**Análise Bayesiana de referência
para modelos de calibração**

Elias Chaibub Neto

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do grau
de
Mestre em Estatística

Área de Concentração: **Estatística**
Orientadora: Profa. Dra. **Márcia D'Elia Branco**

São Paulo - setembro - 2003

Análise Bayesiana de referência para modelos de calibração

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Elias Chaibub Neto e aprovada pela comissão julgadora.

Banca examonadora:

- Prof. Dr. Márcia D'Elia Branco (orientadora) - IME-USP
- Prof. Dr. José Galvão Leite - DEs-UFSCar
- Prof. Dr. Josemar Rodrigues - DEs-UFSCar
- Prof. Dr. Jorge Alberto Aschar (suplente) - ICMC-USP
- Prof. Dr. Luis Gustavo Esteves (suplente) - IME-USP

*A toda minha família
(no sentido amplo).*

Agradecimentos

A Prof. Márcia pela excelente orientação, encorajamento, amizade e constante atenção oferecida nas incontáveis vezes em que procurei sua ajuda.

A todos os professores do IME que de uma forma ou de outra contribuíram para a minha formação estatística. Em especial ao Prof. Sérgio Wechsler, que juntamente com a Prof. Márcia, tanto contribuiu em minha formação bayesiana.

Aos membros da banca examinadora, Prof. José Galvão Leite, Prof. Josemar Rodrigues e Prof. Luis Gustavo Esteves pelas sugestões de trabalhos futuros e de pequenas modificações que aprimoraram o texto. Ao CNPq pelo apoio financeiro durante a elaboração deste trabalho.

Aos meus pais por todo o incentivo e liberdade proporcionados para que eu sempre buscasse minha realização em todos os aspectos da vida. A minha esposa por todo carinho, dedicação e paciência durante todos estes anos em que estamos juntos. A minha sogra pela revisão ortográfica deste trabalho.

Aos meus amigos da graduação e da pós-graduação que tornaram estes anos de IME mais agradáveis.

Resumo

No presente trabalho apresentamos o método de construção de prioris de referência desenvolvido por Berger & Bernardo (1989, 1992a, 1992b) e o aplicamos aos modelos estatísticos de regressão binária, regressão binomial, calibração binomial e calibração linear. Obtivemos prioris de referência próprias para os modelos de regressão binária, regressão binomial e calibração binomial. Provamos que a distribuição a posteriori de referência associada ao modelo de calibração linear é própria e que a distribuição marginal a posteriori de x_0 só possui momentos finitos até a ordem $p - 1$. Apresentamos esquemas computacionais para a implementação da inferência a posteriori para todos os modelos, assim como um estudo de simulação para avaliar o desempenho da priori de referência associada ao modelo de calibração binomial.

Abstract

In this work we present Berger & Bernardo's method for the construction of reference priors (Berger and Bernardo, 1989, 1992a, 1992b), as well as its application to the binary regression, binomial regression, binomial calibration and linear calibration statistical models. The reference priors for the binary and binomial regression and binomial calibration turned out to be proper. We proved that the reference posterior for the linear calibration model is proper and the posterior marginal distribution of x_0 have finite moments of order equal or smaller than $p - 1$. We also present computational schemes for the implementation of posterior inferences, and a simulation study to access the performance of the binomial calibration reference prior.

Sumário

1	Introdução	3
2	Prioris de referência de Berger & Bernardo	5
2.1	Introdução	5
2.2	O método de Berger & Bernardo	11
2.2.1	Caso uniparamétrico	12
2.2.2	Parâmetros de perturbação	21
2.2.3	O caso multiparamétrico	27
3	Análise de referência para o modelo de calibração linear	32
3.1	Introdução	32
3.2	O modelo de calibração linear	33
3.3	A matriz de informação de Fisher associada ao modelo de calibração linear .	34
3.4	As prioris de referência e de Jeffreys associadas ao modelo de calibração linear	36
3.4.1	A priori de referência segundo a abordagem com 2-grupos	36
3.4.2	A priori de referência segundo a abordagem com m -grupos	45
3.4.3	A priori de Jeffreys	49
3.5	Resultados relacionados na literatura	54
3.6	Distribuições condicionais completas para o amostrador de Gibbs	57
4	As prioris-g de Zellner	59
4.1	Introdução	59
4.2	A priori-g	59
5	Análise de referência para modelos de regressão binária e binomial	62
5.1	Introdução	62
5.2	Regressão em dados binários	63

5.3	O modelo de regressão binária	64
5.4	A priori de referência para β no modelo de regressão binária com ligação <i>probit</i>	64
5.5	A estrutura normal latente do modelo <i>probit</i> de regressão binária	70
5.6	O amostrador de Gibbs para o modelo de regressão binária	72
5.7	O modelo de regressão binomial	73
5.8	A priori de referência para o modelo <i>probit</i> de regressão binomial	73
5.9	A estrutura normal latente do modelo <i>probit</i> de regressão binomial	74
5.10	Distribuições condicionais completas para o amostrador de Gibbs no modelo de regressão binomial	75
6	Análise de referência para o modelo de calibração binomial	76
6.1	Introdução	76
6.2	O modelo binomial de calibração	76
6.3	A priori de referência para o modelo de calibração binomial	77
6.4	A estrutura normal latente do modelo <i>probit</i> de calibração binomial	86
6.5	Distribuições condicionais completas para o amostrador de Gibbs	87
6.6	Estudo de simulação	88
7	Considerações finais e perspectivas de trabalhos futuros	99
A	Amostrador de Gibbs: modelo de calibração linear	101
B	Amostrador de Gibbs: modelos de regressão binária e binomial	105
B.1	O modelo de regressão binária	105
B.2	O modelo de regressão binomial	108
C	Amostrador de Gibbs: modelo de calibração binomial	111

Capítulo 1

Introdução

O presente estudo possui dois objetivos principais: o primeiro é apresentar o método de construção de *prioris* de referência desenvolvido por Berger e Bernardo (1989, 1992a, 1992b); o segundo é desenvolver *prioris* de referência para os modelos estatísticos de regressão binária, regressão binomial, calibração binomial e calibração linear.

É de nosso conhecimento, que entre os modelos estudados neste trabalho, apenas o de calibração linear já foi abordado na literatura. Liseo (1993) e Ghosh, Carlin e Srivastava (1995) trabalharam com casos particulares do modelo linear; enquanto Kubokawa e Robert (1994) trabalharam com o modelo de calibração reparametrizado como um problema de estimação da razão das médias de duas distribuições normais. Deste modo os resultados relativos ao modelo de calibração linear desenvolvidos neste trabalho também são originais.

No capítulo 2 faremos uma revisão geral a respeito do método de Berger e Bernardo. Seguindo Smith e Bernardo (1994) e Bernardo e Ramón (1998), abordaremos o método de um ponto de vista heurístico, procurando enfocar mais as idéias subjacentes à construção do método, do que as condições técnicas que tornariam a teoria matematicamente rigorosa. Primeiramente apresentaremos uma discussão informal sobre a motivação, interpretação e filosofia por trás desta metodologia. Em seguida apresentaremos uma série de proposições que serão utilizadas nas construções desenvolvidas nos capítulos subsequentes.

No capítulo três apresentaremos a construção da *priori* de referência associada ao modelo de calibração linear através de duas abordagens diferentes. Na primeira abordagem (Berger e Bernardo, 1989) a *priori* é obtida separando-se os parâmetros em dois grupos distintos; o grupo dos parâmetros de interesse e o grupo dos parâmetros de perturbação. Nosso interesse em apresentar a *priori* obtida a partir desta abordagem é compará-la com a *priori* obtida por Kubokawa e Robert (1994). No entanto, Berger e Bernardo atualmente não recomendam

a utilização desta abordagem, uma vez que foram encontradas situações onde ela produzia resultados insatisfatórios. A segunda abordagem recomendada atualmente pelos autores (Berger e Bernardo, 1992b), representa um refinamento da abordagem anterior e baseia-se na separação do vetor de parâmetros em diversos grupos, ordenados segundo sua importância inferencial.

Além de apresentar a construção das *prioris* de referência (segundo as duas abordagens) para o modelo de calibração linear, provaremos que as respectivas distribuições a posteriori são próprias e que as posteriores marginais para x_0 apresentam momentos finitos somente até a ordem $p - 1$. Construiremos também a *priori* de Jeffreys associada a este modelo e mostraremos que neste caso a distribuição a posteriori também é própria. Por fim, apresentaremos um esquema computacional para a implementação da inferência a posteriori obtida pelo uso da *priori* de referência construída através da segunda abordagem.

No quarto capítulo apresentaremos, brevemente, uma classe de distribuições não-subjetivas, embora informativas, desenvolvidas por Zellner (1986) e denominadas de *prioris-g*. Dedicamos um capítulo a estas *prioris*, pois como veremos mais tarde, as *prioris* de referência desenvolvidas para os modelos de regressão binária e binomial pertencem a classe das *prioris-g*.

No capítulo cinco desenvolveremos a construção da *priori* de referência associada aos modelos de regressão binária e binomial. Basearemos nossa construção em uma proposição apresentada no capítulo 2, que garante que a *priori* de referência é invariante sob transformações um-a-um. Deste modo induziremos a *priori* de referência para os parâmetros da regressão, a partir da *priori* de referência associada as probabilidades de sucesso nestes modelos. Mostraremos também, que as *prioris* de referência para os modelos de regressão binária e binomial são idênticas. Além disso apresentaremos esquemas computacionais para os dois modelos, utilizando o algoritmo proposto por Albert e Chib (1993), para a geração da distribuição a posteriori através do amostrador de Gibbs.

No sexto capítulo desenvolveremos a *priori* de referência associada ao modelo de calibração binomial, assim como, um esquema computacional para a implementação da inferência a posteriori. Neste modelo nem todas as distribuições condicionais completas tem formas conhecidas e portanto utilizaremos um algoritmo híbrido, conhecido como Metrópolis dentro do amostrador de Gibbs para gerar da distribuição a posteriori. Neste capítulo procederemos também um estudo de simulação comparando a performance da *priori* de referência, contra a *priori* “não-subjetiva” usualmente empregada para este modelo. Implementaremos todos algoritmos MCMC em S-Plus. Finalmente no capítulo sete apresentaremos algumas considerações finais e perspectivas de trabalhos futuros.

Capítulo 2

Prioris de referência de Berger & Bernardo

2.1 Introdução

Sob a perspectiva bayesiana, a distribuição a posteriori para a quantidade de interesse representa a mais completa inferência que podemos fazer a respeito desta quantidade. A distribuição a posteriori combina a informação contida nos dados com a informação prévia que possuíamos sobre a quantidade de interesse. A utilização de prioris que de algum modo espelhem uma falta de conhecimento a respeito da quantidade de interesse, tem sido uma constante na história da inferência bayesiana e seu uso remete aos trabalhos de pioneiros como Bayes (1763) e Laplace (1774, 1812). A busca por uma metodologia que capture de uma maneira bem definida o processo de aprendizagem bayesiana (atualização de conhecimento), de maneira que a informação a priori tenha o menor efeito possível (quando comparada aos dados) sobre a posteriori, tem sido um campo de pesquisa bastante ativo em estatística bayesiana (veja por exemplo, Jeffreys 1946 e 1961, Jaynes 1968, Berger e Bernardo 1989 e 1992a e 1992b).

Entre as muitas propostas encontradas na literatura, o método de análise de referência tem se mostrado bastante promissor. Este método foi inicialmente proposto por Bernardo (1979) e posteriormente aprimorado por Berger e Bernardo (B&B) em uma série de trabalhos (Berger e Bernardo 1989, 1992a, 1992b). O método de análise de referência generaliza muitos dos métodos alternativos propostos e inclui, como casos especiais, as soluções apropriadas que tais métodos apresentam em contextos específicos. Por exemplo, em problemas

uniparamétricos onde o espaço paramétrico é composto por um número finito de pontos, a priori de referência coincide com a priori de máxima entropia proposta por Jaynes (1968). Do mesmo modo, sabe-se que a priori de Jeffreys sempre produz inferências adequadas no caso uniparamétrico contínuo regular (Bernardo, 1997) e neste contexto a priori de referência se reduz à de Jeffreys. No caso multiparamétrico contínuo regular, a priori de referência coincide com as soluções *ad hoc* propostas por Jeffreys como alternativas a sua priori multiparamétrica (o próprio Jeffreys reconheceu que sua priori multiparamétrica não produzia resultados coerentes). Um outro ponto em favor do método de referência, é a sua flexibilidade em também incorporar casos que dispensam certas condições de regularidade exigidas pelos demais métodos (Bernardo, 1997).

A noção de uma priori não-informativa, isto é, uma priori que descreve falta de conhecimento a respeito da quantidade de interesse, tem sido o motivo de intermináveis debates dentro da escola bayesiana. Todavia atualmente, esta noção de não-informatividade está sendo substituída pela idéia de uma distribuição a posteriori dominada pelos dados (Bernardo, 1997). Em outras palavras, as prioris não-subjetivas são interpretadas como prioris onde a contribuição dos dados é dominante na construção da posteriori. Note que isto não significa que a priori não-subjetiva seja uma descrição de ignorância. Na verdade, toda priori mesmo a que tem o menor efeito possível sobre a distribuição a posteriori, reflete alguma forma de conhecimento. Estas prioris não-subjetivas apenas tentam tornar preciso o tipo de conhecimento a priori que, para um dado modelo, produz uma distribuição a posteriori dominada pelos dados (Bernardo, 1997). Note que a posteriori deve ser dominada pelos dados mesmo para amostras pequenas, uma vez que para amostras grandes a distribuição a posteriori é usualmente dominada pelos dados qualquer que seja a priori.

As prioris não-subjetivas tipicamente não dependem dos dados, mas dependem do modelo probabilístico que gerou os dados. A maior crítica ao uso destas prioris é feita por bayesianos subjetivistas que argumentam que a priori deve ser a expressão honesta do conhecimento que uma pessoa possui a respeito da quantidade de interesse e não uma função do modelo. A utilização de prioris não-subjetivas que dependem do modelo estatístico pode levar à violação do Princípio da Verossimilhança. O princípio é violado uma vez que a inferência leva em consideração o espaço amostral inteiro, isto é, todos os possíveis resultados que poderiam ter sido observados, mas não foram, ao invés de levar em conta apenas os resultados que, de fato, foram observados. Este problema leva muitos bayesianos a considerar que métodos bayesianos objetivos representam um novo paradigma de inferência. Todavia, existem duas fortes motivações para a utilização de prioris de referência. A primeira argumenta que a

utilização de *prioris* que produzem *posteriores* dominadas pelos dados deve ser vista como uma parte necessária de uma saudável análise de sensibilidade da escolha da *priori*. Na realidade, a utilização de *prioris* de referência permite uma resposta à seguinte importante questão de comunicação científica: O que alguém poderia concluir a partir dos dados se seu conhecimento a *priori* fosse tal que a distribuição a *posteriori* para a quantidade de interesse fosse dominada pelos dados? De posse da resposta a esta questão poderíamos conduzir a análise de sensibilidade comparando a distribuição a *posteriori* de referência com uma distribuição a *posteriori* subjetiva de modo a acessar a importância relativa das nossas opiniões iniciais na inferência final. A segunda motivação para o uso de *prioris* de referência tem um caráter mais pragmático e argumenta que em análises bayesianas de complexos modelos multiparamétricos é praticamente impossível “elicitare” detalhadamente uma *priori* subjetiva multivariada. Deste modo a utilização de uma *priori* de referência representa uma alternativa mais viável na prática (Bernardo, 1998).

Uma característica comum observada em *prioris* de referência, é que via de regra, embora existam exceções, quando o espaço paramétrico é limitado, as *prioris* de referência são próprias, isto é, integram um, e portanto são realmente medidas de probabilidade. Por outro lado, embora também existam exceções, quando o espaço paramétrico não é limitado as *prioris* de referência são em geral impróprias (*prioris* cuja integral é infinita).

Segundo Bernardo (1998), as *prioris* de referência não devem ser interpretadas como distribuições de probabilidade (o que de fato é impossível quando elas são impróprias). Tecnicamente estas *prioris* são apenas funções positivas, formalmente utilizadas no teorema de Bayes, para obtenção de distribuições a *posteriori* não-subjetivas dominadas pelos dados.

A utilização de *prioris* impróprias pode levar a comportamentos patológicos da distribuição a *posteriori* como paradoxos de marginalização, inconsistência forte e até mesmo a *posteriores* impróprias. Tais patologias podem ocorrer pois o uso de *prioris* impróprias viola o cálculo de probabilidades e deste modo perde-se a proteção da coerência (no sentido de obedecer as leis do cálculo de probabilidades) que é automática com *prioris* próprias.¹

Os paradoxos de marginalização, primeiramente evidenciados por Stone e Dawid (1972) e Dawid, Stone e Zidek (1973), foram inicialmente apresentados como um alerta aos perigos do uso de *prioris* impróprias em inferência bayesiana. Resumidamente, os paradoxos de marginalização podem ser caracterizados pela violação do seguinte comportamento de

¹Aqui devemos ressaltar, que as *prioris* desenvolvidas a partir do método de B&B parecem ser imunes a estas patologias. De fato, estes problemas tem sido observados em *prioris* derivadas a partir de métodos alternativos ao de B&B.

marginalização: considere um modelo paramétrico $p(\mathbf{x} \mid \phi, \lambda)$, onde ϕ é a quantidade de interesse. Se a distribuição a posteriori para a quantidade de interesse $p_1(\phi \mid \mathbf{x})$ é da forma $p_1(\phi \mid \mathbf{x}) = p_1(\phi \mid \mathbf{t})$ onde $\mathbf{t} = \mathbf{t}(\mathbf{x})$ é uma função de $\mathbf{x}$, e a distribuição amostral de $\mathbf{t}$, $p(\mathbf{t} \mid \phi, \lambda) = p(\mathbf{t} \mid \phi)$, depende apenas de ϕ e não de λ , então a posteriori $p_2(\phi \mid \mathbf{t})$ obtida a partir do modelo marginal $p(\mathbf{t} \mid \phi)$ deve coincidir com a posteriori $p_1(\phi \mid \mathbf{x})$ obtida a partir do modelo original. Todavia, este comportamento é muitas vezes violado quando utilizamos prióris impróprias. Stone e Dawid (1972) apresentaram dois exemplos desta violação e Dawid, Stone e Zidek (1973) mostraram que esses paradoxos são uma constante em modelos multiparamétricos.

No entanto o método proposto por B&B parece ser imune a esta patologia. De fato, apesar de não ser possível provar que construção de B&B não conduza a paradoxos de marginalização (uma vez que não existe uma teoria geral que caracterize as condições exatas que conduzam a este problema) o fato é que, mesmo tendo sido posta a prova por mais de 30 anos, nenhum paradoxo de marginalização foi encontrado utilizando a metodologia de B&B (Bernardo, 1997). Em todos os exemplos em que paradoxos de marginalização apareciam ao utilizar prióris impróprias, o método de B&B produziu prióris (quase sempre impróprias) que não conduziram a este problema².

Um outro exemplo de comportamento patológico associado ao uso de prióris impróprias é conhecido como o fenômeno de inconsistência forte (Stone, 1976). Este tipo de comportamento é caracterizado pela discrepância entre a distribuição a posteriori de θ e entre a probabilidade de cobertura de θ^3 . Para uma mesma região C , a probabilidade a posteriori $P(C \mid x)$ pode ser completamente diferente de $P(C \mid \theta)$.

Todavia, prióris de referência parecem ser imunes a este problema também. Bernardo e Smith (1994) discutem que, por distinguir cuidadosamente entre parâmetros de interesse e perturbação, o método de B&B evita este tipo de inconsistência. Segundo Bernardo (2001), uma análise frequentista do comportamento de procedimentos bayesianos pode ser clarificadora uma vez que representa uma interessante ponte entre inferência clássica e bayesiana. Em particular foi observado que o comportamento frequentista do método de B&B é geralmente muito bom. De fato um intervalo de probabilidade $1 - \alpha$ da distribuição a posteriori de uma quantidade de interesse (obtida a partir de uma priori de referência) é muitas vezes idêntico ou muito próximo a um intervalo de confiança com nível de significância α calculado

²Veja Bernardo e Smith (1994), pg 362, exemplo 5.26, para um exemplo.

³A probabilidade de cobertura de θ é calculada a partir do modelo estatístico e representa a probabilidade de um intervalo C conter θ , isto é, $P(C \mid \theta)$.

a partir do modelo paramétrico associado a quantidade de interesse (Bernardo, 1998).

Uma outra propriedade muito prezada em prioris não-subjetivas diz respeito a invariância. De fato existem dois tipos distintos de invariância que todo método de construção de prioris não-subjetivas deve ser capaz de satisfazer. O primeiro diz respeito a invariância sobre transformações um-a-um da quantidade de interesse. Formalmente temos que se $\phi = g(\theta)$ é uma transformação um-a-um de θ e $\pi_\phi(\phi)$ e $\pi_\theta(\theta)$ são as respectivas prioris não-subjetivas, então

$$\pi_\phi(\phi) = \pi_\theta(g^{-1}(\phi)) \left| \frac{\partial g^{-1}(\phi)}{\partial \phi} \right|,$$

ou seja, as duas prioris estão logicamente relacionadas através do jacobiano da transformação. Informalmente esta propriedade garante que a inferência feita através de uma posteriori não-subjetiva não depende da parametrização escolhida.

O outro tipo de invariância diz respeito a invariância sob estatísticas suficientes. Neste caso, se t é uma estatística suficiente associada ao modelo $p(x | \phi)$, então a posteriori não-subjetiva $\pi(\phi | x)$ coincide com a posteriori não-subjetiva $\pi(\phi | t)$ obtida a partir do modelo $p(t | \phi)$. Informalmente isto significa que é indiferente utilizarmos a distribuição dos dados ou a das estatísticas suficientes para a obtenção de uma priori não-subjetiva, uma vez que as distribuições a posteriori coincidirão nos dois casos.

Segundo Bernardo (1998), todo método de construção de prioris não-subjetivas deve ser capaz de satisfazer as propriedades descritas acima (invariância, marginalização consistente, bom comportamento frequentista) e também a propriedade de generalidade. Já vimos que as prioris de referência satisfazem as três primeiras propriedades. E no que se refere a generalidade (isto é, o método deve ser aplicável a qualquer problema inferencial bem definido) temos que o método de B&B é superior aos demais visto que seu funcionamento dispensa certas condições de regularidade exigidas pelos demais métodos e por este motivo pode ser aplicado a um maior leque de situações.⁴

Agora cabem algumas considerações a respeito da filosofia do método de B&B. Bernardo (1979) iniciou o método de construção de prioris de referência seguindo a tradição de Laplace e Jeffreys. Esta tradição segue uma linha pragmática onde os resultados são o mais impor-

⁴Veja Bernardo e Smith (1994) pg 311, exemplo 5.12, para um exemplo onde a priori de referência associada ao modelo uniforme foi obtida sem fazer uso da suposição de normalidade assintótica da distribuição a posteriori (e portanto, sem fazer uso da matriz de informação de Fisher). Observe que para o modelo uniforme, o suporte de $p(x | \theta)$ depende de θ e portanto uma das condições de regularidade necessárias para a obtenção da matriz de informação de Fisher não está sendo satisfeita. Deste modo, vemos que ao contrário do método de referência, o método de Jeffreys não pode ser aplicado a este problema.

tante; o método deve funcionar. Deste modo, Laplace simplesmente utilizava uma priori constante para θ ($p(\theta) = 1$), uma vez que ela funcionava excepcionalmente bem para os problemas que ele investigou. Todavia, outros pesquisadores perceberam que esta priori não funcionava bem para amostras muito pequenas e também apresentava inconsistência nas respostas à um mesmo problema, quando se utilizavam parametrizações diferentes (tipicamente uma priori constante para um parâmetro não continuava constante após uma transformação). Estes novos problemas fizeram com que Jeffreys propusesse sua famosa priori. Esta priori é invariante a transformações paramétricas e funciona muito bem em problemas uniparamétricos. No entanto, o próprio Jeffreys encontrou problemas com o seu método no caso multiparamétrico e para solucioná-los, propôs algumas modificações *ad hoc* para sua priori multiparamétrica. Bernardo (1979) eliminou a necessidade destas modificações ao dividir o espaço paramétrico em parâmetros de interesse e de perturbação. No entanto, o método original proposto em Bernardo (1979) foi sendo progressivamente refinado e redefinido como resposta a um contínuo processo de derivação de contra-exemplos às priors de referência. Desta maneira, a medida que contra-exemplos (problemas onde o método não produzia boas respostas) apareciam, B&B refinavam o método introduzindo novas condições e redefinindo algoritmos. Isto explica a aparente arbitrariedade encontrada em alguns detalhes do método. De fato, onde diferentes escolhas eram possíveis, foi pelo estudo extensivo dos exemplos de aplicações que a ambiguidade foi solucionada (Berger e Bernardo, 1992a).

Uma consequência direta desta abordagem pragmática adotada por B&B na construção de seu método, é que muitas vezes nem mesmo os próprios autores possuem uma intuição do porquê determinada regra produz bons resultados. Como um exemplo podemos citar a discussão feita por McCulloch e os autores em Berger e Bernardo (1992a). Nesta discussão McCulloch pergunta se os autores possuem alguma intuição do porquê, ao se levar em conta a ordenação dos parâmetros o método produz bons resultados. B&B simplesmente responderam que

“Nós realmente não sabemos a resposta. O exame de exemplos como o de Neyman-Scott revela o problema de se levar em consideração todos os parâmetros conjuntamente, mas nosso *insight* não é muito mais profundo do que isto. Mais geralmente, um de nós dois nunca teve certeza do porque o método funciona, e tem continuamente se mostrado muito satisfeito e surpreso com o seu sucesso. Mas no final, devemos concordar com o a asserção de McCulloch ‘e no entanto o método gera priors interessantes’.”.

No entanto, a discussão acima não deve ser interpretada como uma admissão de que o método

de B&B é fundamentalmente *ad hoc*, longe disto, o método é muito bem fundamentado em teoria da informação e é matematicamente rigoroso. A idéia central por trás do método é muito atraente e permanece inalterada desde que foi proposta inicialmente por Bernardo em 1979. Em termos gerais, ela pode ser descrita informalmente através da seguinte idéia: a quantidade de informação a respeito de uma quantidade de interesse θ , que esperamos aprender com um experimento conduzido com a intensão de prover informação a respeito de θ , é obviamente uma função do nosso conhecimento a priori a respeito de θ . Neste sentido, se já temos um bom conhecimento a priori sobre θ então esperamos não aprender muita coisa com os dados obtidos pelo experimento; por outro lado, se nosso conhecimento a respeito de θ é escasso, então os dados podem nos ensinar muito. Em outras palavras, quanto maior for a quantidade de informação a priori disponível, menor será a quantidade de informação fornecida pelos dados. Um experimento infinitamente grande poderia eventualmente prover toda a informação que ainda não temos a respeito da quantidade de interesse θ . Bernardo denominou esta quantidade de *missing information*. Assim, é natural definir uma priori que expresse pouco conhecimento, ou melhor, que torne a posteriori dominada pelos dados, como a priori que maximize a *missing information* sobre a quantidade de interesse. No entanto, como a *missing information* é definida como um limite, que não é necessariamente finito, a priori de referência é definida como um tipo especial de limite de uma sequência de prioris que maximizam a informação esperada de um experimento cada vez maior. O funcional que quantifica a *missing information* é representado pela esperança da medida de divergência de Kullback-Liebler entre duas distribuições de probabilidade.

Na próxima seção vamos expor formalmente as idéias descritas acima apresentando a construção formal da priori de referência para os casos uniparamétrico, biparamétrico e multiparamétrico. Também apresentaremos uma série de propriedades das prioris de referência que serão utilizadas por nós nos capítulos subsequentes.

2.2 O método de Berger & Bernardo

Nesta seção iniciaremos a construção formal das prioris de referência de B&B. Seguindo Bernardo e Smith (1994) e Bernardo e Ramón (1998), enfocaremos o método de um ponto de vista heurístico, no sentido de que não explicitaremos as condições técnicas que tornariam a teoria matematicamente rigorosa. Nosso objetivo é tornar claro as idéias subjacentes a construção do método. O leitor interessado em technicalidades deve consultar Berger e Bernardo (1989, 1992a, 1992b) e Berger et al (1989). As proposições e definições enunciadas

próximas seções foram extraídas principalmente de Bernardo e Smith (1994) e Bernardo e Ramón (1998).

2.2.1 Caso uniparamétrico

Considere um experimento ε que consiste da observação de um conjunto de dados $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ ⁵, $\mathbf{x} \in \mathbf{X}$ de um modelo $p(\mathbf{x} | \phi)$, $\phi \in \Phi \subset \mathfrak{R}$. A quantidade de informação esperada a respeito de ϕ fornecida pelo experimento ε , quando nosso conhecimento prévio a respeito de ϕ é representado por $p(\phi)$, é definida como

$$I\{\varepsilon, p(\phi)\} = \int_{\mathbf{X}} p(\mathbf{x}) \int_{\Phi} p(\phi | \mathbf{x}) \log \left\{ \frac{p(\phi | \mathbf{x})}{p(\phi)} \right\} d\phi d\mathbf{x}, \quad (2.1)$$

onde

$$p(\mathbf{x}) = \int_{\Phi} p(\mathbf{x} | \phi) p(\phi) d\phi, \quad (2.2)$$

$$p(\phi | \mathbf{x}) = p(\mathbf{x} | \phi) p(\phi) / p(\mathbf{x}) \quad (2.3)$$

são, respectivamente, as distribuições preditiva e posteriori. A quantidade de informação esperada possui algumas propriedades que a tornam particularmente apropriada como uma medida de informação. Entre elas podemos ressaltar que $I\{\varepsilon, p(\phi)\}$ é uma quantidade não negativa invariante quanto a transformações um-a-um, e que também permanece invariante se uma estatística suficiente é usada no lugar de uma amostra aleatória.⁶

Vamos agora considerar a quantidade de informação que esperamos obter através de um experimento $\varepsilon(k)$, que consiste de k replicações independentes e imaginárias $\mathbf{z}_k = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ do experimento original ε , tal que

$$I\{\varepsilon(k), p(\phi)\} = \int_{\mathbf{X}^k} p(\mathbf{z}_k) \int_{\Phi} p(\phi | \mathbf{z}_k) \log \left\{ \frac{p(\phi | \mathbf{z}_k)}{p(\phi)} \right\} d\phi d\mathbf{z}_k. \quad (2.4)$$

A medida que k aumenta, o experimento $\varepsilon(k)$ torna-se cada vez mais informativo e quando $k \rightarrow \infty$, temos que $I\{\varepsilon(k), p(\phi)\}$ proverá toda a informação que ainda não possuímos, isto é, *missing information* a respeito de ϕ associada a $p(\phi)$.

⁵Frequentemente é assumido que os dados representam uma amostra aleatória de n observações de $p(x | \phi)$. Neste caso, $p(\mathbf{x} | \phi) = \prod_{j=1}^n p(x_j | \phi)$, $I\{\varepsilon, p(\phi)\} = n \int_{\mathbf{X}} p(\mathbf{x}) \int_{\Phi} p(\phi | \mathbf{x}) \log \left\{ \frac{p(\phi | \mathbf{x})}{p(\phi)} \right\} d\phi d\mathbf{x}$. Todavia esta suposição não é necessária para a método de B&B, que pode ser utilizado com dados correlacionados.

⁶Veja Lindley (1956) para uma interpretação de $I\{\varepsilon, p(\phi)\}$ como uma medida de informação, e para demonstrações de suas propriedades. Veja também Kullback (1959) e Soofi (2000), para interpretações de $I\{\varepsilon, p(\phi)\}$.

Uma forma natural de definir uma priori não-subjetiva que produza uma distribuição a posteriori dominada pelos dados é maximizar a *missing information*. Desta maneira, a priori de referência é definida como um tipo especial de limite (a ser explicitado em breve) no qual, a medida que $k \rightarrow \infty$, a sequência de prioris $\pi_k(\phi)$, que maximizam $I\{\varepsilon(k), p(\phi)\}$, converge para a priori de referência $\pi(\phi)$ ⁷. Neste ponto, é importante notar que este limite não está sendo tomado em relação ao tamanho da amostra e portanto, não implica em que a metodologia só funcione assintoticamente.

Agora, observe que, sob certas condições de regularidade, a quantidade esperada de informação $I\{\varepsilon(k), p(\phi)\}$ pode ser manipulada algebricamente tal que (2.4) pode ser reescrita como

$$\int_{\mathbf{X}^k} p(\mathbf{z}_k) \left(\int_{\Phi} p(\phi | \mathbf{z}_k) \log p(\phi | \mathbf{z}_k) d\phi - \int_{\Phi} p(\phi | \mathbf{z}_k) \log p(\phi) d\phi \right) d\mathbf{z}_k. \quad (2.5)$$

Denotando por $H[p(\phi | \mathbf{z}_k)]$ a entropia de Shannon (Shannon, 1948), que é dada por $H[p(\phi | \mathbf{z}_k)] = - \int p(\phi | \mathbf{z}_k) \log p(\phi | \mathbf{z}_k) d\phi$, temos que (2.5) pode ser reescrita como

$$- \int_{\mathbf{X}^k} p(\mathbf{z}_k) H[p(\phi | \mathbf{z}_k)] d\mathbf{z}_k - \int_{\mathbf{X}^k} p(\mathbf{z}_k) \int_{\Phi} \frac{p(\phi, \mathbf{z}_k)}{p(\mathbf{z}_k)} \log p(\phi) d\phi d\mathbf{z}_k \quad (2.6)$$

$$= - \int_{\mathbf{X}^k} p(\mathbf{z}_k) H[p(\phi | \mathbf{z}_k)] d\mathbf{z}_k - \int_{\mathbf{X}^k} \int_{\Phi} p(\phi, \mathbf{z}_k) \log p(\phi) d\phi d\mathbf{z}_k, \quad (2.7)$$

e trocando a ordem de integração

$$= - \int_{\mathbf{X}^k} p(\mathbf{z}_k) H(p(\phi | \mathbf{z}_k)) d\mathbf{z}_k - \int_{\Phi} \log p(\phi) \left(\int_{\mathbf{X}^k} p(\phi, \mathbf{z}_k) d\mathbf{z}_k \right) d\phi \quad (2.8)$$

$$= - \int_{\mathbf{X}^k} p(\mathbf{z}_k) H[p(\phi | \mathbf{z}_k)] d\mathbf{z}_k - \int_{\Phi} p(\phi) \log p(\phi) d\phi \quad (2.9)$$

$$= H[p(\phi)] - \int_{\mathbf{X}^k} p(\mathbf{z}_k) H[p(\phi | \mathbf{z}_k)] d\mathbf{z}_k \quad (2.10)$$

$$= H[p(\phi)] + \int_{\mathbf{X}^k} \int_{\Phi} p(\mathbf{z}_k) p(\phi | \mathbf{z}_k) \log p(\phi | \mathbf{z}_k) d\phi d\mathbf{z}_k. \quad (2.11)$$

Como $p(\mathbf{z}_k)p(\phi | \mathbf{z}_k) = p(\phi, \mathbf{z}_k) = p(\mathbf{z}_k | \phi)p(\phi)$, temos

⁷Observe que foi preciso definir a priori de referência como um tipo de limite pois, em geral, se ϕ é contínuo, $\lim_{k \rightarrow \infty} I\{\varepsilon(k), p(\phi)\} = \infty$, de modo que não podemos trabalhar diretamente com este funcional. Observe ainda que este resultado era esperado uma vez que uma quantidade infinita de informação é necessária para conhecermos o verdadeiro valor de um número real (Bernardo, 1979).

$$= H[p(\phi)] + \int_{\mathbf{X}^k} \int_{\Phi} p(\mathbf{z}_k | \phi) p(\phi) \log p(\phi | \mathbf{z}_k) d\phi d\mathbf{z}_k \quad (2.12)$$

e trocando novamente a ordem de integração

$$= H[p(\phi)] + \int_{\Phi} p(\phi) \int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log p(\phi | \mathbf{z}_k) d\mathbf{z}_k d\phi \quad (2.13)$$

$$= \int_{\Phi} p(\phi) (-\log p(\phi) + \log \left\{ \exp \left(\int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log p(\phi | \mathbf{z}_k) d\mathbf{z}_k \right) \right\}) d\phi \quad (2.14)$$

$$= \int_{\Phi} p(\phi) \log \left\{ \frac{\exp \left(\int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log p(\phi | \mathbf{z}_k) d\mathbf{z}_k \right)}{p(\phi)} \right\} d\phi. \quad (2.15)$$

Deste modo, temos que $I\{\varepsilon(k), p(\phi)\}$ pode ser reescrita como

$$I\{\varepsilon(k), p(\phi)\} = \int_{\Phi} p(\phi) \log \left\{ \frac{f_k(\phi)}{p(\phi)} \right\} d\phi \quad (2.16)$$

onde $f_k(\phi) = \exp \left\{ \int p(\mathbf{z}_k | \phi) \log p(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\}$.

Agora vamos mostrar que $I\{\varepsilon(k), p(\phi)\}$ é maximizada se, e somente se, $p(\phi) \propto f_k(\phi)$. Para tanto, vamos utilizar um teorema do cálculo de variações (veja Sveshnikov, 1978) que garante que para encontrar uma função $y = y(x)$ que maximiza ou minimiza a integral

$$I = \int_a^b \Phi(x, y) dx \quad (2.17)$$

sob as restrições

$$\int_a^b \psi_s(x, y) dx = c_s, \quad s = 1, \dots, m, \quad (2.18)$$

é necessário resolver a equação de Euler

$$\frac{\partial \Phi_1}{\partial y} = 0, \quad (2.19)$$

onde $\Phi_1 = \Phi + \sum \lambda_s \psi_s$ e λ_s são constantes obtidas a partir das restrições.

Em nosso caso, queremos maximizar a equação (2.16), sob a restrição $\int p(\phi) d\phi = 1$. Deste modo,

$$\Phi_1 = p(\phi) \log \left\{ \frac{f_k(\phi)}{p(\phi)} \right\} + \lambda p(\phi). \quad (2.20)$$

Derivando em relação a $p(\phi)$

$$\frac{\partial \Phi_1}{\partial p(\phi)} = \log f_k(\phi) - \log p(\phi) - 1 + \lambda \quad (2.21)$$

e igualando a zero temos

$$\log f_k(\phi) - \log p(\phi) = 1 - \lambda, \quad (2.22)$$

e portanto

$$p(\phi) = e^{1-\lambda} f_k(\phi). \quad (2.23)$$

Logo,

$$p(\phi) \propto f_k(\phi). \quad (2.24)$$

Observe que $p(\phi) \propto f_k(\phi)$ maximiza a integral $I\{\varepsilon(k), p(\phi)\}$ uma vez que $I\{\varepsilon(k), p(\phi)\}$ é uma função côncava se vista como uma função de $p(\phi)$ (veja Lindley, 1956).

Deste modo, temos que

$$p(\phi) \propto \exp \left\{ \int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log p(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\}. \quad (2.25)$$

Todavia, devemos perceber que a relação acima representa uma solução implícita para o problema de maximização uma vez que o termo a direita depende da distribuição a posteriori $p(\phi | \mathbf{z}_k)$ que é uma função de $p(\phi)$.

No entanto, este problema pode ser contornado se considerarmos uma aproximação assintótica $q(\phi | \mathbf{z}_k)$ para a distribuição a posteriori. Observe que esta aproximação é muito razoável, uma vez que estamos apenas interessados em valores grandes de k (lembre que o método é baseado na idéia de *missing information*, que é definida como a medida $I\{\varepsilon(k), p(\phi)\}$ quando $k \rightarrow \infty$).

A aproximação assintótica resolve o problema pois assintoticamente a posteriori é independente de $p(\phi)$, uma vez que em grandes amostras a posteriori é dominada pelos dados⁸. Deste modo, e sob condições de regularidade, esperamos que a sequência de funções positivas $\{\pi_1(\phi), \pi_2(\phi), \dots\}$ induza, pelo uso formal do teorema de Bayes, uma sequência de distribuições a posteriori $\{\pi_1(\phi | \mathbf{x}), \pi_2(\phi | \mathbf{x}), \dots\}$ ⁹ onde

⁸Desde que a regra de Cromwell seja respeitada, isto é, não adotamos probabilidades zero ou um para qualquer evento.

⁹Observe que a sequência de distribuições a posteriori foi obtida a partir da sequência de priors e do modelo paramétrico $p(\mathbf{x} | \phi)$, onde $\mathbf{x}$ representa uma amostra de dados realmente observados. Perceba que $\mathbf{x}$ pode ter um tamanho qualquer, não sendo necessariamente grande. Queremos frisar que o componente assintótico do método diz respeito as replicações imaginárias $\mathbf{z}_k$ do experimento ε (que foram utilizadas como parte do argumento na construção da sequência de funções positivas $\{\pi_1(\phi), \pi_2(\phi), \dots\}$) e não aos dados realmente observados.

$$\pi_k(\phi | \mathbf{x}) = \frac{p(\mathbf{x} | \phi)\pi_k(\phi)}{\int_{\Phi} p(\mathbf{x} | \phi)\pi_k(\phi)d\phi}. \quad (2.26)$$

e onde, para $k_0 \in \mathcal{N}$ suficientemente grande,

$$\pi_k(\phi) = \exp \left\{ \int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log q(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\}, \quad k > k_0. \quad (2.27)$$

A distribuição a posteriori de referência $\pi(\phi | \mathbf{x})$ é definida como um tipo especial de limite da sequência $\{\pi_1(\phi | \mathbf{x}), \pi_2(\phi | \mathbf{x}), \dots\}$. Na verdade, trata-se de um limite no sentido informacional, mais precisamente, de um limite da divergência logarítmica entre duas distribuições de probabilidade definido como

$$\lim_{k \rightarrow \infty} \left[\int_{\Phi} \pi_k(\phi | \mathbf{x}) \log \left\{ \frac{\pi_k(\phi | \mathbf{x})}{\pi(\phi | \mathbf{x})} \right\} d\phi \right] = 0. \quad (2.28)$$

A priori de referência $\pi(\phi)$ é então definida a partir de

$$\pi(\phi | \mathbf{x}) = \frac{p(\mathbf{x} | \phi)\pi(\phi)}{\int_{\Phi} p(\mathbf{x} | \phi)\pi(\phi)d\phi}, \quad \phi \in \Phi, \quad (2.29)$$

isto é, é definida como a função positiva que quando combinada ao modelo paramétrico produz, via teorema de Bayes, a distribuição a posteriori de referência.

Observe que a priori de referência $\pi(\phi)$ definida acima pode ser descrita como o limite da sequência $\{\pi_1(\phi), \pi_2(\phi), \dots\}$ no sentido preciso de que o limite da divergência logarítmica da sequência de distribuições a posteriori $\{\pi_1(\phi | \mathbf{x}), \pi_2(\phi | \mathbf{x}), \dots\}$ é a distribuição a posteriori de referência $\pi(\phi | \mathbf{x})$, obtida a partir de $\pi(\phi)$ através do uso formal do teorema de Bayes. A esta forma especial de convergência Bernardo e Ramón (1998) denotam

$$\pi(\phi) = \text{blim}_{k \rightarrow \infty} \{\pi_k(\phi)\}. \quad (2.30)$$

Este tipo de convergência torna preciso o sentido no qual uma priori $\pi(\phi)$, possivelmente imprópria, pode ser descrita como um limite de uma sequência $\{\pi_1(\phi), \pi_2(\phi), \dots\}$ de funções

positivas e próprias¹⁰. Após estes esclarecimentos segue a definição da priori de referência.

Definição 2.1 (Distribuições de referência uniparamétricas) .

Seja $\mathbf{x}$ os dados obtidos em um experimento ε que consiste de uma observação de $p(\mathbf{x} | \phi)$, $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi \subseteq \mathfrak{R}$. Seja $\mathbf{z}_k = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ o resultado de k replicações independentes e imaginárias de ε , e defina

$$\pi_k(\phi) = \exp \left\{ \int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log q(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\},$$

onde $q(\phi | \mathbf{z}_k)$ é uma aproximação assintótica para a distribuição a posteriori $p(\phi | \mathbf{z}_k)$.

A distribuição a posteriori de referência $\pi(\phi | \mathbf{x})$ é definida por

$$\lim_{k \rightarrow \infty} \left[\int_{\Phi} \pi_k(\phi | \mathbf{x}) \log \left\{ \frac{\pi_k(\phi | \mathbf{x})}{\pi(\phi | \mathbf{x})} \right\} d\phi \right] = 0,$$

assumindo que o limite exista e onde $\pi_k(\phi | \mathbf{x})$ é dado por

$$\pi_k(\phi | \mathbf{x}) = \frac{p(\mathbf{x} | \phi) \pi_k(\phi)}{\int_{\Phi} p(\mathbf{x} | \phi) \pi_k(\phi) d\phi}.$$

A priori de referência é então definida como uma função positiva, que para todo $\phi \in \Phi$, satisfaz a equação

$$\pi(\phi | \mathbf{x}) = \frac{p(\mathbf{x} | \phi) \pi(\phi)}{\int_{\Phi} p(\mathbf{x} | \phi) \pi(\phi) d\phi}. \quad \diamond$$

¹⁰Pode parecer uma complicação desnecessária definir a priori de referência indiretamente a partir da sequência de posteriores, ao invés de defini-la diretamente a partir da sequência $\{\pi_1(\phi), \pi_2(\phi), \dots\}$. No entanto, uma definição direta a partir desta sequência pode conduzir a problemas. Por exemplo, a sequência de prioris $\pi_k(\phi) = \text{Beta}(1/k, 1/k)$ vai convergindo para uma distribuição simétrica cada vez mais pesada em direção ao 0 e ao 1 a medida que k cresce, e por fim quando $k \rightarrow \infty$, seu limite é a distribuição discreta $\pi(\phi = 0) = \pi(\phi = 1) = \frac{1}{2}$. Esta priori discreta é claramente insatisfatória uma vez que ϕ é um parâmetro contínuo. Todavia, a definição indireta da priori de referência, baseada na posteriori de referência, produz uma priori bastante razoável. Suponha, por exemplo, que observamos r sucessos em n ensaios de *Bernoulli*(ϕ), então a distribuição a posteriori de referência é dada por $\pi(\phi | r) = \text{Beta}(r, n - r)$, uma vez que a sequência de posteriores $\pi_k(\phi | r) \propto \phi^r (1 - \phi)^{n-r} \phi^{\frac{1}{k}-1} (1 - \phi)^{\frac{1}{k}-1} = \phi^{r+\frac{1}{k}-1} (1 - \phi)^{n-r+\frac{1}{k}-1}$ converge para $\phi^{r-1} (1 - \phi)^{n-r-1}$, quando $k \rightarrow \infty$. Observe que esta posteriori, por sua vez, induz a priori operacional $\pi(\phi) \propto \phi^{-1} (1 - \phi)^{-1}$ (Bernardo, 1979).

Observe que a priori de referência é definida indiretamente, como uma função positiva que quando combinada ao modelo paramétrico, via teorema de Bayes, produz a distribuição a posteriori de referência. Deste modo, está claro, que esta definição não é operacional.

No que segue, enunciaremos algumas proposições associadas as prioris de referência uniparamétricas, que nos permitem trabalhar de forma operacional com estas prioris. Em primeiro lugar, vamos apresentar a forma explícita para a priori de referência uniparamétrica.

Proposição 2.1 (A forma explícita da priori de referência) .

A priori de referência para ϕ relativa a um experimento que consiste de uma observação de $p(\mathbf{x} | \phi)$, $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi \subseteq \mathfrak{R}$, é dada, contanto que a convergência no sentido informacional seja verificada, por

$$\pi(\phi) = c \lim_{k \rightarrow \infty} \frac{\pi_k(\phi)}{\pi_k(\phi_0)}, \quad \phi \in \Phi$$

onde $c > 0$, $\phi_0 \in \Phi$ e

$$\pi_k(\phi) = \exp \left\{ \int_{\mathbf{X}^k} p(\mathbf{z}_k | \phi) \log q(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\},$$

com $\mathbf{z}_k = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ uma amostra aleatória de $p(\mathbf{x} | \phi)$, e $q(\phi | \mathbf{z}_k)$ uma aproximação assintótica para a distribuição a posteriori $p(\phi | \mathbf{z}_k)$.

Prova: Veja Bernardo e Smith (1994) página 307.

Vamos agora considerar a forma explícita da priori de referência no caso particular em que o espaço paramétrico é discreto. Como veremos na proposição que segue, quando o espaço paramétrico é discreto, a priori de referência é sempre uma distribuição uniforme discreta, qualquer que seja o modelo paramétrico.

Proposição 2.2 (Priori de referência no caso discreto) .

Seja $\mathbf{x}$ o resultado de uma observação de $p(\mathbf{x} | \phi)$, onde $\phi \in \Phi = \{\phi_1, \dots, \phi_M\}$. Então, qualquer função da forma $\pi(\phi_i) = a$, $a > 0$, $i = 1, \dots, M$, é uma priori de referência e a distribuição a posteriori de referência é dada por

$$\pi(\phi_i | \mathbf{x}) = c(\mathbf{x})p(\mathbf{x} | \phi), \quad i = 1, \dots, M$$

onde $c(\mathbf{x})$ é a constante de normalização.

Prova: Veja Bernardo e Smith (1994) página 307.

Note que neste caso, a priori de referência e a priori de máxima entropia proposta por Jaynes (1968) coincidem. É interessante observar que apenas em contextos de espaços paramétricos discretos, a priori de máxima entropia produz resultados sensíveis (Bernardo, 1997). A idéia por trás do método de Jaynes é que, quanto maior a entropia de uma distribuição, maior é a sua medida de incerteza, e consequentemente, menor é a quantidade de informação contida na distribuição. Quando o espaço paramétrico é discreto, $I\{\varepsilon(k), p(\phi)\} = H[p(\phi)]$ de modo que maximizar a medida de entropia é equivalente a maximizar a *missing information*.

Quando consideramos amostras aleatórias de distribuições que admitem estatísticas suficientes de dimensão fixa, é natural indagar se a priori de referência derivada a partir da distribuição das estatísticas suficientes, coincide com a priori de referência obtida a partir da amostra aleatória. A próxima proposição estabelece que, de fato, as prioris coincidem.

Proposição 2.3 (Compatibilidade com estatísticas suficientes) .

Seja $\mathbf{e}_n$, $n \geq 1$, o experimento que consiste da observação de uma amostra aleatória $\mathbf{x}_1, \dots, \mathbf{x}_n$ de $p(\mathbf{x} | \phi)$, $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi$, onde, para todo n , $\mathbf{t}_n = \mathbf{t}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ é uma estatística suficiente. Então, para todo n , a priori de referência obtida considerando $\mathbf{x}_1, \dots, \mathbf{x}_n$ coincide com a priori obtida através de $\mathbf{t}_n$.

Prova: Veja Bernardo e Smith (1994) página 309.

Informalmente, a proposição acima significa que é indiferente utilizarmos a distribuição dos dados ou das estatísticas suficientes para a obtenção da priori de referência.

Dado um modelo paramétrico $p(\mathbf{x} | \theta)$, podemos reparametrizá-lo como $p(\mathbf{x} | \phi)$ onde $\phi = \phi(\theta)$, para qualquer transformação um-a-um $g : \Theta \rightarrow \Phi$. Deste modo é natural indagar se as prioris de referência para θ e ϕ , obtidas respectivamente a partir de $p(\mathbf{x} | \theta)$ e $p(\mathbf{x} | \phi)$, são consistentes no sentido de que estão relacionadas através do jacobiano da transformação. A próxima proposição estabelece este resultado.

Proposição 2.4 (Invariância quanto a transformações um-a-um) .

Suponha que $\pi_\theta(\theta)$ e $\pi_\phi(\phi)$ sejam prioris de referência derivadas, respectivamente, a partir de um experimento que consiste da observação de $p(\mathbf{x} | \theta)$, com $\mathbf{x} \in \mathbf{X}$, $\theta \in \Theta$ e a partir de uma observação de $p(\mathbf{x} | \phi)$ com $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi$, onde $\phi = g(\theta)$ e $g : \Theta \rightarrow \Phi$ é um mapeamento um-a-um. Então, para $c > 0$ e para todo $\phi \in \Phi$:

1. $\pi_\phi(\phi) = c\pi_\theta(g^{-1}(\phi))$, se Φ é discreta;
2. $\pi_\phi(\phi) = c\pi_\theta(g^{-1}(\phi))|J_\phi|$, onde $J_\phi = \frac{\partial g^{-1}(\phi)}{\partial \phi}$ é o jacobiano da transformação.

Prova: Veja Bernardo e Smith (1994) página 310.

Informalmente, a proposição acima garante que a priori de referência não depende da parametrização escolhida.

Tipicamente, sob condições de regularidade, a distribuição a posteriori assintótica $q(\phi | \mathbf{z}_{kn})$, que corresponde a k -replicações de um experimento ε_n envolvendo uma amostra aleatória de tamanho n de $p(\mathbf{x} | \phi)$, só dependerá de $\mathbf{z}_{kn}$ através de uma estimativa assintoticamente suficiente e consistente de ϕ . Nestes casos, a priori de referência pode ser facilmente identificada a partir da forma da distribuição a posteriori assintótica.

Proposição 2.5 (Priori de ref. para um estimador consistente) .

Seja ε_n um experimento que consiste da observação de uma amostra aleatória $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ de $p(\mathbf{x} | \phi)$, $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi \subseteq \mathfrak{R}$, e seja $\mathbf{z}_{kn}$ o resultado de k replicações de ε_n . Se existe $\hat{\phi}_{kn} = \hat{\phi}_{kn}(\mathbf{z}_{kn})$ tal que, com probabilidade um

$$\lim_{k \rightarrow \infty} \hat{\phi}_{kn} = \phi$$

e quando $k \rightarrow \infty$,

$$\int p(\mathbf{z}_{kn} | \phi) \log \frac{q(\phi | \mathbf{z}_{kn})}{q(\phi | \hat{\phi}_{kn})} d\mathbf{z}_{kn} \rightarrow 0,$$

então, para $c > 0$, $\phi_0 \in \Phi$, a priori de referência é definida como

$$\pi(\phi) = c \lim_{k \rightarrow \infty} \frac{\pi_{kn}(\phi)}{\pi_{kn}(\phi_0)},$$

onde

$$\pi_{kn}(\phi) = q(\phi | \hat{\phi}_{kn}) \Big|_{\hat{\phi}_{kn}=\phi}.$$

Prova: Veja Bernardo e Smith (1994) página 312.

Sob condições de regularidade, a distribuição a posteriori assintótica de ϕ converge para a distribuição normal. Neste caso é possível obter uma caracterização da priori de referência diretamente a partir do modelo paramétrico associado a ϕ .

Proposição 2.6 (Priori de referência sob normalidade assintótica) .

Seja ε_n um experimento que consiste da observação de uma amostra aleatória $\mathbf{x}_1, \dots, \mathbf{x}_n$ de $p(\mathbf{x} \mid \phi)$, $\mathbf{x} \in \mathbf{X}$, $\phi \in \Phi \subseteq \mathfrak{R}$. Se a distribuição a posteriori assintótica de ϕ , dado k replicações do experimento ε_n , é normal com variância $1/knh(\hat{\phi}_{kn})$, onde $\hat{\phi}_{kn}$ é um estimador consistente de ϕ e $h(\cdot)$ é a informação de Fisher associada ao modelo $p(\mathbf{x} \mid \phi)$, então a priori de referência tem a forma

$$\pi(\phi) \propto \{h(\phi)\}^{1/2}.$$

Prova: Veja Bernardo e Smith (1994) página 314.

Observe que neste caso a priori de referência coincide com a priori de Jeffreys. É interessante notar que as condições onde a priori de Jeffreys sempre produz bons resultados, coincidem exatamente com as condições da proposição acima (Bernardo, 1997). Este fato, aliado ao resultado análogo obtido no caso discreto, mostra que o método de referência parece sempre incluir como casos especiais, as soluções apropriadas que os demais métodos de construção de prioris não-subjetivas produzem em contextos específicos.

2.2.2 Parâmetros de perturbação

No desenvolvimento feito até aqui, consideramos apenas o caso uniparamétrico. Nesta subseção abordaremos o método de B&B, considerando o caso bidimensional, com $\theta = (\phi, \lambda)$, $\phi \in \Phi$, $\lambda \in \Lambda$, onde ϕ é o parâmetro de interesse e λ o parâmetro de perturbação. Neste caso, o modelo paramétrico é dado por

$$p(\mathbf{x} \mid \phi, \lambda), \quad (\phi, \lambda) \in \Phi \times \Lambda \subseteq \mathfrak{R} \times \mathfrak{R}. \quad (2.31)$$

Veremos a seguir, que a priori de referência para este problema bidimensional pode ser obtida aplicando-se iterativamente a metodologia desenvolvida para o caso uniparamétrico.

A idéia é construir a priori de referência em dois passos. Primeiramente assumimos que ϕ é conhecido, e trabalhamos condicionalmente em relação a este parâmetro. Deste modo, caímos em um contexto unidimensional e podemos utilizar a solução já apresentada na seção anterior para obter a priori de referência condicional $\pi(\lambda \mid \phi)$. De posse desta priori, e lembrando que a priori de referência conjunta pode ser decomposta como

$$\pi(\phi, \lambda) = \pi(\phi)\pi(\lambda \mid \phi), \quad (2.32)$$

precisamos apenas especificar a priori de referência marginal de ϕ , para obter a priori de referência para o problema bidimensional.

Em princípio, a priori condicional $\pi(\lambda \mid \phi)$ poderia ser utilizada para obter o modelo marginal de ϕ através de

$$\begin{aligned} p(\mathbf{x} \mid \phi) &= \int p(\mathbf{x}, \lambda \mid \phi) d\lambda \\ &= \int p(\mathbf{x} \mid \phi, \lambda) \pi(\lambda \mid \phi) d\lambda, \end{aligned} \quad (2.33)$$

e a partir deste modelo marginal, poderíamos obter a priori de referência marginal para ϕ (observe que estaríamos novamente em um contexto unidimensional). Todavia, existe uma séria complicação a esta abordagem. Como já vimos anteriormente, na maioria das vezes, a priori de referência é imprópria, e portanto não pode ser utilizada para eliminar por integração o parâmetro de perturbação do modelo original. Deste modo, a abordagem acima só é útil nos casos menos comuns, quando a priori de referência é própria.

Para contornar este problema B&B sugeriram a seguinte estratégia: identifique uma sequência crescente $\{\Lambda_i\}$ de subconjuntos de Λ , $\bigcup \Lambda_i = \Lambda$, que pode depender de ϕ , e tal que, para cada Λ_i , a priori de referência condicional $\pi(\lambda \mid \phi)$ restrita a Λ_i , possa ser normalizada dando origem a priori de referência própria $\pi_i(\lambda \mid \phi)$. Então, para cada i , obtenha o modelo marginal $p_i(\mathbf{x} \mid \phi)$ e uma priori marginal $\pi_i(\phi)$. A priori de referência $\pi(\phi, \lambda)$ é então obtida como um limite quando $i \rightarrow \infty$. Este procedimento será agora formalizado.

Definição 2.2 (Distr. ref. dado um parâmetro de perturbação) .

Seja $\mathbf{x}$ o resultado de um experimento ε que consiste de uma observação de $p(\mathbf{x} \mid \phi, \lambda)$, $\mathbf{x} \in \mathbf{X}$, $(\phi, \lambda) \in \Phi \times \Lambda \subseteq \mathbb{R} \times \mathbb{R}$. A distribuição a posteriori de referência $\pi(\phi \mid \mathbf{x})$, para o parâmetro de interesse ϕ , relativa ao experimento ε e a sequência crescente de subconjuntos de Λ , $\{\Lambda_i(\phi)\}$, $\phi \in \Phi$, $\bigcup_i \Lambda_i(\phi) = \Lambda$, são definidos como o resultado do seguinte procedimento:

1. *aplicando a definição 1.1 ao modelo $p(\mathbf{x} \mid \phi, \lambda)$, com ϕ fixo, obtenha a priori de referência condicional $\pi(\lambda \mid \phi)$ para Λ ;*
2. *para cada ϕ , normalize $\pi(\lambda \mid \phi)$ dentro de cada $\Lambda_i(\phi)$ para obter uma sequência de priors próprias, $\pi_i(\lambda \mid \phi)$;*
3. *use esta sequência de priors para obter a sequência de modelos integrados*

$$p_i(\mathbf{z}_k \mid \phi) = \int_{\Lambda_i(\phi)} p(\mathbf{z}_k \mid \phi, \lambda) \pi_i(\lambda \mid \phi) d\lambda$$

para $\mathbf{z}_k$ resultante de k replicações do experimento ε ;

4. use esta sequência de modelos para derivar a sequência de prioris de referência

$$\pi_i(\phi) = c \lim_{k \rightarrow \infty} \frac{\pi_k(\phi)}{\pi_k(\phi_0)}, \quad i = 1, 2, \dots,$$

onde

$$\pi_k(\phi) = \exp \left\{ \int p_i(\mathbf{z}_k | \phi) \log q(\phi | \mathbf{z}_k) d\mathbf{z}_k \right\},$$

e para os dados $\mathbf{x}$, obtenha a sequência correspondente de distribuições a posteriori de referência

$$\pi_i(\phi | \mathbf{x}) \propto \pi_i(\phi) \int_{\Lambda_i(\phi)} p(\mathbf{x} | \phi, \lambda) \pi_i(\lambda | \phi) d\lambda; \quad i = 1, 2, \dots,$$

5. defina $\pi(\phi | \mathbf{x})$ como a distribuição tal que

$$\lim_{i \rightarrow \infty} \left[\int_{\Phi} \pi_i(\phi | \mathbf{x}) \log \left\{ \frac{\pi_i(\phi | \mathbf{x})}{\pi(\phi | \mathbf{x})} \right\} d\phi \right] = 0.$$

A priori de referência relativa a ordenação dos parâmetros (ϕ, λ) , é qualquer função positiva $\pi(\phi, \lambda)$, tal que

$$\pi(\phi | \mathbf{x}) \propto \int p(\mathbf{x} | \phi, \lambda) \pi(\phi, \lambda) d\lambda.$$

Tipicamente, a priori $\pi(\phi, \lambda)$ pode ser obtida simplesmente como

$$\pi(\phi, \lambda) = c \lim_{i \rightarrow \infty} \frac{\pi_i(\phi) \pi_i(\lambda | \phi)}{\pi_i(\phi_0) \pi_i(\lambda | \phi_0)}. \quad \diamond$$

Uma desvantagem desta abordagem, é que ela claramente requer a escolha da forma dos subconjuntos Λ_i e, em alguns casos, a priori de referência pode depender desta escolha (veja Berger e Bernardo, 1989). No entanto, os autores argumentam, que para cada problema específico, uma forma adequada para a sequência sempre transparece naturalmente.

Segundo Berger e Bernardo (1989), a forma “natural” da sequência $\{\Lambda_i\}$ deve espelhar de alguma forma a noção de não-informatividade a respeito do parâmetro de perturbação relativamente a parametrização utilizada. Por exemplo, quando a priori de referência condicional $\pi(\lambda | \phi)$ for uma uniforme em $\mathfrak{R}$, a forma pouca informativa natural para Λ_i , recomendada pelos autores, é o conjunto $\Lambda_i = [-i, i]$.

Nos casos onde a priori condicional não é uniforme em $\mathfrak{R}$, devemos procurar uma parametrização onde ela o seja, e a partir desta parametrização, induzir a forma natural de Λ_i na parametrização original. Como exemplo, considere o caso do modelo normal com média

(μ) e variância (σ) desconhecidos, e onde o parâmetro de interesse é o μ . Neste caso, temos que

$$\pi(\sigma \mid \mu) \propto \sigma^{-1}$$

e fazendo a transformação $\sigma \rightarrow \log \sigma$

$$\pi(\sigma \mid \mu) \propto \pi(\log \sigma \mid \mu) \left| \frac{\partial \log \sigma}{\partial \sigma} \right| \Rightarrow \sigma^{-1} \propto \pi(\log \sigma \mid \mu) \sigma^{-1} \Rightarrow \pi(\log \sigma \mid \mu) \propto 1.$$

Deste modo, na parametrização $\log \sigma$ a priori é uma uniforme em $\Re$, e portanto a forma “natural” para os elementos da sequência $\{\Lambda_i\}$ é $\Lambda_i = [-i, i]$. Segue desta observação que a forma natural na parametrização σ é dada por $[e^{-i}, e^i]$, uma vez que $\log \sigma = i$, implica que $\sigma = e^i$.

Bernardo (1997b) apresenta uma forma operacional de encontrar a parametrização canônica, isto é, a parametrização onde a priori de referência condicional é uma distribuição uniforme na reta. A idéia é considerar uma transformação do tipo

$$\psi(\lambda) = F(\lambda) + c, \quad (2.34)$$

onde

$$F(\lambda) = \int_{-\infty}^{\lambda} \pi(t \mid \phi) dt, \quad (2.35)$$

e $\pi(\cdot \mid \phi)$ é a priori de referência condicional na parametrização original.

Observe que na parametrização canônica a priori de referência condicional é sempre uma uniforme em $\Re$ uma vez que

$$\pi(\psi \mid \phi) = \pi(\lambda \mid \phi) \left| \frac{\partial \psi}{\partial \lambda} \right|^{-1} = \pi(\lambda \mid \phi) |\pi(\lambda \mid \phi)|^{-1} = 1 \quad (2.36)$$

É comum encontrarmos na literatura estatística, versões de um modelo paramétrico $p(\mathbf{x} \mid \phi, \lambda)$ reparametrizado como $p(\mathbf{x} \mid \phi, \psi)$, onde $(\phi, \lambda) \rightarrow (\phi, \psi)$ é uma transformação um-a-um. Intuitivamente é razoável esperar que a distribuição marginal a posteriori para o parâmetro de interesse, seja a mesma qualquer que seja a parametrização utilizada para o parâmetro de perturbação. A próxima proposição garante que isto de fato acontece.

Proposição 2.7 (Invariância quanto ao parâmetro de perturbação) *Seja ε o experimento que consiste de uma observação de $p(\mathbf{x} \mid \phi, \lambda)$, $(\phi, \lambda) \in \Phi \times \Lambda \subseteq \Re \times \Re$, e seja ε' um experimento que consiste de uma observação de $p(\mathbf{x} \mid \phi, \psi)$, $(\phi, \psi) \in \Phi \times \Psi \subseteq \Re \times \Re$,*

onde $(\phi, \lambda) \rightarrow (\phi, \psi)$ é uma transformação um-a-um, com $\psi = g_\phi(\lambda)$. Então, as distribuições a posteriori de referência marginais para ϕ , relativas a $[\varepsilon, \Lambda_i(\phi)]$ e $[\varepsilon', \Psi_i(\phi)]$, onde $\Psi_i(\phi) = g_\phi(\Lambda_i(\phi))$, são idênticas.

Prova: Veja Bernardo e Smith (1994) página 325.

Alternativamente, podemos querer manter a forma do parâmetro de perturbação, mas redefinir a forma do parâmetro de interesse segundo uma transformação um-a-um. Intuitivamente esperamos que as distribuições a posteriori relativas as diferentes parametrizações, estejam relacionadas através do elemento jacobiano adequado. A próxima proposição estabelece que isto de fato acontece.

Proposição 2.8 (Invariância quanto a transformações um-a-um) .

Seja ε o experimento que consiste de uma observação de $p(\mathbf{x} \mid \phi, \lambda)$, $\phi \in \Phi$, $\lambda \in \Lambda$, e seja ε' um experimento que consiste de uma observação de $p(\mathbf{x} \mid \gamma, \lambda)$, $\gamma \in \Gamma$, $\lambda \in \Lambda$, onde $\gamma = g(\phi)$. Então, condicionalmente aos dados $\mathbf{x}$, as distribuições de referência a posteriori para ϕ e γ , relativas a $[\varepsilon, \Lambda_i(\phi)]$ e $[\varepsilon', \Lambda_i(\gamma)]$, $\Lambda_i(\gamma) = \Lambda_i\{g(\phi)\}$ são relacionadas por:

1. $\pi_\gamma(\gamma \mid \mathbf{x}) = \pi_\phi(g^{-1}(\gamma) \mid \mathbf{x})$, se Φ for discreto;
2. $\pi_\gamma(\gamma \mid \mathbf{x}) = \pi_\phi(g^{-1}(\gamma) \mid \mathbf{x}) |J_{g^{-1}}(\gamma)|$, onde $J_{g^{-1}}(\gamma) = \frac{\partial g^{-1}(\gamma)}{\partial \gamma}$.

Prova: Veja Bernardo e Smith (1994) página 326.

Na proposição (2.6) vimos que a forma explícita da priori de referência pode ser caracterizada diretamente a partir do modelo paramétrico, quando este último for assintoticamente normal. Na próxima proposição estenderemos este resultado para o modelo bidimensional.

Proposição 2.9 (Priori ref. bivariadas sob normalidade assintótica) *Seja ε_n um experimento onde observa-se uma amostra aleatória $\mathbf{x}_1, \dots, \mathbf{x}_n$ proveniente de $p(\mathbf{x} \mid \phi, \lambda)$, $(\phi, \lambda) \in \Phi \times \Lambda \subseteq \mathbb{R} \times \mathbb{R}$, e seja $\{\Lambda_i(\phi)\}$ uma sequência convenientemente definida de subconjuntos de λ . Suponha que a distribuição a posteriori conjunta assintótica de (ϕ, λ) , dado k replicações do experimento ε_n , é normal multivariada com vetor de médias $(\hat{\phi}_{kn}, \hat{\lambda}_{kn})$ e matriz de precisão*

$$\mathbf{H}(\hat{\phi}_{kn}, \hat{\lambda}_{kn}) = \begin{pmatrix} h_{11}(\hat{\phi}_{kn}, \hat{\lambda}_{kn}) & h_{12}(\hat{\phi}_{kn}, \hat{\lambda}_{kn}) \\ h_{21}(\hat{\phi}_{kn}, \hat{\lambda}_{kn}) & h_{22}(\hat{\phi}_{kn}, \hat{\lambda}_{kn}) \end{pmatrix} = \begin{pmatrix} \hat{h}_{11} & \hat{h}_{12} \\ \hat{h}_{21} & \hat{h}_{22} \end{pmatrix}$$

onde $\mathbf{H}(\phi, \lambda)$ é a matriz de informação de Fisher e $(\hat{\phi}_{kn}, \hat{\lambda}_{kn})$ é uma estimativa consistente de (ϕ, λ) . Então

$$\pi(\lambda | \phi) \propto \{h_{22}(\phi, \lambda)\}^{1/2};$$

$$\pi(\phi, \lambda) = \pi(\lambda | \phi) \lim_{i \rightarrow \infty} \left\{ \frac{\pi_i(\phi) c_i(\phi)}{\pi_i(\phi_0) c_i(\phi_0)} \right\}, \quad \phi_0 \in \Phi,$$

define a priori de referência relativa a parametrização ordenada (ϕ, λ) , onde

$$\pi_i(\phi) \propto \exp \left\{ \int_{\Lambda_i(\phi)} \pi_i(\lambda | \phi) \log \left(\{h_\phi(\phi, \lambda)\}^{1/2} \right) d\lambda \right\},$$

com

$$\pi_i(\lambda | \phi) = c_i(\phi) \pi(\lambda | \phi) = \frac{\pi(\lambda | \phi)}{\int_{\Lambda_i(\phi)} \pi(\lambda | \phi) d\lambda},$$

e

$$h_\phi = h_{11} - h_{12} h_{22}^{-1} h_{21},$$

onde ϕ é o parâmetro de interesse e λ é o parâmetro de perturbação.

Prova: Veja Bernardo e Smith (1994) página 326.

Em muitos casos, $h_\phi(\phi, \lambda)$ e $h_{22}(\phi, \lambda)$ podem ser fatorados em funções de ϕ e λ . Se nestes casos, $\{\Lambda_i\}$ não depende de ϕ , torna-se muito fácil encontrar a forma da priori de referência.

Corolário 2.1 *Suponha que sob as condições da proposição (2.9), escolhemos uma sequência crescente adequada de subconjuntos $\{\Lambda_i\}$ de Λ , que não dependem de ϕ , e suponha também que*

$$\{h_\phi(\phi, \lambda)\}^{1/2} = a_1(\phi) b_1(\lambda), \quad \{h_{22}(\phi, \lambda)\}^{1/2} = a_2(\lambda) b_2(\phi).$$

Então a priori de referência relativa ao vetor paramétrico ordenado (ϕ, λ) é dada por

$$\pi(\phi, \lambda) \propto a_1(\phi) a_2(\lambda).$$

Prova: Veja Bernardo e Smith (1994) página 325.

Aqui vale a pena ressaltar que este corolário será utilizado na derivação da priori de referência associada ao modelo de calibração linear.

2.2.3 O caso multiparamétrico

Na subseção anterior apresentamos a construção da priori de referência para o caso bidimensional onde $\theta = (\phi, \lambda)$ representava um vetor ordenado, cujos parâmetros de interesse e perturbação eram respectivamente ϕ e λ . A priori de referência para esta parametrização ordenada foi construída através de um condicionamento, e da aplicação da metodologia desenvolvida para o caso unidimensional em cada componente da decomposição $\pi(\phi) \pi(\lambda | \phi)$.

Nesta subseção apresentaremos uma extensão desta idéia para o caso multiparamétrico, com $\theta = (\theta_1, \dots, \theta_k)$. Aqui a priori de referência será construída através de condicionamentos sucessivos, sempre aplicando a metodologia do caso uniparamétrico para a obtenção de cada componente da decomposição

$$\pi(\theta_1, \theta_2, \dots, \theta_k) = \pi(\theta_k | \theta_1, \dots, \theta_{k-1}) \cdots \pi(\theta_2 | \theta_1) \pi(\theta_1). \quad (2.37)$$

Vamos agora apresentar o algoritmo para a construção da priori de referência multidimensional no caso regular. Aqui, apresentaremos o algoritmo em sua forma mais geral, seguindo de perto Berger e Bernardo (1992b). Nesta forma mais geral, ao invés de separar e ordenar o vetor paramétrico θ de dimensão k em k grupos distintos, θ é dividido e ordenado em m -grupos de dimensão n_j , $j = 1, \dots, m$.¹¹ É importante observar que nesta abordagem ordenamos os m diferentes grupos segundo sua maior importância inferencial¹², mas que, dentro de cada grupo, a ordenação paramétrica não é relevante (Berger e Bernardo, 1992b).

Antes de apresentar o algoritmo dos m -grupos vamos introduzir um pouco de notação. Seja

$$p(\mathbf{x} | \theta), \quad \theta \in \Theta \subset \mathbb{R}^k \quad (2.38)$$

o modelo paramétrico. Assumimos que a matriz de informação de Fisher

$$\mathbf{H}(\theta) = -E_{\mathbf{x}|\theta} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(\mathbf{x} | \theta) \right\} \quad (2.39)$$

existe e possui posto completo, de modo que

$$\mathbf{S}(\theta) = \mathbf{H}^{-1}(\theta) \quad (2.40)$$

também existe.

¹¹Note que quando $n_j = 1$ para todo j , estamos no caso especial com k grupos

¹²Por ordenação segundo importância inferencial queremos dizer que ordenamos o vetor de parâmetros de modo a refletir nosso interesse a respeito dos parâmetros do modelo. Em particular, o parâmetro de maior interesse deve vir em primeiro lugar na ordenação.

Assumimos que os θ_i , $i = 1, \dots, k$, são separados em m grupos de tamanhos $n_1, \dots, n_m$, e que esses grupos são dados por

$$\begin{aligned}\boldsymbol{\theta}_{(1)} &= (\theta_1, \dots, \theta_{n_1}) \\ \boldsymbol{\theta}_{(2)} &= (\theta_{n_1+1}, \dots, \theta_{n_1+n_2}) \\ &\vdots \\ \boldsymbol{\theta}_{(i)} &= (\theta_{N_{i-1}+1}, \dots, \theta_{N_i}) \\ &\vdots \\ \boldsymbol{\theta}_{(m)} &= (\theta_{N_{m-1}+1}, \dots, \theta_k)\end{aligned}$$

onde $N_j = n_1 + n_2 + \dots + n_j$, para $j = 1, \dots, m$. Também definimos

$$\boldsymbol{\theta}^{[j]} = (\theta_1, \dots, \theta_j), \quad \boldsymbol{\theta}_{[j]} = (\theta_{j+1}, \dots, \theta_k).$$

Escrevemos $\mathbf{S}(\boldsymbol{\theta})$ como

$$\mathbf{S}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \cdots & \mathbf{A}_{1m} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} & \mathbf{A}_{23} & \cdots & \mathbf{A}_{2m} \\ \mathbf{A}'_{13} & \mathbf{A}'_{23} & \mathbf{A}_{33} & \cdots & \mathbf{A}_{3m} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{A}'_{1m} & \mathbf{A}'_{2m} & \mathbf{A}'_{3m} & \cdots & \mathbf{A}_{mm} \end{pmatrix}, \quad (2.41)$$

onde $\mathbf{A}_{ij}$ é uma matriz $n_i \times n_j$.

Definimos $\mathbf{S}_j(\boldsymbol{\theta})$ como a submatriz $N_j \times N_j$ superior a esquerda de $\mathbf{S}(\boldsymbol{\theta})$. Por exemplo, $\mathbf{S}_2(\boldsymbol{\theta})$ é dada por

$$\mathbf{S}_2(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix}$$

e $\mathbf{S}_m(\boldsymbol{\theta}) \equiv \mathbf{S}(\boldsymbol{\theta})$.

Definimos $\mathbf{h}_j(\boldsymbol{\theta})$ como sendo a matriz $n_j \times n_j$ inferior a esquerda de $\mathbf{H}_j(\boldsymbol{\theta}) = \mathbf{S}_j^{-1}(\boldsymbol{\theta})$. Por exemplo, seja

$$\mathbf{H}_2(\boldsymbol{\theta}) = \mathbf{S}_2^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{pmatrix},$$

então temos que $\mathbf{h}_2(\boldsymbol{\theta}) = \mathbf{B}_{22}$. Observe que $\mathbf{h}_1(\boldsymbol{\theta}) = \mathbf{H}_1(\boldsymbol{\theta}) = \mathbf{A}_{11}^{-1}$, e que se $\mathbf{S}(\boldsymbol{\theta})$ é uma matriz bloco diagonal, isto é $\mathbf{A}_{ij} \equiv \mathbf{0}$ para $i \neq j$, então $\mathbf{h}_j(\boldsymbol{\theta}) \equiv \mathbf{A}_{jj}^{-1}$ para todo $j = 1, \dots, m$.

Finalmente, assumamos que

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_{(1)} \times \boldsymbol{\Theta}_{(2)} \times \cdots \times \boldsymbol{\Theta}_{(m)},$$

onde

$$\boldsymbol{\Theta}_{(i)} = \Theta_{N_{i-1}+1} \times \cdots \times \Theta_{N_i},$$

com $\boldsymbol{\theta}_{(i)} \in \boldsymbol{\Theta}_{(i)}$. Denote por $\{\boldsymbol{\Theta}_{(i)}^l\}$, $l = 1, 2, \dots$, a sequência crescente de subconjuntos de $\boldsymbol{\Theta}_{(i)}$ e defina

$$\boldsymbol{\Theta}_{[j]}^l = \boldsymbol{\Theta}_{(j+1)}^l \times \cdots \times \boldsymbol{\Theta}_{(m)}^l$$

Proposição 2.10 (Priori ref. ordenada sob normalidade assintótica) .

Utilizando a notação acima e sob condições de regularidade análogas as da proposição (2.9), a priori de referência $\pi(\boldsymbol{\theta})$, relativa a ordenação $\{\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \dots, \boldsymbol{\theta}_{(m)}\}$, é dada por

$$\pi(\boldsymbol{\theta}) = \lim_{l \rightarrow \infty} \pi^l(\boldsymbol{\theta}),$$

onde $\pi^l(\boldsymbol{\theta})$ é definida pela seguinte recursão:

1. Para $j = m$, e $\boldsymbol{\theta}_{(m)} \in \boldsymbol{\Theta}_{(m)}^l$,

$$\begin{aligned} \pi_m^l(\boldsymbol{\theta}_{[m-1]} \mid \boldsymbol{\theta}^{[m-1]}) &= \pi_m^l(\boldsymbol{\theta}_{(m)} \mid \boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \dots, \boldsymbol{\theta}_{(m-1)}) \\ &= \frac{|\mathbf{h}_m(\boldsymbol{\theta})|^{1/2}}{\int_{\boldsymbol{\Theta}_{(m)}^l} |\mathbf{h}_m(\boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}_{(m)}} \end{aligned}$$

2. Para $j = m-1, m-2, \dots, 2$, e $\boldsymbol{\theta}_{(j)} \in \boldsymbol{\Theta}_{(j)}^l$,

$$\pi_j^l(\boldsymbol{\theta}_{[j-1]} \mid \boldsymbol{\theta}^{[j-1]}) = \pi_{j+1}^l(\boldsymbol{\theta}_{[j]} \mid \boldsymbol{\theta}^{[j]}) \frac{\exp\{E_j^l[\log |\mathbf{h}_j(\boldsymbol{\theta})|^{1/2}]\}}{\int_{\boldsymbol{\Theta}_{(j)}^l} \exp\{E_j^l[\log |\mathbf{h}_j(\boldsymbol{\theta})|^{1/2}]\} d\boldsymbol{\theta}_{(j)}}$$

onde

$$E_j^l[\log |\mathbf{h}_j(\boldsymbol{\theta})|^{1/2}] = \int_{\boldsymbol{\Theta}_{[j]}^l} \log |\mathbf{h}_j(\boldsymbol{\theta})|^{1/2} \pi_{j+1}^l(\boldsymbol{\theta}_{[j]} \mid \boldsymbol{\theta}^{[j]}) d\boldsymbol{\theta}_{[j]}$$

3. Para $j = 1$, $\boldsymbol{\theta}_{[0]} = \boldsymbol{\theta}$, com $\boldsymbol{\theta}^{[0]}$ vazio, e

$$\pi^l(\boldsymbol{\theta}) = \pi_1^l(\boldsymbol{\theta}_{[0]} \mid \boldsymbol{\theta}^{[0]}).$$

Prova: Veja Bernardo e Smith (1994) página 334.

O Corolário (2.1) apresentado na subseção anterior, também pode ser facilmente generalizado para o caso multidimensional. No próximo corolário apresentamos esta extensão.

Corolário 2.2 *Se a sequência crescente $\{\Theta_{(j)}^l\}$ relativa ao parâmetro de perturbação $\theta_{(j)}$ for independente dos demais parâmetros $\theta_{(i)}$, $i \neq j$, e as funções $\mathbf{h}_j(\boldsymbol{\theta})$ se fatoram na forma*

$$|\mathbf{h}_j(\boldsymbol{\theta})|^{1/2} = a_j(\theta_{(j)}) b_j(\theta_{(1)}, \dots, \theta_{(j-1)}, \theta_{(j+1)}, \dots, \theta_{(m)}), \quad j = 1, \dots, m$$

então

$$\pi(\theta_{(j)} \mid \theta_{(1)}, \dots, \theta_{(j-1)}) \propto a_j(\theta_{(j)}), \quad j = 1, \dots, m$$

e a priori de referência é dada por

$$\pi(\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(m)}) \propto \prod_{j=1}^m a_j(\theta_{(j)}).$$

Prova: Veja Bernardo e Ramón (1998).

Neste ponto cabem alguns comentários históricos a respeito da evolução do método de referência. Antes de B&B proporem o algoritmo apresentado acima (Berger e Bernardo, 1992b), os autores recomendavam simplesmente separar o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ em apenas dois grupos: o grupo dos parâmetros de interesse, e o grupo dos parâmetros de perturbação (Berger e Bernardo, 1989)¹³.

Embora a abordagem com 2-grupos seja mais facilmente computável, B&B foram forçados a considerar o refinamento desta abordagem uma vez que foram encontrados casos onde a abordagem com 2-grupos produziu prioris de referência com comportamentos não satisfatórios (Berger e Bernardo, 1992b). Aqui temos um exemplo claro da filosofia por trás da evolução do método de B&B; como já salientamos anteriormente, a medida que contra-exemplos foram aparecendo, o método foi sendo revisitado e progressivamente refinado. Atualmente, B&B recomendam a separação do vetor de parâmetros em k -grupos, isto é, a utilização do algoritmo com m -grupos, no caso particular em que cada grupo é composto por um único parâmetro (Bernardo e Ramón, 1998). Segundo Berger e Bernardo (1992b), devemos agrupar parâmetros somente quando houver algum motivo especial para isto.

¹³Observe que esta abordagem representa o caso particular $m = 2$ do algoritmo geral com m grupos.

A abordagem com k -grupos tem se mostrado muito eficiente. Desde que foi proposta, ela tem sido testada no conjunto de contra-exemplos desenvolvidos contra prioris não-subjetivas, mas até agora não parece ter produzido resultados insatisfatórios. Estes resultados certamente não garantem que a priori com k -grupos sempre apresente bom comportamento, mas o fato é que o seu sucesso tem sido impressionante (Berger e Bernardo, 1992a).

Todavia, esta abordagem (assim com a baseada em m -grupos) também acarreta uma nova complicação ao método de B&B; a priori de referência pode depender da ordenação do vetor de parâmetros. Portanto, um mesmo modelo paramétrico pode gerar diversas prioris de referência, cada uma associada a uma diferente ordenação do vetor de parâmetros. Deste modo, ao contrário do que Jeffreys acreditava, não existe uma única priori não-subjetiva associada a cada modelo paramétrico (Bernardo, 1997).

Para cada problema inferencial específico, Berger e Bernardo (1992b) recomendam ordenar o vetor de parâmetros segundo sua importância inferencial. Em outras palavras, deve-se ordenar o vetor de parâmetros de modo a refletir o interesse a respeito dos parâmetros do modelo. Por exemplo, uma ordenação como $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ significa que o parâmetro de maior interesse é o θ_1 , enquanto θ_k é o parâmetro de menor interesse.

A maior dificuldade em relação as ordenações, é que muitas vezes não é fácil decidir qual ordenação devemos adotar, sobretudo em relação aos parâmetros de perturbação. Nos casos onde somos incapazes de decidir qual ordenação dos parâmetros de perturbação utilizar, Berger e Bernardo (1992b) recomendam obter as prioris de referência associadas a cada uma das possíveis ordenações dos parâmetros de perturbação, e então escolher a priori que apresente o melhor comportamento inferencial. Todavia, Bernardo (1997) argumenta que na maioria das vezes a priori de referência é invariante com respeito a ordenação dos parâmetros de perturbação. De fato, observamos este fenômeno na derivação da priori de referência para o problema de calibração linear.

Capítulo 3

Análise de referência para o modelo de calibração linear

3.1 Introdução

O problema estatístico de calibração linear, também conhecido como regressão inversa, é motivado pela comparação de dois métodos de mensuração. Considere por exemplo, dois aparelhos de medição, A e B, onde o instrumento A é extremamente preciso, mas de alto custo, enquanto o instrumento B, ao contrário, é menos preciso e de baixo custo. O problema estatístico é recuperar as medidas “verdadeiras” (instrumento A), a partir das medidas observadas no instrumento B. Problemas deste tipo aparecem com frequência na indústria e nas ciências experimentais que dependem em grande parte de mensurações precisas, eficientes e viáveis economicamente.

Neste capítulo, abordaremos o problema de calibração linear sob a ótica da análise bayesiana de referência. Apresentaremos a construção das prioris de referência através de duas abordagens diferentes. Na primeira abordagem, a priori é obtida separando-se os parâmetros em dois grupos distintos: o grupo dos parâmetros de interesse, e o grupo dos parâmetros de perturbação. Nosso interesse em apresentar a priori obtida a partir desta abordagem é compará-la com a priori obtida por Kubokawa e Robert (1994) para um modelo relacionado ao modelo de calibração linear. De fato, B&B atualmente não recomendam a utilização desta abordagem, uma vez que foram encontradas situações onde esta abordagem produzia resultados insatisfatórios.

A segunda abordagem representa um refinamento da abordagem anterior e baseia-se na

separação do vetor de parâmetros em diversos grupos ordenados segundo sua importância inferencial, conforme descrito na seção 2.2.3.

Neste capítulo, apresentaremos a construção das prioris de referência para o modelo de calibração linear e provaremos que as respectivas distribuições a posteriori são próprias e as distribuições a posteriori marginais para x_0 apresentam momentos finitos até a ordem $p - 1$, onde p é a dimensão do vetor de respostas. Apresentaremos também a priori de Jeffreys associada ao modelo e mostraremos que também neste caso a distribuição a posteriori é própria.

Por fim obteremos as distribuições condicionais completas para a geração de amostras da distribuição a posteriori de referência (obtida a partir da segunda abordagem) via amostrador de Gibbs.

3.2 O modelo de calibração linear

Nesta seção consideramos o modelo de calibração linear com vetor de respostas de dimensão p . Neste modelo, cada covariável está associada a p variáveis resposta. Trata-se de um modelo multivariado, embora a quantidade de interesse, o x_0 , seja um escalar.

O modelo é dado por

$$\mathbf{y}_i = \boldsymbol{\alpha} + \beta x_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I}_p), \quad i = 1, \dots, n \quad (3.1)$$

$$\mathbf{y}_{0j} = \boldsymbol{\alpha} + \beta x_0 + \boldsymbol{\epsilon}_{0j}, \quad \boldsymbol{\epsilon}_{0j} \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I}_p), \quad j = 1, \dots, k \quad (3.2)$$

onde (3.1) e (3.2) representam os experimentos de calibração e de predição, respectivamente. Observe que consideramos a possibilidade de haver repetições no experimento de predição. Os vetores de erros $\boldsymbol{\epsilon}_i$ e $\boldsymbol{\epsilon}_{0j}$ são independentes, σ^2 é desconhecido, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ e $\beta = (\beta_1, \dots, \beta_p)'$ são vetores de parâmetros da regressão e $x_1, x_2, \dots, x_n$ são quantidades conhecidas, denominadas covariáveis. A função de verossimilhança do modelo é dada por

$$L_{\mathbf{y}, \mathbf{X}}(x_0, \boldsymbol{\alpha}, \beta, \sigma^2) \propto (\sigma^{-2})^{(n+k)p/2} \exp \left\{ - \left[\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\alpha} - \beta x_i)' (\mathbf{y}_i - \boldsymbol{\alpha} - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \beta x_0)' (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \beta x_0) \right] / 2\sigma^2 \right\} \quad (3.3)$$

3.3 A matriz de informação de Fisher associada ao modelo de calibração linear

Para a obtenção da priori de referência faremos uso da suposição de normalidade assintótica da distribuição a posteriori. Deste modo, precisamos calcular a matriz de informação de Fisher.

Considerando o $\log$ da função de verossimilhança (3.3), temos

$$l(\theta) = \log [L_{\mathbf{y}, \mathbf{X}}(\theta)] \propto -\frac{(n+k)p}{2} \log \sigma^2 - \frac{\sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)'(\mathbf{y}_i - \alpha - \beta x_i)}{2\sigma^2} - \frac{\sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)'(\mathbf{y}_{0j} - \alpha - \beta x_0)}{2\sigma^2}. \quad (3.4)$$

As derivadas parciais de primeira ordem são dadas por:

$$\begin{aligned} \frac{\partial l(\theta)}{\partial x_0} &= \frac{\sum_{j=1}^k \beta'(\mathbf{y}_{0j} - \alpha - \beta x_0)}{\sigma^2}, \\ \frac{\partial l(\theta)}{\partial \alpha} &= \frac{\sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)}{\sigma^2}, \\ \frac{\partial l(\theta)}{\partial \beta} &= \frac{\sum_{i=1}^n x_i(\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k x_0(\mathbf{y}_{0j} - \alpha - \beta x_0)}{\sigma^2}, \\ \frac{\partial l(\theta)}{\partial \sigma^2} &= -\frac{(n+k)p}{2\sigma^2} + \frac{\sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)'(\mathbf{y}_i - \alpha - \beta x_i)}{2\sigma^4} + \\ &\quad + \frac{\sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)'(\mathbf{y}_{0j} - \alpha - \beta x_0)}{2\sigma^4}. \end{aligned}$$

As derivadas parciais de segunda ordem são:

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial^2 x_0} &= -\frac{k\beta'\beta}{\sigma^2}, & \frac{\partial^2 l(\theta)}{\partial \alpha \partial x_0} &= -\frac{k\beta}{\sigma^2}, \\ \frac{\partial^2 l(\theta)}{\partial \beta \partial x_0} &= \frac{\sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)}{\sigma^2}, & \frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial x_0} &= -\frac{\sum_{j=1}^k \beta'(\mathbf{y}_{0j} - \alpha - \beta x_0)}{\sigma^4}, \\ \frac{\partial^2 l(\theta)}{\partial \alpha_i \partial \alpha_j} &= -\frac{n+k}{\sigma^2} \delta_{ij}, & \frac{\partial^2 l(\theta)}{\partial \beta_i \partial \alpha_j} &= -\frac{kx_0 + \sum_{i=1}^n x_i}{\sigma^2} \delta_{ij}, \end{aligned}$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2 \partial \alpha} = -\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i) + \sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)}{\sigma^4},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} = -\frac{kx_0^2 + \sum_{i=1}^n x_i^2}{\sigma^2} \delta_{ij},$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2 \partial \beta} = -\frac{\sum_{i=1}^n x_i (y_i - \alpha - \beta x_i) + \sum_{j=1}^k x_0 (y_{0j} - \alpha - \beta x_0)}{\sigma^4},$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2} &= \frac{(n+k)p}{2\sigma^4} + \frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)'(y_i - \alpha - \beta x_i)}{\sigma^6} \\ &\quad + \frac{\sum_{j=1}^k (y_{0j} - \alpha - \beta x_0)'(y_{0j} - \alpha - \beta x_0)}{\sigma^6}, \end{aligned}$$

onde δ_{ij} , é o delta de Kronecker, definido como

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}.$$

Deste modo, os elementos da matriz de informação de Fisher são dados por:

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 x_0} \right] = \frac{k\beta'\beta}{\sigma^2},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha \partial x_0} \right] = \frac{k\beta}{\sigma^2},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta \partial x_0} \right] = \frac{kx_0\beta}{\sigma^2},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2 \partial x_0} \right] = 0,$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_i \partial \alpha_j} \right] = \frac{n+k}{\sigma^2} \delta_{ij},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_i \partial \alpha_j} \right] = \frac{kx_0 + \sum_{i=1}^n x_i}{\sigma^2} \delta_{ij},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2 \partial \alpha} \right] = 0,$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} \right] = \frac{kx_0^2 + \sum_{i=1}^n x_i^2}{\sigma^2} \delta_{ij},$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2 \partial \alpha} \right] = 0,$$

$$E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma^2} \right] = \frac{(n+k)p}{2\sigma^4}.$$

Assim, a matriz de informação de Fisher associada a este modelo é expressa como:

$$\mathbf{H}(\theta) = \begin{pmatrix} k\beta'\beta/\sigma^2 & k\beta'/\sigma^2 & kx_0\beta'/\sigma^2 & 0 \\ k\beta/\sigma^2 & (n+k)/\sigma^2\mathbf{I}_p & (kx_0 + \sum_{i=1}^n x_i)/\sigma^2\mathbf{I}_p & \mathbf{0} \\ kx_0\beta/\sigma^2 & (kx_0 + \sum_{i=1}^n x_i)/\sigma^2\mathbf{I}_p & (kx_0^2 + \sum_{i=1}^n x_i^2)/\sigma^2\mathbf{I}_p & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{0} & (n+k)p/2\sigma^4 \end{pmatrix}. \quad (3.5)$$

3.4 As prioris de referência e de Jeffreys associadas ao modelo de calibração linear

Na presente seção construímos as prioris de referência segundo a abordagem com 2-grupos e a abordagem com m -grupos. Apresentamos também a priori de Jeffreys e demonstramos que, em todos os casos, as distribuições a posteriori marginais para x_0 são próprias. No caso das prioris de referência provamos também, que a distribuição a posteriori marginal para x_0 possui momentos finitos somente até a ordem $p - 1$.

3.4.1 A priori de referência segundo a abordagem com 2-grupos

Nesta abordagem, proposta por Berger e Bernardo (1989), o grupo dos parâmetros de interesse é composto apenas pelo x_0 , enquanto o grupo dos parâmetros de perturbação é dado por $\Lambda = (\alpha, \beta, \sigma^2)$. Observe que neste modelo, Λ tem dimensão $2p + 1$.

Segundo esta abordagem, obtivemos o seguinte resultado:

Teorema 3.1 *A priori de referência segundo a abordagem 2-grupos é dada por*

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto (\sigma^{-2})^{p+1} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2},$$

onde $x_0 \in \mathbb{R}$, $\alpha \in \mathbb{R}^p$, $\beta \in \mathbb{R}^p$, $\sigma^2 \in \mathbb{R}^+$ e $c_x = \sum_{i=1}^n (x_i - \bar{x})^2$.

Prova:

Primeiramente vamos particionar a matriz de informação de Fisher (equação 3.5) da seguinte maneira:

$$\mathbf{H}(\theta) = \left(\begin{array}{c|c} \mathbf{h}_{11} & \mathbf{h}_{12} \\ \hline \mathbf{h}_{21} & \mathbf{h}_{22} \end{array} \right), \quad (3.6)$$

onde

$$\mathbf{h}_{11} = k\beta'\beta/\sigma^2,$$

$$\mathbf{h}_{12} = \mathbf{h}'_{21} = (k\beta'/\sigma^2, kx_0\beta/\sigma^2, 0),$$

$$\begin{aligned} \mathbf{h}_{22} &= \left(\begin{array}{cc|cc} (n+k)/\sigma^2 \mathbf{I}_p & & (kx_0 + \sum_{i=1}^n x_i)/\sigma^2 \mathbf{I}_p & & \mathbf{0} \\ & & \hline (kx_0 + \sum_{i=1}^n x_i)/\sigma^2 \mathbf{I}_p & & (kx_0^2 + \sum_{i=1}^n x_i^2)/\sigma^2 \mathbf{I}_p & & \mathbf{0} \\ & & \mathbf{0} & & \mathbf{0} & & (n+k)p/2\sigma^4 \end{array} \right) \\ &= \left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right). \end{aligned} \quad (3.7)$$

Seguindo a notação da Corolário (2.1) do capítulo 2 e fixando x_0 por um momento, temos que a priori de referência condicional é dada por

$$\pi(\Lambda \mid x_0) \propto |\mathbf{h}_{22}|^{1/2}, \quad (3.8)$$

onde $|\mathbf{h}_{22}|$ pode ser facilmente calculado, uma vez que $\mathbf{A}_{22}$ é uma matriz diagonal, como

$$|\mathbf{h}_{22}| = |\mathbf{A}_{22}| |\mathbf{A}_{11.2}| \quad \text{onde} \quad \mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}. \quad (3.9)$$

Efetuando-se as contas temos que

$$|\mathbf{A}_{22}| = \frac{(n+k)p (kx_0^2 + \sum_{i=1}^n x_i^2)^p}{2(\sigma^2)^{p+2}}, \quad (3.10)$$

e que

$$|\mathbf{A}_{11.2}| = \frac{a^p}{(\sigma^2)^p (kx_0^2 + \sum_{i=1}^n x_i^2)^p}, \quad (3.11)$$

onde $a = (n + k)c_x + nk(x_0 - \bar{x})^2$. Deste modo, temos que

$$|\mathbf{h}_{22}| = \frac{a^p(n + k)p}{2(\sigma^2)^{2p+2}}. \quad (3.12)$$

Assim,

$$|\mathbf{h}_{22}|^{1/2} = (\sigma^{-2})^{p+1} \left\{ \frac{a^p(n + k)p}{2} \right\}^{1/2} = a_2(\alpha, \beta, \sigma^2) b_2(x_0). \quad (3.13)$$

Vamos agora, calcular $\mathbf{h}_{x_0} = \mathbf{h}_{11} - \mathbf{h}_{12}\mathbf{h}_{22}^{-1}\mathbf{h}_{21}$. Para tanto, precisamos primeiramente inverter a matriz $\mathbf{h}_{22}$. Observe que $\mathbf{h}_{22}^{-1}$ pode ser facilmente obtida como

$$\mathbf{h}_{22}^{-1} = \left(\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array} \right), \quad (3.14)$$

onde

$$\begin{aligned} \mathbf{B}_{11} &= \mathbf{A}_{11.2}^{-1}, & \mathbf{B}_{12} &= -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}, \\ \mathbf{B}_{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1}, & \mathbf{B}_{22} &= \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}. \end{aligned}$$

uma vez que só precisamos inverter as matrizes diagonais $\mathbf{A}_{22}$ e $\mathbf{A}_{11.2}$.

Após algumas manipulações algébricas temos que

$$\mathbf{h}_{22}^{-1} = \left(\begin{array}{c|c|c} \begin{array}{c} \sigma^2(kx_0^2 + \sum_{i=1}^n x_i^2)/a\mathbf{I}_p \\ p \times p \end{array} & \begin{array}{c} -\sigma^2(kx_0 + \sum_{i=1}^n x_i)/a\mathbf{I}_p \\ p \times p \end{array} & \begin{array}{c} \mathbf{0} \\ p \times 1 \end{array} \\ \hline \begin{array}{c} -\sigma^2(kx_0 + \sum_{i=1}^n x_i)/a\mathbf{I}_p \\ p \times p \end{array} & \begin{array}{c} \sigma^2(n + k)/a\mathbf{I}_p \\ p \times p \end{array} & \begin{array}{c} \mathbf{0} \\ p \times 1 \end{array} \\ \hline \begin{array}{c} \mathbf{0} \\ 1 \times p \end{array} & \begin{array}{c} \mathbf{0} \\ 1 \times p \end{array} & \begin{array}{c} 2\sigma^4/(n + k)p \\ 1 \times 1 \end{array} \end{array} \right). \quad (3.15)$$

Efetando os cálculos temos que

$$\mathbf{h}_{x_0} = \frac{nc_x k \beta' \beta}{a\sigma^2}, \quad (3.16)$$

de modo que

$$|\mathbf{h}_{x_0}|^{1/2} = \left\{ \frac{nc_x k \beta' \beta}{\sigma^2} \right\}^{1/2} a^{-1/2} = a_1(x_0) b_1(\alpha, \beta, \sigma^2). \quad (3.17)$$

Aplicando o Corolário (2.1) temos que a priori de referência é dada por:¹

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto a_1(x_0) a_2(\alpha, \beta, \sigma^2) = a^{-1/2} (\sigma^{-2})^{p+1}. \quad (3.18)$$

E lembrando que $a = (n+k)c_x + nk(x_0 - \bar{x})^2$ temos que

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto (\sigma^{-2})^{p+1} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad \triangle$$

Observe que esta priori de referência é claramente imprópria. Integrando esta priori em x_0 , α , β e σ^2 , temos

$$\underbrace{\left(\int_{\mathbb{R}^p} d\alpha \right)}_{=\infty} \underbrace{\left(\int_{\mathbb{R}^p} d\beta \right)}_{=\infty} \underbrace{\left(\int_0^\infty (\sigma^{-2})^{p+1} d\sigma^2 \right)}_{=\infty} \left(\int_{-\infty}^{+\infty} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2} dx_0 \right).$$

Até mesmo o termo

$$\left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2},$$

que apresenta a forma do núcleo de uma distribuição da família das t-student, mas com caudas mais pesadas que uma Cauchy centrada em $\bar{x}$ e com parâmetro de escala $c_x(n+k)/nk$, diverge quando integrado em relação a x_0 . Deste modo, torna-se essencial provar que a distribuição a posteriori, obtida a partir da priori acima, é própria.

O próximo resultado nos garante que a distribuição a posteriori obtida através desta priori de referência é própria embora seus momentos de ordem superior a $p-1$ sejam infinitos.

Teorema 3.2 *A distribuição a posteriori conjunta obtida através da priori de referência derivada pela abordagem com 2-grupos é própria.*

Prova:

¹Uma escolha natural para a sequência $\{\Lambda_i\}$ é dada por $\Lambda_i = (\alpha_i, \beta_i, \sigma_i^2)$ onde $\alpha_i = [-i, i]^p$, $\beta_i = [-i, i]^p$ e $\sigma_i^2 = [e^{-i}, e^i]$. Esta escolha é natural uma vez que cada componente dos vetores α e β pertence a $\mathbb{R}$, de modo o um conjunto que espelha conhecimento vago a respeito destes parâmetros de perturbação é dado por $[-i, i]$. E em relação ao σ^2 , já vimos anteriormente, que o conjunto que espelha conhecimento vago para este parâmetro é o $[e^{-i}, e^i]$. Observando que a sequência de conjuntos Λ_i não depende de x_0 , e que $|\mathbf{h}_{x_0}|^{1/2} = a_1(x_0)b_1(\Lambda)$ e $|\mathbf{h}_{22}|^{1/2} = a_2(\Lambda)b_2(x_0)$, temos que a priori de referência é dada por $\pi(x_0, \Lambda) \propto a_1(x_0) a_2(\Lambda)$.

Para demonstrar este resultado, precisamos em primeiro lugar encontrar a distribuição a posteriori marginal de x_0 . Para tanto, eliminaremos por integração, todos os demais parâmetros da distribuição a posteriori.

A distribuição a posteriori de referência $\pi(x_0, \alpha, \beta, \sigma^2 \mid \mathbf{y}, \mathbf{X})$ é dada por

$$\begin{aligned} &\propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} (\sigma^{-2})^{\frac{(n+k+2)p+2}{2}} \exp \left\{ - \left[\sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' \times \right. \right. \\ &\quad \times (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' (\mathbf{y}_{0j} - \alpha - \beta x_0) \left. \right] / 2\sigma^2 \Big\}. \end{aligned} \quad (3.19)$$

Primeiramente, eliminaremos o σ^2 . Para tanto, vamos reescrever a equação (3.19) como

$$\propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} \times \underbrace{(\sigma^{-2})^{\frac{(n+k+2)p+4}{2}-1} \exp \left\{ -(l_1/2)\sigma^{-2} \right\}}_{(I)} \quad (3.20)$$

onde

$$l_1 = \sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' (\mathbf{y}_{0j} - \alpha - \beta x_0).$$

Observe que (I) representa o núcleo de uma distribuição

$$\text{Gama} \left(\frac{(n+k+2)p+4}{2}, \frac{l_1}{2} \right). \quad (3.21)$$

Deste modo, integrando em σ^{-2} temos que

$$\begin{aligned} \pi(x_0, \alpha, \beta \mid \mathbf{y}, \mathbf{X}) &\propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} \left\{ \sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' \times \right. \\ &\quad \times (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' (\mathbf{y}_{0j} - \alpha - \beta x_0) \Big\}^{-\frac{(n+k+2)p+4}{2}}. \end{aligned} \quad (3.22)$$

Agora, vamos eliminar por integração o α . Após algumas manipulações algébricas podemos reescrever (3.22) como

$$\propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} l_3^{-\frac{(n+k+2)p+4}{2}} \times \underbrace{\left(1 + \frac{1}{v_1} \frac{(\alpha - l_2)'(\alpha - l_2)}{l_3/(n+k)v_1}\right)^{-\frac{v_1+p}{2}}}_{(II)}, \quad (3.23)$$

onde

$$v_1 = 4 + (n + k + 1)p,$$

$$l_2 = \frac{\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)}{n + k},$$

$$l_3 = \sum_{i=1}^n (\mathbf{y}_i - \beta x_i)' (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)' (\mathbf{y}_{0j} - \beta x_0) - (n + k) l_2' l_2.$$

Observe que l_3 não é função de α e que (II) representa o núcleo de uma distribuição $t_{v_1}(l_2, l_3/(n + k)v_1)$. Deste modo, integrando com respeito a α temos que $\pi(x_0, \beta | \mathbf{y}, \mathbf{X})$

$$\begin{aligned} & \propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n + k)c_x}\right)^{-1/2} \left\{ (n + k) \left[\sum_{i=1}^n (\mathbf{y}_i - \beta x_i)' (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)' (\mathbf{y}_{0j} - \beta x_0) \right] \right. \\ & \left. - \left(\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0) \right)' \left(\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0) \right) \right\}^{-\frac{(n+k+1)p+4}{2}}. \quad (3.24) \end{aligned}$$

Agora, vamos eliminar por integração o β . Após uma série de manipulações algébricas, (3.24) pode ser reescrita como

$$\begin{aligned} \pi(x_0, \beta | \mathbf{y}, \mathbf{X}) & \propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n + k)c_x}\right)^{-1/2} \left(\frac{a l_6 - (\mathbf{l}_4 + \mathbf{l}_5 x_0)' (\mathbf{l}_4 + \mathbf{l}_5 x_0)}{a} \right)^{-\frac{(n+k+1)p+4}{2}} \times \\ & \times \left(1 + \frac{1}{v_2} \frac{(\beta - (\mathbf{l}_4 + \mathbf{l}_5 x_0)/a)' (\beta - (\mathbf{l}_4 + \mathbf{l}_5 x_0)/a)}{[a l_6 - (\mathbf{l}_4 + \mathbf{l}_5 x_0)' (\mathbf{l}_4 + \mathbf{l}_5 x_0)]/(a^2 v_2)}\right)^{-\frac{v_2+p}{2}}, \quad (3.25) \end{aligned}$$

onde

$$v_2 = (n + k)p + 4,$$

$$l_5 = n \sum_{j=1}^k \mathbf{y}_{0j} - k \sum_{i=1}^n \mathbf{y}_i,$$

$$l_6 = (n + k) \left(\sum_{i=1}^n \mathbf{y}_i' \mathbf{y}_i + \sum_{j=1}^k \mathbf{y}_{0j}' \mathbf{y}_{0j} \right) - \left(\sum_{i=1}^n \mathbf{y}_i + \sum_{j=1}^k \mathbf{y}_{0j} \right)' \left(\sum_{i=1}^n \mathbf{y}_i + \sum_{j=1}^k \mathbf{y}_{0j} \right),$$

$$l_4 = (n+k) \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \left(\sum_{i=1}^n y_i + \sum_{j=1}^k y_{0j} \right).$$

Observe que a , l_4 , l_5 e l_6 não são funções de β , e que (3.25) representa o núcleo de uma distribuição t -student multivariada

$$t_{v_2} \left(\frac{l_4 + l_5 x_0}{a}, \frac{a l_6 - (l_4 + l_5 x_0)'(l_4 + l_5 x_0)}{a^2 v_2} \mathbf{I}_p \right).$$

Deste modo, integrando em relação a β temos que $\pi(x_0 | \mathbf{y}, \mathbf{X})$

$$\propto a^{\frac{(n+k-1)p+4}{2}} \{a l_6 - (l_4 + l_5 x_0)'(l_4 + l_5 x_0)\}^{-\frac{(n+k)p+4}{2}} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad (3.26)$$

Lembrando que $a = (n+k)c_x + nk(x_0 - \bar{x})^2$ temos

$$\pi(x_0 | \mathbf{y}, \mathbf{X}) \propto a^{\frac{(n+k-1)p+3}{2}} \{a l_6 - (l_4 + l_5 x_0)'(l_4 + l_5 x_0)\}^{-\frac{(n+k)p+4}{2}}. \quad (3.27)$$

Após algumas manipulações algébricas podemos reescrevê-la como

$$\pi(x_0 | \mathbf{y}, \mathbf{X}) \propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} \left(1 + \frac{1}{v} \frac{(x_0 - l_7)^2}{l_8/(nkl_6 - l'_5 l_5)v} \right)^{-\frac{v+1}{2}}, \quad (3.28)$$

onde

$$v = (n+k)p + 3,$$

$$l_7 = \frac{nkl_6 \bar{x} + l'_4 l_5}{nkl_6 - l'_5 l_5},$$

$$l_8 = nk\bar{x}^2 l_6 + (n+k)c_x l_6 - l'_4 l_4 - (nkl_6 - l'_5 l_5) l_7^2.$$

Vamos agora, mostrar que a distribuição a posteriori marginal de x_0 é própria. Primeiramente, observe que (3.28) é proporcional a uma função de x_0 multiplicada pelo núcleo de uma distribuição t -student. Deste modo, temos que a integral de $\pi(x_0 | \mathbf{y}, \mathbf{X})$ em relação a x_0 , é proporcional a esperança desta função de x_0 em relação a distribuição t -student. Explicitamente, temos

$$\int_{-\infty}^{+\infty} \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 \propto E_{x_0} \left[\left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} \right], \quad (3.29)$$

onde

$$x_0 \sim t_v \left(l_7, \frac{l_8}{(nkl_6 - l'_5 l_5)v} \right). \quad (3.30)$$

Agora, utilizando o resultado de que uma distribuição *t-student* só possui momentos de ordem estritamente menor do que seus graus de liberdade (Johnson, Kotz e Balakrishnan, 1995) podemos facilmente mostrar que (3.29) é finita. Para tanto, basta notar que

$$\left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} = a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_{(n+k-1)p+3} x_0^{(n+k-1)p+3},$$

pode ser reescrito como um polinômio de grau $(n+k-1)p+3$. E como

$$E \left[\sum_{i=0}^{(n+k-1)p+3} a_i x_0^i \right] = \sum_{i=0}^{(n+k-1)p+3} a_i E \left[x_0^i \right],$$

e o número de graus de liberdade $v = (n+k)p+3$ de (3.30) é estritamente maior que $(n+k-1)p+3$, temos que (3.29) pode ser reescrita como uma combinação linear de momentos finitos e portanto,

$$\int_{-\infty}^{+\infty} \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 \propto \sum_{i=0}^{(n+k-1)p+3} a_i E \left[x_0^i \right] < \infty. \quad \triangle$$

Vamos agora mostrar que apesar da distribuição a posteriori marginal de x_0 ser própria, ela possui momentos finitos somente até a ordem $p-1$.

Corolário 3.1 *A distribuição a posteriori marginal de x_0 possui momentos finitos somente até a ordem $p-1$.*

Prova:

Para tanto, basta mostrar que

$$E[x_0^p | \mathbf{y}, \mathbf{X}] = \int_{-\infty}^{+\infty} x_0^p \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 = \infty, \quad (3.31)$$

e que

$$E[x_0^{p-1} | \mathbf{y}, \mathbf{X}] = \int_{-\infty}^{+\infty} x_0^{p-1} \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 < \infty. \quad (3.32)$$

Observando que

$$E[x_0^p | \mathbf{y}, \mathbf{X}] \propto E_{x_0} \left[x_0^p \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} \right], \quad (3.33)$$

onde

$$x_0 \sim t_v \left(l_7, \frac{l_8}{v(nkl_6 - l'_5 l_5)} \right), \quad (3.34)$$

e que

$$x_0^p \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} = a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_{(n+k)p+3} x_0^{(n+k)p+3},$$

representa um polinômio de ordem $(n+k)p+3$, temos que $E[x_0^p | \mathbf{y}, \mathbf{X}]$ pode ser reescrita como uma combinação linear de momentos de ordem menor e igual a $(n+k)p+3$ da distribuição (3.34). Todavia, como (3.34) possui $v = (n+k)p+3$, temos que o momento de ordem $(n+k)p+3$ diverge, e portanto $E[x_0^p | \mathbf{y}, \mathbf{X}] = \infty$. Lembrando que, quando o momento de ordem i de uma distribuição não existe, então todos os demais momentos de ordem superior a i também não existem, temos que $\pi(x_0 | \mathbf{y}, \mathbf{X})$ não possui nenhum momento de ordem igual ou superior a p .

Agora, observando que

$$E[x_0^{p-1} | \mathbf{y}, \mathbf{X}] \propto E_{x_0} \left[x_0^{p-1} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} \right], \quad (3.35)$$

e que

$$x_0^{p-1} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k-1)p+3}{2}} = a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_{(n+k)p+2} x_0^{(n+k)p+2},$$

representa um polinômio de ordem $(n+k)p+2$, temos que $E[x_0^{p-1} | \mathbf{y}, \mathbf{X}]$ pode ser reescrita como uma combinação linear de momentos (da distribuição 3.34) de ordem menor e igual a $(n+k)p+2$. Todavia, como (3.34) possui $v = (n+k)p+3$, temos que todos momentos de ordem menor ou igual a $(n+k)p+2$ são finitos, e portanto $E[x_0^{p-1} | \mathbf{y}, \mathbf{X}] < \infty$. Deste modo, fica provado que $\pi(x_0 | \mathbf{y}, \mathbf{X})$ é própria, mas só possui momentos finitos de ordem menor ou igual a $p-1$. $\triangle$

Observe que quando $p = 1$, a distribuição marginal de x_0 é própria, mas não possui esperança nem qualquer outro momento de ordem superior.

3.4.2 A priori de referência segundo a abordagem com m -grupos

Nesta subseção vamos apresentar a construção da priori de referência baseada na abordagem com m -grupos. Por uma questão de economia de espaço só apresentaremos a construção associada a ordenação

$$\{x_0\}; \{\alpha\}; \{\beta\}; \{\sigma^2\}.$$

onde $\{\alpha\} = \{\alpha_1, \dots, \alpha_p\}$ e $\{\beta\} = \{\beta_1, \dots, \beta_p\}$. Todavia, em nosso trabalho de pesquisa estudamos também as demais cinco possíveis ordenações de $\{\alpha\}$, $\{\beta\}$ e $\{\sigma^2\}$ no caso especial em que $p = 1$. Como ocorre frequentemente, em relação a ordenação dos parâmetros de perturbação, obtivemos sempre a mesma priori em todas as ordenações.

Observe que neste modelo é razoável agrupar os parâmetros $\{\alpha_1, \dots, \alpha_p\}$ em $\{\alpha\}$ e $\{\beta_1, \dots, \beta_p\}$ em $\{\beta\}$ e trabalhar em um contexto com 4-grupos, ao invés de trabalhar com cada um dos $2p + 2$ parâmetros separadamente. Observe ainda que não foi necessário considerar diferentes ordenações dentro dos grupos $\{\alpha\}$ e $\{\beta\}$, uma vez que a priori de referência é invariante com respeito a ordenações dentro de grupos (Berger e Bernardo, 1992b).

Utilizando a abordagem com m -grupos, obtivemos o seguinte resultado:

Teorema 3.3 *A priori de referência associada as ordenações*

$$\begin{aligned} &\{x_0\}; \{\alpha\}; \{\beta\}; \{\sigma^2\} \quad \{x_0\}; \{\alpha\}; \{\sigma^2\}; \{\beta\} \quad \{x_0\}; \{\beta\}; \{\alpha\}; \{\sigma^2\} \\ &\{x_0\}; \{\beta\}; \{\sigma^2\}; \{\alpha\} \quad \{x_0\}; \{\sigma^2\}; \{\alpha\}; \{\beta\} \quad \{x_0\}; \{\sigma^2\}; \{\beta\}; \{\alpha\} \end{aligned}$$

não depende de p e é dada por

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto \sigma^{-2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2},$$

onde $x_0 \in \mathbb{R}$, $\alpha \in \mathbb{R}^p$, $\beta \in \mathbb{R}^p$, $\sigma^2 \in \mathbb{R}^+$ e $c_x = \sum_{i=1}^n (x_i - \bar{x})^2$.

Prova:

Em primeiro lugar temos que $\mathbf{H}_4(\theta)$ é igual a matriz de informação de Fisher (equação 3.5).

Deste modo,

$$h_4(\theta) = \frac{(n+k)p}{2\sigma^4}, \quad (3.36)$$

de modo que $|h_4(\theta)|^{1/2}$ pode ser fatorado como

$$|h_4(\theta)|^{1/2} = \sigma^{-2} \{(n+k)p/2\}^{1/2} = a_4(\sigma^2) b_4(x_0, \alpha, \beta), \quad (3.37)$$

onde

$$a_4(\sigma^2) = \sigma^{-2} \quad (3.38)$$

Precisamos agora calcular $S_4(\theta) = H_4^{-1}(\theta)$. Lembrando que $H_4(\theta)$ pode ser particionada como em (3.6) temos que

$$S_4(\theta) = \left(\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array} \right), \quad (3.39)$$

onde

$$\begin{aligned} C_{11} &= \mathbf{h}_{11.2}^{-1}, & C_{12} &= -\mathbf{h}_{11.2}^{-1} \mathbf{h}_{12} \mathbf{h}_{22}^{-1}, \\ C_{21} &= -\mathbf{h}_{22}^{-1} \mathbf{h}_{21} \mathbf{h}_{11.2}^{-1}, & C_{22} &= \mathbf{h}_{22}^{-1} \mathbf{h}_{21} \mathbf{h}_{11.2}^{-1} \mathbf{h}_{12} \mathbf{h}_{22}^{-1} + \mathbf{h}_{22}^{-1}, \\ \mathbf{h}_{11.2} &= \mathbf{h}_{x_0} = \mathbf{h}_{11} - \mathbf{h}_{12} \mathbf{h}_{22}^{-1} \mathbf{h}_{21}. \end{aligned}$$

Lembrando que $\mathbf{h}_{x_0}$ que já foi calculado anteriormente (equação 3.16) e vale

$$\mathbf{h}_{x_0} = \mathbf{h}_{11.2} = \frac{nc_x k \beta' \beta}{a \sigma^2}, \quad (3.40)$$

temos que após alguma álgebra $S_4(\theta)$ é dada por

$$S_4(\theta) = \left(\begin{array}{c|c|c|c} a\sigma^2/nc_x k \beta' \beta & -\sigma^2 \beta' b / nc_x \beta' \beta & -\sigma^2 \beta' c / nc_x \beta' \beta & 0 \\ \hline 1 \times 1 & 1 \times p & 1 \times p & 1 \times 1 \\ \hline -\sigma^2 \beta' b / nc_x \beta' \beta & f(x_0, \beta, \sigma^2) & g(x_0, \beta, \sigma^2) & \mathbf{0} \\ \hline p \times 1 & p \times p & p \times p & p \times 1 \\ \hline -\sigma^2 \beta' c / nc_x \beta' \beta & g(x_0, \beta, \sigma^2) & h(x_0, \beta, \sigma^2) & \mathbf{0} \\ \hline p \times 1 & p \times p & p \times p & p \times 1 \\ \hline 0 & \mathbf{0} & \mathbf{0} & 2\sigma^4/(n+k)p \\ \hline 1 \times 1 & 1 \times p & 1 \times p & 1 \times 1 \end{array} \right), \quad (3.41)$$

onde

$$\begin{aligned} b &= \sum_{i=1}^n x_i^2 - x_0 \sum_{i=1}^n x_i, & c &= nx_0 - \sum_{i=1}^n x_i, \\ d &= \sum_{i=1}^n x_i^2, & e &= \sum_{i=1}^n x_i, \end{aligned}$$

$$f(x_0, \beta, \sigma^2) = \frac{\sigma^2 k \beta \beta' b^2}{a n c_x \beta' \beta} + \frac{\sigma^2 (k x_0^2 + d)}{a} \mathbf{I}_p,$$

$$g(x_0, \beta, \sigma^2) = \frac{\sigma^2 k \beta \beta' b c}{a n c_x \beta' \beta} - \frac{\sigma^2 (k x_0 + e)}{a} \mathbf{I}_p,$$

$$h(x_0, \beta, \sigma^2) = \frac{\sigma^2 k \beta \beta' c^2}{a n c_x \beta' \beta} + \frac{\sigma^2 (n + k)}{a} \mathbf{I}_p.$$

Precisamos agora encontrar $h_3(\theta)$, e mostrar que ele se fatora como

$$|h_3(\theta)|^{1/2} = a_3(\beta) b_3(x_0, \alpha, \sigma^2). \quad (3.42)$$

Para tanto, precisamos primeiramente inverter a matriz

$$\mathbf{S}_3(\theta) = \begin{pmatrix} \begin{array}{c|c|c} a\sigma^2 / n c_x k \beta' \beta & -\sigma^2 \beta' b / n c_x \beta' \beta & -\sigma^2 \beta' c / n c_x \beta' \beta \\ \hline 1 \times 1 & 1 \times p & 1 \times p \end{array} & \begin{array}{c|c} -\sigma^2 \beta' b / n c_x \beta' \beta & f(x_0, \beta, \sigma^2) \\ \hline p \times 1 & p \times p \end{array} & \begin{array}{c|c} -\sigma^2 \beta' c / n c_x \beta' \beta & g(x_0, \beta, \sigma^2) \\ \hline p \times 1 & p \times p \end{array} \\ \hline \begin{array}{c|c} -\sigma^2 \beta' b / n c_x \beta' \beta & g(x_0, \beta, \sigma^2) \\ \hline p \times 1 & p \times p \end{array} & \begin{array}{c|c} -\sigma^2 \beta' c / n c_x \beta' \beta & h(x_0, \beta, \sigma^2) \\ \hline p \times 1 & p \times p \end{array} \end{pmatrix}, \quad (3.43)$$

onde $f(x_0, \beta, \sigma^2)$, $g(x_0, \beta, \sigma^2)$ e $h(x_0, \beta, \sigma^2)$ não são matrizes diagonais.

Observe que não é fácil inverter $\mathbf{S}_3(\theta)$ uma vez que não é possível decompô-la de tal modo que só precisamos inverter uma matriz diagonal. Por este motivo, optamos por utilizar o programa Mathematica 4.0 para calcular $\mathbf{S}_3^{-1}(\theta)$ e o determinante de $\mathbf{h}_3(\theta)$. Desta maneira temos que

$$|\mathbf{h}_3(\theta)|^{1/2} = 1 \times \left\{ \frac{\{k x_0^2 + \sum_{i=1}^n x_i^2\}^p}{\sigma^{2p}} \right\}^{1/2} = a_3(\beta) b_3(x_0, \alpha, \sigma^2), \quad (3.44)$$

$$\text{onde } a_3(\beta) = 1.$$

Precisamos agora encontrar $\mathbf{h}_2(\theta)$ a partir de $\mathbf{S}_2(\theta)$ e mostrar que

$$|\mathbf{h}_2(\theta)|^{1/2} = a_2(\alpha) b_2(x_0, \beta, \sigma^2). \quad (3.45)$$

No entanto, observando que

$$S_2(\theta) = \left(\begin{array}{c|c} a\sigma^2/nc_x k\beta'\beta & -\sigma^2\beta'b/nc_x\beta'\beta \\ \hline 1 \times 1 & 1 \times p \\ \hline -\sigma^2\beta'b/nc_x\beta'\beta & f(x_0, \beta, \sigma^2) \\ \hline p \times 1 & p \times p \end{array} \right) \quad (3.46)$$

não é função de α , percebemos que

$$|\mathbf{h}_2(\theta)|^{1/2} = 1 \times |\mathbf{h}_2(\theta)|^{1/2} = a_2(\alpha)b_2(x_0, \beta, \sigma^2), \quad (3.47)$$

$$\text{onde } a_2(\alpha) = 1.$$

Portanto não é preciso inverter $S_2(\theta)$ e calcular $\mathbf{h}_2(\theta)$ explicitamente.

Precisamos agora encontrar $\mathbf{h}_1(\theta)$ a partir de $S_1(\theta)$ e mostrar que

$$|\mathbf{h}_1(\theta)|^{1/2} = a_1(x_0)b_1(\alpha, \beta, \sigma^2). \quad (3.48)$$

Observando que,

$$S_1(\theta) = \frac{a\sigma^2}{nc_x k\beta'\beta} \Rightarrow \mathbf{h}_1(\theta) = \mathbf{H}_1(\theta) = S_1^{-1}(\theta) = \frac{nc_x k\beta'\beta}{a\sigma^2}, \quad (3.49)$$

temos que

$$|\mathbf{h}_1(\theta)|^{1/2} = a^{-1/2} \left\{ \frac{nc_x k\beta'\beta}{\sigma^2} \right\}^{1/2} = a_1(x_0)b_1(\alpha, \beta, \sigma^2), \quad (3.50)$$

$$\text{onde } a_1(x_0) = a^{-1/2}.$$

E lembrando que $a = (n+k)c_x + nk(x_0 - \bar{x})^2$ temos que

$$a_1(x_0) \propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad (3.51)$$

Aplicando o Corolário (2.2)², temos que a priori de referência é dada por

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto a_1(x_0) a_2(\alpha) a_3(\beta) a_4(\sigma^2),$$

²Assim como no Teorema (3.1) neste caso também é natural escolhermos a sequência crescente $\Lambda_i = (\alpha_i, \beta_i, \sigma_i^2)$ tal que $\alpha_i = [-i, i]^p$, $\beta_i = [-i, i]^p$ e $\sigma_i^2 = [e^{-i}, e^i]$. Observe que as sequências de conjuntos α_i , β_i e σ_i^2 não dependem do parâmetro de interesse (x_0) nem de α , β e σ^2 . Desta maneira podemos utilizar o Corolário (2.2).

isto é,

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto \sigma^{-2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad \triangle$$

Observe novamente, que esta priori é imprópria. Deste modo, precisamos mostrar que a distribuição a posteriori de referência é própria.

Teorema 3.4 *A distribuição a posteriori conjunta obtida através da priori de referência apresentada no Teorema (3.3) é própria.*

Corolário 3.2 *A distribuição a posteriori marginal de x_0 possui momentos finitos somente até a ordem $p - 1$.*

Provas:

A demonstração destes resultados é análoga a feita para o Teorema (3.2) e Corolário (3.1). De fato, basta substituir o valor n por $n - 2$ em todas as equações derivadas nas provas do Teorema (3.2) e do Corolário (3.1), que o resultado segue. $\triangle$

3.4.3 A priori de Jeffreys

Nesta subseção vamos apresentar a construção da priori de Jeffreys. Observe que esta priori pode ser vista como o caso particular da abordagem com m -grupos quando $m = 1$, isto é, quando todos os parâmetros são agrupados em um único grupo.

Teorema 3.5 *A priori de Jeffreys associada ao modelo de calibração linear é dada por*

$$\pi_J(x_0, \alpha, \beta, \sigma^2) \propto \frac{(\beta' \beta)^{1/2}}{\sigma^{2p+3}} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{(p-1)/2},$$

onde $x_0 \in \mathbb{R}$, $\alpha \in \mathbb{R}^p$, $\beta \in \mathbb{R}^p$, $\sigma^2 \in \mathbb{R}^+$ e $c_x = \sum_{i=1}^n (x_i - \bar{x})^2$.

Prova:

Para demonstrar este resultado, basta calcularmos o determinante da matriz de informação de Fisher (equação 3.5). Particionando a matriz de informação de Fisher como em (3.6) temos que

$$|\mathbf{H}(\theta)| = |\mathbf{h}_{22}| |\mathbf{h}_{11.2}|, \quad (3.52)$$

onde $|\mathbf{h}_{11.2}| = \mathbf{h}_{11.2} = \mathbf{h}_{x_0} = \mathbf{h}_{11} - \mathbf{h}_{12}\mathbf{h}_{22}^{-1}\mathbf{h}_{21}$ já foi calculado (equação 3.16), e observando que $|\mathbf{h}_{22}|$ também já foi calculado (equação 3.12) temos que

$$|\mathbf{H}(\boldsymbol{\theta})| = |\mathbf{h}_{22}||\mathbf{h}_{11.2}| = \frac{a^p(n+k)p}{2(\sigma^2)^{2p+2}} \frac{nc_x\boldsymbol{\beta}'\boldsymbol{\beta}}{a\sigma^2}. \quad (3.53)$$

Deste modo

$$|\mathbf{H}(\boldsymbol{\theta})|^{1/2} \propto \frac{(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}a^{(p-1)/2}}{2^{1/2}(\sigma^2)^{(2p+3)/2}}, \quad (3.54)$$

e lembrando que $a = (n+k)c_x + nk(x_0 - \bar{x})^2$, segue que

$$\pi_J(x_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2) \propto \frac{(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}}{\sigma^{2p+3}} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{(p-1)/2}. \quad \triangle$$

Observe que esta priori também é imprópria. Todavia, o próximo resultado nos garante que a distribuição a posteriori marginal para x_0 obtida através desta priori é própria.

Teorema 3.6 *A distribuição a posteriori marginal para x_0 obtida através da priori de Jeffreys é própria.*

Prova:

Em primeiro lugar, precisamos obter a distribuição a posteriori marginal para o x_0 . Para tanto, eliminaremos por integração os demais parâmetros do modelo. Aqui não apresentaremos em detalhes a derivação das distribuições marginais, uma vez que esta derivação é análoga a feita na demonstração do Teorema (3.2).

A distribuição a posteriori $\pi(x_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X})$ é proporcional a

$$\begin{aligned} & (\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{\frac{p-1}{2}} (\sigma^{-2})^{\frac{(n+k+2)p+3}{2}} \exp \left\{ - \left[\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}x_i)' \times \right. \right. \\ & \left. \left. \times (\mathbf{y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \boldsymbol{\beta}x_0)' (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \boldsymbol{\beta}x_0) \right] / 2\sigma^2 \right\}. \end{aligned} \quad (3.55)$$

Utilizando resultados de integração da família das gamas temos que

$$\begin{aligned} \pi(x_0, \alpha, \beta \mid \mathbf{y}, \mathbf{X}) &\propto (\beta' \beta)^{1/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{p-1}{2}} \left\{ \sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' \times \right. \\ &\times (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' (\mathbf{y}_{0j} - \alpha - \beta x_0) \left. \right\}^{-\frac{(n+k+2)p+5}{2}}. \end{aligned} \quad (3.56)$$

Utilizando resultados de integração da família das distribuições *t*-student multivariadas temos que

$$\begin{aligned} \pi(x_0, \beta \mid \mathbf{y}, \mathbf{X}) &\propto (\beta' \beta)^{1/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{p-1}{2}} \left\{ (n+k) \left[\sum_{i=1}^n (\mathbf{y}_i - \beta x_i)' \times \right. \right. \\ &\times (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)' (\mathbf{y}_{0j} - \beta x_0) \left. \right] - \left(\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0) \right)' \\ &\times \left(\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0) \right) \left. \right\}^{-\frac{(n+k+1)p+5}{2}}. \end{aligned} \quad (3.57)$$

Utilizando, novamente, resultados de integração da família das distribuições *t*-student multivariadas temos que $\pi(x_0 \mid \mathbf{y}, \mathbf{X})$ é proporcional a

$$E \left[(\beta' \beta)^{1/2} \right] \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}} \left(1 + \frac{1}{v} \frac{(x_0 - l_7)^2}{l_8/(nkl_6 - l_5' l_5)v} \right)^{-\frac{v+1}{2}}, \quad (3.58)$$

onde $v = (n+k)p + 4$ e

$$E \left[(\beta' \beta)^{1/2} \right] = \int_{\beta} c' (\beta' \beta)^{1/2} \left(1 + \frac{1}{v_2} \frac{(\beta - \mathbf{m})' (\beta - \mathbf{m})}{s/v_2} \right)^{-\frac{v_2+p}{2}} d\beta, \quad (3.59)$$

$$c' = \frac{\Gamma(\frac{v_2+p}{2})}{\Gamma(\frac{v_2}{2}) v_2^{p/2} \pi^{p/2} |s \mathbf{I}_p|^{1/2}}$$

$$\mathbf{m} = \frac{l_4 + l_5 x_0}{a},$$

$$s = \frac{a l_6 - (l_4 + l_5 x_0)' (l_4 + l_5 x_0)}{a^2 v_2},$$

$$v_2 = (n+k)p + 5.$$

Observe que (3.59) é uma função de x_0 . Vamos agora mostrar que $\pi(x_0 | \mathbf{y}, \mathbf{X})$ é própria. Em primeiro lugar observe que

$$\int_{-\infty}^{+\infty} \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 \propto E_{x_0} \left[E [(\beta' \beta)^{1/2}] \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}} \right], \quad (3.60)$$

onde

$$x_0 \sim t_v \left(l_7, \frac{l_8}{v(nkl_6 - l'_5 l_5)} \right). \quad (3.61)$$

E lembrando que a raiz quadrada é uma função côncava, segue pela desigualdade de Jensen (James, 1996) que

$$E [(\beta' \beta)^{1/2}] \leq E [(\beta' \beta)]^{1/2}. \quad (3.62)$$

Utilizando o resultado (Fang e Zhang 1990, pg 42)

$$E [\mathbf{x}' \mathbf{A} \mathbf{x}] = \text{tr}(\mathbf{A} \Sigma) + \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}, \quad (3.63)$$

onde a distribuição do vetor $\mathbf{x}$ tem vetor de médias $\boldsymbol{\mu}$ e matriz de covariância Σ , temos que para $\mathbf{A} = \mathbf{I}_p$

$$\begin{aligned} E [\beta' \beta] &= \text{tr}(s \mathbf{I}_p) + \mathbf{m}' \mathbf{m} = \\ &= \frac{\{(n+k)c_x + nk(x_0 - \bar{x})^2\}l_6 - m_3(x_0 + m_1)^2 - m_2}{\{(n+k)c_x + nk(x_0 - \bar{x})^2\}^2 v} p + \\ &\quad + \frac{m_3(x_0 + m_1)^2 + m_2}{\{(n+k)c_x + nk(x_0 - \bar{x})^2\}^2} \end{aligned} \quad (3.64)$$

onde

$$m_1 = \frac{l'_4 l_5}{l'_5 l_5}, \quad m_2 = l'_4 l_4 - \frac{(l'_4 l_5)^2}{l'_5 l_5}, \quad m_3 = l'_5 l_5,$$

não são funções de x_0 . Deste modo, temos que $E [\beta' \beta]$ corresponde a um polinômio de grau -2, quando visto como uma função de x_0 . Assim, temos que

$$E [(\beta' \beta)^{1/2}] \leq \underbrace{E [\beta' \beta]^{1/2}}_{\text{polinômio de grau -1}},$$

e portanto

$$E [(\beta' \beta)^{1/2}] \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}} \leq \underbrace{E [(\beta' \beta)]^{1/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}}}_{\text{polinômio de grau } (n+k)p+3}.$$

Lembrando que uma distribuição *t-student* só possui momentos finitos de ordem estritamente menor que seu número de graus de liberdade, temos que

$$E_{x_0} \left[E [(\beta' \beta)]^{1/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}} \right] < \infty$$

uma vez que a expressão acima pode ser reescrita como uma combinação linear de momentos de ordem menor ou igual a $(n+k)p+3$ de uma distribuição *t-student* com $v = (n+k)p+4$ graus de liberdade. Consequentemente

$$E_{x_0} \left[E [(\beta' \beta)^{1/2}] \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{(n+k)p+4}{2}} \right] < \infty$$

isto é,

$$\int_{-\infty}^{+\infty} \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 < \infty. \quad \triangle$$

Observe que no caso particular $p = 1$ a equação (3.59) se reduz a

$$E[\beta] = \int_{\beta} (\beta^2)^{1/2} \frac{\Gamma(\frac{v_2+1}{2})}{\Gamma(\frac{v_2}{2}) v_2^{1/2} \pi^{1/2} s^{1/2}} \left(1 + \frac{1}{v_2} \frac{(\beta - m)^2}{s/v_2} \right)^{-\frac{v_2+1}{2}} d\beta = m = \frac{l_4 + l_5 x_0}{a}. \quad (3.65)$$

E como $a = (n+k)c_x + nk(x_0 - \bar{x})^2$ é um polinômio de grau 2 quando visto como função de x_0 , temos que $E[\beta]$ representa um polinômio de grau -1 e deste modo, temos que (3.58) é proporcional a esperança de um polinômio de grau $n+k+3$ em relação a uma distribuição *t-student* com $v = n+k+4$ graus de liberdade. Assim, para $p = 1$ temos que $\pi(x_0 | \mathbf{y}, \mathbf{X})$ é própria. O primeiro momento desta distribuição é dado por

$$E[x_0 | \mathbf{y}, \mathbf{X}] = \int_{-\infty}^{+\infty} x_0 \pi(x_0 | \mathbf{y}, \mathbf{X}) dx_0 \propto E_{x_0} \left[x_0 \frac{l_4 + l_5 x_0}{a} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{\frac{n+k+4}{2}} \right] = \infty,$$

uma vez que a expressão acima pode ser reescrita como a esperança de um polinômio de grau $n + k + 4$. Portanto, segue que $\pi(x_0 | \mathbf{y}, \mathbf{X})$ não possui nenhum momento finito.

Observe ainda que para $p = 1$, a priori de referência obtida através da abordagem com 2-grupos é expressa por

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto \sigma^{-4} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad (3.66)$$

A priori de referência obtida através da abordagem com 4-grupos é dada por

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto \sigma^{-2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad (3.67)$$

A priori de Jeffreys é expressa como

$$\pi_J(x_0, \alpha, \beta, \sigma^2) \propto |\beta|/\sigma^5. \quad (3.68)$$

3.5 Resultados relacionados na literatura

Ghosh, Carlin e Srivastava (1995) apresentaram uma lista completa das prioris de referência associadas a todos possíveis agrupamentos e ordenações dos parâmetros do modelo de calibração linear univariada.

Os autores trabalharam com o modelo reparametrizado como

$$L_{\mathbf{y}}(x_0, \alpha_0, \beta, \sigma^2) \propto (\sigma^{-2})^{(n+k)/2} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (y_i - \alpha_0 - \beta(x_i - \bar{x}))^2 + \sum_{j=1}^k (y_{0j} - \alpha_0 - \beta(x_0 - \bar{x}))^2 \right) \right\} \quad (3.69)$$

onde $\alpha_0 = \alpha + \beta\bar{x}$. No entanto, como α_0 é uma transformação um-a-um de α , temos que a priori obtida pelos autores pode ser transformada para a parametrização utilizada por nós através de

$$\pi(x_0, \alpha, \beta, \sigma^2) \propto \pi(x_0, \alpha_0, \beta, \sigma^2) \left| \frac{\partial(x_0, \alpha_0, \beta, \sigma^2)}{\partial(x_0, \alpha, \beta, \sigma^2)} \right|. \quad (3.70)$$

Como era esperado, ao efetuar a transformação acima as prioris obtidas por Ghosh *et al* coincidiram exatamente com as prioris obtidas no presente trabalho (de fato, o jacobiano da transformação vale 1, e $\pi(x_0, \alpha, \beta, \sigma^2) = \pi(x_0, \alpha_0, \beta, \sigma^2)$).

Ghosh *et al* ainda demonstraram que as distribuições a posteriori marginais para x_0 (obtidas a partir das prioris de referência, e também da priori de Jeffreys) são próprias, embora não possuam nenhum momento finito. Em nosso trabalho, demonstramos uma generalização deste resultado para as prioris de referência, a saber, que a posteriori marginal para x_0 é própria, mas possui momentos finitos somente até a ordem $p - 1$. Observe que quando $p = 1$, nossos resultados coincidem com os de Ghosh *et al*.

Liseo (1993) obteve a priori de referência para o modelo de calibração linear univariado no caso particular em que $\alpha = 0$ e σ^2 é conhecido. Kubokawa e Robert (1994) trabalharam com o modelo de calibração linear multivariado, mas reparametrizado como um problema de estimação da razão das médias de duas distribuições normais.

Partindo do modelo original (equações 3.1 e 3.2) Kubokawa e Robert obtiveram, primeiramente, o modelo reduzido às estatísticas suficientes

$$\bar{y} \sim N_p \left(\alpha + \beta \bar{x}, \frac{\sigma^2}{n} \mathbf{I}_p \right), \quad \bar{y}_0 \sim N_p \left(\alpha + \beta x_0, \frac{\sigma^2}{k} \mathbf{I}_p \right), \quad (3.71)$$

$$\hat{\beta} \sim N_p \left(\beta, \frac{\sigma^2}{c_x} \mathbf{I}_p \right), \quad s \sim \text{Gama} \left(\frac{q}{2}, \frac{1}{2\sigma^2} \right), \quad (3.72)$$

onde $\hat{\beta}$, $\bar{y}$, $\bar{y}_0$ e s são mutualmente independentes, e

$$q = (n + k - 3)p, \quad \hat{\beta} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})/c_x,$$

$$\bar{x} = \sum_{i=1}^n x_i/n, \quad \bar{y} = \sum_{i=1}^n y_i/n, \quad \bar{y}_0 = \sum_{j=1}^k y_{0j}/k, \quad c_x = \sum_{i=1}^n (x_i - \bar{x})^2.$$

A partir do modelo reduzido às estatísticas suficientes, Kubokawa e Robert consideraram a seguinte reparametrização

$$\beta^* = \sqrt{c_x} \beta \quad x_0^* = \frac{x_0 - \bar{x}}{\sqrt{c_x(\frac{1}{n} + \frac{1}{k})}} \quad (3.73)$$

$$\mathbf{w} = \sqrt{c_x} \hat{\beta} \quad \mathbf{z} = \frac{\bar{y}_0 - \bar{y}}{\sqrt{\frac{1}{n} + \frac{1}{k}}} \quad (3.74)$$

onde $\mathbf{w}$, $\mathbf{z}$ e s são independentes e distribuídos como

$$\mathbf{w} \sim N_p(\boldsymbol{\beta}^*, \sigma^2 \mathbf{I}_p), \quad \mathbf{z} \sim N_p(x_0^* \boldsymbol{\beta}^*, \sigma^2 \mathbf{I}_p), \quad s \sim \text{Gama}\left(\frac{q}{2}, \frac{1}{2\sigma^2}\right). \quad (3.75)$$

Desta maneira, a nova função de verossimilhança, $L_{\mathbf{w}, \mathbf{z}, s}(x_0^*, \boldsymbol{\beta}^*, \sigma^2)$, associada a este modelo reparametrizado e reduzido é dada por

$$\propto (\sigma^{-2})^{p+q/2} \exp \left\{ -\frac{1}{2\sigma^2} [s + (\mathbf{w} - \boldsymbol{\beta}^*)'(\mathbf{w} - \boldsymbol{\beta}^*) + (\mathbf{z} - x_0^* \boldsymbol{\beta}^*)'(\mathbf{z} - x_0^* \boldsymbol{\beta}^*)] \right\}. \quad (3.76)$$

Observe que este modelo é equivalente ao modelo associado ao problema de estimação da razão das médias de duas distribuições normais, onde o x_0 faz o papel de ψ , o parâmetro que representa a razão das médias (veja Bernardo, 1977). Observe ainda, que o parâmetro α não aparece neste modelo.

Utilizando a abordagem dos 2-grupos, Kubokawa e Robert obtiveram a seguinte priori de referência associada ao modelo (3.76)

$$\pi^k(x_0^*, \boldsymbol{\beta}^*, \sigma^2) \propto (\sigma^{-2})^{(p+2)/2} (1 + x_0^{*2})^{-1/2}. \quad (3.77)$$

Transformando novamente a priori acima para a antiga parametrização, isto é,

$$\pi^k(x_0, \boldsymbol{\beta}, \sigma^2) \propto \pi^k(x_0^*, \boldsymbol{\beta}^*, \sigma^2) \left| \frac{\partial(x_0^*, \boldsymbol{\beta}^*, \sigma^2)}{\partial(x_0, \boldsymbol{\beta}, \sigma^2)} \right| \quad (3.78)$$

temos que

$$\begin{aligned} \pi^k(x_0, \boldsymbol{\beta}, \sigma^2) &\propto (\sigma^{-2})^{(p+2)/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2} \left| \sqrt{\frac{nk}{n+k}} \right| \\ &\propto (\sigma^{-2})^{(p+2)/2} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \end{aligned} \quad (3.79)$$

Observe que a priori acima é diferente da priori obtida por nós segundo a abordagem dos 2-grupos para o modelo original de calibração linear, expressa como

$$\pi(x_0, \alpha, \boldsymbol{\beta}, \sigma^2) \propto (\sigma^{-2})^{p+1} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2}. \quad (3.80)$$

De fato, este resultado era esperado uma vez que, além de efetuar transformações nos parâmetros do modelo (3.73), Kubokawa e Robert efetuaram uma transformação no próprio modelo ao reparametrizá-lo como

$$\mathbf{w} = \sqrt{c_x} \hat{\boldsymbol{\beta}}, \quad \mathbf{z} = \frac{\bar{y}_0 - \bar{y}}{\sqrt{\frac{1}{n} + \frac{1}{k}}}.$$

Deste modo, acabaram obtendo a priori de referência associada ao modelo para o problema de estimação da razão das médias de duas distribuições normais, ao invés do modelo de calibração linear.

Apesar do modelo utilizado por Kubokawa e Robert estar relacionado ao modelo original (de calibração linear) as prioris de referência obtidas nos dois modelos não coincidem uma vez que as prioris de referência são invariantes quanto a transformações um-a-um dos parâmetros do modelo e quanto a estatísticas suficientes, mas não quanto a transformações do próprio modelo paramétrico.

3.6 Distribuições condicionais completas para o amostrador de Gibbs

No apêndice A obtivemos as distribuições condicionais completas para a geração da distribuição a posteriori. Todavia a distribuição condicional $p(x_0 \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X})$ não possui uma forma conhecida. Para contornar este problema, utilizamos o mesmo artifício empregado por Kubokawa e Robert (1994), e introduzimos uma variável latente t que nos permitiu gerar amostras de x_0 , a partir da distribuição condicional aumentada $p(x_0, t \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X})$. De fato, introduzimos um novo passo de amostrador de Gibbs para gerar de $p(x_0, t \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X})$ dentro de cada ciclo do amostrador de Gibbs para geração da distribuição a posteriori.

Explicitamente, as distribuições condicionais completas são dadas por

- $\sigma^{-2} \mid x_0, \alpha, \beta, \mathbf{y}, \mathbf{X} \sim \text{Gama} \left(\frac{(n+k)p+4}{2}, \frac{l_1}{2} \right)$
- $\alpha \mid x_0, \beta, \sigma^2, \mathbf{y}, \mathbf{X} \sim N_p \left(l_2, \frac{\sigma^2}{n+k} \mathbf{I}_p \right)$
- $\beta \mid x_0, \alpha, \sigma^2, \mathbf{y}, \mathbf{X} \sim N_p \left(l_3, \frac{\sigma^2}{kx_0^2 + \sum_{i=1}^n x_i^2} \mathbf{I}_p \right)$
- $t \mid x_0, \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X} \sim \text{Gama} \left(\frac{1}{2}, 1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)$
- $x_0 \mid t, \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X} \sim N(m, v^2)$

onde

$$l_1 = \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\alpha} - \beta x_i)' (\mathbf{y}_i - \boldsymbol{\alpha} - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \beta x_0)' (\mathbf{y}_{0j} - \boldsymbol{\alpha} - \beta x_0),$$

$$l_2 = \frac{\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)}{n + k},$$

$$l_3 = \frac{\sum_{i=1}^n x_i (\mathbf{y}_i - \boldsymbol{\alpha}) + x_0 \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha})}{kx_0^2 + \sum_{i=1}^n x_i^2},$$

$$m = \frac{2\sigma^2 nkt\bar{x}^2 + c_x(n+k)\boldsymbol{\beta}' \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha})}{2\sigma^2 nkt + c_x(n+k)k\boldsymbol{\beta}'\boldsymbol{\beta}},$$

$$v^2 = \frac{\sigma^2(n+k)c_x}{2\sigma^2 nkt + c_x(n+k)k\boldsymbol{\beta}'\boldsymbol{\beta}}.$$

Capítulo 4

As prioris-g de Zellner

4.1 Introdução

Neste capítulo introduzimos uma nova classe de prioris, conhecidas como prioris-g, originalmente desenvolvida por Zellner (1986) como um procedimento para obter distribuições não-subjetivas informativas para o vetor de regressores lineares (β) em modelos de regressão linear múltiplo (MRLM).

Vários procedimentos para acessar distribuições a priori informativas a partir de informação subjetiva para os parâmetros do MRLM já foram desenvolvidos. Em geral, tais procedimentos baseiam-se em resultados de conjugação. Embora, estes procedimentos sejam bastante úteis para facilitar a elicitação de prioris para os parâmetros do MRLM, em muitos casos, é difícil elicitar os hiperparâmetros destas prioris. Por exemplo, pode ser muito difícil especificar os hiperparâmetros de uma priori para uma matriz de covariâncias (Zellner, 1986).

A motivação de Zellner em desenvolver a nova classe de prioris-g, foi obter uma priori não-subjetiva e informativa que pudesse ter utilidade prática nas situações mencionadas acima (Zellner, 1986).

4.2 A priori-g

Informalmente, as prioris-g correspondem a uma distribuição a posteriori “conceitual” ou “imaginária”. Isto é, partindo de uma amostra “imaginária” e de uma priori de referência, obtemos uma distribuição a posteriori, que é por sua vez, utilizada como a distribuição a priori para o problema inferencial real.

Vamos agora mostrar a construção formal da priori-g apresentada por Zellner (1986). Primeiramente considere o MRLM padrão

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{com} \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \mathbf{I}_n) \quad (4.1)$$

onde $\mathbf{I}_n$ é a matriz identidade $n \times n$; $\mathbf{y}$ é o vetor $n \times 1$ de respostas y_i , $i = 1, \dots, n$; $\mathbf{X}$ é a matriz $n \times m$ de covariáveis, onde as linhas representam as amostras e as colunas representam as variáveis; $\boldsymbol{\beta}$ é o vetor de parâmetros da regressão; e $\boldsymbol{\epsilon}$ é o vetor $n \times 1$ de erros $\epsilon_i \sim N(0, 1)$ independentes. Observe que em geral assume-se que $\epsilon_i \sim N(0, \sigma^2)$ onde a variância σ^2 é desconhecida. No entanto, assumimos aqui que σ^2 é conhecido e vale 1, pois este é o nosso caso de interesse.

Neste caso, a função de verossimilhança do MRLM é dada por

$$\begin{aligned} L_{\mathbf{y}, \mathbf{X}}(\boldsymbol{\beta}) &\propto \exp \{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\} \\ &\propto \exp \{-(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})/2\} \end{aligned} \quad (4.2)$$

onde $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Antes de observar $\mathbf{y}$, considere a amostra “conceitual” ou “imaginária” $\mathbf{y}_0$ gerada por

$$\mathbf{y}_0 = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_0 \quad \text{com} \quad \boldsymbol{\epsilon}_0 \sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n) \quad (4.3)$$

A matriz de covariáveis $\mathbf{X}$ é a mesma de (4.1), mas como a amostra é “conceitual”, permitimos que σ_0^2 seja diferente de 1 (lembre que tomamos $\sigma^2 = 1$) e assumimos que $g\sigma_0^2 = 1$, $0 < g < \infty$, onde g é conhecido. Usando a priori de referência $p(\boldsymbol{\beta}) \propto \text{cte}$, a distribuição a posteriori “conceitual” é dada por

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbf{D}_0) &\propto \exp \{-g(\mathbf{y}_0 - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}_0 - \mathbf{X}\boldsymbol{\beta})/2\} \\ &\propto \exp \{-(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_0)' g \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_0)/2\} \end{aligned} \quad (4.4)$$

onde $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_0$ e $\mathbf{D}_0$ a amostra “conceitual”.

Agora, suponha que conhecemos antecipadamente o valor de $\boldsymbol{\beta}$ e o denotamos $\boldsymbol{\beta}_a$. Assumindo que

$$\boldsymbol{\beta}_a = E(\boldsymbol{\beta} \mid \mathbf{D}_0) = \hat{\boldsymbol{\beta}}_0, \quad (4.5)$$

temos que (4.4) é dada por

$$p(\boldsymbol{\beta} \mid \mathbf{D}_0) \propto \exp \{-(\boldsymbol{\beta} - \boldsymbol{\beta}_a)' g \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \boldsymbol{\beta}_a)/2\}. \quad (4.6)$$

Lembrando que esta distribuição a posteriori “conceitual” foi desenvolvida antes de observarmos o vetor de respostas $\mathbf{y}$, podemos interpretá-la como uma distribuição a priori, a ser utilizada para calcular a distribuição a posteriori de β dado o vetor de respostas reais $\mathbf{y}$. Deste modo,

$$p(\beta \mid \theta_0) \propto \exp \{-(\beta - \beta_a)' g \mathbf{X}' \mathbf{X} (\beta - \beta_a) / 2\} \quad (4.7)$$

é a priori- g onde $\theta_0 = (\beta_a, g)$.

Com relação a estas prioris, Zellner comenta que em situações onde o valor de g é desconhecido pode-se introduzir uma priori gama, informativa ou não, para este parâmetro, e é possível, ainda, ter $g = g(n)$, por exemplo, $g \propto 1/n$ ou $g \propto \log(n)/n$, de modo que a precisão da priori- g fica dependente do tamanho da amostra.

As prioris- g são caracterizadas por uma matriz de covariância da forma $g^{-1}(\mathbf{X}'\mathbf{X})^{-1}$ onde $0 < g < \infty$. Deste modo, estas prioris compartilham a mesma estrutura de variância e covariância da função de verossimilhança, embora permitam um maior ou menor grau de difusão, dependendo do valor de g . Em outras palavras, valores pequenos de g tornam a priori difusa, enquanto valores grandes a tornam mais fechada e informativa. Vemos assim, que g funciona como um parâmetro que define a escala de variação da priori- g . Observe que quando g vale 1, a priori- g possui exatamente a mesma estrutura de variância e covariância da função de verossimilhança.

Do ponto de vista de aplicações, o maior interesse nesta classe de prioris decorre de sua praticidade em especificar prioris difusas ($g \ll 1$) que, pelo menos em teoria, evitam problemas significativos de conflito entre a priori e a verossimilhança (West, 2000). Este tipo de problema é bastante comum quando utilizam-se prioris vagas, ainda que próprias, em modelos de grande dimensão. As prioris- g evitam este tipo de conflito uma vez que a verossimilhança “orienta” a forma da priori- g , isto é, a priori- g possui sua estrutura de variância e covariância proporcional a estrutura da função de verossimilhança (West, 2000).

No próximo capítulo, mostraremos que as prioris de referência associadas aos modelos de regressão binária e binomial, correspondem a uma priori- g . É interessante notar, que apesar da priori de referência ser mais difusa que a função de verossimilhança, ela ainda é razoavelmente informativa. Este resultado demonstra, como já foi discutido anteriormente, que as prioris de referência não são necessariamente difusas, e portanto “não-informativas”. De fato, em alguns casos, as prioris de referência, isto é, prioris que são dominadas pelos dados mesmo para amostras pequenas, podem ser “informativas”.

Capítulo 5

Análise de referência para modelos de regressão binária e binomial

5.1 Introdução

Neste capítulo estudaremos o problema de regressão com dados binários e binomiais. Em regressão binária os resultados podem ser classificados em duas categorias distintas que, por conveniência, denominaremos de 0 caso a resposta pertença à primeira categoria e 1 caso pertença à segunda. Uma vasta literatura em estatística, biometria e econometria tem sido produzida em análise de dados binários, veja por exemplo, Johnson e Albert (2000). A abordagem frequentista é baseada na teoria assintótica, e por esse motivo é questionável em amostras pequenas. Na abordagem bayesiana uma solução bastante popular para este problema foi proposta por Albert e Chib (1993). Nesta abordagem são introduzidas variáveis latentes, cuja função é facilitar a implementação da inferência a posteriori via métodos MCMC, que incorporam as variáveis latentes como dados perdidos a serem imputados e inferidos juntamente com o vetor de coeficientes da regressão (β).

Do ponto de vista de aplicações, a grande maioria dos trabalhos baseados na abordagem bayesiana utilizam uma priori uniforme para β . Esta escolha é motivada pelo uso desta priori em modelos lineares, onde a uniforme é a priori de referência (e também a de Jeffreys) associada ao modelo linear. No presente capítulo, vamos obter a priori de referência para β associada ao modelo de regressão binária. Além disso, apresentaremos um esquema computacional, utilizando o algoritmo proposto por Albert e Chib (1993), para a geração da distribuição a posteriori via amostrador de Gibbs.

Estudaremos também o modelo de regressão com respostas binomiais, que pode ser visto como uma extensão do modelo de dados binários. Mostraremos que a priori de referência associada ao modelo de regressão binomial é idêntica a priori de referência para o modelo de regressão binária. Apresentaremos também um esquema computacional, baseado no algoritmo de Albert e Chib, para a obtenção da distribuição a posteriori através do amostrador de Gibbs. No próximo capítulo, veremos que a priori de referência associada a estes dois modelos é obtida como um passo intermediário na construção da priori de referência para o modelo de calibração binomial.

Na próxima seção introduziremos rapidamente a idéia por trás dos modelos de regressão com dados binários seguindo Johnson e Albert (2000).

5.2 Regressão em dados binários

Em modelos de regressão binária estamos interessados em encontrar uma função que relacione as variáveis resposta (binárias) z_i 's com as variáveis explicativas x_i 's.

Talvez a suposição mais simples que podemos fazer a respeito da relação entre as variáveis resposta e as explicativas é que

$$p_i = \beta_1 + \beta_2 x_i, \quad \text{onde} \quad p_i = Pr(z_i = 1), \quad (5.1)$$

isto é, estamos supondo que a probabilidade de sucesso da variável resposta é uma função linear da variável explicativa (para efeitos didáticos estamos adotando aqui um dos modelos lineares mais simples; utilizando apenas uma única variável explicativa). Todavia, a utilização de uma reta para relacionar p_i com x_i não é adequada para se fazer previsões a respeito de p_i , uma vez que valores fora da amplitude dos valores de x_i que foram utilizados na obtenção da reta de mínimos quadrados, poderiam produzir valores de p_i maiores do que um ou até mesmo negativos.

Uma solução mais adequada para este problema é modelar as probabilidades de sucesso usando funções que assumam valores somente no intervalo $[0, 1]$. Uma classe natural de funções com esta propriedade é a das funções de distribuição acumuladas (f.d.a).

Deste modo, temos que p_i pode ser modelado como

$$p_i = F(\beta_1 + \beta_2 x_i) \quad (5.2)$$

onde F é denominada função de ligação.

Existe um grande número de f.d.a's que podem ser usadas para modelar as probabilidades de sucesso e sua escolha é, na maioria das vezes, arbitrária. Por este motivo, é uma prática corriqueira testar diversas funções e escolher a que melhor se adeque aos dados (Johnson e Albert, 1999). As funções de ligação utilizadas com maior frequência são a normal padrão (também conhecida como ligação *probit*) e a logística (também conhecida como ligação *logit*).

Na próxima seção vamos introduzir formalmente o modelo de regressão binária associado a m variáveis explicativas.

5.3 O modelo de regressão binária

Suponha que n variáveis aleatórias binárias independentes $z_1, z_2, \dots, z_n$ são observadas, onde z_i tem distribuição de Bernoulli com probabilidade de sucesso p_i , isto é,

$$\begin{array}{c|cc} z_i & 1 & 0 \\ \hline \text{prob} & p_i & 1 - p_i \end{array}.$$

Suponha que p_i está relacionado a um conjunto de covariáveis através de um modelo de regressão binária definido como

$$p_i = F(\mathbf{x}_i' \boldsymbol{\beta}), \quad i = 1, \dots, n \quad (5.3)$$

onde $\boldsymbol{\beta}$ é um vetor $m \times 1$ de parâmetros desconhecidos, $\mathbf{x}_i' = (x_{i1}, x_{i2}, \dots, x_{im})$ é um vetor conhecido de covariáveis e F é a função de ligação.

Neste modelo, a função de verossimilhança é dada por

$$L_{\mathbf{z}, \mathbf{X}}(\boldsymbol{\beta}) = \prod_{i=1}^n [F(\mathbf{x}_i' \boldsymbol{\beta})]^{z_i} [1 - F(\mathbf{x}_i' \boldsymbol{\beta})]^{1-z_i}, \quad (5.4)$$

onde $\mathbf{z}$ é o vetor de dados binários (vetor de 0's e 1's) e $\mathbf{X} = (\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_n')'$ é a matriz $n \times m$ de covariáveis (observe que $\mathbf{x}_i'$ é uma linha da matriz $\mathbf{X}$).

5.4 A priori de referência para $\boldsymbol{\beta}$ no modelo de regressão binária com ligação *probit*

Nesta seção vamos construir a priori de referência associada ao modelo *probit* de regressão binária. Neste caso, $F = \Phi$, onde Φ é a f.d.a. de uma $N(0, 1)$. Em nossa construção, ao invés

de obtermos a priori de referência para β diretamente através do modelo (5.4), partimos do modelo reparametrizado

$$L_{\mathbf{z}}(\mathbf{p}) = \prod_{i=1}^n p_i^{z_i} (1 - p_i)^{1-z_i}, \quad (5.5)$$

e induzimos a priori de referência em β , a partir da priori de referência dos p_i 's.

Esta indução foi feita em dois passos:

1. a partir da priori de referência para $\mathbf{p}$, induzimos a priori de referência para o vetor de preditores lineares $(\mathbf{X}\beta)$ e
2. a partir desta última, induzimos a priori de referência para β .

Explicitamente, consideramos as seguintes transformações:

$$t_1 : \begin{array}{ccc} \mathbf{p} & \longrightarrow & \mathbf{X}\beta \\ [0, 1]^n & \longrightarrow & \mathbb{R}^n \end{array} \quad \text{onde} \quad t_1(u_1, \dots, u_n) = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

$$t_2 : \begin{array}{ccc} \mathbf{X}\beta & \longrightarrow & \beta \\ \mathbb{R}^n & \longrightarrow & \mathbb{R}^m \end{array} \quad \text{onde} \quad t_2(\mathbf{v}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{v})$$

Observe que esta construção é lícita, uma vez que tanto t_1 quanto t_2 são transformações um-a-um, e já vimos no capítulo 2 que as prioris de referência são invariantes quanto a este tipo de transformações.

Utilizando o roteiro acima, obtivemos o seguinte resultado:

Teorema 5.1 *A priori de referência associada ao modelo de regressão binária é aproximadamente dada por*

$$\beta \sim N_m(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

onde $\sigma^2 = \pi^2/2$ e $\beta \in \mathbb{R}^m$.

Prova:

Em primeiro lugar, vamos encontrar a matriz de informação de Fisher associada ao modelo reparametrizado

$$L_{\mathbf{z}}(\mathbf{p}) = \prod_{i=1}^n p_i^{z_i} (1 - p_i)^{1-z_i}. \quad (5.6)$$

O logaritmo da função de verossimilhança é dado por

$$l(\mathbf{p}) = \sum_{i=1}^n z_i \log p_i + (1 - z_i) \log (1 - p_i). \quad (5.7)$$

As derivadas de primeira e segunda ordem são dadas por

$$\frac{\partial l}{\partial p_i} = \frac{z_i}{p_i} - \frac{1 - z_i}{1 - p_i}, \quad \frac{\partial^2 l}{\partial p_i^2} = -\frac{z_i}{p_i^2} - \frac{1 - z_i}{(1 - p_i)^2}, \quad \frac{\partial^2 l}{\partial p_i \partial p_j} = 0, \quad i \neq j,$$

de modo que

$$E \left[-\frac{\partial^2 l}{\partial p_i \partial p_j} \right] = 0 \quad \text{para } i \neq j$$

e

$$E \left[-\frac{\partial^2 l}{\partial p_i^2} \right] = \frac{1}{p_i} + \frac{1}{1 - p_i} = p_i^{-1}(1 - p_i)^{-1}.$$

Assim, a matriz de informação de Fisher é expressa como

$$H(\mathbf{p}) = \begin{pmatrix} p_1^{-1}(1 - p_1)^{-1} & 0 & \cdots & 0 \\ 0 & p_2^{-1}(1 - p_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n^{-1}(1 - p_n)^{-1} \end{pmatrix}. \quad (5.8)$$

Observe que neste modelo reparametrizado, nosso parâmetro de interesse é o vetor $\mathbf{p}$ inteiro. Deste modo, para a obtenção da priori de referência, agrupamos os parâmetros $p_1, \dots, p_n$ em um único grupo correspondente ao parâmetro de interesse. Note que a priori de referência obtida a partir de um único grupo corresponde a priori de Jeffreys.

Deste modo, a priori de referência é dada por

$$\pi(\mathbf{p}) \propto |H(\mathbf{p})|^{1/2} \propto \prod_{i=1}^n \frac{1}{\pi} p_i^{1/2-1} (1 - p_i)^{1/2-1}. \quad (5.9)$$

Observe que cada p_i é independentemente distribuído como uma Beta(1/2, 1/2).

Vamos agora, com base na priori de referência para $\mathbf{p}$, induzir uma priori de referência para β . Na realidade, primeiramente induziremos uma priori de referência para o vetor de preditores lineares ($\mathbf{X}\beta$) e então, a partir desta priori, induziremos a priori de referência para β .

Lembrando que $p_i = \Phi(\mathbf{x}_i' \beta)$ e que o vetor de preditores lineares

$$\mathbf{X}\beta = \begin{pmatrix} \mathbf{x}_1' \beta \\ \mathbf{x}_2' \beta \\ \vdots \\ \mathbf{x}_n' \beta \end{pmatrix} \quad (5.10)$$

é uma transformação um-a-um do vetor $\mathbf{p}$

$$\mathbf{p} = \begin{pmatrix} \Phi(\mathbf{x}_1'\beta) \\ \Phi(\mathbf{x}_2'\beta) \\ \vdots \\ \Phi(\mathbf{x}_n'\beta) \end{pmatrix} \quad (5.11)$$

isto é, $\mathbf{X}\beta = \Phi^{-1}(\mathbf{p})$, utilizamos a transformação

$$\pi_{\mathbf{X}\beta}(\mathbf{X}\beta) = \pi_{\mathbf{p}}(\Phi(\mathbf{X}\beta)) \left| \mathbf{J}_{\mathbf{p} \rightarrow \mathbf{X}\beta} \right| \quad (5.12)$$

para induzir a priori de referência no vetor de preditores lineares, onde $\left| \mathbf{J}_{\mathbf{p} \rightarrow \mathbf{X}\beta} \right|$ (o jacobiano da transformação vetorial) é dado por

$$\begin{aligned} \left| \mathbf{J}_{\mathbf{p} \rightarrow \mathbf{X}\beta} \right| &= \begin{vmatrix} \frac{\partial \Phi(\mathbf{x}_1'\beta)}{\partial (\mathbf{x}_1'\beta)} & \frac{\partial \Phi(\mathbf{x}_1'\beta)}{\partial (\mathbf{x}_2'\beta)} & \cdots & \frac{\partial \Phi(\mathbf{x}_1'\beta)}{\partial (\mathbf{x}_n'\beta)} \\ \frac{\partial \Phi(\mathbf{x}_2'\beta)}{\partial (\mathbf{x}_1'\beta)} & \frac{\partial \Phi(\mathbf{x}_2'\beta)}{\partial (\mathbf{x}_2'\beta)} & \cdots & \frac{\partial \Phi(\mathbf{x}_2'\beta)}{\partial (\mathbf{x}_n'\beta)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Phi(\mathbf{x}_n'\beta)}{\partial (\mathbf{x}_1'\beta)} & \frac{\partial \Phi(\mathbf{x}_n'\beta)}{\partial (\mathbf{x}_2'\beta)} & \cdots & \frac{\partial \Phi(\mathbf{x}_n'\beta)}{\partial (\mathbf{x}_n'\beta)} \end{vmatrix} \\ &= \begin{vmatrix} \phi(\mathbf{x}_1'\beta) & 0 & \cdots & 0 \\ 0 & \phi(\mathbf{x}_2'\beta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi(\mathbf{x}_n'\beta) \end{vmatrix} = \prod_{i=1}^n \phi(\mathbf{x}_i'\beta). \end{aligned}$$

Deste modo, lembrando que $\phi(\cdot)$ é a função de densidade da normal padrão, temos que

$$\pi(\mathbf{X}\beta) = \prod_{i=1}^n \frac{1}{\pi} [\Phi(\mathbf{x}_i'\beta)]^{1/2-1} [1 - \Phi(\mathbf{x}_i'\beta)]^{1/2-1} \frac{1}{\sqrt{2\pi}} \exp \{ -(\mathbf{x}_i'\beta)^2/2 \} \quad (5.13)$$

Observando o desenho de cada componente desta densidade (veja a figura 5.1), notamos uma flagrante semelhança com uma densidade normal centrada em zero. Esta semelhança sugere a possibilidade de obtermos uma ótima aproximação normal para esta densidade. Observe que não estamos falando de uma aproximação em um sentido assintótico, mas sim de uma simples aproximação de funções. Com isso queremos dizer simplesmente que as funções de densidade são muito parecidas. De fato, cada componente da função de densidade exata para o vetor de preditores lineares $\mathbf{X}\beta$ tem uma densidade muito parecida com a função de

densidade de uma normal centrada em zero e com variância σ^2 (que será calculada a seguir). Por este motivo podemos aproximar a densidade exata do vetor de preditores lineares pela densidade de um produto de distribuições $N(0, \sigma^2)$.

Vamos agora encontrar o valor para o σ^2 . Como cada componente da densidade original está centrado no zero, vamos aproximá-los por normais também centradas no zero e com variância σ^2 estimada da seguinte maneira: Em primeiro lugar, igualamos a densidade exata a densidade de uma $N(0, \sigma^2)$,

$$\frac{1}{\pi \sqrt{\Phi(\mathbf{x}_i' \boldsymbol{\beta})[1 - \Phi(\mathbf{x}_i' \boldsymbol{\beta})]}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mathbf{x}_i' \boldsymbol{\beta})^2}{2}} \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathbf{x}_i' \boldsymbol{\beta})^2}{2\sigma^2}} \quad (5.14)$$

em seguida consideramos $\mathbf{x}_i' \boldsymbol{\beta} = 0$ de modo que

$$\frac{1}{\pi \sqrt{1/2(1 - 1/2)}} \frac{1}{\sqrt{2\pi}} \approx \frac{1}{\sqrt{2\pi}\sigma} \Rightarrow \sigma \approx \frac{\pi}{2}. \quad (5.15)$$

Desta maneira, temos que,

$$\mathbf{x}_i' \boldsymbol{\beta} \sim N(0, \pi^2/4) \quad (5.16)$$

isto é, a priori de referência para cada componente do vetor de preditores lineares é, aproximadamente uma $N(0, \pi^2/4)$. Aqui os símbolos $\sim$ e $\approx$ são notações para distribuição aproximada e valor aproximado, respectivamente.

De fato, a partir da figura (5.1),

Figura 5.1: Densidades exata e aproximada

podemos apreciar a precisão desta aproximação. Na figura (5.1) a linha contínua representa a densidade exata, e a linha tracejada representa a densidade normal aproximada¹.

Desta forma, temos que a priori de referência para $\mathbf{X}\beta$ é muito bem aproximada pela densidade

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}(\pi/2)} \exp \left\{ -\frac{(\mathbf{x}_i' \beta)^2}{2(\pi^2/4)} \right\}. \quad (5.17)$$

Portanto, temos que

$$\mathbf{X}\beta \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad \text{onde } \sigma^2 = \pi^2/4. \quad (5.18)$$

Como $\mathbf{X}\beta$ é uma transformação um-a-um de β , vamos induzir a priori de referência para β a partir da priori de $\mathbf{X}\beta$.

Primeiramente observe que se $\mathbf{X}'\mathbf{X}$ é uma matriz não singular e portanto inversível, então

$$[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}\beta = \beta,$$

de modo que

$$\begin{aligned} \mathbf{X}\beta &\sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \Rightarrow \\ [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}\beta &= \beta \sim N_m(\mathbf{0}, [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\sigma^2 \mathbf{I}_n[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']) \\ &\sim N_m(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\ &\sim N_m(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \end{aligned}$$

Portanto, a priori de referência para β é dada por

$$\beta \sim N_m(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}), \quad (5.19)$$

onde $\sigma^2 = \pi^2/4$. △

¹Uma maneira mais objetiva de estimar o valor de σ^2 que aproxima melhor a densidade normal da densidade exata é calcular o valor de σ^2 que minimiza a distância simétrica de Kullback-Liebler entre a densidade exata e a densidade normal aproximada. A distância simétrica de Kullback-Liebler é definida como $J[f_1 : f_2] = K[f_1 : f_2] + K[f_2 : f_1]$, onde $K[f_1 : f_2] = \int f_1(x) \log \{f_1(x)/f_2(x)\} dx$ é a medida de divergência de Kullback-Liebler entre f_1 e f_2 (Soofi, 2000). Observe que ao contrário de $K[:]$, $J[:]$ é uma medida simétrica, isto é, $J[f_1 : f_2] = J[f_2 : f_1]$. Em nosso caso denotamos a densidade exata por f_1 e a densidade aproximada por f_2 . Utilizando integração de Monte Carlo, obtivemos que o valor de σ que minimiza $J[f_1 : f_2]$ é dado por 1.54, enquanto que o valor estimado através das equações (5.14) e (5.15) é $\pi/2 \approx 1.57$.

Observe que esta priori é própria² e portanto imune a patologias como paradoxos de marginalização, inconsistência forte e obtenção de posterioris impróprias. Perceba também que esta priori corresponde a uma priori- g de Zellner onde $g = \frac{4}{\pi^2} \approx 0.406$. É muito interessante notar que a priori de referência para β no modelo de regressão binária com ligação *probit* corresponda a uma distribuição originalmente desenvolvida por Zellner (1986) como um procedimento para obter distribuições não-subjetivas informativas para β em modelos de regressão linear múltiplo. O fato de uma priori para o modelo de regressão linear múltiplo estar relacionada a uma priori para o modelo de regressão probit não é inesperado, uma vez que podemos mostrar (veja a próxima seção) que existe um modelo de regressão linear múltiplo latente ao modelo probit de regressão binária. No entanto, é interessante que a priori de referência no modelo binário corresponda a uma priori razoavelmente informativa para o modelo de regressão linear múltiplo.

Observando que um valor de $g = 1$ corresponde a uma priori- g cuja matriz de covariância está exatamente na mesma escala da estrutura de variação e covariação da função de verossimilhança, percebemos que um valor de $g = 0.406$ ainda corresponde a uma priori informativa, embora um pouco mais difusa que a função de verossimilhança. Deste modo, fica claro que no contexto deste modelo, a priori de referência é razoavelmente informativa, ilustrando a discussão feita no segundo capítulo onde argumentamos que as prioris de referência não devem ser interpretadas como prioris “vagas”, mas sim como prioris dominadas pelos dados. Este modelo mostra claramente que prioris dominadas pelos dados podem ser razoavelmente “precisas”.

5.5 A estrutura normal latente do modelo *probit* de regressão binária

Para a implementação da inferência a posteriori exploramos a estrutura normal subjacente de variáveis latentes apresentada pelo modelo *probit*. Utilizamos a construção padrão proposta por Albert e Chib (1993), uma vez que ela permite uma implementação rotineira de métodos MCMC para a obtenção das inferências a posteriori.

Para introduzir a estrutura normal de variáveis latentes dentro do modelo de regressão

²Embora normalmente a priori de referência seja imprópria quando o espaço paramétrico é infinito, para este modelo obtivemos uma priori própria.

binária, introduzimos, para cada observação i , a variável contínua w_i , tal que

$$w_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n \quad (5.20)$$

onde $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são independentes e identicamente distribuídos segundo uma $N(0, 1)$. Observe que $w_i \sim N(\mathbf{x}_i' \boldsymbol{\beta}, 1)$.

Assumimos que

$$z_i = \begin{cases} 0, & \text{se } w_i \leq 0 \\ 1, & \text{se } w_i > 0 \end{cases} \quad (5.21)$$

Ou seja, $w_i > 0$ se e somente se a i -ésima observação é um sucesso. Deste modo temos que

$$\begin{aligned} p_i &= \Pr(z_i = 1) = \Pr(w_i > 0) = 1 - \Pr(w_i \leq 0) \\ &= 1 - \int_{-\infty}^0 (2\pi)^{-1/2} \exp \left\{ -(w_i - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2 \right\} dw_i \end{aligned}$$

e fazendo a mudança de variáveis

$$\begin{aligned} u_i &= w_i - \mathbf{x}_i' \boldsymbol{\beta} \Rightarrow du_i = dw_i \\ w_i &= -\infty \Rightarrow u_i = -\infty \\ w_i &= 0 \Rightarrow u_i = -\mathbf{x}_i' \boldsymbol{\beta} \end{aligned}$$

temos

$$p_i = 1 - \int_{-\infty}^{-\mathbf{x}_i' \boldsymbol{\beta}} (2\pi)^{-1/2} \exp \left\{ -u_i^2 / 2 \right\} du_i = 1 - \Phi(-\mathbf{x}_i' \boldsymbol{\beta})$$

e como $\Phi(\cdot)$ é simétrica ao redor do zero temos que

$$p_i = 1 - \Phi(-\mathbf{x}_i' \boldsymbol{\beta}) = \Phi(\mathbf{x}_i' \boldsymbol{\beta}). \quad (5.22)$$

Desta maneira, vemos que a estrutura subjacente normal de variáveis latentes produz um modelo equivalente ao *probit*.

Utilizando esta estrutura latente normal a função de verossimilhança original (equação 5.4) é substituída pela função de verossimilhança aumentada

$$L_{\mathbf{w}, \mathbf{z}, \mathbf{X}}(\boldsymbol{\beta}) = \prod_{i=1}^n (I_{(w_i > 0)} I_{(z_i = 1)} + I_{(w_i \leq 0)} I_{(z_i = 0)}) (2\pi)^{-1/2} \exp \left\{ -(w_i - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2 \right\} \quad (5.23)$$

e em forma matricial

$$L_{\mathbf{w}, \mathbf{z}, \mathbf{X}}(\boldsymbol{\beta}) \propto \exp \left\{ -(\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{w} - \mathbf{X}\boldsymbol{\beta}) / 2 \right\} \prod_{i=1}^n (I_{(w_i > 0)} I_{(z_i = 1)} + I_{(w_i \leq 0)} I_{(z_i = 0)}). \quad (5.24)$$

Deste modo, a distribuição a posteriori $\pi(\beta, \mathbf{w} \mid \mathbf{z}, \mathbf{X})$ é proporcional a

$$\exp \left\{ -\beta' \mathbf{X}' \mathbf{X} \beta / 2\sigma^2 \right\} \exp \left\{ -(\mathbf{w} - \mathbf{X}\beta)' (\mathbf{w} - \mathbf{X}\beta) / 2 \right\} \prod_{i=1}^n (I_{(w_i > 0)} I_{(z_i = 1)} + I_{(w_i \leq 0)} I_{(z_i = 0)}). \quad (5.25)$$

A partir desta distribuição podemos facilmente obter as condicionais completas para a implementação do amostrador de Gibbs.

5.6 O amostrador de Gibbs para o modelo de regressão binária

A idéia de introduzir variáveis latentes em um modelo, de modo a facilitar a utilização de algoritmos numéricos para geração de amostras de distribuições de probabilidade, foi inicialmente proposta por Tanner e Wong (1987), cujo algoritmo ficou conhecido como *Data Augmentation* (DA). Neste trabalho, utilizamos esta idéia e introduzimos variáveis latentes que espelham a estrutura normal subjacente ao modelo *probit*. Todavia, interpretando as variáveis latentes (quantidades desconhecidas) como mais um parâmetro do modelo, notamos que o algoritmo DA é equivalente ao amostrador de Gibbs (veja Tanner, 1996).

No apêndice B, apresentamos a derivação das distribuições condicionais completas para o amostrador de Gibbs. A saber:

- $\beta \mid \mathbf{z}, \mathbf{w}, \mathbf{X} \sim N_n \left(\frac{\sigma^2}{1+\sigma^2} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{w}, \frac{\sigma^2}{1+\sigma^2} (\mathbf{X}' \mathbf{X})^{-1} \right)$, onde $\sigma^2 = \pi^2/4$
- $w_i \mid \beta, \mathbf{z}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \beta, 1) & \text{truncada à esquerda do zero, se } z_i = 1 \\ N(\mathbf{x}_i' \beta, 1) & \text{truncada à direita do zero, se } z_i = 0 \end{cases}$.

Albert e Chib (1993), utilizando uma priori uniforme para β obtiveram distribuições condicionais completas idênticas as acima, exceto pelo termo $\frac{\sigma^2}{1+\sigma^2}$, que não aparece na distribuição condicional de

$$\beta \mid \mathbf{z}, \mathbf{w}, \mathbf{X} \sim N_n \left((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{w}, (\mathbf{X}' \mathbf{X})^{-1} \right) \quad (5.26)$$

obtida pelos autores.

Observe que se tomarmos $\sigma^2 \rightarrow \infty$ então $\frac{\sigma^2}{1+\sigma^2} \rightarrow 1$ e temos que as condicionais completas tornam-se idênticas nos dois casos. De fato, a priori de referência para β converge para a priori utilizada por Albert e Chib (uma uniforme na reta) quando $\sigma^2 \rightarrow \infty$.

5.7 O modelo de regressão binomial

Dados binomiais são obtidos através do agrupamento de dados binários. Mais especificamente, o modelo de regressão com respostas binomiais pode ser visto como um modelo de dados binários onde diversos elementos de uma amostra observada estão associados a um mesmo valor da variável explicativa. Deste modo, podemos definir uma variável discreta (binomial) y_i como o número de 1's em k_i respostas associadas a uma quantidade fixada x_i , $i = 1, 2, 3, \dots$. Observe que y_i nada mais é do que a soma de k_i variáveis binárias que estão associadas a um mesmo valor de x_i , isto é,

$$y_i = \sum_{j=1}^{k_i} z_{ij} \quad (5.27)$$

onde z_{ij} é uma variável binária que assume o valor 1 com probabilidade $p_i = \Pr(z_{ij} = 1) = F(\mathbf{x}_i' \beta)$.

O modelo de regressão binomial é definido da seguinte maneira,

$$y_i | x_i, \beta \stackrel{\text{ind}}{\sim} \text{Binomial}(k_i, F(\mathbf{x}_i' \beta)), \quad i = 1, 2, \dots, n \quad (5.28)$$

onde β é um vetor $m \times 1$ de parâmetros desconhecidos e $\mathbf{x}_i' = (x_{i1}, x_{i2}, \dots, x_{im})$ é um vetor conhecido de covariáveis.

Neste modelo temos que a função de verossimilhança é dada por

$$L_{\mathbf{y}, \mathbf{X}}(\beta) = \prod_{i=1}^n \binom{k_i}{y_i} [F(\mathbf{x}_i' \beta)]^{y_i} [1 - F(\mathbf{x}_i' \beta)]^{k_i - y_i} \quad (5.29)$$

onde $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ é o vetor de dados binomiais e $\mathbf{X} = (\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_n')'$ é a matriz $n \times m$ de covariáveis. Observe que $\mathbf{x}_i'$ é uma linha da matriz $\mathbf{X}$ e que esta matriz é idêntica a matriz de covariáveis associada ao modelo de regressão binária.

5.8 A priori de referência para o modelo *probit* de regressão binomial

Assim como na construção da priori de referência para o modelo de regressão binária, ao invés de obtermos a priori de referência para β diretamente através modelo (5.29), optamos por obter a priori primeiramente para os p_i 's via o modelo

$$L_{\mathbf{y}}(\mathbf{p}) \propto \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{k_i - y_i}, \quad (5.30)$$

para então induzir a priori de referência em β a partir da priori obtida para $\mathbf{p} = (p_1, p_2, \dots, p_n)'$.

A matriz de informação de Fisher associada a este modelo é muito parecida com a matriz associada ao modelo de regressão binária e vale

$$H(\mathbf{p}) = \begin{pmatrix} k_1 p_1^{-1} (1 - p_1)^{-1} & 0 & \cdots & 0 \\ 0 & k_2 p_2^{-1} (1 - p_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n p_n^{-1} (1 - p_n)^{-1} \end{pmatrix} \quad (5.31)$$

Desta maneira, a priori de referência para $\mathbf{p}$ é dada por

$$\pi(\mathbf{p}) \propto \prod_{i=1}^n k_i p_i^{-1/2} (1 - p_i)^{-1/2} \propto \prod_{i=1}^n p_i^{-1/2} (1 - p_i)^{-1/2}. \quad (5.32)$$

Observe que (5.32) é idêntica a priori de referência para $\mathbf{p}$ no modelo de regressão binária. Portanto a priori de referência para β induzida a partir da priori para $\mathbf{p}$ é mesma. Assim, temos que no modelo de regressão binomial, a priori de referência também tem distribuição

$$\beta \sim N_m(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (5.33)$$

onde $\sigma^2 = \pi^2/4$.

5.9 A estrutura normal latente do modelo *probit* de regressão binomial

Analogamente ao modelo de regressão binária introduzimos uma estrutura normal de variáveis latentes dentro do modelo de regressão binomial.

Para tanto, introduzimos, para cada observação i , a variável contínua w_{ij} , tal que

$$w_{ij} = \mathbf{x}_i' \beta + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, k_i \quad (5.34)$$

onde $\epsilon_{1j}, \epsilon_{2j}, \dots, \epsilon_{nj}$ são independentes e identicamente distribuídos segundo uma $N(0, 1)$. Observe que $w_{ij} \sim N(\mathbf{x}_i' \beta, 1)$.

Assumindo que

$$z_{ij} = \begin{cases} 0, & \text{se } w_{ij} \leq 0 \\ 1, & \text{se } w_{ij} > 0 \end{cases}, \quad (5.35)$$

onde $y_i = \sum_{j=1}^{k_i} z_{ij}$. Ou seja, $w_{ij} > 0$ se e somente se a ij -ésima observação é um sucesso. Analogamente ao caso binário, temos que no modelo binomial $p_i = Pr(z_{ij} = 1) = Pr(w_{ij} > 0) = \Phi(\mathbf{x}_i' \boldsymbol{\beta})$. Desta maneira, vemos que a estrutura normal de variáveis latentes subjacente ao modelo de regressão binomial também produz um modelo equivalente ao *probit*.

Utilizando esta estrutura latente normal, a função de verossimilhança original (5.29) pode ser substituída pela função de verossimilhança aumentada

$$L_{\mathbf{w}, \mathbf{y}, \mathbf{X}}(\boldsymbol{\beta}) = \prod_{i=1}^n \prod_{j=1}^{k_i} (I_{(w_{ij} > 0)} I_{(z_{ij}=1)} + I_{(w_{ij} \leq 0)} I_{(z_{ij}=0)}) \times (2\pi)^{-1/2} \exp \left\{ -(w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2 \right\} \quad (5.36)$$

E a distribuição a posteriori é dada por

$$\begin{aligned} \pi(\boldsymbol{\beta}, \mathbf{w} \mid \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \right\} \times \left(\exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{y_i} (w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2 \right\} I_{(w_{ij} > 0)} + \right. \\ \left. + \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=y_i+1}^{k_i} (w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2 \right\} I_{(w_{ij} \leq 0)} \right) \end{aligned} \quad (5.37)$$

A partir desta distribuição podemos facilmente obter as condicionais completas para a implementação do amostrador de Gibbs.

5.10 Distribuições condicionais completas para o amostrador de Gibbs no modelo de regressão binomial

As distribuições condicionais completas para o amostrador de Gibbs no modelo de regressão binomial são dadas por:

- $\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X} \sim N_m(\sigma^2[\mathbf{X}'(\sigma^2 \mathbf{K} + \mathbf{I})\mathbf{X}]^{-1} \mathbf{X}' \mathbf{K} \bar{\mathbf{w}}, [\mathbf{X}'(\sigma^2 \mathbf{K} + \mathbf{I})\mathbf{X}]^{-1})$, onde $\sigma^2 = \pi^2/4$
- $w_{ij} \mid \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \boldsymbol{\beta}, 1) & \text{truncada à esquerda do zero, se } z_{ij} = 1 \\ N(\mathbf{x}_i' \boldsymbol{\beta}, 1) & \text{truncada à direita do zero, se } z_{ij} = 0 \end{cases}$.

Veja o apêndice B, para a derivação destas distribuições.

Capítulo 6

Análise de referência para o modelo de calibração binomial

6.1 Introdução

No capítulo 3 abordamos o problema de calibração sob hipótese de linearidade, normalidade e homocedasticidade dos erros. Neste capítulo, estudaremos o problema de calibração com dados binomiais.

Um exemplo de um problema de calibração com dados binomiais é o de dosimetria citogenética. Neste problema, temos que o x_i representa uma dose de radiação fixada e y_i o número de células com aberrações cromossômicas entre as k_i células que sofreram a exposição a dose de radiação x_i , $i = 1, \dots, n$. Observe que cada célula que sofreu exposição a mesma dose de radiação x_i representa uma variável binária que assume o valor 1 caso um sucesso, isto é, uma aberração cromossômica tenha ocorrido, e 0 caso contrário.

O objetivo da dosimetria citogenética é estimar uma dose de radiação desconhecida x_0 , baseando-se no número de células que sofreram aberrações, y_0 , em k_0 células analisadas, e na curva de calibração (veja Branco *et al*, 2003).

6.2 O modelo binomial de calibração

O modelo de calibração binomial é definido da seguinte maneira,

$$y_i | x_i, \beta \stackrel{ind}{\sim} \text{Binomial}(k_i, F(\mathbf{x}_i' \beta)), \quad i = 0, 1, \dots, n, \quad (6.1)$$

onde

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad (6.2)$$

e

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad (6.3)$$

de modo que $\mathbf{x}_i' \boldsymbol{\beta} = \beta_1 + \beta_2 x_i$

Assume-se que os y_i 's são independentes, isto é, a frequência observada das respostas associadas a um determinado valor da variável explicativa, é independente da frequência observada em respostas associadas a outro valor de covariável. Portanto, a função de verossimilhança é dada por

$$L_{\mathbf{y}, \mathbf{X}}(x_0, \boldsymbol{\beta}) = \prod_{i=0}^n \binom{k_i}{y_i} [F(\mathbf{x}_i' \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}_i' \boldsymbol{\beta})]^{k_i - y_i}, \quad (6.4)$$

onde $\mathbf{y} = (y_0, y_1, \dots, y_n)'$. Observando a expressão acima fica claro que não é fácil determinar a forma analítica da distribuição a posteriori para x_0 qualquer que seja a priori utilizada. Deste modo, torna-se imperativo o uso de métodos numéricos para realizar-se a inferência a posteriori.

6.3 A priori de referência para o modelo de calibração binomial

No modelo de calibração binomial identificamos claramente um parâmetro de interesse, o x_0 , e um parâmetro de perturbação, o $\boldsymbol{\beta} = (\beta_1, \beta_2)$.

Por uma questão de conveniência matemática, consideraremos a seguinte reparametrização do modelo de calibração binomial; $(x_0, \boldsymbol{\beta}) \rightarrow (z_0, \boldsymbol{\beta})$, onde $z_0 = \beta_1 + \beta_2 x_0 = \mathbf{x}_0' \boldsymbol{\beta}$. Neste modelo reparametrizado z_0 representa o parâmetro de interesse, e não é interpretado como uma função de $\boldsymbol{\beta}$.

Reescrevendo o modelo de calibração binomial na nova parametrização temos

$$L_{\mathbf{y}, \mathbf{X}}(z_0, \boldsymbol{\beta}) \propto [F(z_0)]^{y_0} [1 - F(z_0)]^{k_0 - y_0} \prod_{i=1}^n [F(\mathbf{x}_i' \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}_i' \boldsymbol{\beta})]^{k_i - y_i}. \quad (6.5)$$

Como pode ser visto pela equação (6.5), ao fazermos esta reparametrização estamos induzindo uma independência entre os parâmetros z_0 e β , uma vez que z_0 não é interpretado como uma função de β . Esta estrutura de independência torna muito mais fácil a obtenção da priori de referência para (z_0, β) , a qual, por sua vez, é utilizada para a indução da priori de referência em (x_0, β) através do jacobiano da transformação $(z_0, \beta) \rightarrow (x_0, \beta)$. Como veremos mais a frente, a estrutura de dependência entre x_0 e β é recuperada pela transformação jacobiana.

Assim como no capítulo anterior, ao invés de obtermos a priori de referência para β diretamente através da equação (6.5), optamos por obter a priori de referência primeiramente para as probabilidades de sucesso

$$(p_0, p_1, \dots, p_n) = (F(z_0), F(\mathbf{x}_1' \beta), \dots, F(\mathbf{x}_n' \beta)), \quad (6.6)$$

via modelo reparametrizado

$$L_{\mathbf{y}}(p_0, \mathbf{p}) \propto p_0^{y_0} (1 - p_0)^{k_0 - y_0} \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{k_i - y_i}, \quad (6.7)$$

para então, induzir a priori de referência para os preditores lineares. Somente então, induzimos a priori de referência em β , a partir da priori obtida para $(\mathbf{x}_1' \beta, \dots, \mathbf{x}_n' \beta)$. Observe que o z_0 não participa da indução da priori para β , uma vez que na reparametrização (z_0, β) o z_0 não é considerado como uma função de β .

Em síntese, a obtenção da priori de referência para (x_0, β) é dada através do seguinte esquema de induções:

$$(p_0, p_1, \dots, p_n) \xrightarrow{t_1} (z_0, \mathbf{x}_1' \beta, \dots, \mathbf{x}_n' \beta) \xrightarrow{t_2} (z_0, \beta) \xrightarrow{t_3} (x_0, \beta),$$

onde as transformações são dadas explicitamente por

$$\begin{aligned} t_1 : \begin{array}{l} \mathbf{p} \longrightarrow (z_0, \mathbf{X}\beta) \\ [0, 1]^{n+1} \longrightarrow \Re^{n+1} \end{array} & \quad \text{onde} \quad t_1(u_0, \dots, u_n) = (\Phi^{-1}(u_0), \dots, \Phi^{-1}(u_n)), \\ \\ t_2 : \begin{array}{l} (z_0, \mathbf{X}\beta) \longrightarrow (z_0, \beta) \\ \Re^{n+1} \longrightarrow \Re^{m+1} \end{array} & \quad \text{onde} \quad t_2[w, \mathbf{v}] = [w, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{v})], \\ \\ t_3 : \begin{array}{l} (z_0, \beta) \longrightarrow (x_0, \beta) \\ \Re^{m+1} \longrightarrow \Re^{m+1} \end{array} & \quad \text{onde} \quad t_3[w, \mathbf{v}] = \left[\frac{w - \beta_1}{\beta_2}, \mathbf{v} \right]. \end{aligned}$$

Observe que em cada um dos três passos de indução usamos as proposições (2.8) e (2.7) enunciadas no capítulo 2, que garantem que a priori de referência é invariante sob

transformações um-a-um dos parâmetros de interesse e perturbação. Esta construção é lícita uma vez que os preditores lineares são uma transformação um-a-um das probabilidades de sucesso, o vetor β é uma transformação um-a-um do vetor de preditores lineares e x_0 é uma transformação um-a-um de z_0 .

Com base neste esquema de induções obtivemos o seguinte resultado:

Teorema 6.1 *A priori de referência associada ao modelo de calibração binomial é aproximadamente dada por*

$$\pi(x_0, \beta) \approx |\beta_2| \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sigma^2} \exp \left\{ -\frac{\beta' \mathbf{X}'\mathbf{X} \beta}{2\sigma^2} \right\},$$

onde $\sigma^2 = \pi^2/4$, $x_0 \in \mathbb{R}$ e $\beta \in \mathbb{R}^2$.

Prova:

A matriz de informação de Fisher associada a equação (6.7) é idêntica a matriz do modelo de regressão binomial, exceto por sua maior dimensão, e vale

$$H(p_0, \mathbf{p}) = \begin{pmatrix} k_0 p_0^{-1} (1 - p_0)^{-1} & 0 & \cdots & 0 \\ 0 & k_1 p_1^{-1} (1 - p_1)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n p_n^{-1} (1 - p_n)^{-1} \end{pmatrix} \quad (6.8)$$

Trabalhando condicionalmente em relação a p_0 e agrupando todos demais parâmetros de perturbação em um único grupo, temos que a priori de referência condicional $\pi(\mathbf{p} \mid p_0)$ é idêntica a priori de Jeffreys, sendo calculada como a raiz quadrada do determinante da matriz obtida pela eliminação da primeira linha e coluna de (6.8). Explicitamente,

$$\pi(\mathbf{p} \mid p_0) \propto \prod_{i=1}^n p_i^{-1/2} (1 - p_i)^{-1/2}. \quad (6.9)$$

Observe que esta priori é própria e portanto pode ser utilizada para obtermos o modelo marginal $L_{\mathbf{y}}(p_0) = \int_{\mathbf{p}} L_{\mathbf{y}}(p_0, \mathbf{p}) \pi(\mathbf{p} \mid p_0) d\mathbf{p}$, que é proporcional a

$$\int_{p_1} \cdots \int_{p_n} p_0^{y_0} (1 - p_0)^{k_0 - y_0} \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{k_i - y_i} \times \prod_{i=1}^n p_i^{-1/2} (1 - p_i)^{-1/2} dp_1 \cdots dp_n \quad (6.10)$$

$$\propto p_0^{y_0} (1 - p_0)^{k_0 - y_0} \prod_{i=1}^n \int_{p_i} p_i^{y_i - 1/2} (1 - p_i)^{k_i - y_i - 1/2} dp_i \propto p_0^{y_0} (1 - p_0)^{k_0 - y_0}. \quad (6.11)$$

A priori de referência associada ao modelo marginal (6.11) é uma $Beta(1/2, 1/2)$. Deste modo,

$$\pi(p_0, \mathbf{p}) = \pi(p_0)\pi(\mathbf{p} | p_0) \propto p_0^{-1/2} (1 - p_0)^{-1/2} \prod_{i=1}^n p_i^{-1/2} (1 - p_i)^{-1/2}. \quad (6.12)$$

Observe que esta priori conjunta pode ser fatorada em

$$\pi(p_0, \mathbf{p}) = \pi(p_0)\pi(\mathbf{p}), \quad (6.13)$$

onde

$$\pi(p_0) \propto p_0^{1/2-1} (1 - p_0)^{1/2-1} \quad e \quad \pi(\mathbf{p}) \propto \prod_{i=1}^n p_i^{1/2-1} (1 - p_i)^{1/2-1}.$$

Vamos agora, com base na priori de referência para $(p_0, \mathbf{p})$ induzir uma priori de referência para $(z_0, \boldsymbol{\beta})$. Observe que os passos da indução da priori de referência para $(\mathbf{x}_1' \boldsymbol{\beta}, \dots, \mathbf{x}_n' \boldsymbol{\beta})$ a partir de $\pi(\mathbf{p})$, e da subsequente indução da priori de referência para $\boldsymbol{\beta}$ a partir da priori dos preditores lineares, são idênticos aos feitos no modelo de regressão binomial e portanto temos que

$$\boldsymbol{\beta} \sim N_2(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (6.14)$$

onde $\sigma^2 = \pi^2/4$.

Do mesmo modo, já vimos no capítulo 5 que uma priori $Beta(1/2, 1/2)$ para um p_i , induz uma densidade para o preditor linear que pode ser muito bem aproximada por uma distribuição $N(0, \pi^2/4)$. Assim, temos que $\pi(p_0)$ induz uma priori $N(0, \pi^2/4)$ em z_0 .

Desta maneira, temos que

$$\pi(z_0, \boldsymbol{\beta}) = N(0, \sigma^2) \times N_2(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (6.15)$$

Com base nesta priori, induziremos a priori de referência para $(x_0, \boldsymbol{\beta})$. Observe que

$$\pi(x_0, \boldsymbol{\beta}) = \pi(z_0, \boldsymbol{\beta}) \left| \frac{\partial(z_0, \boldsymbol{\beta})}{\partial(x_0, \boldsymbol{\beta})} \right| \quad (6.16)$$

onde

$$\frac{\partial(z_0, \boldsymbol{\beta})}{\partial(x_0, \boldsymbol{\beta})} = \begin{vmatrix} \frac{\partial z_0}{\partial x_0} & \frac{\partial z_0}{\partial \beta_1} & \frac{\partial z_0}{\partial \beta_2} \\ \frac{\partial \beta_1}{\partial x_0} & \frac{\partial \beta_1}{\partial \beta_1} & \frac{\partial \beta_1}{\partial \beta_2} \\ \frac{\partial \beta_2}{\partial x_0} & \frac{\partial \beta_2}{\partial \beta_1} & \frac{\partial \beta_2}{\partial \beta_2} \end{vmatrix} = \begin{vmatrix} \beta_2 & 1 & x_0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \beta_2 \quad (6.17)$$

Deste modo, a priori de referência para (x_0, β) é dada por

$$\pi(x_0, \beta) \approx \frac{|\beta_2|}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \frac{|\mathbf{X}' \mathbf{X}|^{1/2}}{2\pi\sigma^2} \exp \left\{ -\frac{\beta' \mathbf{X}' \mathbf{X} \beta}{2\sigma^2} \right\}. \quad \triangle$$

Observe que a estrutura de dependência entre x_0 e β foi recuperada, visto que não é possível fatorar $\pi(x_0, \beta)$ em uma função apenas de x_0 e em outra exclusivamente de β . Uma outra característica muito interessante desta priori é que devido ao termo $|\beta_2|$, ela possui densidade zero no ponto $\beta_2 = 0$. De fato, esta priori é simétrica, bimodal, tem média $(0, 0, 0)$ e tem massa zero no ponto $\beta_2 = 0$. Na figura (6.1) apresentamos o desenho da priori de referência para $\beta_1 = 0$, $\sum_{i=1}^n x_i^2 = 1$ e $\sigma^2 = 6.25$. Observe que em relação ao x_0 , a distribuição a priori conjunta tem caudas mais pesadas do que para β . Como veremos mais adiante nesta seção, a distribuição marginal de x_0 é aproximadamente uma Cauchy centrada em 0 e com parâmetro de escala igual a $\sum_{i=1}^n x_i^2$, enquanto a de β é aproximadamente normal com média 0 e com variância igual a $\sigma^2 / \sum_{i=1}^n x_i^2$.

Figura 6.1: Priori de referência para o modelo de calibração binária

Embora a primeira vista possa parecer estranho que esta priori seja bimodal devemos nos lembrar de que as probabilidades de sucesso $p_i = \Phi(\beta_1 + \beta_2 x_i)$ valem $p_i = \Phi(0) = 1/2$

no ponto $(\beta_1, \beta_2) = (0, 0)$, e de que a priori de referência para uma probabilidade de sucesso é uma $Beta(1/2, 1/2)$, cujo ponto de menor densidade é exatamente o $1/2$. Desta maneira, vemos que esta bimodalidade da priori de referência apenas espelha, como era de se esperar, a estrutura da priori de referência para probabilidades de sucesso.

Vamos agora, proceder uma breve comparação entre a priori de referência e a priori usualmente empregada na literatura para este modelo. Usualmente, a priori considerada para x_0 é uma priori normal centrada em $\bar{x}$ e com grande variância (priori “vaga”). Para β o usual é uma priori uniforme (Branco, et al 2003). A escolha da priori uniforme para β é motivada pelo seu uso em modelos lineares, enquanto a escolha da priori com grande variância reflete a escolha de uma distribuição “vaga” mas ainda assim própria. É interessante notar que nestas escolhas, x_0 e β são independentes. Isto entra em conflito com a estrutura da função de verossimilhança original (equação 6.4), que mostra claramente uma estrutura de dependência entre x_0 e β .

Como era de se esperar, a priori de referência preserva esta estrutura de dependência e por este motivo, não é direta a sua comparação com priori “usual”¹. No entanto, é possível ganhar alguma intuição sobre a relação destas duas propostas estudando as distribuições marginais da priori de referência.

O próximo resultado mostra que, marginalmente, tanto o β quanto o x_0 possuem distribuições de probabilidade conhecidas.

Corolário 6.1 *As distribuições de probabilidade marginais de β e x_0 são aproximadamente dadas por*

$$\begin{aligned}\beta &\sim N_2(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \\ x_0 &\sim Cauchy\left(\bar{x}, \frac{(n+1)c_x}{n}\right)\end{aligned}$$

onde $c_x = \sum_{i=1}^n (x_i - \bar{x})^2$.

Prova:

Vamos primeiramente obter a distribuição marginal de β . Integrando a priori conjunta em x_0

$$\pi(\beta) = \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sqrt{2\pi}\sigma^3} |\beta_2| \exp\left\{-\frac{\beta'\mathbf{X}'\mathbf{X}\beta}{2\sigma^2}\right\} \underbrace{\int_{-\infty}^{+\infty} \exp\left\{-\frac{(\beta_1 + \beta_2 x_0)^2}{2\sigma^2}\right\} dx_0}_{(I)}, \quad (6.18)$$

¹Neste capítulo, quando mencionarmos priori “vaga”, estaremos nos referindo a priori marginal de x_0 , que é própria mas possui grande dispersão. Quando mencionarmos priori “usual”, estaremos nos referindo a priori conjunta de x_0 e β , composta pela multiplicação da priori “vaga” pela priori imprópria $\pi(\beta) \propto 1$.

e como

$$\frac{(\beta_1 + \beta_2 x_0)^2}{2\sigma^2} = \frac{\beta_2^2 (x_0 - (-\frac{\beta_1}{\beta_2}))^2}{2\sigma^2} = \frac{(x_0 - (-\frac{\beta_1}{\beta_2}))^2}{2\frac{\sigma^2}{\beta_2^2}}, \quad (6.19)$$

temos que

$$(I) = \int_{-\infty}^{+\infty} \exp \left\{ -\frac{(x_0 - (-\frac{\beta_1}{\beta_2}))^2}{2\frac{\sigma^2}{\beta_2^2}} \right\} dx_0 = \frac{\sqrt{2\pi}\sigma}{\sqrt{\beta_2^2}}. \quad (6.20)$$

Portanto,

$$\pi(\beta) = \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sqrt{2\pi}\sigma^3} |\beta_2| \exp \left\{ -\frac{\beta'\mathbf{X}'\mathbf{X}\beta}{2\sigma^2} \right\} \frac{\sqrt{2\pi}\sigma}{\sqrt{\beta_2^2}}. \quad (6.21)$$

Lembrando que $\sqrt{\beta_2^2} = |\beta_2|$ temos

$$\pi(\beta) = \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sigma^2} \exp \left\{ -\frac{\beta'\mathbf{X}'\mathbf{X}\beta}{2\sigma^2} \right\}, \quad (6.22)$$

isto é, a distribuição marginal para β é uma $N_2(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. Observe que esta é exatamente a priori de referência para o problema de regressão binomial.

Vamos agora encontrar a distribuição marginal para x_0 . Integrando a priori conjunta em β ,

$$\pi(x_0) = \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sqrt{2\pi}\sigma^3} \underbrace{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\beta_2| \exp \left\{ -\frac{(\mathbf{x}_0'\beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta'\mathbf{X}'\mathbf{X}\beta}{2\sigma^2} \right\} d\beta_1 d\beta_2}_{(II)}, \quad (6.23)$$

e como $(\mathbf{x}_0'\beta)^2 = \beta'\mathbf{x}_0\mathbf{x}_0'\beta$ temos que

$$\exp \left\{ -\frac{(\mathbf{x}_0'\beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta'\mathbf{X}'\mathbf{X}\beta}{2\sigma^2} \right\} = \exp \left\{ -\frac{\beta'(\mathbf{x}_0\mathbf{x}_0' + \mathbf{X}'\mathbf{X})\beta}{2\sigma^2} \right\} \quad (6.24)$$

$$= \exp \left\{ -\frac{\beta'\mathbf{X}_0'\mathbf{X}_0\beta}{2\sigma^2} \right\}, \quad (6.25)$$

onde

$$\mathbf{X}_0 = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \quad (6.26)$$

Consequentemente,

$$\beta' \mathbf{X}_0' \mathbf{X}_0 \beta = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} n+1 & \sum_{i=0}^n x_i \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad (6.27)$$

$$= (n+1)\beta_1^2 + 2\beta_1\beta_2 \sum_{i=0}^n x_i + \beta_2^2 \sum_{i=0}^n x_i^2. \quad (6.28)$$

Deste modo

$$\begin{aligned} (II) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\beta_2| \exp \left\{ -\frac{(n+1)\beta_1^2 + 2\beta_1\beta_2 \sum_{i=0}^n x_i + \beta_2^2 \sum_{i=0}^n x_i^2}{2\sigma^2} \right\} d\beta_1 d\beta_2 \\ &= \int_{-\infty}^{+\infty} |\beta_2| \exp \left\{ -\frac{\beta_2^2 \sum_{i=0}^n x_i^2}{2\sigma^2} \right\} \underbrace{\left(\int_{-\infty}^{+\infty} \exp \left\{ -\frac{(n+1)\beta_1^2 + 2\beta_1\beta_2 \sum_{i=0}^n x_i}{2\sigma^2} \right\} d\beta_1 \right)}_{(III)} d\beta_2, \end{aligned} \quad (6.29)$$

completando o quadrado e fazendo algumas manipulações algébricas, temos que (III) pode ser reescrita como

$$(III) = \exp \left\{ \frac{\beta_2^2 (\sum_{i=0}^n x_i)^2}{2(n+1)\sigma^2} \right\} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{(\beta_1 - (-\frac{\beta_2 \sum_{i=0}^n x_i}{n+1}))^2}{2\frac{\sigma^2}{n+1}} \right\} d\beta_1 \quad (6.30)$$

$$= \exp \left\{ \frac{\beta_2^2 (\sum_{i=0}^n x_i)^2}{2(n+1)\sigma^2} \right\} \frac{\sqrt{2\pi}\sigma}{\sqrt{n+1}}. \quad (6.31)$$

Temos então que (II) pode ser escrito como

$$(II) = \int_{-\infty}^{+\infty} |\beta_2| \exp \left\{ -\frac{\beta_2^2 \sum_{i=0}^n x_i^2}{2\sigma^2} \right\} \frac{\sqrt{2\pi}\sigma}{\sqrt{n+1}} \exp \left\{ \frac{\beta_2^2 (\sum_{i=0}^n x_i)^2}{2(n+1)\sigma^2} \right\} d\beta_2. \quad (6.32)$$

Após algumas manipulações algébricas podemos reescrevê-lo como

$$(II) = \frac{\sqrt{2\pi}\sigma}{\sqrt{n+1}} \underbrace{\int_{-\infty}^{+\infty} |\beta_2| \exp \left\{ -\beta_2^2/2v^2 \right\} d\beta_2}_{(IV)}, \quad (6.33)$$

onde $v^2 = \frac{(n+1)\sigma^2}{(n+1)\sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2}$. Observe que

$$(IV) = \int_{-\infty}^0 -\beta_2 \exp \left\{ -\beta_2^2/2v^2 \right\} d\beta_2 + \int_0^{+\infty} \beta_2 \exp \left\{ -\beta_2^2/2v^2 \right\} d\beta_2 \quad (6.34)$$

$$= 2 \int_0^{+\infty} \beta_2 \exp \left\{ -\beta_2^2/2v^2 \right\} d\beta_2. \quad (6.35)$$

Fazendo a mudança de variáveis

$$\begin{aligned}\frac{\beta_2}{v} = s &\Rightarrow \frac{1}{v}d\beta_2 = ds \Rightarrow d\beta_2 = vds \\ \beta_2 = \infty &\Rightarrow s = \infty \\ \beta_2 = 0 &\Rightarrow s = 0\end{aligned}$$

temos

$$= 2v^2 \int_0^{+\infty} s \exp\{-s^2/2\} ds. \quad (6.36)$$

Fazendo uma nova mudança de variáveis

$$\begin{aligned}s^2 = u &\Rightarrow 2sds = du \\ s = \infty &\Rightarrow u = \infty \\ s = 0 &\Rightarrow u = 0\end{aligned}$$

temos

$$= v^2 \int_0^{+\infty} \exp\{-u/2\} du = v^2(-2 \exp\{-u/2\})\Big|_0^\infty = 2v^2 \quad (6.37)$$

isto é,

$$(IV) = 2v^2 = 2 \frac{(n+1)\sigma^2}{(n+1) \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2}. \quad (6.38)$$

Deste modo, temos que (II) é dado por

$$(II) = \frac{\sqrt{2\pi}\sigma}{\sqrt{n+1}} \frac{2(n+1)\sigma^2}{(n+1) \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2} \quad (6.39)$$

e

$$\pi(x_0) = \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{2\pi\sqrt{2\pi}\sigma^3} \frac{\sqrt{2\pi}\sigma}{\sqrt{n+1}} \frac{2(n+1)\sigma^2}{(n+1) \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2} \quad (6.40)$$

$$= \frac{|\mathbf{X}'\mathbf{X}|^{1/2}}{\pi} \frac{\sqrt{n+1}}{(n+1) \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2} \quad (6.41)$$

Agora observe que podemos reescrever

$$(n+1) \sum_{i=0}^n x_i^2 - (\sum_{i=0}^n x_i)^2 = (n+1) \sum_{i=1}^n x_i^2 + (n+1)x_0^2 - (\sum_{i=1}^n x_i + x_0)^2 \quad (6.42)$$

$$= n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 + \sum_{i=1}^n x_i^2 + nx_0^2 - 2x_0 \sum_{i=1}^n x_i, \quad (6.43)$$

e como $nc_x = n \sum_{i=1}^n (x_i - \bar{x})^2 = n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2$ temos

$$= nc_x + \sum_{i=1}^n x_i^2 + nx_0^2 - 2x_0n\bar{x}, \quad (6.44)$$

somando e subtraindo $n\bar{x}^2$

$$= nc_x + \sum_{i=1}^n x_i^2 + n(x_0 - \bar{x})^2 - n\bar{x}^2 \quad (6.45)$$

$$= \frac{n^2 c_x + n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n} + n(x_0 - \bar{x})^2 \quad (6.46)$$

$$= (n+1)c_x + n(x_0 - \bar{x})^2 \quad (6.47)$$

$$= (n+1)c_x \left(1 + \frac{n(x_0 - \bar{x})^2}{(n+1)c_x} \right). \quad (6.48)$$

Portanto,

$$\pi(x_0) = \frac{(nc_x)^{1/2}}{\pi} \frac{\sqrt{n+1}}{(n+1)c_x \left(1 + \frac{n(x_0 - \bar{x})^2}{(n+1)c_x} \right)} \quad (6.49)$$

$$= \frac{1}{\pi \sqrt{(n+1)c_x/n}} \left(1 + \frac{(x_0 - \bar{x})^2}{(n+1)c_x/n} \right)^{-1}. \quad (6.50)$$

Deste modo, $x_0 \sim \text{Cauchy}(\bar{x}, \frac{(n+1)c_x}{n})$.

△

Comparando as distribuições marginais da priori de referência e da priori “usual” percebemos que, pelo menos para o β , a priori de referência é marginalmente muito mais informativa do que a priori marginal de β (uma uniforme na reta). Já, com respeito ao x_0 , a priori de referência marginal é uma Cauchy, e por este motivo, apresenta caudas mais pesadas que a priori “vaga” (que é uma normal) e portanto, para um mesmo parâmetro de escala, é menos precisa que a sua concorrente na literatura.

6.4 A estrutura normal latente do modelo *probit* de calibração binomial

Introduzindo a estrutura normal latente já apresentada no modelo de regressão binomial substituímos a função de verossimilhança original (equação 6.4) pela função de verossimi-

lança aumentada $L_{\mathbf{w}, \mathbf{y}, \mathbf{X}}(x_0, \beta)$ dada por

$$\prod_{i=0}^n \prod_{j=1}^{y_i} (I_{(w_{ij} > 0)} I_{(z_{ij}=1)} + I_{(w_{ij} \leq 0)} I_{(z_{ij}=0)}) \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(w_{ij} - \mathbf{x}_i' \beta)^2}{2} \right\}. \quad (6.51)$$

Assim, a distribuição a posteriori de referência é expressa como

$$\begin{aligned} \pi(x_0, \beta, \mathbf{w} \mid \mathbf{y}, \mathbf{X}) &\propto |\beta_2| \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta' \mathbf{X}' \mathbf{X} \beta}{2\sigma^2} \right\} \times \\ &\times \left(\exp \left\{ -\frac{1}{2} \sum_{i=0}^n \sum_{j=1}^{y_i} (w_{ij} - \mathbf{x}_i' \beta)^2 \right\} I_{(w_{ij} > 0)} + \exp \left\{ -\frac{1}{2} \sum_{i=0}^n \sum_{j=y_i+1}^{k_i} (w_{ij} - \mathbf{x}_i' \beta)^2 \right\} I_{(w_{ij} \leq 0)} \right). \end{aligned} \quad (6.52)$$

Na próxima seção apresentaremos um esquema computacional para a geração de amostras desta distribuição a posteriori.

6.5 Distribuições condicionais completas para o amostrador de Gibbs

Através da inspeção da distribuição a posteriori é fácil perceber que, devido ao termo $|\beta_2|$, a distribuição condicional de β não possui uma forma conhecida. No entanto, as demais distribuições condicionais possuem formas conhecidas e podem ser geradas diretamente. Nesta situação, é recomendável o uso de um algoritmo híbrido para a geração da distribuição a posteriori. Muller (1991) recomenda que a distribuição condicional com forma desconhecida seja gerada através de um passo de Metropolis-Hastings dentro de um ciclo do amostrador de Gibbs. Este algoritmo híbrido é conhecido como Metropolis dentro do amostrador de Gibbs.

Fazendo uso deste algoritmo híbrido, temos que as distribuições condicionais completas são dadas por

- $x_0 \mid \beta, \mathbf{y}, \mathbf{w}, \mathbf{X} \sim N \left(\frac{\sigma^2 k_0 \bar{w}_0 - (1 + \sigma^2 k_0) \beta_1}{(1 + \sigma^2 k_0) \beta_2}, \frac{\sigma^2}{(1 + \sigma^2 k_0) \beta_2^2} \right)$
- utilizamos um passo do algoritmo de Metropolis-Hastings para gerar de

$$p(\beta \mid x_0, \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto |\beta_2| \exp \left\{ -\frac{(\beta - \mathbf{m})' \Sigma^{-1} (\beta - \mathbf{m})}{2\sigma^2} \right\}$$

- $w_{ij} \mid x_0, \beta, \mathbf{y}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \beta, 1) & \text{truncada à esquerda do zero, se } z_{ij} = 1 \\ N(\mathbf{x}_i' \beta, 1) & \text{truncada à direita do zero, se } z_{ij} = 0 \end{cases}$
e pode ser gerada através de $w_{ij} = \mathbf{x}_i' \beta + \Phi^{-1}(z_{ij} \Phi(-\mathbf{x}_i' \beta) + u_i(z_{ij} + (1 - 2z_{ij}) \Phi(-\mathbf{x}_i' \beta)))$

A derivação destas distribuições condicionais completas está apresentada no apêndice C.

6.6 Estudo de simulação

Nesta seção apresentamos um estudo de simulação comparando a priori de referência com outra priori não-subjetiva usualmente empregada na literatura. Trata-se de um estudo preliminar, cujo objetivo é apresentar algumas impressões a respeito do desempenho inferencial da priori de referência.

Por simplicidade, implementamos as simulações no modelo de calibração binomial com intercepto zero. Neste caso, a priori de referência é dada por

$$\pi(x_0, \beta) \approx \frac{\sqrt{\sum_{i=1}^n x_i^2}}{2\pi\sigma^2} |\beta| \exp \left\{ -\frac{\beta^2(x_0^2 + \sum_{i=1}^n x_i^2)}{2\sigma^2} \right\}. \quad (6.53)$$

Marginalmente,

$$x_0 \sim \text{Cauchy} \left(0, \sum_{i=1}^n x_i^2 \right), \quad (6.54)$$

$$\beta \sim N \left(0, \sigma^2 / \sum_{i=1}^n x_i^2 \right). \quad (6.55)$$

As distribuições condicionais completas para a implementação do amostrador de Gibbs são expressas como

- $x_0 \mid \beta, w_{ij}, \mathbf{y}, \mathbf{X} \sim N(m_0, v_0^2),$
- $p(\beta \mid x_0, w_{ij}, \mathbf{y}, \mathbf{X}) \propto |\beta| \exp \left\{ -\frac{(\beta - m)^2}{2v^2} \right\},$
- $w_{ij} = \beta x_i + \Phi^{-1}(z_{ij} \Phi(-\beta x_i) + u_i(z_{ij} + (1 - 2z_{ij}) \Phi(-\beta x_i))),$

onde

$$\begin{aligned} m_0 &= \frac{\sigma^2 k_0 \bar{w}_0}{(1 + \sigma^2 k_0) \beta}, & v_0^2 &= \frac{\sigma^2}{(1 + \sigma^2 k_0) \beta^2}, \\ m &= \frac{\sigma^2 \sum_{i=0}^n k_i x_i \bar{w}_i}{\sum_{i=0}^n (1 + \sigma^2 k_i) x_i^2}, & v^2 &= \frac{\sigma^2}{\sum_{i=0}^n (1 + \sigma^2 k_i) x_i^2}. \end{aligned}$$

Como a distribuição condicional $p(\beta \mid x_0, w_{ij}, \mathbf{y}, \mathbf{X})$ não possui uma forma conhecida, utilizamos um passo do algoritmo de Metropolis-Hastings dentro de cada ciclo do amostrador de Gibbs para a geração desta distribuição alvo. Escolhemos, como distribuição proposta, $q(\cdot \mid \cdot) = N(m, v^2)$ e deste modo a probabilidade de aceitação do algoritmo é dada por

$$\alpha(a, b) = \min \left\{ 1, \frac{f(b)q(a \mid b)}{f(a)q(b \mid a)} \right\} = \min \left\{ 1, \frac{|b|}{|a|} \right\}, \quad (6.56)$$

uma vez que

$$\frac{f(b)q(a \mid b)}{f(a)q(b \mid a)} = \frac{c_1 |b| \exp \left\{ -\frac{(b-m)^2}{2v^2} \right\} c_2 \exp \left\{ -\frac{(a-m)^2}{2v^2} \right\}}{c_1 |a| \exp \left\{ -\frac{(a-m)^2}{2v^2} \right\} c_2 \exp \left\{ -\frac{(b-m)^2}{2v^2} \right\}} = \frac{|b|}{|a|}. \quad (6.57)$$

A expressão $\alpha(a, b)$ representa a probabilidade do algoritmo aceitar o novo valor b (gerado a partir da distribuição proposta) como sendo distribuído segundo a distribuição alvo. O a representa o último valor aceito pelo algoritmo. A figura 6.1 mostra claramente que a distribuição proposta utilizada neste estudo é capaz de gerar eficientemente amostras da distribuição alvo $p(\beta \mid x_0, w_{ij}, \mathbf{y}, \mathbf{X})$. Nesta implementação utilizamos $m = 1$, $v^2 = 2$ e a taxa de aceitação do algoritmo foi de 0.593.

A priori não-subjetiva usualmente empregada na literatura é dada por

$$\pi_v(x_0, \beta) = \pi_v(\beta) \pi_v(x_0) = 1 \times N(\bar{x}, s^2), \quad (6.58)$$

onde s^2 é relativamente grande. Seguindo Branco *et al* (2003), adotamos neste trabalho, $s^2 = 10$. As distribuições condicionais completas associadas a distribuição a posteriori “usual” possuem formas conhecidas e são expressas por

- $x_0 \mid \beta, w_{ij}, \mathbf{y}, \mathbf{X} \sim N(m_{*0}, v_{*0}^2),$
- $\beta \mid x_0, w_{ij}, \mathbf{y}, \mathbf{X} \sim N(m_*, v_*^2),$
- $w_{ij} = \beta x_i + \Phi^{-1}(z_{ij} \Phi(-\beta x_i) + u_i(z_{ij} + (1 - 2z_{ij}) \Phi(-\beta x_i))),$

onde

$$m_{*0} = \frac{\bar{x} + 10k_0 \bar{w}_0 \beta}{1 + 10k_0 \beta^2}, \quad v_{*0}^2 = \frac{10}{1 + 10k_0 \beta^2},$$

$$m_* = \frac{\sum_{i=0}^n k_i x_i \bar{w}_i}{\sum_{i=0}^n k_i x_i^2}, \quad v_*^2 = \frac{1}{\sum_{i=0}^n k_i x_i^2}.$$

Figura 6.2: Geração da distribuição condicional de β através do algoritmo de Metrópolis-Hastings.

No presente estudo de simulação, avaliamos os erros empíricos quadráticos e absolutos associados aos parâmetros x_0 e β . O erro quadrático empírico é definido como

$$Eqm(\theta) = \frac{\sum_{i=1}^n (\bar{\theta} - \theta)^2}{n}, \quad (6.59)$$

onde $\bar{\theta}$ é uma estimativa da média de uma amostra da distribuição a posteriori de θ . Analogamente, o erro absoluto empírico é definido por

$$Eam(\theta) = \frac{\sum_{i=1}^n |\tilde{\theta} - \theta|}{n}, \quad (6.60)$$

onde $\tilde{\theta}$ representa uma estimativa da mediana de uma amostra da distribuição a posteriori de θ . Observe que nas definições do erro quadrático e do erro absoluto utilizamos, respectivamente, a média e a mediana como estimadores de θ , uma vez que a média é o estimador ótimo (estimador de Bayes) em relação a função de perda quadrática, enquanto a mediana é o estimador ótimo em relação a função de perda absoluta.

Na avaliação do desempenho destas duas propostas, consideramos $k_i = k$ para todo i e utilizamos $k = 10$ e $k = 50$. Também consideramos diferentes valores para as covariáveis ($\mathbf{x}$) e avaliamos casos onde o valor de x_0 estava contido dentro e fora da amplitude de variação das covariáveis. Ao todo, analisamos 16 diferentes situações, a saber:

1. $k = 10$, $\mathbf{x}_1 = (-0.131, -0.063, 0.000, 0.063, 0.131)$ e $x_0 = 0.031$,
2. $k = 10$, $\mathbf{x}_1 = (-0.131, -0.063, 0.000, 0.063, 0.131)$ e $x_0 = 0.169$,
3. $k = 10$, $\mathbf{x}_2 = (-0.263, -0.127, 0.000, 0.127, 0.263)$ e $x_0 = 0.063$,
4. $k = 10$, $\mathbf{x}_2 = (-0.263, -0.127, 0.000, 0.127, 0.263)$ e $x_0 = 0.337$,
5. $k = 10$, $\mathbf{x}_3 = (-0.063, 0.000, 0.063, 0.131, 0.210)$ e $x_0 = 0.031$,
6. $k = 10$, $\mathbf{x}_3 = (-0.063, 0.000, 0.063, 0.131, 0.210)$ e $x_0 = -0.096$,
7. $k = 10$, $\mathbf{x}_4 = (-0.127, 0.000, 0.127, 0.263, 0.421)$ e $x_0 = 0.063$,
8. $k = 10$, $\mathbf{x}_4 = (-0.127, 0.000, 0.127, 0.263, 0.421)$ e $x_0 = -0.193$.

Nas demais 8 situações apenas trocamos $k = 10$ por $k = 50$.

Em cada uma das 16 situações estudadas, avaliamos 100 repetições, isto é, obtivemos 100 amostras diferentes da distribuição a posteriori de referência e da distribuição a posteriori “usual”. Em cada uma das repetições foram efetuadas 5500 iterações do amostrador de Gibbs, das quais as primeiras 5000 (burn-in) foram descartadas. Para cada uma das amostras (de tamanho 500), calculamos o erro quadrático e o erro absoluto médio, o mínimo, o máximo, a mediana, a média e a amplitude do intervalo de credibilidade com 0.9 de probabilidade. O diagnóstico de convergência das cadeias de Markov foi feito através dos gráficos de médias ergódicas e também através dos gráficos da própria cadeia. Toda a programação foi efetuada no *S-Plus 2000*.

Os resultados deste estudo estão apresentados em cinco tabelas. As Tabelas 6.1, 6.3, 6.4 e 6.5 apresentam os erros empíricos quadráticos e absolutos associados aos parâmetros x_0 e β , nos casos em que $k = 10$ e 50.

Considerando os erros empíricos absolutos da Tabela 6.1, percebemos que em todos os casos a priori de referência apresenta um desempenho melhor do que a priori “usual”. Isto ocorre tanto quando o x_0 original está dentro da amplitude de variação das covariáveis, quanto quando o x_0 original está fora da amplitude de variação das covariáveis. No entanto, com relação aos erros quadráticos os resultados se invertem.

Tabela 6.1: Erros quadráticos e absolutos médios para o x_0 ($k = 10$)

$k = 10$	escala	covariáveis	erros	dentro		fora	
				referência	usual	referência	usual
$\bar{x} = 0$	0.042	$\mathbf{x}_1$	E_{qm}	0.443	0.340	166.152	0.730
			E_{am}	0.092	0.176	0.128	0.302
	0.170	$\mathbf{x}_2$	E_{qm}	1.570	0.091	0.840	0.736
			E_{am}	0.158	0.194	0.258	0.364
$\bar{x} \neq 0$	0.069	$\mathbf{x}_3$	E_{qm}	0.086	0.046	380.013	0.082
			E_{am}	0.089	0.115	0.085	0.123
	0.278	$\mathbf{x}_4$	E_{qm}	1.172	0.197	1.001	0.163
			E_{am}	0.194	0.234	0.237	0.263

Na Tabela 6.2 apresentamos as médias (em relação as 100 amostras simuladas) de algumas estatísticas descritivas obtidas para as distribuições a posteriori de x_0 quando $k = 10$. Considerando esta tabela, percebemos que a distribuição a posteriori de referência parece herdar as caudas pesadas da priori de referência (lembre que marginalmente x_0 tinha distribuição Cauchy) e acaba produzindo máximos e mínimos muito maiores que os da distribuição a posteriori “usual”. Agora, lembrando que a média é uma estimativa menos robusta que a mediana em relação a “outliers”, percebemos que este parece ser o motivo pelo qual o erro quadrático (calculado com base em estimativas da média) é maior que o erro absoluto (calculado com base em estimativas da mediana).

Analisando ainda a Tabela 6.2, notamos que (exceto em dois casos) a amplitude de variação do intervalo de probabilidade de 0.90 foi sempre menor na distribuição a posteriori de referência marginal de x_0 , que na distribuição a posteriori marginal “usual”. Este fato mostra que apesar das caudas mais pesadas, a distribuição marginal a posteriori de referência parece ser um pouco mais concentrada em torno da sua moda, que a sua concorrente. Os dois casos onde a amplitude de variação do intervalo de probabilidade foram maiores para a posteriori marginal de referência ocorreram devido a amostras que geraram “outliers”

Tabela 6.2: Médias das estatísticas descritivas de 100 amostras de x_0 ($k = 10$)

x_0	$k = 10$		mínimo	mediana	média	amplitude	máximo
0.031	$\mathbf{x}_1$	referência	-28.591	0.034	-0.066	1.815	18.817
	dentro	usual	-5.646	0.107	0.111	3.111	5.907
0.199	$\mathbf{x}_1$	referência	-240.573	0.122	0.308	8.362	291.992
	fora	usual	-4.612	0.334	0.376	2.896	5.592
-0.096	$\mathbf{x}_2$	referência	-53.416	-0.067	0.035	2.569	104.228
	dentro	usual	-6.061	-0.077	-0.081	3.547	5.637
0.337	$\mathbf{x}_2$	referência	-14.394	0.358	0.435	3.385	14.612
	fora	usual	-5.227	0.493	0.511	3.655	6.286
0.031	$\mathbf{x}_3$	referência	-12.219	0.035	0.073	0.973	11.453
	dentro	usual	-4.621	0.048	0.070	2.050	4.906
-0.096	$\mathbf{x}_3$	referência	-225.544	-0.108	2.083	15.899	376.572
	fora	usual	-5.052	-0.151	-0.191	2.274	4.444
0.063	$\mathbf{x}_4$	referência	-22.731	0.073	0.119	2.030	20.667
	dentro	usual	-4.156	0.102	0.147	2.334	4.570
-0.193	$\mathbf{x}_4$	referência	-19.108	-0.206	-0.331	2.898	14.095
	fora	usual	-5.261	-0.283	-0.301	2.668	4.496

realmente extremos. Abaixo apresentamos as estatísticas descritivas associadas a duas destas amostras.

x_0 original	mínimo	mediana	média	amplitude	máximo
0.169	-7.871	0.275	111.364	375.279	5501.181
-0.096	-20993.596	-0.153	192.816	1425.400	34128.750

Estes dados, mostram ainda o quanto a média é sensível a “outliers”, em comparação com a grande robustez da mediana.

Um fato interessante a respeito da priori de referência é a sua dependência com respeito aos valores assumidos pelas covariáveis. De fato, a distribuição a priori marginal de x_0 é aproximadamente uma Cauchy centrada em zero e com parâmetro de escala dado por $\sum_{i=1}^n x_i^2$. Analisando os dados da Tabela 6.1, percebemos que a medida que o parâmetro de escala² da

²Os valores dos parâmetros de escala da distribuição marginal de x_0 estão apresentados na Tabela 6.1.

distribuição a priori marginal de x_0 aumenta, os erros absolutos também aumentam. Este fato era esperado, uma vez que aumentando a dispersão da distribuição a posteriori, aumentamos automaticamente o erro associado a estimativa (pontual) da mediana. Em relação ao erro quadrático, este padrão não foi mantido, devido a grande sensibilidade das estimativas de médias com respeito a “outliers”.

Com respeito a comparação dos erros em relação aos casos “dentro” e “fora”, percebemos que parece existir uma tendência para menores erros no caso em que o valor do x_0 original está dentro da amplitude de variação das covariáveis. De fato, para a distribuição a posteriori usual, este padrão sempre foi respeitado, tanto para os erros quadráticos, quanto para os erros absolutos. Contudo, em relação a distribuição a posteriori de referência, mesmo quando consideramos os erros absolutos, encontramos uma excessão a este padrão. Estes resultados parecem mostrar que apesar da tendência a maiores erros no caso “fora” também seja vista na distribuição a posteriori de referência, esta última parece ser mais robusta que a distribuição a posteriori “usual”.

Em relação as covariáveis estarem centradas ou não em zero, não notamos diferenças entre as diferentes propostas. Este fato é curioso, pois esperavamos que os erros da distribuição a posteriori de referência aumentassem quando $\bar{x} \neq 0$, uma vez que a priori de referência marginal para x_0 está centrada no zero.

Tabela 6.3: Erros quadráticos e absolutos médios para o x_0 ($k = 50$)

$k = 50$	escala	covariáveis	erros	dentro		fora	
				referência	usual	referência	usual
$\bar{x} = 0$	0.042	$\mathbf{x}_1$	E_{qm}	0.004	0.006	0.004	0.013
			E_{am}	0.043	0.046	0.040	0.047
	0.170	$\mathbf{x}_2$	E_{qm}	0.010	0.016	0.043	0.069
			E_{am}	0.070	0.077	0.124	0.143
$\bar{x} \neq 0$	0.069	$\mathbf{x}_3$	E_{qm}	0.002	0.002	0.004	0.005
			E_{am}	0.037	0.038	0.041	0.044
	0.278	$\mathbf{x}_4$	E_{qm}	0.010	0.011	0.016	0.021
			E_{am}	0.072	0.075	0.086	0.093

Observando a Tabela 6.3, percebemos que em todos os casos os erros associados a distribuição a posteriori de referência foram sempre ligeiramente menores que os erros associados a distribuição a posteriori usual. Em relação ao aumento do parâmetro de escala, ao com-

portamento “dentro $\times$ fora” e ao comportamento “ $\bar{x} = 0 \times \bar{x} \neq 0$ ”, todos os resultados foram análogos aos obtidos para $k = 10$.

Comparando os casos $k = 10$ e $k = 50$ percebemos que os erros foram muito menores para $k = 50$. Este fato indica que para k grande as estimativas pontuais da média e da mediana tornam-se mais acuradas, como esperado. Percebemos, ainda, através da análise da Tabela 6.3, que para valores grandes de k , as diferenças entre as distribuições a posteriori de referência e “usual” (em relação aos erros) foram muito pequenas. Isso indica que a função de verossimilhança já começa a dominar a priori para $k = 50$.

Analisando a Tabela 6.4, notamos que em todos os casos, tanto o erro quadrático médio quanto o erro absoluto médio, foram sempre menores para a distribuição a posteriori de referência. Percebemos também, que em todos os casos, o erro quadrático foi sempre maior que o erro absoluto. Este fato parece indicar que a mediana é um estimador pontual melhor do que a média para o β também.

Tabela 6.4: Erros quadráticos e absolutos médios para o β ($k = 10$)

$k = 10$	escala	covariáveis	erros	dentro		fora	
				referência	usual	referência	usual
$\bar{x} = 0$	58.197	$\mathbf{x}_1$	E_{qm}	4.142	6.228	3.338	5.441
			E_{am}	1.513	2.338	1.386	2.058
	14.549	$\mathbf{x}_2$	E_{qm}	1.003	1.433	0.761	1.120
			E_{am}	0.829	1.156	0.684	0.965
$\bar{x} \neq 0$	35.512	$\mathbf{x}_3$	E_{qm}	2.420	3.932	3.169	5.021
			E_{am}	1.178	1.685	1.378	1.942
	8.878	$\mathbf{x}_4$	E_{qm}	0.716	1.052	0.720	1.036
			E_{am}	0.659	0.866	0.651	0.843

Assim como para o x_0 , a distribuição a priori marginal de β também depende dos valores assumidos pelas covariáveis. De fato, marginalmente, β tem uma distribuição aproximadamente normal, centrada em zero, e com variância igual a $\sigma^2 / \sum_{i=1}^n x_i^2$, onde $\sigma^2 = \pi^2/4$. Deste modo, vemos que a medida que aumentamos os valores das covariáveis, aumentamos também $\sum_{i=1}^n x_i^2$ e portanto, diminuimos o parâmetro de escala da distribuição marginal de β .³ Observando a Tabela 6.4, percebemos que a distribuição a posteriori de referência

³É interessante notar que a medida que aumentamos o valor de $\sum_{i=1}^n x_i^2$, diminuimos a dispersão da distribuição marginal de β , porém aumentamos a dispersão da distribuição marginal de x_0 .

marginal de β espelha esta propriedade da priori de referência, uma vez que a medida que diminuimos o valor do parâmetro de escala (a variância da distribuição de referência marginal para β), diminuimos também os erros associadas as estimativas pontuais de média e mediana de amostras da distribuição a posteriori marginal de referência de β .

Com relação ao comportamento “dentro $\times$ fora”, não parece haver nenhuma tendência bem definida. Já em relação ao comportamento “ $\bar{x} = 0 \times \bar{x} \neq 0$ ”, parece que os erros tendem a ser menores no caso $\bar{x} \neq 0$.

Analisando a Tabela 6.5 percebemos que para $k = 50$ temos resultados análogos aos obtidos para $k = 10$ (embora nem sempre os erros quadráticos tenham sido maiores que os erros absolutos). E assim como para o x_0 , em todos os casos, os erros obtidos para $k = 50$ foram sempre menores do que os erros para $k = 10$. Embora, neste caso, as diferenças entre as distribuições a posteriori de referência e “usual” não sejam tão pequenas quanto no caso do parâmetro x_0 , também aqui a função de verossimilhança parece começar a dominar a priori.

Tabela 6.5: Erros quadráticos e absolutos médios para o β ($k = 50$)

$k = 50$	escala	covariáveis	erros	dentro		fora	
				referência	usual	referência	usual
$\bar{x} = 0$	58.197	$\mathbf{x}_1$	E_{qm}	0.810	1.050	0.596	0.739
			E_{am}	0.724	0.807	0.591	0.639
	14.549	$\mathbf{x}_2$	E_{qm}	0.201	0.252	0.228	0.273
			E_{am}	0.361	0.403	0.394	0.432
$\bar{x} \neq 0$	35.512	$\mathbf{x}_3$	E_{qm}	0.487	0.533	0.617	0.654
			E_{am}	0.564	0.584	0.615	0.636
	8.878	$\mathbf{x}_4$	E_{qm}	0.137	0.152	0.141	0.165
			E_{am}	0.289	0.304	0.304	0.327

Agora, cabem alguns comentários a respeito do diagnóstico de convergência. Como mencionado anteriormente, analisamos os gráficos das médias ergódicas e os gráficos das cadeias Markov. Observamos que para $k = 50$ a convergência da cadeia é mais rápida e mais saudável (a cadeia se move mais livremente) que para $k = 10$. No entanto, mesmo nos casos em que $k = 10$, a cadeia parece convergir para a distribuição estacionária.

Todavia, em estudos de simulação complementares (não apresentados neste trabalho) encontramos casos onde a cadeia parecia não convergir para a distribuição estacionária.

Quando utilizamos valores de x_0 originais muito fora da amplitude de variação das covariáveis e $k = 10$, obtivemos convergências muito ruins para a distribuição a posteriori “usual”, a ponto de não confiarmos nas estimativas dos erros. Já a convergência da distribuição a posteriori de referência marginal de x_0 não pareceu ser muito afetada, de modo que as estimativas dos erros foram confiáveis.

Uma outra situação onde nos deparamos com problemas de convergência, ocorreu quando utilizamos $k = 10$ e valores grandes para as covariáveis. Neste caso, a convergência da distribuição a posteriori de referência foi ruim.

Parece natural que a convergência da distribuição a posteriori de referência piore a medida que aumentamos os valores das covariáveis uma vez que a dispersão da priori de referência marginal de x_0 aumenta também. Embora não tenha sido possível testar este ponto para valores pequenos de k , conduzimos um estudo de simulação tomando $k = 100$ e valores de covariáveis tais que $\sum_{i=1}^n x_i^2 = 10.035$. A convergência do algoritmo foi boa, e o que é mais importante, mesmo para $k = 100$ a verossimilhança não pareceu dominar a priori⁴ e deste modo foi possível fazer a comparação entre as distribuições a posteriori de referência e “usual”. Neste estudo de simulação, mesmo o erro absoluto médio de x_0 foi maior para a distribuição a posteriori de referência que para a distribuição a posteriori “usual”. Este resultado era esperado uma vez que, marginalmente, estamos utilizando uma priori de referência Cauchy de dispersão 10, enquanto a priori “vaga” é uma normal de variância 10. O fato da priori “vaga” ser mais concentrada em torno de sua mediana que a priori de referência marginal de x_0 , mostra que a escolha da variância igual a 10 fez com que a priori “vaga” fosse mais informativa que a priori de referência.

Observe que o fato da priori “usual” ter tido um desempenho melhor que a priori de referência não quer dizer que a priori “usual” seja melhor que a priori de referência. Isto simplesmente indica que a escolha de variância igual a 10 para a priori “vaga” é uma escolha informativa, apesar de a primeira vista parecer não-informativa.

Neste ponto, cabe um comentário a respeito das interpretações destes resultados de simulação. Devemos observar que todas as comparações analíticas que fizemos entre a priori “usual” e a priori de referência foram conduzidas com base nas distribuições a priori marginais de x_0 e β . Deste modo, todas as comparações analíticas que fizemos, servem somente para dar uma intuição a respeito do verdadeiro comportamento da priori conjunta.

Em termos gerais, podemos tirar as seguintes conclusões a respeito do presente estudo

⁴Este fato parece mostrar que os valores assumidos pelas covariáveis são importantes para determinar o tamanho da amostra onde a contribuição da priori torna-se negligível.

de simulação:

- Em relação ao parâmetro x_0 , a mediana é um estimador muito melhor que a média. Em relação ao β , a mediana ainda parece ser melhor que a média, embora a diferença entre os dois estimadores não seja muito grande.
- Quando as covariáveis assumem valores pequenos, o desempenho da priori de referência é melhor do que o da priori “usual”, e nestes casos valores de k da ordem de 50 parecem ser suficientes para que a verossimilhança domine a priori.
- Quando as covariáveis assumem valores grandes, o desempenho da priori “usual” parece ser melhor que o da priori de referência, e parece não haver dominação da priori mesmo para valores de $k = 100$. No entanto, são necessários mais estudos para sustentar melhor estas conclusões.
- Para valores pequenos de k , ainda precisamos resolver alguns problemas computacionais, de modo a garantir a convergência do algoritmo em uma maior gama de situações. Já em relação a valores grandes de k , o algoritmo utilizado no presente estudo é eficiente.

Capítulo 7

Considerações finais e perspectivas de trabalhos futuros

No presente trabalho consideramos a metodologia de B&B para a construção de prioris de referência e sua aplicação a alguns modelos estatísticos. Ao contrário do método de Jeffreys, no método de B&B é preciso distinguir os parâmetros de interesse dos parâmetros de perturbação. Esta característica torna a aplicação do método de B&B particularmente interessante para os modelos de calibração linear e calibração binomial, já que nestes modelos existe uma clara distinção entre o parâmetro de interesse e os parâmetros de perturbação.

De um ponto de vista didático, os quatro modelos estatísticos abordados neste trabalho foram bastante instrutivos pois ilustraram diversas características associadas às prioris de referência. Podemos citar, por exemplo, que:

- obtivemos prioris de referência impróprias (calibração linear) e próprias (regressão binária, regressão binomial e calibração binomial);
- obtivemos prioris de referência “não-informativas” (calibração linear) e “informativas” (regressão binária e regressão binomial);
- conduzimos a construção da priori de referência diretamente a partir do modelo estatístico original (calibração linear) e também, indiretamente, a partir de um modelo reparametrizado (regressão binária, regressão binomial e calibração binomial).

Entre as questões que ainda necessitam de maiores esclarecimentos podemos destacar que:

- De um ponto de vista teórico, uma questão que ainda permanece em aberto no modelo de calibração linear, é a existência dos momentos na distribuição a posteriori obtida a partir da priori de Jeffreys. No presente trabalho, provamos apenas que a distribuição a posteriori de Jeffreys é própria e que, no caso particular em que $p = 1$, esta distribuição não possui nenhum momento finito. Todavia, para um p maior do que 1 não sabemos se a distribuição a posteriori de Jeffreys não possui nenhum momento finito ou se ela apresenta um comportamento análogo ao da distribuição a posteriori de referência, que possui momentos finitos até a ordem $p - 1$. Uma resposta a esta questão seria de considerável interesse teórico uma vez que a utilização da priori de referência seria certamente preferível à priori de Jeffreys, caso esta última realmente não possuía nenhum momento finito para qualquer valor de p .
- Caso não seja possível provar que a distribuição a posteriori de Jeffreys não possui nenhum momento finito, pretendemos proceder um estudo de simulação comparando os desempenhos da priori de referência e da priori de Jeffreys.
- Em relação ao modelo de calibração binomial, pretendemos resolver os problemas de convergência que o nosso atual algoritmo tem enfrentado quando k é pequeno e, deste modo, proceder um estudo de simulação mais extenso, avaliando o desempenho inferencial da priori de referência de maneira mais segura.

Apêndice A

Amostrador de Gibbs: modelo de calibração linear

Considere primeiramente a distribuição a posteriori

$$\begin{aligned} \pi(x_0, \alpha, \beta, \sigma^2 \mid \mathbf{y}, \mathbf{X}) &\propto \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} (\sigma^{-2})^{\frac{(n+k)p+2}{2}} \times \\ &\times \exp \left\{ - \left[\sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' \times \right. \right. \\ &\left. \left. \times (\mathbf{y}_{0j} - \alpha - \beta x_0) \right] / 2\sigma^2 \right\}, \end{aligned} \quad (\text{A.1})$$

obtida a partir da priori de referência construída através da abordagem com 4-grupos.

A distribuição condicional para σ^{-2} é dada por

$$p(\sigma^{-2} \mid x_0, \beta, \mathbf{y}, \mathbf{X}) \propto (\sigma^{-2})^{\frac{(n+k)p+4}{2}-1} \exp \left\{ -\frac{l_1}{2} \sigma^{-2} \right\}, \quad (\text{A.2})$$

onde

$$l_1 = \sum_{i=1}^n (\mathbf{y}_i - \alpha - \beta x_i)' (\mathbf{y}_i - \alpha - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)' (\mathbf{y}_{0j} - \alpha - \beta x_0).$$

Portanto,

$$\sigma^{-2} \mid x_0, \alpha, \beta, \mathbf{y}, \mathbf{X} \sim \text{Gama} \left(\frac{(n+k)p+4}{2}, \frac{l_1}{2} \right). \quad (\text{A.3})$$

A distribuição condicional para α é dada por

$$p(\alpha \mid x_0, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{l_1}{2\sigma^2} \right\}, \quad (\text{A.4})$$

que após alguma álgebra pode ser reescrita como

$$p(\alpha \mid x_0, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{(\alpha - \mathbf{l}_2)'(\alpha - \mathbf{l}_2)}{2\sigma^2/(n+k)} \right\}, \quad (\text{A.5})$$

onde

$$\mathbf{l}_2 = \frac{\sum_{i=1}^n (\mathbf{y}_i - \beta x_i) + \sum_{j=1}^k (\mathbf{y}_{0j} - \beta x_0)}{n+k}.$$

Deste modo,

$$\alpha \mid x_0, \beta, \sigma^2, \mathbf{y}, \mathbf{X} \sim N_p \left(\mathbf{l}_2, \frac{\sigma^2}{n+k} \mathbf{I}_p \right). \quad (\text{A.6})$$

A distribuição condicional para β é dada por

$$p(\beta \mid x_0, \alpha, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{l_1}{2\sigma^2} \right\}, \quad (\text{A.7})$$

que após alguma álgebra pode ser reescrita como

$$p(\beta \mid x_0, \alpha, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{(\beta - \mathbf{l}_3)'(\beta - \mathbf{l}_3)}{2\sigma^2/(kx_0^2 + \sum_{i=1}^n x_i^2)} \right\}, \quad (\text{A.8})$$

onde

$$\mathbf{l}_3 = \frac{\sum_{i=1}^n x_i (\mathbf{y}_i - \alpha) + x_0 \sum_{j=1}^k (\mathbf{y}_{0j} - \alpha)}{kx_0^2 + \sum_{i=1}^n x_i^2}.$$

Deste modo,

$$\beta \mid x_0, \alpha, \sigma^2, \mathbf{y}, \mathbf{X} \sim N_p \left(\mathbf{l}_3, \frac{\sigma^2}{kx_0^2 + \sum_{i=1}^n x_i^2} \mathbf{I}_p \right). \quad (\text{A.9})$$

A distribuição condicional de x_0 , $p(x_0 \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X})$, é proporcional a

$$\left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2} \exp \left\{ -\frac{\sum_{j=1}^k (\mathbf{y}_{0j} - \alpha - \beta x_0)'(\mathbf{y}_{0j} - \alpha - \beta x_0)}{2\sigma^2} \right\}, \quad (\text{A.10})$$

que por sua vez, é proporcional a

$$\left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right)^{-1/2} \exp \left\{ -\frac{(x_0 - l_4)^2}{2\sigma^2/k\beta'\beta} \right\}, \quad (\text{A.11})$$

onde

$$l_4 = \frac{\beta' \sum_{j=1}^k (y_{0j} - \alpha)}{k\beta'\beta}.$$

Observe que (A.11) não possui uma forma conhecida. Deste modo, seguindo Kubokawa e Robert (1994), vamos introduzir uma variável latente em (A.11). Primeiramente note que

$$\begin{aligned} \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} &\propto \Gamma(1/2) \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right)^{-1/2} = \\ &= \int_0^\infty t^{\frac{1}{2}-1} \exp \left\{ - \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right) t \right\} dt. \end{aligned} \quad (\text{A.12})$$

Deste modo, podemos reescrever (A.11) como

$$p(x_0 \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \int_0^\infty t^{\frac{1}{2}-1} \exp \left\{ - \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right) t \right\} \exp \left\{ - \frac{(x_0 - l_4)^2}{2\sigma^2/k\beta'\beta} \right\} dt. \quad (\text{A.13})$$

E como

$$p(x_0 \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) = \int_0^\infty p(x_0, t \mid \beta, \sigma^2, \mathbf{y}, \mathbf{X}) dt, \quad (\text{A.14})$$

temos que

$$p(x_0, t \mid \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto t^{\frac{1}{2}-1} \exp \left\{ - \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right) t \right\} \exp \left\{ - \frac{(x_0 - l_4)^2}{2\sigma^2/k\beta'\beta} \right\}. \quad (\text{A.15})$$

Agora, note que podemos utilizar novamente o amostrador de Gibbs para gerar x_0 a partir desta distribuição condicional. Observe que

$$p(t \mid x_0, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto t^{\frac{1}{2}-1} \exp \left\{ - \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right) t \right\}. \quad (\text{A.16})$$

Portanto,

$$t \mid x_0, \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X} \sim \text{Gama} \left(\frac{1}{2}, 1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x} \right). \quad (\text{A.17})$$

Observe também que

$$p(x_0 \mid t, \alpha, \beta, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ - \left(1 + \frac{nk(x_0 - \bar{x})^2}{(n+k)c_x}\right) t \right\} \exp \left\{ - \frac{(x_0 - l_4)^2}{2\sigma^2/k\beta'\beta} \right\}, \quad (\text{A.18})$$

que por sua vez, após algumas manipulações algébricas, pode ser reescrito como

$$p(x_0 | t, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{(x_0 - m)^2}{2v^2} \right\}, \quad (\text{A.19})$$

onde

$$m = \frac{2\sigma^2 n k t \bar{x} + c_x(n+k)\boldsymbol{\beta}' \sum_{j=1}^k (\mathbf{y}_{0j} - \boldsymbol{\alpha})}{2\sigma^2 n k t + c_x(n+k)k\boldsymbol{\beta}'\boldsymbol{\beta}},$$

$$v^2 = \frac{\sigma^2(n+k)c_x}{2\sigma^2 n k t + c_x(n+k)k\boldsymbol{\beta}'\boldsymbol{\beta}}.$$

Deste modo

$$x_0 | t, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X} \propto N(m, v^2). \quad (\text{A.20})$$

Apêndice B

Amostrador de Gibbs: modelos de regressão binária e binomial

Neste apêndice apresentaremos as distribuições condicionais completas para a implementação do amostrador de Gibbs.

B.1 O modelo de regressão binária

Considere a distribuição a posteriori de referência para o modelo de regressão binária

$$\begin{aligned} \pi(\beta, \mathbf{w} \mid \mathbf{z}, \mathbf{X}) &\propto \exp \left\{ -\beta' \mathbf{X}' \mathbf{X} \beta / 2\sigma^2 \right\} \exp \left\{ -(\mathbf{w} - \mathbf{X}\beta)'(\mathbf{w} - \mathbf{X}\beta) / 2 \right\} \times \\ &\times \prod_{i=1}^n (I_{(w_i > 0)} I_{(z_i = 1)} + I_{(w_i \leq 0)} I_{(z_i = 0)}). \end{aligned} \quad (\text{B.1})$$

Vamos encontrar, primeiramente, a distribuição condicional de $\beta \mid \mathbf{z}, \mathbf{W}, \mathbf{X}$. Observe que

$$\begin{aligned} p(\beta \mid \mathbf{z}, \mathbf{w}, \mathbf{X}) &\propto \exp \left\{ -\beta' \mathbf{X}' \mathbf{X} \beta / 2\sigma^2 \right\} \exp \left\{ -(\mathbf{w} - \mathbf{X}\beta)'(\mathbf{w} - \mathbf{X}\beta) / 2 \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \left[(1 + \sigma^2) \beta' \mathbf{X}' \mathbf{X} \beta - 2\sigma^2 \beta' \mathbf{X}' \mathbf{w} + \sigma^2 \mathbf{w}' \mathbf{w} \right] \right\}, \end{aligned}$$

e como $\sigma^2 \mathbf{w}' \mathbf{w}$ não depende de β ,

$$\propto \exp \left\{ -\frac{1}{2} \left[\left(\frac{1 + \sigma^2}{\sigma^2} \right) \beta' \mathbf{X}' \mathbf{X} \beta - 2\beta' \mathbf{X}' \mathbf{w} \right] \right\}.$$

Denotando $\left(\frac{1+\sigma^2}{\sigma^2}\right) \mathbf{X}'\mathbf{X} = \Sigma^{-1}$ temos

$$\propto \exp \left\{ -\frac{1}{2} \underbrace{\left[\beta' \Sigma^{-1} \beta - 2\beta' \mathbf{X}' \mathbf{w} \right]}_{(I)} \right\}.$$

Agora, igualando (I) à forma quadrática $(\beta - \mu)' \Sigma^{-1} (\beta - \mu)$

$$(I) \propto (\beta - \mu)' \Sigma^{-1} (\beta - \mu) = \beta' \Sigma^{-1} \beta - 2\beta' \Sigma^{-1} \mu + \mu' \Sigma^{-1} \mu,$$

e desconsiderando o termo $\mu' \Sigma^{-1} \mu$ que não depende de β , temos que

$$(I) \propto \beta' \Sigma^{-1} \beta - 2\beta' \Sigma^{-1} \mu,$$

de modo que

$$\beta' \mathbf{X}' \mathbf{w} = \beta' \Sigma^{-1} \mu.$$

Multiplicando os dois lados da equação primeiramente por $(\beta\beta')^{-1}\beta$ e em seguida por Σ , temos que

$$\mu = \Sigma \mathbf{X}' \mathbf{w} \quad \text{isto é,} \quad \mu = \left(\frac{\sigma^2}{1 + \sigma^2} \right) (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{w}.$$

Assim, a distribuição condicional de β é

$$\beta \mid \mathbf{z}, \mathbf{w}, \mathbf{X} \sim N_n \left(\frac{\sigma^2}{1 + \sigma^2} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{w}, \frac{\sigma^2}{1 + \sigma^2} (\mathbf{X}' \mathbf{X})^{-1} \right), \quad (\text{B.2})$$

onde $\sigma^2 = \pi^2/4$.

Vamos agora, encontrar a distribuição condicional para w_i , para cada i fixo ($i = 1, \dots, n$). Observe que

$$p(w_i \mid \beta, \mathbf{z}, \mathbf{X}) \propto \{I_{(w_i > 0)} I_{(z_i = 1)} + I_{(w_i \leq 0)} I_{(z_i = 0)}\} \times \exp \left\{ -(w_i - \mathbf{x}_i' \beta)^2 / 2 \right\}, \quad (\text{B.3})$$

ou seja,

$$w_i \mid \beta, \mathbf{z}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \beta, 1) & \text{truncada à esquerda do zero, se } z_i = 1 \\ N(\mathbf{x}_i' \beta, 1) & \text{truncada à direita do zero, se } z_i = 0 \end{cases}.$$

Todavia, podemos facilmente gerar destas normais truncadas através do gerador aleatório obtido a seguir.

Em primeiro lugar, observe que $p(w_i \mid \beta, \mathbf{z}, \mathbf{X})$

$$\propto (\exp \left\{ -(w_i - \mathbf{x}_i' \beta)^2 / 2 \right\}) I_{(w_i > 0)} I_{(z_i = 1)} + (\exp \left\{ -(w_i - \mathbf{x}_i' \beta)^2 / 2 \right\}) I_{(w_i \leq 0)} I_{(z_i = 0)}, \quad (\text{B.4})$$

fazendo a mudança de variável

$$s_i = w_i - \mathbf{x}_i' \boldsymbol{\beta} \Rightarrow ds_i = dw_i$$

$$w_i = -\infty \Rightarrow s_i = -\infty$$

$$w_i = 0 \Rightarrow s_i = -\mathbf{x}_i' \boldsymbol{\beta}$$

$$w_i = +\infty \Rightarrow s_i = +\infty$$

temos

$$p(s_i | \boldsymbol{\beta}, \mathbf{z}, \mathbf{X}) \propto (\exp \{-s_i^2/2\}) I_{(s_i > -\mathbf{x}_i' \boldsymbol{\beta})} I_{(z_i=1)} + (\exp \{-s_i^2/2\}) I_{(s_i \leq -\mathbf{x}_i' \boldsymbol{\beta})} I_{(z_i=0)}, \quad (\text{B.5})$$

isto é,

$$s_i | \boldsymbol{\beta}, \mathbf{z}, \mathbf{X} \sim \begin{cases} N(0, 1) & \text{truncada à esquerda do zero, se } z_i = 1 \\ N(0, 1) & \text{truncada à direita do zero, se } z_i = 0 \end{cases}.$$

Utilizando a relação $X = F^{-1}(F(a) + U(F(b) - F(a)))$, onde U tem distribuição $Uniforme(0, 1)$ e $a < b$, $a \in \Re$, $b \in \Re$, podemos verificar facilmente que X tem distribuição F truncada a esquerda de a e a direita de b . Aplicando este resultado ao nosso caso, temos que

$$\begin{aligned} s_i &= \Phi^{-1}(\Phi(-\mathbf{x}_i' \boldsymbol{\beta}) + u_i(\Phi(+\infty) - \Phi(-\mathbf{x}_i' \boldsymbol{\beta}))) \\ &= \Phi^{-1}(\Phi(-\mathbf{x}_i' \boldsymbol{\beta}) + u_i(1 - \Phi(-\mathbf{x}_i' \boldsymbol{\beta}))) \quad \text{se } z_i = 1, \end{aligned}$$

e

$$\begin{aligned} s_i &= \Phi^{-1}(\Phi(-\infty) + u_i(\Phi(-\mathbf{x}_i' \boldsymbol{\beta}) - \Phi(-\infty))) \\ &= \Phi^{-1}(u_i \Phi(-\mathbf{x}_i' \boldsymbol{\beta})) \quad \text{se } z_i = 0. \end{aligned}$$

E como $w_i = \mathbf{x}_i' \boldsymbol{\beta} + s_i$ temos que

$$w_i = \begin{cases} \mathbf{x}_i' \boldsymbol{\beta} + \Phi^{-1}(\Phi(-\mathbf{x}_i' \boldsymbol{\beta}) + u_i(1 - \Phi(-\mathbf{x}_i' \boldsymbol{\beta}))) & \text{se } z_i = 1 \\ \mathbf{x}_i' \boldsymbol{\beta} + \Phi^{-1}(u_i \Phi(-\mathbf{x}_i' \boldsymbol{\beta})) & \text{se } z_i = 0 \end{cases}. \quad (\text{B.6})$$

que pode ser reescrito como

$$w_i = \mathbf{x}_i' \boldsymbol{\beta} + \Phi^{-1}(z_i \Phi(-\mathbf{x}_i' \boldsymbol{\beta}) + u_i(z_i + (1 - 2z_i) \Phi(-\mathbf{x}_i' \boldsymbol{\beta}))).$$

B.2 O modelo de regressão binomial

Considere a distribuição a posteriori de referência para o modelo de regressão binomial

$$\begin{aligned} \pi(\boldsymbol{\beta}, \mathbf{w} \mid \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \right\} \times \left(\exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{y_i} (w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2 \right\} I_{(w_{ij} > 0)} + \right. \\ \left. + \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=y_i+1}^{k_i} (w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2 \right\} I_{(w_{ij} \leq 0)} \right). \end{aligned} \quad (\text{B.7})$$

A partir desta distribuição podemos facilmente obter as distribuições condicionais completas. Primeiramente, vamos encontrar a distribuição condicional de $\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X}$. Observe que

$$p(\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta} / 2\sigma^2 \right\} \exp \left\{ -\frac{1}{2} \underbrace{\sum_{i=1}^n \sum_{j=1}^{k_i} (w_{ij} - \mathbf{x}_i' \boldsymbol{\beta})^2}_I \right\},$$

$$(I) = \sum_{j=1}^{k_i} [w_{ij}^2 - 2w_{ij}\mathbf{x}_i' \boldsymbol{\beta} + (\mathbf{x}_i' \boldsymbol{\beta})^2],$$

e como $\sum_{j=1}^{k_i} w_{ij}^2$ não depende de $\boldsymbol{\beta}$ e $\bar{w}_i = \sum_{j=1}^{k_i} w_{ij} / k_i$

$$(I) \propto -2\mathbf{x}_i' \boldsymbol{\beta} k_i \bar{w}_i + k_i (\mathbf{x}_i' \boldsymbol{\beta})^2,$$

e somando e subtraindo $k_i \bar{w}_i^2$ temos

$$(I) \propto k_i [\bar{w}_i^2 - 2\bar{w}_i \mathbf{x}_i' \boldsymbol{\beta} + (\mathbf{x}_i' \boldsymbol{\beta})^2] = k_i (\bar{w}_i - \mathbf{x}_i' \boldsymbol{\beta})^2.$$

Portanto

$$p(\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \right\} \exp \left\{ -\sum_{i=1}^n k_i (\bar{w}_i - \mathbf{x}_i' \boldsymbol{\beta})^2 / 2 \right\},$$

e em notação matricial

$$p(\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \right\} \exp \left\{ -\frac{(\bar{\mathbf{w}} - \mathbf{X}' \boldsymbol{\beta})' \mathbf{K} (\bar{\mathbf{w}} - \mathbf{X}' \boldsymbol{\beta})}{2} \right\},$$

onde

$$\bar{\mathbf{w}} = (\bar{w}_1, \bar{w}_2, \dots, \bar{w}_n)',$$

$$\mathbf{K} = \begin{pmatrix} k_1 & 0 & \cdots & 0 \\ 0 & k_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n \end{pmatrix}.$$

Deste modo,

$$p(\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \underbrace{[\sigma^2(\bar{\mathbf{w}} - \mathbf{X}\boldsymbol{\beta})' \mathbf{K} (\bar{\mathbf{w}} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}]}_{II} \right\},$$

e

$$\begin{aligned} (II) &= \sigma^2 \bar{\mathbf{w}}' \mathbf{K} \bar{\mathbf{w}} - \sigma^2 \bar{\mathbf{w}}' \mathbf{K} \mathbf{X} \boldsymbol{\beta} - \sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} + \sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta} \\ &\propto -2\sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} + \sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta} \\ &= -2\sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} + \boldsymbol{\beta}' \underbrace{\mathbf{X}' (\sigma^2 \mathbf{K} + \mathbf{I}) \mathbf{X}}_{\Sigma^{-1}} \boldsymbol{\beta} \\ &= \underbrace{-2\sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} + \boldsymbol{\beta} \Sigma^{-1} \boldsymbol{\beta}}_{(III)}. \end{aligned}$$

Igualando (III) a forma quadrática $(\boldsymbol{\beta} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})$, temos

$$(III) \propto (\boldsymbol{\beta} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) = \boldsymbol{\beta}' \Sigma^{-1} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \Sigma^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu},$$

e como $\boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu}$ não depende de $\boldsymbol{\beta}$, temos que

$$(III) = -2\sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} + \boldsymbol{\beta} \Sigma^{-1} \boldsymbol{\beta} = \boldsymbol{\beta} \Sigma^{-1} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \Sigma^{-1} \boldsymbol{\mu}$$

de modo que

$$\sigma^2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} = \boldsymbol{\beta}' \Sigma^{-1} \boldsymbol{\mu}.$$

Multiplicado os dois lados da equação primeiramente por $(\boldsymbol{\beta} \boldsymbol{\beta}')^{-1} \boldsymbol{\beta}$ e em seguida por Σ , temos que

$$\boldsymbol{\mu} = \sigma^2 \Sigma \mathbf{X}' \mathbf{K} \bar{\mathbf{w}} = \sigma^2 [\mathbf{X}' (\sigma^2 \mathbf{K} + \mathbf{I}) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{K} \bar{\mathbf{w}}.$$

Desta maneira,

$$\boldsymbol{\beta} \mid \mathbf{w}, \mathbf{y}, \mathbf{X} \sim N_m \left(\sigma^2 [\mathbf{X}' (\sigma^2 \mathbf{K} + \mathbf{I}) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{K} \bar{\mathbf{w}}, [\mathbf{X}' (\sigma^2 \mathbf{K} + \mathbf{I}) \mathbf{X}]^{-1} \right). \quad (\text{B.8})$$

Agora vamos encontrar a distribuição condicional de $w_{ij} \mid \beta, \mathbf{y}, \mathbf{X}$. Temos que para cada i , j fixo

$$p(w_{ij} \mid \beta, \mathbf{y}, \mathbf{X}) \propto \{I_{(w_{ij} > 0)}I_{(z_{ij}=1)} + I_{(w_{ij} \leq 0)}I_{(z_{ij}=0)}\} \exp \left\{ -(w_{ij} - \mathbf{x}_i' \beta)^2 / 2 \right\},$$

ou seja,

$$w_{ij} \mid \beta, \mathbf{y}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \beta, 1) & \text{truncada à esquerda do zero, se } z_{ij} = 1 \\ N(\mathbf{x}_i' \beta, 1) & \text{truncada à direita do zero, se } z_{ij} = 0 \end{cases} \quad (\text{B.9})$$

Analogamente ao desenvolvimento feito para o modelo de regressão binária, temos que w_{ij} pode ser gerada diretamente através de

$$w_{ij} = \mathbf{x}_i' \beta + \Phi^{-1}(z_{ij} \Phi(-\mathbf{x}_i' \beta) + u_i(z_{ij} + (1 - 2z_{ij}) \Phi(-\mathbf{x}_i' \beta))).$$

Apêndice C

Amostrador de Gibbs: modelo de calibração binomial

Considere a distribuição a posteriori de referência para o modelo de calibração binomial

$$\begin{aligned} \pi(x_0, \beta, \mathbf{w} \mid \mathbf{y}, \mathbf{X}) &\propto |\beta_2| \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta' \mathbf{X}' \mathbf{X} \beta}{2\sigma^2} \right\} \times \\ &\times \left(\exp \left\{ -\frac{1}{2} \sum_{i=0}^n \sum_{j=1}^{y_i} (w_{ij} - \mathbf{x}_i' \beta)^2 \right\} I_{(w_{ij} > 0)} + \exp \left\{ -\frac{1}{2} \sum_{i=0}^n \sum_{j=y_i+1}^{k_i} (w_{ij} - \mathbf{x}_i' \beta)^2 \right\} I_{(w_{ij} \leq 0)} \right). \end{aligned} \quad (\text{C.1})$$

Primeiramente, vamos encontrar a distribuição condicional de $\beta \mid x_0, \mathbf{w}, \mathbf{y}, \mathbf{X}$. Observe que

$$\begin{aligned} p(\beta \mid x_0, \mathbf{w}, \mathbf{y}, \mathbf{X}) &\propto |\beta_2| \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta' \mathbf{X}' \mathbf{X} \beta}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2} \sum_{i=0}^n \sum_{j=1}^{k_i} (w_{ij} - \mathbf{x}_i' \beta)^2 \right\} \\ &\propto |\beta_2| \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\beta' \mathbf{X}' \mathbf{X} \beta}{2\sigma^2} \right\} \exp \left\{ -\sum_{i=0}^n k_i (\bar{w}_i - \mathbf{x}_i' \beta)^2 / 2 \right\} \\ &= |\beta_2| \exp \left\{ -\frac{\beta' \mathbf{X}'_0 \mathbf{X}_0 \beta}{2\sigma^2} \right\} \exp \left\{ -\frac{(\bar{\mathbf{w}}_0 - \mathbf{X}_0 \beta)' \mathbf{K}_0 (\bar{\mathbf{w}}_0 - \mathbf{X}_0 \beta)}{2} \right\}, \end{aligned}$$

onde

$$\bar{\mathbf{w}}_0 = (\bar{w}_0, \bar{w}_1, \dots, \bar{w}_n)', \quad (\text{C.2})$$

$$\mathbf{K}_0 = \begin{pmatrix} k_0 & 0 & \cdots & 0 \\ 0 & k_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n \end{pmatrix}. \quad (\text{C.3})$$

Procedendo analogamente, ao desenvolvimento já realizado para o modelo de regressão binomial temos que

$$p(\beta \mid x_0, \mathbf{w}, \mathbf{y}, \mathbf{X}) \propto |\beta_2| \exp \left\{ -\frac{(\beta - \mathbf{m})' \Sigma^{-1} (\beta - \mathbf{m})}{2\sigma^2} \right\}, \quad (\text{C.4})$$

onde

$$\mathbf{m} = \sigma^2 [\mathbf{X}_0' (\sigma^2 \mathbf{K}_0 + \mathbf{I}) \mathbf{X}_0]^{-1} \mathbf{X}_0' \mathbf{K}_0 \bar{\mathbf{w}}_0 \quad \text{e} \\ \Sigma^{-1} = \mathbf{X}_0' (\sigma^2 \mathbf{K}_0 + \mathbf{I}) \mathbf{X}_0,$$

isto é, a distribuição condicional completa é proporcional ao núcleo de uma $N_2(\mathbf{m}, \sigma^2 \Sigma)$ multiplicado por $|\beta_2|$.

Para gerar desta distribuição condicional (distribuição alvo) utilizaremos um passo do algoritmo de Metropolis-Hastings. Para tanto, precisamos especificar uma distribuição proposta de fácil geração e determinar a probabilidade de aceitação de um valor gerado (α) segundo a equação

$$\alpha = \min \left\{ 1, \frac{f(y)q(x \mid y)}{f(x)q(y \mid x)} \right\} \quad (\text{C.5})$$

onde $f(\cdot)$ é o núcleo da distribuição alvo e $q(\cdot \mid \cdot)$ é a distribuição proposta, y é o novo valor da distribuição proposta e x é um valor da distribuição alvo.

Em relação a distribuição condicional completa para w_{ij} , já vimos no desenvolvimento para o modelo de regressão binomial, que ela é dada por

$$w_{ij} \mid x_0, \beta, \mathbf{y}, \mathbf{X} \sim \begin{cases} N(\mathbf{x}_i' \beta, 1) & \text{truncada à esquerda do zero, se } z_{ij} = 1 \\ N(\mathbf{x}_i' \beta, 1) & \text{truncada à direita do zero, se } z_{ij} = 0 \end{cases}.$$

e pode ser facilmente gerada através de

$$w_{ij} = \mathbf{x}_i' \beta + \Phi^{-1}(z_{ij} \Phi(-\mathbf{x}_i' \beta) + u_i(z_{ij} + (1 - 2z_{ij}) \Phi(-\mathbf{x}_i' \beta))). \quad (\text{C.6})$$

Vamos agora, encontrar a distribuição marginal completa para x_0 . Observe que

$$p(x_0 \mid \beta, \mathbf{y}, \mathbf{w}, \mathbf{X}) \propto \exp \left\{ -\frac{(\mathbf{x}_0' \beta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{k_0} (w_{0j} - \mathbf{x}_0' \beta)^2 \right\} \quad (\text{C.7})$$

$$\begin{aligned}
&= \exp \left\{ -\frac{(\beta_1 + \beta_2 x_0)^2}{2\sigma^2} - \frac{k_0(\bar{w}_0 - \mathbf{x}_0' \boldsymbol{\beta})^2}{2} \right\} \\
&\propto \exp \left\{ -\frac{(1 + \sigma^2 k_0)\beta_2^2 x_0^2 - 2[\sigma^2 k_0 \bar{w}_0 \beta_2 - (1 + \sigma^2 k_0)\beta_1 \beta_2]x_0}{2\sigma^2} \right\} \\
&= \exp \left\{ -\frac{x_0^2 - 2x_0 \frac{\sigma^2 k_0 \bar{w}_0 \beta_2 - (1 + \sigma^2 k_0)\beta_1 \beta_2}{(1 + \sigma^2 k_0)\beta_2^2}}{2 \frac{\sigma^2}{(1 + \sigma^2 k_0)\beta_2^2}} \right\} \\
&\propto \exp \left\{ -\frac{(x_0 - m_0)^2}{2v_0^2} \right\},
\end{aligned}$$

onde

$$m_0 = \frac{\sigma^2 k_0 \bar{w}_0 - (1 + \sigma^2 k_0)\beta_1}{(1 + \sigma^2 k_0)\beta_2}, \quad \text{e} \quad v_0^2 = \frac{\sigma^2}{(1 + \sigma^2 k_0)\beta_2^2}.$$

Portanto temos que

$$x_0 \mid \boldsymbol{\beta}, \mathbf{y}, \mathbf{w}, \mathbf{X} \sim N(m_0, v_0^2). \quad (\text{C.8})$$

Referências Bibliográficas

- [1] Albert, J. H. e Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88**, 669-679.
- [2] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. Lond.* **53**, 370-418.
- [3] Berger, J. O. e Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200-207.
- [4] Berger, J. O. e Bernardo, J. M. (1992a). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid, e A. F. M. Smith, eds). Oxford University Press, 35-60 (com discussão).
- [5] Berger, J. O. e Bernardo, J. M. (1992b). Ordered group reference priors with application to the multinomial problem. *Biometrika* **79**, 25-37.
- [6] Berger, J. O. , Bernardo, J. M. e Mendonza, M. (1989). On priors that maximize expected information. *Recent Developments in Statistics and Their Applications*. (J. P. Klein e J. C. Lee eds). 1-20. Seoul: Freedom Academy.
- [7] Bernardo, J. M. (1977). Inferences about the ratio of normal means: A Bayesian approach to the Fieller-Creasy problem. *Recent Developments in Statistics*. (Barra et al eds.). 345-350. Amsterdam: North Holland.
- [8] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B* **41**, 113-147 (com discussão).
- [9] Bernardo, J. M. (1997). Non-informative priors do not exist. *J. Statist. Planning Inference* **65**, 159-189 (com discussão).

- [10] Bernardo, J. M. (1997b). Na discussão de Gutiérrez-Peña e Smith (1997). *Test* **6**, 70-71.
- [11] Bernardo, J. M. (2001). *Curso Métodos Bayesianos Objetivos*. Seminário Chileno de Estatística Bayesiana.
- [12] Bernardo, J. M. e Ramón, J. M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* **47**, 101-135.
- [13] Bernardo, J. M. e Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.
- [14] Branco, M. D. , Bolfarine, H. , Iglesias, P. e Arellano-Valle, R. B. (2003). Bayesian and classical solutions for binomial cytogenetic dosimetry problem. *Journal of Data Science* **1**, 65-82.
- [15] Dawid, A. P. , Stone, M. e Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. R. Statist. Soc. B* **35**, 189-233, (com discussão).
- [16] Fang, K. T. e Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. Beijing: Springer-Verlag.
- [17] Ghosh, M. Carlin, B. P. e Srivastava, M. S. (1995). Probability matching priors for linear calibration. *Test* **4**, 2, 333-357.
- [18] James, B. R. (1996). *Probabilidade: Um curso em nível intermediário*. Impa: Projeto Euclides.
- [19] Jaynes, E. T. (1968). Prior Probabilities. *IEEE Trans. Syst. Sci. Cyb.* **4**, 227-291.
- [20] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. A* **186**, 453-461.
- [21] Jeffreys, H. (1961). *Theory of Probability*, 3ª ed. Oxford: Oxford University Press.
- [22] Johnson, V. E. e Albert, J. H. (2000). *Ordinal Data Modeling*. New York: Springer.
- [23] Johnson, N. L. , Kotz, S. e Balakrishnan, N. (1995). *Continuous Univariate Distributions* vol 2, 2nd edição. New York: Wiley.
- [24] Kubokawa, T. e Robert, C. (1994). New perspectives on linear calibration. *J. Multivariate Analysis* **50**, 178-200.

- [25] Kullback, S. (1959). *Information Theory and Statistics*, New York: Wiley.
- [26] Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Mem. Acad. R. Sci. Présentés par Divers Savans* **6**, 621-656. (traduzido em *Statist. Sci.* **1**, 359-378).
- [27] Laplace, P. S. (1812). *Théorie Analytique des Probabilités*. Paris: Courcier.
- [28] Lindley, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27**, 986-1005.
- [29] Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295-304.
- [30] Muller, P. (1991). Metropolis based posterior integration schemes. *Technical Report, Statistics Department Purdue University*.
- [31] Shannon, C. E. (1948). *A Mathematical Theory of Communication*. Urbana, IL.: Univ. Illinois Press.
- [32] Soofi, E. S. (2000). Principal information theoretic approaches. *J. Amer. Statist. Assoc.* **95**, 1349-1353.
- [33] Stone, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71**, 114-125, (com discussão).
- [34] Stone, M. e Dawid, A. P. (1972). Un-Bayesian implications of improper Bayesian inference in routine statistical problems. *Biometrika* **59**, 369-375.
- [35] Sveshnikov, A. A. (1978). *Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions*. New York: Dover.
- [36] Tanner, M. A. (1993). *Tools of Posterior Distributions and Likelihood Functions*, 2nd edição. New York: Springer-Verlag.
- [37] Tanner, M. A., e Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (com discussão). *J. Amer. Statist. Assoc.* **82**, 528-550.
- [38] West, M. (2000). Bayesian regression analysis in the "Large p, Small n" paradigm. *ISDS Discussion Paper #00-22*.

- [39] Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g -prior distributions. *Bayesian Inference and Decision Techniques* (P. Goel e A. Zellner, eds). Elsevier Science Publishers, 233-243.