

UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia de Presidente Prudente
Programa de Pós-Graduação em Matemática Aplicada e Computacional

Análise Bayesiana para a distribuição Exponencial-Logarítmica

Lívia Matos Garcia
Orientador: Prof. Dr. Fernando Antonio Moala

Presidente Prudente, outubro de 2013

UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia de Presidente Prudente
Programa de Pós-Graduação em Matemática Aplicada e Computacional

Análise Bayesiana para a distribuição Exponencial-Logarítmica

Lívia Matos Garcia

Orientador: Prof. Dr. Fernando Antonio Moala

Dissertação apresentada ao Programa de Pós-Graduação da UNESP para obtenção do título de Mestre em Matemática Aplicada e Computacional.

Presidente Prudente, outubro de 2013

*Aos meus pais,
Vicente e Maria Lúcia*

Agradecimentos

Nada seria sem Aquele que formou os céus e a Terra, e por sua infinita bondade me concedeu o dom da vida. Agradeço a Deus por ter me dado sabedoria e me ajudado a passar pelos obstáculos que surgiram durante a caminhada. Agradeço a Ele por sua infinita misericórdia quando me ouviu clamando e pedindo auxílio. É certo que palavras são poucas para expressar minha gratidão, mas jamais poderia deixar de agradecer primeiramente a Deus por ter alcançado esta grande vitória.

Aos meus pais, obrigada pelo incentivo e pelo sacrifício para que eu chegasse até aqui. Vocês sempre me mostraram o caminho correto a seguir e com muita dedicação me fizeram caminhar para o sucesso.

Laan Felipe, te agradeço pelo apoio inconstante, mesmo longe. Você foi o meu ombro amigo nos momentos difíceis e suas palavras me fizeram seguir em frente. Obrigada.

Agradeço também ao meu professor Fernando pela paciência em me ensinar e por todo tempo que foi dedicado a mim. Este trabalho jamais teria chegado ao fim sem a sua tão grande e importante contribuição.

Aos meus colegas de mestrado, obrigada pela amizade, pela ajuda, e pelo bom-humor que acompanha todos da nossa "salinha da Pós". Agradeço especialmente os meus colegas Fábio (Expo) e Clóvis, pelas inúmeras ajudas que recebi de vocês. Muito obrigada.

Resumo

A Exponencial-Logarítmica ($EL(p, \beta)$) é uma distribuição para modelos de sobrevivência com taxa de falha decrescente. Esta distribuição pode ser usada para estudar os tempos de vida de organismos, materiais, dispositivos, etc, em ciências biológicas e na engenharia. Neste trabalho foram utilizadas as abordagens Clássica e Bayesiana para inferir sobre os parâmetros do modelo com conjunto de dados completos e censurados.

A abordagem Bayesiana requer a seleção de distribuições a priori para os parâmetros do modelo. No caso onde há informação dos dados, foram escolhidas distribuições a priori não-informativas. Por outro lado, quando há poucos dados e/ou são dados censurados, torna-se necessária uma priori informativa obtida a partir das informações de um especialista. Neste trabalho propôs-se uma priori informativa elicitada através de informações de especialistas e derivada da aproximação de Laplace.

Portanto, a análise Bayesiana foi realizada considerando os dois tipos de distribuição a priori: informativa e não-informativa. As distribuições a priori não-informativas usadas foram: priori de Jeffreys (Jeffreys (1967)), priori de Referência (Berger and Bernardo (1992)), priori de Máxima Informação dos Dados (Zellner (1977)) e priori derivada da função Cópula (Achcar et al. (2010)). Estas distribuições a priori também foram comparadas com outras distribuições a priori comuns, tais como Beta, Gama e Uniforme.

A fim de avaliar o desempenho das distribuições a priori, foi apresentado um estudo comparativo utilizando dados simulados a partir da distribuição $EL(p, \beta)$ e um conjunto de dados reais introduzido por Lawless (1982).

Utilizou-se o algoritmo MCMC para obter uma amostra de valores da posteriori conjunta, a fim de extrair características das distribuições posteriores marginais, tais como médias a posteriori, moda e intervalos de credibilidade.

Abstract

The Exponential-Logarithmic, denoted by $EL(p, \beta)$, is a lifetime distribution with decreasing failure rate. This distribution can be used to study the lengths of organisms, devices, materials, etc., in the biological and engineering sciences. In this dissertation we use classical and Bayesian approaches to make inferences for the parameters of the model under complete and censored data set.

Bayesian approach requires the selection of prior distributions for all parameters of the model. In this case, we will seek to choose a noninformative prior that provides best estimation when there is absence of information or with large data set and uncensored data. On the other hand, when there is few data to use or presence of censored data, an informative prior obtained from the expert's information is necessary. We propose an informative prior elicited from the expert's opinion and derived through Laplace's approximation.

Thus, we carry out the Bayesian estimation by considering the two types of prior distributions. Different noninformative prior distributions are used as Jeffreys (Jeffreys (1967)), Reference (Berger and Bernardo (1992)), maximal data information prior (Zellner (1977)) and prior derived from copula function (Achcar et al. (2010)). These priors are also compared with other common priors such as beta, gamma, and uniform distributions.

A comparative study to evaluate the performance of the prior distributions through simulated data from the $EL(p, \beta)$ distribution and a practical data set introduced by Lawless (1982) presented.

We also need to appeal to the MCMC algorithm to obtain a sample of values of and from the joint posterior in order to extract characteristics of marginal posterior distributions such as Bayes estimator, mode and credible intervals.

Sumário

Lista de Figuras	iii
Lista de Tabelas	iv
1 Introdução	1
2 Distribuição Exponencial-Logarítmica	4
2.1 Função Densidade de Probabilidade	4
2.1.1 Propriedades	5
2.1.2 Funções de sobrevivência e de risco	8
2.1.3 Geração de Amostras Aleatórias	10
3 Estimação de máxima verossimilhança para p e β	11
4 Análise Bayesiana	15
4.1 Priori de Jeffreys	15
4.2 Produto de distribuições a priori independentes	16
4.3 MCMC para a $EL(p, \beta)$	17
4.4 Análise de dados simulados	17
4.5 Análise com Conjunto de Dados Reais	23
5 Distribuições a priori MDIP & Gama e de Referência	26
5.1 Priori MDIP & Gama	26
5.2 Priori de Referência	28
5.3 Análise utilizando dados simulados	31
5.4 Análise com Conjunto de Dados Reais	34
6 Uso de Cópulas como priori	37
6.1 Função Cópula	37
6.2 Função Cópula Farlie-Gumbel-Morgenstern	38
6.3 Priori Cópula	39
6.4 Análise Utilizando Dados Simulados	40
6.5 Análise com Conjunto de Dados Reais	42

7	Priori Informativa: Elicitação	44
7.1	Introdução à Elicitação	44
7.2	Elicitação da confiabilidade	45
7.3	Análise utilizando Elicitação	49
8	Análise de Dados Censurados	51
8.1	Introdução	51
8.2	Função de Verossimilhança para Dados Censurados	53
8.3	Amostra simulada	54
8.4	Máxima Verossimilhança utilizando dados censurados	54
8.5	Análise de Dados Censurados para a $EL(p, \beta)$	55
9	Conclusões	57
10	Propostas Futuras	59
	Referências Bibliográficas	60
A	Apêndice	62
A.1	Demonstrações	62
A.2	Aproximação de Laplace	64
A.3	Critério de Geweke	65

Lista de Figuras

2.1	Densidades da Exponencial Logarítmica para $p = 0.5$ e $\beta = 1, 2$ e 3	6
2.2	Densidades da Exponencial Logarítmica para $\beta = 2$ e $p = 0.2, 0.5$ e 0.8 . .	6
2.3	Função de Risco da distribuição Exponencial Logarítmica para $\beta = 2$ e $p = 0.2, 0.5$ e 0.8	9
4.1	Output do MCMC e plots de autocorrelação para priori de Jeffreys consi- derando $n = 20$	18
4.2	Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 20$	19
4.3	Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 50$	20
4.4	Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 70$	21
4.5	Verossimilhança, densidades a priori e a posteriori do parâmetro p para $n = 10, 20$ e 50	22
4.6	Verossimilhança, densidades a priori e a posteriori do parâmetro β para $n = 10, 20$ e 50	22
4.7	Output do MCMC e plots de autocorrelação utilizando a priori de Jeffreys	23
4.8	Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)	25
5.1	Gráficos das marginais a posteriori para $n = 20$	33
5.2	Gráficos das marginais a posteriori para $n = 50$	33
5.3	Gráficos das marginais a posteriori para $n = 70$	34
5.4	Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)	36
6.1	Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 50$	41
6.2	Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)	43

Lista de Tabelas

4.1	EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$ e $\beta = 2$.	19
4.2	Intervalos bayesianos com 95% de confiança para as marginais das distribuições a posteriori em relação a p e β .	20
4.3	Resultados obtidos da posteriori para o parâmetro p .	23
4.4	Resultados obtidos da posteriori para o parâmetro β .	24
4.5	Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.	24
5.1	EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.	32
5.2	EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.	32
5.3	Resultados obtidos da posteriori para o parâmetro p (dados de Lawless).	35
5.4	Resultados obtidos da posteriori para o parâmetro β (dados de Lawless).	35
5.5	Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.	35
6.1	EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.	40
6.2	EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.	40
6.3	Estimativas (média) e desvio-padrão da correlação obtida através da priori Cópula.	41
6.4	Resultados obtidos da posteriori para o parâmetro p (dados de Lawless).	42
6.5	Resultados obtidos da posteriori para o parâmetro β (dados de Lawless).	42
6.6	Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.	43
7.1	EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.	50
7.2	EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.	50
8.1	Estimativas (média) Bayesianas e desvio padrão para $p = 0.5$, $\beta = 2$ e 10% dos dados censurados.	56
8.2	Estimativas (média) Bayesianas e desvio padrão para $p = 0.5$, $\beta = 2$ e 20% dos dados censurados.	56

Introdução

A distribuição Exponencial-Logarítmica proposta por Tahmasbi and Rezaei (2008) consiste em uma distribuição com dois parâmetros p e β definida na análise de sobrevivência com taxa de falha decrescente. Em geral, espera-se que, com o passar do tempo, o comportamento de componentes (ou organismos) apresente taxa de falha decrescente quando este comportamento é caracterizado por "work-hardening"(em termos de engenharia) ou "imunidade"(em termos biológicos). Além disso, essa distribuição pode ser usada para estudar os tempos de vida de organismos, dispositivos, materiais, etc, em ciências biológicas e na engenharia.

A densidade da distribuição Exponencial-Logarítmica, denotada por $EL(p, \beta)$, é dada por:

$$f(x|p, \beta) = \left(-\frac{1}{\ln(p)} \right) \frac{\beta(1-p)e^{-\beta x}}{1 - (1-p)e^{-\beta x}} \quad (1.1)$$

com $x > 0$, $p \in (0, 1)$ sendo o parâmetro de forma e $\beta > 0$ o parâmetro de escala.

A $EL(p, \beta)$ pode ser uma boa alternativa para modelar taxas de falha devido à forma simples das suas funções de sobrevivência e risco, em comparação com outras distribuições com taxa de falha decrescente, como a Gama e a Weibull. Tahmasbi and Rezaei (2008) também observam que os riscos iniciais e a longo prazo para essa distribuição são ambos finitos em contraste com aqueles da distribuição Weibull, ou seja, para a $EL(p, \beta)$ tem-se riscos iguais a: $h(0) = \frac{-\beta(1-p)}{p \ln p}$ e $h(\infty) = \beta$, enquanto que para a distribuição Weibull os riscos são dados por: $h(0) = \infty$ e $h(\infty) = 0$.

É importante citar que, para que exista o Estimador de Máxima Verossimilhança (EMV) para a distribuição $EL(p, \beta)$, a seguinte condição deve estar satisfeita: $n/2 < \sum_{i=1}^n e^{-\beta x_i}$. Observa-se que para número de amostras pequeno, ou seja, para $n < 20$, essa condição não é satisfeita e, portanto, o EMV não existe. Com a análise Bayesiana,

mostrou-se que esse problema é eliminado, pois não há restrição alguma nem mesmo para amostras pequenas.

Várias propriedades da distribuição $EL(p, \beta)$, tais como: tempo de falha médio, geradores de momentos, moda e variância são apresentados em Tahmasbi and Rezaei (2008). Eles mostram também a estimação clássica dos parâmetros, suas variâncias e covariâncias assintóticas. Entretanto, não há trabalho algum proposto na literatura que mostre a estimação de seus parâmetros sob um enfoque Bayesiano, mesmo para dados sem censuras.

Na prática, os dados relacionados com o tempo de falha de um componente eletrônico são geralmente censurados ou sujeitos a "stress", por exemplo, alta temperatura, alta voltagem, etc. Os dados são "censurados" quando existem elementos na amostra para os quais somente o tempo máximo e mínimo são conhecidos. A censura ou a aceleração dos dados são comuns na prática devido ao longo tempo de espera até que ocorra uma falha. Quanto maior sofisticação tecnológica na construção do componente, maior será a necessidade de censura do seu tempo de vida. Com dados censurados é importante desenvolver técnicas para prever o tempo real de falha do componente censurado ou prever o tempo de vida dos componentes que não estão na amostra (observações futuras). Dentre os objetivos deste trabalho, pretende-se utilizar dados não-censurados e censurados com diferentes distribuições a priori para análise da distribuição $EL(p, \beta)$.

Os principais objetivos desse trabalho foram:

- Desenvolver técnicas Bayesianas com distribuições a priori informativas e não-informativas para estimar os parâmetros da distribuição Exponencial - Logarítmica.
- Realizar um estudo do efeito das distribuições a priori nas distribuições a posteriori marginalizadas nos dois parâmetros, p e β .
- Utilizar os métodos Monte Carlo Markov Chain (MCMC) e amostrador de Gibbs para gerar as distribuições marginais a posteriori dos parâmetros, como também a função de confiabilidade, e para prever o comportamento de observações futuras da mesma distribuição.
- Além dos dados simulados, mostrar aplicações a dados reais para efeito de enriquecimento do trabalho e consolidação da importância da distribuição.
- Comparar as distribuições a priori não-informativas Jeffreys (Jeffreys (1967)), de Referência (Bernardo (1991)), MDIP (Zellner (1977)) e Cópula (Achcar et al. (2010)).
- Realizar a elicitación dos hiperparâmetros das distribuições a priori informativas.
- Analisar a distribuição $EL(p, \beta)$ sob a presença de dados censurados.

O trabalho foi dividido da seguinte forma:

- No Capítulo 2 foi realizado um estudo mais aprofundado das principais propriedades da distribuição Exponencial- Logarítmica, tais como: funções de sobrevivência e de

risco, esperança, variância, geradora de momentos, além da estimação clássica dos parâmetros através do método da máxima verossimilhança.

- O Capítulo 3 mostra a análise estatística, sob o enfoque clássico, para a distribuição Exponencial-Logarítmica.
- O Capítulo 4 dá início à análise Bayesiana para a distribuição $EL(p, \beta)$. Neste capítulo foram utilizadas distribuições a priori não - informativas, incluindo a priori de Jeffreys, para estimar os parâmetros p e β . Este capítulo foi baseado no artigo Moala and Garcia (2013).
- No Capítulo 5 foram exploradas as distribuições a priori MDIP(Zellner (1977)) e de Referência (Bernardo (1991)). Elas foram comparadas às distribuições a priori não-informativas analisadas no Capítulo 5.
- No Capítulo 6 foi feita a análise Bayesiana para a distribuição $EL(p, \beta)$ utilizando a priori Cópula(Achcar et al. (2010)). Foram explicitados detalhes da obtenção da priori e realizada uma comparação com as demais distribuições a priori estudadas.
- No Capítulo 7 foi detalhado o método de elicitación dos hiperparâmetros das distribuições a priori Beta e Gama.
- No Capítulo 8 foi apresentada a análise Bayesiana para a distribuição Exponencial-Logarítmica utilizando dados censurados.
- Por fim, no Capítulo 9 foram feitas conclusões e considerações baseadas neste trabalho.

Distribuição Exponencial-Logarítmica

2.1 Função Densidade de Probabilidade

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória gerada da distribuição Exponencial cuja função densidade de probabilidade é dada por:

$$f_Y(y; \beta) = \beta e^{-\beta y}, \quad y > 0 \quad \text{e} \quad \beta > 0. \quad (2.1)$$

Suponha agora Z uma amostra aleatória gerada da distribuição Logarítmica, com função massa de probabilidade igual a:

$$P_Z(z; p) = \frac{(1-p)^z}{-z \ln p}, \quad z = 1, 2, \dots \quad \text{e} \quad 0 < p < 1. \quad (2.2)$$

Assumindo que as variáveis aleatórias Y e Z são independentes, é possível definir a variável aleatória $X = \min\{Y_1, \dots, Y_z\}$. Então, $f_{X|Z}(x|z; \beta) = \beta z e^{-\beta x z}$ e a marginal de X é dada por:

$$f_X(x; p, \beta) = \left(-\frac{1}{\ln(p)} \right) \left(\frac{\beta(1-p)e^{-\beta x}}{1 - (1-p)e^{-\beta x}} \right) \quad x \geq 0. \quad (2.3)$$

onde p é o parâmetro de forma e β é o de escala.

A construção da distribuição é dada da seguinte forma: Seja $Y|\beta \sim \text{Exp}(\beta)$ e $Z|p \sim \text{Logarítmica}(p)$. Então,

$$f_Y(y; \beta) = \beta e^{-\beta y} \quad \text{e} \quad F_Y(y; \beta) = 1 - e^{-\beta y} \quad \text{para} \quad y > 0 \quad \text{e} \quad \beta > 0$$

$$\text{e} \quad P_Z(z; p) = \frac{(1-p)^z}{-z \ln p}, \quad z = 1, 2, \dots \quad \text{e} \quad p \in (0, 1).$$

Seja $X = \min\{Y_1, Y_2, \dots, Y_z\}$. Então X tem fdp dada por:

$$f_X(x) = z[1 - F_Y(y)]^{z-1} f_Y(y).$$

Para maiores detalhes veja Mood et al. (1974).

Logo, $f_{X|Z}(x) = z(e^{-\beta x})^{z-1} \beta e^{-\beta x}$, que implica em $f_{X|Z}(x) = \beta z e^{-\beta z x}$, com $\beta > 0$ e $z = 1, 2, \dots$

Utilizando o último resultado encontra-se a função conjunta $f(x, z)$, dada por:

$$f(x, z) = f_Z(z) f_{X|Z}(x).$$

Assim,

$$f(x, z) = \frac{(1-p)^z}{-z \ln p} \beta z e^{-\beta z x}.$$

Por fim, a marginal de X é obtida como:

$$f_X(x) = \sum_{z=1}^{\infty} f(x, z) = \sum_{z=1}^{\infty} \frac{(1-p)^z}{-z \ln p} \beta z e^{-\beta z x} = \frac{-\beta}{\ln p} \sum_{z=1}^{\infty} \left(\frac{1-p}{e^{\beta x}} \right)^z. \quad (2.4)$$

Utilizando um resultado de série infinita, tem-se: $\sum_{x=1}^{\infty} a^x = \frac{a}{1-a}$ para $0 < a < 1$, e tomando $a = \frac{1-p}{e^{\beta x}}$, obtém-se:

$$\sum_{z=1}^{\infty} \left(\frac{1-p}{e^{\beta x}} \right)^z = \frac{\frac{1-p}{e^{\beta x}}}{1 - \left(\frac{1-p}{e^{\beta x}} \right)} = \frac{(1-p)e^{-\beta x}}{1 - (1-p)e^{-\beta x}}. \quad (2.5)$$

Substituindo (2.5) em (2.4), chega-se ao resultado final:

$$f_X(x) = \left(-\frac{1}{\ln p} \right) \left(\frac{\beta(1-p)e^{-\beta x}}{1 - (1-p)e^{-\beta x}} \right), \quad x > 0, \quad 0 < p < 1 \quad e \quad \beta > 0.$$

Outra forma de obter a distribuição Exponencial-Logarítmica é dada no Apêndice A1 deste trabalho.

Observe que se $\lim_{p \rightarrow 1} p = 1$, ela se reduz à distribuição Exponencial com parâmetro β .

Os gráficos da função densidade de probabilidade da $EL(p, \beta)$ são apresentados nas figuras (2.1) e (2.2) a seguir, considerando diferentes valores dos parâmetros.

2.1.1 Propriedades

Tahmasbi and Rezaei (2008) mostram diversas propriedades da distribuição Exponencial-Logarítmica. A seguir são explicitadas algumas delas.

P1: Função distribuição acumulada da $EL(p, \beta)$:

$$F_X(x; p, \beta) = 1 - \frac{\ln(1 - (1-p)e^{-\beta x})}{\ln p}. \quad (2.6)$$

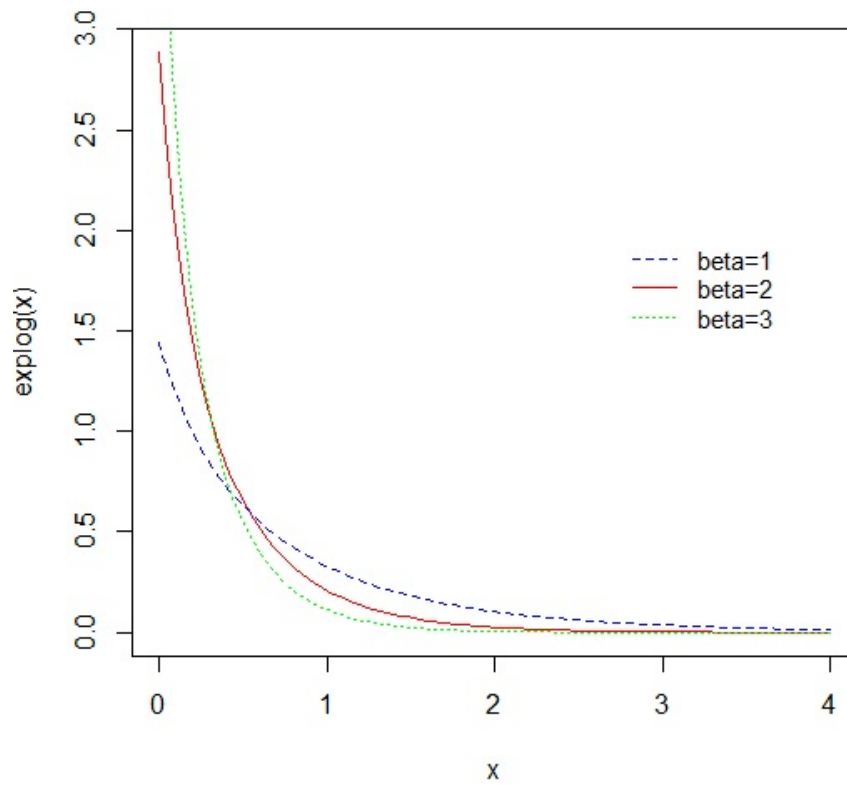


Figura 2.1: Densidades da Exponencial Logarítmica para $p = 0.5$ e $\beta = 1, 2$ e 3

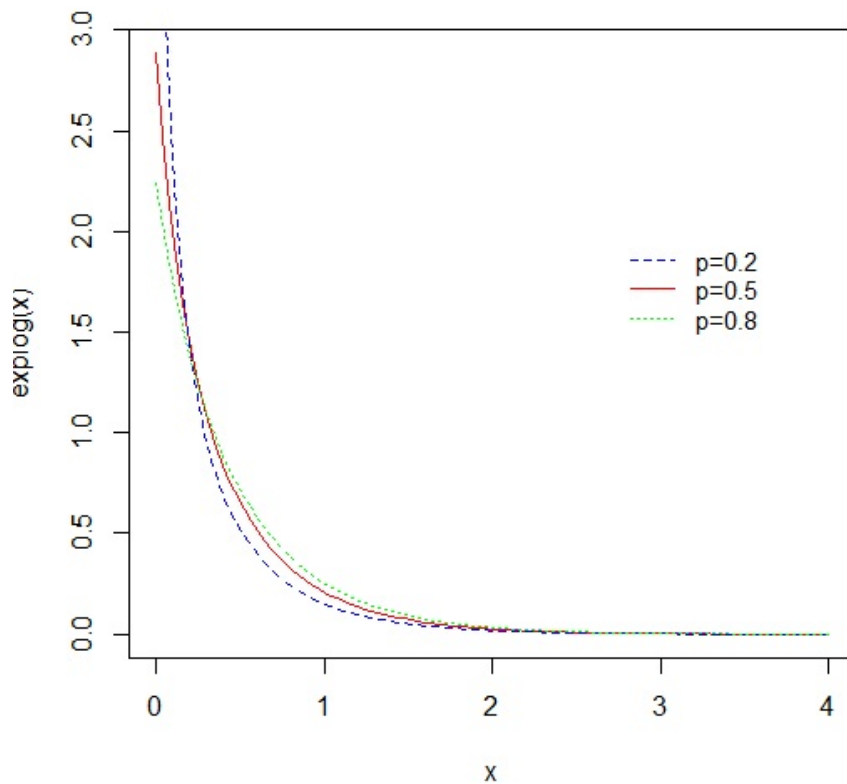


Figura 2.2: Densidades da Exponencial Logarítmica para $\beta = 2$ e $p = 0.2, 0.5$ e 0.8

Seja $f(x|p, \beta)$ a função densidade de probabilidade da $EL(p, \beta)$ dada em (2.3). Da definição de função distribuição acumulada, tem-se:

$$F_X(x; p, \beta) = \int_0^y -\frac{1}{\ln p} \frac{\beta(1-p)e^{-\beta t}}{1-(1-p)e^{-\beta t}} dx = \frac{-\beta(1-p)}{\ln p} \int_0^y \frac{e^{-\beta x}}{1-(1-p)e^{-\beta x}} dx. \quad (2.7)$$

Utilizando o método de substituição, é possível tomar $u = 1 - (1-p)e^{-\beta x}$ e sua derivada $du = (1-p)\beta e^{-\beta x} dx$, ou seja, $e^{-\beta x} dx = \frac{du}{(1-p)\beta}$. Aplicando na equação (2.7), tem-se:

$$\begin{aligned} F_X(x; p, \beta) &= -\left(\frac{\beta(1-p)}{\ln p}\right) \left(\frac{1}{(1-p)\beta}\right) \int_p^{1-(1-p)e^{-\beta y}} \frac{1}{u} du = -\frac{1}{\ln p} \ln u \Big|_p^{1-(1-p)e^{-\beta y}} \\ &= -\frac{1}{\ln p} [\ln(1-(1-p)e^{-\beta y}) - \ln p] = 1 - \frac{\ln(1-(1-p)e^{-\beta y})}{\ln p}, \end{aligned}$$

provando-se, portanto, a função distribuição acumulada.

P2: Mediana da EL(p, β):

$$m_d = \frac{\ln(1+\sqrt{p})}{\beta}. \quad (2.8)$$

Para obtenção da mediana tem-se:

$$F_X(m_d; p, \beta) = 1 - \frac{\ln(1-(1-p)e^{-\beta m_d})}{\ln p} = 0.5, \quad \text{isto é,} \quad 0.5 \ln p = \ln(1-(1-p)e^{-\beta m_d}).$$

Aplicando a função exponencial em ambos os lados chega-se à expressão

$$e^{\ln p^{0.5}} = 1 - (1-p)e^{-\beta m_d}, \quad \text{ou seja,} \quad \sqrt{p} = 1 - (1-p)e^{-\beta m_d} \quad \text{ou} \quad e^{-\beta m_d} = \frac{1-\sqrt{p}}{(1-p)}.$$

Ainda, aplicando a função logarítmica obtém-se

$$-\beta m_d = \ln\left(\frac{1-\sqrt{p}}{1-p}\right) \quad \text{que pode ser reescrito como} \quad -\beta m_d = \ln\left(\frac{1-\sqrt{p}}{(1-\sqrt{p})(1+\sqrt{p})}\right),$$

e implica no resultado final: $m_d = \frac{\ln(1+\sqrt{p})}{\beta}$.

P3: A função geradora de momentos é determinada através de (2.3) e dada pela seguinte expressão:

$$M_X(t) = E(e^{tX}) = -\frac{\beta(1-p)}{\ln p(\beta-t)} \text{hipergeo}_{2,1} \left(\left[1, \frac{\beta-t}{\beta}\right], \left[\frac{2\beta-t}{\beta}\right], 1-p \right), \quad (2.9)$$

onde $\text{hipergeo}(\cdot)$ é a função hipergeométrica. Esse resultado pode ser demonstrado

utilizando-se conhecimentos sobre aproximações de integrais por séries infinitas. A função hipergeométrica também é conhecida como função hipergeométrica generalizada de Barnes's e é dada por:

$$F_{p,q}(n, d, \lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k \prod_{i=1}^p \Gamma(n_i + k) \Gamma^{-1}(n_i)}{\Gamma(k+1) \prod_{i=1}^q \Gamma(d_i + k) \Gamma^{-1}(d_i)} \quad (2.10)$$

onde $n = [n_1, n_2, \dots, n_p]$, p é o número de coordenadas de n , $d = [d_1, d_2, \dots, d_q]$ e q é o número de coordenadas de d . Além disso, a função hipergeométrica generalizada é dada em vários softwares matemáticos, como o Maple.

P4: Os momentos de X são determinados derivando $M_X(t)$ e dados por:

$$E(X^r | p, \beta) = \frac{r! \text{polylog}(r+1, 1-p)}{\beta^r \ln p}, \quad r \in N, \quad (2.11)$$

onde $\text{polylog}(\cdot)$ é a função polylogarítmica definida como:

$$\text{polylog}(a, z) = \sum_{k=1}^{\infty} \frac{z^k}{k^a}. \quad (2.12)$$

P5 A média e a variância da distribuição $EL(p, \beta)$ utilizando (2.11) são dadas, respectivamente, por

$$E(X) = -\frac{\text{polylog}(2, 1-p)}{\beta \ln p} \quad e \quad (2.13)$$

$$\text{Var}(X) = -\frac{2\text{polylog}(3, 1-p)}{\beta^2 \ln p} - \frac{\text{polylog}^2(2, 1-p)}{\beta^2 \ln^2 p}. \quad (2.14)$$

Outra forma de demonstrar a Esperança é dada no apêndice deste trabalho.

2.1.2 Funções de sobrevivência e de risco

Utilizando (2.3) e (2.6), as funções de sobrevivência e de risco da $EL(p, \beta)$ podem ser encontradas. Elas são dadas, respectivamente, por:

$$S(x) = \frac{\ln(1 - (1-p)e^{-\beta x})}{\ln p} \quad (2.15)$$

e

$$h(x) = \frac{-\beta(1-p)e^{-\beta x}}{(1 - (1-p)e^{-\beta x}) \ln(1 - (1-p)e^{-\beta x})}. \quad (2.16)$$

A função de sobrevivência $S(x)$ pode ser facilmente encontrada utilizando a equação (2.6) e sua definição: $S(x) = 1 - F_X(x)$. Portanto:

$$S(x) = 1 - \left(1 - \frac{\ln(1 - (1 - p)e^{-\beta x})}{\ln p} \right) = \frac{\ln(1 - (1 - p)e^{-\beta x})}{\ln p}.$$

Da definição de função de risco dada por:

$$h(x) = \frac{f(x)}{S(x)},$$

e utilizando as equações (2.3) e (2.15) tem-se:

$$h(x) = \frac{\frac{1}{\ln(p)} \cdot \frac{\beta(1-p)e^{-\beta x}}{1 - (1-p)e^{-\beta x}}}{\frac{\ln(1 - (1-p)e^{-\beta x})}{\ln(p)}} = \frac{-\beta(1-p)e^{-\beta x}}{(1 - (1-p)e^{-\beta x}) \ln(1 - (1-p)e^{-\beta x})}.$$

Essas duas funções têm formas mais simples que as funções de outras distribuições que possuem taxa de risco decrescente, como a distribuição Gama.

A função de risco é exibida na figura (2.3) a seguir para diferentes valores dos parâmetros.

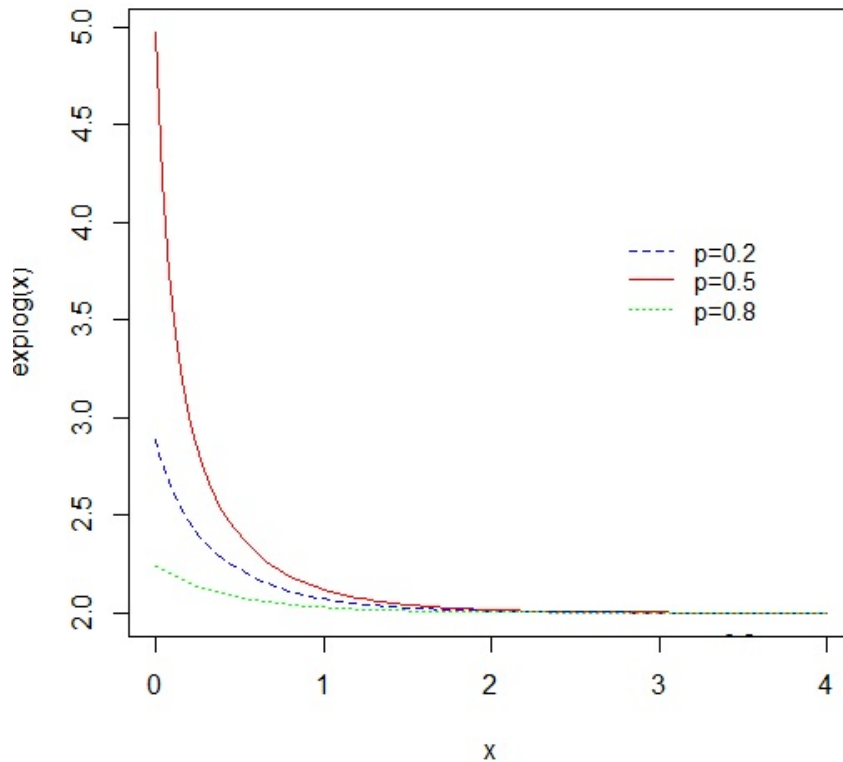


Figura 2.3: Função de Risco da distribuição Exponencial Logarítmica para $\beta = 2$ e $p = 0.2, 0.5$ e 0.8

Os riscos iniciais e a longo prazo são dados, respectivamente, por: $h(0) = \frac{-\beta(1-p)}{p \ln p}$ e $h(\infty) = \beta$. Eles são ambos finitos em contraste com os da distribuição Weibull, que são

iguais a: $h(0) = \infty$ e $h(\infty) = 0$.

2.1.3 Geração de Amostras Aleatórias

Para gerar a distribuição $EL(p, \beta)$, foi utilizado um método simples baseado na amostragem da transformação inversa. Seja U uma amostra aleatória da distribuição uniforme no intervalo $(0,1)$. Então a variável:

$$X = \frac{1}{\beta} \ln \left(\frac{1-p}{1-p^{1-U}} \right).$$

tem distribuição EL com parâmetros p e β .

O método da transformação inversa se baseia no fato da função distribuição acumulada estar dentro do intervalo $(0,1)$, ou seja, no mesmo intervalo de um número aleatório, gerado por um gerador básico. Para encontrar X , basta encontrar a inversa de $F_X(x|p, \beta)$. Assim, utilizando a equação (2.6) e tomando U como uma amostra aleatória da distribuição $U(0, 1)$, pode-se encontrar:

$$F_X(x|p, \beta) = U, \quad \text{ou seja,} \quad 1 - \frac{\ln(1 - (1-p)e^{-\beta X})}{\ln p} = U.$$

$$\text{Portanto,} \quad \ln(1 - (1-p)e^{-\beta X}) = (1-U) \ln p \quad \text{ou} \quad 1 - (1-p)e^{-\beta X} = p^{1-U}.$$

$$\text{Assim,} \quad 1 - p^{-U+1} = (1-p)e^{-\beta X}, \quad \text{ou seja,} \quad \frac{1-p^{-U+1}}{1-p} = e^{-\beta X}.$$

$$\text{Aplicando logaritmo tem-se:} \quad \ln \left(\frac{1-p^{-U+1}}{1-p} \right) = -\beta X.$$

Dessa maneira chega-se à fórmula para geração:

$$X = \frac{1}{\beta} \ln \left(\frac{1-p}{1-p^{1-U}} \right).$$

Estimação de máxima verossimilhança para p e β

Suponha uma amostra aleatória X_1, X_2, \dots, X_n de tamanho n da distribuição $EL(p, \beta)$, onde os parâmetros p e β são desconhecidos. A fim de solucionar esse problema, é preciso encontrar os Estimadores de Máxima Verossimilhança (EMV) para esses parâmetros.

Os Estimadores de Máxima Verossimilhança para os parâmetros p e β são valores que maximizam a função de verossimilhança dada por

$$L(p, \beta|x) \propto \frac{1}{(\ln p)^n} \frac{\beta^n (1-p)^n e^{-\beta \sum_{i=1}^n x_i}}{\prod_{i=1}^n (1 - (1-p)e^{-\beta x_i})}. \quad (3.1)$$

Para determinar os EMV para p e β , seja o logaritmo da função de verossimilhança baseado na amostra observada X :

$$\begin{aligned} l(p, \beta; x_{obs}) &= \log L(p, \beta|x) = \\ &= -n \ln(-\ln(p)) + n \ln(1-p) + n \ln(\beta) - \beta \sum_{i=1}^n x_i - \sum_{i=1}^n \ln[1 - (1-p)e^{-\beta x_i}]. \end{aligned} \quad (3.2)$$

Então, os estimadores de máxima verossimilhança (EMV) para os parâmetros p e β são obtidos como solução das derivadas de $l(p, \beta; x_{obs})$ dadas por:

$$\frac{\partial l}{\partial p} = -\frac{n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}} = 0, \quad (3.3)$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}} = 0. \quad (3.4)$$

Para encontrar as estimativas através do EMV, Tahmasbi and Rezaei (2008) determinam as condições para a existência e singularidade do EMV para p e β quando o outro parâmetro é dado ou conhecido. Essas condições são expressas a seguir, nos Teoremas (1) e (2).

Teorema 1. *Suponha que $r(p; \beta, x)$ denota a expressão em (3.3), onde β é o parâmetro conhecido. Então para $n/2 < \sum_{i=1}^n e^{-\beta x_i}$, a equação $r(p; \beta, x) = 0$ tem pelo menos uma raiz.*

Demonstração. Seja $r(p, \beta|x) = -\frac{n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}}$.

Suponha β conhecido. Então p_0 é uma raiz de $r(p, \beta|x)$, para $0 < p_0 < 1$ se, e somente se, $r(0) < 0$ e $r(1) > 0$ ou $r(0) > 0$ e $r(1) < 0$.

Calculemos $r(0)$.

$$r(0) = \lim_{p \rightarrow 0} r(p) = \lim_{p \rightarrow 0} \left[-\frac{n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}} \right] = -n \lim_{p \rightarrow 0} \frac{1}{p \ln p} - \frac{n}{1} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - e^{-\beta x_i}}. \text{ Mas } \lim_{p \rightarrow 0} \frac{1}{p \ln p} = +\infty. \text{ Logo, } r(0) = +\infty.$$

Então, deve-se mostrar que $r(1) < 0$ para que haja raiz em $r(p)$.

De fato,

$$\begin{aligned} r(1) &= \lim_{p \rightarrow 1} r(p) = \lim_{p \rightarrow 1} \left[-\frac{n}{p \ln p} - \frac{n}{1-p} - \sum_{i=1}^n \frac{e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}} \right] = \\ &= \lim_{p \rightarrow 1} \left[-n \left(\frac{1-p+p \ln p}{p(1-p) \ln p} \right) \right] - \sum_{i=1}^n e^{-\beta x_i} \frac{1}{\lim_{p \rightarrow 1} [1 - (1-p)e^{-\beta x_i}]} \end{aligned} \quad (3.5)$$

Avaliemos o primeiro limite de (3.5). Aplicando a regra de L'hôpital tem-se:

$$\begin{aligned} \lim_{p \rightarrow 1} \left[\frac{1-p+p \ln p}{p(1-p) \ln p} \right] &= \lim_{p \rightarrow 1} \left[\frac{-1 + \ln p + 1}{(1-2p) \ln p + p(1-p)\frac{1}{p}} \right] = \lim_{p \rightarrow 1} \left[\frac{\ln p}{(1-2p) \ln p + 1-p} \right] = \\ &= \lim_{p \rightarrow 1} \left[\frac{\frac{1}{p}}{-2 \ln p + (1-2p)\frac{1}{p} - 1} \right] = \frac{1}{0 + \frac{1-2}{1} - 1} = -\frac{1}{2}. \end{aligned}$$

O 2º limite da equação é dado por:

$$\lim_{p \rightarrow 1} \frac{1}{1 - (1-p)e^{-\beta x_i}} = \frac{1}{1} = 1.$$

Logo, de (3.5) tem-se:

$$r(1) = \lim_{p \rightarrow 1} r(p) = -n \left(-\frac{1}{2} \right) - \sum_{i=1}^n e^{-\beta x_i}.$$

Assim, $r(1) < 0$ se, e somente se, $\frac{n}{2} - \sum e^{-\beta x_i} < 0$, ou seja, $\frac{n}{2} < \sum_{i=1}^n e^{-\beta x_i}$. \square

Teorema 2. *Suponha que $g(\beta; p, x)$ denota a expressão em (3.4), e $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, onde p é o parâmetro conhecido. Então, para um $p \in (0, 1)$ dado, a raiz de $g(\beta; p, x)$ dentro do intervalo $[p\bar{x}^{-1}, \bar{x}^{-1}]$ é única.*

Demonstração. Seja $g(\beta; p, x) = \frac{n}{\beta} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}$ e p conhecido.

Se existe uma raiz de $g(\beta; p, x)$ dentro do intervalo $[p\bar{x}^{-1}, \bar{x}^{-1}]$, então a função g para um desses pontos resulta em um valor positivo e no outro ponto um valor negativo.

Assim, para o ponto \bar{x}^{-1} , a função g se comporta como a seguir:

$$\begin{aligned} g(\bar{x}^{-1}; p, x) &= \frac{n}{\bar{x}^{-1}} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{x_i}{\bar{x}}}}{1-(1-p)e^{-\frac{x_i}{\bar{x}}}} \\ g(\bar{x}^{-1}; p, x) &= n \frac{\sum_{i=1}^n x_i}{n} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{x_i}{\bar{x}}}}{1-(1-p)e^{-\frac{x_i}{\bar{x}}}} \\ g(\bar{x}^{-1}; p, x) &= - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{x_i}{\bar{x}}}}{1-(1-p)e^{-\frac{x_i}{\bar{x}}}} \end{aligned}$$

Sabendo que $x_i \geq 0$, $(1-p) \geq 0$, $0 < e^{-\frac{x_i}{\bar{x}}} < 1$ e, conseqüentemente, $1-(1-p)e^{-\frac{x_i}{\bar{x}}} > 0$, $\sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{x_i}{\bar{x}}}}{1-(1-p)e^{-\frac{x_i}{\bar{x}}}} > 0$. Portanto, $g(\bar{x}^{-1}; p, x) < 0$.

Considere agora o ponto $p\bar{x}^{-1}$:

$$\begin{aligned} g(p\bar{x}^{-1}; p, x) &= \frac{n}{p\bar{x}^{-1}} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}} \\ g(p\bar{x}^{-1}; p, x) &= n \frac{\sum_{i=1}^n x_i}{pn} - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}} \\ g(p\bar{x}^{-1}; p, x) &= \sum_{i=1}^n x_i \left(\frac{1-p}{p} \right) - \sum_{i=1}^n \frac{(1-p)x_i e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}}. \end{aligned}$$

Necessariamente, $g(p\bar{x}^{-1}; p, x) > 0$, isto é,

$$\sum_{i=1}^n x_i (1-p) \left[\frac{1}{p} - \frac{e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}} \right] > 0.$$

É fato que $x_i(1-p) > 0$. Portanto, resta provar que $\frac{1}{p} - \frac{e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}} > 0$. Assim,

$$\frac{1}{p} > \frac{e^{-\frac{px_i}{\bar{x}}}}{1-(1-p)e^{-\frac{px_i}{\bar{x}}}} \quad \text{implica em} \quad p < \frac{1-(1-p)e^{-\frac{px_i}{\bar{x}}}}{e^{-\frac{px_i}{\bar{x}}}}.$$

Além disso, $p < e^{\frac{px_i}{\bar{x}}} - (1-p)$ é igual a $1 < e^{\frac{px_i}{\bar{x}}}$.

Portanto, tem-se que $g(p\bar{x}^{-1}; p, x) > 0$.

□

Tahmasbi and Rezaei (2008) também mostram em detalhes que a matriz Informação de Fisher da $El(p, \beta)$ é dada por:

$$I(p, \beta) = n \begin{bmatrix} -\frac{\ln p + 1}{(p \ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2 \ln p + 3}{2(1-p)^2 \ln p} & \frac{1-p+p \ln p}{2\beta p(1-p) \ln p} \\ \frac{1-p+p \ln p}{2\beta p(1-p) \ln p} & -\frac{\text{dilog}(p)}{\beta^2 \ln p} \end{bmatrix}, \quad (3.6)$$

onde $\text{dilog}(\cdot)$ é a função dilogarítmica definida como $\text{dilog}(x) = \int_1^x \frac{\ln t}{1-t} dt$.

Portanto, os estimadores de máxima verossimilhança \hat{p} e $\hat{\beta}$ possuem distribuição normal conjunta assintótica dada por

$$(\hat{p}, \hat{\beta}) \sim N_2[(p, \beta), I^{-1}(p, \beta)] \quad \text{para } n \rightarrow \infty. \quad (3.7)$$

Análise Bayesiana

Neste capítulo serão apresentadas algumas distribuições a priori não-informativas e ilustrações numéricas para comparará-las. Além disso será detalhado o método MCMC para o caso da distribuição $EL(p, \beta)$.

A inferência Bayesiana é um método eficaz para a solução de diversos problemas e bastante popular hoje em dia. Este tipo de inferência estatística baseia-se na distribuição a posteriori dos parâmetros estudados. Para mais detalhes veja Colosimo and Giolo (2006).

Para a $EL(p, \beta)$, a distribuição a posteriori é dada por:

$$p(p, \beta|x) \propto \pi(p, \beta)L(p, \beta|x). \quad (4.1)$$

A distribuição a priori representa o conhecimento ou o estado de incerteza sobre os parâmetros p e β antes do experimento e a posteriori descreve a informação atualizada sobre (p, β) depois dos dados \mathbf{x} serem observados.

Para avançar com a análise Bayesiana é necessário especificar distribuições a priori para os parâmetros. Para a $EL(p, \beta)$, assumiram-se diferentes distribuições a priori, as quais serão apresentadas nas próximas seções.

Este Capítulo foi baseado no artigo Moala and Garcia (2013) publicado pela revista Quality Engineering em maio de 2013.

4.1 Priori de Jeffreys

A priori de Jeffreys (1967), comumente utilizada na literatura, é uma distribuição a priori que desempenha um importante papel na Inferência Bayesiana. Ela é dada por:

$$\pi(p, \beta) \propto \sqrt{\det I(p, \beta)}, \quad (4.2)$$

onde $I(p, \beta)$ é a matriz informação de Fisher dada em (3.6).

A priori de Jeffreys é comumente utilizada devido à sua propriedade de invariância sob transformação um-a-um dos parâmetros p e β . Box and Tiao (1973) explicam a obtenção da priori de Jeffreys em termos da verossimilhança "data translated". Portanto, para os parâmetros da EL (p e β) a priori de Jeffreys é dada por:

$$\pi(p, \beta) \propto \frac{1}{\beta} \sqrt{h(p)}, \quad (4.3)$$

onde

$$h(p) = \left(\frac{\ln p + 1}{(p \ln p)^2} - \frac{1}{(1-p)^2} - \frac{(1-4p)/p^2 - 2 \ln p + 3}{2(1-p)^2 \ln p} \right) \frac{d \log(p)}{\ln p} - \left(\frac{1-p+p \ln p}{2p(1-p) \ln p} \right)^2.$$

Uma possível alternativa é considerar a priori não-informativa da seguinte forma: $\pi(p, \beta) = \pi(p|\beta)\pi(\beta)$. Achcar et al. (2010) considera uma outra versão da priori de Jeffreys dada por:

$$\pi(p, \beta) \propto \sqrt{E \left(-\frac{\partial^2}{\partial p^2} \ln L \right)} \pi(\beta), \quad (4.4)$$

onde $E \left(-\frac{\partial^2}{\partial p^2} \ln L \right)$ é dado por (3.6) e $\pi(\beta)$ é uma priori não-informativa dada por $\pi(\beta) \propto \frac{1}{\beta}$. Deste modo,

$$\pi(p, \beta) \propto \frac{1}{\beta} \sqrt{I_{pp}}, \quad (4.5)$$

onde $I_{pp} = -\frac{\ln p + 1}{(p \ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2 \ln p + 3}{2(1-p)^2 \ln p}$, $\beta > 0$, $0 < p < 1$.

4.2 Produto de distribuições a priori independentes

Uma outra maneira de se representar a priori não-informativa é dada pelo produto das distribuições a priori independentes. É mais fácil encontrar uma priori para cada parâmetro com condição de independência do que considerar uma priori conjunta para todos os parâmetros simultaneamente. Por isso, para a Exponencial-Logarítmica foi considerado o produto das seguintes distribuições a priori:

$$\pi_p(p) \sim \text{Beta}(a_p, b_p) \quad \text{e} \quad \pi_\beta(\beta) \sim \text{Gama}(a_\beta, b_\beta) \quad (4.6)$$

com hiperparâmetros a_p, b_p, a_β e b_β assumindo valores específicos. A priori uniforme $U(0, 1)$ para o parâmetro p também é apropriada, pois $p \in (0, 1)$.

Então, a distribuição a posteriori conjunta para os parâmetros p e β é proporcional ao

produto da função de verossimilhança (3.1) com a priori $\pi(p, \beta)$ dada em (4.2) ou (4.6), e resulta em:

$$p(p, \beta|x) \propto \frac{1}{\ln p^n} \left(\frac{\beta^n (1-p)^n e^{-\beta \sum_{i=1}^n x_i}}{\prod_{i=1}^n (1 - (1-p)e^{-\beta x_i})} \right) \pi(p, \beta). \quad (4.7)$$

4.3 MCMC para a $EL(p, \beta)$

Como não é possível encontrar uma expressão analítica para a marginal da distribuição a posteriori para cada parâmetro e, portanto, não é possível extrair suas características, tais como estimativas Bayesianas e intervalos de credibilidade, o algoritmo MCMC torna-se indispensável. Através dele é possível obter uma amostra de valores de p e β a partir da posteriori conjunta. Os detalhes da implementação do algoritmo MCMC utilizados neste trabalho são dados abaixo.

- Escolha valores iniciais p_0 e β_0 ;
- No passo $i + 1$, é dado um novo valor p_{i+1} condicionado ao atual p_i da distribuição Beta $B(bp_i/(1-p_i), b)$;
- O candidato p_{i+1} será aceito com probabilidade dada pelo algoritmo de Metropolis - Hastings

$$\alpha(p_i, p_{i+1}) = \min \left\{ 1, \frac{B(bp_i/(1-p_i), b)p(p_{i+1}, \beta_i|x)}{B(bp_{i+1}/(1-p_{i+1}), b)p(p_i, \beta_i|x)} \right\};$$

- Amostre o valor novo β_{i+1} da distribuição Gama $G(\beta_i/d, d)$;
- O candidato β_{i+1} também será aceito com probabilidade dada pelo algoritmo de Metropolis - Hastings

$$\alpha(\beta_i, \beta_{i+1}) = \min \left\{ 1, \frac{G(\beta_i/d, d)p(p_{i+1}, \beta_{i+1}|x)}{G(\beta_{i+1}/d, d)p(p_{i+1}, \beta_i|x)} \right\}.$$

As distribuições propostas para os parâmetros p e β , foram escolhidas de forma a obter uma boa mistura das cadeias. A convergência das amostras do MCMC são avaliadas utilizando o critério de Raftery and Lewis (1992).

4.4 Análise de dados simulados

Nesta seção foram feitas análises utilizando as distribuições a priori apresentadas anteriormente. Elas foram baseadas em três conjuntos de 100 amostras de tamanhos $n = 20$, 50 e 70 geradas da distribuição Exponencial-Logarítmica. Os valores utilizados para os parâmetros foram iguais a $p = 0.5$ e $\beta = 2$.

Foi necessário recorrer ao algoritmo MCMC para obter as marginais das distribuições a posteriori, os estimadores Bayesianos e os intervalos de credibilidade. O *software* utilizado

para as análises foi o R. Para encontrar os EMV, utilizou-se a função "maxNR" do R, que é baseada no método de Newton-Raphson.

A cadeia para o MCMC foi executada com 15000 iterações, com período de *burn-in* de 5000 e convergência monitorada por seus *output*, critério de Geweke e pelos plots de autocorrelação, como mostra a figura (4.1) para a priori de Jeffreys. Os plots do MCMC sugerem que foi alcançada a convergência com uma taxa de aceitação em torno de 30-40%.

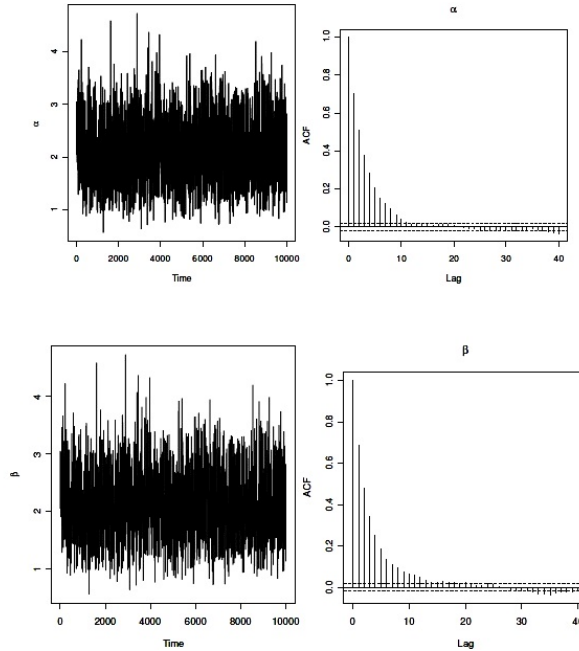


Figura 4.1: Output do MCMC e plots de autocorrelação para priori de Jeffreys considerando $n = 20$

Inicialmente foram examinadas as performances dos pares de distribuições a priori $U(0, 1)$ (que é igual a $Beta(1, 1)$) & $Gama(0.01, 0.01)$, e $Beta(0.5, 0.5)$ & $Gama(0.01, 0.01)$ para os parâmetros p e β , respectivamente, supondo-se independência entre os parâmetros para essas distribuições a priori. Também foi examinada a priori de Jeffreys, que representa uma estrutura de dependência entre os parâmetros. Na Tabela (4.1), as médias das distribuições a posteriori de cada priori são comparadas juntamente com o EMV.

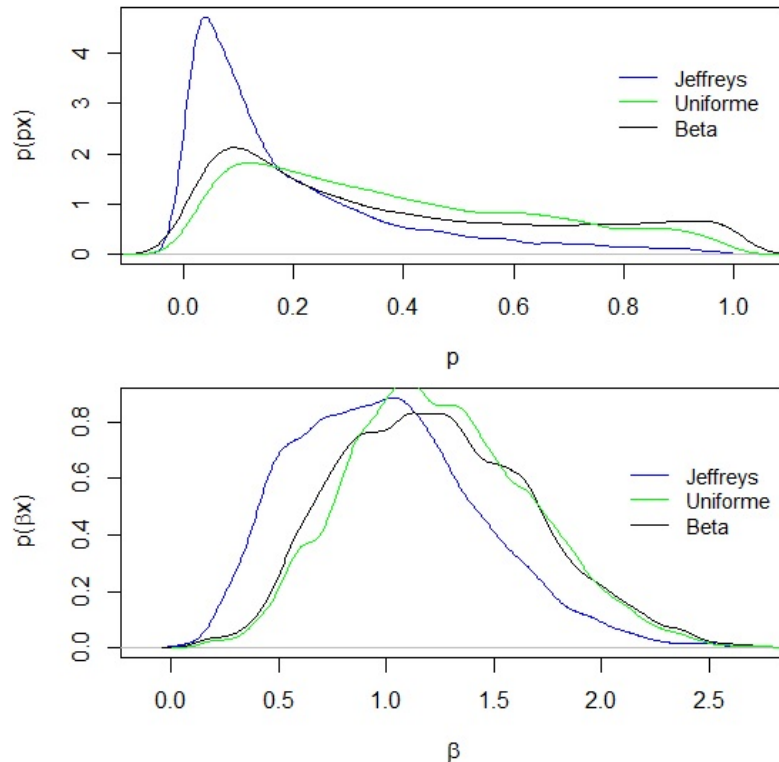
Examinando a Tabela (4.1), foi observado que a abordagem Bayesiana proporciona melhores estimativas para os parâmetros p e β do que o EMV. As estimativas Bayesianas para as distribuições a priori examinadas estão próximos dos valores verdadeiros $p = 0.5$ e $\beta = 2$. As distribuições a priori *Beta* e *Uniforme* produzem estimativas Bayesianas bem melhores que a priori de Jeffreys quando n é pequeno. Além disso, o EMV melhora muito pouco quando n aumenta.

Na Tabela (4.1) a distribuição *Beta* usada para p é $Beta(0.5, 0.5)$, isto é, $\pi(p) \propto \frac{1}{\sqrt{p(1-p)}}$.

Os intervalos Bayesianos de 95% de credibilidade para cada parâmetro p e β são também obtidos através do algoritmo MCMC e mostrados na Tabela (4.2). A priori *Beta* produz intervalos com maior cobertura e, por isso, ela poderia ser considerada, a

Tabela 4.1: EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$ e $\beta = 2$.

p	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	0.1515 (0.0348)	0.5757 (0.1339)	0.5069 (0.1267)	0.3560 (0.11161)
$n = 50$	0.2022 (0.0263)	0.5749 (0.1568)	0.5446 (0.1263)	0.4282 (0.1430)
$n = 70$	0.3129 (0.0503)	0.5768 (0.1823)	0.5784 (0.1175)	0.4307 (0.1338)
β	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	1.0497 (0.2281)	2.1606 (0.6354)	1.9904 (0.5476)	1.7339 (0.4691)
$n = 50$	1.6784 (0.2356)	2.0093 (0.3779)	2.0534 (0.3789)	1.8499 (0.3668)
$n = 70$	1.9497 (0.2243)	2.0257 (0.3818)	2.0649 (0.2839)	1.8264 (0.2819)

Figura 4.2: Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 20$.

princípio, como a priori mais indicada. Por outro lado, quando o tamanho da amostra é muito pequeno (como é o caso dos estudos de confiabilidade), os intervalos de confiança para p , baseados na abordagem clássica (EMV), poderiam incluir números negativos ou maiores que 1, embora não pudessem ser negativos ou superiores a 1. Portanto, o uso de propriedades assintóticas não é uma boa escolha para se construir intervalos de confiança.

As figuras (4.2), (4.3) e (4.4) mostram as densidades marginais a posteriori para os dois parâmetros, p e β , com amostras de tamanho $n = 20, 50$ e 70 .

A figura (4.2) mostra claramente as diferenças e similaridades entre as distribuições a priori dos parâmetros p e β quando $n = 20$. É possível observar que as distribuições a priori para o parâmetro p possuem notável assimetria. Além disso, a marginal da posteriori resultante da priori de Jeffreys se desloca significativamente para perto do zero.

Tabela 4.2: Intervalos bayesianos com 95% de confiança para as marginais das distribuições a posteriori em relação a p e β .

p	Beta/Gama	Uniform/Gama	Jeffreys
$n = 20$	(0.2232, 0.7720)	(0.2216, 0.6901)	(0.1502, 0.8157)
$n = 50$	(0.2050, 0.8064)	(0.3013, 0.7284)	(0.2050, 0.8064)
$n = 70$	(0.1502, 0.8157)	(0.3275, 0.7372)	(0.2232, 0.7720)
β	Beta/Gama	Uniform/Gama	Jeffreys
$n = 20$	(1.3443, 3.8373)	(1.1835, 3.2901)	(1.3727, 2.8090)
$n = 50$	(1.3538, 2.7381)	(1.3687, 2.7040)	(1.3538, 2.7381)
$n = 70$	(1.3727, 2.8090)	(1.5741, 2.7547)	(1.3443, 3.8373)

As distribuições a priori Beta & Gama e Uniforme & Gama produzem densidades a posteriori muito similares e bastante *flat* dentro do intervalo (0, 1).

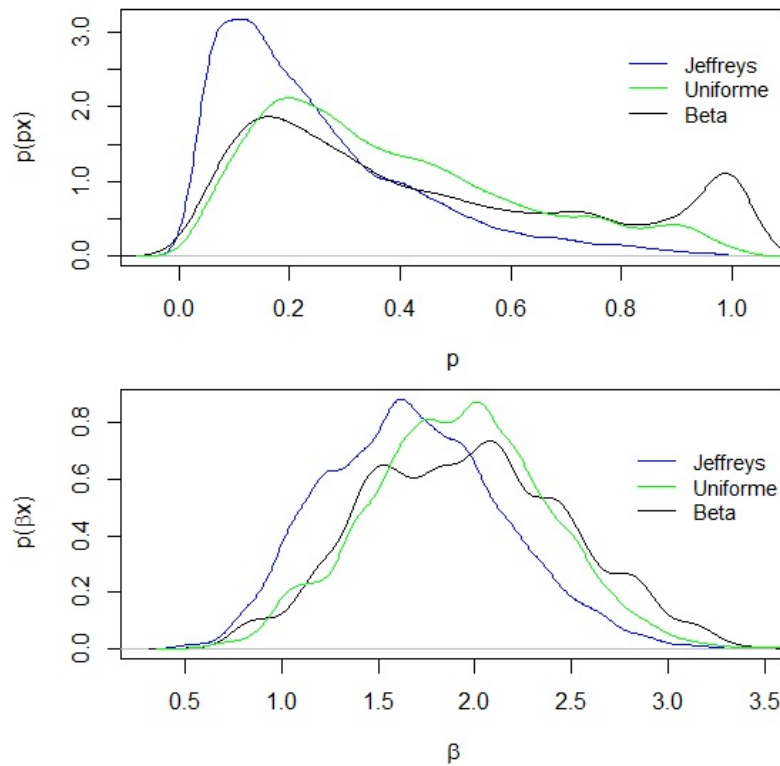


Figura 4.3: Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 50$

Pode-se observar na Figura (4.2) que as densidades a posteriori do parâmetro β não se diferenciam muito. Quando n aumenta, nota-se na Figura (4.3) que todas as densidades a posteriori se aproximam mais do verdadeiro valor de *beta* e tendem a se tornar idênticas.

Quando o tamanho da amostra n é moderadamente grande, não há diferença entre as distribuições a priori para ambos os parâmetros e, portanto, a escolha delas torna-se irrelevante, como pode ser visto na Figura (4.4).

Para entender a bimodalidade da densidade a posteriori para o parâmetro p ocorrida na Figura (4.2), (4.3) e (4.4) (quando a priori *Beta* é utilizada), deve-se observar as

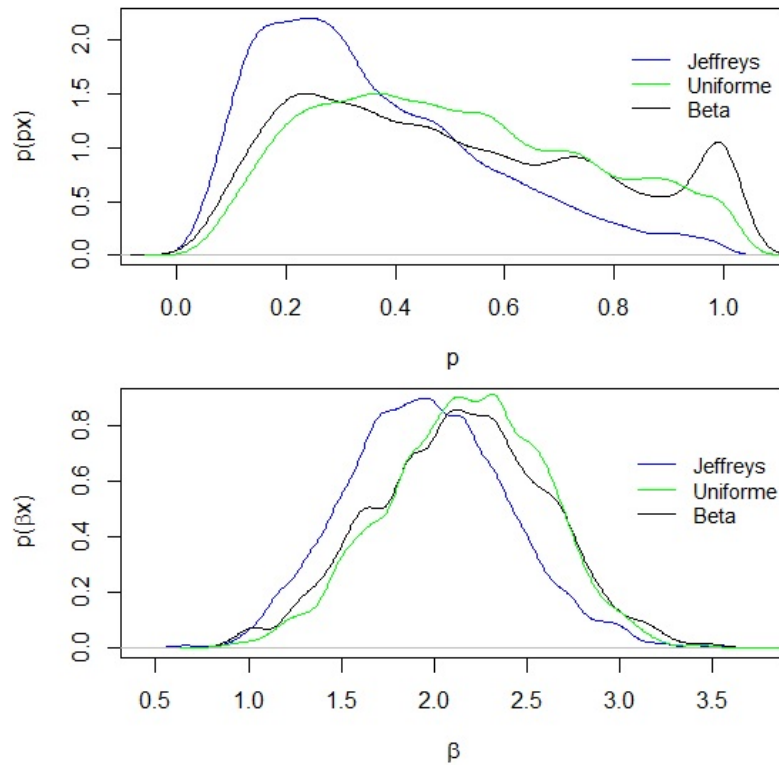


Figura 4.4: Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 70$

figuras (4.5) e (4.6). Elas exibem os perfis da verossimilhança, a priori perfilada de Beta e a densidade a posteriori para p com $n = 10, 20$ e 50 , mostrando como a inferência Bayesiana comporta-se: com mais ou menos informação da verossimilhança e da priori.

Na figura (4.5), foram utilizados os valores $a = 0.5$ e $b = 0.5$ como hiperparâmetros da priori Beta e considerados $n = 10, 20$ e 50 . Na figura (4.6), a priori é a Beta também, agora com hiperparâmetros iguais a $a = 2$ e $b = 3$.

No primeiro painel da Figura (4.5), a condição para a existência do EMV proposta por Tahmasbi and Rezaei (2008) e dada por $\frac{n}{2} < \sum_{i=1}^n e^{\beta x_i}$, não foi alcançada, e o plot mostra que a função de verossimilhança não possui um ponto de máximo no intervalo $(0,1)$, ou seja, não há EMV para o parâmetro p . Em contrapartida, o estimador de Bayes é obtido para p , embora produza uma estimativa ruim.

No painel 2 da Figura (4.5), as condições são satisfeitas. Consequentemente, há um EMV e ele é único, além de ser próximo do verdadeiro valor. Com a estimação Bayesiana, a priori $B(0.5, 0.5)$ atribui elevada probabilidade nos extremos do intervalo $(0, 1)$, que combinada com a verossimilhança resulta em uma posteriori bimodal com as modas próximas dos extremos.

Já no painel 3 da Figura (4.5), aumentando a informação dos dados, a verossimilhança atinge um ponto de máximo reduzindo substancialmente o peso da priori na densidade a posteriori. Portanto, a bimodalidade tende a desaparecer.

A Figura (4.6) mostra que para uma priori adequada como a $B(2, 3)$ com moda próxima

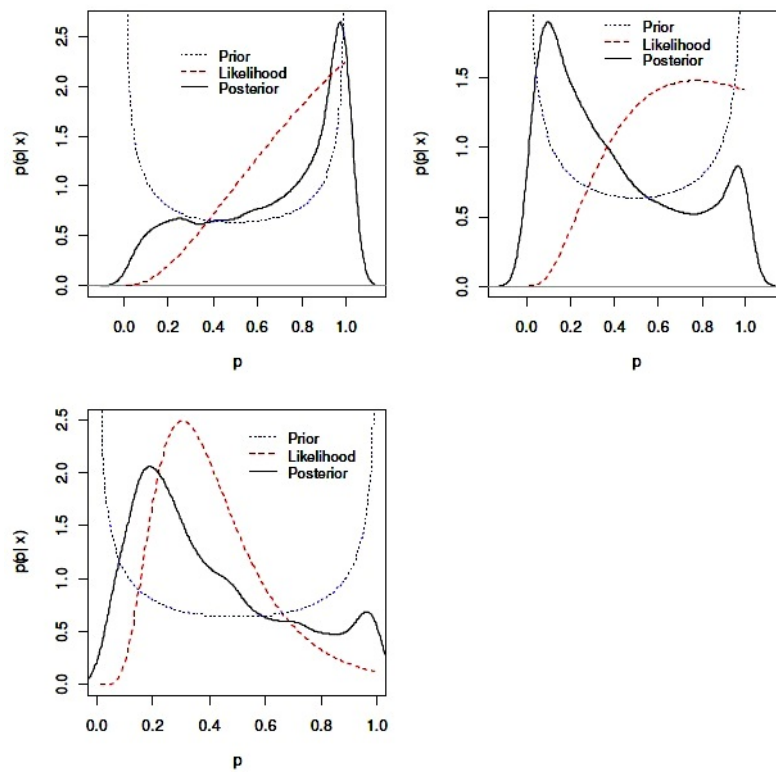


Figura 4.5: Verossimilhança, densidades a priori e a posteriori do parâmetro p para $n = 10, 20$ e 50 .

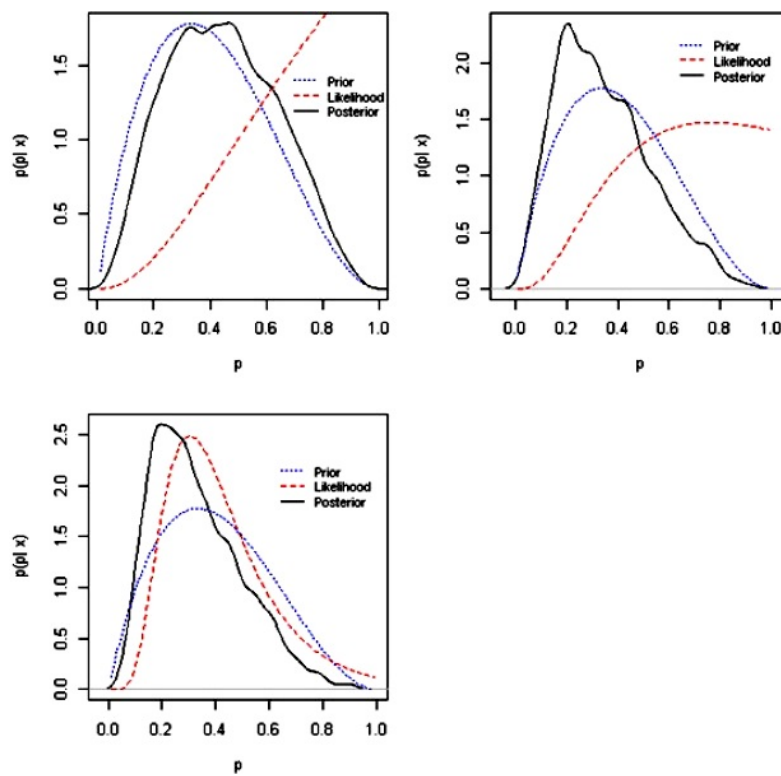


Figura 4.6: Verossimilhança, densidades a priori e a posteriori do parâmetro β para $n = 10, 20$ e 50 .

Tabela 4.3: Resultados obtidos da posteriori para o parâmetro p .

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.3198	0.2915	(0.0152, 0.9779)
Uniforme/Gama	0.3467	0.2594	(0.0296, 0.9333)
Jeffreys	0.1633	0.1891	(0.0368, 0.7409)

do verdadeiro valor p , a bimodalidade da densidade a posteriori não ocorre para nenhum tamanho de amostra.

4.5 Análise com Conjunto de Dados Reais

Nesta seção, foram considerados dados relativos ao tempo de vida de um certo tipo de isolante elétrico exposto a um estresse de tensão constante. Esses dados foram introduzidos por Lawless (1982) e são não-censurados. Além disso, representam o tempo de vida (em minutos) até a ocorrência de falha igual a: 0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89.

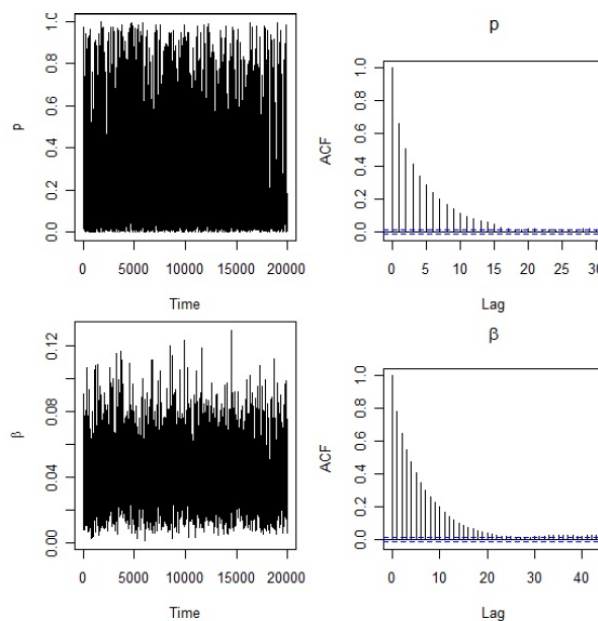


Figura 4.7: Output do MCMC e plots de autocorrelação utilizando a priori de Jeffreys

Lawless utilizou esse conjunto de dados para ilustrar o ajuste do modelo Weibull na análise de Confiabilidade. Neste trabalho, foi utilizada a distribuição Exponencial - Logarítmica com densidade (2.3) para analisar os dados.

Os estimadores de máxima verossimilhança e seus respectivos desvios-padrão (em parêntesis) para p e β são dados por $\hat{p} = 0.0978(0.3127)$ e $\hat{\beta} = 0.0393(0.0069)$, com intervalo de confiança de 95% igual a: $(-0.1536; 0.3492)$ e $(0.00197; 0.07662)$ para p e β , respectivamente. Note que o limite inferior do intervalo para o parâmetro p é negativo, sendo mais um indício de que o estimador de máxima verossimilhança não é um bom estimador a ser utilizado para estimar os parâmetros da distribuição Exponencial-Logarítmica.

Tabela 4.4: Resultados obtidos da posteriori para o parâmetro β .

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.0483	0.0187	(0.0165, 0.0873)
Uniforme/Gama	0.0513	0.0177	(0.0195, 0.0879)
Jeffreys	0.0396	0.0182	(0.0259, 0.0796)

Tabela 4.5: Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.

	AIC	BIC	DIC
Weibull	231.9720	236.0553	140.830
EL	141.8041	145.8874	139.167

Para a análise Bayesiana do conjunto de dados foram assumidas distribuições a priori (4.3), (4.6) e o par Uniforme & Gama para p e β , respectivamente.

Utilizando-se o MCMC foi executada uma cadeia com 25000 iterações e descartadas as primeiras 5000 como *burn-in*. Os plots do MCMC e o critério de Geweke (1992) (veja Apêndice A3) sugerem que houve convergência e o algoritmo mostrou um grau de aceitação em torno de 35%, considerando a priori de Jeffreys. O output do MCMC é mostrado na Figura (4.7).

Os sumários da posteriori considerando as diferentes distribuições a priori são dados nas Tabelas (4.3) e (4.4).

Para os resultados das Tabelas (4.3) e (4.4) observou-se que a priori de Jeffreys resultou em estimativas menores em comparação com as distribuições a priori Uniforme & Gama e Beta& Gama, principalmente para o parâmetro p .

A comparação das marginais das densidades a posteriori são dadas na Figura (4.8). É importante notar nesse exemplo que a marginal das densidades a posteriori na Figura (4.2) e (4.8) são similares tanto para dados simulados quanto para os dados de Lawless, com tamanho de amostras iguais a $n = 20$ e 19 , respectivamente. Este resultado confirma que o método Bayesiano proposto neste trabalho, para estimar os parâmetros da distribuição Exponencial-Logarítmica, é o mais apropriado.

Para determinar a distribuição mais apropriada para ajustar esse conjunto de dados foram utilizados os critérios de seleção de modelo AIC, BIC e DIC, que são dados na Tabela (4.5) para a distribuição Exponencial-Logarítmica e para a distribuição Weibull. Deve-se observar que para a distribuição Exponencial-Logarítmica foram utilizadas as distribuições a priori Uniforme(0, 1) para p e Gama(0.01, 0.01) para β e para a distribuição Weibull foi utilizada para ambos parâmetros a priori Gama(0.01, 0.01).

Analisando a tabela (4.5) pode-se concluir que a Exponencial-Logarítmica proporciona um melhor ajuste para os dados, pois os valores para o AIC e BIC são pequenos em comparação com os valores para a Weibull. Considerando o critério DIC, as duas distribuições são apropriadas para ajustar os dados. No entanto, se é necessário escolher entre

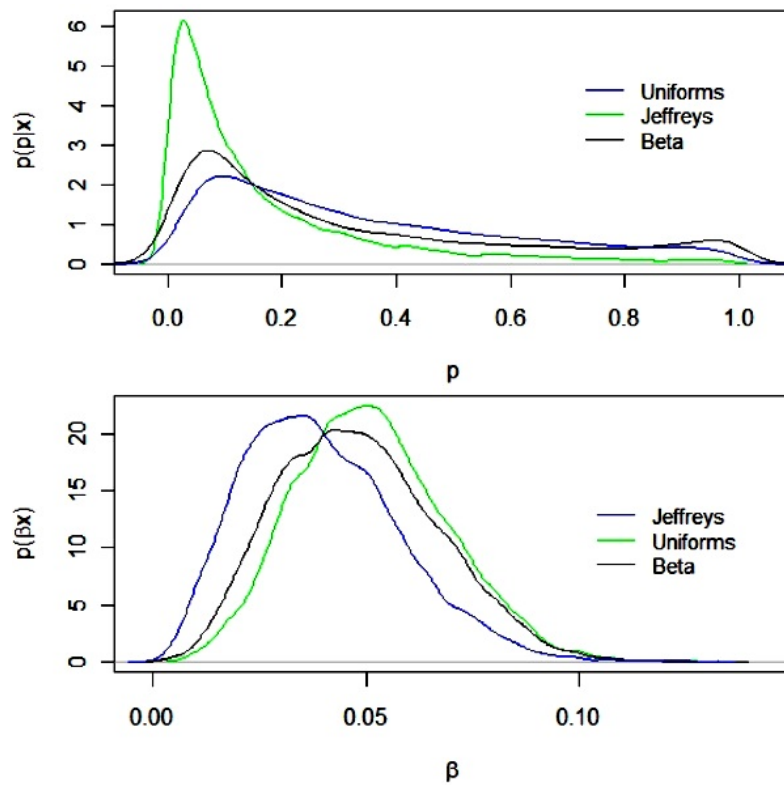


Figura 4.8: Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)

as duas distribuições através do DIC, então será escolhida a Exponencial-Logarítmica, pois possui menor valor para ele.

Distribuições a priori MDIP & Gama e de Referência

Nesse capítulo duas importantes distribuições a priori não-informativas foram examinadas para os parâmetros desconhecidos p e β : a priori MDIP (Maximal Data Information) e a priori de Referência. O objetivo foi comparar essas distribuições a priori com a priori de Jeffreys, já estudada anteriormente. A motivação para este estudo vem da importância do uso de distribuições a priori não-informativas, como a MDIP, que não têm sido frequentemente consideradas na literatura Bayesiana. Os estudos comparativos entre as distribuições a priori foram realizados para diferentes tamanhos de amostra.

5.1 Priori MDIP & Gama

A priori de Máxima Informação dos Dados (MDIP) foi proposta por Zellner (1977) e também é apropriada para ser utilizada no problema de inferência da distribuição Exponencial-Logarítmica.

Quando um experimento é realizado, espera-se que as informações sobre o parâmetro fornecidas pelos dados sejam maiores do que as obtidas pela priori; caso contrário, não haveria justificativa para a realização do mesmo. Portanto, é desejável uma distribuição a priori $\pi(\phi)$ que proporcione o máximo de ganho de informação fornecida pelos dados. Partindo dessa ideia, Zellner (1977) propôs uma priori que maximiza a informação média da densidade dos dados em relação à densidade a priori.

Seja

$$H(\phi) = \int_a^b f(x|\phi) \ln f(x|\phi) dx \quad (5.1)$$

dada como a entropia negativa da densidade $f(x|\phi)$, ou seja, a medida da informação em

$f(x|\phi)$. Assim, o seguinte critério funcional é empregado na abordagem MDIP:

$$G[\pi(\phi)] = \int_a^b H(\phi)\pi(\phi)d\phi - \int_a^b \pi(\phi) \ln \pi(\phi)d\phi. \quad (5.2)$$

$G[\pi(\phi)]$ é maximizada pela seleção de $\pi(\phi)$ sujeito à $\int_a^b \pi(\phi)d\phi = 1$.

A solução é então uma priori própria dada por

$$\pi(\phi) = k \exp\{H(\phi)\}, \quad a \leq \phi \leq b, \quad (5.3)$$

onde $k^{-1} = \int_a^b \exp\{H(\phi)\}d\phi$ é a constante de normalização.

Portanto, a priori MDIP dá ênfase na informação da densidade dos dados ou função de verossimilhança, ou seja, a sua informação é fraca em comparação com a informação dos dados. Mais informações sobre a construção desta priori podem ser encontradas em Zellner (1984). Zellner (1990) mostra diversas propriedades e condições adicionais da priori MDIP.

Para se obter uma priori conjunta $\pi(p, \beta)$ tal que seja própria, isto é, $\int \int \pi(p, \beta)dpd\beta = 1$, consideraremos β conhecido. Teremos então uma versão "parcial" da priori proposta por Zellner (1977) pois o parâmetro β assume valores dentro de um intervalo ilimitado, condição não satisfeita para uma priori MDIP própria. Portanto, será utilizada a priori não-informativa Gama(0.01, 0.01) para o parâmetro β .

Teorema 3. *Considerando o parâmetro β conhecido, a priori MDIP para o parâmetro p é dada por:*

$$\pi_Z(p|\beta) = \left(-\frac{1}{\ln p}\right) \frac{(1-p)\exp\left\{-\frac{\text{polylog}(2,1-p)}{\ln p}\right\}}{\sqrt{p}}, \quad (5.4)$$

onde $\text{polylog}(z, \alpha) = \sum_{k=1}^{\infty} \frac{z^k}{k^\alpha}$.

Demonstração. Primeiramente, é necessário encontrar a medida de informação,

$$H(p|\beta) = \int_0^{\infty} \left(-\frac{1}{\ln p}\right) \frac{\beta(1-p)e^{-\beta t}}{1-(1-p)e^{-\beta t}} \ln \left(\left(-\frac{1}{\ln p}\right) \frac{\beta(1-p)e^{-\beta t}}{1-(1-p)e^{-\beta t}} \right) dt. \quad (5.5)$$

Calculando a integral acima tem-se:

$$H(p|\beta) = \ln \left(\frac{1}{-\ln p} \beta(1-p) \right) - \beta E(T) - E[\ln(1-(1-p)e^{-\beta T})].$$

Sabendo que a média da distribuição $EL(p, \beta)$ é $E(T) = \frac{\text{polylog}(2,1-p)}{\beta \ln p}$ e $E[\ln(1-(1-p)e^{-\beta T})] = \frac{1}{2} \ln p$, tem-se,

$$H(p|\beta) = \ln \left(\frac{\beta(1-p)}{-\ln p} \right) - \frac{\text{polylog}(2,1-p)}{\ln p} - \frac{1}{2} \ln p. \quad (5.6)$$

Utilizando a equação $H(p)$ em (5.3), a priori MDIP para o parâmetro p é dada por:

$$\pi_Z(p|\beta) \propto \exp\{H(p)\} \propto \left(-\frac{1}{\ln p}\right) \frac{\beta(1-p)\exp\left\{-\frac{\text{polylog}(2,1-p)}{\ln p}\right\}}{\sqrt{p}}. \quad (5.7)$$

□

Uma vez obtida uma priori própria $\pi(p|\beta)$ pode-se determinar uma priori conjunta para p e β como:

$$\pi(p, \beta) = \pi(p|\beta)\pi(\beta). \quad (5.8)$$

Basta então especificar uma priori própria para β . Considere uma priori Gama com hiperparâmetros "c" e "d". Desta forma, a priori conjunta MDIP & Gama será dada por:

$$\pi(p, \beta) = \left(-\frac{1}{\ln p}\right) \frac{\beta(1-p)\exp\left\{-\frac{\text{polylog}(2,1-p)}{\ln p}\right\}}{\sqrt{p}} \frac{c}{\Gamma(d)} (c\beta)^{d-1} e^{-c\beta} \quad (5.9)$$

A correspondente posteriori é dada por:

$$p(p, \beta|x) \propto \frac{1}{(\ln p)^{n+1}} \frac{\beta^{n+d}(1-p)^{n+1}\exp\left\{-\left(\beta(c + \sum_{i=1}^n x_i) + \frac{\text{polylog}(2,1-p)}{\ln p}\right)\right\}}{\sqrt{p} \prod_{i=1}^n (1 - (1-p)e^{-\beta x_i})} \frac{c^d}{\Gamma(d)}, \quad (5.10)$$

em que $c = d = 0.01$, parâmetros da priori Gama(c,d).

5.2 Priori de Referência

Outra priori bem conhecida da classe de distribuições a priori não-informativas é a priori de Referência, descrita primeiramente por Bernardo (1979) e desenvolvida por Berger and Bernardo (1992).

A ideia é obter uma priori $\pi(\phi)$ que maximiza a informação a posteriori esperada sobre os parâmetros. Uma medida natural da informação esperada sobre ϕ fornecida pelos dados \mathbf{x} é dada por

$$I(\phi) = E_x[K(p(\phi|\mathbf{x}), \pi(\phi))], \quad (5.11)$$

em que

$$K(p(\phi|\mathbf{x}), \pi(\phi)) = \int_{\Phi} p(\phi|\mathbf{x}) \log \frac{p(\phi|\mathbf{x})}{\pi(\phi)} d\phi \quad (5.12)$$

é a distância de Kullback-Leibler. Então a priori de Referência é definida como a priori $\pi(\phi)$ que maximiza a distância esperada de Kullback-Leibler entre a distribuição a posteriori e a distribuição a priori $\pi(\phi)$, tomada em relação aos dados experimentais.

A densidade a priori $\pi(\phi)$ que maximiza o funcional (5.11) é obtida através de cálculo de variação, e a solução não é explícita. Contudo, quando a posteriori $p(\phi|\mathbf{x})$ é assintoticamente normal, essa abordagem leva à priori de Jeffreys com apenas um parâmetro. Se, por outro lado, o interesse está em um dos parâmetros, sendo os restantes parâmetros *nuisances*, a situação é bem diferente, e a priori de Referência apropriada não é uma priori de Jeffreys multivariada. Bernardo argumenta que quando existem parâmetros *nuisances*, a priori de Referência dependerá de qual parâmetro é considerado como sendo de principal interesse.

Neste trabalho foi aplicada a versão do algoritmo de Berger and Bernardo (1992) para uma simples porém importante situação que é frequentemente encontrada na prática.

Seja $f(x|\theta_1, \theta_2), (\theta_1, \theta_2) \in \Phi \times \Lambda \subseteq \mathfrak{R}^2$ uma distribuição de probabilidade com 2 parâmetros θ_1 e θ_2 , em que θ_1 é o parâmetro de interesse. Suponha que a distribuição a posteriori conjunta de (θ_1, θ_2) é assintoticamente normal com matriz de covariância $S(\theta_1, \theta_2)$, com $S(\theta_1, \theta_2) = I^{-1}(\theta_1, \theta_2)$ e $S_{11}(\theta_1, \theta_2) = \frac{I_{22}(\theta_1, \theta_2)}{\det I(\theta_1, \theta_2)}$.

Se o espaço do parâmetro *nuisance* $\Lambda(\theta_1) = \Lambda$ é independente de θ_1 , e as funções $S_{11}^{-1/2}$ e $I_{22}^{1/2}(\theta_1, \theta_2)$ podem ser fatorizadas na forma

$$S_{11}^{-1/2}(\theta_1, \theta_2) = f_1(\theta_1)g_1(\theta_2) \quad , \quad I_{22}^{1/2}(\theta_1, \theta_2) = f_2(\theta_1)g_2(\theta_2) \quad (5.13)$$

sendo, $\pi(\theta_1) \propto f_1(\theta_1)$ e $\pi(\theta_2|\theta_1) \propto g_2(\theta_2)$, então a priori de Referência relativa ao par de parâmetros (θ_1, θ_2) é dada por

$$\pi_{\theta_1}(\theta_1, \theta_2) = f_1(\theta_1)g_2(\theta_2), \quad (5.14)$$

e, neste caso, não há necessidade de uma aproximação, mesmo se a condição de existência da priori de Referência não estiver satisfeita.

Geralmente, trocar o parâmetro de interesse pode resultar em uma priori diferente, isto é, a priori de Referência depende do parâmetro de interesse.

O método pode ser generalizado para espaços de parâmetros multidimensionais.

Assim, é possível derivar a priori de Referência em relação aos parâmetros da distribuição $EL(p, \beta)$ dada em (2.3).

Teorema 4. • *Se p é considerado um parâmetro de interesse e β é o parâmetro nuisance, então a priori de Referência para a distribuição $EL(p, \beta)$ é dada por,*

$$\pi_p(p, \beta) = \frac{\sqrt{h(p)}}{2\sqrt{\text{dilog}(p)} \ln(p)(-1+p)p^\beta}, \quad (5.15)$$

em que $h(p) = 2\text{dilog}[-\ln(p) + \ln(p)p^2 - 2 + 4p - 2p^2] + \ln(p)[1 - 2p + 2\ln(p)p + p^2 - 2p^2 \ln(p) + p^2 \ln(p)^2]$.

- Se β é o parâmetro de interesse e p é o parâmetro nuisance então a priori de Referência é dada por:

$$\pi_\beta(p, \beta) = \frac{1}{\beta} \sqrt{\frac{-\ln p + 1}{(p \ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2 \ln p + 3}{2(1-p)^2 \ln p}}. \quad (5.16)$$

Demonstração. **a)** A inversa da matriz de Fisher dada em (3.6) é dada por:

$$S(p, \beta) = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{4 \operatorname{dilog}(p)(\ln p)^2 p^2 (-1+p)^2}{h(p)} & S_{p\beta} \\ S_{p\beta} & \frac{2\beta^2(-1+p) \ln p (p \ln p - 2p + 2 + \ln p)}{h(p)} \end{bmatrix} \quad (5.17)$$

onde $h(p) = 2 \operatorname{dilog}[-\ln(p) + \ln(p)p^2 - 2 + 4p - 2p^2] + \ln(p)[1 - 2p + 2 \ln(p)p + p^2 - 2p^2 \ln(p) + p^2 \ln(p)^2]$.

Assim, tem-se que:

$$S_{11}^{-1/2}(p, \beta) = \frac{\sqrt{h(p)}}{2\sqrt{\operatorname{dilog}(p) \ln(p)p(-1+p)}}.$$

Além disso,

$$I_{22}(p, \beta) = -\frac{\operatorname{dilog}(p)}{\beta^2 \ln(p)} \quad \text{implica em} \quad I_{22}(p, \beta)^{1/2} = \frac{1}{\beta} \sqrt{\frac{-\operatorname{dilog}(p)}{\ln(p)}}.$$

É possível encontrar, então, $f_1(p)$ e $g_2(\beta)$, que são iguais a:

$$f_1(p) = \frac{\sqrt{h(p)}}{2\sqrt{\operatorname{dilog}(p) \ln(p)p(-1+p)}} \quad \text{e} \quad (5.18)$$

$$g_2(\beta) = \frac{1}{\beta}. \quad (5.19)$$

Portanto, a priori de Referência é dada por:

$$\pi_p(p, \beta) = \frac{\sqrt{h(p)}}{2\sqrt{\operatorname{dilog}(p) \ln(p)p(-1+p)}} \cdot \frac{1}{\beta}, \quad (5.20)$$

onde $h(p) = 2 \operatorname{dilog}[-\ln(p) + \ln(p)p^2 - 2 + 4p - 2p^2] + \ln(p)[1 - 2p + 2 \ln(p)p + p^2 - 2p^2 \ln(p) + p^2 \ln(p)^2]$.

b) Do mesmo modo, considerando β o parâmetro de interesse e p o parâmetro nuisance, a inversa da matriz de Fisher dada em (3.6) é dada por:

$$S(\beta, p) = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{2\beta^2(-1+p) \ln p (p \ln p - 2p + 2 + \ln p)}{h(p)} & S_{p\beta} \\ S_{p\beta} & \frac{4 \operatorname{dilog}(p)(\ln p)^2 p^2 (-1+p)^2}{h(p)} \end{bmatrix} \quad (5.21)$$

onde $h(p) = 2dilog[-\ln(p) + \ln(p)p^2 - 2 + 4p - 2p^2] + \ln(p)[1 - 2p + 2\ln(p)p + p^2 - 2p^2\ln(p) + p^2\ln(p)^2]$.

Assim, tem-se que:

$$S_{11}^{-1/2}(\beta, p) = \frac{1}{\beta} \sqrt{\frac{h(p)}{2(-1+p)\ln p(p\ln p - 2p + 2 + \ln p)}}.$$

Além disso,

$$I_{22}^{1/2}(\beta, p) = \sqrt{-\frac{\ln p + 1}{(p\ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2\ln p + 3}{2(1-p)^2\ln p}}.$$

Portanto, $f_1(\beta)$ e $g_2(p)$, que são iguais a:

$$f_1(\beta) = \frac{1}{\beta} \quad \text{e} \quad (5.22)$$

$$g_2(p) = \sqrt{-\frac{\ln p + 1}{(p\ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2\ln p + 3}{2(1-p)^2\ln p}}. \quad (5.23)$$

Daí, a priori de Referência é dada por:

$$\pi_p(p, \beta) = \sqrt{-\frac{\ln p + 1}{(p\ln p)^2} + \frac{1}{(1-p)^2} + \frac{(1-4p)/p^2 - 2\ln p + 3}{2(1-p)^2\ln p}} \cdot \frac{1}{\beta}. \quad (5.24)$$

□

5.3 Análise utilizando dados simulados

Nesta seção foi realizada uma comparação entre as distribuições a priori MDIP & Gama e de Referência e as demais distribuições a priori: Jeffreys, Beta & Gama e Uniforme & Gama. Foram utilizados três conjuntos de 100 amostras de tamanhos $n = 20$, 50 e 70 geradas da distribuição Exponencial-Logarítmica. Os valores utilizados para os parâmetros foram $p = 0.5$ e $\beta = 2$. A cadeia para o MCMC foi executada com 15000 iterações, com período de *burn-in* de 5000 e convergência monitorada por seus *output*, critério de Geweke e plots de autocorrelação.

Como mostrado no capítulo anterior, recorreu-se ao algoritmo MCMC no *software* R para obter as marginais das distribuições a posteriori e as estimativas Bayesianas. Para encontrar os Estimadores de Máxima Verossimilhança (EMV), utilizou-se a função "maxNR" do R, que é baseada no método de Newton-Raphson.

Os plots do MCMC sugeriram que foi alcançada a convergência com taxa de aceitação em torno de 30-40%.

Nas Tabelas (5.1) e (5.2) observou-se que a priori de Referência não produz boas estimativas para os dois parâmetros p e β , independente de qual parâmetro é o de interesse.

Tabela 5.1: EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.

p	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	0.1515 (0.0348)	0.5757 (0.1339)	0.5069 (0.1267)	0.3560 (0.1116)
$n = 50$	0.2022 (0.0263)	0.5749 (0.1568)	0.5446 (0.1263)	0.4282 (0.1430)
$n = 70$	0.3129 (0.0503)	0.5768 (0.1823)	0.5784 (0.1175)	0.4307 (0.1338)
p	Ref_p	Ref_β	MDIP/Gama	
$n = 20$	0.3173 (0.1347)	0.6961 (0.1722)	0.5370 (0.1010)	
$n = 50$	0.3911 (0.1263)	0.6850 (0.1925)	0.5694 (0.1213)	
$n = 70$	0.4286 (0.1528)	0.6496 (0.2135)	0.5456 (0.1412)	

Tabela 5.2: EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.

β	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	1.0497 (0.2281)	2.1606 (0.6354)	1.9904 (0.5476)	1.7339 (0.4691)
$n = 50$	1.6784 (0.2356)	2.0093 (0.3779)	2.0534 (0.3789)	1.8499 (0.3668)
$n = 70$	1.9497 (0.2243)	2.0257 (0.3818)	2.0649 (0.2839)	1.8264 (0.2819)
β	Ref_p	Ref_β	MDIP/Gama	
$n = 20$	1.7366 (0.6214)	2.2223 (0.6024)	2.0650 (0.5279)	
$n = 50$	1.8071 (0.3484)	2.1218 (0.4399)	2.0580 (0.3514)	
$n = 70$	1.8541 (0.2892)	2.0884 (0.3728)	2.0241 (0.3312)	

Se p é o parâmetro de interesse, então suas estimativas estão abaixo dos verdadeiros valores (subestimadas); por outro lado, se β é o parâmetro de interesse, então as estimativas estão acima dos verdadeiros valores (sobrestimadas). Mesmo quando n aumenta, há pouca melhora. Observou-se ainda que a priori de Referência quando p é o parâmetro de interesse se aproxima bastante da priori de Jeffreys. Nos gráficos das densidades a posteriori a seguir será possível visualizar melhor essa semelhança.

A priori MDIP & Gama produz estimativas próximas dos verdadeiros valores para os dois parâmetros. Quando comparada às distribuições a priori Beta & Gama e Uniforme & Gama, que também obtiveram boas estimativas, observa-se que a priori de Zellner produz as menores variâncias. Portanto, conclui-se que a priori MDIP & Gama, a princípio, é uma boa priori a ser utilizada na análise Bayesiana da distribuição Exponencial-Logarítmica, assim como as distribuições a priori Beta & Gama e Uniforme & Gama, que também produzem bons resultados.

As figuras (5.1), (5.2) e (5.3) mostram as densidades marginais a posteriori utilizando as distribuições a priori comparadas neste capítulo. Considerou-se os dois parâmetros, p e β , com amostras de tamanho $n = 20, 50$ e 70 . Além disso, para obter as densidades marginais a posteriori, tomou-se apenas uma amostra de cada análise utilizando as diferentes distribuições a priori.

A figura (5.1) evidencia que as densidades marginais a posteriori resultantes das distribuições a priori de Jeffreys e de Referência (considerando p o parâmetro de interesse) se diferenciam totalmente das outras distribuições a priori considerando-se o parâmetro p . Observou-se que essas marginais se deslocam significativamente para perto do zero

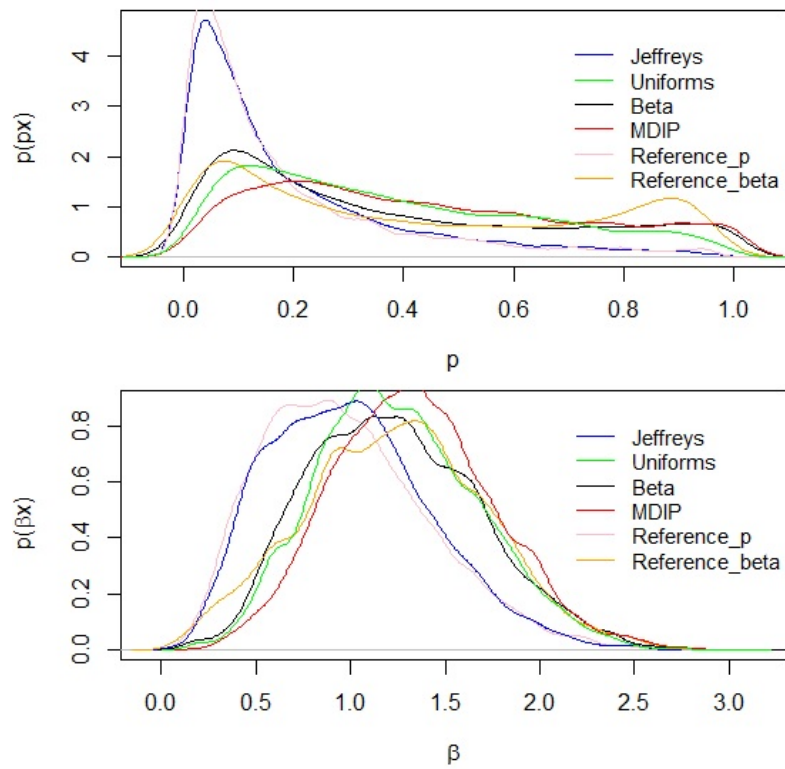


Figura 5.1: Gráficos das marginais a posteriori para $n = 20$

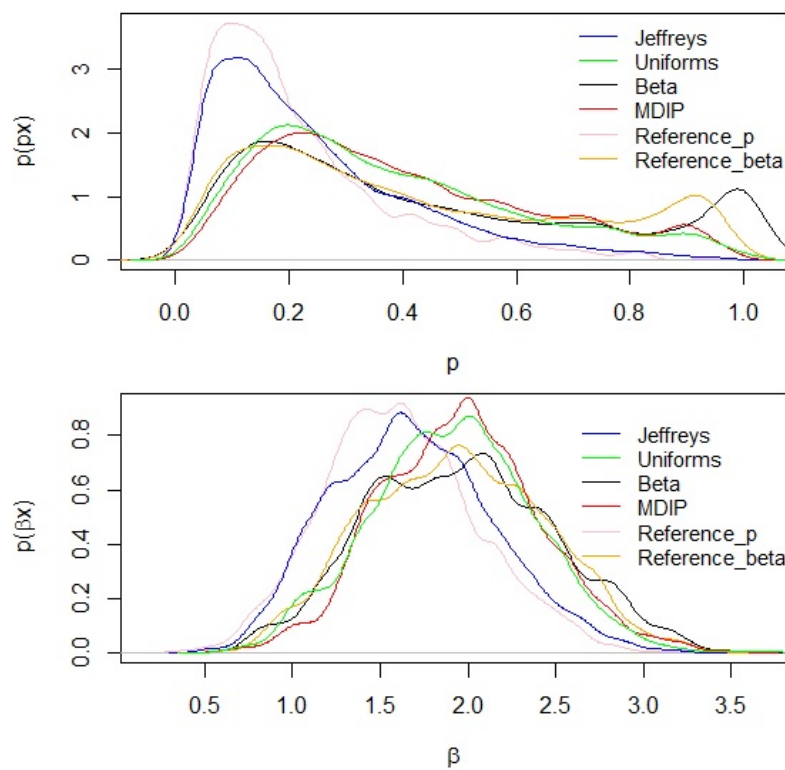


Figura 5.2: Gráficos das marginais a posteriori para $n = 50$

enquanto as demais possuem perfis mais *flat*, ou seja, com alta variabilidade. Outra pri-

ori bastante interessante é a priori de Referência quanto β é o parâmetro de interesse. Observou-se uma leve bimodalidade em sua densidade marginal considerando o parâmetro p da mesma forma como ocorre com a priori Beta & Gama. Quanto ao parâmetro β , notou-se que mesmo com n pequeno não há muita diferença entre os perfis das densidades marginais. À medida que n aumenta, como pode-se observar em (5.2) e (5.3), a diferença entre as distribuições a priori torna-se menor e as densidades marginais tendem a se uniformizar.

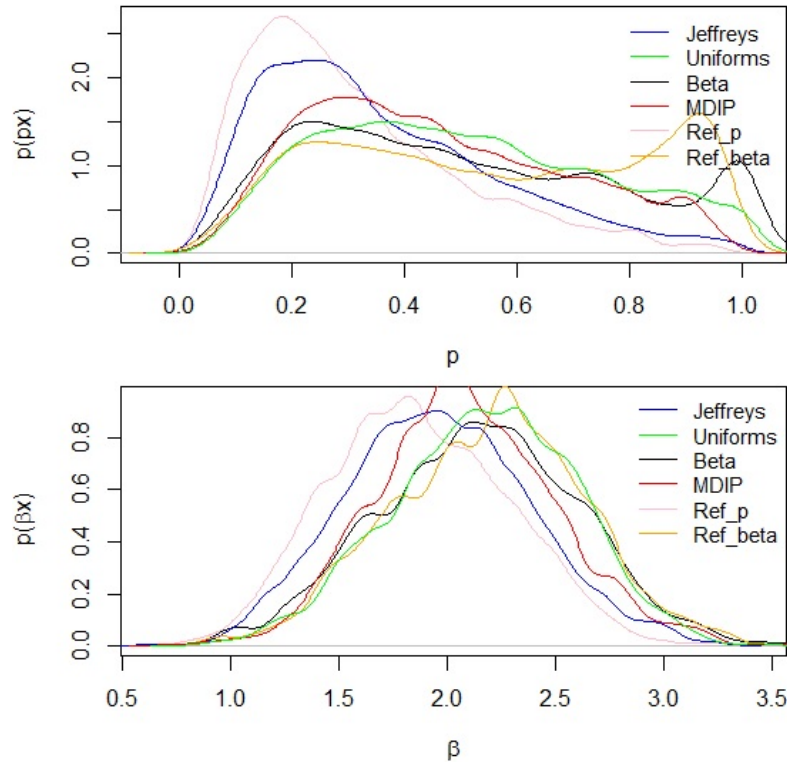


Figura 5.3: Gráficos das marginais a posteriori para $n = 70$

5.4 Análise com Conjunto de Dados Reais

Utilizando o mesmo conjunto de dados reais do Capítulo anterior, foram feitas análises com as distribuições a priori MDIP & Gama e de Referência e os resultados foram comparados com as estimativas produzidas pelas distribuições a priori de Jeffreys, Beta & Gama e Uniforme & Gama.

Observou-se que as estimativas referentes às distribuições a priori de Jeffreys e de Referência mantêm-se similares utilizando o conjunto de dados reais. Além disso, as distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama também produzem estimativas semelhantes.

Na Figura (5.4) nota-se claramente que as distribuições a priori de Jeffreys e de Referência considerando p o parâmetro de interesse mostram-se muito similares entre si e bastante divergentes em relação às outras distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama. Em contrapartida, as distribuições a priori não-informativas

Tabela 5.3: Resultados obtidos da posteriori para o parâmetro p (dados de Lawless).

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.3198	0.2915	(0.0152, 0.9779)
Uniforme/Gama	0.3467	0.2594	(0.0296, 0.9333)
Jeffreys	0.1633	0.1891	(0.0368, 0.7409)
MDIP/Gama	0.3580	0.2679	(0.1359, 0.9413)
Ref_p	0.1291	0.1637	(0.0283, 0.6466)

Tabela 5.4: Resultados obtidos da posteriori para o parâmetro β (dados de Lawless).

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.0483	0.0187	(0.0165, 0.0873)
Uniforme/Gama	0.0513	0.0177	(0.0195, 0.0879)
Jeffreys	0.0396	0.0182	(0.0259, 0.0796)
MDIP/Gama	0.0514	0.0182	(0.0389, 0.0916)
Ref_p	0.0366	0.0171	(0.0242, 0.0749)

Beta & Gama e Uniforme & Gama em conjunto com a priori MDIP & Gama mostram-se bastante *flat* dentro do intervalo $(0, 1)$. Observa-se, porém, que a bimodalidade da priori Beta & Gama é mantida, embora de forma bem sutil.

Quando considerado o parâmetro β , há diferença novamente entre as distribuições a priori de Jeffreys e de Referência e as demais distribuições a priori. As densidades marginais considerando as distribuições a priori de Jeffreys e de Referência estão deslocadas para perto do zero enquanto as demais mantêm-se semelhantes e bastante *flat*.

Para determinar a distribuição mais apropriada para o ajuste desse conjunto de dados, novamente foram calculados os valores para os critérios AIC, BIC e DIC. Foram analisadas as distribuições Weibull e Exponencial-Logarítmica, sendo que para os parâmetros da primeira distribuição foram utilizadas distribuições a priori Gama $(0.01, 0.01)$ para cada parâmetro e para a $EL(p, \beta)$ foram utilizadas as distribuições a priori Beta & Gama, Uniforme & Gama, Jeffreys, MDIP & Gama e de Referência.

Analisando a tabela (5.5) pode-se concluir que a Exponencial-Logarítmica proporciona um melhor ajuste para os dados, pois os valores para o AIC e BIC são pequenos em comparação com os valores para a Weibull. Considerando o critério DIC, as duas distribuições são apropriadas para ajustar os dados. No entanto, se é necessário escolher entre as duas distribuições através do DIC, então será escolhida a Exponencial-Logarítmica, pois possui

Tabela 5.5: Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.

	AIC	BIC	DIC
Weibull	231.9720	236.0553	140.830
Uniforme/Gama	141.8041	145.8874	139.167
Beta/Gama	141.08031	142.9692	139.1063
MDIP/Gama	141.1381	143.027	139.1575
Ref_p	140.2788	142.1677	139.8291

menor valor para ele. Entre as distribuições a priori da distribuição Exponencial-Logarítmica a priori Beta & Gama é que produz menores valores de um modo geral.

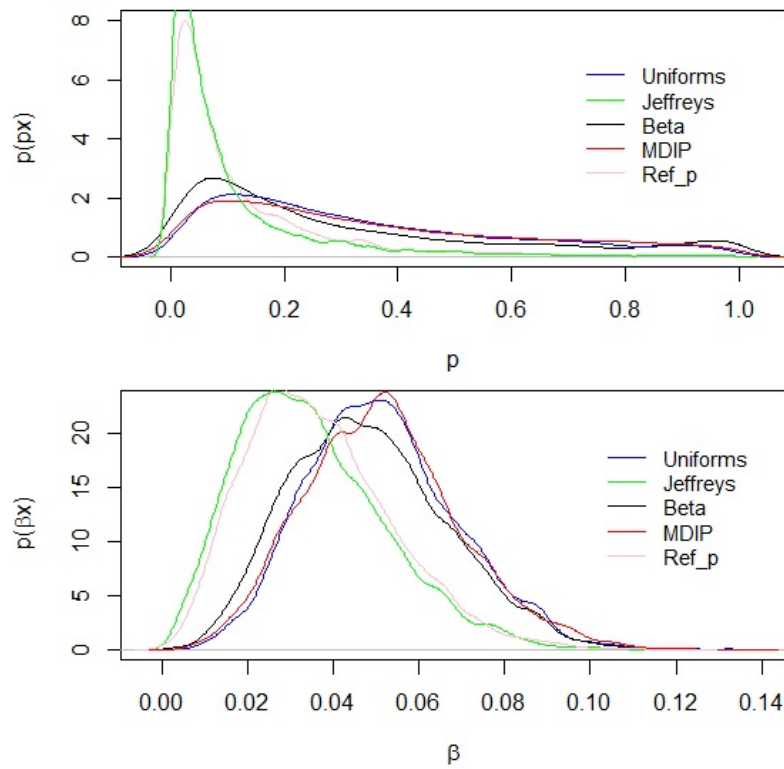


Figura 5.4: Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)

Uso de Cópulas como priori

6.1 Função Cópula

A Cópula é uma função que permite expressar uma distribuição conjunta como uma função de distribuições marginais. A necessidade de modelar estruturas de dependência complexas tem feito da função Cópula uma ferramenta muito utilizada em diversas áreas.

Sklar (1959) mostra que toda distribuição conjunta $F(x), x = (x_1, x_2, \dots, x_k)$ com distribuições marginais $F_i, i = 1, \dots, k$ pode ser escrita como

$$F(x) = C[F_1(x_1), F_2(x_2), \dots, F_k(x_k)],$$

onde C é a função cópula. Portanto, cópulas são funções que combinam funções distribuições marginais a fim de representar funções distribuições multivariadas.

Sklar (1959) ainda afirma que se cada F_i é contínua, então C é única. Além disso, a cópula é estritamente invariante sob transformações nas variáveis $\theta_1, \dots, \theta_k$. Outra vantagem é que distribuições condicionais também podem ser expressas utilizando a função cópula.

Uma cópula bidimensional é uma função $C : [0, 1]^2 \rightarrow [0, 1]$ com as seguintes propriedades:

1. Para todo $u \in [0, 1]$

$$C(0, u) = C(u, 0) = 0.$$

2. Para todo $u \in [0, 1]$

$$C(u, 1) = u \quad \text{e} \quad C(1, u) = u.$$

3. Para todo $(u_1, u_2), (v_1, v_2) \in [0, 1] \times [0, 1]$ com $u_1 \leq v_1$ e $u_2 \leq v_2$:

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0.$$

Para formular uma distribuição multivariada utilizando uma cópula é necessário basear-se na ideia de que uma simples transformação pode ser feita para cada variável marginal de maneira que cada variável marginal transformada tenha uma distribuição uniforme. Uma vez feito isso, a estrutura dependente pode ser expressada como uma distribuição multivariada composta por uniformes.

Assim, há muitas famílias de cópulas que diferem entre si quanto ao grau de dependência que elas representam.

No caso bivariado, sejam T_1 e T_2 duas amostras de variáveis com funções distribuições contínuas F_1 e F_2 . Seja $U = F_1(T_1)$ e $V = F_2(T_2)$, onde U e V possuem distribuições $U(0, 1)$, mas são usualmente dependentes se T_1 e T_2 são dependentes (T_1 e T_2 independentes implica que U e V são independentes). Assim, assumir dependência entre T_1 e T_2 é o mesmo que assumir dependência entre U e V .

Sendo U e V variáveis amostradas da distribuição uniforme, o problema se reduz em especificar uma distribuição bivariada entre duas uniformes, ou seja, uma cópula.

Existem diferentes famílias de cópulas na literatura. Em aplicações financeiras, a cópula mais utilizada é a Gaussiana. A cópula Gaussiana construída através da distribuição normal bivariada é dada por:

$$C_\rho(u, v) = \Phi_\rho[\Phi^{-1}(u), \Phi^{-1}(v)], \quad (6.1)$$

onde Φ_ρ é a função distribuição bivariada de uma distribuição normal padrão bivariada com coeficiente de correlação ρ , dada por:

$$\Phi_\rho(x, y) = \int_{-\infty}^x \int_{-\infty}^y \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z^2 - 2\rho zw + w^2)\right\} dz dw. \quad (6.2)$$

6.2 Função Cópula Farlie-Gumbel-Morgenstern

Diversas funções Cópula poderiam ser utilizadas para obter uma priori para a distribuição Exponencial-Logarítmica. Neste trabalho foi utilizada a Cópula proposta por Morgenstern (1956): Cópula Farlie-Gumbel-Morgenstern, que é dada por:

$$C(u, v) = [1 - e^{\ln(1-u)}][1 - e^{\ln(1-v)}][1 + \rho \exp\{\ln(1-u) + \ln(1-v)\}], \quad (6.3)$$

onde $u = F_1(t_1)$ é a distribuição marginal da variável aleatória T_1 e $v = F_2(t_2)$ é a distribuição marginal da variável aleatória T_2 .

Assim, simplificando a expressão tem-se:

$$C(u, v) = uv[1 + \rho(1 - u)(1 - v)], \quad (6.4)$$

em que $-1 \leq \rho \leq 1$.

Portanto, de (6.4), a função distribuição conjunta para T_1 e T_2 é dada por:

$$F(T_1, T_2 | \rho) = C(F_1(t_1), F_2(t_2)) = F_1(t_1)F_2(t_2)[1 + \rho(1 - F_1(t_1))(1 - F_2(t_2))], \quad (6.5)$$

em que $-1 \leq \rho \leq 1$.

Assim, a densidade conjunta de T_1 e T_2 é dada pela derivada segunda da função acumulada $F(T_1, T_2 | \rho)$ dada anteriormente:

$$f(T_1, T_2 | \rho) = \frac{\partial^2 F(T_1, T_2)}{\partial T_1 \partial T_2}. \quad (6.6)$$

Resultando em:

$$f(T_1, T_2 | \rho) = f_1(t_1)f_2(t_2) + \rho f_1(t_1)f_2(t_2)(1 - 2F_1(t_1))(1 - 2F_2(t_2)), \quad (6.7)$$

onde $f_1(t_1)$ e $f_2(t_2)$ são as densidades marginais de T_1 e T_2 .

6.3 Priori Cópula

Na análise Bayesiana, é possível assumir a função Cópula como uma priori conjunta com dois ou mais parâmetros. Ou seja, a priori conjunta Cópula contará com distribuições a priori individuais para cada parâmetro, além da variável ρ , que mede a correlação entre os parâmetros. Para a distribuição Exponencial-Logarítmica, serão utilizadas a distribuição $U(0, 1)$ para o parâmetro p e a distribuição Gama(0.01, 0.01) para β .

Portanto, de (6.7) podemos escrever:

$$\pi(p, \beta | \rho) = \frac{a}{\Gamma(b)}(a\beta)^{b-1}e^{-a\beta} + \rho \frac{a}{\Gamma(b)}(a\beta)^{b-1}e^{-a\beta}(1 - 2p) \left(1 - 2 \frac{\gamma(a, \beta b)}{\Gamma(a)} \right), \quad (6.8)$$

onde $\gamma(a, y) = \int_0^y t^{a-1}e^{-t}dt$.

Considerando $U(-1, 1)$ como a priori de ρ , temos que a priori Cópula é dada por:

Tabela 6.1: EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.

p	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	0.1515 (0.0348)	0.5757 (0.1339)	0.5069 (0.1267)	0.3560 (0.1116)
$n = 50$	0.2022 (0.0263)	0.5749 (0.1568)	0.5446 (0.1263)	0.4282 (0.1430)
$n = 70$	0.3129 (0.0503)	0.5768 (0.1823)	0.5784 (0.1175)	0.4307 (0.1338)
p	Ref_p	Ref_β	MDIP/Gama	Cópula
$n = 20$	0.3173 (0.1347)	0.6961 (0.1722)	0.5370 (0.1010)	0.5177 (0.1250)
$n = 50$	0.3911 (0.1263)	0.6850 (0.1925)	0.5694 (0.1213)	0.5584 (0.1289)
$n = 70$	0.4286 (0.1528)	0.6496 (0.2135)	0.5456 (0.1412)	0.5695 (0.1388)

Tabela 6.2: EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.

β	EMV	Beta/Gama	Uniform/Gama	Jeffreys
$n = 20$	1.0497 (0.2281)	2.1606 (0.6354)	1.9904 (0.5476)	1.7339 (0.4691)
$n = 50$	1.6784 (0.2356)	2.0093 (0.3779)	2.0534 (0.3789)	1.8499 (0.3668)
$n = 70$	1.9497 (0.2243)	2.0257 (0.3818)	2.0649 (0.2839)	1.8264 (0.2819)
β	Ref_p	Ref_β	MDIP/Gama	Cópula
$n = 20$	1.7366 (0.6214)	2.2223 (0.6024)	2.0650 (0.5279)	1.9854 (0.5649)
$n = 50$	1.8071 (0.3484)	2.1218 (0.4399)	2.0580 (0.3514)	2.0015 (0.3600)
$n = 70$	1.8541 (0.2892)	2.0884 (0.3728)	2.0241 (0.3312)	2.0818 (0.3245)

$$\pi(p, \beta, \rho) = \pi(p, \beta|\rho) \cdot \pi(\rho) =$$

$$\frac{a}{\Gamma(b)} (a\beta)^{b-1} e^{-a\beta} + \rho \frac{a}{\Gamma(b)} (a\beta)^{b-1} e^{-a\beta} (1-2\rho) \left(1 - 2 \frac{\gamma(a, \beta b)}{\Gamma(a)} \right) \cdot \frac{1}{2} \quad (6.9)$$

onde $\gamma(a, y) = \int_0^y t^{a-1} e^{-t} dt$.

6.4 Análise Utilizando Dados Simulados

Nesta seção foi realizada uma comparação entre a priori Cópula (utilizando a $U(-1, 1)$ como priori de ρ) e as demais distribuições a priori já anteriormente estudadas (Jeffreys, Beta & Gama e Uniforme & Gama, MDIP & Gama e de Referência). Foram simuladas 100 amostras de tamanhos $n = 20, 50$ e 70 . Além disso, a cadeia para o MCMC foi executada com 15000 iterações, com período de *burn-in* de 5000 e convergência monitorada por seus *output*, plots de autocorrelação e pelo critério de Geweke.

Utilizou-se o *software* R para executar os algoritmos MCMC a fim de obter as marginais das distribuições a posteriori e as estimativas Bayesianas.

Na Tabela (6.1) observa-se que a priori Cópula produz estimativas bem próximas dos verdadeiros valores para os dois parâmetros, inclusive para tamanhos de amostra pequenos. Comparada às outras distribuições a priori, pode-se concluir que os resultados da priori Cópula estão bem parecidos com os resultados das distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama.

Quanto ao desvio-padrão, nota-se que as estimativas se assemelham muito às estima-

Tabela 6.3: Estimativas (média) e desvio-padrão da correlação obtida através da priori Cópula.

	ρ
$n = 20$	-0.0117 (0.0904)
$n = 50$	-0.0409 (0.0913)
$n = 70$	-0.0479 (0.0984)

tivas das distribuições a priori Uniforme & Gama e MDIP & Gama. Além disso, para o parâmetro p os valores das estimativas pouco mudam com o aumento do tamanho da amostra; em contrapartida, para o parâmetro β as estimativas diminuem significativamente com o aumento de n .

O coeficiente que mede a dependência entre as variáveis p e β também foi estimado e seus valores podem ser observados na Tabela (6.3). Nota-se que as duas variáveis demonstram pouca dependência, o que justifica o fato das estimativas da priori Cópula serem tão próximas das estimativas da priori Uniforme & Gama. Se as variáveis p e β são independentes, a priori conjunta simplifica-se como o produto das distribuições a priori de cada parâmetro, no caso a Uniforme para p , e a Gama para β . Além disso, para n grande, tem-se um aumento no valor de ρ .

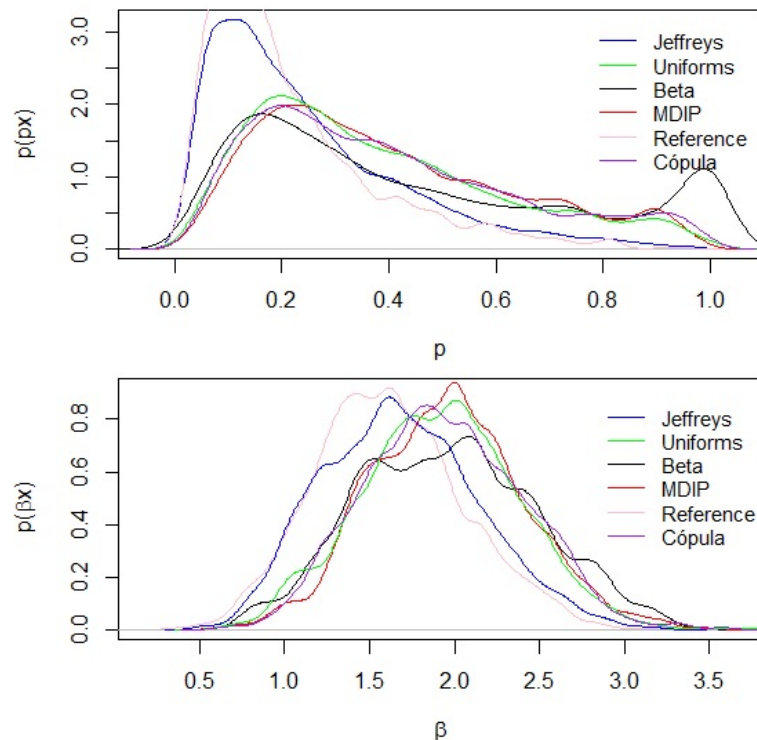


Figura 6.1: Marginais das densidades a posteriori dos parâmetros $p = 0.5$ e $\beta = 2$ com $n = 50$.

Neste capítulo as marginais foram obtidas apenas para $n = 50$ pois nos capítulos anteriores já chegou-se à conclusão de que os perfis tendem a se tornar similares com o aumento de n . Além disso, notou-se que a priori Cópula assemelha-se, para todos os

Tabela 6.4: Resultados obtidos da posteriori para o parâmetro p (dados de Lawless).

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.3198	0.2915	(0.0152, 0.9779)
Uniforme/Gama	0.3467	0.2594	(0.0296, 0.9333)
Jeffreys	0.1633	0.1891	(0.0368, 0.7409)
MDIP/Gama	0.3580	0.2679	(0.1359, 0.9413)
Ref_p	0.1291	0.1637	(0.0283, 0.6466)
Cópula	0.3566	0.2613	(0.1360, 0.9219)

Tabela 6.5: Resultados obtidos da posteriori para o parâmetro β (dados de Lawless).

Priori	Média	Desvio-Padrão	Intervalo 95%
Beta/Gama	0.0483	0.0187	(0.0165, 0.0873)
Uniforme/Gama	0.0513	0.0177	(0.0195, 0.0879)
Jeffreys	0.0396	0.0182	(0.0259, 0.0796)
MDIP/Gama	0.0514	0.0182	(0.0389, 0.0916)
Ref_p	0.0366	0.0171	(0.0242, 0.0749)
Cópula	0.0524	0.0180	(0.0398, 0.0911)

tamanhos da amostra, às distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama, que já haviam sido mais minuciosamente analisadas nos demais capítulos.

6.5 Análise com Conjunto de Dados Reais

Utilizando o conjunto de dados reais de Lawless, foram feitas análises utilizando a priori Cópula e os resultados foram comparados com as estimativas das outras distribuições a priori.

É importante lembrar que a cadeia obtida via MCMC foi executada no *software* R com 15000 iterações e convergência monitorada por seus *output*, plots de autocorrelação e pelo critério de Geweke. Dessa maneira obteve-se as marginais das distribuições a posteriori e as estimativas Bayesianas.

Nota-se, nas tabelas (6.4) e (6.5), que as estimativas das médias dos dois parâmetros e de seus respectivos desvios-padrão utilizando a priori Cópula são similares às médias produzidas pelas distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama.

O coeficiente ρ , que mede a correlação entre os parâmetros p e β , teve como estimativa $\rho = 0.131665$. Observa-se que considerando os dados reais a priori Cópula identificou melhor a dependência entre os parâmetros do que quando foram considerados os dados simulados.

Na Figura (6.2), comprova-se mais uma vez a semelhança da priori Cópula com as distribuições a priori Beta & Gama, Uniforme & Gama e MDIP & Gama. Observa-se que as distribuições a priori de Jeffreys e de Referência considerando p o parâmetro de interesse mantêm-se similares, como visto nos demais capítulos.

Analisando a tabela (6.6) observa-se que os valores do AIC, BIC e DIC utilizando a priori Cópula são semelhantes aos valores das outras distribuições a priori para a distribuição

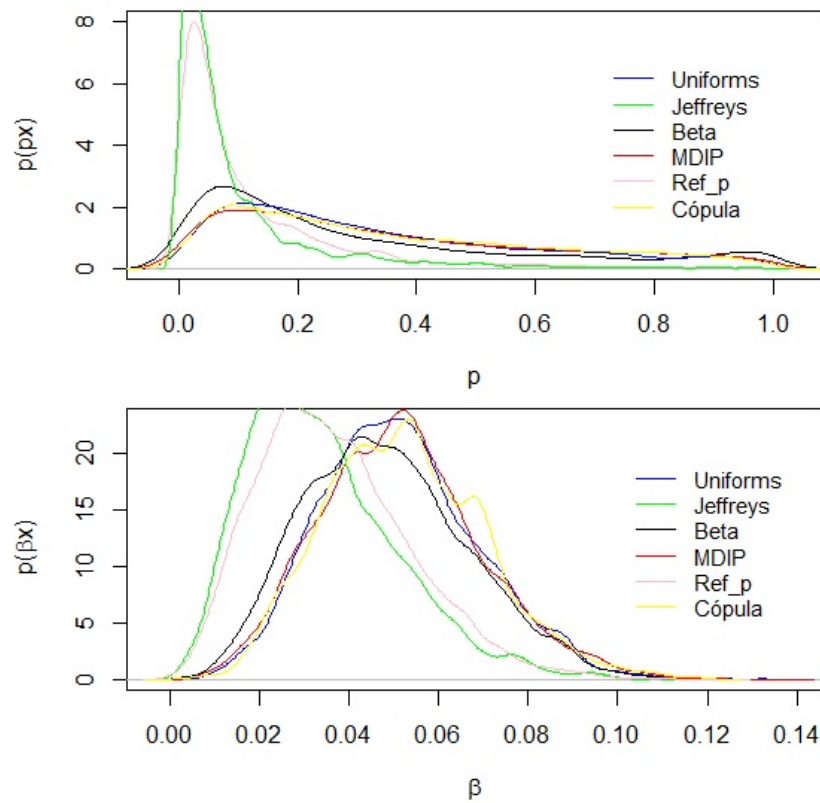


Figura 6.2: Marginais das densidades a posteriori dos parâmetros p e β (dados de Lawless)

Tabela 6.6: Informações baseadas nos critérios de seleção de modelo (AIC, BIC e DIC) para os dados de Lawless.

	AIC	BIC	DIC
Weibull	231.9720	236.0553	140.830
Uniforme/Gama	141.8041	145.8874	139.167
Beta/Gama	141.08031	142.9692	139.1063
MDIP/Gama	141.1381	143.027	139.1575
Ref_p	140.2788	142.1677	139.8291
Cópula	141.0903	142.9792	139.0331

Exponencial- Logarítmica.

Como mostrado anteriormente, pode-se concluir que a Exponencial-Logarítmica proporciona um melhor ajuste para os dados, pois os valores para o AIC e BIC são pequenos em comparação com os valores para a Weibull. Considerando o critério DIC, as duas distribuições são apropriadas para ajustar os dados. No entanto, se é necessário escolher entre as duas distribuições através do DIC, então será escolhida a Exponencial - Logarítmica, pois possui menor valor para ele. Entre as distribuições a priori da distribuição Exponencial-Logarítmica a priori Beta & Gama é que produz menores valores de um modo geral.

Priori Informativa: Elicitação

7.1 Introdução à Elicitação

A Elicitação de prioris tem um importante papel na análise Bayesiana. Em estudos envolvendo dados reais, utilizar a análise Bayesiana é o mais indicado principalmente quando se tem o conhecimento de especialistas. Além disso, é muito importante considerar e obter corretamente a opinião de especialistas a fim de complementar as informações obtidas pela amostra. A elicitação, por sua vez, é o processo de extração desse conhecimento. Com as informações obtidas através da elicitação, pode-se construir distribuições a priori.

Em qualquer análise estatística sempre haverá alguma forma disponível de conhecimento mais aprofundado sobre o assunto além dos dados obtidos pela amostra. Por exemplo, considere que o tempo de vida médio de um componente esteja sendo investigado. Espera-se que se realizem testes em amostras desses componentes a fim de estimar seus tempos de vida médios; porém, além dos estudos realizados nas amostras, é possível obter do projetista as suas expectativas pessoais sobre a vida do componente. Se o conhecimento do projetista a respeito das características dos componentes pode ser representado, então este conhecimento adicional (a priori) pode ser utilizado na inferência Bayesiana.

Quando se investiga fenômenos novos ou raros, os poucos dados podem não ser suficientes para se tomar uma decisão estatística. Como exemplos têm-se os modelos dos danos causados por terremotos, o estudo de doenças e espécies raras, etc. Nem sempre haverá dados suficientes para que se possa ignorar o conhecimento a priori, e conseqüentemente, a opinião do especialista pode ser uma das poucas fontes de informação.

Quando se trabalha com dados reais, além de poucos dados, pode-se encontrar também dados censurados, o que justifica ainda mais a utilização da elicitação. Tal contexto é encontrado principalmente na indústria, onde não é possível obter amostras muito grandes

devido ao encarecimento da pesquisa ou pela demora em ocorrer uma falha. Quando isso acontece, é possível acelerar o processo, o que pode levar a dados não condizentes com a realidade. Neste caso, todas as fontes de conhecimento como a opinião de especialistas devem ser levadas em conta. Vários autores têm buscado o estudo da elicitación da priori $\pi(\phi)$.

Na prática, numa tomada de decisão, a escolha de uma priori deve-se considerar o seu prático manuseio (recursos explícitos e fácil amostragem), que é de grande importância para estudos de sensibilidade. Definições estatísticas como matriz de Fisher inversa são, na maioria das vezes, demasiadamente técnicas e inacessíveis. Portanto, é impraticável esperar que um especialista seja capaz de estabelecer valores da densidade a priori. Sabe-se que não é fácil responder a perguntas como: "qual é a probabilidade da variável de interesse assumir tal valor?". O mais razoável seria pedir valores vinculados à probabilidades, como $P(X > x)$, ou seja, o mais conveniente é perguntar "qual o valor que a variável poderia assumir para uma certa probabilidade?". Além disso, estudos indicam que os momentos de segunda ordem não são recomendados para serem elicitados. A recomendação vem da dificuldade de especialistas com conhecimento em conceitos estatísticos (ou até mesmo estatísticos) fornecerem medidas de variabilidade precisas (como variância e desvio-padrão). Em geral, a variância é subestimada. A escolha mais comum é obter quantis, como a mediana ou os quartis, pois são mais fáceis de explicar aos especialistas. Da mesma forma, geralmente intervalos de credibilidade são bem entendidos e avaliados com bastante precisão, desde que não seja exigido do especialista o fornecimento de intervalos com alta probabilidade de cobertura (por exemplo 95% ou 97,5%). Além disso, a elicitación de quantis permite a avaliação adequada da variância da distribuição a priori do especialista, em vez de elicitá-la diretamente. O'Hagan (1998) dá um exemplo do uso de percentis elicitados para estimação da variância. Portanto, o melhor é estabelecer probabilidades de fácil entendimento para leigos em estatística e pedir a eles que proponham valores correspondentes à essas porcentagens. Desta maneira a informação fornecida pelos especialistas será obtida de forma mais fiel à realidade.

7.2 Elicitación da confiabilidade

No caso da Elicitación da confiabilidade $S(t)$, dada em (2.15), o especialista indica um valor condicionado a uma probabilidade especificada. Ou seja, dada uma probabilidade p , pergunta-se ao especialista qual o valor que ele acredita que $p\%$ dos produtos estão acima desse valor.

Portanto, k valores t_1, t_2, \dots, t_k são escolhidos e o especialista fornece suas probabilidades $S_{t_1}, S_{t_2}, \dots, S_{t_k}$ com

$$S_{t_j} = P(T \geq t_j), \quad j = 1, \dots, k, \quad (7.1)$$

que nada mais é do que a confiabilidade de S dado os parâmetros p e β .

A escolha das probabilidades que serão pedidas é uma tarefa muito importante. É importante levar em conta que as probabilidades devem ser de fácil entendimento para quem não possui conhecimentos em estatística. Neste caso, foram utilizados os percentis 10° , 25° , 50° e 75° .

Considerando as distribuições a priori $Beta(a, b)$ para o parâmetro p e $Gama(c, d)$ para β e supondo independência entre os parâmetros, a expressão em (7.1) pode ser definida como:

$$\begin{aligned} P(T \geq t|a, b, c, d) &= \int_0^\infty \int_0^1 P(T \geq t|a, b, c, d)\pi(p, \beta)dpd\beta \\ &= -\frac{2\pi\Gamma(a+b)d^c}{\Gamma(a)\Gamma(b)\Gamma(c)} \frac{\hat{\beta}^c(1-e^{-\hat{u}})^b}{\sqrt{(b-1)(c-1)}} \exp\left\{-\left[\left(a-\frac{1}{2}\right)\hat{u}+d\hat{\beta}\right]\right\} \frac{1}{\hat{u}} \ln[1-(1-e^{-\hat{u}})e^{-\hat{\beta}t}] \end{aligned} \quad (7.2)$$

onde $\hat{u} = \ln\left[\frac{a+b-1}{2}\right]$ e $\hat{\beta} = \frac{c-1}{d}$, de modo que $a+b > 1$ e $c > 1$.

Para demonstrar a equação em (7.2), seja $S(t|p, \beta)$ a função de sobrevivência dada em (2.15). Além disso, sejam $p \sim Beta(a, b)$ e $\beta \sim Gama(c, d)$, isto é,

$$\pi(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \quad \text{e} \quad \pi(\beta|c, d) = \frac{d}{\Gamma(c)} (d\beta)^{c-1} e^{-d\beta}.$$

Sabe-se que:

$$S(t) = \int_0^1 \int_0^\infty P(T > t|p, \beta)\pi(p, \beta)dpd\beta. \quad (7.3)$$

Supondo p e β independentes, tem-se que $\pi(p, \beta) = \pi(p)\pi(\beta)$, ou seja,

$$\pi(p, \beta) = \frac{\Gamma(a+b)d^c}{\Gamma(a)\Gamma(b)\Gamma(c)} p^{a-1}(1-p)^{b-1}\beta^{c-1}e^{-d\beta}. \quad (7.4)$$

Portanto, de (7.3) tem-se:

$$S(t) = \frac{\Gamma(a+b)d^c}{\Gamma(a)\Gamma(b)\Gamma(c)} \int_0^1 \int_0^\infty \frac{\ln[1-(1-p)e^{-\beta t}]}{\ln p} p^{a-1}(1-p)^{b-1}\beta^{c-1}e^{-d\beta} dpd\beta. \quad (7.5)$$

É fácil notar que a integral em (7.5) não possui solução analítica. Neste caso, a aproximação de Laplace torna-se muito útil pois é um método eficaz para aproximação de integrais complicadas. Veja mais sobre o Método de Laplace no Apêndice A2 deste trabalho.

Para que se possa utilizar o Método de Laplace em (7.5), é preciso fazer uma transformação em p para se obter limites 0 a ∞ . Seja então a transformação $u = -\ln p$ tal que $p \rightarrow 0$ implica em $u \rightarrow \infty$, $p \rightarrow 1$ implica em $u \rightarrow 0$, $p = e^{-u}$ e $dp = -e^{-u} du$.

Assim,

$$\begin{aligned} & \int_0^1 \frac{\ln[1 - (1-p)e^{-\beta t}]}{\ln p} p^{a-1} (1-p)^{b-1} dp \\ &= - \int_0^\infty \frac{\ln[1 - (1-e^{-u})e^{-\beta t}]}{u} e^{-(a-1)u} (1-e^{-u})^{b-1} e^{-u} du. \end{aligned}$$

Portanto,

$$\begin{aligned} & \int_0^1 \int_0^\infty \frac{\ln[1 - (1-p)e^{-\beta t}]}{\ln p} p^{a-1} (1-p)^{b-1} \beta^{c-1} e^{-d\beta} dp d\beta \\ &= - \int_0^\infty \int_0^\infty \frac{\ln[1 - (1-e^{-u})e^{-\beta t}]}{u} (1-e^{-u})^{b-1} e^{-au} \beta^{c-1} e^{-d\beta} du d\beta. \end{aligned}$$

Sejam $f(u, \beta) = -\frac{\ln[1 - (1-e^{-u})e^{-\beta t}]}{u}$ e $n = 1$ pois não há dados em X . Então,

$$\exp\{-h(u, \beta)\} = \exp\{-au + (b-1)\ln(1-e^{-u}) + (c-1)\ln\beta - d\beta\}$$

que implica em $h(u, \beta) = au - (b-1)\ln(1-e^{-u}) - (c-1)\ln\beta + d\beta$.

Calculando as derivadas parciais tem-se:

$$\frac{\partial h(u, \beta)}{\partial u} = a - (b-1)\frac{e^{-u}}{1-e^{-u}} \quad \text{e} \quad \frac{\partial h(u, \beta)}{\partial \beta} = d - \frac{(c-1)}{\beta}.$$

Sabendo que $\frac{\partial h(u, \beta)}{\partial u} = 0$ e $\frac{\partial h(u, \beta)}{\partial \beta} = 0$, então:

$$u = \ln\left[\frac{a+b-1}{a}\right] \quad \text{e} \quad \hat{\beta} = \frac{c-1}{d}.$$

Observa-se que $\hat{\beta}$ é a moda da distribuição Gama.

Agora é preciso calcular as derivadas segundas $\frac{\partial^2 h}{\partial u^2}$, $\frac{\partial^2 h}{\partial \beta^2}$ e $\frac{\partial^2 h}{\partial u \partial \beta}$ para obter a matriz Σ^{-1} , que é a Matriz Hessiana dada a seguir.

$$\Sigma^{-1} = \begin{bmatrix} \frac{\partial^2 h}{\partial u^2} & \frac{\partial^2 h}{\partial u \partial \beta} \\ \frac{\partial^2 h}{\partial u \partial \beta} & \frac{\partial^2 h}{\partial \beta^2} \end{bmatrix}. \quad (7.6)$$

Nota-se que $\det(\Sigma) = \frac{1}{\det(\Sigma^{-1})}$.

Calculando as derivadas segundas é possível obter:

$$\frac{\partial^2 h}{\partial u^2} = \frac{(b-1)e^{-u}}{(1-e^{-u})^2}, \quad \frac{\partial^2 h}{\partial u \partial \beta} = 0 \quad \text{e} \quad \frac{\partial^2 h}{\partial \beta^2} = \frac{c-1}{\beta^2}.$$

Portanto,

$$\Sigma^{-1} = \begin{bmatrix} \frac{(b-1)e^{-u}}{(1-e^{-u})^2} & 0 \\ 0 & \frac{c-1}{\beta^2} \end{bmatrix}. \quad (7.7)$$

Além disso,

$$\det(\Sigma) = \frac{1}{\det(\Sigma^{-1})} = \left(\frac{(b-1)(c-1)e^{-u}}{\beta^2(1-e^{-u})^2} \right)^{-1} = \frac{\beta^2(1-e^{-u})^2}{(b-1)(c-1)e^{-u}}.$$

Considerando $\theta = (a, b, c, d)$ e utilizando a aproximação de Laplace é possível obter:

$$S(t|\theta) = -\frac{\Gamma(a+b)d^c 2\pi}{\Gamma(a)\Gamma(b)\Gamma(c)} \hat{\beta}^{c-1} e^{-d\hat{\beta}} e^{-a\hat{u}} (1-e^{-\hat{u}})^{b-1} \frac{\hat{\beta}(1-e^{-\hat{u}})}{\sqrt{(b-1)(c-1)}} e^{\frac{\hat{u}}{2}} \frac{1}{\hat{u}} \ln[1-(1-e^{-\hat{u}})e^{-\hat{\beta}t}] \quad (7.8)$$

$$S(t|\theta) = -\frac{2\pi\Gamma(a+b)d^c}{\Gamma(a)\Gamma(b)\Gamma(c)} \frac{\hat{\beta}^c(1-e^{-\hat{u}})^b}{\sqrt{(b-1)(c-1)}} \exp\left\{-\left[\left(a-\frac{1}{2}\right)\hat{u} + d\hat{\beta}\right]\right\} \frac{1}{\hat{u}} \ln[1-(1-e^{-\hat{u}})e^{-\hat{\beta}t}] \quad (7.9)$$

onde $\hat{u} = \ln\left[\frac{a+b-1}{2}\right]$ e $\hat{\beta} = \frac{c-1}{d}$, de modo que $a+b > 1$ e $c > 1$.

Portanto, utilizando a equação (7.2) é possível encontrar valores para (a, b, c, d) através dos quatro valores (t_1, t_2, t_3, t_4) , correspondentes aos quatro percentis 10%, 25%, 50% e 75%, indicados pelo especialista. As quatro equações formam um sistema não-linear que será calculado no *software* R utilizando o pacote "nleqslv".

Procedimento para encontrar os hiperparâmetros a, b, c, d considerando o conhecimento de especialistas:

- Fixar diferentes probabilidades $S(t)$ e eliciar t .
- Se forem considerados desconhecidos os quatro hiperparâmetros a, b, c, d , será necessário quatro valores $S(t)$ dado quatro valores t .
- Pode-se considerar desconhecidos apenas "a" e "b", sendo c e d conhecidos. Assim, será preciso apenas dois diferentes valores para $S(t)$ dados t_1 e t_2 .

Simulação para a distribuição Exponencial- Logarítmica:

- Fixar valores para p e β .
- Considerar $S(t_1) = 0.90$, $S(t_2) = 0.75$, $S(t_3) = 0.50$ e $S(t_4) = 0.25$.
- Calcular os valores t_i , $i = 1, 2, 3, 4$ a partir de $S(t_i)$. Ou seja, dado $S(t_i) = S_i$, calcular:

$$t_i = \frac{1}{\beta} \ln \left[\frac{1-p}{1-p^{S_i}} \right].$$

É importante lembrar que na prática o especialista não terá 100% de certeza sobre os valores t_i . Porém, como um especialista, sabe-se também que seu erro ao apontar um valor será pequeno. Portanto, neste trabalho trabalhou-se com um erro de 10% em cima do verdadeiro valor de t_i encontrado através da equação (7.2).

- Haverá, portanto, um sistema não-linear composto por quatro equações e quatro incógnitas a, b, c, d , cujas equações são dadas por:

$$S(t|\theta) = -\frac{\sqrt{2\pi}\Gamma(a+b)d^c}{\Gamma(a)\Gamma(b)\Gamma(c)} \frac{\hat{\beta}^c(1-e^{-\hat{u}})^b}{\sqrt{(b-1)(c-1)}} \exp\left\{-\left[\left(a-\frac{1}{2}\right)\hat{u} + d\hat{\beta}\right]\right\} \cdot \frac{1}{\hat{u}} \ln[1 - (1 - e^{-\hat{u}})e^{-\hat{\beta}t}],$$

onde $\hat{u} = \ln\left[\frac{a+b-1}{2}\right]$ e $\hat{\beta} = \frac{c-1}{d}$, de modo que $a+b > 1$ e $c > 1$.

7.3 Análise utilizando Elicitação

Nesta seção foram comparadas as estimativas utilizando Elicitação com as estimativas das demais distribuições a priori já anteriormente estudadas (Jeffreys, Beta & Gama, Uniforme & Gama, MDIP & Gama, de Referência e Cópula). Para que se pudesse utilizar uma única tabela, foi necessário retirar das comparações a priori de Referência com β sendo o parâmetro de interesse. Isto não prejudicou o trabalho pois as estimativas utilizando esta priori não haviam sido satisfatórias e eram as mais ruins em comparação com as demais distribuições a priori.

Foram simuladas 100 amostras de tamanhos $n = 20, 50$ e 70 no *software* R. A cadeia obtida via MCMC foi executada no *software* R com 15000 iterações e convergência monitorada por seus *output*, plots de autocorrelação e pelo critério de Geweke. Dessa maneira obteve-se as marginais das distribuições a posteriori e as estimativas Bayesianas.

Além disso, o sistema não-linear dado em (7.9) foi resolvido e chegou-se aos seguintes valores dos hiperparâmetros (lembrando que foi inserido um erro de 10% em cima do verdadeiro valor de t_i): $a = 49.54704$, $b = 48.33053$, $c = 34.52137$ e $d = 18.44905$; ou seja, para a simulação foram utilizadas as distribuições a priori informativas $Beta(49.54704, 48.33053)$ para o parâmetro p e $Gama(34.52137, 18.44905)$ para o parâmetro β .

Nas Tabelas (7.1) e (7.2) observou-se que as estimativas utilizando as distribuições a priori informativas se aproximam muito dos verdadeiros valores, principalmente para o parâmetro p , além dos desvios-padrão serem muito menores em comparação com as outras distribuições a priori.

As distribuições a priori não-informativas Uniforme(0, 1) & Gama(0.01, 0.01), Cópula, Beta(0.5, 0.5) & Gama(0.01, 0.01) e MDIP & Gama também produzem estimativas próximas dos verdadeiros valores dos parâmetros, mesmo com desvios-padrão maiores que os das distribuições a priori informativas.

As distribuições a priori de Jeffreys e de Referência (considerando p o parâmetro de interesse) produzem estimativas abaixo dos verdadeiros valores para p e β e possuem formas bem parecidas das marginais a posteriori (ver (5.1)). O Estimador de Máxima Verossimilhança melhora muito pouco quando n aumenta.

Tabela 7.1: EMV, estimativas (média) Bayesianas e desvio padrão para $p = 0.5$.

p	EMV	Beta/Gama	Uniforme/Gama	Jeffreys
$n = 20$	0.1515 (0.0348)	0.5757 (0.1339)	0.5069 (0.1267)	0.3560 (0.1116)
$n = 50$	0.2022 (0.0263)	0.5749 (0.1568)	0.5446 (0.1263)	0.4282 (0.1430)
$n = 70$	0.3129 (0.0503)	0.5768 (0.1823)	0.5784 (0.1175)	0.4307 (0.1338)
p	Ref_p	MDIP/Gama	Cópula	Elicitação
$n = 20$	0.3173 (0.1347)	0.5370 (0.1010)	0.5177 (0.1250)	0.5047 (0.0058)
$n = 50$	0.3911 (0.1263)	0.5694 (0.1213)	0.5584 (0.1289)	0.5053 (0.0072)
$n = 70$	0.4286 (0.1528)	0.5456 (0.1412)	0.5695 (0.1388)	0.5043 (0.0072)

Tabela 7.2: EMV, estimativas (média) Bayesianas e desvio padrão para $\beta = 2$.

β	EMV	Beta/Gama	Uniform/Gama	Jeffreys
$n = 20$	1.0497 (0.2281)	2.1606 (0.6354)	1.9904 (0.5476)	1.7339 (0.4691)
$n = 50$	1.6784 (0.2356)	2.0093 (0.3779)	2.0534 (0.3789)	1.8499 (0.3668)
$n = 70$	1.9497 (0.2243)	2.0257 (0.3818)	2.0649 (0.2839)	1.8264 (0.2819)
β	Ref_p	MDIP/Gama	Cópula	Elicitação
$n = 20$	1.7366 (0.6214)	2.0650 (0.5279)	1.9854 (0.5649)	1.9331 (0.1499)
$n = 50$	1.8071 (0.3484)	2.0580 (0.3514)	2.0015 (0.3600)	1.9292 (0.1317)
$n = 70$	1.8541 (0.2892)	2.0241 (0.3312)	2.0818 (0.3245)	1.9624 (0.1495)

Portanto, mesmo inserindo um erro no verdadeiro valor de t_i , a Elicitação produz boas estimativas para os parâmetros da distribuição Exponencial-Logarítmica e se mostra uma ótima priori a ser utilizada.

Análise de Dados Censurados

8.1 Introdução

Na prática, os estudos que envolvem respostas dependentes do tempo demandam muito desgaste devido à longa duração dos testes. Além disso, nos estudos clínicos de sobrevivência é possível que não haja o término do teste com todas as unidades apresentando falha. Dados incompletos podem ser resultantes de testes de vida onde: (i) critérios de ordem prática ou econômica não permitiram realizar o teste até que todas as unidades falhassem, (ii) algumas unidades perderam-se, por um motivo ou outro, durante o teste, ou (iii) não foi possível registrar o exato momento de ocorrência da falha nas unidades, mas somente um intervalo de tempo que contém esse momento. Esse conjunto de dados incompletos de "tempos-até-falha" é dito "censurado". As circunstâncias que resultam em dados censurados permitem concluir que, em um teste de vida de componentes, a censura pode ou não ser planejada. É importante citar que os dados censurados não devem ser descartados. Isso se dá principalmente porque muitas vezes a amostra composta apenas por dados completos pode gerar conclusões viciadas. Além disso, deve-se considerar as informações sobre o tempo de vida dos pacientes até o momento da perda.

A censura é classificada da seguinte forma: censura à direita, censura à esquerda e censura intervalar. Em um conjunto de dados censurados à direita, existem uma ou mais unidades para as quais só se conhece o limite inferior para o tempo-até-falha. Suponha, por exemplo, um teste onde 20 máquinas são colocadas em uso durante 30 dias. Ao final do teste, 13 máquinas haviam falhado. Nesse caso, o conjunto de dados consiste de 13 tempos-até-falha e 7 observações censuradas à direita, cujas falhas ocorrem em algum momento após os 30 dias. Este é o tipo de censura mais frequente.

Em ensaios de confiabilidade, podem ocorrer três tipos de censura à direita: censura tipo I, censura tipo II e censura aleatória. Nas definições que se seguem, n designa o número total de unidades colocadas em teste e r o número de falhas observadas.

A censura tipo I é aquela em que se estabelece um tempo t_0 para o término do estudo, ou seja, após terminado o tempo determinado, o estudo é finalizado independente do evento ter ocorrido para todos os indivíduos. Assim, após o ensaio, apenas os tempos-até-falha das unidades que falharam antes de t_0 são conhecidos com exatidão. Em um ensaio com censura tipo I, obtém-se um conjunto de dados contendo r ($\leq n$) tempos-até-falha observados e $(n - r)$ tempos-até-falha censurados em t_0 . No número R de falhas observadas em um ensaio com censura tipo I corre-se o risco de que poucas ou nenhuma unidade falhe até o tempo t_0 , sendo esta a desvantagem deste tipo de censura.

Na censura do tipo II é estabelecido um número de falhas r e após a ocorrência da r -ésima falha, o teste é finalizado. Como o número total de falhas r é definido a priori, pode-se escolher um valor de r que garanta uma modelagem estatística satisfatória dos resultados do teste. Na censura tipo II, todas as unidades são ativadas em $t = 0$ e o conjunto de dados obtidos do teste consiste de r tempos-até-falha observados e $(n - r)$ tempos-até-falha censurados. O tempo $T_{(r)}$ de término do teste é aleatório; consequentemente, é impossível prever a sua duração total, sendo esta a desvantagem deste tipo de censura.

Na censura do tipo aleatória, as n unidades em teste são ativadas em momentos distintos e o teste é interrompido no tempo t_0 . Em contrapartida, quando todas as unidades são ativadas em $t = 0$, mas têm sua operação interrompida em momentos distintos no tempo, dá-se o nome também de censura aleatória. Em ambos os casos, os tempos de censura das unidades são aleatórios, podendo ser diferentes entre si ($R_i, i = 1, \dots, n$). Na área médica essa perda decorre de morte do paciente por motivo diferente do evento estudado, ou, por exemplo, quando o paciente deixa de ser acompanhado por não comparecer mais ao hospital. A censura do tipo aleatória é a que mais ocorre na prática médica. Outro exemplo está na área industrial quando se observa a utilização da garantia em produtos manufaturados por um período de tempo. Como os produtos são manufaturados em diferentes momentos ao longo do período de observação, ao interromper-se a observação ao término deste período, os tempos de censura dos produtos que não apresentaram falhas serão diferentes entre si, dependendo de sua data de fabricação. É importante notar que a censura do tipo I e a censura do tipo II são casos particulares da censura aleatória.

A censura à esquerda ocorre quando o tempo em que ocorreu o evento é desconhecido e ocorreu antes do início do estudo. Este tipo de censura é menos frequente que a censura à direita. Como exemplo tem-se um antropólogo que deseja verificar com que idade as crianças da localidade aprendem a realizar uma determinada tarefa. Neste caso, o tempo até "falha" é o tempo transcorrido entre o nascimento e o momento em que a criança desenvolve a habilidade de realizar a tarefa. Crianças já capazes de realizar a tarefa quando da chegada do antropólogo serão observações censuradas à esquerda; em contrapartida, crianças que não haviam aprendido a realizar a tarefa quando da partida do psiquiatra serão observações censuradas à direita.

Uma outra classificação de censura é a censura por intervalo. Neste caso, dados de

tempo-até-falha são agrupadas em intervalos. Esta censura ocorre, normalmente, em investigações onde não é possível determinar o momento da falha com precisão, já que o esquema de coleta dos dados não o permite. Um exemplo são as inspeções periódicas a determinados componentes onde apenas se conhece que o evento ocorreu entre uma inspeção e outra.

Existem pelo menos três abordagens para o tratamento de dados censurados, mas apenas uma delas é válida e prática. A primeira abordagem consiste em ignorar os valores censurados e realizar a análise apenas com os dados de falha observados. Ainda que esse curso de ação simplifique a análise em termos matemáticos, essa não é uma abordagem válida. Se utilizada, por exemplo, em um conjunto de dados censurados à direita, serão exatamente os maiores valores de tempo-até-falha (correspondentes aos valores censurados à direita) que serão excluídos da análise. Neste caso, a modelagem subestimarà o tempo-até-falha das unidades, já que justamente as melhores unidades foram excluídas. A segunda abordagem consiste em simplesmente aguardar até que todos os dados censurados à direita falhem. Ainda que válida estatisticamente, essa abordagem pode não ser prática devido ao tempo total demandado para o teste de vida. A terceira abordagem consiste em tratar os dados censurados probabilisticamente incluindo os valores censurados na função de verossimilhança utilizada para estimar os parâmetros da distribuição que modela os tempos-até-falha; este é o tratamento correto a ser dado à amostra censurada, sendo detalhado na próxima seção.

No restante deste capítulo, os tempos-até-falha das n unidades colocadas em teste serão considerados como estatisticamente independentes e identicamente distribuídos, seguindo uma função de densidade $f(t)$. Ou seja, dizer que os tempos-até-falha são identicamente distribuídos corresponde à suposição de que as unidades são idênticas, isto é, de um mesmo tipo e expostas a aproximadamente aos mesmos estresses ambientais e operacionais. A suposição de independência implica em unidades não afetadas pela operação ou falha de outras unidades. É importante observar que os tempos-até-falha, no caso de dados censurados, devem ser interpretados como tempos-até-falha potenciais, já que as falhas não foram observadas na prática.

8.2 Função de Verossimilhança para Dados Censurados

A função de verossimilhança para um conjunto de dados censurados pode ser escrita de várias formas. Neste trabalho utilizou-se a seguinte: sejam T_1, \dots, T_n variáveis aleatórias que seguem uma distribuição de probabilidade $f(t, \theta)$, onde θ é um parâmetro desconhecido. Nos desenvolvimentos a seguir, os r tempos-até-falha observados são designados por t_1, \dots, t_r , e os dados censurados são designados por t_1^+, \dots, t_{n-r}^+ . No caso de censura à direita do tipo I, $t_1^+ = \dots = t_{n-r}^+ = t_0$. Um total de n unidades foram colocadas em teste.

Considere inicialmente um conjunto completo de dados (isto é, sem censura, tal que $r = n$). A função de verossimilhança associada à amostra é dada por:

$$L(\theta) = \prod_{i=1}^n f(t_i, \theta). \quad (8.1)$$

A função na equação (8.1) informa sobre a possibilidade das variáveis T_1, \dots, T_n assumirem os valores t_1, \dots, t_n ; tal possibilidade é dada pelo valor da função de densidade.

No caso de um conjunto de dados contendo dados censurados, a equação (8.1) é assim reescrita para incluir a porção censurada da amostra:

$$L(\theta) = \prod_{i=1}^r f(t_i, \theta) \prod_{i=1}^{n-r} R(t_i^+, \theta) \quad (8.2)$$

onde $R(t_i^+, \theta)$ é a função de confiabilidade com parâmetro θ avaliado no tempo censurado t_i^+ . A justificativa para a utilização da função de confiabilidade na função de verossimilhança para uma observação censurada à direita é que $R(t_i^+, \theta)$ correspondente à probabilidade da unidade i sobreviver a um tempo t_i^+ .

8.3 Amostra simulada

Para simular uma amostra contendo dados censurados deve-se levar em conta a proporção (p) de dados censurados que a amostra terá e o limite inferior (l_i) do intervalo em que está contida a distribuição estudada. Para facilitar o entendimento, um algoritmo foi esboçado:

1. Simular uma amostra $x = (x_1, x_2, \dots, x_n)$ de tamanho n da distribuição de interesse.
2. Determinar a proporção p de dados censurados.
3. A cada $1/p$ da amostra, faça x_i igual a um número simulado da distribuição Uniforme no intervalo (l_i, x_i) . Por exemplo: se $p = 0.2$, $n = 20$ e $l_i = 0$, tem-se $1/p = 5$, ou seja, se $x_5 = 3.5$ (agora ele será um dado censurado), então: $x_5 = Unif(0, 3.5)$. Da mesma forma faça com x_{10} , x_{15} e x_{20} .

8.4 Máxima Verossimilhança utilizando dados censurados

Considere $x = x_1, x_2, \dots, x_n$ um conjunto de dados censurados do tipo II. Portanto:

$$\delta_i = \begin{cases} 1, & \text{se } x_i \text{ for censurado,} \\ 0, & \text{se } x_i \text{ for não-censurado.} \end{cases}$$

Sabendo que $S(x_i; \theta)$ é a função de sobrevivência para um parâmetro θ da distribuição de interesse, tem-se que a função de verossimilhança é expressa por:

$$L(\theta) = \prod_{i=1}^n [f(x_i; \theta)]^{\delta_i} [S(x_i; \theta)]^{1-\delta_i}, \quad (8.3)$$

a qual também pode ser escrita como:

$$L(\theta) = \prod_{i=1}^n [h(x_i; \theta)]^{\delta_i} S(x_i; \theta), \quad (8.4)$$

onde $h(x_i; \theta)$ é a função de risco.

Portanto, utilizando as equações dadas em (1.1) e (2.15), a função de verossimilhança com dados censurados para a distribuição Exponencial-Logarítmica é dada por:

$$\begin{aligned} L(p, \beta) &= \prod_{i=1}^n [f(x_i; p, \beta)]^{\delta_i} [S(x_i; p, \beta)]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[\left(-\frac{1}{\ln(p)} \right) \left(\frac{\beta(1-p)e^{-\beta x_i}}{1 - (1-p)e^{-\beta x_i}} \right) \right]^{\delta_i} \left[\frac{\ln(1 - (1-p)e^{-\beta x_i})}{\ln p} \right]^{1-\delta_i}. \end{aligned} \quad (8.5)$$

8.5 Análise de Dados Censurados para a $EL(p, \beta)$

Foram simuladas 100 amostras de dados censurados tipo II de tamanhos $n = 20$, 50 e 70 e utilizadas as prioris Beta & Gama, Uniforme & Gama, Cópula e as prioris informativas obtidas através da Elicitação (com erro de 10% aplicado ao verdadeiro valor de t_i) para encontrar as estimativas dadas nas Tabelas (8.1) e (8.2).

A cadeia para o MCMC foi executada com 15000 iterações, com período de *burn-in* de 5000 e convergência monitorada por seus *output*, plots de autocorrelação e pelo critério de Geweke. Utilizou-se o *software* R para executar os algoritmos MCMC a fim de obter as marginais das distribuições a posteriori e as estimativas Bayesianas.

Na Tabela (8.1) foram utilizados 10% dos dados censurados. Já na Tabela (8.2), utilizou-se 20%.

Pode-se observar nas Tabelas que as estimativas dos parâmetros p e β utilizando amostras com 10% dos dados censurados são melhores do que com 20% dos dados censurados. Isso comprova que quanto maior o número de dados censurados, menos eficiente será a estimação dos parâmetros. Além disso, as prioris Beta & Gama e Uniforme & Gama produzem estimativas muito ruins comparadas com a priori Cópula e quando se utiliza a Elicitação.

Os desvios padrão produzidos pela Elicitação são bem menores que os produzidos pelas outras prioris, tanto para o parâmetro p quanto para β .

Além disso, nota-se que com o aumento de n as estimativas continuam ruins ou até mesmo pioram. Deve-se lembrar que a quantidade de dados censurados é proporcional ao

Tabela 8.1: Estimativas (média) Bayesianas e desvio padrão para $p = 0.5$, $\beta = 2$ e 10% dos dados censurados.

p	Beta/Gama	Uniforme/Gama	Cópula	Elicitação
$n = 20$	0.3124 (0.0863)	0.3386 (0.0738)	0.5309 (0.1075)	0.4975 (0.0042)
$n = 50$	0.2997 (0.0578)	0.3091 (0.0557)	0.4983 (0.1309)	0.4853 (0.0055)
$n = 70$	0.2861 (0.0527)	0.2964 (0.0533)	0.4983 (0.1473)	0.4767 (0.0071)
β	Beta/Gama	Uniforme/Gama	Cópula	Elicitação
$n = 20$	2.0138 (0.2281)	2.0547 (0.5118)	2.1785 (0.7456)	1.9372 (0.1343)
$n = 50$	2.1579 (0.7108)	2.1775 (0.7271)	2.0904 (0.3719)	2.0034 (0.1771)
$n = 70$	2.1309 (0.5237)	2.1194 (0.5032)	2.1074 (0.3629)	1.9948 (0.2056)

Tabela 8.2: Estimativas (média) Bayesianas e desvio padrão para $p = 0.5$, $\beta = 2$ e 20% dos dados censurados.

p	Beta/Gama	Uniforme/Gama	Cópula	Elicitação
$n = 20$	0.2732 (0.0557)	0.2965 (0.0529)	0.4577 (0.1367)	0.4884 (0.0048)
$n = 50$	0.2562 (0.0408)	0.2685 (0.0375)	0.4409 (0.1468)	0.4667 (0.0066)
$n = 70$	0.2385 (0.0364)	0.2456 (0.0363)	0.4430 (0.1568)	0.4536 (0.0073)
β	Beta/Gama	Uniforme/Gama	Cópula	Elicitação
$n = 20$	1.9363 (0.5394)	1.9737 (0.5923)	2.1591 (0.5846)	1.9480 (0.1547)
$n = 50$	2.0752 (0.5538)	2.1295 (0.6276)	2.1355 (0.4522)	2.0125 (0.2227)
$n = 70$	2.0596 (0.4994)	2.0788 (0.4957)	2.1554 (0.3858)	2.0133 (0.2641)

tamanho da amostra. Portanto, quanto maior a amostra, mais serão os dados censurados, ou seja, as estimativas tendem a piorar.

Conclusões

O uso da distribuição Exponencial-Logarítmica pode ser uma boa alternativa para analisar dados de sobrevivência devido às funções de sobrevivência e de risco apresentarem formas analíticas fechadas. Isso implica em facilidade analítica e vantagens computacionais para as análises de inferência.

Neste trabalho pôde-se mostrar através de um estudo de simulação que o método de máxima verossimilhança não garante a existência de um valor máximo, a menos que haja condições restritivas. Além disso, é necessária uma amostra de tamanho moderadamente grande ($n = 100$) para que se atinja uma estimação com uma precisão desejável. Por outro lado, análises Bayesianas baseadas em diferentes prioris possibilitam estimativas pontuais e intervalares dos parâmetros mesmo para conjuntos de dados pequenos.

Observou-se que a análise Bayesiana utilizando as prioris não-informativas torna-se muito importante especialmente nas situações em que não se tem opinião de especialistas para construir uma distribuição a priori.

Dentre as prioris não-informativas, a Uniforme parece ser a mais adequada para estimar o parâmetro p já que a priori de Jeffreys não dá boas estimativas e a priori Beta pode levar a uma densidade a posteriori bimodal.

A priori de Referência proposta por Berger and Bernardo (1992), não apresentou boas estimativas independente da escolha do parâmetro de interesse, como foi visto no capítulo 6.

Mostrou-se que a priori MDIP & Gama maximiza a informação proveniente dos dados e produz ótimas estimativas. Isto mostra que a informação na função de verossimilhança é enfatizada pela priori. Uma característica fundamental do procedimento proposto por Zellner é que a priori resultante para o parâmetro p é própria.

Notou-se que o uso de distribuições a priori formada por cópulas para p e β pode facilitar a análise de inferência. A priori gerada pela Cópula Farlie-Gumbel-Morgenstern tem que ser própria, o que é importante do ponto de vista prático. Sua flexibilidade e

tratabilidade analítica sugerem que é um método interessante para representar dependência entre parâmetros. Além disso, as condições refletindo informações a priori sobre as distribuições marginais dos parâmetros podem ser facilmente utilizadas para obtenção da priori conjunta.

Por outro lado, em aplicações médicas ou em engenharia que apresentem dados censurados, o uso de métodos de elicitación da priori pode ter inferências mais precisas, como vistos neste trabalho. Para se obter prioris informativas para os parâmetros p e β considerou-se o par de distribuições a priori Beta & Gama cujos hiperparâmetros foram obtidos via elicitación da função de confiabilidade do modelo Exponencial-Logarítmica usando a Aproximação de Laplace proposta por Kass et al. (1990).

Finalmente, a utilização dos métodos MCMC (Markov Chain Monte Carlo) facilitaram muito os cálculos computacionais.

Propostas Futuras

1. Teminar um 2^o artigo sobre o uso das prioris não-informativas estudadas aqui para estimação dos parâmetros da distribuição Exponencial-Logarítmica.
2. Escrever um 3^o artigo sobre o uso da elicitación para melhor estimação dos parâmeros considerando dados completos e censurados.

Referências Bibliográficas

- Achcar, J. A., Moala, F. A., and Boleta, J. (2010). Generalized exponential distribution: a bayesian approach using mcmc methods. *Submitted*.
- Berger, J. and Bernardo, J. (1992). *On the development of the reference prior method*. Spain: Fourth Valencia International Meeting on Bayesian Statistics.
- Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal Royal Statistical Society*, 41(2):113–147.
- Bernardo, J. M. (1991). *On the Development of the Reference Prior Method. Fourth Valencia International Meeting on Bayesian Statistics*. Spain.
- Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Addison Wesley.
- Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hasting algorithm. *The American Statistician*, 49.
- Colosimo, E. and Giolo, S. (2006). *Análise de Sobrevivência Aplicada*. Brasil: Editora Edgard Blucher.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*. Oxford,UK: Clarendon Press.
- Jeffreys, S. H. (1967). *Theory of probability*. London: Oxford U. Press.
- Kass, R., Tierney, L., and Kadane, J. (1990). The validity of posterior expansions based on laplace’s method. *S. Geisser, J.S. Hodges, S.J. Press, and A. Zellner, Amsterdam: North Holland*, pages 473–488.
- Moala, F. A. and Garcia, L. M. (2013). A bayesian analysis for the parameters of the exponential-logarithmic distribution. *Quality Engineering*, 25:282–291.
- Mood, A., Graybill, F., and Boes, D. (1974). *Introduction to the theory of Statistics*. New York: McGraw Hill.

- Morgensten, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt fur Mathematische Statistik* 8, pages 234–235.
- O’Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Statistician* 47: (1), pages 21–35.
- Raftery, A. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, 7:493–497.
- Rodrigues, J., Balakrishnan, N., Cordeiro, G., and Castro, M. (2011). A unified view on lifetime distributions arising from selection mechanisms. *Computational Statistics and Data Analysis* 55, pages 3311–3319.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *L’Institut de Statistique de L’Université de Paris*, pages 229–231.
- Tahmasbi, R. and Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, 52:3889–3901.
- Tierney, L. and Kadane, J. (1986). Accurate approximations for the posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–6.
- Zellner, A. (1977). *Maximal Data Information Prior Distributions in new methods in the applications of Bayesian Methods*. North-Holland, Amsterdam: A. Aykac e C. Brumat.
- Zellner, A. (1984). *Maximal Data Information Prior Distributions*. U. of Chicago Press: Basic Issues in Econometrics North-Holland.
- Zellner, A. (1990). Bayesian methods and entropy in economics and econometrics. *Kluwer Academic Publishers*, pages 17–31.



Apêndice

A.1 Demonstrações

Outra forma de obter a densidade em (2.3) é dada a seguir.

Seja Z uma variável discreta sob distribuição Logarítmica com função densidade de probabilidade dada por:

$$p_n = P\{Z = z\} = \frac{(1-p)^z}{-z \ln p}, \quad z = 1, 2, \dots \quad (\text{A.1})$$

e $p_0 = P\{Z = 0\} = 0$ com parâmetro $p \in (0, 1)$.

Suponha que a variável aleatória X tenha distribuição Exponencial com função densidade de probabilidade dada por:

$$g(x) = \beta e^{-\beta x}, \quad x > 0 \quad \text{e} \quad \beta > 0, \quad (\text{A.2})$$

e função de sobrevivência

$$S(x) = e^{-\beta x}. \quad (\text{A.3})$$

A função geradora de probabilidade da v.a. Logarítmica é dada por:

$$A_Z(s) = E(s^Z) = \frac{\ln(1-sq)}{\ln(1-q)}, \quad q = 1-p. \quad (\text{A.4})$$

Então,

$$\frac{d}{ds} A_Z(s) = -\frac{q}{\ln(1-q)} \left(\frac{1}{1-sq} \right) \quad (\text{A.5})$$

e portanto,

$$\frac{d}{ds}A_Z(s)|_{s=S(x)} = -\frac{1-p}{\ln(p)} \left(\frac{1}{1-(1-p)e^{-\beta x}} \right). \quad (\text{A.6})$$

De Rodrigues et al. (2011) tem-se que

$$f(x) = \frac{g(x)}{1-p_0} \left\{ \frac{d}{ds}A_Z(s)|_{s=S(x)} \right\} \quad (\text{A.7})$$

que resulta em

$$f(x) = -\beta e^{-\beta x} \frac{1-p}{\ln(p)} \left(\frac{1}{1-(1-p)e^{-\beta x}} \right) \quad (\text{A.8})$$

isto é, a função densidade de probabilidade de X é

$$f(x|p, \beta) = -\left(\frac{1}{\ln(p)} \right) \frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}}, \quad x > 0, \quad (\text{A.9})$$

dependente dos parâmetros $p \in (0, 1)$ e $\beta > 0$.

Outra forma de demonstrar a esperança em (2.13) é dada a seguir.

Seja a definição de Esperança dada em (A.10):

$$E(X) = \int_0^\infty x \left(-\frac{1}{\ln p} \right) \frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}} dx. \quad (\text{A.10})$$

Tomando $u = 1 - (1-p)e^{-\beta x}$, tem-se $du = \beta(1-p)e^{-\beta x} dx$.

De $1 - (1-p)e^{-\beta x} = u$ obtém-se $\frac{1-u}{1-p} = e^{-\beta x}$. Assim, $\ln \left(\frac{1-u}{1-p} \right) = -\beta x$. Portanto,

$$x = -\frac{1}{\beta} \ln \left(\frac{1-u}{1-p} \right).$$

Se $x = 0$ então $u = p$. Da mesma forma, se $x = \infty$, então $u = 1$.

Portanto,

$$E(X) = -\frac{1}{\ln p} \int_p^1 \left[-\frac{1}{\beta} \ln \left(\frac{1-u}{1-p} \right) \right] \frac{1}{u} du = \frac{1}{\beta \ln p} \int_p^1 \frac{1}{u} \ln \left(\frac{1-u}{1-p} \right) du =$$

$$\frac{1}{\beta \ln p} \left(\int_p^1 \frac{\ln(1-u)}{u} du - \ln(1-p) \int_p^1 \frac{du}{u} \right) =$$

$$= \frac{1}{\beta \ln p} \left(\int_p^1 \frac{\ln(1-u)}{u} du - \ln(1-p)(-\ln p) \right)$$

Sabendo que $\text{polylog}(s+1, z) = \int_0^z \frac{L_i(s, t)}{t}$ e $\text{polylog}(2, z) = -\int_0^z \frac{\ln(1-t)}{t} dt$,

$$\int_p^1 \frac{\ln(1-u)}{u} du = \int_0^1 \frac{\ln(1-u)}{u} du - \int_0^p \frac{\ln(1-u)}{u} du = -\text{polylog}(2, 1) + \text{polylog}(2, p) = -\frac{\pi^2}{6} + \text{polylog}(2, p).$$

Portanto,

$$E(X) = \frac{1}{\beta \ln p} \left[-\frac{\pi^2}{6} + \text{polylog}(2, p) + \ln(1-p) \ln p \right].$$

De $\text{polylog}(2, u) + \text{polylog}(2, 1-u) = \frac{\pi^2}{6} - \ln(u) \ln(1-u)$, obtém-se

$$E(X) = \frac{1}{\beta \ln p} [\text{polylog}(2, p) - \text{polylog}(2, p) - \text{polylog}(2, 1-p)] = -\frac{\text{polylog}(2, 1-p)}{\beta \ln p}.$$

A.2 Aproximação de Laplace

É possível encontrar, com frequência, situações em que não é possível resolver integrais analiticamente, isto é, não há como determinar uma fórmula fechada para a integral. Uma possibilidade para se contornar este problema é utilizar métodos numéricos de aproximação. Porém, na Inferência Bayesiana necessita-se 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 and Kadane (1986)).

O Método de Laplace para Aproximação de Integrais é uma técnica muito utilizada na Inferência Bayesiana para o cálculo de posteriori marginais, momentos a posteriori e densidades preditivas. É baseado na expansão de Taylor de 2ª ordem aplicada ao termo exponencial do integrando. Nesta seção será dada uma breve e informal descrição do método e suas propriedades. Para mais detalhes, ver Kass et al. (1990).

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

$$I = \int_0^\infty f(\phi) \exp\{-nh(\phi)\} d\phi \tag{A.11}$$

expandindo h e f sobre $\hat{\phi}$.

A função $f(\hat{\phi})$ é escolhida de modo que $\hat{\phi}$ seja obtido de forma explícita.

Para o caso em que ϕ é um parâmetro unidimensional obtém-se:

$$\int_0^\infty f(\phi) \exp\{-nh(\phi)\} d\phi \cong \left(\frac{2\pi\sigma^2}{n} \right)^{1/2} f(\hat{\phi}) \exp\{-nh(\hat{\phi})\} \tag{A.12}$$

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

Kass et al. (1990) argumentam que:

$$I = \hat{I} \cdot 1 + O(n^{-1}) \quad (\text{A.13})$$

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,

$$\int_0^\infty f(\phi) \exp\{-nh(\phi)\} d\phi \cong \left(\frac{2\pi}{n}\right)^{m/2} \det(\Sigma)^{1/2} f(\hat{\phi}) \exp\{-nh(\hat{\phi})\} \quad (\text{A.14})$$

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

A.3 Critério de Geweke

O critério de Geweke é baseado em um teste composto por duas médias de duas partes da cadeia de Markov: geralmente os primeiros 10% e os últimos 50%. Se as partes (amostras) pertencem à uma distribuição estacionária da cadeia, as médias são iguais e as estatísticas de Geweke possuem uma distribuição normal assintótica.

Geweke (1992) define a estatística do teste como sendo a diferença entre as duas médias dividida pelo erro padrão estimado. A estatística do teste neste trabalho foi encontrada através do *software* R, utilizando o pacote "coda" e o comando "geweke.diag(p)".