PRIORIS NÃO-INFORMATIVAS
PARA O
MODELO DE WEIBULL

Fernando Antonio Moala

Orientador: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos, da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Ciências - Área: Ciências de Computação e Matemática Computacional.

São Carlos

1993

Aos meus pais Waldomiro e Jandyrá, que,
com afinho, lutaram para a realização de
meus estudos.

Aos meus irmãos Paulo (in memoriam),
Miro, Kico e Celo pelo incentivo, carinho
e apoio.

Agradecimentos

Ao Prof. Dr. Josemar Rodrigues pela amizade, pelo incentivo e apoio dado, no início do meu curso de Pós-Graduação e da minha carreira acadêmica, e por sua dedicada e segura orientação.

À Marta Y. Baba pela amizade, carinho e companheirismo, durante este período, dentro e fora do ICMSC, e pelas oportunas discussões e sugestões sobre o tema em questão.

Ao Prof. Dr. Jorge A. Achcar pelas valiosas sugestões e comentários a este trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro recebido.

Agradeço, também, ao Aparecido Doniseti da Costa pelo excelente trabalho de digitação do texto; enfim, a todos que direta ou indiretamente contribuíram para a realização deste estudo.

"It was from studying these flashes of intuition in Jeffreys that I became convinced that there must exist a general formal theory of determination of priors by logical analysis of the prior information and that to develop it is today the top priority research problem of Bayesian theory."

Jaynes (1980)

Resumo

Pela ausência de uma definição precisa do conceito de “Não-Informação”, há na literatura Bayesiana várias formas de formular densidades a priori não-informativas, por exemplo, Jeffreys (1967), Zellner (1984, 1990), Tibshirani (1989) e Bernardo (1979). Um estudo para verificar se estas densidades a priori são equivalentes, isto é, produzem densidades a posteriori iguais, é de grande interesse prático. O objetivo de nosso trabalho é selecionar um modelo que melhor represente um estado de “pouco conhecimento”, a priori, sobre o parâmetro. Este estudo comparativo é feito para o caso específico do modelo de Weibull, com a finalidade de obter informação a respeito da função de confiabilidade R . Uma exposição de cada densidade a priori bem como de suas propriedades também é apresentada. Para início das discussões, apresentamos, no Capítulo 1, os conceitos básicos da Inferência Bayesiana, questões filosóficas, vantagens e desvantagens de sua utilização em uma análise estatística. No Capítulo 2, revisamos algumas propriedades do modelo de Weibull. O Capítulo 3 é dedicado ao estudo das densidades a priori não-informativas onde discutimos sua definição, vantagens e desvantagens de sua utilização e descrevemos os métodos propostos por Jeffreys, Zellner, Tibshirani e Bernardo. No capítulo 4, calculamos as respectivas densidades a posteriori para a função de confiabilidade. Utilizando a Aproximação de Laplace (Kass, Tierney e Kadane, 1990) obtivemos uma forma fechada para essas densidades a posteriori; mais do que isso, encontramos que as densidades a posteriori seguem uma distribuição conhecida, denominada Log-Gama Negativa. Os resultados obtidos via Laplace são comparados graficamente aos resultados de Sinha e Guttman (1988), obtidos via integração numérica. Além disso, um outro resultado importante obtido é que as densidades a posteriori correspondentes às densidades a priori de Zellner, Tibshirani e Bernardo, respectivamente, são iguais. A igualdade das densidades a posteriori correspondentes as duas últimas densidades a priori é devida à coincidência dessas densidades a priori. Finalmente, no Capítulo 5, apresentamos uma ilustração numérica, através de gráficos e interva-

los, para o estudo comparativo entre as densidades a posteriori correspondentes às densidades a priori de Jeffreys e Zellner e uma conclusão do estudo.

Verificamos, também, que o critério de probabilidade de cobertura dos intervalos a posteriori, proposto por Berger(1992), não é suficiente para determinarmos qual densidade a priori seria menos informativa, para valores de R próximos de 0 e 1; para isto sugerimos um critério adicional, através da comparação das amplitudes dos intervalos a posteriori.

Abstract

Because of the absence of a precise definition of the "NonInformation" concept, several ways of formulating noninformative priors have been reported in the Bayesian literature (see, for example, Jeffreys, 1967; Bernardo, 1979; Zellner, 1984, 1990; Tibshirani, 1989). A study carried out to determine whether these priors are equivalent, i.e., whether they produce equal posteriors, is of great practical interest. The objective of our study was to select the model that would best represent a state of "little knowledge", a priori, about the parameter. This comparative study was carried out for the specific case of the Weibull model in order to obtain information about the reliability function R . A description of each prior and of its properties is also presented. To start the discussion, the basic concepts of Bayesian Inference, philosophical questions, advantages and disadvantages of its utilization are presented in a statistical analysis. Chapter 2 is devoted to a review of some properties of the Weibull model. Chapter 3 concerns the study of noninformative priors, with a discussion of their definition, advantages and disadvantages of their use, and with a description of the methods proposed by Jeffreys, Zellner, Tibshirani and Bernardo. In Chapter 4, the respective posteriors for the reliability function are calculated. Using the Laplace approximation (Kass, Tierney and Kadane, 1990), we obtained a closed formula for these posteriors; in addition, we found that the posteriors follow a known distribution denoted Log-Gamma Negative. The results obtained by the Laplace method were graphically compared to those of Sinha and Guttman (1988) obtained by numerical integration. Furthermore, another important result obtained was that the posteriors corresponding to the priors of Zellner, Tibshirani and Bernardo are equal. The equality of the posteriors corresponding to the last two priors is due to the coincidence of these priors. Finally, Chapter 5 presents a numerical illustration based on graphs and intervals for the comparative study of the posteriors corresponding to the priors of Jeffrey and Zellner, as well as a conclusion about the study.

We also noted that the criterion for the probability of a posterior interval coverage proposed by Berger (1992) is not sufficient to determine which prior would be least informative for R values close to 0 and 1. For this, we suggest an additional criterion based on the comparison of the amplitudes of a posterior intervals.

Terminologia e Notação

Variáveis aleatórias são representadas por letras maiúsculas e suas realizações por minúsculas.

A função densidade de probabilidade de uma variável aleatória é representada por $f(\cdot)$, as funções de verossimilhança dos parâmetros envolvidos por $L(\cdot \mid \underline{x})$, as densidades a priori dos parâmetros envolvidos por $\pi(\cdot)$, as densidades a posteriori por $p(\cdot \mid \underline{x})$.

Letras gregas indicam os parâmetros dos modelos.

Em geral, constantes são representadas por letras minúsculas.

Vetores são representados por letras sublinhadas e matrizes por letras maiúsculas em negrito.

As seguintes abreviações são utilizadas:

v.a	variável aleatória
f.d.p	função densidade de probabilidade
d.t	“data translated”
EMV	Estimador de Máxima Verossimilhança
i.i.d.	independente e identicamente distribuída

CONTEÚDO

1 Inferência Bayesiana	01
1.1 Introdução	01
1.2 Análise Clássica versus Inferência Bayesiana	03
1.3 Densidade a Priori	05
2 Modelo de Weibull	06
3 Densidades a Priori Não-Informativas	08
3.1 Introdução	08
3.2 Densidade a Priori de Jeffreys	12
3.2.1 Densidade a Priori Localmente Uniforme	12
3.2.2 Verossimilhança “Data Translated”	12
3.2.3 Intervalo de Verossimilhança de Alta Densidade	13
3.2.4 Densidade a Priori Não-Informativa e Verossimilhança “Data Translated”	14
3.2.5 Verossimilhança Aproximadamente “Data Translated” e Densidade a Priori Aproximadamente Não-Informativa	18
3.2.6 Densidade a Priori de Jeffreys	21
3.2.7 Extensão da Densidade a Priori de Jeffreys para Modelos Multiparamétricos	24
3.2.8 Obtenção da Densidade a Priori de Jeffreys para o Modelo de Weibull	31
3.3 Densidade a Priori de Zellner (MDIP)	34
3.3.1 Introdução	34
3.3.2 Obtenção da Densidade a Priori de Zellner para o Modelo de Weibull	41
3.4 Densidade a Priori de Tibshirani	42

3.4.1	Motivação	42
3.4.2	Introdução	44
3.4.3	Construção da Densidade a Priori de Tibshirani	46
3.4.4	Observações	48
3.4.5	Obtenção da Densidade a Priori de Tibshirani para o Modelo de Weibull	50
3.5	Densidade a Priori de Bernardo	53
3.5.1	Motivação	53
3.5.2	Construção da Densidade a Priori Referencial	54
3.5.3	Densidade a Priori Referencial para Parâmetros de Interesse na presença de Parâmetros “Nuisances”	57
3.5.4	Obtenção da Densidade a Priori de Bernardo para o Modelo de Weibull	63
4	Estudo comparativo das Densidades a Priori	65
4.1	Cálculo da Densidade a Posteriori de R utilizando-se a densidade a priori de Jeffreys	65
4.1.1	Cálculo da Densidade a Priori de Jeffreys nos novos parâmetros ...	65
4.1.2	Cálculo da Densidade a Posteriori de R	66
4.2	Cálculo da Densidade a Posteriori de R utilizando-se a densidade a priori MDIP	70
4.2.1	Cálculo da Densidade a Priori MDIP nos novos parâmetros	70
4.2.2	Cálculo da Densidade a Posteriori de R	72
4.3	Cálculo da Densidade a Posteriori de R utilizando-se a densidade a priori de Tibshirani	73
4.3.1	Cálculo da Densidade a Priori de Tibshirani nos novos parâmetros ..	73
4.3.2	Cálculo da Densidade a Posteriori de R	74
4.4	Cálculo da Densidade a Posteriori de R utilizando-se a densidade a priori de Bernardo	75

4.4.1 Cálculo da Densidade a Priori de Bernardo nos novos parâmetros ..	75
4.4.2 Cálculo da Densidade a Posteriori de R	77
5 Ilustração Numérica	81
5.1 Comparação gráfica	81
5.2 Comparação Pontual e Intervalar	85
5.3 Discussão	95
Apêndice A - Método de Laplace para Aproximação de Integrais	97
Apêndice B - Distribuição Log-Gama Negativa	99
Apêndice C - Reparametrização Ortogonal	101
Apêndice D - Algumas Integrais Usuais	105
Apêndice E - Programas Computacionais	107
Bibliografia	119

Capítulo 1

INFERÊNCIA BAYESIANA

1.1 Introdução

Um problema que freqüentemente surge para o estatístico é o interesse em obter informação a respeito de uma quantidade ou estado da natureza, ϕ , desconhecida. Uma teoria estatística alternativa à Análise Clássica, para solucionar este problema, muito utilizada em problemas de Análise de Confiabilidade, é dada pela Inferência Bayesiana a ser considerada nesta seção.

Com o intuito de introduzir a notação a ser utilizada, consideremos o vetor $\underline{x} = (x_1, x_2, \dots, x_n)$ representando n observações associadas a um modelo probabilístico $f(x | \phi)$ indexado por um parâmetro ϕ desconhecido, pertencente ao espaço paramétrico Φ .

Na Análise Clássica, o parâmetro ϕ é considerado uma constante fixada. Um estimador pontual $\hat{\phi}(\underline{x})$, que é uma função dos dados $\underline{x}$ gerados pelo modelo $f(x | \phi)$, é então escolhido por algum princípio, tal como, estimador de máxima verossimilhança (EMV), variância mínima, mínimos quadrados, ou métodos de momentos. Imaginando todos os possíveis conjuntos de dados $\underline{x}_1, \underline{x}_2, \dots$ que poderiam ser gerados por $f(x | \phi)$ para um dado valor de ϕ , obteríamos a correspondente distribuição amostral de $\hat{\phi}(\underline{x})$. Um intervalo de confiança $100(1 - \gamma)\%$ para ϕ pode ser calculado através das n observações para proporcionar uma idéia de quanto o valor $\hat{\phi}(\underline{x})$ difere do verdadeiro valor ϕ .

Na Inferência Bayesiana, ϕ é uma variável aleatória com uma densidade de probabilidade, $\pi(\phi)$, denominada Densidade a Priori. Esta densidade expressa o estado de conhecimento ou ignorância (incerteza) sobre ϕ , antes da realização do

experimento, isto é, antes da obtenção dos dados amostrais $\underline{x}$.

Uma vez observado $\underline{x}$, uma forma natural de combinar o conhecimento a priori com a informação fornecida pelos dados, é através da distribuição condicional de ϕ dado $\underline{x}$. Pelo Teorema de Bayes, esta distribuição condicional pode ser escrita como:

$$p(\phi | \underline{x}) = c.L(\phi | \underline{x}).\pi(\phi) \quad (1)$$

denominada Densidade a Posteriori de ϕ dado $\underline{x}$, onde c é a constante normalizadora, isto é

$$c^{-1} = \int_{\Phi} L(\phi | \underline{x}).\pi(\phi).d\phi \quad (2)$$

A função de verossimilhança, $L(\phi | \underline{x})$, resume a informação a respeito de ϕ contida nos dados, sendo definida por

$$L(\phi | \underline{x}) = \prod_{i=1}^n f(x_i | \phi) \quad (3)$$

Portanto, as inferências sobre ϕ são feitas utilizando-se a densidade a posteriori $p(\phi | \underline{x})$, por exemplo, os Intervalos Bayesianos para ϕ . Assim como a densidade a priori $\pi(\phi)$ reflete o conhecimento a priori sobre ϕ , antes de uma investigação estatística, a densidade a posteriori $p(\phi | \underline{x})$ reflete a opinião atualizada sobre ϕ , depois dos dados $\underline{x}$ serem observados.

Na Inferência Bayesiana o trabalho do estatístico é o de construir modelos probabilísticos para o parâmetro ϕ e desenvolver, baseado nestes modelos, métodos de inferência sobre o verdadeiro valor de ϕ . Segundo Jaynes (1980), o desenvolvimento de uma teoria geral para a construção de densidades a priori, que quantifiquem adequadamente a informação a priori sobre o parâmetro, constitui um tópico prioritário da Inferência Bayesiana a ser pesquisado.

Como podemos ver, a Inferência Bayesiana não pode ser encarada como um capítulo da inferência estatística clássica, como muitos livros clássicos introduziram. A teoria Bayesiana e a teoria Clássica são filosoficamente opostas, pois, enquanto na teoria Clássica, as inferências estão baseadas na distribuição dos dados, fixado o parâmetro, na Bayesiana se dá exatamente o contrário, onde as inferências estão baseadas na distribuição do parâmetro para um conjunto de dados fixado. Ou seja, na teoria Clássica o parâmetro ϕ é fixo, enquanto na Bayesiana ele é tratado como uma variável aleatória. É importante realçar que todo problema estatístico, seja de teste de hipótese, intervalos de confiança, inferência não-paramétrica e etc, pode ser abordado tanto da forma Clássica quanto da forma Bayesiana. A seguir, apresentaremos algumas vantagens e desvantagens da escolha dessas duas abordagens.

1.2 Análise Clássica versus Inferência Bayesiana

Há distintas diferenças entre os métodos de Inferência Clássica e os métodos Bayesianos, proporcionando vantagens e desvantagens na utilização de um sobre o outro.

A teoria de Inferência Clássica, baseada em hipóteses restritivas, tais como Normalidade e Independência de erros, que proporcionam convenientes soluções matemáticas, dirige-se somente a um limitado número de problemas encontrados na prática. Além disso, uma característica distintiva da Inferência Bayesiana é que ela leva em consideração a informação a priori na análise. Este fato se contrasta com a Análise Clássica em que a informação a priori é considerada somente de modo não intuitivo ou informal, por ex. através de variáveis auxiliares, na teoria da amostragem com estratificação da população, etc. Ou seja, um pesquisador pode querer usar seu julgamento e sua experiência passada que reflita um conhecimento do fenômeno em estudo e a abordagem Bayesiana permite incorporar, formalmente, esta informação ao modelo em estudo. Neste caso, a densidade a priori $\pi(\phi)$ re-

sumiria todo o conhecimento do pesquisador sobre o parâmetro ϕ do modelo, antes dos dados $\underline{x}$ serem obtidos. Já os métodos clássicos utilizam somente dados empíricos correspondentes ao fenômeno estudado.

Por outro lado, a densidade a priori pode incorporar uma noção subjetiva de probabilidade, quando as frequências dos possíveis valores do parâmetro ϕ são desconhecidas. Como consequência, as probabilidades a priori não admitem uma interpretação freqüentista e não estão sujeitas à confirmação experimental. Neste caso, os dados observados podem ser utilizados para estimar a densidade a priori. Isto, porém, tem gerado um questionamento da relevância da utilização da densidade a priori e controvérsias a respeito da atratividade do método Bayesiano.

Se há realmente um modo acessível de quantificar a informação a priori em uma análise Bayesiana, isto dependerá da natureza do problema.

Outra consideração a ser evidenciada é que os métodos baseados na Inferência Bayesiana, quase sempre, requerem menos dados amostrais, para obter os mesmos resultados dos métodos baseados na Inferência Clássica. Na maioria dos casos, isto é uma motivação para se utilizar os métodos Bayesianos, existindo, assim, uma forte inclinação para utilizar informação a priori. Esta é uma consideração muito importante em áreas de aplicação onde os dados amostrais podem ser onerosos e difíceis de serem obtidos, como é o caso da Análise de Sobrevivência. De fato, a partir da década de 70, os pesquisadores intensificaram o uso dos métodos Bayesianos em problemas relacionados à Análise de Dados de Sobrevivência. Mas-
tran e Singpurwalla (1978), Martz e Waller (1982) dão exemplos de aplicações em problemas reais.

A facilidade de interpretações, obtidas através dos gráficos das densidades a posteriori do parâmetro de interesse, é mais uma vantagem dos métodos Bayesianos.

Até que ponto tais ganhos podem ser realmente aceitos é realçado por Box e Tiao (1973), . . . “Somente a Inferência Bayesiana parece proporcionar flexibilidade suficiente, para solucionar problemas complexos, livre do impedimento de limitações puramente técnicas . . .”

Em nosso trabalho, utilizamos os métodos Bayesianos, para o caso específico do modelo de Weibull, com a finalidade de obter informação a respeito da função de confiabilidade R .

1.3 Densidades a priori

Como visto anteriormente, a parte fundamental de qualquer análise Bayesiana é a densidade a priori. Alguns métodos para se obter densidade a priori têm sido propostos na literatura.

O método pode ser subjetivo (escola de De Finetti-Savage), quando é baseado na determinação de probabilidade subjetiva de subconjuntos (eventos) do espaço paramétrico. Na realidade, as densidades a priori resultantes deste método incorporam uma opinião pessoal do pesquisador, a respeito do parâmetro, proveniente de sua experiência ou de sua intuição. Métodos subjetivos são, geralmente, mais difíceis de se aplicar na prática.

O método também pode ser objetivo (escola de Laplace-Jeffreys) quando:

- i) nenhuma ou pouca informação a priori é disponível. A densidade a priori assim obtida é denominada Densidade a Priori Não-Informativa, ou
- ii) quando a informação a priori é proveniente de estimativas passadas do parâmetro em questão (H. Robbins, 1964). Neste caso, a densidade a priori incorpora a informação que provém de um experimento equivalente realizado anteriormente.

Uma das maneiras de classificar as densidades a priori, muito comum, é dicotomizada em densidades a priori “informativas” e “não-informativas”.

Utiliza-se a Densidade a Priori Informativa quando o pesquisador dispõe de informação sobre o parâmetro proveniente de sua experiência sobre o modelo, considerações teóricas ou de informações contidas em amostras de dados passados.

Um estudo especial a respeito de densidades a priori não-informativas, definidas anteriormente, é apresentado no capítulo 3.

Capítulo 2

MODELO DE WEIBULL

Consideremos o modelo de Weibull dado por:

$$f(x | p, \theta) = \frac{p}{\theta} x^{p-1} \exp \left\{ -\frac{x^p}{\theta} \right\} \quad \text{para todo } x > 0, \quad (4)$$

que depende dos parâmetros $p > 0$ e $\theta > 0$.

Os parâmetros p e θ são referidos como parâmetros de forma e escala respectivamente.

Figura 1 f.d.p. do modelo de Weibull para $\theta = 1$ e $p = 0.5, 1.0, 1.5$ e 3.0

O modelo de Weibull tem média e variância respectivamente,

$$E(X) = \frac{\theta^{1/p}}{p} \Gamma\left(\frac{1}{p}\right)$$

$$Var(X) = \theta^{2/p} \left\{ \frac{2}{p} \Gamma\left(\frac{2}{p}\right) - \left(\frac{1}{p} \Gamma\left(\frac{1}{p}\right) \right)^2 \right\}$$

(5)

Seja X representando o tempo de vida de um equipamento. A Função de Confiabilidade (ou de Sobrevivência) $R = R(t) = P(X > t)$ dada por:

$$R = R(t) = \exp\left\{-\frac{t^p}{\theta}\right\},$$

(6)

define a confiabilidade do equipamento considerado para uma tarefa de duração t , ou ainda, é a probabilidade do equipamento não falhar antes de t .

O modelo de Weibull é uma extensão natural da distribuição exponencial, incluindo-a como caso especial quando $p = 1$.

É talvez o modelo mais utilizado em problemas de confiabilidade e testes de sobrevivência. É adequado ao estudo de tempos de falhas de muitos tipos de manufaturados, por exemplo, válvulas eletrônicas (Kao, 1959), rolamentos ou engrenagens (Lieblein e Zelen, 1956), e isolantes de eletricidade (Nelson, 1972). É também muito utilizado em aplicações biomédicas, como no estudo do tempo de ocorrência de tumores em seres humanos (Whittemore e Altschuler, 1976) ou em cobaias (Pike, 1966 ; Peto e Lee, 1973). Mann (1968), discute uma variedade de situações onde este modelo pode ser utilizado.

Capítulo 3

DENSIDADES A PRIORI NÃO-INFORMATIVAS

3.1 Introdução

Nesta seção introduzimos a filosofia geral das densidades a priori não-informativas. Algumas densidades a priori não-informativas específicas são consideradas nas seções subsequentes.

A densidade a priori não-informativa refere-se ao caso em que relativamente pouca informação é disponível a priori, isto é, a informação a respeito dos parâmetros do modelo não é considerada substancial, se comparada à informação proveniente de uma amostra. Também significa que há um conjunto de valores paramétricos que o pesquisador acredita ser igualmente provável para escolha do parâmetro, ou seja, o pesquisador mostra-se indiferente a um conjunto de valores selecionado do parâmetro.

Há várias razões para se utilizar densidades a priori não-informativas, dentre elas:

- i) pode-se, imediatamente, fazer uma análise Bayesiana para o modelo, uma vez que existe na literatura inúmeras densidades a priori não-informativas propostas;
- ii) as densidades a priori informativas, baseadas em métodos subjetivos, são, frequentemente, mais difíceis de se aplicar na prática;

iii) a realização de um experimento é justificada, se a informação fornecida por este for considerável em relação à informação já conhecida, ou seja, uma investigação científica só é empreendida, se houver um aumento substancial de conhecimento;

iv) se há uma quantidade substancial de dados, a densidade a priori terá pouco efeito. Isto justificaria o uso de densidades a priori não-informativas;

v) ao formular um procedimento estatístico alternativo ou para um estudo comparativo entre metodologias estatísticas, este seria mais facilmente aceito, se utilizássemos densidades a priori não-informativas;

vi) as densidades a priori não-informativas podem ser úteis como ponto de referência para o estudo do efeito das densidades a priori informativas (Inferência Bayesiana Robusta).

Há também, algumas críticas que são feitas à utilização das densidades a priori não-informativas, entre elas:

i) a existência de vários procedimentos para a obtenção dessas densidades a priori na literatura. Essa diversidade de métodos para desenvolver densidades a priori não-informativas concerne à dificuldade de como expressar a noção de “pouca informação”. A maioria dos métodos começam com uma tentativa de medir a quantidade de informação contida na densidade a priori, através da comparação da densidade a priori com a função de verossimilhança ou através do grau de influência que a densidade a priori causa na posteriori.

Por exemplo, Jeffreys (1967) propôs uma densidade a priori não-informativa resultante de um argumento baseado na invariância sob transformações e na Medida de Informação de Fisher. Um método alternativo de determinar densidades a priori não-informativas é baseado no conceito de Medida de Entropia (Zellner,

1984, 1990 e Bernardo, 1979). Outros métodos de se obter densidades a priori não-informativas podem ser desenvolvidos, utilizando-se noções freqüentistas como, cobertura freqüentista assintótica (Tibshirani, 1989) e etc.

ii) o grande número de densidades a priori não-informativas pode causar dificuldades, quando essas não produzem resultados similares e os dados não informam o suficiente, para se chegar a uma conclusão;

Exemplo 1: Suponhamos que $f(x | \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$. Considere as seguintes densidades a priori não-informativas $\pi_1(\theta) = 1$ (Bayes, 1763, Laplace, 1774 e Geisser, 1984), $\pi_2(\theta) = [\theta(1 - \theta)]^{-1}$ (Novick e Hall, 1965, Jaynes, 1968 e Villegas, 1977) e $\pi_3(\theta) = [\theta(1 - \theta)]^{-1/2}$ (Jeffreys, 1967, Box e Tiao, 1973, Akaike, 1978 e Bernardo, 1979). As correspondentes médias a posteriori são, $\mu_1(x) = (x+1)/(n+2)$, $\mu_2(x) = x/n$ e $\mu_3(x) = (x + 1/2)/(n + 1)$, respectivamente. Elas são similares se n for grande.

iii) as densidades a priori não-informativas podem não ser invariantes sob transformações diferenciáveis e um a um;

iv) alguns procedimentos propostos, para obtenção de densidades a priori não-informativas, podem levar a densidades a priori impróprias ($\int_{\Phi} \pi(\phi) d\phi = \infty$);

v) a dificuldade de se medir a informação a priori pode nos levar a um questionamento sobre quanto uma densidade a priori não-informativa pode ser;

Exemplo 2: Se X tem uma distribuição Binomial(n, θ), duas densidades a priori não-informativas, freqüentemente utilizadas são $\pi_1(\theta) = 1$ (Bayes e Laplace) e $\pi_2(\theta) = \text{Beta}(1/2, 1/2)$ (Jeffreys). Cada uma proporciona um tipo de informação: π_1 informa que todos os valores de θ são igualmente prováveis e π_2 que os valores de

θ próximos de 0 ou 1 são mais prováveis a priori. Então estas densidades a priori não são estritamente não-informativas no sentido de expressar ausência de informação a priori.

Diante desses problemas, há duas proposições que são objetos de estudos dos Bayesianos que utilizam densidades a priori não-informativas. A primeira é o desenvolvimento de métodos, para obtenção de densidades a priori não-informativas que não produzam esses potenciais problemas. A segunda proposição é a comparação de diferentes densidades a priori não-informativas existentes na literatura Bayesiana a fim de se discriminar entre estas possíveis densidades a priori. Um procedimento para se obter isto, é através da avaliação frequentista dos intervalos a posteriori do parâmetro. A segunda proposição é o objetivo deste trabalho.

Dedicamos às próximas seções deste capítulo a uma discussão das densidades a priori não-informativas, explorando os métodos para selecioná-las.

3.2 - Densidade a Priori de Jeffreys

Jeffreys (1967) propôs uma densidade a priori não-informativa muito utilizada na literatura. Apresentamos uma interpretação estatística para esta densidade a priori, via função de verossimilhança “data translated”, dada por Box e Tiao (1973).

3.2.1 - Densidade a priori localmente uniforme

Denominamos Densidade a Priori Localmente Uniforme a uma densidade a priori dominada pela verossimilhança e cujos valores são, aproximadamente, constantes sobre a região em que a verossimilhança é apreciável e não assume grandes valores fora deste domínio. Portanto, na região onde a densidade a priori $\pi(\phi)$ é dominada pela verossimilhança,

$$\pi(\phi) \propto cte \quad (7)$$

3.2.2 - Verossimilhança “Data Translated”

Quando os dados afetam somente a locação da curva de verossimilhança, $L(\phi | \underline{x})$, sem mudar a sua forma, dizemos que se trata de uma Verossimilhança “Data Translated” (d.t.), ou seja, para diferentes conjuntos de dados, a curva de verossimilhança é transladada no eixo ϕ . Portanto, para diferentes conjuntos de dados, a função de verossimilhança “data translated” é igualmente informativa em relação a ϕ .

Matematicamente, uma verossimilhança d.t. em ϕ deve ser expressa na forma:

$$L(\phi | \underline{x}) = G(\phi - g(\underline{x})), \quad (8)$$

onde $G(y)$ é uma função conhecida, independente dos dados $\underline{x}$, e $g(\underline{x})$ é uma função de $\underline{x}$.

Figura 2: Verossimilhança padronizada: (a) para média Normal θ e (b) para $k = \theta^{-1}$

3.2.3 - Intervalo de verossimilhança de alta densidade

Seja $L(\phi | \underline{x}) = G(\phi - g(\underline{x}))$ uma função de verossimilhança “data translated” e suponhamos que G seja uma função contínua com um único valor máximo $\hat{G}$. Seja α uma constante positiva arbitrária tal que $0 < \alpha < \hat{G}$. Então para dado α , existem c_1 e c_2 reais, $c_1 < c_2$, independente de $\underline{x}$, tal que

$$G(\phi - g(\underline{x})) > \alpha \text{ para } \phi \text{ no intervalo } g(\underline{x}) + c_1 < \phi < g(\underline{x}) + c_2$$

Este intervalo é chamado Intervalo de Verossimilhança de Alta Densidade.

3.2.4 - Densidade a priori não-informativa e verossimilhança “Data Translated”

Uma densidade a priori não-informativa $\pi(\phi)$ caracteriza uma situação onde se tem pouco conhecimento sobre o parâmetro, antes de realizarmos o experimento. “Pouca informação” a priori, porém, somente tem sentido lógico, quando comparada à informação fornecida pelo experimento. O principal problema é então selecionar uma densidade a priori que forneça “pouca informação” em relação à informação esperada pelo experimento.

Bayes (1763) sugeriu utilizarmos uma densidade a priori localmente uniforme,

$$\pi(\phi) \propto cte,$$

que será considerada não-informativa, se a função de verossimilhança for “data translated”.

De fato, suponhamos que a verossimilhança para o parâmetro desconhecido ϕ seja “data translated”, isto é,

$$L(\phi | \underline{x}) = G(\phi - g(\underline{x})).$$

Isto significa que a curva de verossimilhança para ϕ é completamente determinada a priori, antes de coletarmos os dados, exceto a sua locação, a qual depende do conjunto de dados a ser observado. Então, não teríamos, a priori, preferência por uma curva de verossimilhança sobre outra. Desde que, a informação sobre o parâmetro ϕ está contida na verossimilhança, este estado de indiferença local pode ser representado tomando-se intervalos de verossimilhança de alta densidade; tais

intervalos teriam então, probabilidades iguais de conter ϕ . Mas como a verossimilhança é “data translated” para ϕ , os intervalos de verossimilhança de alta densidade têm todos o mesmo tamanho.

Daí, sendo a densidade a priori dominada pela verossimilhança, então a densidade a priori deve ser localmente uniforme,

$$\pi(\phi) \propto cte,$$

e será considerada não-informativa para ϕ com respeito aos dados.

Figura 3: Densidade a Priori Não-Informativa e Verossimilhança padronizada para
a) média Normal θ e b) $k = \theta^{-1}$.

Apresentaremos um argumento para determinar uma densidade a priori não-informativa para o caso da verossimilhança não ser d.t. no parâmetro ϕ .

A solução é determinar uma transformação $\eta(\phi)$ tal que a verossimilhança $L(\eta | \underline{x})$ seja d.t.

Suponhamos $\eta = \eta(\phi)$ uma transformação um a um e diferenciável de ϕ tal que, em termos de η , a verossimilhança $L(\eta | \underline{x})$ seja “data translated”.

Pelas conclusões anteriores, a densidade a priori localmente uniforme $\pi(\eta)$ é não-informativa, isto é

$$\pi(\eta) \propto cte$$

Pelo Teorema da Transformação de variáveis,

$$\pi(\phi) = \pi(\eta) \left\| \frac{\partial \eta}{\partial \phi} \right\|$$

Daí, a densidade a priori não-informativa para ϕ será

$$\pi(\phi) \propto \left\| \frac{\partial \eta}{\partial \phi} \right\|$$

Definição 1: Seja $\eta(\phi)$ uma transformação um a um e diferenciável de ϕ . Diremos que a densidade a priori

$$\pi(\phi) \propto \left\| \frac{\partial \eta}{\partial \phi} \right\| \quad (9)$$

é não-informativa para o parâmetro ϕ se, em termos de η , a verossimilhança $L(\eta | \underline{x})$ for “data translated”.

Observemos que tal argumento somente é possível, se soubermos em qual métrica a verossimilhança correspondente é d.t.

Exemplo 3: Suponhamos X normalmente distribuída com média μ e variância σ^2 .

i) μ desconhecido e σ^2 conhecido:

A função de verossimilhança é dada por:

$$L(\mu \mid \underline{x}, \sigma) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

Observemos que

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\mu - \bar{x})^2,$$

logo

$$L(\mu \mid \underline{x}, \sigma) \propto \exp \left\{ -\frac{n}{2\sigma^2} (\mu - \bar{x})^2 \right\} \quad (10)$$

Neste caso,

$$G(y) = \exp \left\{ -\frac{n}{2\sigma^2} y^2 \right\}, \quad \eta(\mu) = \mu \text{ e } g(\underline{x}) = \bar{x}.$$

Então, $\pi(\mu) \propto \left\| \frac{\partial \mu}{\partial \mu} \right\|$, isto é,

$$\pi(\mu) \propto cte. \quad (11)$$

ii) μ conhecido e σ^2 desconhecido:

A função de verossimilhança é dada por:

$$L(\sigma \mid \underline{x}, \mu) \propto \frac{1}{\sigma^n} \cdot \exp \left\{ -\frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu)^2 \right\},$$

$$L(\sigma \mid \underline{x}, \mu) \propto \exp \left\{ -n \ln \sigma - n \frac{S^2}{\sigma^2} \right\},$$

$$L(\sigma \mid \underline{x}, \mu) \propto \exp \{ -n \ln \sigma - n \ln S + n \ln S + \exp \{ 2n (\ln \sigma - \ln S) \} \},$$

$$L(\sigma \mid \underline{x}, \mu) \propto \exp \{ -n (\ln \sigma - \ln S) + \exp \{ 2n (\ln \sigma - \ln S) \} \} \quad (12)$$

Aqui $G(y) = \exp\{-n.y + \epsilon^{2ny}\}$, $\eta(\sigma) = \ln \sigma$ e $g(\underline{x}) = \ln S$

Então, $\pi(\sigma) \propto \left\| \frac{\partial \ln \sigma}{\partial \sigma} \right\|$, isto é,

$$\pi(\sigma) \propto \frac{1}{\sigma} \quad (13)$$

Um problema que surge, é que nem sempre conhecemos uma transformação $\eta(\phi)$ para qual a verossimilhança seja d.t. em η , isto é, não conseguimos expressar a verossimilhança na forma $L(\eta | \underline{x}) = G(\eta(\phi) - g(\underline{x}))$, portanto não podemos determinar a densidade a priori $\pi(\phi)$ sugerida em (9).

3.2.5 - Verossimilhança aproximadamente “Data Translated” e densidade a priori aproximadamente não-informativa

Conseguir uma transformação $\eta(\phi)$ tal que a verossimilhança $L(\eta | \underline{x})$ seja “exatamente” d.t., isto é, satisfaça (8), não é, em geral, possível.

Observemos então que, se a amostra for “moderamente grande”, qualquer variação na densidade a priori não afeta a posteriori, isto é, a posteriori é insensível ou robusta em relação a densidade a priori. Neste caso, é suficiente obter uma transformação $\eta(\phi)$ para a qual a verossimilhança seja aproximadamente d.t., e com isso obter uma densidade a priori não-informativa, como no caso da verossimilhança d.t. exata. Para isto, aproximamos a verossimilhança por uma Normal.

Obtenção da transformação para a qual a verossimilhança é aproximadamente d.t

$$\text{Seja, } L^*(\phi | \underline{x}) = L(\hat{\phi} | \underline{x}) \cdot \exp \left\{ -\frac{n}{2} \cdot (\phi - \hat{\phi})^2 \cdot \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \phi^2} \right)_{\hat{\phi}} \right\} \quad (14)$$

uma função de ϕ , onde $\hat{\phi}$ é o estimador de máxima verossimilhança de ϕ .

Se $f(x | \phi)$ obedece a certas condições de regularidade, Johnson (1967,1970) mostrou que para n suficientemente grande, a função de verossimilhança de ϕ é aproximadamente Normal, isto é,

$$L(\phi | \underline{x}) \cong L^*(\phi | \underline{x})$$

e permanece aproximadamente Normal sob uma transformação $\eta(\phi)$ um a um diferenciável de ϕ , ou seja,

$$L(\eta | \underline{x}) \cong L^*(\eta | \underline{x}) \quad (15)$$

Seja

$$J(\hat{\eta}) = \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \eta^2} \right)_{\hat{\eta}}, \quad (16)$$

segue que, se escolhermos uma transformação $\eta(\phi)$ tal que $J(\hat{\eta})$ seja uma constante independente de $\hat{\eta}$, então a função $L^*(\eta | \underline{x})$ será exatamente d.t., consequentemente, a verossimilhança $L(\eta | \underline{x})$ será aproximadamente “data translated”, como desejamos.

Observemos que,

$$J(\hat{\eta}) = \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \eta^2} \right)_{\hat{\eta}} = \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \phi^2} \right)_{\hat{\phi}} \cdot \left(\frac{\partial \phi}{\partial \eta} \right)_{\hat{\phi}}^2 = J(\hat{\phi}) \left(\frac{\partial \phi}{\partial \eta} \right)_{\hat{\phi}}^2$$

Isto é,

$$J(\hat{\eta}) = J(\hat{\phi}) \left(\frac{\partial \phi}{\partial \eta} \right)_{\hat{\phi}}^2 \quad (17)$$

Por momento, suponhamos que $J(\hat{\phi})$ seja expressa somente como função

de $\hat{\phi}$. Em geral, $J(\hat{\phi}) > 0$. Portanto, se escolhermos $\eta(\phi)$ tal que

$$\left\| \frac{\partial \eta}{\partial \phi} \right\| \propto J(\hat{\phi})^{1/2} \quad (18)$$

então $J(\hat{\eta})$ será uma constante independente de $\hat{\eta}$, e a função $L^*(\eta \mid \underline{x})$ será exatamente d.t., consequentemente, a verossimilhança $L(\eta \mid \underline{x})$ será aproximadamente “data translated”.

Então, a transformação $\eta(\phi)$ para a qual a verossimilhança $L(\eta \mid \underline{x})$ é aproximadamente d.t. é dada pela relação

$$\frac{\partial \eta}{\partial \phi} \propto J^{1/2}(\phi) \quad \text{ou} \quad \eta \propto \int_{\phi} J(t)^{1/2} dt \quad (19)$$

e a correspondente densidade a priori não-informativa para ϕ será:

$$\pi(\phi) \propto \left\| \frac{\partial \eta}{\partial \phi} \right\| \propto J^{1/2}(\phi) \quad (20)$$

Resultado 1: Se $f(x \mid \phi)$ obedece a certas condições de regularidade e

$$J(\hat{\phi}) = \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \phi^2} \right)_{\hat{\phi}} > 0 \quad (21)$$

é uma constante dependente apenas de $\hat{\phi}$, então a densidade a priori dada em (20) é não-informativa.

É importante observar que, $J(\hat{\phi})$ é uma função somente de $\hat{\phi}$, se as observações $\underline{x}$ forem geradas por uma densidade $f(x \mid \phi)$ da forma

$$f(x \mid \phi) = z(x)w(\phi)\exp\{c(\phi)u(x)\}, \quad (22)$$

onde o domínio de x não depende de ϕ .

3.2.6 - Densidade a Priori de Jeffreys

Em geral, a densidade $f(x | \phi)$ não pertence à família definida por (22) e a quantidade $J(\hat{\phi}) = \left(-\frac{1}{n} \cdot \frac{\partial^2 \ln L}{\partial \phi^2}\right)_{\hat{\phi}}$ não depende apenas de $\hat{\phi}$. Neste caso, o argumento usado para obtermos uma densidade a priori não-informativa como em (20) pode ser modificado como segue.

Motivação:

Notemos que,

$$-\frac{1}{n} \cdot \frac{\partial^2 \ln L}{\partial \phi^2} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln f(x_i | \phi)}{\partial \phi^2}$$

é a média das n funções idênticas de x_i .

Suponhamos que ϕ_0 seja o verdadeiro valor de ϕ tal que x seja gerado pela densidade $f(x | \phi_0)$. Logo, para n grande, a média converge em probabilidade para a esperança da função, isto é,

$$-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln f(x_i | \phi)}{\partial \phi^2} \rightarrow E_{x|\phi_0} \left\{ -\frac{\partial^2 \ln f(x | \phi)}{\partial \phi^2} \right\}$$

assumindo que a esperança existe.

Definindo $a(\phi, \phi_0) = E_{x|\phi_0} \left\{ -\frac{\partial^2 \ln f(x|\phi)}{\partial \phi^2} \right\}$ então $-\frac{1}{n} \cdot \frac{\partial^2 \ln L}{\partial \phi^2} \cong a(\phi, \phi_0)$.

Daí,

$$\left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \phi^2}\right)_{\hat{\phi}} \cong a(\hat{\phi}, \phi_0).$$

Também, para n grande, o estimador de máxima verossimilhança $\hat{\phi}$ converge em probabilidade para ϕ_0 , isto é, podemos escrever aproximadamente,

$$a(\hat{\phi}, \hat{\phi}) \cong a(\hat{\phi}, \phi_0)$$

$$\text{Logo} \quad \left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \phi^2} \right)_{\hat{\phi}} \cong a(\hat{\phi}, \hat{\phi}) = I_1(\hat{\phi}) \quad (23)$$

$$\text{onde} \quad I_1(\phi) = -E \left\{ \frac{\partial^2 \ln f(x | \phi)}{\partial \phi^2} \right\} = E \left\{ \frac{\partial \ln f(x | \phi)}{\partial \phi} \right\}^2 \quad (24)$$

é chamada Medida de Informação de Fisher, associada à uma única observação x .

Obtenção da densidade a priori de Jeffreys

Consequentemente, se usarmos $I_1(\hat{\eta})$, que depende somente de $\hat{\eta}$, para aproximar $\left(-\frac{1}{n} \frac{\partial^2 \ln L}{\partial \eta^2} \right)_{\hat{\eta}}$ em (15), ou seja

$$L(\eta | \underline{x}) \cong L(\hat{\eta} | \underline{x}) \cdot \exp \left\{ -\frac{n}{2} (\eta - \hat{\eta})^2 I_1(\hat{\eta}) \right\} \quad (25)$$

então, argumentando como no item 3.2.5, encontramos que a métrica $\eta(\phi)$ para qual a verossimilhança $L(\eta | \underline{x})$ é aproximadamente d.t. é

$$\frac{\partial \eta}{\partial \phi} \propto [I_1(\phi)]^{1/2} \text{ ou } \eta \propto \int_{\phi} [I_1(t)]^{1/2} dt \quad (26)$$

Equivalentemente, a densidade a priori não informativa para ϕ será escolhida tal que, localmente,

$$\pi(\phi) \propto [I_1(\phi)]^{1/2}. \quad (27)$$

Definição 2: A densidade a priori dada em (27) é não-informativa, chamada Densidade a Priori de Jeffreys.

Observações:

1- Quando a densidade $f(x | \phi)$ é da forma (22) então $J(\hat{\phi}) = I_1(\hat{\phi})$, portanto as densidades a priori dadas em (20) e (27) são idênticas.

2- Não é difícil verificar que a medida de Informação de Fisher,

$$I_n(\phi) = -E \left(\frac{\partial^2 \ln L(\phi | \underline{x})}{\partial \phi^2} \right)$$

para uma amostra $\underline{x} = (x_1, x_2, \dots, x_n)$ é dada por:

$$I_n(\phi) = n \cdot I_1(\phi). \quad (28)$$

Teorema 1: (Invariância) A densidade a priori de Jeffreys é invariante sob transformação um a um e diferenciável.

prova: Suponhamos $\eta = \eta(\phi)$ uma transformação um a um diferenciável, então sabemos que

$$I_1(\eta) = \left(\frac{\partial \phi}{\partial \eta} \right)^2 \cdot I_1(\phi).$$

Se um princípio de escolha conduziu a uma densidade a priori não-informativa $\pi(\phi)$ para ϕ , o mesmo princípio deveria levar

$$\pi(\eta) = \pi(\phi) \cdot \left\| \frac{\partial \phi}{\partial \eta} \right\|$$

como densidade a priori não-informativa para η .

De fato, isto ocorre com a densidade a priori de Jeffreys, como segue

$$\pi(\eta) = \pi(\phi) \cdot \left\| \frac{\partial \phi}{\partial \eta} \right\| \propto (I_1(\phi))^{1/2} \cdot \left\| \frac{\partial \phi}{\partial \eta} \right\| \propto \left(I_1(\eta) \cdot \left(\frac{\partial \eta}{\partial \phi} \right)^2 \right)^{1/2} \cdot \left\| \frac{\partial \phi}{\partial \eta} \right\| \propto (I_1(\eta))^{1/2}.$$

Como trabalharemos com a densidade a priori de Jeffreys para o modelo de Weibull, que é um modelo biparamétrico, é conveniente, então, estendermos a noção da densidade a priori de Jeffreys para modelos multiparamétricos.

3.2.7 - Extensão da densidade a priori de Jeffreys para modelos multiparamétricos

Se a densidade de $\underline{x}$, que depende do vetor paramétrico $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_k)$, obedece a certas condições de regularidade, então, para uma amostra suficientemente grande, a função de verossimilhança para $\underline{\phi}$ se aproxima de uma distribuição Normal Multivariada e permanece Normal para uma transformação $\underline{\eta}(\underline{\phi})$. Logo, argumentando como no caso uniparamétrico,

$$L(\underline{\eta} \mid \underline{x}) \cong L(\hat{\underline{\eta}} \mid \underline{x}) \cdot \exp \left\{ -\frac{n}{2} \cdot (\underline{\eta} - \hat{\underline{\eta}})' \cdot I_n(\hat{\underline{\eta}}) \cdot (\underline{\eta} - \hat{\underline{\eta}}) \right\}, \quad (29)$$

onde $\hat{\underline{\eta}}$ é o estimador de máxima verossimilhança de $\underline{\eta}$ e $I_n(\underline{\eta}) = E \left(-\frac{\partial^2 \ln L}{\partial \eta_i \partial \eta_j} \right)$ é a matriz de Informação de Fisher $k \times k$ associada à amostra $\underline{x}$, dada por

$$I_n(\underline{\eta}) = (I_{ij}(\underline{\eta})) \text{ com } I_{ij}(\underline{\eta}) = E \left(\frac{\partial^2 \ln L(\underline{\eta} \mid \underline{x})}{\partial \eta_i \partial \eta_j} \right) \quad (30)$$

onde a esperança é calculada com respeito à densidade $f(x \mid \underline{\eta})$ dos dados.

Logo, nosso interesse é obter uma transformação $\underline{\eta}(\underline{\phi})$ tal que as regiões de verossimilhança

$$(\underline{\eta} - \hat{\underline{\eta}})' I_n(\hat{\underline{\eta}})(\underline{\eta} - \hat{\underline{\eta}}) \leq cte \quad (31)$$

ou equivalentemente, os contornos de verossimilhança,

$$(\underline{\eta} - \hat{\underline{\eta}})' I_n(\hat{\underline{\eta}})(\underline{\eta} - \hat{\underline{\eta}}) = cte \quad (32)$$

tenham a mesma forma para diferentes $\hat{\underline{\eta}}$, conseqüentemente, a verossimilhança $L(\underline{\eta} | \underline{x})$ será aproximadamente “data translated”. Isto só será possível, se $I_n(\hat{\underline{\eta}})$ for uma matriz de constantes independentes de $\hat{\underline{\eta}}$.

Exemplo 4: Seja $\underline{y} = (y_1, y_2, \dots, y_n)$ um conjunto de v.a.’s. tal que $y_i \sim N\left(\sum_{j=1}^n \theta_j x_{ji}, \sigma^2\right), i = 1, \dots, n$ onde os x ’s são constantes conhecidas. Seja X uma matriz $(X_{ij}), n \times k$ destas constantes. Suponhamos que o vetor paramétrico $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ seja desconhecido e σ^2 conhecido.

Então, a verossimilhança pode ser escrita como,

$$L(\underline{\theta} | \sigma, \underline{y}) \propto \exp\left\{-\frac{1}{2\sigma^2}(\underline{\theta} - \hat{\underline{\theta}})' X' X (\underline{\theta} - \hat{\underline{\theta}})\right\} \quad (33)$$

que está na forma Normal Multivariada centrada em $\hat{\underline{\theta}}$ e tendo matriz de covariância $\sigma^2(X'X)^{-1}$. Portanto, $I_n(\hat{\underline{\theta}}) = X'X$ independe de $\hat{\underline{\theta}}$, conseqüentemente, a verossimilhança é “data translated”.

Os contornos da verossimilhança são da forma

$$(\underline{\theta} - \hat{\underline{\theta}})' X' X (\underline{\theta} - \hat{\underline{\theta}}) = cte \quad (34)$$

e são mostrados pela figura abaixo

Figura 4: Contornos da função de verossimilhança, para diferentes conjuntos de dados, com $k = 2$.

Então, a idéia de pouca informação a priori pode ser expressa pela densidade a priori localmente uniforme

$$\pi(\underline{\theta}) \propto cte. \quad (35)$$

Exemplo 5: Consideremos a escolha de uma densidade a priori em relação à amostra de uma distribuição Normal $N(\theta, \sigma^2)$, onde θ e σ são desconhecidos.

A verossimilhança de (θ, σ) é

$$L(\theta, \sigma \mid \underline{x}) \propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\}$$

que pode ser escrita como,

$$L(\theta, \sigma | \underline{x}) \propto \exp \left\{ -\frac{n}{2} \left(\frac{\theta - \bar{x}}{2s^2} \right)^2 \exp \{ -2(\ln \sigma - \ln s) \} \right\} \\ \exp \left\{ -n(\ln \sigma - \ln s) - \left(\frac{n-1}{2} \right) \exp \{ -2(\ln \sigma - \ln s) \} \right\} \quad (36)$$

A matriz de informação para a transformação $\underline{\eta}(\theta, \sigma) = (\theta, \ln \sigma)$ é

$$I_n(\hat{\theta}, \ln \hat{\sigma}) = n \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & 2 \end{bmatrix} \quad (37)$$

onde um de seus elementos depende de $\hat{\sigma}$, portanto, a verossimilhança não será aproximadamente d.t. para a transformação $\underline{\eta}(\theta, \sigma) = (\theta, \ln \sigma)$ (ver figura 5).

Figura 5: Contornos da função de verossimilhança de $(\theta, \ln \sigma)$ para quatro diferentes amostras com $n = 10$ observações.

Como obter então, uma transformação $\underline{\eta}(\underline{\phi})$, tal que as regiões de verossimilhança

$$(\underline{\eta} - \hat{\underline{\eta}})' \cdot I_n(\hat{\underline{\eta}}) \cdot (\underline{\eta} - \hat{\underline{\eta}}) < cte \quad (38)$$

tenham a mesma forma?

Para obtermos uma transformação $\underline{\eta}(\underline{\phi})$ tal que $I_n(\hat{\underline{\eta}})$ seja uma matriz de constantes independentes de $\hat{\underline{\eta}}$, teríamos que resolver a equação

$$I_n(\underline{\eta}) = A \cdot I_n(\underline{\phi}) \cdot A' \quad (39)$$

onde A é a matriz $k \times k$ das derivadas parciais,

$$A = \left[\frac{\partial(\phi_1, \dots, \phi_k)}{\partial(\eta_1, \dots, \eta_k)} \right]$$

cuja solução é complicada.

Como solução para este problema, Jeffreys sugere procurarmos uma transformação $\underline{\eta}(\underline{\phi})$ para qual as regiões dadas em (38) tenham volumes iguais para diferentes $\hat{\underline{\eta}}$. Desde que, $\|I_n(\hat{\underline{\eta}})\|^{1/2}$ mede o volume da região de verossimilhança (38), a exigência acima é equivalente a procurar uma transformação para qual $\|I_n(\hat{\underline{\eta}})\|$ seja independente de $\hat{\underline{\eta}}$.

Para obtermos uma transformação $\underline{\eta}$ que satisfaça esta exigência, primeiramente notemos que, de (39)

$$\|I_n(\underline{\eta})\| = \|A\|^2 \cdot \|I_n(\underline{\phi})\| \quad (40)$$

Portanto, $\|I_n(\hat{\underline{\eta}})\|$ será uma constante independente de $\hat{\underline{\eta}}$, se $\underline{\eta}$ for tal que,

$$\left\| \frac{\partial(\phi_1, \dots, \phi_k)}{\partial(\eta_1, \dots, \eta_k)} \right\| \propto \|I_n(\underline{\phi})\|^{-1/2} \quad (41)$$

consequentemente, a densidade a priori não informativa $\pi(\underline{\eta})$ será localmente uniforme para $\underline{\eta}$, e a correspondente densidade a priori não-informativa de Jeffreys para $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_k)$ será

$$\pi(\underline{\phi}) \propto \|I_n(\underline{\phi})\|^{1/2} \quad (42)$$

Observação: A versão multiparamétrica da densidade a priori de Jeffreys que conduz a $\pi(\underline{\phi}) \propto \|I_n(\underline{\phi})\|^{1/2}$ pode levar a uma transformação na qual a verossimilhança não seja transladada pelos dados. Para efeito ilustrativo, a figura (6) mostra que regiões de verossimilhança podem ter o mesmo volume e ainda assim, serem de formas muito diferentes.

Figura 6: Regiões de verossimilhança de diferentes formas tendo o mesmo volume.

Uma dificuldade adicional para utilização da densidade a priori de Jeffreys para modelos multiparamétricos surge, quando parâmetros de diferentes espécies são considerados simultaneamente. Neste caso, pode não ser interessante procurarmos uma transformação que produza regiões de verossimilhança com mesmo volume.

Exemplo 6: Suponhamos que os parâmetros Localização-Escala de uma distribuição Normal sejam considerados desconhecidos simultaneamente.

Como visto anteriormente (ver figura 6), para a transformação $\underline{\eta}(\theta, \sigma) = (\theta, \ln \sigma)$, as regiões de verossimilhança não têm formas iguais para diferentes $\underline{\eta}(\hat{\theta}, \ln \hat{\sigma})$ e nem os volumes dessas regiões são iguais. De fato,

$$\|I_n(\hat{\theta}, \ln \hat{\sigma})\| \propto \frac{1}{\hat{\sigma}^2} \quad (43)$$

que não é independente de $\hat{\sigma}^2$.

Para esta transformação obtemos uma densidade a priori, aproximadamente, localmente uniforme (ver Box e Tiao), ou seja

$$\pi(\theta, \ln \sigma) \propto cte \quad (44)$$

Consequentemente,

$$\pi(\theta, \sigma) \propto \frac{1}{\sigma} \quad (45)$$

Utilizando-se, porém, a densidade a priori de Jeffreys para (θ, σ) obtemos,

$$\pi(\theta, \sigma) \propto \frac{1}{\sigma^2} \quad (46)$$

pois,

$$I_n(\theta, \sigma) = n \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}. \quad (47)$$

A densidade a priori dada em (46) não leva em consideração a independência dos parâmetros. Para casos, como os parâmetros localização-escala θ e σ da Normal, Jeffreys propõe incorporarmos a informação, a priori, de independência dos parâmetros, obtendo, para este exemplo, a densidade a priori (45).

Observação: Como no caso uniparamétrico, a densidade a priori de Jeffreys, para vários parâmetros, também é invariante sob transformações um a um diferenciáveis.

3.2.8 - Obtenção da densidade a priori de Jeffreys para o modelo de Weibull

Determinemos a matriz de Informação de Fisher para os parâmetros do modelo de Weibull.

Neste caso a matriz é determinada da seguinte maneira:

$$I_1(p, \theta) = \begin{bmatrix} -E \left\{ \frac{\partial^2}{\partial p^2} \ln f \right\} & -E \left\{ \frac{\partial^2}{\partial p \partial \theta} \ln f \right\} \\ -E \left\{ \frac{\partial^2}{\partial p \partial \theta} \ln f \right\} & -E \left\{ \frac{\partial^2}{\partial \theta^2} \ln f \right\} \end{bmatrix} \quad (48)$$

onde os elementos da matriz são calculados abaixo:

$$\ln f(x | p, \theta) = \ln p - \ln \theta + (p-1) \ln x - \frac{x^p}{\theta}$$

$$\frac{\partial^2}{\partial p^2} \ln f(x | p, \theta) = -\frac{1}{p^2} - \frac{x^p}{\theta} (\ln x)^2$$

$$\frac{\partial^2}{\partial \theta^2} \ln f(x | p, \theta) = \frac{1}{\theta^2} - \frac{2x^p}{\theta^3}$$

$$\frac{\partial^2}{\partial p \partial \theta} \ln f(x | p, \theta) = \frac{x^p}{\theta^2} \ln x$$

Para simplificar as integrações, façamos $u = \frac{x^p}{\theta}$ e utilizemos a função definida por

$$\psi^{(i)}(n) = \int_0^\infty e^{-u} \cdot u^{n-1} (\ln u)^{i+1} \cdot du \quad i = 0, 1, 2, \dots, \quad (49)$$

onde

$$\psi^{(0)} = \psi \quad \text{e} \quad \psi(n) = \frac{1}{n-1} + \psi(n-1), \quad n > 1. \text{ Logo,}$$

$$\begin{aligned} -E \left\{ \frac{\partial^2}{\partial p^2} \ln f \right\} &= \int_0^\infty \left(\frac{\partial^2}{\partial p^2} \ln f \right) \cdot f(x | p, \theta) dx = \int_0^\infty \left\{ \frac{1}{p^2} + \frac{x^p}{\theta} (\ln x)^2 \right\} \cdot \frac{p}{\theta} \cdot x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx \\ &= \frac{1}{p^2} \cdot \int_0^\infty \frac{p}{\theta} x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx + \int_0^\infty \frac{x^p}{\theta} \cdot (\ln x)^2 \cdot \frac{p}{\theta} x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx \\ &= \frac{1}{p^2} \left\{ 1 + (\ln \theta)^2 \int_0^\infty u \cdot e^{-u} du + 2 \ln \theta \cdot \int_0^\infty u \cdot \ln u \cdot e^{-u} du + \int_0^\infty u \cdot (\ln u)^2 \cdot e^{-u} du \right\}, \\ -E \left\{ \frac{\partial^2}{\partial p^2} \ln f \right\} &= \frac{1}{p^2} \left(1 + (\ln \theta)^2 + 2\psi(2) \ln \theta + \psi'(2) \right) \end{aligned} \quad (50)$$

Analogamente, em relação a θ temos:

$$\begin{aligned} -E \left\{ \frac{\partial^2}{\partial \theta^2} \ln f \right\} &= \int_0^\infty \left(\frac{\partial^2}{\partial \theta^2} \ln f \right) \cdot f(x | p, \theta) dx = \int_0^\infty \left\{ -\frac{1}{\theta^2} + \frac{2x^p}{\theta^3} \right\} \cdot \frac{p}{\theta} \cdot x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx \\ &= -\frac{1}{\theta^2} \int_0^\infty \frac{p}{\theta} x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx + \int_0^\infty \frac{2x^p}{\theta^3} \cdot \frac{p}{\theta} x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx, \\ -E \left\{ \frac{\partial^2}{\partial \theta^2} \ln f \right\} &= -\frac{1}{\theta^2} + \frac{2}{\theta^2} \int_0^\infty u e^{-u} du = \frac{1}{\theta^2}. \end{aligned} \quad (51)$$

E em relação a p e θ temos:

$$\begin{aligned} -E \left\{ \frac{\partial^2}{\partial p \partial \theta} \ln f \right\} &= -\int_0^\infty \left(\frac{\partial^2}{\partial p \partial \theta} \ln f \right) \cdot f(x | p, \theta) dx = \int_0^\infty \frac{x^p}{\theta^2} \ln x \cdot \frac{p}{\theta} x^{p-1} \cdot \exp \left\{ -\frac{x^p}{\theta} \right\} dx \\ -E \left\{ \frac{\partial^2}{\partial p \partial \theta} \ln f \right\} &= -\frac{1}{p\theta} \left(\ln \theta \cdot \int_0^\infty u \cdot e^{-u} du + \int_0^\infty u \cdot \ln u \cdot du \right) = -\frac{1}{p\theta} (\ln \theta + \psi(2)) \end{aligned}$$

Portanto:

$$I_1(p, \theta) = \begin{bmatrix} \frac{1}{p^2} (1 + (\ln \theta)^2 + 2\psi(2)\ln \theta + \psi'(2)) & -\frac{1}{p\theta} (\ln \theta + \psi(2)) \\ -\frac{1}{p\theta} (\ln \theta + \psi(2)) & \frac{1}{\theta^2} \end{bmatrix} \quad (52)$$

Observemos que os parâmetros p e θ não são ortogonais, pois, a matriz de Fisher não é diagonal (Cox e Reid, 1987).

A densidade a priori de Jeffreys, para os parâmetros do modelo de Weibull, obtida por:

$$\pi(p, \theta) \propto \|I_1(p, \theta)\|^{1/2},$$

é então

$$\pi(p, \theta) \propto \left\{ \frac{1}{\theta^2 p^2} (1 + \psi'(2) - \psi^2(2)) \right\}^{1/2},$$

Portanto

$$\pi(p, \theta) \propto \frac{1}{p \cdot \theta} \quad (53)$$

3.3 - Densidade a Priori de Zellner (MDIP)

3.3.1 - Introdução

A idéia básica da densidade a priori de Zellner é maximizar a informação proveniente dos dados, em relação à informação conhecida, a priori, sobre o parâmetro. Para medir esta informação, Zellner utiliza a medida

$$H(X) = \int p(x) \ln p(x) dx, \quad (54)$$

onde $p(x)$ é a f.d.p. da v.a. X .

Shannon (1948) define $U = -H$ como a Entropia Negativa da v.a. X .

Seja X uma variável aleatória com f.d.p. $f(x | \phi)$, $x \in R_x$ ($R_x \subseteq \mathbb{R}$), $\phi \in [a, b]$ é o parâmetro e $\pi(\phi)$ a densidade a priori a ser determinada.

Definição 3: A informação a priori sobre o parâmetro ϕ , antes da realização do experimento, é dada por:

$$H(\phi) = \int_a^b \pi(\phi) \ln \pi(\phi) d\phi, \quad (55)$$

que será denominada “Informação Passada”.

$H(\phi)$ é interpretada como a quantidade de informação, sobre o parâmetro ϕ , fornecida pela v.a. ϕ .

Determinemos a medida que representa a informação proveniente dos dados, isto é, após realizarmos o experimento.

Definição 4: Suponhamos que $\phi = \phi_0$ seja um valor observado para o parâmetro, então a quantidade de informação fornecida pela v.a. X , isto é, pelos dados, é dada por

$$H_{\phi_0}(X) = \int_{R_x} f(x | \phi_0) \ln f(x | \phi_0) dx. \quad (56)$$

Como não conhecemos o verdadeiro valor de ϕ , tomamos a média das medidas $H_{\phi}(X)$ obtidas para cada valor ϕ fixado. Obtemos assim, o quanto de informação, em média, a variável aleatória X (os dados) informa sobre o parâmetro ϕ . Esta informação será medida por

$$E(H_{\phi}(X)) = \int_a^b H_{\phi}(X) \pi(\phi) d\phi \quad (57)$$

que será denominada “Informação Presente”.

A seguir mostraremos que a medida global (total) da informação sobre ϕ é a soma da “informação passada” com a “informação presente”.

Definição 5: Se $P(x, \phi)$ é a f.d.p. conjunta de X e ϕ , então a medida de informação global fornecida pelo vetor aleatório (X, ϕ) é dada por:

$$H(X, \phi) = \int_a^b \int_{R_x} P(x, \phi) \ln P(x, \phi) dx d\phi \quad (58)$$

Lema 1

$$H(X, \phi) = E(H_{\phi}(X)) + H(\phi) \quad (59)$$

prova: A f.d.p. conjunta $P(x, \phi)$ pode ser escrita como:

$$P(x, \phi) = \pi(\phi) f(x | \phi),$$

daí, substituindo $P(x, \phi)$ em (58) obtemos

$$H(X, \phi) = \int_a^b \int_{R_x} \pi(\phi) f(x | \phi) \ln (\pi(\phi) f(x | \phi)) dx d\phi.$$

Desenvolvendo o logarítmo, temos:

$$H(X, \phi) = \int_a^b \int_{R_x} \{f(x | \phi) \ln f(x | \phi) dx\} \pi(\phi) d\phi + \int_a^b \left\{ \int_{R_x} f(x | \phi) dx \right\} \pi(\phi) \ln \pi(\phi) d\phi,$$

Daí,

$$H(X, \phi) = \int_a^b H_\phi(X) \pi(\phi) d\phi + \int_a^b \pi(\phi) \ln \pi(\phi) d\phi$$

□

Observação: Pelo Lema 1, temos que

$$\int_a^b \int_{R_x} P(x, \phi) \ln P(x, \phi) dx d\phi = \int_a^b H_\phi(X) \cdot \pi(\phi) d\phi + \int_a^b \pi(\phi) \ln \pi(\phi) d\phi. \quad (60)$$

$H(X, \phi)$ é a soma da “informação passada”, $H(\phi)$, com a “informação presente”, $E(H_\phi(X))$. Daí, observemos que $H(X, \phi) \geq H(\phi)$. Isto era de se esperar uma vez que estamos acrescentando a informação dos dados.

Com isso, o ganho na informação fornecida pelos dados através do experimento sobre a informação a priori do parâmetro é:

$$G = E(H_\phi(X)) - H(\phi). \quad (61)$$

Uma fórmula análoga a (61) é,

$$G = \int_a^b H_\phi(X) \pi(\phi) d\phi - \int_a^b \pi(\phi) \ln \pi(\phi) d\phi, \quad (62)$$

conhecida como Medida de Informação de Lindley.

É de nosso interesse que os dados informem mais a respeito do parâmetro do que a informação a priori. Caso isto não ocorresse, não haveria justificativa para a realização do experimento.

Desejamos obter uma densidade a priori $\pi(\phi)$ que maximize G , isto é, torne o ganho na informação fornecida pelos dados sobre a informação a priori do parâmetro o maior possível, portanto, maximize a informação dos dados.

Definição 6: A densidade a priori de Zellner, também chamada de “Maximal Data Information Prior” (MDIP), é uma f.d.p., $\pi(\phi)$, que maximiza G , definida em (62), na variável $\pi(\phi)$ sujeito à condição:

$$\int_a^b \pi(\phi) d\phi = 1 \quad (63)$$

A densidade a Priori de Zellner é aquela que fornece menos informação, a respeito do parâmetro, comparada à informação proveniente dos dados.

O seguinte teorema proporciona uma fórmula simples para a densidade a priori de Zellner.

Teorema 2: A densidade a priori de Zellner normalizada é dada por:

$$\pi(\phi) = c \exp \{H_\phi(X)\} \quad a \leq \phi \leq b, \quad (64)$$

onde $c^{-1} = \int_a^b \exp \{H_\phi(X)\} d\phi$ é a constante normalizadora.

prova: Temos que maximizar a função $U = G - \lambda \cdot \left\{ \int_a^b \pi(\phi) d\phi - 1 \right\}$ onde λ é o multiplicador de Lagrange.

$$\text{Então, } \frac{\partial U}{\partial \pi} = 0 \text{ se, e somente se, } H_\phi(X) - \ln \pi(\phi) - 1 - \lambda = 0$$

e a solução é dada por,

$$\pi(\phi) = c \exp \{ H_\phi(X) \}.$$

□

Observações:

1 - Se $H_\phi(X) = cte$ então $\pi(\phi)$ é uma densidade uniforme.

2 - Para o caso em que a f.d.p. dos dados envolve mais do que um parâmetro, digamos um vetor de parâmetros $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_k)$, G é definida como segue:

$$G = \int \int_{R_{\underline{\phi}}} \dots \int H_{\underline{\phi}}(X) \pi(\underline{\phi}) d\phi_1 d\phi_2 \dots d\phi_k - \int \int_{R_{\underline{\phi}}} \dots \int \pi(\underline{\phi}) \ln \pi(\underline{\phi}) d\phi_1 d\phi_2 \dots d\phi_k \quad (65)$$

onde $R_{\underline{\phi}}$ é uma região finita para $\underline{\phi}$ e

$$H_{\underline{\phi}}(X) = \int_{R_x} f(x | \underline{\phi}) \ln f(x | \underline{\phi}) dx. \quad (66)$$

Procedendo como no Teorema 2, a densidade a priori de Zellner será

$$\pi(\underline{\phi}) = c \exp \{ H_{\underline{\phi}}(X) \} \quad \underline{\phi} \in R_{\underline{\phi}}, \quad (67)$$

onde c é uma constante normalizadora.

3- Se outras condições são impostas, refletindo uma dada informação inicial, por exemplo $\int \phi^i \pi(\phi) d\phi = \mu_i$, $i = 1, 2, \dots, m$, então a densidade a priori de Zellner é dada por:

$$\pi(\phi) \propto \exp \left\{ H_\phi(X) + \lambda_1 \phi + \lambda_2 \phi^2 + \dots + \lambda_m \phi^m \right\}, \quad (68)$$

onde λ_i 's são os multiplicadores de Lagrange, os quais podem ser obtidos, a partir dos valores dos momentos a priori μ_i , $i = 1, 2, \dots, m$.

Se, por exemplo, $m = 2$ e $H_\phi(X) = cte$, então $\pi(\phi)$ tem uma forma normal.

4- É possível obter a densidade a priori de Jeffreys, através da formulação proposta por Zellner. Zellner (1993) observou que, substituindo a entropia $H_\phi(X)$ por $\ln \|I(\phi)\|^{1/2}$ no funcional (62), obteríamos a densidade a priori proposta por Jeffreys. Para se chegar a este resultado, bastaria então, maximizar

$$G_J[\pi(\phi)] = \int \ln \|I(\phi)\|^{1/2} \pi(\phi) d\phi - \int \pi(\phi) \ln \pi(\phi) d\phi \quad (69)$$

com respeito a $\pi(\phi)$, sujeito a $\int \pi(\phi) d\phi = 1$, obtendo como solução,

$$\pi_J(\phi) \propto \|I(\phi)\|^{1/2}. \quad (70)$$

Esta densidade a priori tem a forma da densidade a priori de Jeffreys, porém, com a vantagem de ser uma densidade a priori própria. Isto se contrasta com a densidade a priori de Jeffreys, que em muitos casos, é imprópria.

5- Há portanto, uma questão em aberto, relacionada a qual medida de informação deva ser usada, $H_\phi(X)$ ou $\ln \|I(\phi)\|^{1/2}$. Para contornar este problema, Rodriguez (1991 a, b) sugere uma atitude conciliatória, combinando a medida de entropia e a medida de Fisher, bastando tomar a média ponderada das medidas $H_\phi(X)$ e $\ln \|I(\phi)\|^{1/2}$ obtendo, assim, o seguinte critério funcional,

$$G_R[\pi(\phi)] = \int \left(w H_\phi(X) + (1 - w) \ln \|I(\phi)\|^{1/2} \right) \pi(\phi) d\phi - \int \pi(\phi) \ln \pi(\phi) d\phi \quad (71)$$

com $0 \leq w \leq 1$

Maximizando G_R , com a condição de $\pi(\phi)$ ser própria, resulta a seguinte densidade a priori:

$$\pi_R(\phi) \propto \|I(\phi)\|^{(1-w)/2} e^{w H_\phi(X)}, \quad (72)$$

que é uma média geométrica ponderada da densidade a priori de Jeffreys, $\|I(\phi)\|^{1/2}$, e da densidade a priori de Zellner, $\exp \{H_\phi(X)\}$.

Veja Rodriguez (1991 a, b) para maiores detalhes, que a densidade a priori (72) tem as mesmas propriedades de invariância da densidade a priori de Jeffreys.

6- A densidade a priori proposta por Zellner em (64) possui propriedades limitadas de invariância, sendo invariante apenas sob transformações lineares de X e ϕ , e não para toda transformação diferenciável um a um. Ver a seguir caso do modelo de Weibull.

3.3.2 Obtenção da Densidade a Priori de Zellner para o modelo de Weibull

Para o modelo de Weibull, temos que:

$$H_{(p,\theta)}(X) = \int_0^\infty \frac{p}{\theta} x^{p-1} \exp\left\{-\frac{x^p}{\theta}\right\} \left(\ln p - \ln \theta + (p-1) \ln x - \frac{x^p}{\theta}\right) dx. \quad (73)$$

Para o cálculo da integral acima, façamos $u = \frac{x^p}{\theta}$. Logo,

$$H_{(p,\theta)}(X) = \ln p - \ln \theta + (p-1) \frac{1}{p} \int_0^\infty (\ln \theta + \ln u) e^{-u} du - \int_0^\infty u e^{-u} du,$$

$$H_{(p,\theta)}(X) = \ln p - \frac{1}{p} \ln \theta + \frac{(p-1)}{p} \psi(1) - 1.$$

Portanto,

$$H_{(p,\theta)}(X) = \ln (p \cdot \theta^{-1/p}) - \frac{1}{p} \psi(1) + \psi(1) - 1. \quad (74)$$

Concluindo, a densidade a priori de Zellner, para os parâmetros p e θ , é dada por:

$$\pi(p, \theta) \propto p \theta^{-1/p} \exp\left\{-\frac{1}{p} \psi(1)\right\} \quad (75)$$

3.4 - Densidade a Priori de Tibshirani

3.4.1 - Motivação

Welch e Peers (1963) mostram que, para um modelo uniparamétrico, a densidade a priori de Jeffreys $\pi(\phi)$ produz um intervalo a posteriori unimodal para ϕ tendo cobertura fixada a menos de $O(n^{-1})$, contudo, para um modelo multiparamétrico sob a presença de parâmetros “nuisances”, a densidade a priori de Jeffreys não produz, necessariamente, intervalos a posteriori com essa probabilidade de cobertura, para todo funcional $\delta = t(\underline{\phi})$, como mostra o seguinte exemplo:

Exemplo 7: Sejam X_1 e X_2 duas variáveis independentes, com distribuição Normal $N(\theta_1, 1)$ e $N(\theta_2, 1)$, respectivamente e $\delta = t(\theta_1, \theta_2) = \theta_1 \cdot \theta_2$ o parâmetro de interesse.

Suponhamos que sejam retiradas uma amostra $\underline{x}_1$ de n_1 observações da população $N(\theta_1, 1)$ e uma amostra $\underline{x}_2$ de n_2 observações da população $N(\theta_2, 1)$, independentemente.

A densidade a priori de Jeffreys $\pi(\theta_1, \theta_2)$ é localmente uniforme.

Consideremos a transformação um a um e diferenciável $(\theta_1, \theta_2) \rightarrow (\delta, \lambda)$ onde $\delta = \theta_1 \cdot \theta_2$ e $\lambda = \theta_2$. Então, pela propriedade de invariância da densidade a priori de Jeffreys, obtemos

$$\pi(\delta, \lambda) \propto \frac{1}{\lambda} \quad (76)$$

A densidade a posteriori conjunta para os novos parâmetros é dada por,

$$p(\delta, \lambda \mid \underline{x}) \propto \frac{1}{\lambda} \cdot \exp \left\{ -\frac{1}{2} \left(\left(\frac{\delta}{\lambda} - \bar{x}_1 \right)^2 + (\lambda - \bar{x}_2)^2 \right) \right\} \quad (77)$$

Daí, temos que a densidade a posteriori marginal para o parâmetro δ é

$$p(\delta \mid \underline{x}) \propto \int_0^\infty \frac{1}{\lambda} \cdot \exp \left\{ -\frac{1}{2} \left(\left(\frac{\delta}{\lambda} - \bar{x}_1 \right)^2 + (\lambda - \bar{x}_2)^2 \right) \right\} d\lambda \quad (78)$$

Aplicando-se a Aproximação de Laplace (Apêndice A) em (78), obtemos uma forma fechada para a densidade a posteriori. Daí, a densidade a posteriori marginal normalizada para o parâmetro de interesse é então,

$$p(\delta \mid \underline{x}) = \frac{1}{\sqrt{2\pi\bar{x}_2}} \cdot \exp \left\{ -\frac{1}{2\bar{x}_2^2} (\delta - \bar{x}_1\bar{x}_2)^2 \right\}, \quad (79)$$

isto é, $\delta \mid \underline{x} \sim N(\bar{x}_1\bar{x}_2, \bar{x}_2^2)$.

A figura 7 nos mostra que a densidade a priori de Jeffreys para (θ_1, θ_2) produz uma densidade a posteriori, concentrando maior massa onde $\|\theta_1\theta_2\|$ é pequeno. Portanto, os intervalos a posteriori para $\delta = \theta_1.\theta_2$ serão muito pequenos, consequentemente, a probabilidade de cobertura será bem menor que α , a medida que $\delta \rightarrow 0$, isto é, está nos informando de que δ está próximo de zero. Esta informação só pode ser proveniente da densidade a priori, pois a verossimilhança é d.t. Logo, a densidade a priori é informativa.

Figura 7: Distribuição a posteriori para $\delta = \theta_1 \theta_2$ para diferentes conjuntos de dados.

Tibshirani propõe um procedimento para obtenção de uma densidade a priori não-informativa para um vetor paramétrico $\underline{\phi}$ tal que, para um coeficiente fixado α , o intervalo para o parâmetro de interesse $\delta = t(\underline{\phi})$, tenha probabilidade de cobertura aproximadamente α .

3.4.2- Introdução

Seja $\underline{X} = (X_1, X_2, \dots, X_n)$ um vetor aleatório, onde X_i 's são v.as i. i. d. com fdp $f(x_i | \underline{\phi})$ e $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_k)$ um vetor paramétrico. Suponhamos que o parâmetro δ , obtido pela transformação $\delta = t(\underline{\phi})$, seja o parâmetro de interesse.

A idéia básica é construir uma densidade a priori “não-informativa” $\pi(\underline{\phi})$ tal que, fixado um coeficiente α , por exemplo 90%, 95% ou 99%, o intervalo de credibilidade $\alpha\%$, $S_\alpha(\underline{X})$, para o parâmetro δ , obtido da seguinte forma

$$Pr_{\delta|\underline{x}} \{\delta \in S_\alpha(\underline{X})\} = \int_{s_\alpha(\underline{x})} p(\delta | \underline{x}).d\delta = \alpha \quad (80)$$

tenha probabilidade de cobertura aproximadamente α , isto é,

$$Pr_{\underline{x}|\delta} \{\delta \in S_\alpha(\underline{X})\} = \int_{s_\alpha(\underline{x})} f(x | \delta).dx = \alpha + o(n^{-1}) \quad (81)$$

para todo δ .

Definição 5: Dizemos que uma densidade a priori qualquer $\pi(\underline{\phi})$ é não-informativa, no sentido dado por Tibshirani (1989), se e somente se, $\pi(\underline{\phi})$ produz intervalos de credibilidade para o parâmetro δ com probabilidade de cobertura independente de δ .

Há pelo menos duas razões para utilizarmos este procedimento:

- (a) proporciona um método, para obter intervalos de confiança;
- (b) devido a forma como $\pi(\underline{\phi})$ foi construída, ela será útil para comparações dos resultados obtidos na Inferência Bayesiana e na Inferência Clássica.

Notemos que a condição (80) é dada em termos Bayesianos e a condição (81) em termos Frequentistas. A condição (81) garante que a densidade a posteriori de δ seja mais aberta. De fato, a probabilidade de cobertura do intervalo de credibilidade $S_\alpha(\underline{X})$ para δ , sendo próxima do α fixado, implica em um intervalo grande, conseqüentemente, em uma densidade a posteriori de δ mais aberta.

Stein (1985) determinou uma condição suficiente, para se obter uma densidade a priori $\pi(\underline{\phi})$ que proporcione um intervalo a posteriori para δ com probabi-

lidade de cobertura α a menos de $O(n^{-1})$, obtendo para o exemplo 7,

$$\pi(\theta_1, \theta_2) \propto (\theta_1^2 + \theta_2^2)^{1/2} \quad (82)$$

Tibshirani (1989) utilizou a condição de Stein, explorando a noção de parâmetros ortogonais, para produzir um método simples de se obter $\pi(\underline{\phi})$. Especificamente, ele mostra através de uma parametrização ortogonal, que a classe das densidades a priori, satisfazendo a condição de Stein, é proporcional à raiz quadrada da informação de Fisher do parâmetro de interesse vezes uma função arbitrária dos parâmetros “nuisance”.

3.4.3- Construção da densidade a priori de Tibshirani

Sejam $I(\underline{\phi})$ a Matriz de Informação de Fisher para $\underline{\phi}$ e

$$\nabla t(\underline{\phi}) = \left(\frac{\partial t(\underline{\phi})}{\partial \phi_1}, \frac{\partial t(\underline{\phi})}{\partial \phi_2}, \dots, \frac{\partial t(\underline{\phi})}{\partial \phi_k} \right) \quad (83)$$

o gradiente de $t(\underline{\phi})$. Para simplicidade dos cálculos, Tibshirani sugere usarmos o vetor padronizado $\underline{\eta}(\underline{\phi})$, onde

$$\underline{\eta}(\underline{\phi}) = (\eta_1(\underline{\phi}), \eta_2(\underline{\phi}), \dots, \eta_k(\underline{\phi}))^t = k \cdot I(\underline{\phi})^{-1/2} \cdot \nabla t(\underline{\phi}) \quad (84)$$

e k é um número tal que,

$$\underline{\eta}(\underline{\phi})^t \cdot I(\underline{\phi}) \cdot \underline{\eta}(\underline{\phi}) = 1, \text{ para todo } \underline{\phi} \quad (85)$$

Seja α um número, $0 \leq \alpha \leq 1$ e para α fixado, seja

$$S_\alpha(\underline{X}) = \{ \underline{\phi} \text{ tal que } \underline{\eta}(\hat{\underline{\phi}})^t \cdot I(\hat{\underline{\phi}})^{1/2} \cdot (\underline{\phi} - \hat{\underline{\phi}}) < \Phi^{-1}(\alpha) \} \quad (86)$$

uma região a posteriori normal para $t(\underline{\phi})$, onde $\underline{\phi}$ é a moda da densidade a posteriori para $\underline{\phi}$ e $\Phi^{-1}(\alpha)$ é o α^0 percentil da distribuição normal padronizada.

Desejamos obter uma densidade a priori $\pi(\underline{\phi})$ tal que

$$Pr_{\underline{\phi}|\underline{x}} \{ \underline{\phi} \in S_\alpha(\underline{X}) \} = \alpha \quad (87)$$

Stein (1985) afirma que

$$Pr_{\underline{x}|\underline{\phi}} \{ \underline{\phi} \in S_\alpha(\underline{X}) \} = \alpha + O(n^{-1}) \quad (88)$$

para todo $\underline{\phi}$, e obtém a seguinte condição suficiente para $\pi(\underline{\phi})$ satisfazer (87)

$$\pi(\underline{\phi}) \sum_{i=1}^k \frac{\partial \eta_i(\underline{\phi})}{\partial \phi_i} + \sum_{i=1}^k \eta_i(\underline{\phi}) \frac{\partial \pi(\underline{\phi})}{\partial \phi_i} = 0. \quad (89)$$

Segundo Stein, esta condição pode ser também necessária, embora ele não prove isto.

Com o intuito de se obter uma forma simples para $\pi(\underline{\phi})$, Tibshirani propõe usarmos uma reparametrização ortogonal do parâmetro $\underline{\phi}$ por $(\delta, \underline{\lambda})$, onde $\delta = t(\underline{\phi})$ é o parâmetro de interesse e $\underline{\lambda} = (\lambda_1, \dots, \lambda_{k-1})$ é o parâmetro “nuisance”. Logo, a matriz de Informação de Fisher $k \times k$ é dada por:

$$I(\delta, \underline{\lambda}) = \begin{bmatrix} I_{\delta\delta}(\delta, \underline{\lambda}) & 0 \\ 0 & I_{\underline{\lambda}\underline{\lambda}}(\delta, \underline{\lambda}) \end{bmatrix} \quad (90)$$

Portanto,

$$\underline{\eta}(\delta, \underline{\lambda}) = (\eta_1(\delta, \underline{\lambda}), \eta_2(\delta, \underline{\lambda}), \dots, \eta_k(\delta, \underline{\lambda}))^t \propto I(\delta, \underline{\lambda})^{-1/2} \cdot \nabla \delta, \quad (91)$$

onde $\nabla t(\underline{\phi}) = \nabla \delta = (\frac{\partial \delta}{\partial \delta}, \frac{\partial \delta}{\partial \lambda_1}, \dots, \frac{\partial \delta}{\partial \lambda_{k-1}}) = (1, 0, 0, \dots, 0)$. Segue que

$$\underline{\eta}(\delta, \underline{\lambda}) \propto (I_{\delta\delta}(\delta, \underline{\lambda})^{-1/2}, 0, 0, \dots, 0)^t \quad (92)$$

Supondo que $\pi(\delta, \underline{\lambda})$ satisfaz a condição de Stein (89), temos que,

$$\pi(\delta, \underline{\lambda}) \cdot \frac{\partial \eta_1(\delta, \underline{\lambda})}{\partial \delta} + \pi(\delta, \underline{\lambda}) \sum_{i=1}^{k-1} \frac{\partial \eta_{i+1}(\delta, \underline{\lambda})}{\partial \lambda_i} + \eta_1(\delta, \underline{\lambda}) \frac{\partial \pi(\delta, \underline{\lambda})}{\partial \delta} + \sum_{i=1}^{k-1} \eta_{i+1}(\delta, \underline{\lambda}) \frac{\partial \pi(\delta, \underline{\lambda})}{\partial \lambda_i} = 0$$

A condição de Stein determina completamente $\pi(\delta, \underline{\lambda})$ somente na direção de δ (ver (92)).

Dai, $\pi(\delta, \underline{\lambda}), \frac{\partial I_{\delta\delta}(\delta, \underline{\lambda})^{-1/2}}{\partial \delta} + I_{\delta\delta}(\delta, \underline{\lambda})^{-1/2} \frac{\partial \pi(\delta, \underline{\lambda})}{\partial \delta} \propto 0$, ou

$$\frac{1}{\pi(\delta, \underline{\lambda})} \frac{\partial \pi(\delta, \underline{\lambda})}{\partial \delta} \propto - \frac{1}{I_{\delta\delta}(\delta, \underline{\lambda})^{-1/2}} \cdot \frac{\partial I_{\delta\delta}(\delta, \underline{\lambda})^{-1/2}}{\partial \psi} \quad (93)$$

Resolvendo em π a equação diferencial acima, temos que

$$\pi(\delta, \underline{\lambda}) \propto g(\underline{\lambda}) \cdot I_{\delta\delta}(\delta, \underline{\lambda})^{1/2}, \quad (94)$$

onde $g(\underline{\lambda})$ é uma função arbitrária.

A correspondente densidade a priori para $\underline{\phi}$ é dada por:

$$\pi(\underline{\phi}) \propto g(\underline{\lambda}) \cdot I_{\delta\delta}(\delta, \underline{\lambda})^{1/2} \cdot \|J(\delta, \underline{\lambda})\|, \quad (95)$$

onde $\|J(\delta, \underline{\lambda})\|$ é o Jacobiano da transformação de $(\delta, \underline{\lambda})$ para $\underline{\phi}$.

3.4.4 Observações

1- A não unicidade da parametrização ortogonal é refletida na função $g(\underline{\lambda})$ em (94). Para o caso de um único parâmetro “nuisance” λ , uma possível solução para a escolha de $g(\lambda)$ é requerer que $\pi(\delta, \lambda)$ satisfaça a condição de Stein para λ , com δ tomado como parâmetro “nuisance”. Este método, porém, se aplica em alguns casos, como no exemplo abaixo, mas não é válido em geral.

Exemplo 8: Média e Variância da Normal. A média μ e o desvio-padrão σ da normal são ortogonais, e é fácil mostrar que a densidade a priori para (μ, σ) , com μ sendo parâmetro de interesse, é $\pi(\mu, \sigma) \propto g_1(\sigma)$ e, com σ sendo parâmetro de interesse, é

$$\pi(\mu, \sigma) \propto \frac{g_2(\mu)}{\sigma} \text{ onde } g_1(.) > 0 \text{ e } g_2(.) > 0 \text{ são funções arbitrárias.}$$

Requerendo que $g_1(\sigma) = \frac{g_2(\mu)}{\sigma}$ temos que $\pi(\mu, \sigma) \propto \frac{1}{\sigma}$.

Caso o parâmetro “nuisance” $\underline{\lambda}$ seja um vetor, pode-se ortogonalizar totalmente os parâmetros e então requerer que π satisfaça a condição de Stein, quando qualquer um dos parâmetros for o parâmetro de interesse, contudo, ortogonalização total nem sempre é possível (Cox e Reid, 1987).

2 - Na parametrização ortogonal, a densidade a priori de Jeffreys satisfaz a condição de Stein, se $I_{\underline{\lambda}\underline{\lambda}}(\delta, \underline{\lambda})$ não envolver δ . Se além disso, $I_{\delta\delta}(\delta, \underline{\lambda})$ não envolver $\underline{\lambda}$, então a densidade a priori de Jeffreys satisfaz a condição de Stein para ambos δ e $\underline{\lambda}$, como requer acima. Veremos no capítulo 4 que, para o modelo de Weibull sugerido, isto não se verifica.

3 - Embora pareça natural dizer que parâmetros ortogonais são a priori independentes (ver Apêndice C), isto não, necessariamente, ocorre no procedimento estudado aqui.

Exemplo 9: Sejam Y_1 e Y_2 variáveis Normais independentes, com médias θ_1 e θ_2 , respectivamente e variâncias unitárias. Suponhamos que $\delta = \theta_1 \cdot \theta_2$ seja o parâmetro de interesse. O parâmetro $\lambda = \theta_1^2 - \theta_2^2$ é ortogonal a δ , então, a densidade a priori $\pi(\delta, \lambda)$ é proporcional a $g(\lambda) \cdot (\lambda^2 + 4\delta^2)^{-1/4}$

4 - Devido a dificuldades de ortogonalização, às vezes impossível, a formulação de tal densidade a priori pode ser impraticável. Quando isto não ocorre, a densidade a priori aplica-se muito bem ao problema.

5 - Box e Tiao sugerem tomarmos precauções na integração sob parâmetros “nuisances”, porém, com a utilização da parametrização ortogonal na formulação da densidade a priori de Tibshirani, não é preciso nos preocuparmos com a influência do parâmetro λ na inferência sobre δ (ver Apêndice C).

3.4.5 Obtenção da densidade a priori de Tibshirani para o modelo de Weibull

Para o modelo de Weibull dado em (4), determinaremos uma reparametrização ortogonal (δ, λ) onde $\delta = p$ é o parâmetro de interesse e λ é o parâmetro “nuisance” a ser determinado.

O parâmetro ortogonal λ é obtido resolvendo a equação diferencial (ver Apêndice C):

$$i_{\theta\theta}^* \cdot \frac{\partial \theta}{\partial p} = -i_{\theta p}^* \quad (96)$$

Para os parâmetros do modelo de Weibull, tem-se,

$$\frac{1}{\theta^2} \cdot \frac{\partial \theta}{\partial p} = \frac{1}{p\theta} (\ln \theta + \psi(2)) \quad (97)$$

Separando as variáveis, (97) fica da seguinte forma,

$$\frac{1}{\theta (\ln \theta + \psi(2))} \partial \theta = \frac{1}{p} \partial p.$$

Integrando ambos os lados obtemos,

$$\ln (\ln \theta + \psi(2)) = \ln p + a(\lambda),$$

onde $a(\lambda)$ é uma função arbitrária de λ .

Escolhendo-se $a(\lambda) = \ln(\ln \lambda)$ obtém-se, como solução da equação diferencial (97), o parâmetro “nuisance” λ ortogonal δ ,

$$\lambda = \theta^{1/p} \cdot \exp \left\{ \frac{\psi(2)}{p} \right\}. \quad (98)$$

A matriz de Informação de Fisher, para os parâmetros ortogonais, é obtida por, $I(\delta, \lambda) = J' \cdot I(p, \theta) \cdot J$, onde $I(p, \theta)$ é dada em (52) e J é a matriz Jacobiana dada em (99), isto é.

$$J = \begin{bmatrix} 1 & 0 \\ \lambda^\delta \cdot \ln \lambda \cdot \exp \{-\psi(2)\} & \delta \cdot \lambda^{\delta-1} \cdot \exp \{-\Gamma'(2)\} \end{bmatrix}. \quad (99)$$

Portanto,

$$I(\delta, \lambda) = \begin{bmatrix} \frac{1}{\delta^2} (1 - \psi^2(2) + \psi'(2)) & 0 \\ 0 & \frac{\delta^2}{\lambda^2} \end{bmatrix}. \quad (100)$$

De (100), obtém-se,

$$\pi(\delta, \lambda) \propto g_1(\lambda) \cdot \left\{ \frac{1}{\delta^2} (1 - \psi^2(2) + \psi(2)) \right\}^{1/2}.$$

Ou seja, a densidade a priori de Tibshirani para (δ, λ) , onde δ é o parâmetro de interesse e $g_1(\lambda)$ é uma função arbitrária, é dada por,

$$\pi(\delta, \lambda) \propto g_1(\lambda) \cdot \frac{1}{\delta} \quad (101)$$

Procedendo como sugerido na observação 4 desta seção, determinemos uma forma única para a densidade a priori $\pi(\delta, \lambda)$. Considerando neste caso, λ como parâmetro de interesse, obtém-se

$$\pi(\delta, \lambda) \propto g_2(\delta) \cdot \left(\frac{\delta^2}{\lambda^2} \right)^{1/2} \text{ onde } g_2(\delta) : \text{função arbitrária.}$$

Ou seja,

$$\pi(\delta, \lambda) \propto g_2(\delta) \cdot \frac{1}{\lambda}. \quad (102)$$

Exigindo, $g_1(\lambda) \cdot \frac{1}{\delta} = g_2(\delta) \cdot \frac{1}{\lambda}$, obtemos que

$$g_1(\lambda) = \frac{1}{\lambda} \quad \text{e} \quad g_2(\delta) = \frac{1}{\delta},$$

portanto

$$\pi(\delta, \lambda) \propto \frac{1}{\delta \cdot \lambda}. \quad (103)$$

A densidade a priori de Tibshirani para os parâmetros iniciais (p, θ) é então,

$$\pi(p, \theta) \propto \frac{1}{p^2 \cdot \theta} \quad (104)$$

3.5 - Densidade a Priori de Bernardo

3.5.1- Motivação

Nesta seção revisamos o método para construção de densidades a priori não-informativas proposto por Bernardo (1979). Para mais detalhes sobre este método, ver Bernardo (1979), Berger e Bernardo (1991) e Berger (1992).

Por simplicidade, consideremos um parâmetro simples ϕ pertencente a um compacto $\Phi \subseteq \mathfrak{R}$.

A idéia é definir uma densidade a priori não-informativa, $\pi(\phi)$, que “maximize a informação sobre ϕ proveniente dos dados $\underline{x}$ ”.

Uma medida natural da informação esperada sobre ϕ proporcionada por $\underline{x}$, quando $\pi(\phi)$ é a densidade a priori, é (Shannon, 1948; Lindley, 1956)

$$I^\phi = E_x D \{p(\phi | \underline{x}), \pi(\phi)\}, \quad (105)$$

onde D é a “Distância de Kullback-Liebler” entre as densidades $p(\phi | \underline{x})$ e $\pi(\phi)$, definida por

$$D \{p(\phi | \underline{x}), \pi(\phi)\} = \int_{\Phi} p(\phi | \underline{x}) \ln \frac{p(\phi | \underline{x})}{\pi(\phi)} d\phi, \quad (106)$$

e E_x é a esperança com respeito à densidade marginal, $m(x)$, de X . Portanto,

$$I^\phi = \int \left\{ \int_{\Phi} p(\phi | \underline{x}) \ln \frac{p(\phi | \underline{x})}{\pi(\phi)} d\phi \right\} m(x) dx, \quad (107)$$

onde $m(x) = \int_{\Phi} f(x | \phi) \pi(\phi) d\phi$ e $p(\phi | \underline{x}) = \frac{f(\underline{x} | \phi) \pi(\phi)}{m(x)}$.

A densidade a priori $\pi(\phi)$ que maximiza I^ϕ , segundo Bernardo, deverá satisfazer a equação implícita

$$\pi(\phi) \propto \exp \left\{ \int f(x | \phi) \ln p(\phi | \underline{x}) dx \right\} \quad (108)$$

3.5.2- Construção da densidade a priori referencial

Infelizmente, a obtenção da densidade a priori, através de I^ϕ , não é muito satisfatória, uma vez que a densidade a priori $\pi(\phi)$ obtida pode ser discreta, mesmo quando ϕ é um parâmetro contínuo (Berger, Bernardo e Mendoza, 1989). Uma densidade a priori não-informativa discreta para um parâmetro contínuo obviamente não é muito atraente. Além disso, é muito difícil resolver a equação implícita (108) para $\pi(\phi)$.

Diante destas dificuldades, Bernardo (1979) considerou uma variante deste método, definindo

$$I_t^\phi = E_{Z_t} D \{p(\phi | Z_t), \pi(\phi)\}, \quad (109)$$

onde $Z_t = \{\underline{X}_1, \dots, \underline{X}_t\}$ é um vetor aleatório resultante de t replicações independentes identicamente distribuídas do experimento original, isto é, z_t consiste de t amostras $\{\underline{x}_1, \dots, \underline{x}_t\}$. Daí, Z_t terá densidade dada por

$$L(z_t | \phi) = \prod_{i=1}^t f(\underline{x}_i | \phi). \quad (110)$$

Quando $t \rightarrow \infty$, z_t proporcionará informação precisa sobre o valor de ϕ , e neste caso $I_\infty^\phi = \lim I_t^\phi$ será interpretado como sendo uma medida da quantidade de informação perdida sobre ϕ quando utilizamos a densidade a priori $\pi(\phi)$, porém, nos casos contínuos, frequentemente tem-se $I_\infty^\phi = +\infty$, para toda densidade a priori. Isto

é de se esperar, uma vez que seria necessário uma quantidade infinita de informação para se conhecer exatamente um número real, contudo, Bernardo sugere encontrar, para cada t , a densidade a priori $\pi_t(\phi)$ que maximiza I_t^ϕ , e então passar ao limite em t .

Usando um argumento variacional, pode-se mostrar que, sob certas condições de regularidade, para garantir as operações envolvidas, $\pi_t(\phi)$ satisfaz

$$\pi_t(\phi) \propto \exp \left\{ \int L(z_t | \phi) \ln p(\phi | z_t) dz_t \right\}. \quad (111)$$

Definição 6: Definimos a Densidade a Priori de Bernardo, também conhecida como Densidade a Priori Referencial $\pi(\phi)$, como sendo,

$$\pi(\phi) = \lim \pi_t(\phi) \quad (112)$$

assumindo que o limite existe.

Por $\lim \pi_t(\phi) = \pi(\phi)$, entendemos que a correspondente seqüência de funções de distribuições converge para a função de distribuição do limite em todos seus pontos de continuidade.

Como observado em (108), (111) somente define π_t implicitamente, contudo, quando $t \rightarrow \infty$, ambas $L(z_t | \phi)$ e $p(\phi | z_t)$ convergirão para suas distribuições assintóticas, e (111) tornar-se-á uma equação explícita, obtendo-se o seguinte resultado.

Caso Regular Contínuo

Resultado 2: Se $p(\phi | z_t)$ é assintoticamente normal, então

$$\pi_t(\phi) \propto \exp \left\{ \int L(z_t | \phi) \ln p(\phi | z_t) dz_t \right\} \xrightarrow{t \rightarrow \infty} I(\phi)^{1/2}, \quad (113)$$

onde $I(\phi)$ é a Informação de Fisher correspondente à f.d.p $L(z_t | \phi)$, isto é, a densidade a priori referencial convergirá para a densidade a priori de Jeffreys.

prova: No caso regular, para valores grandes de t , $p(\phi | z_t)$ pode ser aproximada por uma distribuição $N(\hat{\phi}_t, \frac{1}{t} S(\hat{\phi}_t))$, com $S(\phi)$ sendo a inversa da informação de Fisher $I(\phi)$.

Com um abuso de notação,

$$\begin{aligned} \pi_t(\phi) &\propto \exp \left\{ \int L(z_t | \phi) \ln N(\hat{\phi}_t, \frac{1}{t} S(\hat{\phi}_t)) dz_t \right\} \\ &= \exp \left\{ \int L(\hat{\phi}_t | \phi) \ln N(\hat{\phi}_t, \frac{1}{t} S(\hat{\phi}_t)) dz_t \right\} \\ &= \exp \left\{ \int L(\hat{\phi}_t | \phi) \left(\ln \sqrt{\frac{t}{2\pi}} - \frac{1}{2} \ln S(\hat{\phi}_t) - \frac{t}{2S(\hat{\phi}_t)} (\hat{\phi}_t - \phi)^2 \right) d\hat{\phi}_t \right\} \end{aligned} \quad (114)$$

Mas $L(\hat{\phi}_t | \phi)$ pode ser aproximada, para t grande, por uma normal $N(\hat{\phi}_t, \frac{1}{t} S(\phi))$, tal que (114) torna-se

$$\begin{aligned} \pi_t(\phi) &\propto \sqrt{\frac{t}{2\pi}} \exp \left\{ \int N(\hat{\phi}_t, \frac{1}{t} S(\phi)) \left(-\frac{1}{2} \ln S(\hat{\phi}_t) - \frac{t}{2S(\hat{\phi}_t)} (\hat{\phi}_t - \phi)^2 \right) d\hat{\phi}_t \right\} \\ &\cong \sqrt{\frac{t}{2\pi}} \exp \left\{ -\frac{1}{2} \ln S(\phi) \right\} \exp \{-1\} \end{aligned}$$

A aproximação para a integral sobre o primeiro termo segue do fato que $N(\hat{\phi}_t, \frac{1}{t} S(\phi))$ é convergente para um ponto concentrado em ϕ .

Então, temos que, para t grande, $\pi_t(\phi)$ é aproximadamente proporcional a

$$\exp\left\{-\frac{1}{2}\ln S(\phi)\right\} = S(\phi)^{-1/2} = \sqrt{I(\phi)}$$

que é a densidade a priori referencial. □

Observação: O argumento pode ser facilmente estendido para o caso multivariado, obtendo-se a densidade a priori multivariada de Jeffreys somente para inferências simultâneas sobre todos os parâmetros, isto é, desde que não haja parâmetros “nuisances” e a forma usual da normalidade assintótica possa ser garantida.

3.5.3- Densidade a priori referencial para parâmetros de interesse na presença de parâmetros “nuisances”

Motivação

Se, contudo, estamos interessados em um dos parâmetros, sendo os parâmetros restantes “nuisances”, a situação é bastante diferente, e a densidade a priori referencial apropriada não é mais a densidade a priori de Jeffreys multivariada.

Apresentamos em detalhes o caso de dois parâmetros. Para o caso geral, ver Berger e Bernardo (1991).

Suponhamos o modelo $f(x | \eta)$ com $\eta = (\phi, \lambda)$ sendo o vetor paramétrico desconhecido, onde ϕ é o parâmetro de interesse e λ o parâmetro “nuisance”.

Neste caso, o método proposto por Bernardo consiste de dois estágios. Procede-se, primeiramente, determinando-se $\pi(\lambda | \phi)$, a densidade a priori referen-

cial condicional para λ assumindo que ϕ é dado. Isto é feito exatamente como no caso univariado proposto anteriormente.

Daí, integrando-se $L(z_t | \phi, \lambda)$, ponderado por $\pi(\lambda | \phi)$ com respeito a λ , obtem-se o modelo marginal $L_1(z_t | \phi)$, para o qual a densidade a priori referencial $\pi(\phi)$ pode ser calculada (quando $t \rightarrow \infty$).

Para determinarmos a densidade a priori referencial $\pi(\phi)$, observemos que a informação esperada sobre ϕ proporcionada pelos dados, quando a densidade a priori de (ϕ, λ) é dada por (108), é definida por

$$I_t^\phi = \int \left\{ \int_{\Phi} p(\phi | z_t) \ln \frac{p(\phi | z_t)}{\pi(\phi)} d\phi \right\} m(z_t) dz_t, \quad (115)$$

com $p(\phi | z_t) = \frac{L_1(z_t | \phi) \pi(\phi)}{m(z_t)}$ onde $m(z_t) = \int L(z_t | \phi, \lambda) \pi_\phi(\phi, \lambda) d\phi d\lambda$ e $L_1(z_t | \phi) = \int L(z_t | \phi, \lambda) \pi(\lambda | \phi) d\lambda$.

Para qualquer dada densidade a priori condicional $\pi(\lambda | \phi)$, a informação esperada sobre ϕ pode ser calculada de (115). Usando o argumento anterior, a densidade a priori referencial $\pi(\phi)$, para o parâmetro de interesse, pode ser obtida como o limite de

$$\pi_t(\phi) \propto \exp \left\{ - \int L_1(z_t | \phi) \ln p(\phi | z_t) dz_t \right\} \quad (116)$$

A densidade a priori referencial $\pi(\phi)$ assim obtida, geralmente, dependerá de $\pi(\lambda | \phi)$.

Observações:

- 1- Observe que a informação esperada sobre ϕ depende da densidade a priori global $\pi_\phi(\phi, \lambda)$ e não somente da correspondente marginal $\pi(\phi)$.

2- A densidade a priori condicional do parâmetro “nuisance” $\pi(\lambda \mid \phi)$ pode ser escolhida por opiniões pessoais ou conhecimento objetivo empírico prévio, por ex., utilizando-se a densidade a priori de Jeffreys.

Definição 7: A densidade a priori referencial conjunta (global) com ϕ parâmetro de interesse é obtida por

$$\pi_{\phi}(\phi, \lambda) = \pi(\phi) \pi(\lambda \mid \phi) \quad (117)$$

3- É importante distinguir, previamente, o parâmetro de interesse ϕ do parâmetro global (ϕ, λ) , pois, uma mudança no parâmetro de interesse poderá produzir um resultado diferente.

Caso regular contínuo - Algoritmo

Seja $\underline{\eta} = (\phi, \lambda)$ sendo ϕ o parâmetro de interesse e λ o parâmetro “nuisance”. O método proposto por Bernardo consiste de dois estágios.

Primeiramente, determina-se a densidade a priori referencial condicional normalizada, $\pi(\lambda \mid \phi)$, para λ assumindo que ϕ é dado. Do argumento usado no caso univariado, uma vez que ϕ está sendo tratado nos cálculos como conhecido, tem-se que

$$\pi(\lambda \mid \phi) = k \cdot \sqrt{I_{\lambda\lambda}(\phi, \lambda)}, \quad (118)$$

onde $I_{\lambda\lambda}(\phi, \lambda)$ é elemento da matriz de informação de Fisher e k é a constante

normalizadora, isto é,

$$k = \frac{1}{\int \sqrt{I_{\lambda\lambda}(\phi, \lambda)} d\lambda} \quad (119)$$

O segundo passo é encontrar a densidade a priori $\pi(\phi)$, obtida pelo mesmo tipo de argumento assintótico utilizado em (114), isto é, notando que a densidade a posteriori marginal assintótica de ϕ é $N(\phi | \hat{\phi}, \frac{1}{t} S_1(\hat{\eta}))$, onde $S_1(\underline{\eta})$ é o elemento de ordem 1×1 da matriz inversa, $S(\underline{\eta})$, de $I(\underline{\eta})$ e $\hat{\eta}$, é o EMV.

Resultado 3:
$$\pi(\phi) \propto \exp \left\{ -\frac{1}{2} \int \pi(\lambda | \phi) \ln \left\| \frac{I_{\lambda\lambda}(\phi, \lambda)}{\det I(\phi, \lambda)} \right\| d\lambda \right\}. \quad (120)$$

onde $I(\phi, \lambda)$ é a Matriz de Informação de Fisher correspondente à f.d.p. $L(z_t | \phi, \lambda)$

prova: Partindo de $\pi_t(\phi) \propto \exp \{ - \int L_1(z_t | \phi) \ln p(\phi | z_t) dz_t \}$

$$\cong \exp \left\{ \int L_1(z_t | \phi) \ln N(\phi | \hat{\phi}, \frac{1}{t} S_1(\hat{\eta})) dz_t \right\}$$

$$= \exp \left\{ \int L_1(\hat{\eta} | \phi) \ln N(\phi | \hat{\phi}, \frac{1}{t} S_1(\hat{\eta})) d\hat{\eta} \right\}$$

Mas $L_1(\hat{\eta} | \phi) = \int L(\hat{\eta} | \phi, \lambda) \pi(\lambda | \phi) d\lambda$

Portanto, $\pi_t(\phi) \cong \exp \left\{ \int \pi(\lambda | \phi) \left(\int L(\hat{\eta} | \phi, \lambda) \ln N(\phi | \hat{\phi}, \frac{1}{t} S_1(\hat{\eta})) d\hat{\eta} \right) d\lambda \right\}$.

O último passo segue do mesmo argumento utilizado em (114), obtendo a expressão para $\pi(\phi)$ dada em (120), isto é,

$$\pi(\phi) \propto \exp \left\{ -\frac{1}{2} \int \pi(\lambda | \phi) \ln \|S_1(\phi, \lambda)\| d\lambda \right\}. \quad (121)$$

A matriz inversa da matriz de informação de Fisher é

$$S(\phi, \lambda) = \begin{bmatrix} \frac{I_{\lambda\lambda}(\phi, \lambda)}{\det I(\phi, \lambda)} & -\frac{I_{\lambda\phi}(\phi, \lambda)}{\det I(\phi, \lambda)} \\ -\frac{I_{\phi\lambda}(\phi, \lambda)}{\det I(\phi, \lambda)} & \frac{I_{\phi\phi}(\phi, \lambda)}{\det I(\phi, \lambda)} \end{bmatrix} \quad (122)$$

e conseqüentemente, temos que $S_1(\phi, \lambda) = \frac{I_{\lambda\lambda}(\phi, \lambda)}{\det I(\phi, \lambda)}$ □

A densidade a priori referencial global para (ϕ, λ) é então

$$\pi_\phi(\phi, \lambda) \propto \pi(\phi) \pi(\lambda | \phi) \quad (123)$$

Observações:

- 1- Se o parâmetro global (ϕ, λ) for considerado de interesse, então o método de Bernardo reduz-se à densidade a priori de Jeffreys.
- 2- A dependência da densidade a priori, na escolha do parâmetro de interesse, parece-nos razoável. O fato de um particular parâmetro ser considerado como parâmetro de interesse é uma informação relevante e afetaria a escolha da densidade a priori.
- 3- A densidade a priori referencial não é invariante sob transformações um a um dos parâmetros. Segundo Berger e Bernardo (1991) os estudos sobre invariância, ainda, estão sendo analisados.
- 4- Pode haver casos em que a densidade a priori referencial não seja adequada, como por exemplo, quando há mais do que um parâmetro “nuisance” no modelo. Para

mais detalhes, ver Berger e Bernardo (1991) e Berger (1992)).

5- Se δ é o parâmetro de interesse e ψ é um parâmetro “nuisance” ortogonal, então a densidade a priori referencial $\pi(\delta, \psi)$ é aquela sugerida pelo argumento da cobertura frequentista assintótica, a densidade a priori sugerida por Tibshirani. Notemos que o método da densidade a priori referencial é importante para aplicação do método proposto por Tibshirani, para se determinar uma densidade a priori não-informativa, uma vez que especifica como a função do parâmetro “nuisance” $g(\psi)$ em (94) seria escolhida, $g(\psi)$ seria a densidade a priori referencial $\pi(\psi)$. Isto não é especificado no método de Tibshirani.

Exemplo 10: Consideremos a reparametrização ortogonal (δ, λ) dos parâmetros (p, θ) , determinada em (98), onde $\delta = p$ e $\lambda = \theta^{1/p} \exp\left\{\frac{\Gamma'(2)}{p}\right\}$.

Se δ é o parâmetro de interesse, então a densidade a priori referencial conjunta é

$$\pi_\delta(\delta, \lambda) \propto \pi(\delta) \sqrt{I_{\lambda\lambda}(\delta, \lambda)} = \frac{1}{\delta} \frac{1}{\lambda},$$

onde $\pi(\lambda | \delta) \propto \sqrt{I_{\lambda\lambda}(\delta, \lambda)}$ é a densidade a priori condicional e $\pi(\delta) \propto \frac{1}{\delta}$ é a densidade a priori referencial para o parâmetro δ .

Agora, a densidade a priori sugerida por Tibshirani com λ sendo o parâmetro de interesse é determinada como segue

$$\pi_\lambda(\delta, \lambda) \propto g(\delta) \sqrt{I_{\lambda\lambda}(\delta, \lambda)} = \frac{1}{\delta} \frac{1}{\lambda},$$

Concluimos então que a densidade a priori de Bernardo satisfaz a condição de Stein tomando-se λ como parâmetro de interesse, onde $g(\delta)$ é a densidade a priori referencial $\pi(\delta)$.

Como a densidade a priori referencial é $\pi(\delta, \lambda) \propto \frac{1}{\delta\lambda}$, independente do parâmetro de interesse, então ela satisfaz a condição de Stein globalmente.

3.5.4 Obtenção da densidade a priori de Bernardo para o modelo de Weibull

Determinaremos a densidade a priori de Bernardo para os parâmetros do modelo de Weibull dado em (4), onde p será considerado o parâmetro de interesse e θ o parâmetro “nuisance”.

Pelo método proposto por Bernardo, determinemos a densidade a priori de Jeffreys para θ , assumindo que p seja dado, como segue,

$$\pi(\theta | p) = k \sqrt{I_{\theta\theta}(p, \theta)} \text{ onde } k = \frac{1}{\int \sqrt{I_{\theta\theta}(p, \theta)} d\theta} \quad (124)$$

De (52), temos que

$$\pi(\theta | p) = k \frac{1}{\theta} \text{ onde } k = \frac{1}{\int \frac{1}{\theta} d\theta} \quad (125)$$

O segundo passo é determinar a densidade a priori $\pi(p)$ que é obtida por

$$\pi(p) \propto \exp \left\{ -\frac{1}{2} \int \pi(\theta | p) \ln \left\| \frac{I_{\theta\theta}(p, \theta)}{\det I(p, \theta)} \right\| d\theta \right\}. \quad (126)$$

Substituindo os devidos valores na expressão acima obtemos

$$\pi(p) \propto \exp \left\{ -\frac{1}{2} \int k \frac{1}{\theta} \ln \left\| \frac{1/\theta^2}{[1 + \psi'(2) - \psi^2(2)]/p^2\theta^2} \right\| d\theta \right\}, \quad (127)$$

ou ainda,

$$\pi(p) \propto \exp \left\{ -\frac{1}{2} \ln \left(\frac{p^2}{1 + \psi'(2) - \psi^2(2)} \right) k \int \frac{1}{\theta} d\theta \right\}.$$

Portanto, a densidade a priori referencial resultante para p é

$$\pi(p) \propto \frac{1}{p}, \quad (128)$$

e a densidade a priori referencial global para (p, θ) obtida de $\pi_p(p, \theta) \propto \pi(\theta | p)\pi(p)$ é,

$$\pi_p(p, \theta) \propto \frac{1}{p \theta}. \quad (129)$$

De forma análoga, obtemos a densidade a priori referencial, considerando-se θ parâmetro de interesse e p “nuisance”. Neste caso, a densidade a priori é,

$$\pi_\theta(p, \theta) \propto \frac{1}{p \theta \sqrt{1 + (\ln \theta)^2 + 2\psi(2)\ln \theta + \psi'(2)}} \quad (130)$$

Capítulo 4

ESTUDO COMPARATIVO DAS DENSIDADES A PRIORI NÃO-INFORMATIVAS

4.1 - Cálculo da densidade a posteriori de R utilizando-se a densidade a priori de Jeffreys

Observemos que a função de confiabilidade R (6) depende dos parâmetros p e θ , portanto, R é um parâmetro com uma densidade a posteriori $p(R | \underline{x})$.

Para simplificar os cálculos posteriores, Zellner (1990) sugere utilizar a reparametrização (β, γ) , dada a seguir

$$\beta = \theta^{1/p} \text{ e } \gamma = \frac{1}{p}, \quad (131)$$

para os parâmetros do modelo de Weibull definido em (4).

4.1.1- Cálculo da densidade a priori de Jeffreys nos novos parâmetros

Como a densidade a priori de Jeffreys é invariante sob a transformação (131) então, nos novos parâmetros (β, γ) , a densidade a priori pode ser determinada,

utilizando-se o teorema da Mudança de Variáveis como segue:

$$\pi(\beta, \gamma) = \pi(p, \theta) \cdot \|J\| \text{ onde } \pi(p, \theta) \propto \frac{1}{p\theta} \text{ e}$$

$$J = \begin{bmatrix} -\frac{1}{\gamma^2} & 0 \\ -\frac{1}{\gamma^2} \beta^{1/\gamma} \ln \beta & \frac{1}{\gamma} \beta^{1/\gamma-1} \end{bmatrix} \quad (132)$$

Portanto, a densidade a priori de Jeffreys para (β, γ) é

$$\pi(\beta, \gamma) \propto \frac{1}{\beta \gamma^2} \quad (133)$$

4.1.2- Cálculo da densidade a posteriori de R

Para determinarmos a densidade a posteriori da função de confiabilidade R , calculamos, primeiramente, a densidade a posteriori conjunta $p(\beta, \gamma | \underline{x})$, daí transformamos, adequadamente, (β, γ) em (R, W) e então determinamos a densidade a posteriori marginal $p(R | \underline{x})$.

A densidade a posteriori conjunta $p(\beta, \gamma | \underline{x})$ é proporcional ao produto da função de verossimilhança nos parâmetros (β, γ)

$$L(\beta, \gamma | \underline{x}) = \frac{1}{\gamma^n \beta^{n/\gamma}} \prod x_i^{1/\gamma-1} \exp \left\{ - \sum \left(\frac{x_i}{\beta} \right)^{1/\gamma} \right\}. \quad (134)$$

pela densidade a priori de Jeffreys (133) . Logo,

$$p(\beta, \gamma | \underline{x}) \propto \frac{1}{\gamma^{n+2} \beta^{1+n/\gamma}} \prod x_i^{1/\gamma-1} \exp \left\{ - \sum \left(\frac{x_i}{\beta} \right)^{1/\gamma} \right\}. \quad (135)$$

A Função de Confiabilidade R , em relação aos novos parâmetros, é dada por,

$$R = \exp \left\{ - \left(\frac{t}{\beta} \right)^{1/\gamma} \right\} \quad (136)$$

Para determinarmos a densidade a posteriori $p(R, W \mid \underline{x})$, utilizamos a transformação

$$R = \exp \left\{ - \left(\frac{t}{\beta} \right)^{1/\gamma} \right\} \text{ e } W = \gamma \quad (137)$$

A matriz Jacobiana é dada por,

$$J = \begin{bmatrix} 0 & 1 \\ \frac{t.W}{R(\ln \frac{1}{R})^{W+1}} & \frac{t \ln(\ln \frac{1}{R})}{(\ln \frac{1}{R})^W} \end{bmatrix} \quad (138)$$

e, conseqüentemente, temos como Jacobiano da transformação $\|J\| = \frac{t.W}{(\ln \frac{1}{R})^{W+1} . R}$

A densidade a posteriori conjunta de (R, W) é então,

$$p(R, W \mid \underline{x}) \propto \frac{1}{W^{n+1} . t^{n/W}} \cdot \prod x_i^{1/W-1} \cdot \left(\ln \frac{1}{R} \right)^{n-1} . R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) \quad (139)$$

Integrando, com respeito a W , obtemos a densidade a posteriori marginal de R ,

$$p(R \mid \underline{x}) \propto \left(\ln \frac{1}{R} \right)^{n-1} \int_0^\infty \frac{1}{W^{n+1} . t^{n/W}} \cdot \prod x_i^{1/W-1} \cdot \left(\ln \frac{1}{R} \right)^{n-1} . R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) dW \quad (140)$$

Sinha e Guttman (1988) determinam graficamente a densidade a posteriori (140) utilizando integração numérica. Em nosso estudo, procuramos obter uma forma fechada para a densidade a posteriori acima, aplicando a Aproximação de Laplace (ver Apêndice A). Para isto determinemos $f(W)$ e $h(W)$ na integral dada em (140).

A integral

$$\int_0^\infty \frac{1}{W^{n+1} \cdot t^{n/W}} \cdot \prod x_i^{1/W-1} \cdot R\left(\sum \left(\frac{x_i}{t}\right)^{1/W} - 1\right) dW \quad (141)$$

pode ser escrita como

$$\int_0^\infty \prod x_i^{1/W-1} \cdot R\left(\sum \left(\frac{x_i}{t}\right)^{1/W} - 1\right) \cdot \exp\left\{-(n+1) \cdot \ln W - \frac{n}{W} \cdot \ln t\right\} dW.$$

$$\text{Escolhendo, } f(W) = \prod x_i^{1/W-1} \cdot R\left(\sum \left(\frac{x_i}{t}\right)^{1/W} - 1\right)$$

$$\text{e } h(W) = (n+1) \cdot \ln W + \frac{n}{W} \cdot \ln t$$

$$\text{temos que, } h'(W) = \frac{n+1}{W} - \frac{n}{W^2} \ln t,$$

$$h''(W) = -\frac{n+1}{W^2} + 2 \cdot \frac{n}{W^3} \ln t \quad \text{e} \quad \dot{W} = \frac{n}{n+1} \cdot \ln t$$

Aplicando-se a Aproximação de Laplace, a integral (141) pode ser aproximada por,

$$\sqrt{\frac{2\pi}{n}} \prod x_i^{1/\hat{W}-1} \cdot R\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right) \cdot \exp\left\{-(n+1) \cdot \ln \hat{W} - \frac{n}{\hat{W}} \cdot \ln t\right\} \quad (142)$$

$$\text{com } \sigma = \frac{n \cdot \ln t}{(n+1)^{3/2}}$$

Consequentemente,

$$p(R | \underline{x}) \propto \left(\ln \frac{1}{R} \right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} - 1 \right) \quad (143)$$

Desde que, $\int_0^1 \left(\ln \frac{1}{R} \right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} - 1 \right) dR = \left(\sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} \right)^{-n} \cdot \Gamma(n)$ (Apêndice D), a densidade a posteriori, normalizada, da função de confiabilidade R é:

$$p(R | \underline{x}) = \frac{\left(\sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} \right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R} \right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} - 1 \right). \quad (144)$$

onde

$$\hat{W} = \frac{n}{n+1} \ln t.$$

R tem, então, uma distribuição a posteriori Log-Gama Negativa (Apêndice B), utilizando-se uma densidade a priori de Jeffreys, isto é,

$$R | \underline{x} \sim LGN \left(n; \sum \left(\frac{x_i}{t} \right)^{1/\hat{W}} \right) \text{ onde } \hat{W} = \frac{n}{n+1} \cdot \ln t$$

Obtemos assim, não só uma expressão para a densidade a posteriori, mas também conseguimos determinar uma distribuição conhecida para a função de confiabilidade R .

Uma comparação gráfica entre o método de Aproximação de Laplace, utilizado para obtermos uma forma explícita da integral (141), e o método de integração numérica de Simpson, é apresentada na figura 8.

Figura 8: Posteriores para R via Laplace e Método de Simpson para a) $n = 5$, $t = 2$, $p = 2$ e $\theta = 4$ e b) $n = 25$, $t = 1.2$, $p = 4$ e $\theta = 2$.

4.2- Cálculo da densidade a posteriori de R utilizando a densidade a priori de Zellner

4.2.1- Cálculo da densidade a priori de Zellner nos novos parâmetros

Como a densidade a priori de Zellner não é invariante sob transformação um a um dos parâmetros, substituímos então esses novos parâmetros na f.d.p. (4) e calculamos

$$\pi(\beta, \gamma) \propto \exp \left\{ H_{(\beta, \gamma)}(X) \right\}.$$

A fdp de x nos parâmetros β e γ é

$$f(x | \beta, \gamma) = \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} \quad (145)$$

A Medida de Informação $H_{(\beta, \gamma)}(X)$ é determinada a seguir,

$$H_{(\beta, \gamma)}(X) = \int_0^\infty \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} \cdot \ln\left\{\frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\}\right\} dx \quad (146)$$

Desenvolvendo o logarítmo em (146) tem-se que:

$$\begin{aligned} H_{(\beta, \gamma)}(X) = & -\ln\gamma \cdot \int_0^\infty \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} dx - \frac{1}{\gamma} \ln\beta \cdot \int_0^\infty \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} dx + \\ & + \left(\frac{1}{\gamma} - 1\right) \cdot \int_0^\infty \ln x \cdot \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} dx - \int_0^\infty \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \frac{x^{1/\gamma}}{\beta^{1/\gamma}} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} dx. \end{aligned}$$

Sabendo que $\int_0^\infty \frac{\beta^{-1/\gamma}}{\gamma} \cdot x^{1/\gamma-1} \cdot \exp\left\{-\frac{x^{1/\gamma}}{\beta^{1/\gamma}}\right\} dx = 1$ e fazendo uma mudança de variável $u = \frac{x^{1/\gamma}}{\beta^{1/\gamma}}$ nas duas últimas integrais obtemos,

$$H_{(\beta, \gamma)} = -\ln\gamma - \frac{1}{\gamma} \ln\beta + \left(\frac{1}{\gamma} - 1\right) \cdot \ln\beta + \gamma \cdot \left(\frac{1}{\gamma} - 1\right) \cdot \int_0^\infty \ln u \cdot e^{-u} du \cdot \int_0^\infty u \cdot e^{-u} du.$$

De $\int_0^\infty \ln u \cdot e^{-u} du = \psi(1)$ e $\int_0^\infty u \cdot e^{-u} du = 1$ resulta,

$$H_{(\beta, \gamma)} = -\ln\gamma - \ln\beta - \gamma \cdot \psi(1) + \psi(1) - 1 \quad (147)$$

Portanto, na nova parametrização, a densidade a priori de Zellner fica como,

$$\pi(\beta, \gamma) \propto \frac{1}{\gamma\beta} \cdot \exp\{-\psi(1) \cdot \gamma\} \quad (148)$$

Observação: Observemos que $\pi(\beta, \gamma)$ é um produto de Gammas, isto é,

$$\pi(\beta, \gamma) \propto \beta^{r_1-1} \exp(-q_1\beta) \gamma^{r_2-1} \exp(-q_2\gamma), \text{ onde } r_1 = r_2 = q_1 = 0 \text{ e } q_2 = \psi(1)$$

4.2.2 - Cálculo da densidade a posteriori de R

Nesta seção e nas seções 4.3 e 4.4, determinamos a densidade a posteriori da função de confiabilidade R , de forma análoga ao que foi feito na seção anterior.

Portanto, utilizando-se a densidade a priori (148), obtemos a densidade a posteriori conjunta para (β, γ)

$$p(\beta, \gamma | \underline{x}) \propto \frac{1}{\gamma^{n+1} \beta^{n/\gamma+1}} \prod x_i^{1/\gamma-1} \cdot \exp \left\{ - \sum \left(\frac{x_i}{\beta} \right)^{1/\gamma} - \psi(1) \cdot \gamma \right\} \quad (149)$$

Determinemos a densidade a posteriori $p(R, W | \underline{x})$ na parametrização $R = \exp \left\{ - \left(\frac{t}{\beta} \right)^{1/\gamma} \right\}$ e $W = \gamma$. Então,

$$p(R, W | \underline{x}) \propto \frac{1}{W^n \cdot t^n / W} \cdot \left(\ln \frac{1}{R} \right)^{n-1} \cdot \prod x_i^{1/W-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) \cdot \exp \{ -\psi(1)W \} \quad (150)$$

Integrando, com respeito a W , obtemos a densidade a posteriori marginal de R como segue,

$$p(R | \underline{x}) \propto \left(\ln \frac{1}{R} \right)^{n-1} \cdot \int_0^\infty \frac{1}{W^n \cdot t^n / W} \cdot \prod x_i^{1/W-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) \cdot \exp \{ -\psi(1)W \} dW \quad (151)$$

Aplicando-se a Aproximação de Laplace à (151) e escolhendo

$$f(W) = \prod x_i^{1/W-1} .R\left(\sum\left(\frac{x_i}{t}\right)^{1/W}-1\right) .exp\{-\psi(1)W\} \quad e \quad h(W) = -n.ln W - \frac{n}{W}.lnt$$

temos:

$$p(R | \underline{x}) \propto \left(\ln \frac{1}{R}\right)^{n-1} .R\left(\sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}-1\right) \text{ onde } \hat{W} = lnt \quad (152)$$

Desde que, $\int_0^1 \left(\ln \frac{1}{R}\right)^{n-1} .R\left(\sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}-1\right) dR = \left(\sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^{-n} .\Gamma(n)$, a densidade a posteriori, normalizada, da função de confiabilidade R é então,

$$p(R | \underline{x}) = \frac{\left(\sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} .\left(\ln \frac{1}{R}\right)^{n-1} .R\left(\sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}-1\right) \quad (153)$$

onde $\hat{W} = lnt$

Novamente, R tem uma distribuição a posteriori Log-Gama Negativa, isto é,

$$R | \underline{x} \sim LGN\left(n; \sum\left(\frac{x_i}{t}\right)^{1/\hat{W}}\right), \text{ onde } \hat{W} = lnt$$

4.3 - Cálculo da densidade a posteriori de R utilizando a densidade a priori de Tibshirani

4.3.1- Cálculo da densidade a priori de Tibshirani nos novos parâmetros

Como a densidade a priori de Tibshirani é invariante sob a transformação (131) então, nos novos parâmetros (β, γ) , a densidade a priori pode ser determinada utilizando-se o teorema da Mudança de Variáveis, $\pi(\beta, \gamma) = \pi(p, \theta) \cdot \|J\|$, onde $\pi(p, \theta) \propto \frac{1}{p^2 \cdot \theta}$ e $\|J\| = \frac{1}{\gamma^3} \beta^{1/\gamma-1}$.

Portanto, a densidade a priori de Tibshirani, nos novos parâmetros, é

$$\pi(\beta, \gamma) \propto \frac{1}{\beta \cdot \gamma} \quad (154)$$

4.3.2- Cálculo da densidade a posteriori de R

A densidade a posteriori conjunta $p(\beta, \gamma \mid \underline{x})$ correspondente à densidade a priori (154) é

$$p(\beta, \gamma \mid \underline{x}) \propto \frac{1}{\gamma^{n+1} \beta^{n/\gamma+1}} \cdot \prod x_i^{1/\gamma-1} \cdot \exp \left\{ - \sum \left(\frac{x_i}{\beta} \right)^{1/\gamma} \right\} \quad (155)$$

Determinemos a densidade a posteriori na parametrização (R, W) onde $R = \exp \left\{ - \left(\frac{t}{\beta} \right)^{1/\gamma} \right\}$ e $W = \gamma$. Daí,

$$p(R, W \mid \underline{x}) \propto \frac{1}{W^n \cdot t^{n/W}} \cdot \left(\ln \frac{1}{R} \right)^{n-1} \cdot \prod x_i^{1/W-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) \quad (156)$$

Integrando, com respeito a W , obtemos a densidade a posteriori marginal de R

$$p(R \mid \underline{x}) \propto \left(\ln \frac{1}{R} \right)^{n-1} \cdot \int_0^\infty \frac{1}{W^n \cdot t^{n/W}} \cdot \prod x_i^{1/W-1} \cdot R \left(\sum \left(\frac{x_i}{t} \right)^{1/W} - 1 \right) dW \quad (157)$$

e a seguir, aplicando-se a Aproximação de Laplace em (157), tem-se

$$p(R | \underline{x}) \propto \left(\ln \frac{1}{R}\right)^{n-1} R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right) \quad (158)$$

onde $\hat{W} = \ln t$.

Desde que, $\int_0^1 \left(\ln \frac{1}{R}\right)^{n-1} R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right) dR = \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^{-n} \Gamma(n)$, a densidade a posteriori, normalizada, da função de confiabilidade R é então,

$$p(R | \underline{x}) = \frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right) \quad (159)$$

onde $\hat{W} = \ln t$

Portanto, como em (144) e (153), R tem uma distribuição a posteriori Log-Gama Negativa, neste caso, utilizando-se a densidade a priori de Tibshirani. Além disso, concluímos de (153) e (159), que tanto faz utilizarmos a densidade a priori de Zellner ou a densidade a priori de Tibshirani que as densidades a posteriori serão as mesmas.

4.4 - Cálculo da densidade a posteriori de R utilizando a densidade a priori de Bernardo

4.4.1 - Cálculo da densidade a priori de Bernardo nos novos parâmetros

Assim como a densidade a priori de Zellner, a densidade a priori de Bernardo também não é invariante sob a transformação (131) então, nos novos parâmetros

(β, γ) , esta deverá ser calculada passo a passo, utilizando-se o procedimento proposto por Berger/Bernardo. Seja β o parâmetro de interesse, então a densidade a priori $\pi(\gamma | \beta)$ é obtida da seguinte forma,

$$\pi(\gamma | \beta) = k \cdot \sqrt{I_{\gamma\gamma}(\beta, \gamma)} \text{ onde } k = \frac{1}{\int \sqrt{I_{\gamma\gamma}(\beta, \gamma)} d\gamma} \quad (160)$$

onde

$$I(\beta, \gamma) = \begin{bmatrix} \frac{1}{\gamma^2 \beta^2} & \frac{\psi(2)}{\gamma^2 \beta} \\ \frac{\psi(2)}{\gamma^2 \beta} & \frac{1}{\gamma^2} (1 + \psi'(2)) \end{bmatrix} \quad (161)$$

De (161), a densidade a priori condicional resulta

$$\pi(\gamma | \beta) = k \cdot \frac{1}{\gamma} \text{ onde } k = \frac{1}{\int \frac{1}{\gamma} d\gamma} \quad (162)$$

A densidade a priori referencial marginal para β , é obtida da seguinte expressão,

$$\pi(\beta) \propto \exp \left\{ -\frac{1}{2} \int \pi(\gamma | \beta) \cdot \ln \left\| \frac{I_{\gamma\gamma}(\beta, \gamma)}{\det I(\beta, \gamma)} \right\| d\gamma \right\} \quad (163)$$

resultando em,

$$\pi(\beta) \propto \exp \left\{ -\frac{1}{2} \int k \cdot \frac{1}{\gamma} \cdot \ln \left\| \gamma^2 \beta^2 \frac{1 + \psi'(2)}{1 + \psi'(2) - \psi^2(2)} \right\| d\gamma \right\} \quad (164)$$

$$\text{Daí, } \pi(\beta) \propto \exp \left\{ -\frac{1}{2} \cdot \ln \beta^2 \cdot k \cdot \int \frac{1}{\gamma} d\gamma \right\} \cdot \exp \left\{ -\frac{1}{2} \cdot k \cdot \int \frac{1}{\gamma} \cdot \ln \gamma^2 d\gamma \right\}$$

Como a segunda exponencial não depende de β , $\pi(\beta)$ reduz-se a

$$\pi(\beta) \propto \exp \left\{ -\frac{1}{2} \cdot \ln \beta^2 \right\}$$

Conseqüentemente,

$$\pi(\beta) \propto \frac{1}{\beta} \quad (165)$$

Portanto, a densidade a priori referencial global (conjunta) para os novos parâmetros (β, γ) considerando-se β como parâmetro de interesse, obtida do produto $\pi_\beta(\beta, \gamma) \propto \pi(\gamma | \beta) \cdot \pi(\beta)$ é,

$$\pi_\beta(\beta, \gamma) \propto \frac{1}{\beta \cdot \gamma} \quad (166)$$

Similarmente, se estamos interessados em γ , sendo β o parâmetro “nuisance”, podemos usar um argumento análogo para obtermos

$$\pi_\gamma(\beta, \gamma) \propto \frac{1}{\beta \cdot \gamma} \quad (167)$$

como densidade a priori referencial para fazermos inferência sobre γ .

Concluimos de (166) e (167) que, independente do parâmetro de interesse, a densidade a priori conjunta para β e γ é

$$\pi(\beta, \gamma) \propto \frac{1}{\beta \cdot \gamma} \quad (168)$$

que coincide com a densidade a priori (154) dada por Tibshirani, conseqüentemente, as densidades a posteriori serão iguais.

Observação: Observemos que $\pi(\beta, \gamma) = \pi(\beta) \cdot \pi(\gamma)$, logo β e γ são independentes.

4.4.2- Cálculo da densidade a posteriori de R

A densidade a posteriori da função de confiabilidade R , correspondente a densidade a priori de Bernardo (168), é a mesma obtida em (159) utilizando-se a densidade a priori de Tibshirani, isto é,

$$p(R | \underline{x}) = \frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right) \quad (169)$$

onde $\hat{W} = \ln t$

Portanto, R tem uma distribuição a posteriori Log-Gama Negativa, utilizando-se uma densidade a priori de Bernardo. Isto é,

$$R | \underline{x} \sim LGN \left(n; \sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} \right)$$

Notemos que esta densidade a posteriori também coincide com a densidade a posteriori correspondente à densidade a priori de Zellner (153). Uma comparação das densidades a posteriori obtidas, para as diferentes densidades a priori propostas, é dada na tabela a seguir.

Tabela 1: Comparação das densidades a posteriori $p(R | \underline{x})$ e respectivas densidades a priori $\pi(\beta, \gamma)$.

Densidades a Priori	$\pi(\beta, \gamma)$	$p(R \underline{x})$	$\hat{W}$
Jeffreys	$\frac{1}{\beta\gamma^2}$	$\frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right)$	$\frac{n}{n+1} \ln t$
Zellner	$\frac{1}{\gamma\beta} \exp\{\psi(1)\gamma\}$	$\frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right)$	$\ln t$
Tibshirani	$\frac{1}{\beta\gamma}$	$\frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right)$	$\ln t$
Bernardo	$\frac{1}{\beta\gamma}$	$\frac{\left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}}\right)^n}{\Gamma(n)} \cdot \left(\ln \frac{1}{R}\right)^{n-1} \cdot R \left(\sum \left(\frac{x_i}{t}\right)^{1/\hat{W}} - 1\right)$	$\ln t$

A utilização das densidades a priori estudadas neste trabalho, para o caso da função de confiabilidade do modelo de Weibull, permite-nos concluir que, para n grande, podemos escolher qualquer uma das densidades a priori propostas, uma vez que as densidades a posteriori convergirão para a mesma distribuição. Além disso, independentemente do valor de n , as densidades a priori de Zellner, Tibshirani e Bernardo produzem densidades a posteriori idênticas. Para n pequeno, um estudo comparativo dessas densidades a priori com a densidade a priori de Jeffreys através de ilustração numérica (gráficos e intervalos) se faz necessário e será abordado no próximo capítulo.

Zellner (1984) apresenta uma tabela com sua densidade a priori aplicada a algumas distribuições contínuas univariadas dadas a seguir. À tabela apresentada por Zellner, acrescentamos as densidades a priori de Jeffreys, Tibshirani e Bernardo para os modelos sugeridos por Zellner, assim como os resultados para o modelo de Weibull obtidos neste trabalho e os resultados para os modelos Gama e Valor Extremo (ver tabela 2). A tabela 2 mostra que nos casos univariados, as densidades a priori de Jeffreys, Zellner e Bernardo coincidem. A aplicação da priori de Tibshirani não é justificada neste caso. Em problemas multivariados, porém, esta coincidência não é verificada, principalmente, quando se trata de problemas envolvendo o modelo de Weibull e Gama.

Tabela 2: Priori não-informativa de Jeffreys, Zellner, Tibshirani e Bernardo

Distribuição	Priori de Jeffreys	Priori de Zellner	Priori de Tibshirani	Priori de Bernardo
1. Exponencial				
$f(x \theta) = \frac{1}{\theta} \exp\{-x/\theta\}$	$\pi(\theta) \propto \frac{1}{\theta}$	$\pi(\theta) \propto \frac{1}{\theta}$	—	$\pi(\theta) \propto \frac{1}{\theta}$
2. Normal				
$2a. f(x \theta) = \frac{1}{\sqrt{2\pi}} \exp\{-(x - \theta)^2/2\}$	$\pi(\theta) \propto \text{const}$	$\pi(\theta) \propto \text{const}$	—	$\pi(\theta) \propto \text{const}$
$2b. f(x \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-x^2/2\sigma^2\}$	$\pi(\sigma) \propto \frac{1}{\sigma}$	$\pi(\sigma) \propto \frac{1}{\sigma}$	—	$\pi(\sigma) \propto \frac{1}{\sigma}$
$2c. f(x \theta, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma^2}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$
3. Valor Extremo				
$f(x \phi, \varphi) = \frac{1}{\varphi} \exp\left\{\frac{x-\phi}{\varphi} - \exp\left(\frac{x-\phi}{\varphi}\right)\right\}$	$\pi(\phi, \varphi) \propto \frac{1}{\varphi^2}$	$\pi(\phi, \varphi) \propto \frac{1}{\varphi}$	$\pi(\phi, \varphi) \propto \frac{1}{\varphi}$	$\pi(\phi, \varphi) \propto \frac{1}{\varphi}$
4. Formas Gerais ($-\infty < x < \infty$)				
$4a. f(x \theta) = f(x - \theta)$	$\pi(\theta) \propto \text{const}$	$\pi(\theta) \propto \text{const}$	—	$\pi(\theta) \propto \text{const}$
$4b. f(x \sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$	$\pi(\sigma) \propto \frac{1}{\sigma}$	$\pi(\sigma) \propto \frac{1}{\sigma}$	—	$\pi(\sigma) \propto \frac{1}{\sigma}$
$4c. f(x \theta, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma^2}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$	$\pi(\theta, \sigma) \propto \frac{1}{\sigma}$
5. Gama				
$f(x \alpha, k) = \frac{1}{\Gamma(k)} \cdot \left(\frac{x}{\alpha}\right)^{k-1} \exp\left\{-\frac{x}{\alpha}\right\}$	$\pi(\alpha, k) \propto (k\alpha^2\psi'(k) - \frac{1}{\alpha})^{1/2}$	$\pi(\alpha, k) \propto \frac{1}{\alpha\Gamma(k)} \exp\left\{-\frac{(k-1)\psi(k)}{\Gamma(k)}\right\}$	$\pi(\alpha, k) \propto \sqrt{\psi'(k)}$	—
6. Weibull				
$6a. f(x p, \theta) = \frac{p}{\theta} x^{p-1} \exp\left\{-\frac{x^p}{\theta}\right\}$	$\pi(p, \theta) \propto \frac{1}{p\theta}$	$\pi(p, \theta) \propto p\theta^{-\frac{1}{p}} \exp\left\{-\frac{\psi(1)}{p}\right\}$	$\pi(p, \theta) \propto \frac{1}{p^{\frac{1}{p}}\theta}$	$\pi_p(p, \theta) \propto \frac{1}{p\theta}$
$6b. f(x \beta, \gamma) = \frac{\theta^{-\frac{1}{\gamma}}}{\gamma} x^{\frac{1}{\gamma}-1} \exp\left\{-\left(\frac{x}{\beta}\right)^{\frac{1}{\gamma}}\right\}$ onde $\gamma = \frac{1}{p}$ e $\beta = \theta^{1/p}$.	$\pi(\beta, \gamma) \propto \frac{1}{\beta\gamma^2}$	$\pi(\beta, \gamma) \propto \frac{1}{\beta\gamma} \exp\{-\psi(1)\gamma\}$	$\pi(\beta, \gamma) \propto \frac{1}{\beta\gamma}$	$\pi_{\theta}(p, \theta) \propto \frac{1}{p\theta\sqrt{1+(1+\theta)^2+2\psi(2)(1+\theta)+\psi'(2)}}$
$6c. f(x \delta, \lambda) = e^{\psi(2)} \frac{\delta}{\lambda^2} x^{\delta-1} \exp\left\{-e^{\psi(2)}\left(\frac{x}{\lambda}\right)^{\delta}\right\}$ onde $\delta = p$ e $\lambda = \theta^{1/p}$. $\exp\left\{\frac{\psi(2)}{p}\right\}$ são ortogonais.	$\pi(\delta, \lambda) \propto \frac{1}{\lambda}$	$\pi(\delta, \lambda) \propto \frac{\delta}{\lambda} \exp\left\{-\frac{\psi(2)-\psi(1)}{p}\right\}$	$\pi(\delta, \lambda) \propto \frac{1}{\lambda^2}$	$\pi(\delta, \lambda) \propto \frac{1}{\delta\lambda}$

Capítulo 5

ILUSTRAÇÃO NUMÉRICA

Vimos no Capítulo 4 que as densidades a priori propostas por Zellner, Tibshirani e Bernardo proporcionam densidades a posteriori, aplicadas à função de confiabilidade do modelo Weibull, coincidentes, portanto, essas densidades a priori são coerentes do ponto de vista de não-informação. Este resultado também se verifica quando comparamos essas densidades a priori com a densidade a priori de Jeffreys, desde que consideremos n suficientemente grande. Isto pode não se verificar quando n for pequeno, sendo necessário um estudo numérico para comparação. Consideremos o estudo comparativo entre as densidades a priori de Jeffreys e Zellner.

5.1. Comparação gráfica

Na figura 9, as densidades a posteriori para R são comparadas para (a) $n = 5$, (b) $n = 10$ e (c) $n = 25$, fixado o tempo $t = 2$ e os parâmetros $p = 2$ e $\theta = 4$, do modelo de Weibull dado em (4), valores estes sugeridos por Sinha/Guttman (1988). A densidade a priori recomendada é a densidade a priori proposta por Zellner pois é menos concentrada do que a densidade a priori de Jeffreys quando n é pequeno. Quando n é grande, as duas densidades a priori produzem densidades a posteriori coincidentes, confirmando-se a afirmação acima.

Figura 9 Densidades a posteriori de R com $t = 2$, $p = 2$ e $\theta = 4$, para (a) $n = 5$, (b) $n = 10$ e (c) $n = 25$.

Na figura 10, as densidades a posteriori para R são agora comparadas para diferentes valores de t . Verificamos que, independente do valor do tempo t , a densidade a priori de Zellner é menos informativa, portanto a densidade a priori recomendada. O mesmo resultado é verificado quando variamos os valores dos parâmetros p e θ do modelo Weibull, como mostra a figura 11.

Figura 10 Distribuições a posteriori de R com $n = 5$, $p = 2$ e $\theta = 4$ fixados, para
(a) $t = 0.001$, (b) $t = 1.2$, (c) $t = 4.5$ e (d) $t = 6$.

Figura 11 Distribuições a posteriori de R com $n = 5$ e $t = 2$ fixados, para
 (a) $p = 1, \theta = 4$, (b) $p = 2, \theta = 0.5$, (c) $p = 1, \theta = 100$ e
 (d) $p = 0.1, \theta = 0.5$.

5.2 - Comparação Pontual e Intervalar

Na tabela 3, diferentes estimadores pontuais, como média e moda a posteriori e o estimador de máxima verossimilhança (EMV), são comparados para diferentes valores de n . O EMV de R foi calculado usando o Método de Newton-Raphson (ver Apêndice E).

Tabela 3. Médias e modas a posteriori, estimador de máxima verossimilhança e o valor real da função de confiabilidade R para $t = 2$, $p = 2$ e $\theta = 4$

n	R	Modas a posteriori		Médias a posteriori		EMV
		Jeffreys	Zellner	Jeffreys	Zellner	
05	0.367879	0.420782	0.386197	0.441003	0.415286	0.4272874
10	0.367879	0.248813	0.243943	0.284686	0.28059	0.2604744
15	0.367879	0.144983	0.153760	0.179742	0.187752	0.0922286
20	0.367879	0.274706	0.273276	0.290964	0.289649	0.2882142
25	0.367879	0.342288	0.340001	0.351062	0.348914	0.3677271

Podemos considerar que não há diferença entre as médias a posteriori correspondente às densidades a priori de Jeffreys e Zellner, o mesmo ocorrendo com respeito à moda a posteriori.

Um critério proposto por Berger (1992), para a comparação da qualidade das densidades a priori é dado pelo estudo das probabilidades de cobertura freqüentista dos intervalos a posteriori.

A tabela 4 mostra as probabilidades de cobertura freqüentista dos intervalos a posteriori 95% e 90% da função de confiabilidade R para as densidades a priori de Jeffreys e Zellner.

Tabela 4. Probabilidades de cobertura freqüentista, dos intervalos a posteriori 95% e 90% de R para diferentes valores de n e fixados $p = 2$, $\theta = 4$ e $t = 2$.

n	intervalos a posteriori 95%		intervalos a posteriori 90%	
	Jeffreys	Zellner	Jeffreys	Zellner
05	98%	99%	95%	97%
10	99%	99%	97%	99%
15	100%	100%	95%	95%
20	98%	98%	97%	98%
25	99%	99%	96%	96%

O estudo comparativo das probabilidades de cobertura dos intervalos a posteriori 95% e 90% sugere que a densidade a priori de Zellner tem probabilidades de cobertura maiores do que a densidade a priori de Jeffreys para n pequeno e propriedades idênticas para n grande. Isto é uma forte evidência de que a densidade a priori de Jeffreys produz uma densidade a posteriori para R mais concentrada. Este resultado comprova o observado na comparação gráfica.

Um outro critério para comparação de densidades a priori não-informativas, sugerido neste trabalho, é dado pela comparação do tamanho dos intervalos a posteriori, onde verificamos a porcentagem de intervalos correspondentes a densidade a priori de Zellner que tem tamanho maior do que o intervalo correspondente a Jeffreys. Além disso, a diferença média dos comprimentos dos intervalos gerados são analisados. Os resultados obtidos são apresentados na Tabela 5.

Tabela 5: Porcentagem de intervalos a posteriori 95% e 90% para R via densidade a priori de Jeffreys que tem tamanho menor que os intervalos via Zellner e a diferença média entre os tamanhos dos intervalos via Jeffreys e Zellner, para diferentes valores de n e fixados $p = 2$, $\theta = 4$ e $t = 2$.

n	tamanho dos intervalos Jeffreys < Zellner		diferença média dos comprimentos dos interv.	
	95%	90%	95%	90%
05	70%	71%	0.008567	0.008031
10	73%	73%	0.001927	0.001710
15	-	-	0.000713	0.000621
20	-	-	0.000283	0.000246
25	-	-	0.000191	0.000164

A Tabela 5 vem confirmar os resultados obtidos pela cobertura freqüentista dados na Tabela 4. Para $n \leq 10$ a densidade a priori de Zellner é menos informativa, uma vez que, dos intervalos a posteriori gerados via Jeffreys e Zellner, aproximadamente, 70% dos intervalos de Zellner produziram amplitudes maiores, portanto o gráfico da densidade a posteriori será mais aberto. Isto também é comprovado através da existência de uma substancial diferença média entre os comprimentos dos intervalos via Jeffreys e Zellner. Para $n \geq 15$, verificamos uma diferença média entre os tamanhos dos intervalos gerados inferior a 0.001, o que não justificou a comparação dos tamanhos dos intervalos, podendo portanto serem considerados coincidentes, ou seja, as densidades a priori são coerentes do ponto de vista de não-informação.

A seguir, faremos um estudo comparativo das densidades a priori, verificando seu comportamento em relação ao tempo t .

Na tabela 6, os estimadores pontuais são agora comparados, para diferentes valores do tempo t , com $n = 5$, $p = 2$ e $\theta = 4$ fixados.

Tabela 6. Médias e modas a posteriori, estimador de máxima verossimilhança e o valor real da função de confiabilidade R para $n = 5$, $p = 2$ e $\theta = 4$.

t	R	Moda a posteriori		Média a posteriori		EMV
		Jeffreys	Zellner	Jeffreys	Zellner	
10^{-4}	1.0	6.91×10^{-5}	4.767×10^{-3}	0.069314	0.104168	0.9999999
0.8	0.852144	0.123292	0.074885	0.228435	0.190526	0.8585627
1.2	0.697676	0.998172	0.993670	0.997717	0.992112	0.7216486
1.5	0.569783	0.823257	0.750705	0.797221	0.723437	0.6091187
2.0	0.367879	0.420782	0.386197	0.441003	0.415286	0.4272874
2.5	0.209610	0.236526	0.241622	0.308787	0.312340	0.2746732
3.5	0.0467706	0.108530	0.138442	0.217330	0.239570	0.0881504
4.5	0.0063297	0.066319	0.100651	0.183193	0.211269	0.0204233
6.0	1.234×10^{-4}	0.041067	0.075279	0.159343	0.190857	0.0012633

A Tabela 6 mostra não haver similaridade entre as estimativas pontuais, existindo uma diferença significativa entre as medidas a posteriori e o EMV, principalmente, nos pontos próximos de $R = 0$, $R = 1$ e $t = 1$. Nestes pontos, onde é impossível definir a densidade a posteriori, o EMV é um estimador consistente de R , porém, para valores de t em torno de 2 e, conseqüentemente, valores de R pertencentes ao domínio da densidade a posteriori, os estimadores Bayesianos são mais consistentes do que o EMV.

Tabela 7. Probabilidades de cobertura freqüentista dos intervalos a posteriori para R com $p = 2$, $\theta = 4$ e $n = 5$.

t	R	intervalos a posteriori 95%		intervalos a posteriori 95%	
		Jeffreys	Zellner	Jeffreys	Zellner
10^{-4}	1	0%	0%	0%	0%
0.8	0.852144	16%	12%	13%	9%
1.2	0.697676	5%	11%	5%	8%
1.5	0.569783	66%	86%	53%	76%
2.0	0.367879	99%	100%	96%	99%
2.5	0.20961	100%	100%	97%	99%
3.5	0.0467706	86%	79%	72%	60%
4.5	0.0063297	31%	11%	17%	3%
6.0	1.23×10^{-4}	0%	0%	0%	0%

Como foi observado na estimação pontual, valores de R e t próximos aos pontos onde a densidade a posteriori, não é definida, acarretam probabilidades de cobertura muito menores do que o nível de credibilidade indica, porém, um outro fato relevante pode ser verificado. Observemos que, para valores de t entre 3 e 5, as probabilidades de cobertura correspondentes a densidade a priori de Jeffreys são maiores que as coberturas correspondentes a Zellner, embora consideremos a densidade a priori de Zellner menos informativa. Para garantirmos a veracidade da afirmação, propomos o critério da comparação do tamanho dos intervalos a posteriori, utilizado na Tabela 5.

Tabela 8: Porcentagem de intervalos a posteriori 95% e 90% para R via densidade a priori de Jeffreys que tem tamanho menor que os intervalos via Zellner e a diferença média entre os tamanhos dos intervalos via Jeffreys e Zellner, para diferentes valores de t e fixados $p = 2$, $\theta = 4$ e $n = 5$.

t	tamanho dos intervalos Jeffreys < Zellner		diferença média dos comprimentos dos interv.	
	95%	90%	95%	90%
10^{-4}	100%	100%	0.074240	0.070626
0.8	84%	83%	0.034311	0.028631
1.2	100%	100%	0.034321	0.028967
1.5	91%	91%	0.050010	0.043761
2.0	70%	71%	0.008567	0.008031
2.5	91%	92%	0.012901	0.012488
3.5	100%	100%	0.025849	0.025050
4.5	100%	100%	0.035286	0.034192
6.0	100%	100%	0.039869	0.038634

Pela tabela acima, podemos então comprovar realmente que a densidade a posteriori correspondente a densidade a priori de Zellner produz intervalos maiores do que a densidade a priori de Jeffreys, ou seja, que a densidade a priori de Zellner é menos informativa, como afirmamos anteriormente. Como explicar, então, a ocorrência de coberturas freqüentistas menores para a densidade a priori de Zellner obtidas na Tabela 7?

Se observarmos os estimadores pontuais da Tabela 6, verificaremos que a moda a posteriori se aproxima muito do verdadeiro valor de R , para os valores de t em que a cobertura é maior com a densidade a priori de Jeffreys, isto é, a moda é um estimador consistente e portanto, a densidade a posteriori será mais concentrada próxima ao verdadeiro valor de R . Logo, seus intervalos a posteriori,

mesmo tendo comprimento menor, conterão R mais vezes. Os resultados obtidos em nosso estudo, evidenciam que o critério de cobertura freqüentista, proposto por Berger, não é suficiente para a comparação de densidades a priori não-informativas, quando utilizamos o modelo de Weibull para os dados, sendo necessário, então, a utilização do critério da comparação dos tamanhos dos intervalos.

A seguir, realizaremos um estudo comparativo das densidades a priori com respeito aos valores dos parâmetros p e θ do modelo de Weibull, considerando distintos conjuntos de dados, cada um consistindo de $n = 5$ observações.

Tabela 9. Médias e modas a posteriori, estimadores de máxima verossimilhança e o valor real da função de confiabilidade R para $n = 5$ e $t = 2$.

p	θ	R	Modas a posteriori		Médias a posteriori		EMV
			Jeffreys	Zellner	Jeffreys	Zellner	
4.0	1000.0	0.984123	0.775653	0.7277340	0.748248	0.701091	0.988406
4.0	16.0	0.367879	0.437916	0.4255010	0.453954	0.444556	0.835069
4.0	8.0	0.135469	0.038985	0.0692550	0.157168	0.185738	0.0312386
4.0	6.0	0.069443	0.123846	0.1550000	0.228846	0.251519	0.0658976
4.0	3.0	0.004823	0.152440	0.1817040	0.249684	0.270464	0.0784009
2.0	200.0	0.980198	0.968614	0.9448830	0.961375	0.932959	0.9948379
2.0	4.0	0.367879	0.224226	0.2303190	0.300213	0.304460	0.2606677
2.0	2.0	0.135335	0.189420	0.2048780	0.275888	0.286712	0.2080063
2.0	0.5	0.0003354	0.001248	0.0061550	0.088070	0.108245	0.0510630
1.0	100.0	0.980199	0.999274	0.9975730	0.999093	0.996970	0.9987629
1.0	4.0	0.606531	0.860421	0.7991100	0.837020	0.772106	0.713126
1.0	1.0	0.135335	0.0185565	0.0406843	0.132047	0.158955	0.043631
1.0	0.5	0.0183156	0.0917911	0.0935794	0.204317	0.205733	0.161038
0.1	2.0	0.585150	0.999680	0.9981240	0.999488	0.997657	—
0.1	0.5	0.117238	0.998523	0.9944930	0.998155	0.993135	—

Como observado anteriormente, valores de R próximos a $R = 1$ e $R = 0$, o estimador recomendado seria o EMV.

Observamos, também que, para valores do parâmetro p próximos de zero, os estimadores a posteriori não produzem boas estimativas, mesmo R assumindo valores no domínio da densidade a posteriori. Além disso, é impossível a obtenção do EMV de R .

Tabela 10. Probabilidades de cobertura freqüentista dos intervalos a posteriori para R com $t = 2$ e $n = 5$.

p	θ	R	Intervalos a posteriori 90%		Intervalos a posteriori 90%	
			Jeffreys	Zellner	Jeffreys	Zellner
4.0	1000.0	0.984123	0%	0%	0%	0%
4.0	16.0	0.367879	100%	100%	100%	100%
4.0	8.0	0.135469	99%	99%	94%	94%
4.0	6.0	0.069443	94%	88%	78%	72%
4.0	3.0	0.004822	35%	9%	19%	3%
2.0	200.0	0.980198	91%	41%	85%	22%
2.0	4.0	0.367879	99%	100%	96%	99%
2.0	2.0	0.135335	98%	98%	91%	95%
2.0	0.5	0.000335	95%	74%	91%	66%
1.0	100.0	0.980199	0%	5%	0%	3%
1.0	4.0	0.606531	54%	69%	47%	60%
1.0	1.0	0.135335	81%	85%	75%	80%
1.0	0.5	0.018315	62%	73%	54%	64%
0.1	2.0	0.585150	2%	2%	2%	2%
0.1	0.5	0.117238	9%	9%	8%	8%

Quando variamos os valores p e θ , verificamos novamente os resultados obtidos na Tabela 8, onde a cobertura freqüentista, além de não ser a desejada, obtemos cobertura maior com a densidade a priori de Jeffreys do que com a densidade a priori de Zellner, devido à consistência da moda a posteriori (ver Tabela 9).

A Tabela 10 nos mostra que a probabilidade de cobertura dos intervalos a posteriori para R depende dos parâmetros p e θ do modelo de Weibull, portanto os intervalos a posteriori 95% e 90% para R , não podem ser considerados intervalos de confiança.

Pela Tabela 11 fica evidenciado que a densidade a priori de Zellner produz intervalos a posteriori de comprimentos maiores do que a densidade a priori de Jeffreys, mesmo para os casos em que a cobertura freqüentista foi menor, nos levando a concluir que a densidade a priori de Zellner é a menos informativa, portanto, a densidade a priori recomendada.

Tabela 11. Porcentagem de intervalos a posteriori 95% e 90% para R via densidade a priori de Jeffreys, que tem tamanho menor que os intervalos via Zellner e a diferença média entre os tamanhos dos intervalos via Jeffreys e Zellner, para diferentes valores de p e θ fixados $t = 2$ e $n = 5$.

p	θ	Tamanho dos intervalos Jeffreys < Zellner		Diferença média dos comprimentos dos interv.	
		95%	90%	95%	90%
4.0	1000.0	100%	100%	0.062338	0.053639
4.0	16.0	91%	91%	0.002363	0.002258
4.0	8.0	97%	97%	0.013112	0.012677
4.0	6.0	100%	100%	0.016592	0.016060
4.0	3.0	100%	100%	0.038568	0.037370
2.0	200.0	100%	100%	0.044751	0.037576
2.0	4.0	70%	71%	0.008567	0.008031
2.0	2.0	91%	91%	0.027866	0.026526
2.0	0.5	100%	100%	0.067849	0.055532
1.0	100.0	100%	100%	0.004673	0.003893
1.0	4.0	100%	100%	0.065215	0.056481
1.0	1.0	80%	80%	0.025159	0.021157
1.0	0.5	98%	97%	0.034500	0.026044
0.1	2.0	99%	100%	0.003091	0.002597
0.1	0.5	88%	88%	0.022233	0.019166

Concluimos o estudo das densidades a priori não-informativas para o caso específico da função de confiabilidade aplicada ao modelo de Weibull, tratadas neste trabalho, que as densidades a priori de Zellner, Tibshirani e Bernardo produzem densidades a posteriori menos concentradas do que a densidade a priori de Jeffreys,

portanto, são menos informativas; contudo, merece também destacarmos as vantagens e desvantagens da utilização dessas densidades a priori no caso geral, como veremos a seguir.

5.3 - Discussão

A densidade a priori proposta por Jeffreys além de ser universalmente conhecida é invariante sob reparametrizações (diferenciáveis e um a um), e tem proporcionado sucesso em problemas unidimensionais. Contudo, nota-se algumas dificuldades com a densidade a priori quando o parâmetro é multidimensional. No caso dos parâmetros θ e σ da Normal, Jeffreys ignora o uso da regra geral (42) em termos de julgamento a priori de que θ e σ são independentes para obter a densidade a priori conjunta $1/\sigma$.

A utilização da densidade a priori de Tibshirani requer uma informação a priori de qual será o parâmetro nuisance, além de uma reparametrização ortogonal, contudo, ortogonalização total nem sempre é possível (Cox e Reid, 1987). Devido a estas dificuldades de ortogonalização, às vezes impossível, a formulação de tal densidade a priori pode ser impraticável. Uma outra ressalva à utilização dessa densidade a priori é que, embora pareça natural dizer que os parâmetros ortogonais são, a priori, independentes (ver Apêndice C), isto não ocorre necessariamente no procedimento proposto por Tibshirani.

A densidade a priori referencial proposta por Bernardo é muito pouco utilizada devido a algumas dificuldades como:

- não-unicidade da densidade a priori, devido à escolha do parâmetro de interesse,
- o método de Bernardo depende do atual parâmetro de interesse;
- caso o interesse seja o parâmetro global, a densidade a priori de Bernardo coincide com a densidade a priori de Jeffreys, para a qual enunciamos as dificuldades cima;

- não-existência. Veja, por exemplo, o problema das componentes da variância (Berger e Bernardo, 1989).

Mostramos que a densidade a priori proposta por Zellner maximiza a informação proveniente dos dados. Isto mostra que a informação na função de verossimilhança é enfatizada pela densidade a priori. Por este motivo, a utilização da densidade a priori de Zellner seria mais facilmente aceita por não-Bayesianos. Além disso, as condições que reflitam alguma informação inicial e que devam ser consideradas, podem facilmente ser incorporadas para se obter a densidade a priori de Zellner, isto é, a densidade a priori de Zellner pode ser relativamente não-informativa ou informativa pelo uso de condições representando informações disponíveis. Uma característica fundamental do procedimento proposto por Zellner é a obtenção de uma densidade a priori própria, o que não se verifica nos outros procedimentos. A densidade a priori de Zellner não requer a existência de condições de regularidade, como requerem as densidades a priori de Jeffreys e Tibshirani.

Quanto à invariância, a densidade a priori proposta por Zellner é invariante para uma classe de transformações muito menor que a densidade a priori de Jeffreys, porém, Rao (1987) escreveu “A escolha de uma métrica naturalmente depende de um particular problema sob investigação, e invariância pode, ou não, ser relevante”.

Apêndice A

MÉTODO DE LAPLACE PARA APROXIMAÇÃO DE INTEGRAIS

Freqüentemente, não conseguimos resolver as integrais analiticamente, isto é, não somos capazes de determinar uma fórmula fechada para a integral. Uma possibilidade, para contornarmos este problema, é utilizarmos métodos numéricos de aproximação. Na Inferência Bayesiana, contudo, necessitamos de uma expressão para tais integrais e não do valor numérico da integral.

Um dos métodos de aproximação mais populares para integrais, é dado pelo Método de Laplace (Tierney e Kadane, 1986).

O Método de Laplace para Aproximação de Integrais constitui uma técnica muito utilizada na Inferência Bayesiana, para o cálculo de densidades a posteriori marginais, momentos a posteriori e densidades preditivas.

É baseado na expansão de Taylor de 2ª ordem, aplicada ao termo exponencial do integrando. Neste apêndice, damos uma breve e informal descrição do método e suas propriedades (para mais detalhes, ver Kass, Tierney e Kadane, 1990).

Seja $h : R^m \rightarrow R$ uma função diferenciável de $\underline{\phi}$ com $-h$ tendo um único ponto de máximo $\hat{\underline{\phi}}$; o Método de Laplace aproxima uma integral da forma,

$$I = \int_0^\infty f(\underline{\phi}) \exp\{-nh(\underline{\phi})\} d\underline{\phi} \quad (A1)$$

expandindo h e f sobre $\hat{\underline{\phi}}$, onde $f(\underline{\phi})$ é uma função um a um.

Para o caso em que $\underline{\phi}$ é um parâmetro unidimensional obtemos:

$$\int_0^\infty f(\phi) \exp\{-nh(\phi)\} d\phi \cong \left(\frac{2\pi\sigma^2}{n}\right)^{1/2} f(\hat{\underline{\phi}}) \exp\{-nh(\hat{\underline{\phi}})\} \quad (A2)$$

onde $\sigma = h''(\hat{\phi})^{-1/2}$

Kass, Tierney e Kadane argumentam que

$$I = \hat{I} [1 + O(n^{-1})] \quad (A3)$$

onde $\hat{I} = \left(\frac{2\pi\sigma^2}{n}\right)^{1/2} f(\hat{\phi}) \exp\{-nh(\hat{\phi})\}$

Para o caso m-dimensional,

$$\int_0^\infty f(\underline{\phi}) \exp\{-nh(\underline{\phi})\} d\underline{\phi} \cong \left(\frac{2\pi}{n}\right)^{m/2} \det(\Sigma)^{1/2} f(\hat{\phi}) \exp\{-nh(\hat{\phi})\} \quad (A4)$$

onde $\Sigma^{-1} = D^2h(\hat{\phi})$ (a Hessiana de h para $\hat{\phi}$)

Na realidade, o Método de Laplace transforma um problema de integração num problema de maximização.

A aproximação será muito precisa, quando a distribuição a posteriori for aproximadamente normal. Caso isto não ocorra, para conseguirmos boas aproximações, será necessário uma boa escolha da parametrização, especialmente para amostras de tamanho pequeno. Achcar (1990) obteve uma fórmula para checar o erro da Aproximação de Laplace e também uma reparametrização para se chegar a aproximações precisas.

Apêndice B

DISTRIBUIÇÃO LOG-GAMA NEGATIVA

A distribuição Log-Gama Negativa tem sido muito utilizada em análise de confiabilidade de sistemas por Mastran (1976), Martz e Waller (1982).

A variável aleatória X tem distribuição LOG-GAMA NEGATIVA com parâmetros k e λ , se $Y = \ln\left(\frac{1}{X}\right)$ tem distribuição Gama, que denotamos por $Y \sim \Gamma(k, \lambda)$, cuja função de densidade é dada por

$$f(y | \lambda, k) = \frac{\lambda^k}{\Gamma(k)} \cdot y^{k-1} \cdot e^{-y\lambda} \quad (B1)$$

Por outro lado, a função de densidade de probabilidade da variável aleatória $X = e^{-Y}$ é dada por,

$$f(x | \lambda, k) = \frac{\lambda^k}{\Gamma(k)} \cdot \left(\ln \frac{1}{x}\right)^{k-1} x^{\lambda-1}, \quad 0 < x < 1, \quad (B2)$$

e denotada por $X \sim LGN(k, \lambda)$.

As seguintes medidas descritivas relacionadas à v.a. X são dadas abaixo:

$$\text{i) média} = \left(1 + \frac{1}{\lambda}\right)^{-k}$$

$$\text{ii) variância} = \left(1 + \frac{2}{\lambda}\right)^{-k} - \left(1 + \frac{1}{\lambda}\right)^{-2k} \quad (B3)$$

$$\text{iii) moda} = \exp\left\{-\frac{k-1}{\lambda-1}\right\}$$

Suponhamos que a Função de Confiabilidade R tenha distribuição a posteriori Log-Gama Negativa com parâmetros k e λ dada por (B2), e densidade a posteriori $p(R | \underline{x})$, então o intervalo a posteriori $100(1 - \alpha)\%$, (R_*, R^*) , para R pode ser obtido usando-se a distribuição χ^2 . De fato, o intervalo a posteriori (R_*, R^*) é definido por

$$P(R_* \leq R \leq R^* | \underline{x}) = 1 - \alpha \quad (B4)$$

Daí

$$P(2.\lambda.\ln \frac{1}{R^*} \leq 2.\lambda.\ln \frac{1}{R} \leq 2.\lambda.\ln \frac{1}{R_*} | \underline{x}) = 1 - \alpha \quad (B5)$$

Mas como $2.\lambda.\ln \frac{1}{R}$ tem uma distribuição $\chi^2_{(2k)}$, então o intervalo desejado tem limite inferior e superior,

$$R_* = \exp \left\{ -\frac{\chi^2_{\alpha/2}(2k)}{2\lambda} \right\} \quad \text{e} \quad R^* = \exp \left\{ -\frac{\chi^2_{1-\alpha/2}(2k)}{2\lambda} \right\}, \quad (B6)$$

respectivamente.

Apêndice C

REPARAMETRIZAÇÃO ORTOGONAL

Introdução

Neste apêndice apresentaremos algumas considerações genéricas sobre os parâmetros ortogonais, úteis à formulação da densidade a priori de Tibshirani (para maiores detalhes, ver Cox e Reid, 1987). Em seu artigo, Cox e Reid abordam o problema de ortogonalização de parâmetros de uma forma geral, explorando a conexão entre ortogonalidade de parâmetros e a teoria assintótica da inferência condicional.

Considere uma variável aleatória X com densidade $f(x | \underline{\phi})$ onde $\underline{\phi} = (\phi_1, \dots, \phi_p)$ é um vetor de parâmetros desconhecido e $L(\underline{\phi} | \underline{x})$ a função de verossimilhança.

Seja $\underline{\phi} = (\underline{\delta}, \underline{\lambda})$ uma partição tal que $\underline{\delta} = (\delta_1, \delta_2, \dots, \delta_k)$ e $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{p-k})$

Definição: O vetor de parâmetros $\underline{\delta}$ é ORTOGONAL ao vetor de parâmetros $\underline{\lambda}$ se os elementos da matriz de Informação de Fisher $I(\underline{\phi})$ satisfazem a seguinte relação:

$$i_{\delta_s \lambda_t} = \frac{1}{n} E_{\underline{\phi}} \left(- \frac{\partial^2 \log L(\underline{\phi} | \underline{x})}{\partial \delta_s \partial \lambda_t} \right) = 0 \quad (C1)$$

para $s = 1, \dots, k$ e $t = 1, \dots, p - k$. Neste caso, $I(\underline{\phi})$ será dada por

$$I(\underline{\delta}, \underline{\lambda}) = \begin{bmatrix} I_1(\underline{\delta}, \underline{\delta}) & \mathbf{0}_1 \\ \mathbf{0}_2 & I_2(\underline{\lambda}, \underline{\lambda}) \end{bmatrix} \quad (C2)$$

onde $I_1(\underline{\delta}, \underline{\delta})$ e $I_2(\underline{\lambda}, \underline{\lambda})$ são matrizes de Fisher $k \times k$ e $(p-k) \times (p-k)$ respectivamente, $\mathbf{0}_1$ e $\mathbf{0}_2$ são matrizes nulas $(p-k) \times (p-k)$ e $k \times k$ respectivamente.

Se a expressão (C1) for válida para todo $\underline{\phi}$ no espaço paramétrico, a ortogonalidade é dita GLOBAL ou TOTAL e a matriz de Fisher é dada por (C3); se o for somente para $\underline{\phi} = \underline{\phi}_0$, então os vetores $\underline{\phi}_1$ e $\underline{\lambda}$ são LOCALMENTE ORTOGONAIS em $\underline{\phi}$ e neste caso a matriz de Fisher é dada por (C2).

$$I(\underline{\delta}, \underline{\lambda}) = \begin{bmatrix} i_{\delta_1 \delta_1} & \mathbf{0} \\ \mathbf{0} & i_{\lambda_p \lambda_p} \end{bmatrix} \quad (C3)$$

Observações:

1- A definição de ortogonalidade pode ser expressa para mais do que dois conjuntos de parâmetros.

2- Ortogonalidade Local $\rightarrow$ matriz de Informação de Fisher bloco-diagonal.

Ortogonalidade Global $\rightarrow$ matriz de Informação de Fisher diagonal.

Se $\underline{\phi} = (\delta, \lambda)$, então δ é ortogonal a $\lambda \Leftrightarrow I(\delta, \lambda)$ for diagonal

Exemplo: Os parâmetros de locação-escala θ e σ da distribuição Normal são ortogonais.

3- Enquanto ortogonalidade local pode ser sempre alcançada, ortogonalidade global é possível somente em casos especiais.

Conseqüências da ortogonalidade:

Seja $\underline{\delta}$ e $\underline{\lambda}$ parâmetros ortogonais, Cox e Reid resumiam as conseqüências da ortogonalidade do seguinte modo:

- i) os estimadores de máxima verossimilhança $\hat{\underline{\delta}}$ e $\hat{\underline{\lambda}}$ são assintoticamente independentes;
- ii) o desvio-padrão assintótico de $\hat{\underline{\delta}}, \sigma_{\hat{\underline{\delta}}}$, não depende de $\underline{\lambda}$ ser ou não fixado;
- iii) a resolução numérica para $(\underline{\delta}, \underline{\lambda})$ é facilitada, por exemplo, em regressão não linear;
- iv) o estimador de máxima verossimilhança de $\underline{\delta}$ quando $\underline{\lambda}$ é fixado, $\hat{\underline{\delta}}(\underline{\lambda})$, não varia muito com $\underline{\lambda}$;

Observação: A propriedade iv) diz que, na parametrização ortogonal, não é preciso nos preocuparmos com a influência do parâmetro $\underline{\lambda}$ na inferência sobre $\underline{\delta}$. Portanto, a ortogonalidade é uma forma de controlar a influência do parâmetro “nuisance” na estimativa do parâmetro de interesse.

- v) se $\hat{\underline{\delta}}(\underline{\lambda}) = \hat{\underline{\delta}}$ para todo $\underline{\lambda}$ fixado, então $\underline{\delta}$ e $\underline{\lambda}$ são ortogonais.

Obtenção de parâmetros ortogonais

Se os parâmetros não forem naturalmente ortogonais, é possível, em alguns casos, obter uma reparametrização que produza a ortogonalidade desejada. Cox e

Reid abordam o problema via equações diferenciais, discutindo a solução quando $k = 1$ e o interesse recai sobre δ , sendo $\underline{\lambda}$ o vetor de parâmetros “nuisance”.

Suponhamos o vetor $\underline{\theta} = (\theta_1, \underline{\theta}_2)$ com θ_1 escalar sendo o parâmetro de interesse e $\underline{\theta}_2 = (\theta_{21}, \theta_{22}, \dots, \theta_{2p})$ o vetor de parâmetros “nuisance”, com $q = p - 1$. Originalmente, a função de verossimilhança L é escrita em termos de $\theta_1, \theta_{21}, \theta_{22}, \dots, \theta_{2p}$ e o objetivo é reparametrizar o modelo obtendo $(\delta, \lambda_1, \lambda_2, \dots, \lambda_p)$, com $\delta = \theta_1$ ortogonal a λ_t , $t = 1, 2, \dots, q$. O critério proposto consiste de um sistema com q equações da forma

$$\sum_{r=1}^q i_{\theta_{2r}\theta_{2s}}^* \frac{\partial \theta_{2r}}{\partial \theta_1} = -i_{\theta_1\theta_{2s}}^*, \quad s = 1, \dots, q. \quad (C4)$$

onde os i^* são os elementos da matriz de informação calculados na parametrização $(\theta_1, \underline{\theta}_2)$ e a suposição de que a transformação de $(\theta_1, \underline{\theta}_2) \rightarrow (\delta, \underline{\lambda})$ tenha Jacobiano não nulo.

Desta forma, o conjunto original $(\theta_1, \theta_{21}, \theta_{22}, \dots, \theta_{2p})$ é reparametrizado para $(\delta, \lambda_1, \lambda_2, \dots, \lambda_p)$. Quando o sistema admite solução, esta é arbitrária, sendo interessante escolher $\underline{\lambda}$ com uma interpretação conveniente.

Apêndice D

ALGUMAS INTEGRAIS USUAIS

$$\Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du, \quad k > 0 \quad (D1)$$

$$\Gamma(1 + \theta) = \int_0^\infty u^\theta e^{-u} du \quad (D2)$$

$$\Gamma(n) = (n-1)! \quad , n \in \mathbb{Z} \quad (D3)$$

$$\Gamma'(1) = \int_{-\infty}^{+\infty} u \exp\{u - e^{-u}\} du = \int_0^\infty \ln w e^{-w} dw = -\gamma \quad (D4)$$

onde $\gamma = 0.5772\dots$

$$\Gamma''(1) = \int_{-\infty}^{+\infty} u^2 \exp\{u - e^{-u}\} du = \int_0^\infty (\ln w)^2 e^{-w} dw = \frac{\pi^2}{6} + \gamma^2 \quad (D5)$$

$$\int_0^\infty \frac{e^{\theta u} u^{k-1} e^{-u}}{\Gamma(k)} du = (1 - \theta)^{-k} \quad (D6)$$

$$\frac{1}{\Gamma(k)} \int_{-\infty}^{+\infty} e^{\theta u} \exp\{k u - e^u\} du = \frac{1}{\Gamma(k)} \int_0^\infty w^{\theta+k-1} e^{-w} dw = \frac{\Gamma(\theta+k)}{\Gamma(k)} \quad (D7)$$

$$\int_0^\infty u^{p-1} e^{-au^k} du = \frac{1}{k} a^{-p/k} \Gamma\left(\frac{p}{k}\right) \quad (D8)$$

$$\int_0^1 \left(\ln \frac{1}{w}\right)^{p-1} w^{a-1} dw = a^{-p} \Gamma(p) \quad (D9)$$

$$\int_0^\infty u^{-(p+1)} e^{-au^{-k}} du = \frac{1}{k} a^{-p/k} \Gamma\left(\frac{p}{k}\right) \quad (D10)$$

$$\psi(k) = \frac{d}{dk} \ln \Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)} \quad (D11)$$

$$\psi'(k) = \frac{d^2}{dk^2} \ln \Gamma(k) \quad (\text{D12})$$

$$\psi^{(i)}(k) = \int_0^\infty u^{k-1} (\ln u)^{i+1} \epsilon^{-u} du \quad (\text{D13})$$

$$\psi(z+1) = \psi(z) + \frac{1}{z} \quad , \quad z > 0 \quad (\text{D14})$$

$$\psi'(z+1) = \psi'(z) + \frac{1}{z^2} \quad , \quad z > 0 \quad (\text{D15})$$

$$\psi(1) = \Gamma'(1) = -\gamma \quad (\text{D16})$$

$$\psi(2) = \Gamma'(2) = -\gamma + \frac{1}{2} \quad (\text{D17})$$

$$\Gamma'(2) = \psi(1) + \frac{1}{2} = -\gamma + \frac{1}{2} \quad (\text{D18})$$

Apêndice E

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Este Apêndice contém os programas do Minitab 8.2 e Eureka 1.0 relativos às análises de dados discutidas no capítulo 5.

PROGRAMA 1:

a) Obtenção das densidades a posteriori de R , cálculo dos coeficientes, correspondentes às densidades a priori de Jeffreys e Zellner.

b) Obtenção das medidas descritivas a posterioris, correspondentes às densidades a priori de Jeffreys e Zellner.

MTB> #k1 é o tamanho da amostra n

MTB> #k20 é o valor do tempo t

MTB> #k21 é o valor do parâmetro $a = p$

MTB> #k22 é o valor do parâmetro $b = \theta^{1/p}$

MTB> #k25 é o valor de $\Gamma(k1) = (k1 - 1)!$

MTB>

MTB> #cálculo dos coeficientes k5 e k6 da densidade a posteriori

MTB> # $p(R|x) = k6 * ((\ln(1/R))^{k1-1}) * R^{k5}$

MTB>

MTB> #cálculo das medidas descritivas

MTB> #k7 é a média a posteriori

```

MTB> #k8 é a variância a posteriori
MTB> #k9 é a moda a posteriori
MTB>
MTB> Let k1=5
MTB> Let k20=2
MTB> Let k22=2
MTB> Let k25=24
MTB> # gerando uma amostra aleatória de tamanho k1 da distribuição Weibull
MTB> random k1 c1;
SUBC> weibull k21 k22.
MTB> print c1
MTB>
MTB> # cálculo da densidade a posteriori e das medidas descritivas via densidade
MTB> # a priori de Jeffreys
MTB> let c2=c1/k20
MTB> let k2=k1/(k1+1)
MTB> let k3=k2*log(k20)
MTB> let c3=c2**(1/k3)
MTB> let k4=sum(c3)
MTB> let k5=k4-1
MTB> let k6=(k4**k1)/k25
MTB> print k5 k6
MTB> #cálculo das medidas descritivas
MTB> let k7=(1+1/k4)**(-k1)
MTB> let k8=(1+2/k4)**(-k1)-k7**2
MTB> let k9=exp(-(k1-1)/k5)
MTB> print k7 k8 k9
MTB>
MTB> #cálculo da densidade a posteriori e das medidas descritivas via densidade

```



```

MTB> # a priori de Zellner
MTB> let k3=loge(k20)
MTB> let c3=c2**(1/k3)
MTB> let k4=sum(c3)
MTB> let k5=k4-1
MTB> let k6=(k4**k1)/k25
MTB> print k5 k6
MTB> #cálculo das medidas descritivas
MTB> #k7 é a média a posteriori
MTB> #k8 é a variância a posteriori
MTB> #k9 é a moda a posteriori
MTB> let k7=(1+1/k4)**(-k1)
MTB> let k8=(1+2/k4)**(-k1)-k7**2
MTB> let k9=exp(-(k1-1)/k5)
MTB> print k7 k8 k9

```

PROGRAMA 2: Obtenção dos intervalos a posteriori 95% e 90%, para a função de confiabilidade R , correspondentes às densidades a priori de Jeffreys e Zellner, com a finalidade de comparar as respectivas probabilidades de cobertura.

```
MTB> #k1 é o tamanho da amostra  $n$ 
MTB> #k20 é o valor do tempo  $t$ 
MTB> #k21 é o valor do parâmetro  $a = p$ 
MTB> #k22 é o valor do parâmetro  $b = \theta^{1/p}$ 
MTB> #k25 é o valor de  $\Gamma(k1) = (k1 - 1)!$ 
MTB> #K10 é o verdadeiro valor de  $R$ 
MTB> #k15 é um contador
MTB> #k16 é a porcentagem de cobertura do intervalo a posteriori 95% via den-
MTB> # sidade a priori de Jeffreys
MTB> #k17 é a porcentagem de cobertura do intervalo a posteriori 90% via den-
MTB> # sidade a priori de Jeffreys
MTB> #k18 é a porcentagem de cobertura do intervalo a posteriori 95% via den-
MTB> # sidade a priori de Zellner
MTB> #k19 é a porcentagem de cobertura do intervalo a posteriori 90% via den-
MTB> # sidade a priori de Zellner
MTB> #Interv é umaMacro para cálculo dos intervalos a posteriori em cada si-
MTB> # mulação
MTB>
MTB> let k1=5
MTB> let k20=2
MTB> let k21=2
MTB> let k22=2
MTB> let k25=24
MTB> let k10=exp(-(k20/k22)**k21)
```

```

MTB>
MTB> let k15=1
MBT> exec 'interv.' 100
MTB> let c18=(c10<k10)&(c11>k10)
MTB> let c19=(c12<k10)&(c13>k10)
MTB> let c20=(c14<k10)&(c15>k10)
MBT> let c21=(c16<k10)&(c17>k10)
MTB> let k16=sum(c18)/100
MTB> let k17=sum(c19)/100
MTB> let k18=sum(c20)/100
MBT> let k19=sum(c21)/100
MTB> print k16 k17 k18 k19

```

File Interv

```

MTB> #gerando uma amostra aleatória de tamanho k1 da distribuição Weibull
MTB> random k1 c1;
SUBC> weibull k21 k22.
MTB> let c2=c1/k20
MTB>
MTB> #cálculo do intervalo a posteriori via densidade a priori de Jeffreys
MBT> let k2=k1/(k1+1)
MTB> let k3=k2*loge(k20)
MTB> let c3=c2**(1/k3)
MTB> let k4=sum(c3)
MBT>
MTB> #k5 é o limite inferior do interv. a posteriori 95% (k5;k6) para R
MTB> #k6 é o limite superior do interv. a posteriori 95% (k5;k6) para R
MTB> let k5=exp(-20.483/(2*k4))

```

```

MTB> let k6=exp(-3.247/(2*k4))
MTB> let c10(k15)=k5
MTB> let c11(k15)=k6
MTB>
MTB> #k7 é o limite inferior do interv. a posteriori 90% (k7;k8) para R
MTB> #k8 é o limite superior do interv. a posteriori 90% (k7;k8) para R
MTB> let k7=exp(-18.307/(2*k4))
MTB> let k8=exp(-3.940/(2*k4))
MTB> let c12(k15)=k7
MTB> let c13(k15)=k8
MTB> #cálculo do intervalo a posteriori via densidade a priori de Zellner
MTB> let k30=loge(k20)
MTB> let c30=c2*(1/k30)
MTB> let k40=sum(c30)
MTB>
MTB> #k50 é o limite inferior do interv. a posteriori 95% (k50;k60) para R
MTB> #k60 é o limite superior do interv. a posteriori 95% (k70;k80) para R
MTB> let k50=exp(-20.483/(2*k40))
MTB> let k60=exp(-3.247/(2*k40))
MTB> let c14(k15)=k50
MTB> let c15(k15)=k60
MTB>
MTB> #k70 é o limite inferior do interv. a posteriori 90% (k70;k80) para R
MTB> #k80 é o limite superior do interv. a posteriori 90% (k70;k80) para R
MTB> let k70=exp(-18.307/(2*k40))
MTB> let k80=exp(-3.940/(2*k40))
MTB> let c16(k15)=k70
MTB> let c17(k15)=k80
MTB> let k15=k15+1

```

PROGRAMA 3: a) Obtenção da porcentagem de intervalos a posteriori 95% e 90% para a função de confiabilidade R , para a qual o intervalo correspondente à densidade a priori de Jeffreys tem tamanho menor do que o correspondente a Zellner e,

b) obtenção da diferença média dos tamanhos dos intervalos correspondentes às densidades a priori de Jeffreys e Zellner.

```
MTB> #k1 é o tamanho da amostra  $n$ 
MTB> #k20 é o valor do tempo  $t$ 
MTB> #k21 é o valor do parâmetro  $a = p$ 
MTB> #k22 é o valor do parâmetro  $b = \theta^{1/r}$ 
MTB> #k15 é um contador
MTB> #k16 é a porcentagem de intervalos a posteriori 95% via densidade a priori
MTB> # de Jeffreys que são menores do que os intervalos via Zellner
MTB> #k17 é a porcentagem de intervalos a posteriori 90% via densidade a priori
MTB> # de Jeffreys que são menores do que os intervalos via Zellner
MTB> #k18 é a diferença média dos intervalos a posteriori 95% via
MTB> # densidade a priori de Jeffreys e Zellner
MTB> #k19 é a diferença média dos intervalos a posteriori 90% via
MTB> # densidade a priori de Jeffreys e Zellner
MTB> #Interv é um arquivo para cálculo dos intervalos a posteriori
MTB> # em cada simulação
MTB>
MTB> let k1=5
MTB> let k20=2
MTB> let k21=2
MTB> let k22=2
MTB> let k15=1
```

```

MTB> exec 'interv.' 100
MTB> let c18=(c10<c12)
MTB> let c19=(c11<c13)
MTB> let k16=sum(c18)/100
MTB> let k17=sum(c19)/100
MTB> print k16 k17
MTB> let c20=abs(c12-c10)
MTB> let c21=abs(c13-c11)
MTB> let k18=sum(c20)/100
MTB> let k19=sum(c21)/100
MTB> print k18 k19

```

File Interv

```

MTB> #gerando uma amostra aleatória de tamanho k1 da distribuição Weibull
MTB> random k1 c1;
SUBC> weibull k21 k22.
MTB> let c2=c1/k20
MTB>
MTB> #cálculo do intervalo a posteriori via densidade a priori de Jeffreys
MBT> let k2=k1/(k1+1)
MTB> let k3=k2*logc(k20)
MTB> let c3=c2**(1/k3)
MTB> let k4=sum(c3)
MBT>
MTB> #k5 é o limite inferior do interv. a posteriori 95% (k5;k6) para R
MTB> #k6 é o limite superior do interv. a posteriori 95% (k5;k6) para R
MTB> let k5=exp(-20.483/(2*k4))
MTB> let k6=exp(-3.247/(2*k4))

```

```

MTB> #c10(k15) é o tamanho do intervalo 95% gerado na k15ª simulação
MTB> let c10(k15)=abs(k6-k5)
MTB>
MTB> #k7 é o limite inferior do interv. a posteriori 90% (k7;k8) para R
MTB> #k8 é o limite superior do interv. a posteriori 90% (k7;k8) para R
MTB> let k7=exp(-18.307/(2*k4))
MTB> let k8=exp(-3.940/(2*k4))
MTB> #c11(k15) é o tamanho do intervalo 90% gerado na k15ª simulação
MTB> let c11(k15)=abs(k8-k7)
MTB>
MTB> #cálculo do intervalo a posteriori via densidade a priori de Zellner
MTB> let k30=loge(k20)
MTB> let c30=c2**(1/k30)
MTB> let k40=sum(c30)
MTB>
MTB> #k50 é o limite inferior do interv. a posteriori 95% (k50;k60) para R
MTB> #k60 é o limite superior do interv. a posteriori 95% (k50;k60) para R
MTB> let k50=exp(-20.483/(2*k40))
MTB> let k60=exp(-3.247/(2*k40))
MTB> #c12(k15) é o tamanho do intervalo 95% gerado na k15ª simulação
MTB> let c12(k15)=abs(k60-k50)
MTB>
MTB> #k70 é o limite inferior do interv. a posteriori 90% (k70;k80) para R
MTB> #k80 é o limite superior do interv. a posteriori 90% (k70;k80) para R
MTB> let k70=exp(-18.307/(2*k40))
MTB> let k80=exp(-3.940/(2*k40))
MTB> #c13(k15) é o tamanho do intervalo 90% gerado na k15ª simulação
MTB> let c13(k15)=abs(k80-k70)
MTB> let k15=k15+1

```

PROGRAMA 4: Obtenção do estimador de Máxima Verossimilhança (EMV) dos parâmetros p e θ , com a finalidade de se obter o EMV da função de confiabilidade R . Este programa determina uma solução numérica para a equação implícita.

$$\frac{1}{p} + \frac{\sum_{i=1}^n \ln x_i}{n} - \frac{\sum_{i=1}^n x_i^p \cdot \ln x_i}{\sum_{i=1}^n x_i^p} = 0 \quad (E1)$$

Nota: $\hat{\theta} = \frac{\sum_{i=1}^n x_i^p}{n}$ e $\hat{R} = \exp\left\{-\frac{t^p}{\theta}\right\}$ são os estimadores de máxima verossimilhança de θ e R respectivamente, onde o EMV $\hat{p}$ é solução da equação E1.

*****

Eureka: The solver, Version 1.0

Tuesday October 5, 1993, 2:12 pm.

Name of Input file: B:\ EMVER.EKA

*****

;dados gerados da weibull com parâmetros $p = 2$ e $\theta = 4$, $n = 5$ e $t = 2$

;C é a raiz da equação implícita, i.é, $C = \hat{p}$

;C1 = $\hat{\theta}$ e C2 = $\hat{R}$

k=ln(1.29665)+ln(3.83614)+ln(0.66075)+ln(2.34744)+ln(1.48015)

k1=k/5

A(x):=1.29665^x+3.83614^x+0.66075^x+2.34744^x+1.48015^x

B(x):=LN(1.29665)*1.29665^x+LN(3.83614)*3.83614^x+LN(0.66075)*0.66075^x+LN(2.34744)*2.34744^x+LN(1.48015)*1.48015^x

L(x):=(1/x)+k1-B(x)/A(x)

L(C)=0

C1=A(C)/5

C2=exp(2^C)/C1

\$ accuracy = 0.000000000000000001

*****

PROGRAMA 5: Obtenção da posteriori de R correspondente à densidade a priori de Jeffreys, dada abaixo, utilizando-se a integração numérica de Simpson, para efeito de comparação com a aproximação de Laplace.

$$p(R|\underline{x}) = c \frac{1}{R} \left(\ln \frac{1}{R} \right)^{n-1} \int_0^\infty \frac{1}{W'^{(n+1)} t^{n/W'}} \prod x_i^{1/W'} R^{\sum (\frac{x_i}{t})^{1/W'}} dW' \quad (E2)$$

onde

$$c = \frac{1}{\Gamma(n) \int_0^\infty \frac{W'^{-(n+1)} \prod x_i^{1/W'}}{\left(\sum x_i^{1/W'} \right)^n} dW'} \quad (E3)$$

*****

Eureka: The solver, Version 1.0

Tuesday October 5, 1993, 2:12 pm.

Name of Input file: B:\ POSTSIMP.EKA

*****

;dados gerados da weibull com parâmetros $p = 2$ e $\theta = 4$, $n = 5$ e $t = 2$

;C é a constante normalizadora da densidade a posteriori

;F é o valor da densidade a posteriori no ponto $R = 0.2$

;T(x) é o integrando dado em E3

;D(x) é o integrando dado em E2

;para limites de integração diferentes dos valores propostos as integrais

;assumem valores aproximadamente zero

K=1.29665*3.83614*0.66075*2.34744*1.48015

A(x):=1.29665^(1/x)+3.83614^(1/x)_

+0.66075^(1/x)+2.34744^(1/x)+1.48015^(1/x)

```
T(x):=(1/x^6)*(K^(1/x))/(A(x)^5)
```

```
Tintl=integ(T(x),x,0.01,5)
```

```
C=1/(24*Tintl)
```

```
K1=K/(2^5)
```

```
B(x):=A(x)/(2^(1/x))
```

```
D(x):=(1/x^6)*(K1^(1/x))*(0.2^(B(x)-1))
```

```
Dintl=integ(D(x),x,0.01,3)
```

```
F=C*((ln(1/0.2))^4)*Dintl
```

```
*****
```

Bibliografia

- ACHCAR, J.A.** (1990). Reparametrization and Accuracy of Laplace Approximations for Posterior Moments. *Revista Matemática e Estatística - UNESP*, v. 8, p. 23-30.
- AKAIKE, H.** (1978). A new look at Bayes procedures. *Biometrika*, 65, 53-59.
- BAYES, T.R.** (1763). Essay towards solving a problem in the doctrine of changes. Reprinted in *Biometrika*, 45 (1958), 243-315.
- BERGER, J. O.** (1992). Objective Bayesian Analysis: Development of the Reference Noninformative Priors. Course on Recent Advances on Bayesian Statistics and Econometrics, Caracas.
- BERGER, J. O. ; BERNARDO, J. M.** (1991). On the Development of the Reference Prior Method. Fourth Valencia International Meeting on Bayesian Statistics. Spain.
- BERGER, J.O. ; BERNARDO, J.M. ; MENDOZA, M.** (1989). On priors that maximize expected information. In *Recent Developments in Statistics and Their Applications*. J. Klein and J. Lee (eds), Freedom Acad. Publ., Seoul.
- BERNARDO, J. M.** (1979). Reference Posterior Distributions for Bayesian Inference. *Journal Royal Statistical Society*, v.41, n.2, p. 113-147.

- BOX, G.E.P. ; TIAO, G.C.** (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley, New York.
- COX, D.R. ; REID, N.** (1987). Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society, Ser. B, v. 49, p. 1-39.
- GEISSER, S.** (1984). On prior distributions for binary trials. American Statistician, 38 244-251.
- JAYNES, E.T.** (1968). Prior Probabilities. IEEE Transactions on Systems Science and Cybernetics, SSC-4, p. 227-241.
- JAYNES, E.T.** (1980). Discussion, in J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, editors, Bayesian Statistics, Valencia, Spain: University Press, 618-629.
- JEFFREYS, H.** (1967). Theory of Probability. 3rd rev. ed., London: Oxford U. Press.
- JOHNSON, R.A.** (1967). An Asymptotic Expansion for Posterior Distributions. Annals of Mathematical Statistic, v. 38, p. 1899.
- JOHNSON, R.A.** (1970). Asymptotic Expansion Associated with Posterior Distributions. Annals of Mathematical Statistic, v. 41, p. 851.
- KASS, R.E. ; TIERNEY, L. ; KADANE, J. B.** (1990). Laplace's Method in Bayesian Analysis. [Relatório Técnico].

- KAO, J.H.K.** (1959). A Graphical estimate of mixed Weibull parameters in life testing electron tubes. *Technometrics*, I, 389-407.
- LAPLACE, P.** (1774). Mémoire sur la probabilité des causes par les événements. *Mem. Acad. R. Sci. Présentés par Divers Savans*, 6, 621-656 (translated in *Statistical Science*, 1, 359-378).
- LAWLESS, J.F.** (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- LEIBLEIN, J. ; ZELLEN, M.** (1956). Statistical investigation of the fatigue life of deep groove ball bearings. *J. of Res. Nat'l-burstd.* 57, 273-315.
- LINDLEY, D.V.** (1956). On a Measure of the Information Provided by an Experiment. *Annals of Mathematical Statistics*, v. 27, p. 986-1005.
- MANN, N.** (1968). Results on statistical estimation and hypotheses testing with application to the Weibul and Extreme Value distribution. Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio.
- MARTZ, H.F. ; WALLER, R.A.** (1982). *Bayesian Reliability Analysis*. Wiley, New York.
- MASTRAN, D.V.** (1976). Incorporating Component and System Test Data into the Same Assessment: A Bayesian Approach. *Operations Research*, v.24, n.3, p. 491-499.
- MASTRAN, D.V. ; SINGPURWALLA, N. D.** (1978). A Bayesian Estimation

of the Reliability of Coherent Structures. Operations Research, v. 26, n. 4.

NELSON, W. B. (1972). Graphical analysis of accelerated life test data with the inverse power law model. IEEE Trans. Reliab, R21, 2-11.

NOVICK, M.R. ; HALL, W.J. (1965) A Bayesian Indifference Procedure. Journal of the American Statistical Association, v. 60, p. 1104-1117.

PETO, R. ; LEE, P (1973). Weibull distributions for continuous carcinogenesis experiments. Biometrics, 29, 457-470.âPIKE, M.C. (1966). A methods of analysis of a certain class of experiments in carcinogenesis. Biometrics, 22, 142-161.

PIKE, M.C. (1966). A methods of anlysis of a certain class of experiments in carcinogenesis. Biometrics, 22, 142-161.

POLSON, N. ; WASSERMAN, L. (1990) Prior Distributions for the Bivariate Binomial. Biometrika. 77, 4, pp. 901-4.

RAIFFA, H. ; SCHLAIFER, R. (1961). Applied Statistical Decision Theory. Harvard University Press, Cambridge, MA.

ROBBINS, H. (1964). The Empirical Bayes Approach to Statistical Decision Problems. The Annals of Mathematical Statistics. 35(1), 1-20.

RODRIGUES, J. ; LOUZADA NETTO, F. (1990). A note on the orthogonal reparametrization in Bayesian inference for the extreme value distribution. Notas do ICMSC-USP, no. 64, São Carlos.

RODRIGUEZ, C. C. (1991a). Entropic Priors. Ms., dept. of Mathematics and

Statistics, State U. of New York, Albany, NY.

RODRIGUEZ, C. C. (1991b). From Euclid to Entropy. In W. T. Grandy and L. H. Schick (1991), 343-348.

SAVAGE, L.J. (1961). The Subjective Basis of Statistical Practice. Department of Statistics, University of Michigan.

SHANNON, C.E. (1948). A Mathematical Theory of Communication. Bell System Tech. J., v. 27, p. 379-423 e 623-656.

SINHA, S.K. ; GUTTMAN, I. (1988). Bayesian Analysis of Life-Testing Problems Involving the Weibull Distribution. Communications in Statistics-Theory and Methods, v.17 , n.2, p. 343-356.

STEIN, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In Sequential Methods in Statistics, Banach center publications, 16. Warsaw: PWN-Polish scientific publisher.

TIBSHIRANI, R. (1989). Noninformative Priors for One Parameters of Many. Biometrika, v. 76, p. 604-608.

TIERNEY, L. ; KADANE, J.B. (1986). Accurate Approximations for the Posterior Moments and Marginal Densities. Journal of the American Statistical Association, v. 81, n. 393, p. 82-6.

VILLEGAS, C. (1977). On the representation of ignorance. J.A.SA., 72, 651-654.

ZELLNER, A. (1971). An Introduction to Bayesian Inference in Econometrics.

Wiley, New York.

ZELLNER, A. (1977). Maximal Data Information Prior Distributions, In New Methods in the applications of Bayesian Methods, A. Aykac e C. Brumat (eds). North-Holland, Amsterdam.

ZELLNER, A. (1984). Maximal Data Information Prior Distributions, Basic Issues in Econometrics, U. of Chicago Press.

ZELLNER, A. (1990). Bayesian Methods and Entropy in Economics and Econometrics. In W.T. Grandy, Jr. and L.H. Schick (eds.), Maximum Entropy and Bayesian Methods, Dordrecht, Netherlands: Kluwer Academic Publishers, p.17-31.

ZELLNER, A. (1993). Models, Prior Information and Bayesian Analysis. [Relatório Técnico].

ZELLNER, A. ; MIN, C.K. (1992). Bayesian Analysis, Model Selection and Prediction. [Relatório Técnico].

ZELLNER, A. ; SINHA, S.K. (1990). A Note on the Prior Distributions of Weibull Parameters, SCIMA vol.19, nos.1-2.

WELCH, B.L. ; PEERS, H.W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. J.R.S.S, B 25, 318-29.

WHITTEMORE, A. ; ALTSCHULER, B. (1976). Lung cancer incidence in cigarette smokers: further analysis of Dall and Hill's data for British Physicians. Biometrics, 32, 805-816.