NEYMAN, J.; SCOTT, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometric*, **16**, 1–16.

- NGUYEN, T.; SAMPSON, A.R. (1991). A note on characterizations of multivariate stable distributions. *Ann. Inst. Statist. Math.*, **43**(4), 793–801.
- NIMMO-SMITH, I. (1979). Linear regressions and sphericity. *Biometrika*, **66**(2), 390–392.
- NISHII, R. (1988). Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivariate Analysis*, **27**, 392–403.
- NOMAKUCHI, K.; SAKATA, T. (1988). Characterizations of the forms of covariance matrix of an elliptically contoured distribution. *Sankhyā : The Indian Journal of Statistics, Series A*, **50**, Part 2, 205–210.
- OLKIN, I.; RUBIN, H. (1964). Multivariate beta distributions and independence properties of the wishart distribution. *The Annals of Mathematic and Statistics*, **35**, 261–269.
- OSIEWALSKI, J.; STEEL, M.F.J. (1993). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**, 345–363.
- OSIEWALSKI, J.; STEEL, M.F.J. (1993). Robust Bayesian inference in l_q -spherical models. *Biometrika*, **80**(2), 456–460.
- PARKE, W. (1986). Pseudo maximum likelihood estimation: the asymptotic distribution. *The Annals of Statistics*, **14**(1), 355–357.
- PARKE, W. (1986). Two modified fine estimators for use in small samples. *International Economic Review*, **27**(2), 465–481.
- PATEFIELD, W.M. (1981). Multivariate linear relationships: maximum likelihood estimation and regression bounds. *J.R. Statist. Soc., B*, **43**(3), 342–352.
- PIERCE, D. STRAM, D.O.; VAETH, M.; SCHAFER, D.W. (1992). The errors-in-variables problem: considerations provided by radiation dose-response analyses of the A-bomb survivor data. *Journal of the American Statistical Association*, **87**(418), 351–359.
- RALESCU, S.; BRANDWEIN, A.C. (1992). Stein estimation for non-normal spherically symmetric location families in three dimensions. *Journal of Multivariate Analysis*, **42**, 35–50.

- RELLES, D.A. (1970). Variance reduction techniques for Monte Carlo sampling from student distributions. *Technometrics*, **12**(3), 499–515.
- RELLES, D.A.; ROGERS, W.H. (1977). Statisticians are fairly robust estimators of location. *Journal of the American Statistical Association*, **72**(357), 107–111.
- ROGERS, W.H.; TUCKEY, J.W. (1972). Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica*, **26**, 211–226.
- von ROSEN, D. (1988). Moments for matrix normal variables. *Statistics*, **19**, 575–583.
- ROSS, W.H. (1987). The expectation of the likelihood ratio criterion. *International Statistical Review*, **55**(3), 315–330.
- RUBIN, D.B. (1983). Iteratively reweighted least squares. In: *Encyclopedia of the Statistical Sciences*, 4. New York: Wiley, 272–275.
- SALAMA, I.A.; KUMARSEN, P. (1992). Spherical uniformity and some characterizations of the Cauchy distribution. *Journal of Multivariate Analysis*, **41**, 212–219.
- SCHMID, C.H.; ROSNER, B. (1993). A Bayesian approach to logistic regression models having measurement error following a mixture distribution. *Statistics and Medicine*, **12**, 1141–1153.
- SCHNEEWEISS, H. (1982). Note on Creasy's confidence limits for the gradient in the linear functional relationship. *Journal of Multivariate Analysis*, **12**, 155–158.
- SCHOENBERG, I.J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.*, **39**, 811–841.
- SCHWEDER. (1976). Some “optimal” methods to detect structural shift or outliers in regression. *Journal of the American Statistical Association*, **71**(354), 491–501.
- SEBER, G.A.F. (1977). *Linear regression analysis*. New York: Wiley.
- SEBER, G.A.F. (1984). *Multivariate observations*. New York: Wiley.
- SEPANSKI, J.H. (1992). Score tests in a generalized linear model with surrogate covariates. *Statistics & Probability*, **15**, 1–10.
- SERFLING, R.J. (1984). Generalized L -, M -, and R -statistics. *The Annals of Statistics*, **12**(1), 76–86.

- SESHADRI, V. (1991). Finite mixtures of natural exponential families. *The Canadian Journal of Statistics*, **19**(4), 437–445.
- SHAPIRO, A.; BROWNE, M.W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association*, **82**(400), 1092–1097.
- SINGH, R.S. (1991). James-Stein rule estimators in linear regression models with multivariate t distributed error. *Austral. J. Statist.*, **33**(2), 145–158.
- SKOVGAARD, I. (1986). A note on the differentiation of cumulants of log likelihood derivatives. *International Statistical Review*, **54**(1), 29–32.
- SMITH, A.F.M. (1981). On random sequences with centred spherical symmetry. *Journal of the Royal Statistical Society , B*, **43**(2), 208–209.
- SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. *Journal of the Royal Statistical Society , B*, **28**, 278–297.
- SUTRADHAR, B.C. (1986). On the characteristic function of multivariate student t distribution. *The Canadian Journal of Statistics*, **14**(4), 329–337.
- SUTRADHAR, B.C. (1988). Author's revision. *The Canadian Journal of Statistics*, **16**, 323.
- SUTRADHAR, B.C.; ALI, M.M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Comm. Statist. - Theor. Meth.*, **15**(2), 429–450.
- SUTRADHAR, B.C.; ALI, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model. *Journal of Multivariate Analysis*, **29**, 155–162.
- SUTRADHAR, B.C. (1993). Score test for the covariance matrix of the elliptical t distribution. *Journal of Multivariate Analysis*, **46**, 1–12.
- TAN, W.Y. (1969). Note on the multivariate and the generalized multivariate beta distributions. *Journal of the American Statistical Association*, 236–241.
- TAYLOR, J.M.G. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics & Data Analysis*, **13**, 33–46.

- TIAO, G.G.; GUTMAN, I. (1965). The inverted Dirichlet distribution with applications. *Journal of the American Statistical Association*, september, 793–805.
- TIWARI, R.C.; JAMMALAMADAKA, S.R.; CHIB, S. (1988). Bayes prediction density and regression estimation. A semiparametric approach *empec.* v.13, 209–222.
- TITTERINGTON, D.M. (1990). Some recent research in the analysis of mixture distributions. *Statistics*, **21**(4), 619–641.
- TONG, Y.L. (1980). *Probabilistic inequalities in multivariate distributions*. Academic Press: New York.
- TRACY, D.S.; SULTAN, S.A. (1993). Third moment of matrix quadratic form. *Statistic & Probability Letters*, **16**, 71–76.
- TRACY, D.S.; SULTAN, S.A. (1993). Higher order moments of multivariate normal distribution using matrix derivatives. *Stochastic Analysis and Applications*, **11**, 337–348.
- TYLER, D.E. (1982). Radial estimates and the test for sphericity. *Biometrika*, **69**(2), 429–436.
- TYLER, D.E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, **70**(2), 411–420.
- TYLER, D.E. (1987). A distribution-free M -estimator of multivariate scatter. *The Annals of Statistics*, **15**(1), 234–251.
- UHLIG, H. (1994). On singular Wishart and singular multivariate beta distributions. *The Annals of Statistics*, **22**(1), 395–405.
- ULLAH, A.; ZINDE-WALSH, V. (1984). On the robustness of LM, LR, and W test in regression models. *Econometrica*, **52**(4), 1055–1066.
- ULLAH, A.; ZINDE-WALSH, V. (1985). Estimation and testing in a regression model with spherically symmetric errors. *Economic Letters*, **17**, 127–132.
- VAN PRAAG, M.S.; WESSELMAN, B.M. (1989). Elliptical multivariate analysis. *Journal of Econometrics*, **41**, 189–203.
- VILLEGAS, C. (1969). Maximum likelihood estimation of a linear functional relationship. *Ann. Math. Statist.*, **40**, 462–466.

- VITALE, R.A. (1993). Orthogonally invariant sequences as Gaussian scale mixtures: an alternate proof. *Journal of Multivariate Analysis*, **45**, 234–238.
- VON ROSEN, D. (1988). Moments for matrix normal variables. *Statistics*, **19**(4), 575–583.
- WEST, M. (1984). Outliers models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society , B*, **46**(3), 431–439.
- WIENS, D.P.; ZHENG, Z. (1986). Robust M -estimators of multivariate location and scatter in the presence of asymmetry. *The Canadian Journal of Statistics*, **14**(2), 161–176.
- WIJSMAN, R.A. (1957). Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis. *Ann. Inst. Statist. Math.*, **10**, 415–423.
- WOLFE, S.J. (1975). On the unimodality of spherically symmetric stable distribution functions. *Journal of Multivariate Analysis*, **5**, 236–242.
- WOLTER, K.M.; FULLER, W.A. (1982). Estimation of nonlinear errors-in-variables models. *The Annals of Statistics*, **10**(2), 539–548.
- WONG, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika*, **76**(1), 141–148.
- WONG, C.S.; LIU, D. (1994). Moments for left elliptically contoured random matrices. *Journal of Multivariate Analysis*, 1–23.
- YAMAGUCHI, K. (1990). Generalized EM algorithm for model with contaminated error term. In *Proceedings of The Seven Japan and Korea Joint Conference of Statistics*, 107–114.
- YUH, L.; HOGG, R.V. (1988). On adaptive M -regression. *Biometrika*, **44**, 433–445.
- ZELLNER, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley.
- ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error terms. *Journal of the American Statistical Association*, **71**(354), 400–405.

ZHAO, Y. (1994). Covariance matrices of quadratic forms in elliptical distributions.
Statistics and Probability Letters, **21**, 131–140.