UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Matemática Aplicada e Computacional

Inferência Bayesiana Não-Paramétrica para Elicitação da Função de Confiabilidade

Débora Luzia Penha

2014

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS E TECNOLOGIA

DÉBORA LUZIA PENHA

Inferência Bayesiana Não-Paramétrica para Elicitação da Função de Confiabilidade

Dissertação apresentada ao Programa de Pós-Graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP para obtenção do título de Mestre em Matemática Aplicada e Computacional.
Orientador: Profº Drº Fernando Antônio Moala

Presidente Prudente
2014

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Presidente Prudente

BANCA EXAMINADORA

PROF. DR. FERNANDO ANTONIO MOALA
ORIENTADOR

PROF. DR. MANOEL IVANILDO SILVESTRE BEZERRA
UNESP/FCT

PROF. DR. FABIO NOGUEIRA DEMARQUI
UFMG

Débora Borges Penha
DÉBORA LUZIA PENHA

Presidente Prudente (SP), 20 de janeiro de 2014.

Resultado: Aprovado

Faculdade de Ciências e Tecnologia
Seção Técnica de Pós-Graduação
Rua Roberto Simonsen, 305 CEP 19060-900 Presidente Prudente SP
Tel 18 3229-5318 fax 18 3223-4519 posgrad@fct.unesp.br

AGRADECIMENTOS

A Deus, Nossa Senhora e a Santa Ana por sempre me auxiliarem nas horas difíceis e guiarem meus caminhos.

Ao meu orientador, Profº Drº Fernando Antônio Moala, por sua dedicação e por tudo o que me ensinou.

Aos membros da banca, especialmente aos examinadores externos por aceitarem o convite honrando-nos com sua presença.

Aos meus pais e a minha irmã que sempre me deram suporte para seguir em frente nessa caminhada tortuosa.

Ao meu namorado, melhor amigo e companheiro de todas as horas, Renato, por todo o carinho, compreensão e ajuda. É seu todo o meu amor.

À Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, pelo apoio financeiro.

Resumo

A elicitação é um processo que permite a incorporação da informação a priori fornecida por um especialista sobre alguma quantidade, desconhecida e de interesse, à informação proveniente dos dados do experimento. Pode ser utilizada em muitas áreas aplicadas do conhecimento, principalmente em situações nas quais os dados experimentais não são tão numerosos devido à dificuldade ou custo para obtê-los. Na abordagem Bayesiana não-paramétrica, a densidade ou a função de interesse podem ser estimadas sem imposição de quaisquer suposições restritivas sobre a sua forma. Assim, os dados permitem determinar a estimativa da função de interesse em vez de condicioná-la a pertencer a uma dada família paramétrica. O objetivo do presente trabalho é realizar uma aplicação do método Bayesiano não-paramétrico de elicitação de prioris proposto por Oakley e O'Hagan (2007), Moala (2009) e Moala e O'Hagan (2010) a fim de estimar a função de confiabilidade, considerada completamente desconhecida em sua forma, baseando-se apenas na informação fornecida pelo especialista.

Palavras-chave: Função de Confiabilidade. Elicitação. Inferência Bayesiana Não-Paramétrica.

Abstract

The elicitation is a process that allows the incorporation of a prior information provided by an expert on some quantity, unknown and of interest, to the information from the experimental data. It can be used in many applied areas of knowledge, especially in situations where experimental data are not numerous because of the difficulty or cost to obtain them. In Bayesian non-parametric approach, the density or the function of interest can be estimated without imposing any restrictive assumptions on its shape. Thus, the data allows determine the estimate of the interested function rather than conditioning it to a given parametric class functions. The aim of this paper is to apply the Bayesian non-parametric elicitation method of priors proposed by Oakley and O'Hagan (2007), Moala (2009) and Moala and O'Hagan (2010) to estimate the reliability function, considered completely unknown in its form, based only on the information provided by the expert.

Key words: Reliability function. Elicitation. Non-Parametric Bayesian Inference.

Sumário

1 Introdução p. 6

2 Confiabilidade p. 9

2.1 Inferência Paramétrica da Função de Confiabilidade . . . . . p. 10

2.2 Inferência Paramétrica da Função de Confiabilidade na presença de censura p. 17

2.2.1 Censura tipo II . . . . . p. 18

2.2.2 Censura tipo I . . . . . p. 22

3 Elicitação p. 28

3.1 O Processo de Elicitação . . . . . p. 29

3.2 Técnicas Paramétricas de Elicitação . . . . . p. 29

3.2.1 Ajustando a distribuição Beta . . . . . p. 30

3.2.2 Outros Métodos de Elicitação . . . . . p. 35

4 Elicitação em Análise de Confiabilidade p. 37

4.1 Elicitação da distribuição Beta . . . . . p. 37

4.2 Elicitação da Confiabilidade da distribuição Weibull . . . . . p. 38

4.2.1 O método . . . . . p. 39

5 Inferência Bayesiana Não-Paramétrica p. 46

5.1 Processos Estocásticos . . . . . p. 47

5.2 Processo Gaussiano . . . . . p. 48

5.2.1 Propriedades da distribuição Normal . . . . . p. 49

5.2.2 Definição de um processo Gaussiano . . . . . p. 50
5.3 Funções de Covariância . . . . . p. 52
5.4 Inferência Bayesiana sobre um Processo Gaussiano . . . . . p. 55

6 Inferência Bayesiana Não-Paramétrica para Elicitação de distribuição a priori p. 61
6.1 Processo Gaussiano como uma representação a priori para $f(.)$ . . . . . p. 62
6.2 Modelando uma apropriada função de covariância . . . . . p. 63
6.3 Distribuições a priori para m, b, η e σ^2 . . . . . p. 64
6.4 Dados Elicitados do Especialista . . . . . p. 65
6.5 Posteriori como atualização da priori . . . . . p. 67
6.5.1 Estimadores a posteriori para os hiperparâmetros . . . . . p. 68

7 Inferência Bayesiana Não-Paramétrica para Elicitação da Função de Confiabilidade p. 73
7.1 Representação a priori para $R(.)$ . . . . . p. 73
7.2 Elicitando Valores da Confiabilidade . . . . . p. 75
7.3 Atualização da priori . . . . . p. 75
7.4 Ilustração Numérica . . . . . p. 76

8 Conclusões p. 80

Referências p. 82

1 Introdução

A possibilidade de incorporar informação a priori fornecida por um especialista sobre alguma quantidade, desconhecida e de interesse, à informação proveniente dos dados do experimento é uma importante vantagem da inferência Bayesiana. O processo de incorporar formalmente esse conhecimento é denominado elicitação.

A utilização da informação do especialista permitirá termos um conhecimento mais atualizado, portanto mais preciso, da quantidade de interesse, justificando ainda mais o uso da inferência Bayesiana.

Os agrônomos e os agricultores têm uma larga experiência nas necessidades do solo para o plantio e colheita de produtos agrícolas como hortifrutigranjeiros, produtos para produção de biocombustíveis, etc. Além disso, devido à contaminação do solo, problemas de erosão e poluição em muitos ambientes urbanos, os especialistas podem manter atualizados os riscos associados aos elementos tóxicos na aplicação de fertilizantes e defensivos agrícolas. Os especialistas podem ter experiência também com avaliação de risco de contaminantes metálicos como cádmio, arsênico, níquel, entre outros, que podem ser associados com subprodutos industriais.

Apesar da grande contribuição dessa informação a uma análise estatística, o conhecimento a priori não é frequentemente considerado nas análises Bayesianas. Justificando a dificuldade de se representar a opinião do especialista subjetivamente ou que, com uma grande quantidade de dados, o conhecimento a priori tende a ter pouco efeito nas inferências finais e as várias técnicas disponíveis para representar não informação a priori, os usuários da inferência Bayesiana frequentemente evitam utilizar o conhecimento a priori disponível em prol do uso de prioris não informativas. Porém, ao investigarmos fenômenos novos ou raros, os poucos dados podem não fornecer informação suficiente para se tomar uma decisão estatística. Um exemplo disto ocorre ao se tentar modelar os danos causados por terremotos, estudo de doenças raras, etc. Nem sempre teremos dados suficientes para poder ignorar o conhecimento a priori, e consequentemente, a opinião do especialista pode

ser uma das poucas fontes de informação.

A elicitação pode ser utilizada em muitas áreas aplicadas do conhecimento, principalmente em situações nas quais os dados experimentais não são tão numerosos devido a dificuldade ou custo para obtê-los.

A análise de confiabilidade é uma área que pode ter importantes aplicações da elicitação. Por exemplo, suponha que estamos investigando o tempo de vida médio de um componente em alguma máquina. Podemos fazer testes através de uma amostra dos componentes para inferir o tempo de vida médio deles, mas o projetista do componente pode ter suas expectativas próprias sobre este desempenho. Se pudermos representar o conhecimento do projetista a respeito das características dos componentes que podem ter influência no tempo de vida por uma distribuição de probabilidade, então este adicional conhecimento (a priori) pode ser utilizado dentro do sistema Bayesiano de inferência. Outros exemplos de elicitação na análise de risco e confiabilidade podem ser encontrados em Cooke (1991) e Bedford e Cooke (2001).

Geralmente, assumimos uma específica família paramétrica de distribuições para a o tempo de falha, como Weibull, lognormal ou gama, tal que a inferência sobre a função de confiabilidade reduz-se ao problema da estimativa dos parâmetros da distribuição do tempo de falha.

Os métodos Bayesianos não-paramétricos são muito apropriados para a análise de confiabilidade, permitindo uma flexível modelagem para a função de confiabilidade, função de risco acumulada ou função de risco, sem as restritivas especificações paramétricas.

Em muitos problemas práticos, os dados apresentam características como multimodalidade, caudas pesadas, etc, que tornam as distribuições usuais inadequadas, preferindo-se assim uma abordagem não-paramétrica para a modelagem da função de confiabilidade em que esta é tratada como uma função completamente desconhecida.

A abordagem Bayesiana não-paramétrica também permite a incorporação de uma informação a priori na estimação da função de confiabilidade. Oakley e O'Hagan (2007) e Moala e O'Hagan (2010) propõem um método Bayesiano não-paramétrico para elicitar distribuições a priori f(.) univariadas e multivariadas, respectivamente, que pode ser aplicado sem muitas restrições e a uma ampla classe de problemas práticos. Além de evitar ter que assumir uma forma paramétrica para a função desconhecida, nos permite medir nossa incerteza sobre ela dado um número limitado de sumários elicitados do especialista e, principalmente, não é difícil de ser implementado.

Portanto, neste trabalho, pretendemos realizar uma aplicação do método Bayesiano não-paramétrico de elicitação de prioris proposto por Oakley e O'Hagan (2007), Moala (2009) e Moala e O'Hagan (2010) a fim de estimar a função de confiabilidade, considerada completamente desconhecida em sua forma, baseado apenas na informação fornecida pelo especialista.

O trabalho está organizado da seguinte forma: No Capítulo 2 é abordada a análise da função de confiabilidade sob o enfoque paramétrico Bayesiano e frequentista, para dados com e sem censura. No Capítulo 3 é discutido o uso da elicitação em análises estatísticas e algumas técnicas de elicitação paramétrica são apresentadas. Um novo método de elicitação aplicado à distribuição Weibull para determinação dos hiperparâmetros da distribuição conjunta a priori é abordado no Capítulo 4. A inferência Bayesiana não-paramétrica é analisada no Capítulo 5, onde o conceito de processo Gaussiano é introduzido. O Capítulo 6 aborda o método Bayesiano não-paramétrico de elicitação univariado proposto por Oakley e O'Hagan(2007) para estimação de uma função densidade qualquer. A estimação da função de confiabilidade utilizando o conhecimento do especialista, usando o método discutido no Capítulo 6, é apresentada no Capítulo 7. Algumas conclusões e sugestões para trabalhos futuros são descritas no Capítulo 8.

2 Confiabilidade

A análise de confiabilidade ou sobrevivência é bastante utilizada na área industrial, médica, financeira, entre outras. A variável resposta é o tempo transcorrido até a ocorrência de algum evento de interesse.

O objetivo da análise de confiabilidade é, estudando dados relacionados aos tempos de falha ou sobrevivência de unidades específicas como pessoas ou peças de máquinas, fornecer uma medida capaz de nos dar uma confiança em relação ao bom funcionamento de um certo produto durante um período de tempo especificado, sob condições de uso pré-estabelecidas (no caso da engenharia) ou em experimentos médicos.

A função de confiabilidade é uma das principais funções probabilísticas usadas para descrever estudos de tempo de vida. Ela é definida como

$$R(t) = P[X \geq t] \qquad t > 0. \tag{2.1}$$

A função de confiabilidade especifica a probabilidade de uma observação não falhar até um certo tempo t. Baseado na análise de confiabilidade, os fabricantes de produtos eletrônicos, por exemplo, podem melhorar a qualidade de seus produtos.

Uma característica dos dados utilizados na análise de confiabilidade é a presença de censura, que é a observação parcial da resposta. Devido a frequentemente elevada confiabilidade dos componentes testados e o custo do tempo para a realização do experimento, os testes de vida são geralmente terminados depois de uma quantidade específica de tempo decorrido (censura tipo I) ou após um determinado número de falhas ter sido observado (censura tipo II).

Existem diversas distribuições de probabilidade que são utilizadas em análise de confiabilidade, algumas delas destacam-se das demais por sua comprovada adequação a várias situações práticas, por exemplo, a distribuição Weibull. Ela é uma das mais importantes distribuições de probabilidade utilizadas em problemas de confiabilidade. Mann (1968) discute uma variedade de situações onde esta distribuição pode ser empregada.

Na estimação dos parâmetros de uma distribuição o método de máxima verossimi-

lhança é uma técnica estatística amplamente utilizada, fornecendo estimativas satisfatórias, principalmente quando o tamanho da amostra é grande. Porém, nos casos em que a estimação da função de confiabilidade é realizada com base em amostras pequenas, o estimador de máxima verossimilhança pode tornar-se impreciso e até mesmo pouco confiável. Assim, o uso da inferência Bayesiana pode ser uma alternativa poderosa na estimação dos parâmetros.

Esse capítulo está organizado da seguinte maneira: Na Seção 2.1, a inferência paramétrica da função de confiabilidade é discutida. A inferência paramétrica da função de confiabilidade na presença de censura é abordada na Seção 2.2.

2.1 Inferência Paramétrica da Função de Confiabilidade

Existem diversas distribuições de probabilidade que são utilizadas na análise estatística de dados de confiabilidade, tais como Exponencial, Log-normal, Weibull, entre outras. Tais distribuições têm se mostrado apropriadas para descrever o tempo de vida de equipamentos e produtos industriais.

A grande popularidade da distribuição Weibull se deve ao fato dela apresentar uma grande variedade de formas. A densidade da distribuição Weibull do tempo de falha da variável X é dada por

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\}, \quad x > 0, \tag{2.2}$$

dependente dos parâmetros $\alpha > 0$ e $\beta > 0$, em que α é o parâmetro de escala e β o de forma da distribuição.

Para a distribuição Weibull definida em (2.2), a função de confiabilidade é dada por

$$R(t) = exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}. \tag{2.3}$$

A estimação da função de confiabilidade pode ser feita através dos métodos Bayesianos ou do método de máxima verossimilhança.

Algumas formas da função de confiabilidade de uma variável X com distribuição Weibull são apresentadas na Figura 1.

Seja $X_1,\ldots,X_n$ uma amostra aleatória de X com função densidade de probabilidade

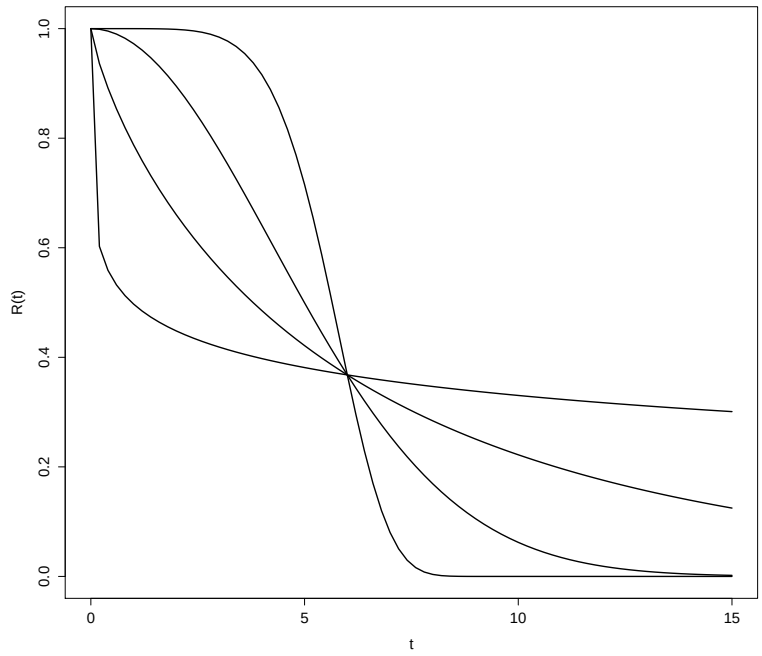


Figura 1: Função de confiabilidade da distribuição Weibull para diversos β's e $\alpha = 8$.

dada em (2.2), então a função de verossimilhança para os parâmetros α e β é dada por

$$L(\alpha, \beta|\mathbf{x}) = \left(\frac{\beta}{\alpha}\right)^n \prod_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\beta-1} exp\left\{-\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta}\right\}. \tag{2.4}$$

Os estimadores de máxima verossimilhança de α e β podem ser obtidos resolvendo simultaneamente as equações $\frac{\partial ln\, L}{\partial \alpha} = 0$ e $\frac{\partial ln\, L}{\partial \beta} = 0$. Dessa forma, a estimativa do parâmetro α é

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^{n} x_i^{\hat{\beta}}}{n}\right)^{\frac{1}{\hat{\beta}}} \tag{2.5}$$

e substituindo (2.5) na equação $\frac{\partial ln\, L}{\partial \beta} = 0$ tem-se

$$\frac{1}{\hat{\beta}} + \frac{\sum_{i=1}^{n} ln\, x_i}{n} - \frac{\sum_{i=1}^{n} x_i^{\hat{\beta}} ln\, x_i}{\sum_{i=1}^{n} x_i^{\hat{\beta}}} = 0 \tag{2.6}$$

Como (2.6) não pode ser resolvida analiticamente para β, utiliza-se alguma aproximação numérica, como o Método de Newton Raphson.

Pela propriedade de invariância dos estimadores de máxima verossimilhança, uma vez

estimados α e β podemos obter a estimativa de $\hat{R} = \hat{R}(t_0)$ que é dada por

$$\hat{R} = exp\left\{-\left(\frac{t_0}{\hat{\alpha}}\right)^{\hat{\beta}}\right\}. \tag{2.7}$$

Uma possível reparametrização encontrada em Moala, Rodrigues e Tomazella (2009) pode ser utilizada para obter as propriedades assintóticas do estimador da confiabilidade, $\hat{R}$. Reparametrizando (α, β) para (R, W), em que $W = \beta$ e $R = exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$ e utilizando (2.4), temos que a função de verossimilhança de (R, W) é dada por

$$L(R, W|\mathbf{x}) = W^n \left(ln\frac{1}{R}\right)^n \prod_{i=1}^{n} y_i^{W-1} R^{\sum_{i=1}^{n} y_i^W}. \tag{2.8}$$

em que $y_i = \frac{x_i}{t}$.

Moala, Rodrigues e Tomazella (2009) obtiveram a matriz de informação de Fisher esperada para os parâmetros (R, W) dada por

$$I(R, W) = \begin{bmatrix} \frac{1}{R^2\left(ln\frac{1}{R}\right)^2} & \frac{\psi(2) - ln\left(ln\frac{1}{R}\right)}{RWln\frac{1}{R}} \\ \frac{\psi(2) - ln\left(ln\frac{1}{R}\right)}{RWln\frac{1}{R}} & \frac{1}{W^2}\left\{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)ln\left(ln\frac{1}{R}\right) + d\right\} \end{bmatrix} \tag{2.9}$$

em que ψ é a função digama, ψ' é a função trigama e $d = \psi'(2) + \psi(2)^2 + 1$.

Sabendo que assintoticamente os estimadores R e W possuem distribuição

$$(\hat{R}, \hat{W}) \sim N_2\Big[(R, W), I^{-1}(R, W)\Big], \tag{2.10}$$

conclui-se que a distribuição assintótica do estimador de máxima verossimilhança $\hat{R}$ é

$$\frac{\sqrt{n}(\hat{R} - R)}{\sqrt{var(\hat{R})}} \overset{a}{\sim} N(0, 1) \tag{2.11}$$

em que $var(\hat{R}) = \frac{6}{n\pi^2}\left(Rln\frac{1}{R}\right)^2\left[ln^2\left(ln\frac{1}{R}\right) - 2\psi(2) - ln\left(ln\frac{1}{R}\right) + d\right]$.

Em uma análise Bayesiana é necessário especificar uma distribuição a priori para os parâmetros. Se não existe informação a priori previamente disponível, então a incerteza sobre os parâmetros pode ser quantificada com distribuições a priori não informativas. Existem diversas distribuições a priori na literatura, algumas das mais conhecidas são as prioris de Jeffreys (1967) e Referência (Bernardo,1979).

Uma distribuição a priori não informativa para representar uma situação em que existe pouca informação previamente disponível sobre os parâmetros foi proposta por Jeffreys (1967). Desde então, essa distribuição a priori tem um importante papel na inferência

Bayesiana. A priori de Jeffreys é dada por

$$\pi(R, W) \propto \sqrt{det\, I(R, W)}. \tag{2.12}$$

Essa distribuição a priori possui a vantagem de ser invariante sob transformações um a um dos parâmetros. Supondo que não existe nenhuma informação a priori sobre R e W e utilizando (2.9) e (2.12), a priori de Jeffreys para (R, W) é dada por

$$\pi_1(R, W) \propto \frac{1}{RWln\frac{1}{R}}. \tag{2.13}$$

Na inferência Bayesiana, todas as quantidades incertas são modeladas em termos da sua distribuição conjunta a priori e posteriormente são atualizadas, pelos dados da amostra, em termos da distribuição conjunta a posteriori. Logo, a distribuição conjunta a posteriori é dada por

$$p_1(R, W|\mathbf{x}) \propto W^{n-1} \left(ln\frac{1}{R}\right)^{n-1} \prod_{i=1}^{n} y_i^{W-1} R^{\sum_{i=1}^{n} y_i^W} \frac{1}{R}. \tag{2.14}$$

A priori de referência é uma outra priori que pode ser empregada na análise de confiabilidade. Ela foi introduzida por Bernardo em 1979 e posteriormente desenvolvida por Berger e Bernardo (1992). O intuito é encontrar uma distribuição a priori não informativa que maximize a informação para os parâmetros proveniente dos dados.

Uma característica importante do método de Berger-Bernardo para construir uma priori não informativa é o diferente tratamento aos parâmetros de interesse e nuisance. Quando existem parâmetros nuisance, a priori de referência irá depender dos parâmetros que foram julgados de interesse.

Suponha que a distribuição conjunta a posteriori de (ϕ, λ) é assintoticamente normal com matriz de covariância $S(\hat{\phi}, \hat{\lambda}) = I^{-1}(\hat{\phi}, \hat{\lambda})$. Assim, sob adequadas condições de regularidade (ver Bernardo (1996) para mais detalhes) a priori conjunta de referência pode ser escrita como o produto de duas funções independentes dos parâmetros como mostra o Teorema 1, descrito em Bernardo (2005).

Teorema 1: Se o espaço do parâmetro nuisance $\Lambda(\phi) = \Lambda$ é independente de ϕ e as funções $S_{11}^{-\frac{1}{2}}(\phi, \lambda)$ e $I_{22}^{\frac{1}{2}}(\phi, \lambda)$ forem fatoradas na forma

$$[S_{11}(\phi, \lambda)]^{-\frac{1}{2}} = g_1(\phi)h_1(\lambda) \quad , \quad [I_{22}(\phi, \lambda)]^{\frac{1}{2}} = g_2(\phi)h_2(\lambda) \tag{2.15}$$

então

$$\pi(\phi) \propto g_1(\phi) \qquad , \qquad \pi(\lambda|\phi) \propto h_2(\lambda) \tag{2.16}$$

e a distribuição a priori de referência considerando ϕ o parâmetro de interesse e λ nuisance é dada por

$$\pi(\lambda, \phi) = g_1(\phi)h_2(\lambda). \tag{2.17}$$

Mais detalhes podem ser encontrados em Bernardo(2005).

Agora, considere a construção da distribuição a priori de referência para a função de confiabilidade $R(t)$ da distribuição Weibull apresentada em (2.2).

A distribuição a priori de referência não é invariante, assim não seria possível fazer inferências sobre a confiabilidade se tivéssemos utilizado a parametrização (α, β). Dessa forma, ela só pôde ser utilizada porque realizamos a reparametrização (R, W). O parâmetro de interesse é $R(t) = exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$ e o parâmetro nuisance escolhido é $W = \beta$.

A matriz de informação de Fisher, dada em (2.9), possui a seguinte matriz inversa

$$S(R, W) = \frac{6}{\pi^2}\begin{bmatrix} R^2\left(ln\frac{1}{R}\right)^2\left[ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)p + d\right] & -WRln\frac{1}{R}\left[\psi(2) - p\right] \\ -WRln\frac{1}{R}\left[\psi(2) - p\right] & W^2 \end{bmatrix} \tag{2.18}$$

em que $d = \psi'(2) + \psi(2)^2 + 1$ e $p = ln\left(ln\frac{1}{R}\right)$. Assim, a priori de referência é determinada da seguinte forma

$$S_{11}^{-\frac{1}{2}}(\phi, \lambda) = \frac{1}{R\left(ln\frac{1}{R}\right)\sqrt{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)p + d}} = g_1(R)h_1(W) \tag{2.19}$$

com $g_1(R) = \frac{1}{R\left(ln\frac{1}{R}\right)\sqrt{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)p + d}}$, tal que

$$\pi(R) = \frac{1}{R\left(ln\frac{1}{R}\right)\sqrt{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)p + d}}. \tag{2.20}$$

Similarmente,

$$I_{22}^{\frac{1}{2}}(\phi, \lambda) = \frac{1}{W}\sqrt{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)p + d} = g_2(R)h_2(W) \tag{2.21}$$

com $h_2(W) = \frac{1}{W}$, e então

$$\pi(W|R) = \frac{1}{W}. \tag{2.22}$$

Portanto, usando o Teorema 1, a distribuição priori conjunta de referência com parâmetro de interesse R é

$$\pi_2(R, W) = \frac{1}{WR\left(ln\frac{1}{R}\right)\sqrt{ln^2\left(ln\frac{1}{R}\right) - 2\psi(2)ln\left(ln\frac{1}{R}\right) + \psi'(2) + \psi(2)^2 + 1}}. \tag{2.23}$$

A distribuição conjunta a posteriori para os parâmetros (R, W) é dada por

$$p_2(R, W|\mathbf{x}) \propto \frac{1}{W^n}\left(ln\frac{1}{R}\right)^n \prod_{i=1}^{n} y_i^{\frac{1}{W}-1} R^{\sum_{i=1}^{n} y_i^{\frac{1}{W}}} \pi(R, W), \tag{2.24}$$

em que $\pi(R, W)$ é uma distribuição conjunta a priori apropriada com $0 < R < 1$ e $W > 0$.

Claramente (2.24) é uma complexa função de R e W e para fazer inferências sobre o parâmetro R é necessário obter as integrais que envolvem a densidade a posteriori, a qual é analiticamente intratável. Dessa forma, o uso de métodos numéricos MCMC (Markov Chain Monte Carlo), como o popular algoritmo de Metropolis-Hastings, torna-se indispensável para a obtenção das estimativas dos parâmetros desconhecidos.

Exemplo 1: A situação apresentada aqui pertence a um famoso exemplo de Lieblein e Zelen (1956) sobre o tempo de fadiga de rolamentos de esferas. Os dados são apresentados na Tabela 1.

Tabela 1: Observações do número de ciclos até a falha

17,88	28,92	33,00	41,52	42,12	45,60
48,48	51,84	51,96	54,12	55,56	67,80
68,64	68,88	84,12	93,12	98,64	105,12
105,84	127,92	128,04	173,40		

Foi utilizado o método MCMC (Monte Carlo via Cadeias de Markov), algoritmo de Metropolis Hastings, para simulação de amostras dos valores de R e W das distribuições marginais a posteriori de (2.14) e (2.24). Como o maior interesse está na estimativa da confiabilidade, apenas os resultados de R serão apresentados.

Uma cadeia com 20000 iterações foi gerada no pacote estatístico R (R Development Core Team, 2012), com o objetivo de simular amostras dos valores de R das distribuição a posteriori. As exigências de convergência da cadeia para a distribuição de equilíbrio foram checadas, como mostram as Figura 2 e 3.

Os estimadores de máxima verossimilhança são: $\hat{\alpha} = 81,88$ e $\hat{\beta} = 2,102$. Substituindo esses valores estimados em (2.7), obtém-se a estimativa da confiabilidade para $t = 100$ que é de aproximadamente 0,22 e seu intervalo de confiança de 95% [0,107 ; 0,391].

A Tabela 2 fornece as estimativas a posteriori do parâmetro R (média) e respectivas variâncias, além dos intervalos de credibilidade de 95% para cada uma das distribuições a priori, Jeffreys e Referência.

De acordo com os resultados apresentados na Tabela 2, a probabilidade do tempo de vida dos rolamentos de esfera ser maior que 100 ciclos é de aproximadamente 0,24.

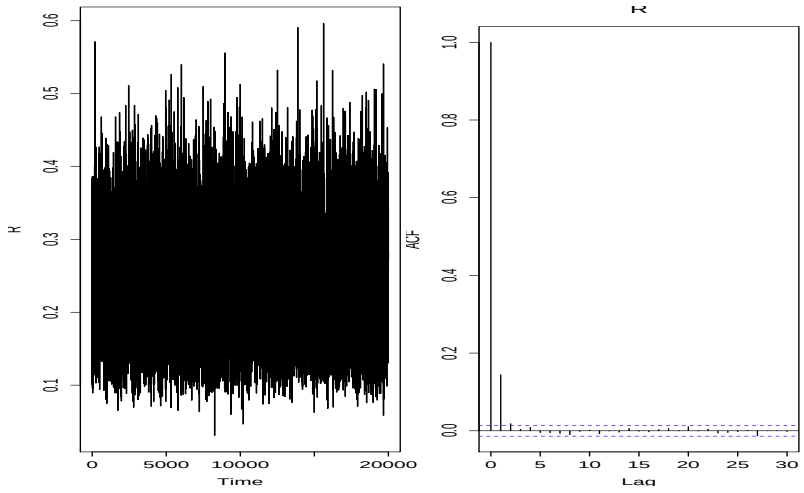


Figura 2: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando a distribuição a priori de Jeffreys.

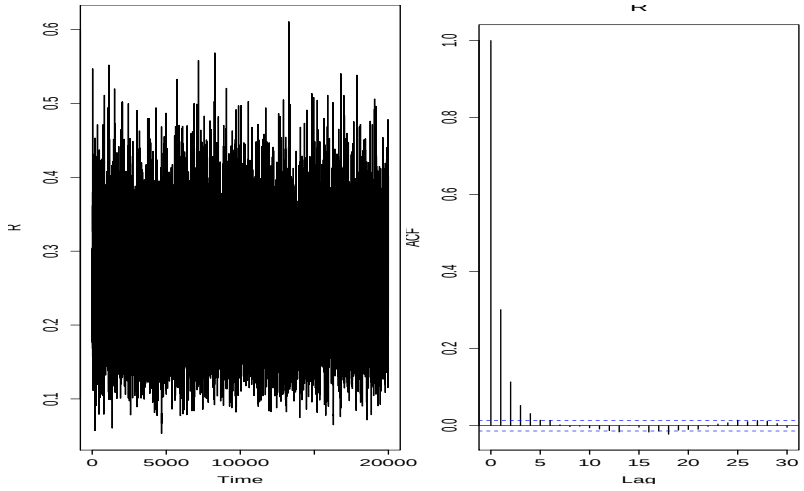


Figura 3: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando a distribuição a priori de Referência.

Esse exemplo já foi discutido por vários autores, incluindo Singpwarla (1988) que ajustou esses dados a uma distribuição Weibull utilizando informação de um especialista para determinar os hiperparâmetros da distribuição conjunta a priori. Para a distribuição a priori do β, o autor considerou a distribuição Gama(θ, φ), dada por

$$\pi(\beta) = \frac{\theta^{\varphi}}{\Gamma(\varphi)}\beta^{\varphi-1}exp\{-\theta\beta\} \qquad \beta > 0, \tag{2.25}$$

como priori para o parâmetro β e para α considerou uma normal truncada

$$\pi(\alpha) = \frac{k}{\sqrt{2\pi}\sigma}exp\left\{-\frac{1}{2}\left(\frac{\alpha e^{\frac{c}{\beta}} - \mu}{\sigma}\right)^2\right\} \qquad \alpha \geq 0, \tag{2.26}$$

em que $c = ln(ln2)$ e $k = \int_{-\frac{\mu}{\sigma}}^{\infty}(2\pi)^{-\frac{1}{2}}exp\{-\frac{u^2}{2}\}du$.

Tabela 2: Estimativas para o parâmetro R para $t = 100$

Estimativas	Jeffreys	Referência
Média	0,239	0,244
Variância	0,005	0,005
I. Cred.	[0,116 ; 0,389]	[0,119 ; 0,404]

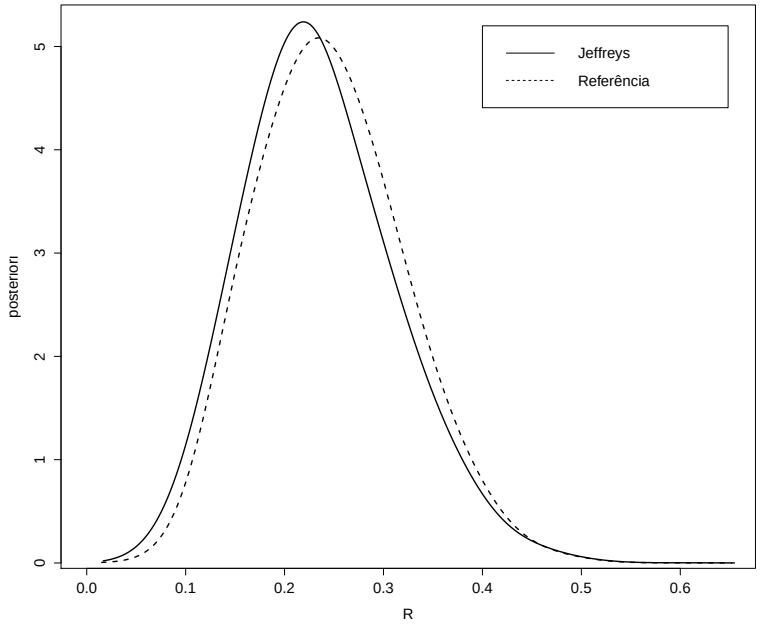


Figura 4: Densidades marginais a posteriori para o parâmetro R.

A multiplicação de (2.25) e (2.26) resulta na distribuição conjunta a priori especificada por ele, cujos hiperparâmetros escolhidos foram $\theta = 0,15$, $\varphi = 2$, $\mu = 80$ e $\sigma = 12$. O autor não estabelece a metodologia para obtenção dos valores considerados para os hiperparâmetros. As estimativas dos parâmetros α e β encontradas foram, respectivamente, 85 e 2,12. Substituindo novamente os valores das estimativas dos parâmetros em (2.7), encontra-se 0,243 como estimativa da confiabilidade em $t = 100$.

Observa-se que as diferentes abordagens (Frequentista e Bayesiana) adotadas para a resolução desse problema produziram resultados similares.

2.2 Inferência Paramétrica da Função de Confiabilidade na presença de censura

Na análise de confiabilidade a presença de observações incompletas ou parciais é frequente. Essas observações censuradas podem ocorrer por uma variedade de razões, dentre elas, a não ocorrência do evento até o término do experimento.

Todos os resultados provenientes de um estudo de confiabilidade devem ser utilizados na análise estatística. As razões são:
i) Apesar de serem incompletas, as observações censuradas fornecem informações sobre o tempo de vida do produto ou componente em estudo;
ii) O descarte das observações censuradas na análise estatística pode levar a conclusões viciadas.

A seguir serão apresentados os procedimentos para a análise de confiabilidade de dados que possuem censura tipo II e censura tipo I.

2.2.1 Censura tipo II

Para dados censurados tipo II, consideramos as n unidades em teste e terminamos o experimento quando ocorre a r-ésima falha (r fixo, $1 \leq r \leq n$). Assim, somente os r menores tempos são observados. O número de falhas r é fixado antes do começo do experimento. Nesse caso, a função de verossimilhança baseada nos r tempos de falha $X_{(1)},\ldots,X_{(r)}$ é dada por

$$L = \frac{n!}{(n-r)!} f(x_{(1)}) \ldots f(x_{(r)}) [S(x_{(r)})]^{n-r}. \tag{2.27}$$

Para tempos de vida de componentes provenientes da distribuição Weibull apresentada em (2.2) com censura tipo II, a função de verossimilhança é dada por

$$L(\alpha, \beta | \mathbf{x}) = \left(\frac{\beta}{\alpha}\right)^r \prod_{i=1}^{r} \left(\frac{x_{(i)}}{\alpha}\right)^{\beta-1} exp\left\{ -\sum_{i=1}^{r} \left(\frac{x_{(i)}}{\alpha}\right)^{\beta} - (n-r)\left(\frac{x_{(r)}}{\alpha}\right)^{\beta} \right\}. \tag{2.28}$$

Os estimadores de máxima verossimilhança, $\hat{\alpha}$ e $\hat{\beta}$, dos parâmetros desconhecidos da distribuição Weibull são dados por

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^{n} x_i^{\hat{\beta}}}{r} \right)^{\frac{1}{\hat{\beta}}} \tag{2.29}$$

e substituindo (2.29) na equação $\frac{\partial ln\, L}{\partial \beta} = 0$ tem-se

$$\frac{1}{\hat{\beta}} + \frac{\sum_{i=1}^{n} ln\, x_i}{r} - \frac{\sum_{i=1}^{n} x_i^{\hat{\beta}} ln\, x_i}{\sum_{i=1}^{n} x_i^{\hat{\beta}}} = 0. \tag{2.30}$$

Novamente algum método numérico como o de Newton-Raphson deve ser utilizado

para resolver a equação (2.30).

A estimativa de máxima verossimilhança da confiabilidade, $\hat{R} = \hat{R}(t_0)$, pode ser obtida através das estimativas de máxima verossimilhança de α e β por $\hat{R} = exp\left\{-\left(\frac{t_0}{\hat{\alpha}}\right)^{\hat{\beta}}\right\}$.

Para uma análise Bayesiana com esquema de censura, podemos considerar diferentes distribuições a priori para a confiabilidade.

Primeiramente, visto que $R(t)$ é uma função de α e β para um tempo t fixado, e (R, W) possui uma relação um-a-um com (α, β), é possível desenvolver um procedimento alternativo de estimação de $R(t)$. Ao invés de especificar uma distribuição a priori para α e β do modelo Weibull, podemos usar uma distribuição que varie no intervalo $[0, 1]$ para representar a distribuição a priori para a confiabilidade $R(t)$, para t fixado.

Então considerando a parametrização (R, W) e os dados censurados tipo II, a função de verossimilhança de R e W pode ser encontrada substituindo os parâmetros α e β da função de verossimilhança da distribuição Weibull em (2.28) por $W = \beta$ e $R = exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$. Assim a função de verossimilhança de (R,W) pode ser escrita como

$$L(R, W|\mathbf{x}) = W^r \left(ln\frac{1}{R}\right)^r \prod_{i=1}^{r} y_{(i)}^{W-1} R^{\sum_{i=1}^{r} y_{(i)}^W - (n-r)y_{(r)}^W}. \tag{2.31}$$

Uma possível distribuição a priori para R poderia ser a distribuição Beta, com densidade dada por

$$\pi(R) = \frac{1}{B(\upsilon, \gamma)} R^{\upsilon-1}(1-R)^{\gamma-1} \qquad 0 < R < 1 \tag{2.32}$$

em que $\upsilon, \gamma > 0$ e $B(\upsilon, \gamma)$ é a função beta com $B(\upsilon, \gamma) = \frac{\Gamma(\upsilon)\Gamma(\gamma)}{\Gamma(\upsilon+\gamma)}$. A distribuição Beta possui uma grande variedade de formas e tem média $\frac{\upsilon}{\upsilon+\gamma}$ e variância $\frac{\upsilon\gamma}{(\upsilon+\gamma)^2(\upsilon+\gamma+1)}$.

A distribuição Beta(0,0) é uma distribuição a priori imprópria e algumas vezes é utilizada para representar o não conhecimento de informações a priori dos valores dos parâmetros. Assumindo independência entre R e W e considerando que a distribuição a priori de R é a Beta(υ, γ) e a de W é uma Gama(a, b) (observe que W é uma variável não negativa), tem-se a distribuição conjunta a priori dada por

$$\pi_3(R, W) \propto R^{\upsilon-1}(1-R)^{\gamma-1} W^{a-1} exp\{-Wb\}, \tag{2.33}$$

com hiperparâmetros $\upsilon > 0$, $\gamma > 0$, $a > 0$ e $b > 0$. Para que a distribuição Gama torne-se não informativa os valores dos hiperparâmetros são geralmente considerados conhecidos com $a = 0.01$ e $b = 0.01$.

Pode-se considerar também a distribuição Log-Gama Negativa como priori para R.

Nesse caso, a densidade com κ e λ como hiperparâmetros é dada por

$$\pi(R) = \frac{\lambda^{\kappa}}{\Gamma(\kappa)} \left(\frac{1}{R}\right)^{\kappa-1} R^{\lambda-1}, \qquad 0 < R < 1, \tag{2.34}$$

denotada por $LGN(\kappa, \lambda)$. A média dessa distribuição é $\left(1+\frac{1}{\lambda}\right)^{-\kappa}$ e a variância é dada por $\left(1+\frac{2}{\lambda}\right)^{-\kappa} - \left(1+\frac{1}{\lambda}\right)^{-2\kappa}$.

Não é difícil mostrar que a função acumulada de R correspondente a (2.34) pode ser escrita como

$$F(R|\kappa, \lambda) = 1 - I\left(\kappa, \lambda\, ln\frac{1}{R}\right), \qquad 0 < R < 1, \tag{2.35}$$

em que $I(u, z)$ é a função gama incompleta.

A distribuição Log-Gama Negativa foi discutida e utilizada em confiabilidade por Springer e Thompson (1965, 1967), Mann(1968) e outros.

A distribuição Log-Gama Negativa e Beta seriam um procedimento alternativo na seleção de distribuições a priori, quando o especialista possui um conhecimento mais aprofundado sobre as propriedades da confiabilidade de um item em estudo. Novamente os parâmetros R e W foram considerados independentes e a distribuição Gama foi atribuída ao parâmetro W. Assim a distribuição conjunta a priori é dada por

$$\pi_4(R, W) \propto \left(\frac{1}{R}\right)^{\kappa-1} R^{\lambda-1} W^{a-1} exp\left\{-Wb\right\}, \qquad 0 < R < 1, W > 0, \tag{2.36}$$

com hiperparâmetros $\kappa > 0$, $\lambda > 0$, $a > 0$ e $b > 0$.

Assumindo dependência entre os parâmetros R e W, pode-se considerar uma distribuição a priori bivariada derivada da função cópula. Um caso especial é dado por Farlie-Gumbel-Morgenstern (ver Morgenstern, 1956) e sua expressão é

$$\pi(R, W|\rho) = f_1(R) f_2(W) + \rho f_1(R) f_2(W)[1 - 2F_1(R)][1 - 2F_2(W)], \tag{2.37}$$

em que $f_1(R)$ e $f_2(W)$ são, respectivamente, as distribuições marginais para R e W; $F_1(R)$ e $F_2(W)$ correspondem as distribuições acumuladas de R e W, respectivamente. Observe que se $\rho = 0$, os parâmetros R e W são independentes.

Nesse trabalho, assumimos distribuição marginal Gama para W e distribuição marginal Log-Gama Negativa para R, resultando na seguinte priori

$$\pi_5(R, W|\rho) \propto \left(\frac{1}{R}\right)^{\kappa-1} R^{\lambda-1} W^{a-1} exp\left\{-\gamma\right\} \left[1 + \rho\left[I\left(\kappa, \lambda\, ln\frac{1}{R}\right) - 1\right]\left[1 - 2I\left(a, \gamma\right)\right]\right., \tag{2.38}$$

com $\gamma = Wb$, $0 < R < 1$, $W > 0$ e $-1 < \rho < 1$.

Nesse caso, a distribuição a posteriori pode ser escrita como

$$p_5(R, W|x) \propto L(R, W|x)\pi(R, W|\rho)\pi(\rho), \tag{2.39}$$

em que $\pi(\rho)$ é a distribuição a priori para ρ.

Muitas distribuições a priori podem ser especificadas para ρ, uma possibilidade é considerar uma distribuição a priori uniforme no intervalo $[-1, 1]$.

Como as distribuições a priori propostas nesse trabalho requerem métodos de simulação para a determinação da distribuição a posteriori, média a posteriori, variância e intervalos estimados utilizou-se o MCMC para gerar amostras da distribuição conjunta a posteriori.

Exemplo 2: Mann e Fertig (1973) apresentam um conjunto de dados provenientes de um teste de tempo de vida de componentes de um avião. Foram colocados em teste 13 componentes, o experimento foi finalizado quando ocorreu a décima falha. Os dados são apresentados na Tabela 3.

Tabela 3: Tempo de vida (em horas) de componentes de um avião

0,22	0,50	0,88	1,00	1,32
1,33	1,54	1,76	2,50	3,00

Para que as distribuições a priori se tornassem não informativas, os hiperparâmetros escolhidos da distribuição Log Gama Negativa foram $\kappa = 0,1$ e $\lambda = 0,01$. Já para a distribuição Beta, foram especificados $\upsilon = 0$ e $\gamma = 0$.

Uma cadeia com 20000 iterações foi gerada para cada distribuição a posteriori de R. As exigências de convergência foram checadas através dos gráficos apresentados nas Figuras 5, 6 e 7.

Tabela 4: Estimativas a posteriori para R em $t = 2$ e intervalos de credibilidade de 95%

Estimativas	Beta	LGN	Cópula
Média	0,212	0,224	0,222
Variância	0,010	0,011	0,011
I.Cred.	[0,056 ; 0,444]	[0,066 ; 0,465]	[0,064 ; 0,457]

De acordo com os resultados obtidos utilizando as distribuições a priori Log-Gama Negativa e Cópula, a probabilidade estimada do tempo de vida de componentes de um avião ser maior do que duas horas é de aproximadamente 0,22. E segundo o resultado da priori Beta, essa probabilidade é de 0,21.

A Figura 8 mostra as densidades marginais a posteriori da confiabilidade.

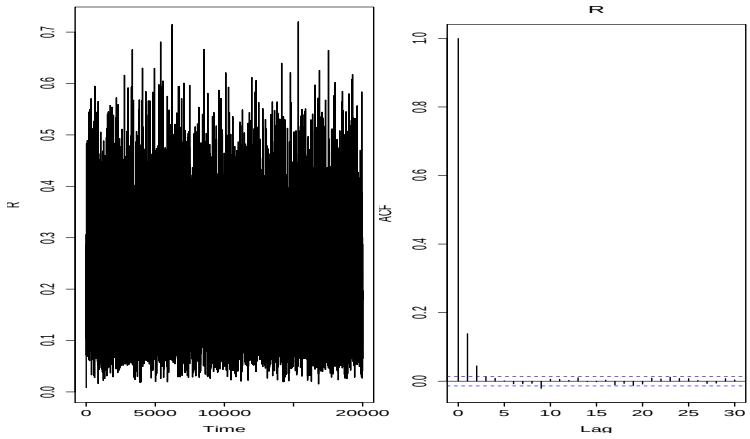


Figura 5: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando como priori a distribuição Beta.

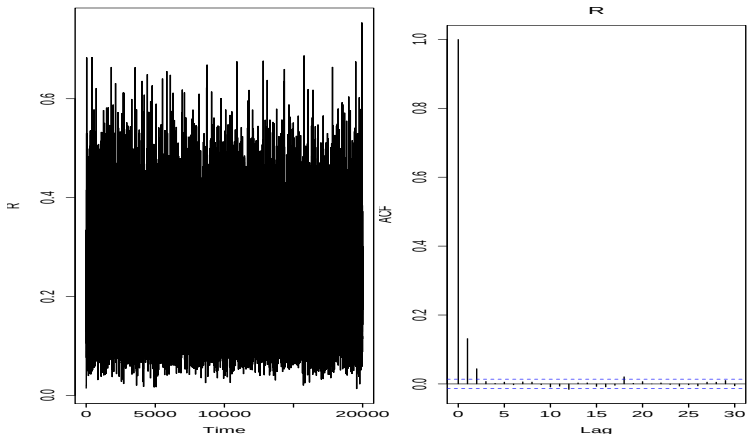


Figura 6: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando como priori a distribuição Log Gama Negativa.

2.2.2 Censura tipo I

No esquema de censura tipo I, o experimento é finalizado após um período pré-estabelecido de tempo, L. Uma forma mais geral de dados com censura tipo I é um experimento em que cada componente testado possui seu específico tempo de censura, pois nem todas as unidades começaram o teste ao mesmo tempo.

Suponha que n itens (componentes eletrônicos, por exemplo) são testados em um experimento de tempo de vida. Os tempos de vida das unidades amostrais são variáveis aleatórias independentes e identicamente distribuídas. Seja $L_1,\ldots,L_n$ os tempos de censura fixados para cada unidade no experimento. O tempo de vida X_i do i-ésimo item será observado se $X_i \leq L_i$, $i = 1,\ldots,n$. Portanto, os dados observados são (t_i, δ_i), em que

$$t_i = min(X_i, L_i) \tag{2.40}$$

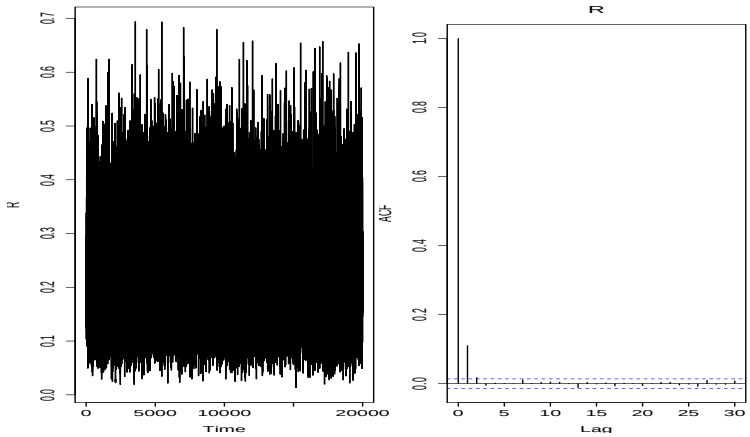


Figura 7: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando a priori Cópula.

e

$$\delta_i = \begin{cases} 1 & \text{se } T_i \leq L_i \\ 0 & \text{se } T_i > L_i. \end{cases}, \tag{2.41}$$

é a variável aleatória indicadora δ_i que mostra se T_i é censurado ou não.

Se os pares (t_i, δ_i) são independentes, então a função de verossimilhança é dada por

$$L = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(L_i)]^{1-\delta_i}, \tag{2.42}$$

onde $S(.)$ é a função de sobrevivência.

Se os tempos de vida dos itens colocados sob teste seguem uma distribuição Weibull dada em (2.2), então a função de verossimilhança para os parâmetros α e β é escrita como

$$L(\alpha, \beta|\mathbf{x}) = \left(\frac{\alpha}{\beta}\right)^r \prod_{i=1}^{n} \left(\frac{x_{(i)}}{\alpha}\right)^{(\beta-1)\delta_i} exp\left\{-\sum_{i=1}^{n} \delta_i \left(\frac{x_{(i)}}{\alpha}\right)^{\beta}\right\} exp\left\{-\sum_{i=1}^{n}(1-\delta_i)\left(\frac{x_{(i)}}{\alpha}\right)^{\beta}\right\}, \tag{2.43}$$

em que $r = \sum_{i=1}^{n} \delta_i$ denota o número de itens que falharam ($\delta_i = 1$).

As equações de verossimilhança para α e β são dadas por

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^{n} \delta_i x_i^{\hat{\beta}} + \sum_{i=1}^{n}(1-\delta_i) L_i^{\hat{\beta}}}{r}\right)^{\frac{1}{\hat{\beta}}} \tag{2.44}$$

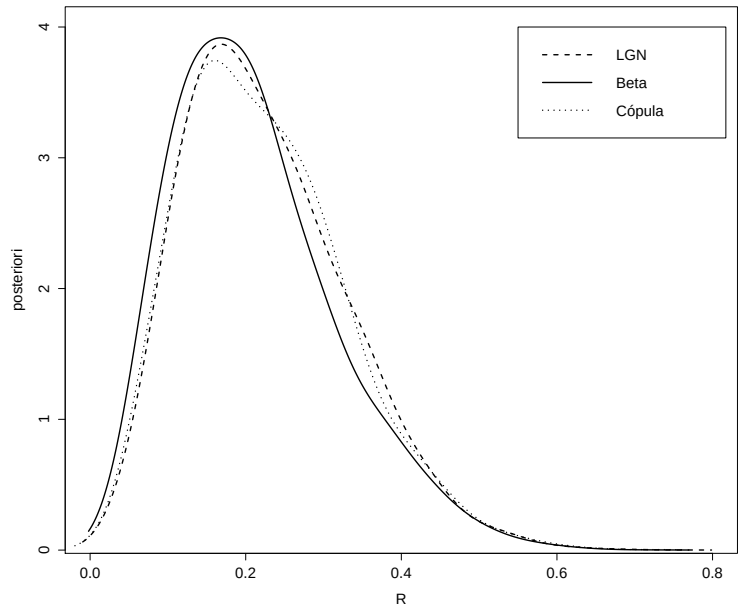


Figura 8: Densidades marginais a posteriori para o parâmetro R

e $\hat{\beta}$ como solução da seguinte equação

$$\frac{1}{\hat{\beta}} + \frac{\sum_{i=1}^{n} \delta_i ln\, x_i}{r} - \frac{\sum_{i=1}^{n} \delta_i x_i^{\hat{\beta}} ln\, x_i + \sum_{i=1}^{n} (1-\delta_i) L_i^{\hat{\beta}} ln\, L_i}{\sum_{i=1}^{n} \delta_i x_i^{\hat{\beta}} + \sum_{i=1}^{n} (1-\delta_i) L_i} = 0. \tag{2.45}$$

Para esse esquema de censura, a função de verossimilhança dos parâmetros (R, W) é dada por

$$L(R, W|\mathbf{x}) = W^r \left(ln \frac{1}{R} \right)^r \prod_{i=1}^{r} y_{(i)}^{(W-1)\delta_i} R^{T^*} \tag{2.46}$$

em que $T^* = \sum_{i=1}^{n} \delta_i y_i^{\hat{\beta}} + \sum_{i=1}^{n} (1-\delta_i) \left(\frac{L_i}{t}\right)^{\hat{\beta}}$.

Para uma análise Bayesiana da confiabilidade com censura tipo I, consideramos as mesmas distribuições a priori discutidas na seção anterior, porém a função de verossimilhança utilizada é dada em (2.46).

Exemplo 3: Bartholomew (1957) descreve uma situação em que peças de um equipamento são instaladas em diferentes instantes. Após algum tempo, algumas peças falharam e as demais continuaram funcionando. O objetivo é estudar a distribuição dos tempos de falha desse tipo de peça de equipamento e estimar a proporção delas que irão falhar dentro de um tempo especificado.

A Tabela 6 apresenta os resultados de 10 equipamentos testados nesse experimento,

que foi finalizado em 31 de agosto.

Tabela 5: Tempo de funcionamento de 10 peças de equipamento

Item	1	2	3	4	5
Data da Instalação	11 de jun.	21 de jun.	22 de jun.	2 de jul.	21 de jul.
Data da Falha	13 de jun.	-	12 de ago.	-	23 de ago.
Tempo de Falha(dias)	2	$\geq$72	51	$\geq$60	33
Item	6	7	8	9	10
Data da Instalação	31 de jul.	31 de jul.	1 ago.	2 de ago.	10 de ago.
Data da Falha	27 de ago.	14 de ago.	25 de ago.	6 de ago.	-
Tempo de Falha(dias)	27	14	24	4	$\geq$21

Observe que existem três itens que não falharam durante o experimento (nº 2, 4 e 10), logo eles são censurados. Os dados da Tabela 6 podem ser escritos de uma outra forma apresentada na Tabela 7.

Tabela 6: Tempo de falha e tempo pré-estabelecido para cada peça de equipamento

Item	1	2	3	4	5	6	7	8	9	10
T_i	2	-	51	-	33	27	14	24	4	-
L_i	81	72	70	60	41	31	31	30	29	21

Uma cadeia com 20000 iterações foi gerada para cada distribuição a posteriori. As exigências de convergência foram checadas através dos gráficos apresentados nas Figuras 9, 10 e 11.

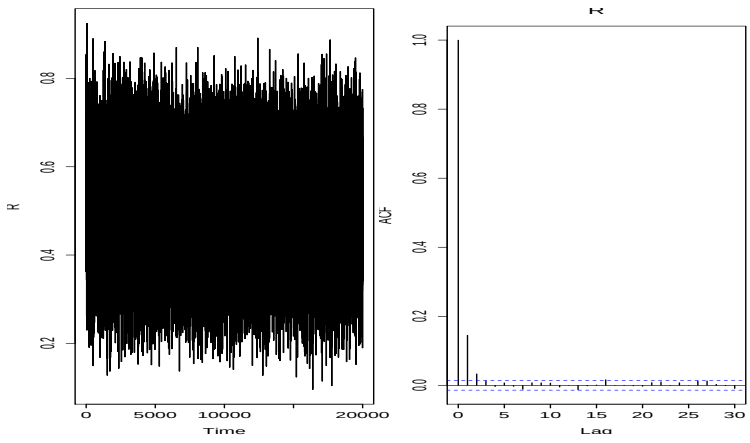


Figura 9: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando como priori a distribuição Beta.

Analisando a Tabela 8, podemos observar que as estimativas da confiabilidade obtidas foram similares, assim como a variância e os intervalos de credibilidade. De acordo com os resultados obtidos, utilizando as prioris Beta e Cópula, a probabilidade estimada do tempo de vida de determinada peça de equipamento ser maior do que trinta dias é 0,49.

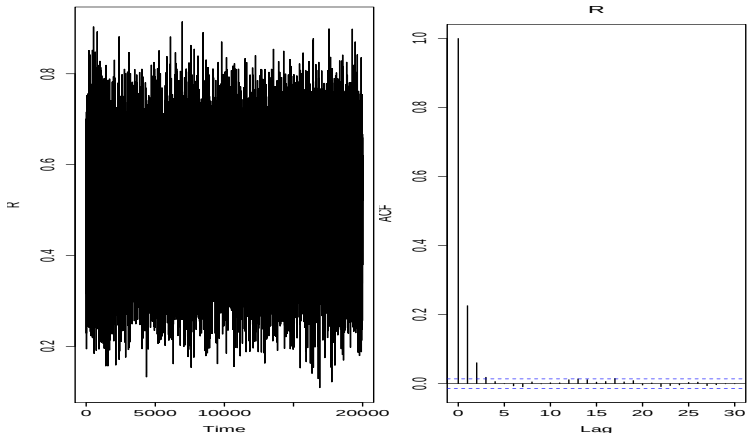


Figura 10: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando como priori a distribuição Log Gama Negativa.

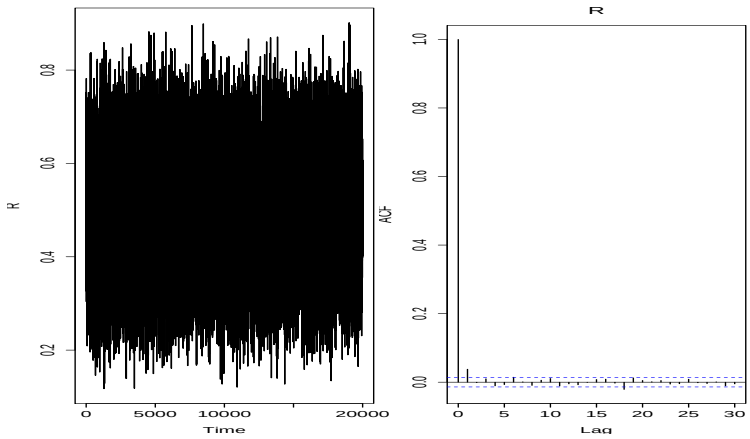


Figura 11: Cadeia gerada pelo MCMC para R e correspondente correlograma considerando a distribuição a priori Cópula.

A Figura 12 ilustra as densidades marginais a posteriori correspondentes as três distribuições a priori em estudo.

Lawless (1982) também discute esse exemplo. Ele ajustou esse conjunto de dados a uma distribuição Exponencial. Pelo método de máxima verossimilhança, a estimativa do parâmetro λ é 0,023 e seu intervalo de confiança de 95% é [0,012 ; 0,107]. A estimativa da confiabilidade R, para $t = 30$, é de 0,506 e o intervalo de confiança de 95%, é [0,07;0,676]. Observe que os resultados obtidos por esse autor são próximos aos obtidos com a análise Bayesina realizada acima.

O Capítulo 2 dessa dissertação resultou em um artigo intitulado *Bayesian estimation of reliability of the eletronic components using censored data from Weibull distribution: different prior distribution*, submetido à revista *International Journal of Industrial Engineering Production Research.*

Tabela 7: Estimativas a posteriori para R com $t = 30$ e intervalo de credibilidade de 95%

Estimativas	Beta	Log Gama Neg.	Cópula
Média	0,491	0,499	0,494
Variância	0,016	0,016	0,015
I.Cred.	[0,246 ; 0,740]	[0,261 ; 0,742]	[0,249 ; 0,739]

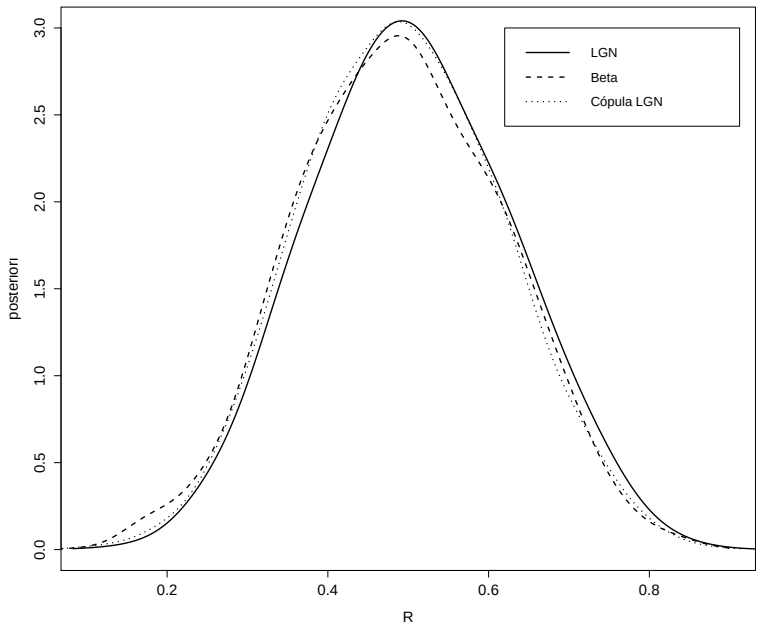


Figura 12: Densidades marginais a posteriori para o parâmetro R no instante $t = 30$.

3 Elicitação

A elicitação é o processo de transformação do conhecimento de um especialista sobre alguma quantidade desconhecida na forma de uma distribuição de probabilidade. No contexto da análise estatística Bayesiana, essa distribuição de probabilidade é geralmente utilizada como distribuição a priori para um ou mais parâmetros desconhecidos de um modelo estatístico.

A elicitação é uma parte importante do paradigma Bayesiano, porém, ainda pouco utilizada. Apesar de trazer grandes contribuições à análise estatística, muitos acreditam que a elicitação é inviável. Um dos fatores é a dificuldade em representar adequadamente a opinião do especialista na forma probabilística, outro é o vício algumas vezes encontrado nas respostas dos especialistas quando questionados sobre situações que envolvem incertezas.

O desafio no processo de elicitação consiste na elaboração de perguntas adequadas que permitam capturar o conhecimento do especialista e amenizem o possível vício de suas respostas. Além disso, é necessário uma fase de treinamento para que o especialista se familiarize com o formato do processo de elicitação.

Existem diversas aplicações das técnicas de elicitação no estudo de situações raras, ou seja, situações em que não existe uma quantidade relevante de dados disponível. Nesses casos, a opinião do especialista é fundamental no processo de decisão. Gustafson et al. (2003) elicitou a opinião de um especialista sobre mudança organizacional, nessa situação não existiam dados disponíveis. Dominitz (1998) elicitou a opinião de um especialista sobre futuras remunerações, novamente não existiam dados. Ang e Buttery (1997) elicitaram avaliações sobre a segurança de usinas nucleares. Laws e O'Hagan (2002) elicitaram a opinião de um especialista sobre erros no processo de auditoria financeira. Alguns outros exemplos podem ser encontrados em Grisley e Kellogg (1982) sobre aplicações na agricultura e Coolen et al. (1992) em aplicações na engenharia.

É assumido que o especialista apenas é capaz de definir algumas medidas de sua densidade, tais como mediana, moda, média e percentis. Para traduzir essas informações na forma de uma distribuição de probabilidade podemos utilizar tanto a metodologia

paramétrica quanto a não-paramétrica. Na elicitação paramétrica, o estatístico ajusta as informações provenientes do especialista a uma distribuição pertencente a alguma família de distribuições paramétricas. A elicitação não-paramétrica é um procedimento mais flexível e será abordada no Capítulo 6.

Esse capítulo está organizado da seguinte maneira: Na Seção 3.1, os estágios do processo de elicitação são descritos. Algumas técnicas de elicitação paramétrica são abordadas na Seção 3.2.

3.1 O Processo de Elicitação

Segundo Garthwaite, Kadane e O'Hagan (2005) o processo de elicitação pode ser dividido em quatro estágios.

1. Consiste na seleção e treinamento do especialista. O treinamento do especialista é realizado para investigar a precisão de suas especificações probabilísticas e familiarizá-lo com o procedimento estatístico, para aqueles que não o conhecem.

2. A elicitação das especificações probabilísticas do especialista. As especificações escolhidas podem ser momentos da distribuição ou qualquer outra quantidade que o estatístico acredita que o especialista possa ser capaz de fornecer. Nesse trabalho, alguns percentis serão elicitados nesse estágio do processo.

3. Ajuste de uma distribuição de probabilidade às especificações probabilísticas fornecidas pelo especialista. A distribuição ajustada é conhecida como densidade estimada do especialista, pois tal distribuição representa a opinião do especialista.

4. Conhecido como *feedback*, os resultados do ajuste são apresentados ao especialista que avalia se sua opinião está sendo bem representada. Se o especialista ficar insatisfeito com os resultados, então retorna-se ao estágio dois e elicita-se mais informações.

O estágio 2 tem sido foco de muitos trabalhos na área da psicometria. Apesar de existir uma vasta literatura sobre os aspectos psicométricos no processo de elicitação, não faz parte do objetivo desse trabalho tal estudo. Uma revisão sobre esses aspectos podem ser encontrada em Daneshkhah (2004).

3.2 Técnicas Paramétricas de Elicitação

Uma das suposições presentes na maior parte das técnicas de elicitação é a de que a densidade do especialista pertence a alguma família paramétrica. Assim, o foco do estudo

torna-se a elicitação dos hiperparâmetros da densidade de probabilidade selecionada.

Uma vez obtidas as especificações probabilísticas do especialista, o processo de elicitação se completa com o ajuste de uma distribuição a esses dados elicitados. A distribuição Uniforme é a distribuição mais simples que pode ser ajustada às crenças do especialista. Nesse caso, o especialista só precisa especificar um intervalo $[a, b]$ no qual ele acredita que o parâmetro esteja. Na maioria das vezes, essa distribuição é bastante criticada por ser muito simples. Ela atribui probabilidade zero para os valores que se encontram fora do intervalo especificado e mesma probabilidade para todos aqueles que pertencem ao intervalo. Assim, valores próximos a extremidade do intervalo possuem mesma probabilidade daqueles que se situam no centro do mesmo.

A distribuição triangular evita o problema de atribuir mesma probabilidade para valores centrais e extremos. Para essa distribuição, o especialista também precisa especificar a moda, denotada por m. Então a distribuição assumida possui densidade

$$f(x) = \begin{cases} 2\frac{x-a}{(b-a)(m-a)} & \text{se} \quad a \leq x \leq m \\ 2\frac{b-x}{(b-a)(b-m)} & \text{se} \quad m \leq x \leq b \end{cases}.$$

O ajuste dessa distribuição à crença do especialista pode ser melhor do que o encontrado com a distribuição uniforme, porém a probabilidade de um valor que não pertence ao intervalo ainda é zero.

Geralmente, métodos de elicitação mais complicados impõem uma estrutura sobre a opinião de um especialista, assumindo que seu conhecimento pode ser representado adequadamente por uma distribuição de probabilidade mais complexa do que a uniforme e a triangular.

Essas distribuições de probabilidade são controladas por um ou mais parâmetros (conhecidos como hiperparâmetros). Assim, o processo de elicitação consiste em escolher adequadamente valores para esses hiperparâmetros que consigam capturar as principais característica da opinião do especialista.

3.2.1 Ajustando a distribuição Beta

A distribuição Beta é bastante utilizada para representar o conhecimento do especialista para determinadas quantidades de interesse, por exemplo $R(t)$. Se X possui distribuição Beta com parâmetros l e k, então sua função densidade de probabilidade é dada por

$$f(x) = \frac{1}{B(l,k)} x^{l-1}(1-x)^{k-1} \qquad 0 < x < 1, \tag{3.1}$$

e

$$E(X) = \frac{l}{l+k} \quad \text{e} \quad Var(X) = \frac{lk}{(l+k)^2(l+k+1)}. \tag{3.2}$$

A distribuição Beta possui dois parâmetros que não devem ser elicitados diretamente, visto que é impraticável esperar que o especialista entenda a função dos parâmetros de uma distribuição de probabilidade. Porém, a média e a variância da distribuição Beta possuem relação direta com seus parâmetros. Assim, elicitando os valores da média e variância podemos determinar os valores de l e k. Contudo, a elicitação dos momentos de uma distribuição é muito criticada (veja Garthwaite, Kadane e O'Hagan, 2005).

Alguns pesquisadores como Beach e Swenson (1966), Peterson e Miller (1964) e Spencer (1961) investigaram a habilidade das pessoas em fornecer medidas de tendência central. Uma amostra de números foi exibida para alguns indivíduos para que estimassem a média, a moda e a mediana. Os pesquisadores observaram que quando a distribuição da amostra era aproximadamente simétrica, as estimativas dessas três medidas eram precisas. Porém, o mesmo não acontecia quando a distribuição da amostra era assimétrica. Por exemplo, os pesquisadores Peterson e Miller (1964) observaram que quando a distribuição era assimétrica à direita, as avaliações da mediana e da moda foram razoavelmente precisas, porém as avaliações da média foram tendenciosas.

A elicitação da variância é ainda mais criticada. Em estudos onde a estimativa da variância foi requisitada ao especialista, os pesquisadores perceberam que as pessoas têm muita dificuldade em entender o significado da variância e atribuir valores numéricos a ela. Assim, a variância não deve ser elicitada diretamente.

Um método apresentado em O'Hagan (1998) utilizado para ajustar qualquer distribuição à opinião do especialista, evita a elicitação da variância. Em um primeiro momento, esse método requer a elicitação da moda ($\hat{m}$), do limite inferior ($\hat{i}$) e do superior ($\hat{s}$) da distribuição do especialista. No próximo estágio as seguintes probabilidades são requeridas do especialista

$$\begin{aligned}
p_1 &= P(\hat{i} < X < \hat{m}), \\
p_2 &= P\left(\hat{i} < X < \frac{\hat{i}+\hat{m}}{2}\right), \\
p_3 &= P\left(\frac{\hat{m}+\hat{s}}{2} < X < \hat{s}\right), \\
p_4 &= P\left(\hat{i} < X < \frac{\hat{i}+3\hat{m}}{4}\right), \\
p_5 &= P\left(\frac{\hat{s}+3\hat{m}}{4} < X < \hat{s}\right).
\end{aligned}$$

As probabilidades foram requisitadas nessa ordem para minimizar o vício. Além disso, a escolha desses intervalos evita que o especialista avalie pequenas probabilidades. Para fornecer uma melhor elicitação do centro da distribuição, os intervalos são menores em torno da moda.

Posteriormente as respostas são convertidas em seis probabilidades, descritas como:

$$\begin{aligned}
q_1 = P\left(\hat{i} < X < \frac{\hat{i}+\hat{m}}{2}\right) &= p_2, \\
q_2 = P\left(\frac{\hat{i}+\hat{m}}{2} < X < \frac{\hat{i}+3\hat{m}}{4}\right) &= p_4 - p_2, \\
q_3 = P\left(\frac{\hat{i}+3\hat{m}}{4} < X < \hat{m}\right) &= p_1 - p_4, \\
q_4 = P\left(\hat{m} < X < \frac{3\hat{m}+\hat{s}}{4}\right) &= 1 - p_1 - p_5, \\
q_5 = P\left(\frac{3\hat{m}+\hat{s}}{4} < X < \frac{\hat{m}+\hat{s}}{2}\right) &= p_5 - p_3, \\
q_6 = P\left(\frac{\hat{m}+\hat{s}}{2} < X < \hat{s}\right) &= p_3.
\end{aligned}$$

Para o caso da distribuição Beta, temos $\hat{i} = 0$ e $\hat{s} = 1$. Para selecionar a distribuição Beta$(\hat{l}, \hat{k})$ que melhor representa a opinião do especialista, podemos utilizar o método descrito a seguir.

- São escolhidos n possíveis valores para $\hat{l}$ e w possíveis valores para $\hat{k}$.
- Apenas os pares $\hat{l}$ e $\hat{k}$ que satisfazem $\frac{\hat{l}_i - 1}{\hat{l}_i + \hat{k}_j - 2} = \hat{m}$, $i = 1, \ldots, n$ e $j = 1, \ldots, w$, são selecionados. Assim, teremos r pares de valores $\hat{l}$ e $\hat{k}$.
- Para os r pares, consideramos que $X \sim$Beta$(\hat{l}_u, \hat{k}_u)$, $u = 1, \ldots, r$ e calculamos $q_1, q_2, \ldots, q_6$.
- Para os r pares, a soma de quadrados das diferenças entre as probabilidades elicitadas e as probabilidades da distribuição Beta$(\hat{l}_u, \hat{k}_u)$ são calculadas.
- Os valores de $\hat{l}$ e $\hat{k}$ que possuem a menor soma de quadrados das diferenças são escolhidos como estimativas de l e k.

Uma revisão sobre métodos que podem ser utilizados para ajustar a distribuição Beta à opinião do especialista é feita em Hughes e Madden (2002). Entre os métodos descritos, podemos destacar o de Fox (1966) e Gross (1971).

No método de Fox (1966) o especialista é solicitado a fornecer uma estimativa para a moda de sua distribuição, denotada por $\hat{m}$ e a probabilidade p de X pertencer ao intervalo $[\hat{m} - Z\hat{m}, \hat{m} + Z\hat{m}]$, em que $0 < Z < 1$ é dado ao especialista. As estimativas dos parâmetros, l e k, são encontradas através das seguintes equações

$$\begin{aligned} \frac{\hat{l}-1}{\hat{l}+\hat{k}-2} &= \hat{m}, \\ \int_{\hat{m}-Z\hat{m}}^{\hat{m}+Z\hat{m}} \frac{1}{B(\hat{l},\hat{k})} x^{\hat{l}-1}(1-x)^{\hat{k}-1}dx &= p. \end{aligned}$$

Tais equações podem ser resolvidas da mesma forma utilizada no método proposto por O'Hagan, porém o número de probabilidades elicitadas é apenas um e não seis.

O método de Gross (1971) sugere que o especialista forneça a média de sua distribuição ($\overline{x}$) e a probabilidade p de X pertencer ao intervalo $[0, Z\overline{x}]$, em que $0 < Z < 1$ é dado ao especialista pelo estatístico. As equações obtidas são

$$\begin{aligned} \frac{\hat{l}}{\hat{l}+\hat{k}} &= \overline{x}, \\ \int_{0}^{Z\overline{x}} \frac{1}{B(\hat{l},\hat{k})} x^{\hat{l}-1}(1-x)^{\hat{k}-1}dx &= p. \end{aligned}$$

E também podem ser resolvidas da mesma forma utilizada no método proposto por O'Hagan, mas os valores de l e k selecionados serão aqueles que satisfazem $\frac{\hat{l}_i}{\hat{l}_i+\hat{k}_j} = \overline{x}$.

Exemplo 4: Suponha que a densidade verdadeira do especialista é a distribuição Beta com parâmetros $l = 4,45$ e $k = 2,05$, dada por

$$f(x) = \frac{1}{B(4,45, 2,05)} x^{3,45}(1-x)^{1,05}. \tag{3.3}$$

Assumimos que o especialista seja capaz de fornecer as especificações probabilísticas exatas de sua distribuição. Assim, é possível fazer a comparação dos métodos propostos por O'Hagan, Fox e Gross.

Utilizando o método de O'Hagan e a densidade (3.3) temos que

$$\begin{aligned} \hat{m} &= 0,77, \\ q_1 = \int_{0}^{\frac{\hat{m}}{2}} f(p)dp &= 0,06, \\ q_2 = \int_{\frac{\hat{m}}{2}}^{\frac{3\hat{m}}{4}} f(p)dp &= 0,20, \\ q_3 = \int_{\frac{3\hat{m}}{4}}^{\hat{m}} f(p)dp &= 0,39, \end{aligned}$$

$$
\begin{aligned}
q_4 = \int_{\hat{m}}^{\frac{3\hat{m}+1}{4}} f(p)dp &= 0,13, \\
q_5 = \int_{\frac{3\hat{m}+1}{4}}^{\frac{\hat{m}+1}{2}} f(p)dp &= 0,11, \\
q_6 = \int_{\frac{\hat{m}+1}{2}}^{1} f(p)dp &= 0,11.
\end{aligned}
$$

Utilizando esse método, os parâmetros da distribuição Beta selecionados foram $\hat{l} = 4,43$ e $\hat{k} = 2,05$.

Utilizando o método de Fox e a densidade apresentada em (3.3) temos

$$\hat{m} = 0,77, \qquad \int_{\frac{9\hat{m}}{10}}^{\frac{11\hat{m}}{10}} f(p)dp = 0,34,$$

com $Z = 0,1$. Os parâmetros escolhidos foram $\hat{l} = 4,08$ e $\hat{k} = 1,9$.

Utilizando o método de Gross e a densidade apresentada em (3.3) temos

$$\overline{x} = 0,68, \qquad \int_{0}^{\frac{\overline{x}}{2}} f(p)dp = 0,45,$$

para $Z = 0,5$. Os parâmetros da distribuição Beta selecionados foram $\hat{l} = 2,29$ e $\hat{k} = 1,1$.

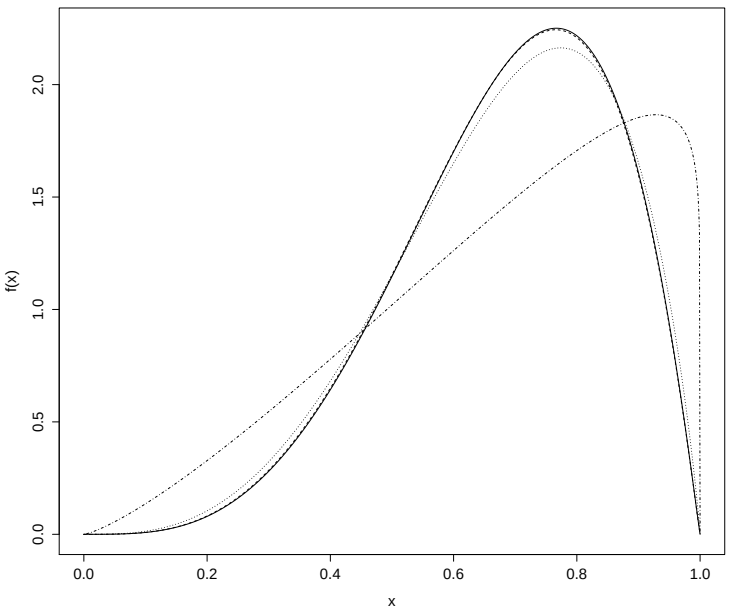


Figura 13: A verdadeira função densidade do especialista (linha sólida), a função densidade resultante do método de O'Hagan (linha tracejada), a função densidade resultante do método de Fox (linha pontilhada) e a função densidade resultante do método de Gross (linha tracejada e pontilhada).

Observando a Figura 13 vemos que o método de Gross apresentou o pior desempenho. Já os métodos de O'Hagan e Fox produziram melhores resultados, as duas distribuições têm a mesma moda da distribuição verdadeira do especialista. Além disso, a assimetria de ambas são similares à assimetria da distribuição verdadeira.

3.2.2 Outros Métodos de Elicitação

Outras três técnicas de elicitação da distribuição de probabilidade do especialista são descritas em Winkler (1967).

1. O primeiro método é conhecido como método da amostra a priori equivalente (*equivalent prior sample method - EPS*). O especialista é solicitado a determinar n e s, tal que seu conhecimento seria aproximadamente equivalente a ter observado, por exemplo, exatamente "s"ursos fêmeas em uma amostra aleatória de tamanho "n"de ursos de determinada espécie. Então a distribuição a priori é uma Beta com parâmetros "s"e "$n - s$".

2. O método da função distribuição acumulada (*cumulative distribution function method - CDF*) requer que o especialista forneça diversas especificações probabilísticas, como a mediana e alguns percentis de sua distribuição. Esses valores são representados graficamente e uma suave distribuição acumulada pode ser retirada através deles. Então o *feedback* é realizado com o especialista para verificar se o mesmo concorda com os valores de outros percentis não elicitados, comprovando assim a adequação da curva ajustada com sua opinião.

3. No método da função densidade de probabilidade (*probability density function method - PDF*) é solicitado ao especialista a moda da sua distribuição e alguns pontos próximos a ela. Posteriormente, a densidade do especialista é esboçada unindo-se os pontos elicitados. Segundo Goshing (2005) a validade desse método é questionável, por exigir que o especialista possua um grande conhecimento em estatística.

Enquanto o método da amostra a priori equivalente é geralmente restrito a elicitação de proporções, os métodos da função distribuição acumulada e da função densidade de probabilidade podem ser utilizados para elicitar a distribuição de qualquer variável aleatória contínua.

Outro método discutido na literatura é o da bisecção. Segundo O'Hagan, Garthwaite e Kadane (2005) esse método consiste na seguinte sequência de perguntas:

1. Seja X a quantidade desconhecida de interesse. Determine um valor (a mediana do especialista) tal que X é igualmente provável ser maior ou menor do que esse ponto.

2. Supondo que o valor de X está abaixo da mediana especificada anteriormente, determine um novo valor (quartil inferior) tal que é igualmente provável que X seja menor ou maior do que esse valor.

3. Supondo que o valor de X está acima da mediana especificada na pergunta 1, determine um novo valor (quartil superior) tal que seja igualmente provável que X seja menor ou maior do que esse valor.

4 *Elicitação em Análise de Confiabilidade*

No contexto de análise de confiabilidade, aplicações da elicitação podem ser encontradas em Martz e Waller (1982), Singpurwalla (1988) e Coolen (1992).

Existem algumas técnicas de elicitação aplicadas à análise de confiabilidade. A técnica de elicitação, proposta por Weiler (1965), para a seleção de uma distribuição Beta é descrita na Seção 4.1. Uma outra técnica, proposta nessa dissertação, que utiliza a elicitação da função de confiabilidade da distribuição Weibull para determinação dos hiperparâmetros da distribuição conjunta a priori é apresentada na Seção 4.2.

4.1 Elicitação da distribuição Beta

O método apresentado por Weiler (1965) é um dos métodos capazes de utilizar as informações sobre média e percentis para a seleção da distribuição Beta (parâmetros l e k) como priori.

Essa técnica é válida nas seguintes situações:

- Quando a esperança da confiabilidade, $E(R)$, e a probabilidade p,

$$P\Big(R > E(R)\Big) = p,$$

são conhecidas.

- Quando as probabilidades, p_a e p_b, dado R_a e R_b,

$$P(R > R_a) = p_a \quad e \quad P(R < R_b) = p_b$$

são conhecidas.

A técnica é implementada através dos seguintes passos:

1. Defina a média a priori da confiabilidade R_1 ($R_1 = E(R)$);

2. Determine o 95° percentil, denotado por R_2. Em outras palavras, determine o valor R_2 tal que, antes da execução do teste, existe uma chance de 5% da confiabilidade R ser maior do que R_2, ou seja, $P(R > R_2) = 0.05$;

3. Determine o 5° percentil, denotado por R_3, $P(R < R_3) = 0.05$. Ou seja, indique o valor R_3 tal que, antes da realização do teste, existe uma chance de 5% da confiabilidade R ser menor do que R_3.

4. Utilize a Tabela C_1 (disponível em Martz e Waller, 1982), e o par de valores (R_1, R_2) para determinar l e k.

5. Utilize a Tabela C_2 (disponível em Martz e Waller, 1982), e o par de valores (R_1, R_3) para determinar l e k.

Geralmente, qualquer par de valores, (R_1, R_2) ou (R_1, R_3) são suficientes para determinar unicamente os parâmetros l e k.

Exemplo 5: Suponha que a média a priori da confiabilidade supostamente seja $0,80$. E acredita-se que a chance da confiabilidade verdadeira a priori exceda 0,98 é de 5%. Então

$$R_1 = 0,80 \quad e \quad R_2 = 0,98. \tag{4.1}$$

A Tabela C_1 mostra que os correspondentes parâmetros são $l = 4,95117$ e $k = 6,18896$. Assim, $R \sim$Beta$(4,95117, 6,18896)$.

Exemplo 6: Suponha que a média a priori da confiabilidade supostamente seja $0,96$. E acredita-se que a chance da confiabilidade verdadeira a priori ser menor do que 0,75 é de 5%. Então

$$R_1 = 0,96 \quad e \quad R_3 = 0,75. \tag{4.2}$$

De acordo com a Tabela C_2, a distribuição Beta com parâmetros $l = 2,55812$ e $k = 2,66470$ é a escolhida como distribuição a priori.

4.2 Elicitação da Confiabilidade da distribuição Weibull

Nesta seção propomos um método alternativo para elicitação da confiabilidade. A metodologia proposta aqui pode ser aplicada à qualquer outra distribuição. Supondo que os dados são provenientes da distribuição Weibull e assumindo distribuições a priori

Gamas para seus parâmetros, elicitamos alguns valores da confiabilidade e então, com base nessas informações, determinamos os hiperparâmetros da distribuição a priori conjunta.

Como já exposto no Capítulo 2, a distribuição Weibull é uma das distribuições mais utilizadas na análise de confiabilidade. Sua função densidade é dada por

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\},$$

dependendo dos parâmetros α e β.

Para a distribuição Weibull, a função de confiabilidade é dada por

$$R(t) = exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}.$$

Sua função de verossimilhança pode ser escrita como

$$L(\alpha, \beta|\mathbf{x}) = \left(\frac{\beta}{\alpha}\right)^{n} \prod_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\beta-1} exp\left\{-\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta}\right\}. \tag{4.3}$$

Pode-se assumir que α e β são independentes e com uma distribuição a priori Gama(a, b) e Gama(c, d), respectivamente. Dessa forma, a distribuição conjunta a priori é dada por

$$\pi(\alpha, \beta) = \frac{bd}{\Gamma(a)\Gamma(b)}(b\alpha)^{a-1}(d\beta)^{c-1}exp\{-(b\alpha + d\beta)\}, \tag{4.4}$$

em que a, b, c e d são os hiperparâmetros. Se não existe informação previamente disponível, os hiperparâmetros podem assumir o valor 0,01.

A distribuição a posteriori é então dada por

$$p(\alpha, \beta|\mathbf{x}) \propto \left(\frac{\beta}{\alpha}\right)^{n} \prod_{i=1}^{n} \left(\frac{x_i}{\alpha}\right)^{\beta-1} exp\left\{-\sum_{i=1}^{n}\left(\frac{x_i}{\alpha}\right)^{\beta}\right\} (b\alpha)^{a-1}(d\beta)^{c-1}exp\{-(b\alpha + d\beta)\}. \tag{4.5}$$

4.2.1 O método

Alguns autores propuseram métodos de elicitação de distribuições conjuntas a priori para α e β, como Kaminskiy e Krivtsov (2005). Porém, a maioria desses métodos requer a elicitação de momentos de segunda ordem ou a elicitação de hiperparâmetros, o que torna sua aplicação prática inviável.

Também é impraticável esperar que o especialista seja capaz de estabelecer valores da densidade a priori, pois sabemos que não é fácil responder perguntas como "qual é

a probabilidade da variável de interesse assumir um certo valor". O mais razoável seria pedirmos probabilidades, como $P(\theta \leq x)$, ou seja, perguntarmos "qual o valor que a variável poderia assumir para uma certa probabilidade"(Qing,1998).

Dessa forma, no processo de elicitação da confiabilidade $R(t)$, pedimos ao especialista um domínio de valores do qual ele acredita que R pertence com uma probabilidade especificada. Por exemplo, se a probabilidade dada é 0,25, então o especialista fornece o valor de "t"para qual ele acredita que existe uma chance de 25% do tempo de vida do item em estudo ser maior do que "t".

Assim, k valores $R_{t_1}, R_{t_2}, \ldots, R_{t_k}$ são escolhidos e então o especialista fornece os correspondentes $t_1, t_2, \ldots, t_k$, tal que

$$R_{t_j} = P(T \geq t_j) \qquad\qquad j = 1, \ldots, k. \tag{4.6}$$

A escolha das probabilidades que serão pedidas é uma tarefa importante. Os percentis geralmente utilizados na literatura são $10^o, 25^o, 50^o$ e 75^o percentis.

Segundo Kadane e Wolfson(1998), consideramos que o especialista é capaz de fornecer valores da confiabilidade, incondicionais a α e β. Assim, assumimos que os valores da confiabilidade provenientes do especialista devem ser definidos em relação a densidade preditiva a priori, ou seja, a densidade a priori dos plausíveis tempos de vida,

$$\begin{aligned} P(T \geq t_i) &= \int_0^\infty \int_0^\infty P(T \geq t_i|\alpha, \beta)\pi(\alpha, \beta)d\alpha d\beta \\ &= \frac{bd}{\Gamma(a)\Gamma(b)} \int_0^\infty \int_0^\infty (b\alpha)^{a-1}(d\beta)^{c-1} exp\{-(b\alpha + d\beta)\} exp\left\{-\left(\frac{t_i}{\alpha}\right)^\beta\right\} d\alpha d\beta \end{aligned} \tag{4.7}$$

com $i = 1, \ldots, 4$. Observe que haverá quatro expressões (4.7), uma para cada percentil elicitado.

Imagine que uma particular escolha de $P(T \geq t_i)$ resultará em (4.7), correspondendo às crenças que o especialista possui sobre a confiabilidade. Como R, apresentado em (2.3), é função de α e β, a escolha de $P(T \geq t_i)$ descreve implicitamente a crença do especialista sobre os parâmetros α e β.

Utilizando a aproximação de Laplace (ver Kadane e Tierney, 1986) na expressão (4.7), que não possui solução analítica, obtém-se

$$R(t_i) = 2\pi \frac{bd}{\Gamma(a)\Gamma(b)} exp\left\{-\left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\beta}}\right\} (b\hat{\alpha})^{a-1}(d\hat{\beta})^{c-1} exp\{-(b\hat{\alpha} + d\hat{\beta})\} \frac{\hat{\alpha}\hat{\beta}}{\sqrt{(a-1)(c-1)}}, \tag{4.8}$$

em que $\hat{\alpha} = \frac{a-1}{b}$ e $\hat{\beta} = \frac{c-1}{d}$ com $a > 1$, $c > 1, i = 1, \ldots, 4$. Dessa forma, obtém-se um sistema não linear de quatro equações e quatro incógnitas a, b, c e d. Assim, é possível determinar os hiperpâremetros da distribuição a priori apresentada em (4.4). Esse sistema não linear pode ser resolvido, no software R, através do pacote "BB"com o comando "BBsolve".

Exemplo 7: Para ilustrar o método proposto, foram geradas amostras de uma distribuição Weibull com parâmetros $\alpha = 4$ e $\beta = 1,3$.

Para ilustrar o caso onde existe informação de um especialista disponível, os percentis $10^o, 25^o, 50^o$ e 75^o foram elicitados utilizando-se o quantil de uma distribuição Weibull com $\alpha = 4$ e $\beta = 1,3$. Assim, as probabilidades obtidas são exatas e o valor dos hiperparâmetros da distribuição conjunta a priori definida em (4.4) são: $a = 8,81, b = 1,95, c = 2,14$ e $d = 0,87$. Porém, é muito difícil que o especialista, apesar de conhecer bastante sobre o assunto, consiga fornecer as probabilidades exatas. Dessa forma, atribui-se um erro aleatório $N(0,1)$ para cada uma das probabilidades. Nesse caso, os valores dos hiperpârametros são: $a = 4,37, b = 0,91, c = 2,27$ e $d = 1,24$.

Para mostrar o caso em que não existe informação prévia, os hiperparâmetros escolhidos são $a = b = c = d = 0,01$, pois sabe-se que a distribuição a priori dada em (4.4) torna-se não informativa.

Foram geradas mil amostras de tamanho 3, mil amostras de tamanho 6 e mil amostras de tamanho 10 de uma distribuição Weibull com parâmetros $\alpha = 4$ e $\beta = 1,3$. Para cada uma das três mil amostras consideradas, uma cadeia com 10000 iterações foi gerada no pacote estatístico R, com o objetivo de simular amostras dos valores de α, β e R. As exigências de convergência foram checadas através dos gráficos de autocorrelação e série de tempo. A Tabela 9 mostra as estimativas dos parâmetros desconhecidos, suas variâncias, erro quadrático médio e a média dos intervalos de credibilidade de 95% para os diferentes tamanhos de amostra.

As estimativas, onde a elicitação e a elicitação com erro foram utilizadas, aproximaram-se mais do verdadeiro valor quando a amostra é de tamanho 3 e apresentaram uma variância menor se comparadas com aquelas obtidas utilizando priori não informativa, como era esperado. À medida que o n cresce, a performance das prioris assemelha-se.

Podemos estimar a função de confiabilidade utilizando os três tipos de prioris - elicitação, elicitação com erro e não informativa - e sabendo que a função de confiabilidade verdadeira é dada por

$$R(t) = exp\left\{-\left(\frac{t}{4}\right)^{1,3}\right\}, \tag{4.9}$$

é possível comparar as funções de confiabilidade estimadas com a verdadeira.

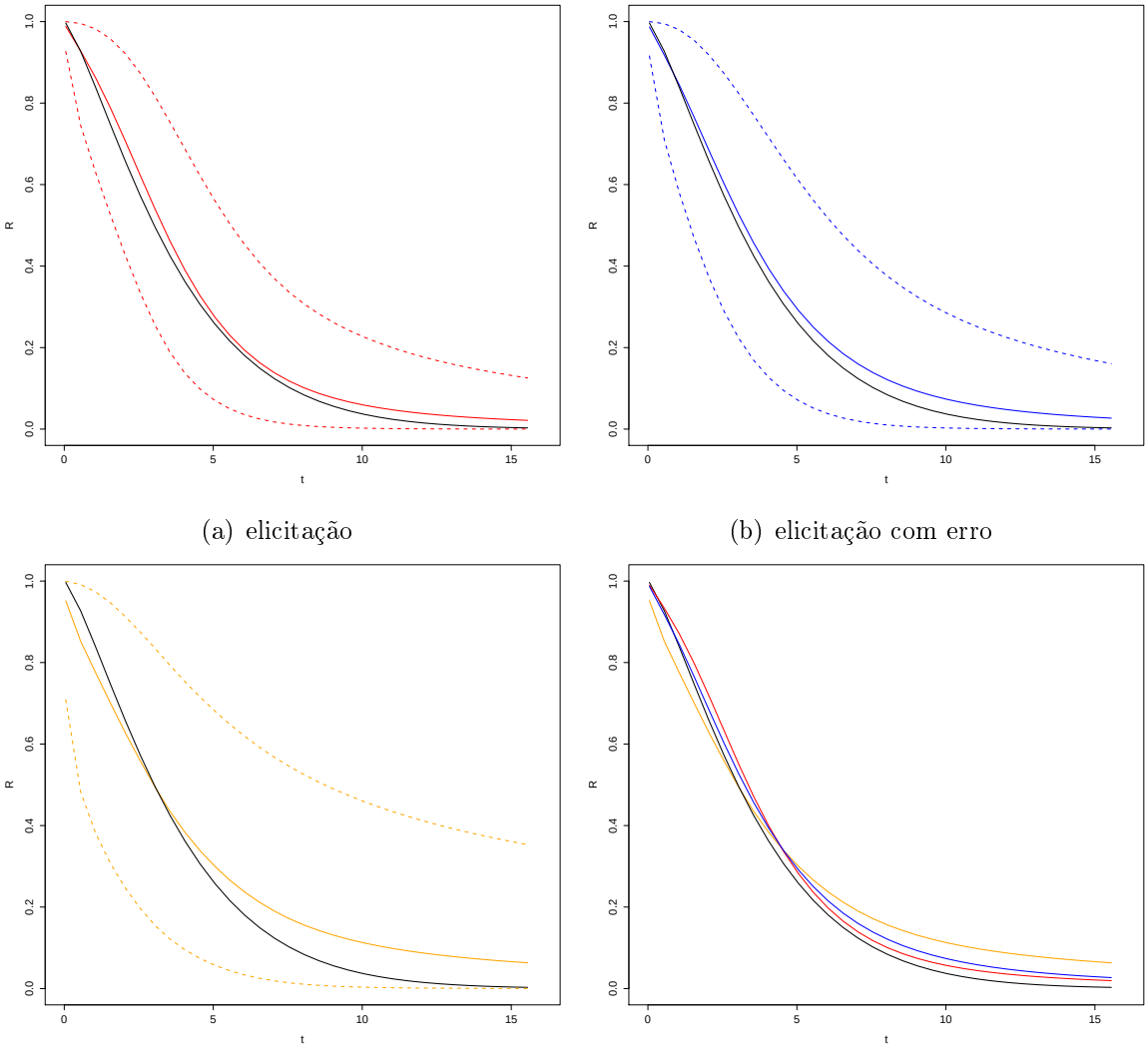
Tabela 8: Estimativas dos parâmetros $\alpha = 4$, $\beta = 1,3$.

amostra	Método		Média	Var	EQM	I.Cred.
n=3	Elicitação	$\hat{\alpha}$	4,37	1,34	0,83	[2,84 ; 6,10]
		$\hat{\beta}$	1,71	0,60	0,51	[0,84 ; 3,11]
	Elicitação com erro	$\hat{\alpha}$	4,45	2,31	1,36	[2,46 ; 6,67]
		$\hat{\beta}$	1,93	0,79	1,00	[0,87 ; 3,92]
	Não Informativa	$\hat{\alpha}$	5,62	82,98	11,92	[1,69 ; 11,37]
		$\hat{\beta}$	2,02	3,80	6,79	[0,55 ; 07,28]
n=6	Elicitação	$\hat{\alpha}$	4,37	1,02	0,78	[2,86 ; 6,06]
		$\hat{\beta}$	1,64	0,28	0,42	[0,94 ; 2,98]
	Elicitação com erro	$\hat{\alpha}$	4,38	1,02	0,78	[2,87 ; 6,06]
		$\hat{\beta}$	1,64	0,28	0,42	[0,93 ; 2,96]
	Não Informativa	$\hat{\alpha}$	4,56	6,95	2,40	[2,12 ; 7,82]
		$\hat{\beta}$	1,49	0,35	6,79	[0,75 ; 3,13]
n=10	Elicitação	$\hat{\alpha}$	4,27	0,75	0,54	[2,98 ; 5,78]
		$\hat{\beta}$	1,52	0,14	0,22	[0,96 ; 2,48]
	Elicitação com erro	$\hat{\alpha}$	4,28	1,00	0,73	[2,82 ; 6,03]
		$\hat{\beta}$	1,47	0,14	0,17	[0,93 ; 2,32]
	Não Informativa	$\hat{\alpha}$	4,21	1,77	1,05	[2,51 ; 6,49]
		$\hat{\beta}$	1,42	0,16	0,21	[0,85 ; 2,44]

Para cada um dos valores das cadeias de α e β, utilizando (4.9), calculamos a confiabilidade R para diversos valores de t. Calculando a média das estimativas da confiabilidade para cada t, obtivemos a função de confiabilidade estimada (linha sólida e colorida), o intervalo de credibilidade de 95%(linha tracejada e colorida) e a função de confiabilidade verdadeira (linha sólida e preta) para amostras de tamanho 3, 6 e 10 apresentadas nas Figuras 14, 15 e 16.

Para $n = 3$, a função de confiabilidade estimada que utiliza a elicitação é a que mais se aproxima da função verdadeira. Ela possui também o menor intervalo de credibilidade. Para $n = 6$, as funções de confiabilidade estimadas tornaram-se mais parecidas umas com as outras. E os intervalos de credibilidade se aproximaram mais da função de confiabilidade estimada. Já para $n = 10$, as funções de confiabilidade estimadas são muito similares.

O processo de elicitação proposto aqui melhora a análise do tempo de vida, principalmente quando a amostra é pequena. Para amostras grandes, em que temos bastante informações provenientes dos dados, a contribuição da elicitação não é tão perceptível.



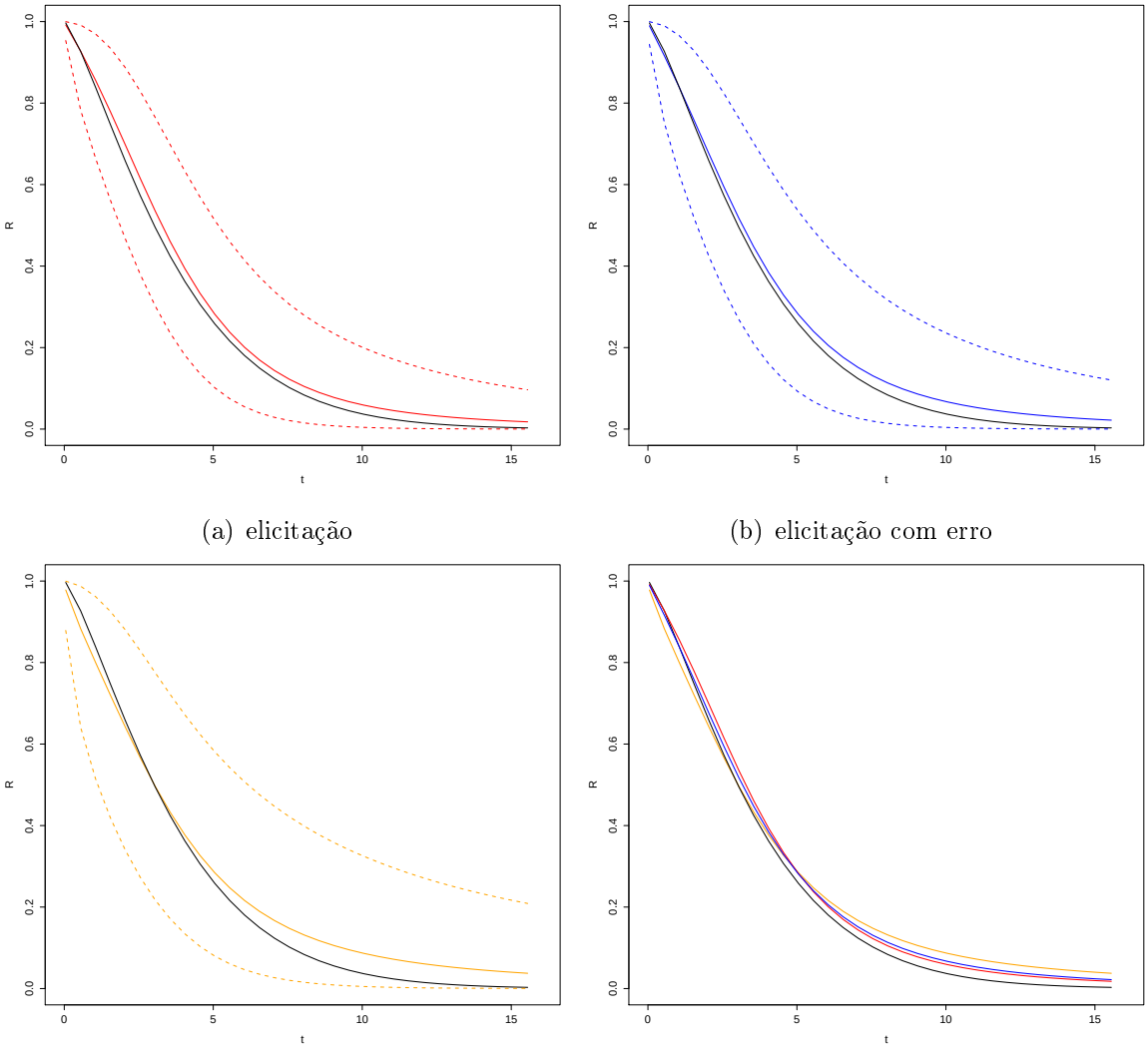
(a) elicitação

(b) elicitação com erro

(c) não informativa

(d) funções de confiabilidade

Figura 14: Funções de confiabilidade estimadas (sólidas e coloridas), intervalo de credibilidade de 95% (tracejadas e coloridas) e função de confiabilidade verdadeira (sólida e preta) para n=3.



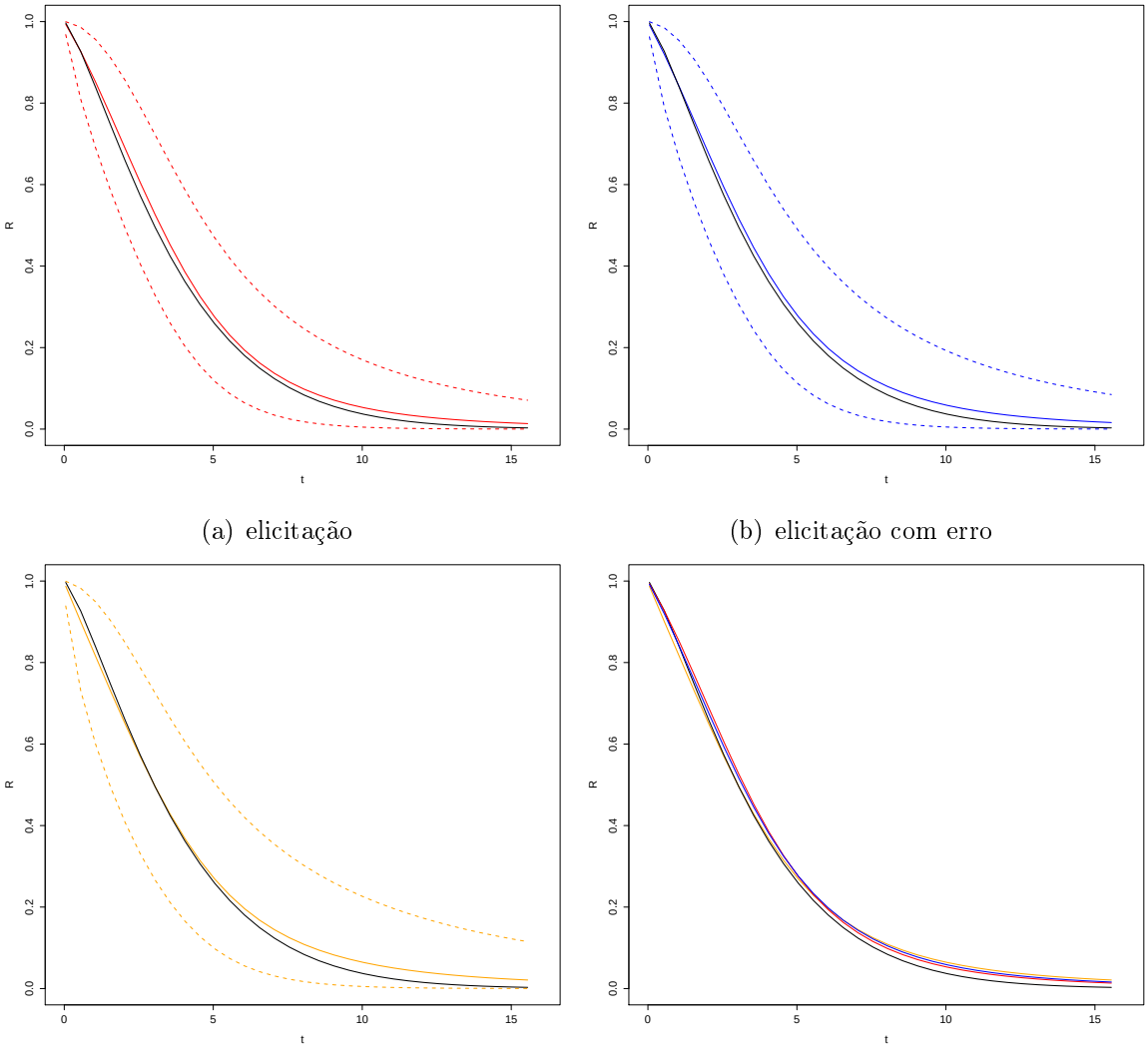
(a) elicitação

(b) elicitação com erro

(c) não informativa

(d) funções de confiabilidade

Figura 15: Funções de confiabilidade estimadas (sólidas e coloridas), intervalo de credibilidade de 95% (tracejadas e coloridas) e função de confiabilidade verdadeira (sólida e preta) para n=6.



(a) elicitação

(b) elicitação com erro

(c) não informativa

(d) funções de confiabilidade

Figura 16: Funções de confiabilidade estimadas (sólidas e coloridas), intervalo de credibilidade de 95% (tracejadas e coloridas) e função de confiabilidade verdadeira (sólida e preta) para n=10.

5 *Inferência Bayesiana Não-Paramétrica*

A inferência paramétrica restringe a densidade (ou qualquer outra função) que desejamos estimar a uma família de dimensão finita, geralmente a família exponencial, que pode ser completamente especificada por alguns parâmetros. Dessa forma, a inferência sobre essa distribuição se reduz a um problema de estimação de parâmetros (MOALA, 2006).

Apesar da inferência paramétrica ser exaustivamente utilizada nas análises estatísticas, as suposições inerentes a essa metodologia limitam drasticamente os tipos de densidades que podem ser representadas. Por exemplo, algumas suposições para o uso da distribuição normal são a unimodalidade e a simetria da função de interesse. Assim, quando a função desconhecida f possui características como multi-modalidade, assimetria, caudas pesadas, sensibilidade a *outliers*, entre outros, o uso da tradicional modelagem paramétrica torna-se difícil ou até mesmo inaceitável.

A necessidade de uma modelagem mais flexível fez com que a inferência Bayesiana não-paramétrica fosse bastante difundida nos últimos anos. Isso se deve a sua capacidade em lidar com complexos problemas e produzir boas inferências que não dependem de especificações paramétricas restritivas. Segundo Rasmussen e William (2006), ao invés de restringir a função desconhecida f a uma classe de funções, como é feito na inferência paramétrica, a inferência não-paramétrica é uma abordagem que (falando de maneira pouco rigorosa) fornece uma probabilidade a priori para todas as possíveis funções, em que altas probabilidades são atribuídas a funções mais prováveis. Assim, funções muito diferentes de nossa convicção a priori não são excluídas, são apenas menos prováveis.

A metodologia não-paramétrica permite uma modelagem flexível da função desconhecida f que pode ser a função de sobrevivência, a função de risco, a função densidade de probabilidade, entre outras. A única suposição existente é de que a função f pertence a uma apropriada classe de funções denotada por $\mathfrak{C}$.

Sob o enfoque Bayesiano, a função f é considerada aleatória. Uma função aleatória é uma função escolhida aleatoriamente de uma família de possíveis funções. A combinação

da metodologia Bayesiana com a estatística não-paramétrica requer a construção de prioris no espaço das funções. Uma priori sobre o espaço das funções pode ser vista como um processo estocástico que assume valores no espaço das funções dado (CHOUDHURI et al, 2005 apud MOALA, 2006).

Diversos tipos de processos estocásticos podem ser utilizados como prioris no espaço de funções. Dentre todos os processos aleatórios, o processo Gaussiano é um dos mais simples e atrativos por possuir características muito úteis na modelagem Bayesiana.

Esse capítulo está organizado da seguinte forma: Na Seção 5.1, são apresentadas algumas noções de processos estocásticos. O processo Gaussiano é discutido na Seção 5.2. Na Seção 5.3 são abordadas algumas funções de covariância que podem ser especificadas em um processo Gaussiano. A utilização da inferência Bayesiana sobre um processo Gaussiano é abordada na Seção 5.4.

5.1 Processos Estocásticos

Segundo Moala (2006), a inferência Bayesiana não-paramétrica refere-se a modelos probabilísticos no espaço de funções, onde as funções desconhecidas são tratadas como funções aleatórias. Assim, é necessário que as distribuições de probabilidade pertençam a uma família de funções em que a distribuição a priori e a posteriori sejam processos estocásticos.

A noção de um processo estocástico é uma extensão do conceito de variável aleatória. O comportamento de uma variável ou vetor aleatório é descrito por uma distribuição de probabilidade. Já as propriedades de funções aleatórias (ou seja, uma distribuição de probabilidade de funções) são descritas por um processo estocástico.

Definição 1: Seja um experimento aleatório definido em um espaço de probabilidade $(\Omega, P, \Im)$. Suponha também que para cada evento ω em Ω, de acordo com certa regra, associamos uma função

$$f(\mathbf{t}, \omega) \qquad \mathbf{t} \in T \subset R^n \tag{5.1}$$

A Figura 17 ilustra a interpretação de um processo estocástico visto como uma família de variáveis aleatórias. Note que a trajetória observada empiricamente não precisa necessariamente passar pelas médias das distribuições probabilísticas, passando ao invés disto por pontos aleatórios pertencentes a distribuição probabilística representativa da variável no dado instante.

Segundo Papoulis (1965), um processo estocástico pode ser visto como uma função de duas variáveis t e ω. Para t fixo, $f(t, \omega)$ é uma variável aleatória com uma distribuição de

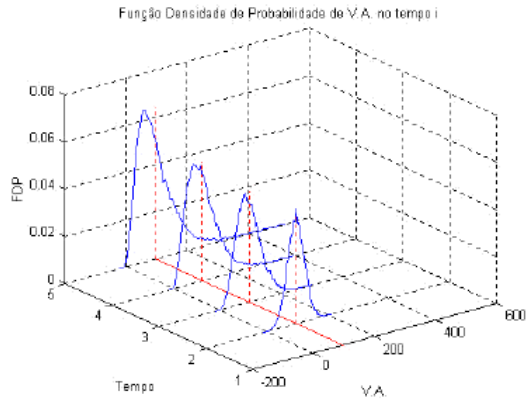


Figura 17: Processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: Camposilvan (2003)

probabilidade. Em contrapartida, se ω é fixo, $f(t, \omega)$ é uma função do tempo e é chamada de função amostral ou realização do processo. Quando t e ω são fixados, $f(t, \omega)$ é um número real.

Dada uma coleção finita de tempos t_1, $t_2, \ldots, t_n$ então $f(t_1), f(t_2), \ldots, f(t_n)$ constituem um conjunto de variáveis aleatórias com distribuição conjunta de probabilidade. O propósito do estudo de processos estocásticos é a especificação de uma distribuição conjunta de probabilidade adequada ao fenômeno em estudo para que se possa prever o comportamento do processo futuro.

Existem diversos tipos de processos estocásticos como o processo Gaussiano, o processo de Poisson, o processo Markoviano, entre outros. O processo de interesse nesse trabalho é o Gaussiano e será apresentado com mais detalhes na próxima seção. Informações sobre os demais processos podem ser encontrados em Papoulis (1965).

5.2 Processo Gaussiano

O processo Gaussiano é um dos processos aleatórios mais importantes encontrados na literatura e suas aplicações estendem-se a diversas áreas da pesquisa. Isso ocorre porque esse processo é capaz de representar diversos fenômenos da natureza apropriadamente. Além disso, a facilidade de manipulação matemática de um processo Gaussiano torna-o atrativo nas pesquisas.

As técnicas que envolvem o processo Gaussiano na realização de inferências sobre uma função desconhecida estão cada vez mais populares. Esses métodos foram introduzidos por O'Hagan (1978), entre outros.

5.2.1 Propriedades da distribuição Normal

Antes de iniciarmos a discussão sobre o processo Gaussiano é importante apresentarmos algumas propriedades da distribuição Normal multivariada que nos auxiliarão no entendimento de resultados referentes a esse processo.

A densidade de probabilidade da distribuição Normal multivariada pode ser escrita como

$$g(\mathbf{x}) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T|\Sigma|^{-1}(\mathbf{x}-\mathbf{m})\right\}, \tag{5.2}$$

em que $\mathbf{m}$ é o vetor de média de dimensão $n \times 1$ e Σ é a matriz de covariância (simétrica e positiva definida) de dimensão $n \times n$.

Considere a partição $(\mathbf{X}_1, \mathbf{X}_2)$ do vetor $\mathbf{X}$, então

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}(x_1) \\ \boldsymbol{\mu}(x_2) \end{pmatrix}, \begin{pmatrix} \Sigma_{x_1x_1} & \Sigma_{x_1x_2} \\ \Sigma_{x_2x_1} & \Sigma_{x_2x_2} \end{pmatrix}\right]. \tag{5.3}$$

A distribuição condicional de $\mathbf{X}_2$ dado $\mathbf{X}_1$ é

$$\Big(\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1\Big) \sim N\Big(\boldsymbol{\mu}(x_2) + \Sigma_{x_2x_1}\Sigma_{x_1x_1}^{-1}\Big(\mathbf{x}_1 - \boldsymbol{\mu}(x_1)\Big), \Sigma_{x_2x_2} - \Sigma_{x_2x_1}\Sigma_{x_1x_1}^{-1}\Sigma_{x_1x_2}\Big). \tag{5.4}$$

A demonstração completa desse resultado pode ser encontrada em Press (1972).

Teorema 2: Se $X|\boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \Sigma)$ sendo Σ conhecido e $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \Sigma_0)$, então $\boldsymbol{\mu}|\mathbf{x} \sim N(\boldsymbol{\mu}_1, \Sigma_1)$ em que

$$\boldsymbol{\mu}_1 = \Sigma_1(\Sigma^{-1}\mathbf{x} + \Sigma_0^{-1}\boldsymbol{\mu}_0) \quad e \quad \Sigma_1^{-1} = \Sigma^{-1} + \Sigma_0^{-1}.$$

Demonstração: $p(\boldsymbol{\mu}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\mu}, \Sigma)p(\boldsymbol{\mu})$

$$\begin{aligned}
&\propto (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}(2\pi)^{-\frac{n}{2}}|\Sigma_0|^{-\frac{1}{2}} \times \\
&\times exp\left\{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^t\Sigma_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right\} \\
&\propto exp\left\{-\frac{1}{2}\Big[(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) + (\boldsymbol{\mu}-\boldsymbol{\mu}_0)^t\Sigma_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\Big]\right\} \\
&= exp\left\{-\frac{1}{2}\Big(\mathbf{x}^t\Sigma^{-1}\mathbf{x} - 2\mathbf{x}^t\Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^t\Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^t\Sigma_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^t\Sigma_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^t\Sigma_0^{-1}\boldsymbol{\mu}_0\Big)\right\} \\
&\propto exp\left\{-\frac{1}{2}\Big[\boldsymbol{\mu}^t(\underbrace{\Sigma^{-1} + \Sigma_0^{-1}}_{\Sigma_1^{-1}})\boldsymbol{\mu} - 2\boldsymbol{\mu}^t(\Sigma^{-1}\mathbf{x} + \Sigma_0^{-1}\boldsymbol{\mu}_0)\Big]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= exp\left\{-\frac{1}{2}\Big[\boldsymbol{\mu}^t\Sigma_1^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^t\Sigma_1^{-1}\underbrace{\Sigma_1(\Sigma^{-1}\mathbf{x}+\Sigma_0^{-1}\boldsymbol{\mu}_0)}_{\boldsymbol{\mu}_1} - \boldsymbol{\mu}_1^t\Sigma^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^t\Sigma^{-1}\boldsymbol{\mu}_1\Big]\right\} \\
&\propto exp\left\{-\frac{1}{2}\Big[\boldsymbol{\mu}^t\Sigma_1^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^t\Sigma_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^t\Sigma^{-1}\boldsymbol{\mu}_1\Big]\right\} \\
&= exp\left\{-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_1)^t\Sigma_1^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_1)\right\},
\end{aligned}
$$

que é o núcleo de uma distribuição Normal com vetor de média $\boldsymbol{\mu}_1 = \Sigma_1(\Sigma^{-1}\mathbf{x}+\Sigma_0^{-1}\boldsymbol{\mu}_0)$ e matriz de covariância $\Sigma_1^{-1} = \Sigma^{-1}+\Sigma_0^{-1}$.

5.2.2 Definição de um processo Gaussiano

Um processo Gaussiano pode ser visto como uma generalização da distribuição Normal sobre um espaço n-dimensional. Da mesma forma como a distribuição Normal é especificada por sua média e variância, um processo Gaussiano é especificado por sua função média e sua função de covariância.

As variáveis aleatórias individuais em um vetor com distribuição Normal são indexadas por suas posições no vetor. Para um processo Gaussiano é o argumento t (da função aleatória $f(t)$) que desempenha o papel de índice, ou seja, para cada valor de t existe uma variável aleatória associada ($f(t)$).

Definição 2: Uma função aleatória $f(t)$ é dita seguir um processo Gaussiano com função média $m(x)$ e função covariância $C(t,t^*)$, se qualquer conjunto finito de valores da função é um vetor aleatório normalmente distribuído, isto é, se para qualquer $t_1, t_2, \ldots, t_n$, $n = 1, 2, \ldots$, tem-se que

$$\mathbf{f} \sim N_n(\mathbf{m}, \Sigma), \tag{5.5}$$

em que

$$\mathbf{f} = \begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix}, \mathbf{m} = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \vdots \\ m(t_n) \end{bmatrix}, \Sigma = \begin{bmatrix} C(t_1,t_1) & C(t_1,t_2) & \ldots & C(t_1,t_n) \\ C(t_2,t_1) & C(t_2,t_2) & \ldots & C(t_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(t_n,t_1) & C(t_n,t_2) & \ldots & C(t_n,t_n) \end{bmatrix}.$$

A notação usual é

$$f(.) \sim PG\Big(m(.), C(.,.)\Big). \tag{5.6}$$

A função média fornece a média do processo ao longo do tempo e é dada por $E\Big(f(t)\Big) = m(t)$. É importante ressaltar que as realizações do processo se localizarão, com maior frequência, em torno da função média.

A função de covariância, dada por $\Sigma(t,t^*) = C(t,t^*) = E\Big[\Big(f(t) - m(t)\Big)\Big(f(t^*) - m(t^*)\Big)\Big]$, pode ser vista como uma medida da dispersão das possíveis realizações do processo.

O Exemplo 8, proposto por Rasmussen (2004), ilustra um processo Gaussiano especificado por sua função média $m(t)$ e função covariância $C(t,t^*)$, que expressa a covariância entre os valores da função $f(.)$ no ponto t e t^*.

Exemplo 8 Considere o Processo Gaussiano com função média

$$m(t) = \frac{1}{4}t^2, \tag{5.7}$$

e função de covariância dada por

$$C(t,t^*) = exp\left\{-\frac{1}{2}(t-t^*)^2\right\}. \tag{5.8}$$

Seleciona-se um conjunto finito de pontos $\mathbf{t} = (t_1,\ldots,t_n)$ e calcula-se o vetor de médias e a matriz de covariância dadas em (5.7) e (5.8), que define uma distribuição Gaussiana multivariada com vetor de médias

$$\mu_i = m(t_i) = \frac{1}{4}t_i^2 \qquad i = 1,\ldots,n$$

e matriz de covariância com elementos

$$\Sigma_{ij} = C(t_i,t_j) = exp\left\{-\frac{1}{2}(t_i-t_j)^2\right\} \qquad i,j = 1,\ldots,n.$$

Assim, a distribuição conjunta dos valores da função $f(t)$ para os correspondentes t's é

$$\begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix} \sim N_n(\boldsymbol{\mu}, \Sigma) \tag{5.9}$$

em que

$$\boldsymbol{\mu} = \begin{bmatrix} \frac{1}{4}t_1^2 \\ \frac{1}{4}t_2^2 \\ \vdots \\ \frac{1}{4}t_n^2 \end{bmatrix} \quad e \quad \Sigma = \begin{bmatrix} exp\left\{-\frac{1}{2}(t_1-t_1)^2\right\} & \ldots & exp\left\{-\frac{1}{2}(t_1-t_n)^2\right\} \\ exp\left\{-\frac{1}{2}(t_2-t_1)^2\right\} & \ldots & exp\left\{-\frac{1}{2}(t_2-t_n)^2\right\} \\ \vdots & \ddots & \vdots \\ exp\left\{-\frac{1}{2}(t_n-t_1)^2\right\} & \ldots & exp\left\{-\frac{1}{2}(t_n-t_n)^2\right\} \end{bmatrix}.$$

A partir da geração de vetores aleatórios da distribuição Normal definida em (5.9), é possível obter funções aleatórias, visto que, temos agora um modelo probabilístico para

$f(t)$.

A Figura 18 mostra três funções retiradas aleatoriamente do processo Gaussiano definido nesse exemplo.

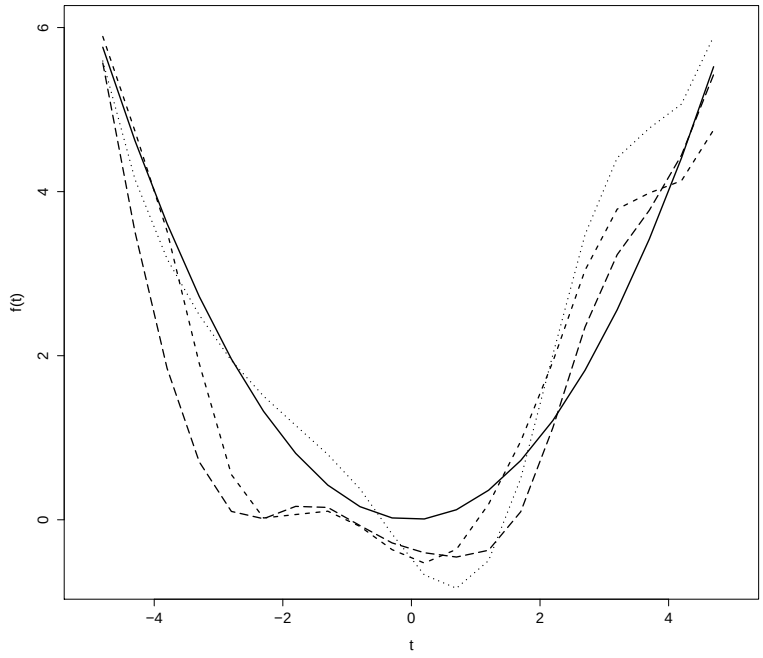


Figura 18: Três funções aleatórias retiradas do processo Gaussiano e sua função média (linha sólida).

5.3 Funções de Covariância

A função de covariância desempenha um importante papel em um processo Gaussiano, pois possui característica única e determina as propriedades das funções amostrais. A escolha adequada da função de covariância é fundamental para uma boa qualidade das estimações e predições.

Nessa seção serão discutidas algumas das funções de covariância mais conhecidas na literatura. As figuras ilustram as formas das funções amostrais provenientes de processos Gaussianos com função média nula e função de covariância $C(t, t^*)$.

Qualquer função $C(t, t^*)$ que aplicada a um conjunto de valores de t resulte em uma matriz P simétrica positiva semi-definida, isto é $v^t P v \geq 0$, pode ser considerada uma função de covariância válida em processos Gaussianos. Essa função mensura quão correlacionado são os valores da função $f(t)$ no ponto t e t^*, além de governar propriedades como estacionariedade, periodicidade, suavidade, entre outras.

Todas as funções de covariância podem ser escritas da seguinte forma

$$C(t, t^*) = \sigma^2 c(t, t^*), \tag{5.10}$$

em que o parâmetro σ^2 controla a variabilidade total do processo.

Uma das funções de covariância mais conhecida é a Laplaciana, dada por

$$C(t, t^*) = \sigma^2 exp\left\{-\frac{1}{2\eta}|t - t^*|\right\}. \tag{5.11}$$

O parâmetro η controla a dependência do processo e também é conhecido como parâmetro de suavidade. Sua função é redimensionar a distância entre t e t^*. Assim quanto maior for η, mais suave será a função $f(.)$.

Essa função de covariância fornece funções amostrais contínuas, porém não diferenciáveis.

A Figura 19 mostra duas funções retiradas aleatoriamente de um processo Gaussiano com função de covariância Laplaciana.

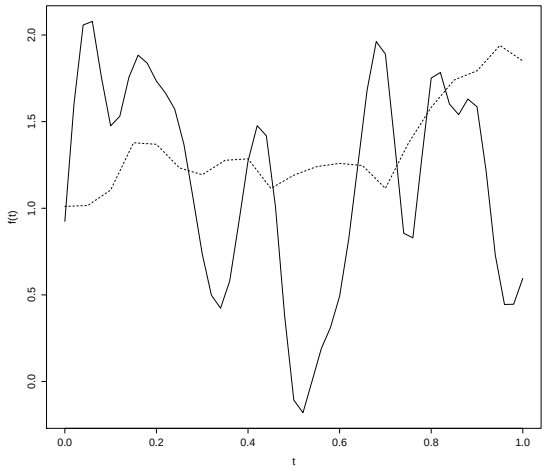


Figura 19: Funções retiradas aleatoriamente de um P.G. com função de covariância Laplaciana com $\sigma = 1$, $\eta = 0,5$ (sólida) e $\eta = 1,3$ (tracejada).

Uma outra função de covariância é a Exponencial Quadrática (ou Gaussiana) que pode ser escrita como

$$C(t, t^*) = \sigma^2 exp\left\{-\frac{1}{2\eta}(t - t^*)^2\right\}. \tag{5.12}$$

As funções amostrais provenientes de um processo Gaussiano com função de covariância dada em (5.12) são contínuas e infinitamente diferenciáveis.

Note que para qualquer par de elementos t e t^*, $f(t)$ e $f(t^*)$ possuirão alta correlação se t e t^* são valores próximos, isto é, se $|t - t^*| \approx 0$ então $exp\left\{-\frac{1}{2\eta}(t - t^*)^2\right\} \approx 1$ (função

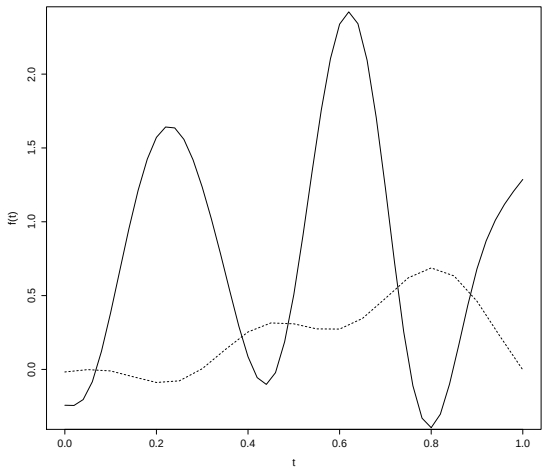


Figura 20: Funções retiradas aleatoriamente de um P.G.com função de covariância Exponencial Quadrática com $\sigma = 1$, $\eta = 0,1$ (sólida) e $\eta = 0,3$ (tracejada).

de covariância Exponencial Quadrática). O mesmo vale para a função de covariância Laplaciana.

Segundo Rasmussen e Williams (2006) apesar da função Exponencial Quadrática ser provavelmente a função de covariância mais utilizada em processos Gaussianos, as funções amostrais resultantes são muito suaves e isso é pouco realístico na modelagem de alguns processos físicos. Dessa forma, em alguns casos, a classe de Matérn é mais recomendada.

As funções de covariância da classe de Matérn são dadas por

$$C(t,t^*) = \sigma^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \left[\frac{\sqrt{2\nu}}{l}|t-t^*|\right]^{\nu} K_\nu \left(\frac{\sqrt{2\nu}}{l}|t-t^*|\right), \tag{5.13}$$

em que $l > 0$, $\nu > 0$ e K_ν é a função de Bessel modificada do segundo tipo de ordem ν que pode ser escrita como

$$K_q(z) = \left(\frac{\pi}{2}\right) \frac{I_{-q}(z) - I_q(z)}{sin(\pi q)},$$

em que $I_q(z) = \left(\frac{z}{2}\right)^q \sum_{p=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^p}{p!\Gamma(q+p+1)}$.

O hiperparâmetro ν controla o grau de diferenciabilidade das funções amostrais da forma Matérn. Assim essas funções são $(\nu - 1)$ diferenciáveis. Se $\nu \to \infty$, a função de covariância Matérn aproxima-se da forma Exponencial Quadrática e se $\nu = \frac{1}{2}$ a função de covariância Matérn assume a forma da Laplaciana.

Um outro exemplo de função de covariância é a Periódica, que pode ser escrita como

$$C(t,t^*) = \sigma^2 exp\left\{-\frac{1}{\eta} sin^2\Big(\gamma(t-t^*)\Big)\right\}. \tag{5.14}$$

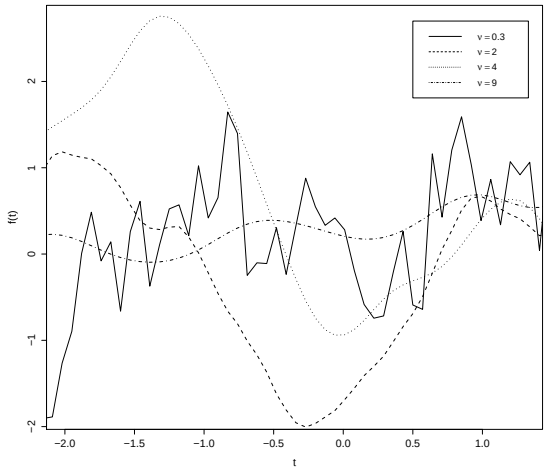


Figura 21: Funções retiradas aleatoriamente de um P.G. com função de covariância Matérn com $\sigma = 1$ e $l = 1$.

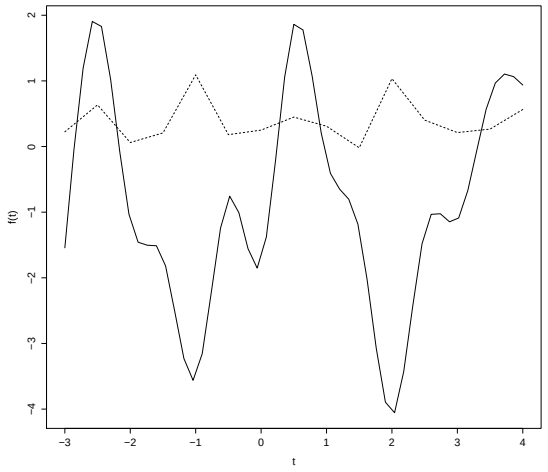


Figura 22: Funções retiradas aleatoriamente de um P.G. com função de covariância Periódica com $\sigma = 1$, $\gamma = 1$ $\eta = 0,1$ (sólida) e $\eta = 0,6$ (tracejada).

O parâmetro γ controla a periodicidade das funções, ou seja, a frequência com que elas irão se repetir através do tempo.

5.4 Inferência Bayesiana sobre um Processo Gaussiano

Seja f a função desconhecida que queremos estimar, podemos assumir que f segue um processo Gaussiano dado por

$$f(t) \sim PG\Big(m(t), C(t, t^*)\Big), \tag{5.15}$$

em que $C(.,.)$ é conhecida.

Seja $\mathbf{y}$ o vetor de valores de $f(.)$ observados em n pontos $t_1, t_2, \ldots, t_n$ em que

$$\mathbf{y} \sim N(\mathbf{m}, \Sigma), \tag{5.16}$$

em que

$$\mathbf{y} = \begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix}, \mathbf{m} = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \vdots \\ m(t_n) \end{bmatrix} e\ \Sigma = \begin{bmatrix} C(t_1,t_1) & \ldots & C(t_1,t_n) \\ C(t_2,t_1) & \ldots & C(t_2,t_n) \\ \vdots & \ddots & \vdots \\ C(t_n,t_1) & \ldots & C(t_n,t_n) \end{bmatrix}.$$

A distribuição conjunta de $\mathbf{y}$ e $f(t)$ pode ser escrita como

$$\begin{pmatrix} \mathbf{y} \\ f(t) \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{m} \\ m(t) \end{pmatrix}, \begin{pmatrix} \Sigma & \mathbf{k}(t) \\ \mathbf{k}(t)^T & C(t,t) \end{pmatrix} \right]$$

em que $\mathbf{k}(t) = Cov(f(t), \mathbf{y}) = Cov(f(t), f(t_i)) = [C(t_1,t), C(t_2,t), \ldots, C(t_n,t)]^T$.

A partir do resultado (5.4) temos que a distribuição condicional de $f(t)$ dado $\mathbf{y}$ é também normal. Assim, a distribuição a posteriori de qualquer p pontos da função é normal multivariada. Dessa forma, a distribuição a posteriori de $f(.)$ é dada por

$$f(.)|\mathbf{y} \sim PG\Big(m(t) + \mathbf{k}(t)^T\Sigma^{-1}(\mathbf{y} - \mathbf{m}), C(t,t) - \mathbf{k}(t)^T\Sigma^{-1}\mathbf{k}(t)\Big), \tag{5.17}$$

que incorpora as observações $f(x_1), \ldots, f(x_n)$.

Esses são os resultados apresentados por O'Hagan (1978) que formam a base da metodologia Bayesiana não-paramétrica na análise de funções desconhecidas utilizando o processo Gaussiano.

Exemplo 9 Considere o Processo Gaussiano com função média

$$m(t) = \frac{1}{12}t^3, \tag{5.18}$$

e função de covariância dada por

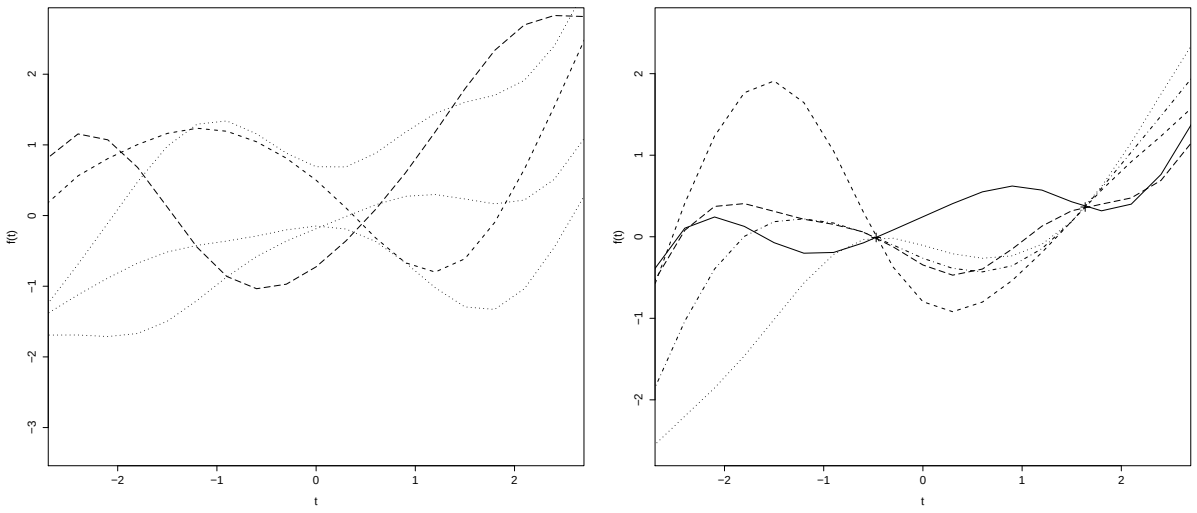
$$C(t, t^*) = exp\left\{-\frac{1}{2}(t - t^*)^2\right\}. \tag{5.19}$$

Então, o processo a priori é dado por

$$f(.) \sim PG\left(\frac{1}{12}t^3, exp\left\{-\frac{1}{2}(t - t^*)^2\right\}\right). \tag{5.20}$$

Utilizando o mesmo procedimento ilustrado no Exemplo 8, algumas funções amostrais

foram retiradas aleatoriamente da distribuição a priori sob funções especificadas pelo processo Gaussiano dado em (5.20) e são mostradas na Figura 23 (a). Essa priori representa a nossa crença sobre o tipo de função que esperamos obter, antes de observar os dados.



(a) *priori* (b) *posteriori*

Figura 23: Funções amostrais provenientes da distribuição a priori definida em (5.20) são mostradas em (a). Em (b) ilustra-se o caso em que duas observações são consideradas (pontos em x). A partir disso, obtém-se aleatoriamente cinco funções amostrais retiradas da posteriori.

Suponha agora que observamos um conjunto de dados $S = \{(t_1, y_1), (t_2, y_2)\}$. Desejamos considerar apenas as funções que passam exatamente por esses pontos. A combinação da priori e dos dados fornece a distribuição a posteriori sobre funções que é dada por

$$f(.)|S \sim PG\Big(m(t) + \mathbf{k}(t)^T\Sigma^{-1}(\mathbf{y} - \mathbf{m}), C(t,t) - \mathbf{k}(t)^T\Sigma^{-1}\mathbf{k}(t)\Big),$$

em que

$$\mathbf{y} = \begin{bmatrix} f(t_1) \\ f(t_2) \end{bmatrix}, \mathbf{m} = \begin{bmatrix} m(t_1) \\ m(t_2) \end{bmatrix} e\ \Sigma = \begin{bmatrix} C(t_1,t_1) & C(t_1,t_2) \\ C(t_2,t_1) & C(t_2,t_2) \end{bmatrix}.$$

Note que a variância a posteriori sempre será menor do que a variância a priori $C(t,t)$, desde que os dados observados forneçam alguma informação. Isso ocorre porque a expressão da variância a posteriori, $C(t,t) - \mathbf{k}(t)^T\Sigma^{-1}\mathbf{k}(t)$, é a variância a priori menos um termo positivo, que depende dos dados.

A forma de se retirar funções amostrais da distribuição a posteriori de $f(.)$ é similar ao processo descrito no Exemplo 8. Simulamos a função em um número finito de pontos para a obtenção de um vetor de dados simulados S' e o adicionamos ao conjunto de dados inicial S. Aplicamos então o processo Gaussiano dado em (5.17) a esses dados combinados. Esses novos pontos são selecionados para melhorar a cobertura da área de interesse (área

que desejamos fazer inferência sobre a forma de $f(.)$). Isso é ilustrado na Figura 24 (b). Note que a incerteza é reduzida próximo as observações.

Segundo Rasmussen e Willians (2006), o processo Gaussiano não é um modelo paramétrico, assim não existe a necessidade de nos preocuparmos se é possível o ajuste desse modelo ao conjunto de dados (como seria o caso se tentássemos, por exemplo, ajustar um modelo de regressão linear em dados com comportamento não linear). Porém, a qualidade desse ajuste depende de diversos fatores como por exemplo, a escolha adequada da função média (seu comportamento deve ser próximo ao da função verdadeira), da função de covariância e seus hiperparâmetros. Vimos na Seção 5.3 que a funções de covariância possuem características únicas e determinam as propriedades das funções amostrais. Dentre as funções de covariância mais utilizadas encontra-se a função Exponencial quadrática, que possui um hiperparâmetro de suavidade (η).

Exemplo 10: Considere $f(t) = t + cos(3t)$ a função verdadeira. Foram obtidas quatro observações dessa função para $t = 1, 2, 3$ e 4.

A opinião a priori do estatístico é

$$f(t) \sim PG\Big(m(t), C(t, t^*)\Big),$$

em que $m(t) = t+1$ e $C(t, t^*) = exp\left\{-\frac{1}{2\eta}(t - t^*)^2\right\}$ com η conhecido. Serão considerados dois valores para η: um bem pequeno, 0.0005 e outro maior, 0.3.

A Figura 24 mostra que a função média a posteriori possui um comportamento diferente da função verdadeira quando o hiperparâmetro de suavidade é muito pequeno. Apenas nos pontos observados as funções média a posteriori e a verdadeira coincidem.

Quando $\eta = 0,3$ a função média a posteriori possui uma forma muito mais próxima da função verdadeira. Segundo Gosling (2005), a função média a posteriori dada em (5.17) é composta pela soma de duas partes: a primeira parte, $m(t)$ representa a opinião a priori do estatístico sobre $f(.)$ e a segunda parte, $\mathbf{k}(t)^T\Sigma^{-1}(\mathbf{y} - \mathbf{m})$, ajusta essa crença de forma que a função média a posteriori passe pelos pontos observados. A segunda parte depende bastante de η; valores muito pequenos implicam que a correlação entre $f(t)$ e $f(t^*)$ é também muito pequena, mesmo quando t e t^* são próximos. Por outro lado, valores relativamente grandes de η implicam que $f(t)$ e $f(t^*)$ são bastante correlacionados, mesmo quando t e t^* são distantes.

A estimação do valor de η não é muito simples, exigindo o uso de distribuições a priori específicas para que a distribuição a posteriori seja própria. Mais detalhes sobre a estimação de η serão descritas no próximo capítulo.

Para ilustrar o efeito da escolha da função $m(t)$, considere os casos em que ao invés de

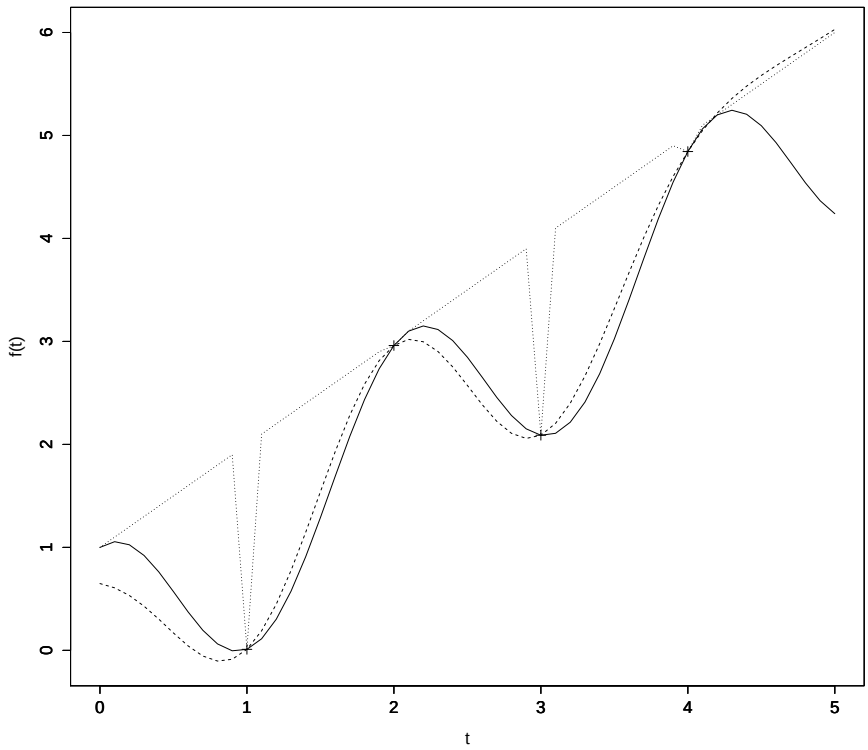


Figura 24: Função média a posteriori para $f(.)$ com $\eta = 0,0005$ (pontilhada), $\eta = 0,3$ (tracejada) e a função verdadeira (sólida).

$m(t) = t+1$, foram selecionadas $m(t) = t+cos(3t)$ (função verdadeira) e $m(t) = \frac{1}{2}t+cos(t)$.

A Figura 25 mostra que quando as crenças a priori do estatístico sobre $f(.)$ são atualizadas, as densidades simuladas resultantes da distribuição a posteriori correspondem exatamente às informações provenientes da função verdadeira, ou seja, os dados observados. Dessa forma, no intervalo em que alguns pontos foram observados, a função média selecionada será alterada não parametricamente para espelhar as informações dos dados. Nas áreas em que não existem informações provenientes dos dados, a função média tem maior impacto.

Até agora consideramos apenas os casos em que os hiperparâmetros que controlam o processo Gaussiano são conhecidos ou inexistentes. A seguir, apresentaremos o caso em que esses hiperparâmetros são desconhecidos e a forma como podem ser estimados através do conjunto de dados.

Quando um processo Gaussiano é utilizado como priori, segue que a distribuição dos dados também é normal. Assim,

$$p(\mathbf{y}|\mathbf{t}, \boldsymbol{\kappa}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}, \tag{5.21}$$

em que $\boldsymbol{\kappa}$ é o vetor de hiperparâmetros.

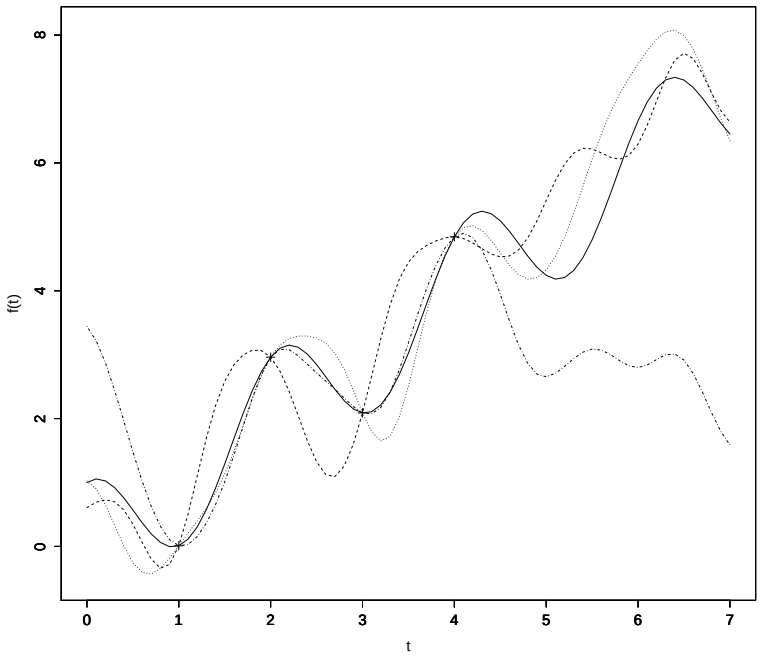


Figura 25: Função verdadeira (sólida) e funções amostrais retiradas aleatoriamente do processo Gaussiano a posteriori quando $m(t) = t + 1$(tracejada) , $m(t) = t + cos(3t)$ (pontilhada) e $m(t) = \frac{1}{2}t + cos(t)$(tracejada e pontilhada).

A densidade a posteriori de $\boldsymbol{\kappa}$ é dada por

$$p(\boldsymbol{\kappa}|\mathbf{t},\mathbf{y}) \propto p(\mathbf{y}|\mathbf{t},\boldsymbol{\kappa})\pi(\boldsymbol{\kappa}), \tag{5.22}$$

onde $\pi(\boldsymbol{\kappa})$ é a distribuição a priori dos hiperparâmetros. Geralmente atribui-se distribuições a priori vagas aos hiperparâmetros.

Na maior parte das vezes a distribuição a posteriori dada em (5.22) não é analiticamente tratável. Assim, as técnicas de Monte Carlo via Cadeias de Markov (MCMC) podem ser utilizadas para obtenção de amostras da distribuição a posteriori (5.22) e estimativas de interesse.

6 *Inferência Bayesiana Não-Paramétrica para Elicitação de distribuição a priori*

No Capítulo 2, vimos que a elicitação é um processo de transformação do conhecimento do especialista sobre alguma quantidade desconhecida na forma de uma distribuição de probabilidade. Até o momento, discutimos apenas métodos de elicitação paramétrica da distribuição a priori, em que ajustamos as informações provenientes do especialista a uma distribuição de probabilidade conhecida. Porém, sabemos que tais métodos podem representar apenas alguns tipos de densidade (geralmente, densidades unimodais, sem caudas pesadas, etc). Além disso, podem ocorrer significativas perdas das características da distribuição do especialista quando a condicionamos a forma de uma distribuição de probabilidade conhecida.

A alternativa não-paramétrica oferece uma maneira mais flexível de elicitação da distribuição a priori do especialista. Não assumimos uma forma paramétrica para $f(.)$, que representa a opinião do especialista, mas sim a consideramos uma função aleatória. Devido a essa flexibilidade, a metodologia não-paramétrica pode ser aplicada em diversas áreas do conhecimento.

Assumimos que o especialista seja capaz de fornecer algumas medidas descritivas de sua distribuição, como média e percentis, que serão os dados utilizados no procedimento Bayesiano não-paramétrico para a identificação da distribuição a priori do especialista.

O procedimento Bayesiano não-paramétrico para elicitação de uma distribuição a priori para o caso univariado foi denvolvido por Oakley e O'Hagan (2007). A elicitação da distribuição a priori pode ser vista como a estimação de um parâmetro na inferência Bayesiana. O objetivo do estatístico é fazer inferências sobre uma função desconhecida $f(.)$, a densidade a priori do especialista. Primeiramente, o estatístico elabora suas próprias crenças a priori sobre $f(.)$ e então solicita ao especialista algumas medidas descritivas de sua distribuição, como média, moda e percentis. As medidas resumo elicitadas são vistas como dados, ou seja, observações de $f(.)$. Assim, atualiza-se as convicções do estatístico

sobre $f(.)$ utilizando esses dados, obtendo-se então a distribuição a posteriori que pode ser considerada a melhor estimativa para $f(.)$.

A distribuição pode ser representada por um processo Gaussiano, permitindo que $f(.)$ tenha qualquer forma. O processo Gaussiano já foi utilizado para fazer inferências sobre uma função desconhecida, alguns exemplos podem ser encontrados em Oakley e O'Hagan(2002), Oakley e O'Hagan(2007).

Esse capítulo está organizado da seguinte forma: na Seção 6.1, a priori do estatístico para $f(.)$ é discutida. Na Seção 6.2, a função de covariância utilizada no procedimento Bayesiano não-paramétrico é apresentada. Na Seção 6.3, as distribuições a priori para os hiperparâmetros são descritas. Os dados elicitados do especialista são apresentados na Seção 6.4 e a atualização da priori é discutida na Seção 6.5.

6.1 Processo Gaussiano como uma representação a priori para $f(.)$

Sob o enfoque não-paramétrico, não assumimos nenhuma forma paramétrica para a função densidade do especialista, denotada por $f(.)$. Nesse caso, uma representação a priori adequada é dada por um processo Gaussiano. Assim, a distribuição a priori de $f(.)$ é dada por

$$f(.)|\boldsymbol{\kappa} \sim PG\Big(g(.), \sigma^2 c(.,.)\Big). \tag{6.1}$$

Sabemos que $f(.)$ é uma função densidade. Dessa forma, uma escolha natural para a esperança a priori de $f(.)$ é uma distribuição de probabilidade, representada por $g(.)$,

$$E(f(.)|\boldsymbol{\kappa}) = g(.). \tag{6.2}$$

A função $g(.)$ é conhecida como densidade *underlying*.

O método de Oakley e O'Hagan considera o caso onde o estatístico escolhe como função média, $g(.)$, uma distribuição normal dada por

$$g(\theta) = \frac{1}{\sqrt{2\pi b}} exp\left\{-\frac{1}{2b}(\theta - m)^2\right\} \tag{6.3}$$

em que m é a média e b a variância da distribuição.

Nesse caso, o estatístico acredita que a função do especialista será unimodal, suave e aproximadamente simétrica. Se essas condições não forem razoáveis, outra distribuição de probabilidade pode ser selecionada. Nesse trabalho, a distribuição normal será assumida por ser apropriada em várias situações e por facilidade na manipulação matemática.

Se modelarmos a razão entre a densidade do especialista e a função média do processo Gaussiano, $h(\theta) = \frac{f(\theta)}{g(\theta)}$, teremos como processo Gaussiano com média constante 1 e função variância dada por

$$Var\left(\frac{f(\theta)}{g(\theta)}|\boldsymbol{\kappa}\right) = \sigma^2, \tag{6.4}$$

para todos os valores de θ. Assim, o hiperparâmetro σ^2 especifica quão próximo a verdadeira função densidade estará da função média a priori $g(.)$. Dessa forma, a razão entre as duas densidades será próxima a 1 para pequenos valores de σ^2, isto é, esperamos que a densidade do especialista se aproxime bastante da densidade *underlying*.

A covariância entre quaisquer dois pontos $h(\theta)$ e $h(\theta')$ pode ser escrita como

$$Cov\{h(\theta), h(\theta')|\boldsymbol{\kappa}\} = \sigma^2 c(\theta, \theta'). \tag{6.5}$$

em que $c(\theta, \theta')$ é a função de correlação. De fato, sabe-se que a função de correlação $c(\theta, \theta')$ pode ser escrita como

$$c(\theta, \theta') = \frac{Cov\{h(\theta), h(\theta')\}}{\sqrt{Var(h(\theta))}\sqrt{Var(h(\theta'))}},$$

e utilizando (6.4) temos que

$$c(\theta, \theta') = \frac{Cov\{h(\theta), h(\theta')\}}{\sigma\sigma}.$$

Portanto

$$Cov\{h(\theta), h(\theta')|\boldsymbol{\kappa}\} = \sigma^2 c(\theta, \theta').$$

Logo, na prática, modelamos a razão $h(\theta) = \frac{f(\theta)}{g(\theta)}$ como um processo Gaussiano estacionário com média constante 1 e função de covariância $Cov\{h(\theta), h(\theta')|\boldsymbol{\kappa}\} = \sigma^2 c(\theta, \theta')$, isto é,

$$h(.)|\boldsymbol{\kappa} \sim PG\Big(1, \sigma^2 c(.,.)\Big). \tag{6.6}$$

6.2 Modelando uma apropriada função de covariância

Sabe-se que

$$\begin{aligned} c(\theta, \theta') &= \frac{Cov\{h(\theta), h(\theta')\}}{\sqrt{Var(h(\theta))}\sqrt{Var(h(\theta'))}} \\ &= \frac{Cov\left\{\frac{f(\theta)}{g(\theta)}, \frac{f(\theta')}{g(\theta')}\right\}}{\sqrt{Var\left(\frac{f(\theta)}{g(\theta)}\right)}\sqrt{Var\left(\frac{f(\theta')}{g(\theta')}\right)}} \end{aligned}$$

$$= \frac{\frac{1}{g(\theta)g(\theta')}Cov\{f(\theta), f(\theta')\}}{\frac{1}{g(\theta)g(\theta')}\sqrt{Var(f(\theta))}\sqrt{Var(f(\theta'))}}.$$

Mas,

$$Var(h(\theta)) = Var\left(\frac{f(\theta)}{g(\theta)}\right) = \frac{1}{g(\theta)^2}Var(f(\theta)).$$

E sabe-se que

$$Var(h(\theta)) = \sigma^2.$$

Então,

$$Var(f(\theta)) = \sigma^2 g(\theta)^2. \tag{6.7}$$

Portanto, a função de covariância entre $f(\theta)$ e $f(\theta')$ pode ser escrita da seguinte forma

$$Cov\{f(\theta), f(\theta')|\boldsymbol{\kappa}\} = \sigma^2 g(\theta)g(\theta')c(\theta, \theta'). \tag{6.8}$$

A função de correlação escolhida é dada por

$$c(\theta, \theta') = exp\left\{-\frac{1}{2\eta}(\theta - \theta')^2\right\}. \tag{6.9}$$

Essa escolha, que foi apresentada em Kennedy e O'Hagan(1996), é matematicamente conveniente. A função $c(\theta, \theta')$, dada em (6.9), expressa a convicção do estatístico que $f(\theta)$ é uma função suave e infinitamente diferenciável.

6.3 Distribuições a priori para m, b, η e σ^2

Para completar o modelo hierárquico é necessário a especificação de uma distribuição a priori para os hiperparâmetros $\boldsymbol{\kappa} = (m, b, \eta, \sigma^2)$, que são considerados desconhecidos. Essa priori irá refletir a crença do estatístico sobre esses hiperparâmetros.

O uso de distribuições a priori informativas para $\pi(\boldsymbol{\kappa})$ poderia interferir nos objetivos da elicitação do conhecimento do especialista. Assim, é preferível que a distribuição a posteriori para $f(.)$ esteja baseada apenas nos dados, isto é, nas informações fornecidas pelo especialista. Por isso, prioris não informativas serão especificadas para m, b e σ^2.

Segundo Moala e O'Hagan(2010), uma possível priori conjunta para os hiperparâme-

tros é dada por

$$\pi(\boldsymbol{\kappa}) \propto \frac{1}{b\sigma^2}\pi(\eta). \tag{6.10}$$

em que $\pi(\eta)$ é a distribuição a priori para o parâmetro de suavidade η fornecida pelo estatístico.

Uma distribuição a priori imprópria para η levaria a uma posteriori imprópria. Por isso, Oakley e O'Hagan(2007) sugerem o uso de uma distribuição a priori para $\eta^* = \frac{\eta}{b}$. Assim, a função de covariância definida em (6.9) pode ser reparametrizada como

$$c(\theta, \theta') = exp\left\{-\frac{1}{2b\eta^*}(\theta - \theta')^2\right\}, \tag{6.11}$$

em que $\eta^* = \frac{\eta}{b}$.

Para encontrar uma distribuição a priori adequada para η^*, Oakley e O'Hagan (2007) fizeram um estudo de simulação e concluíram que a distribuição a priori adequada é

$$log(\eta^*) \sim N(0,65; 0,25^2). \tag{6.12}$$

Assim, a distribuição conjunta para m, b, η^* e σ^2 é dada por

$$\pi(\boldsymbol{\kappa}^*) \propto \frac{exp\left\{-8log(\eta^* - 0,65)^2\right\}}{b\sigma^2}, \tag{6.13}$$

onde $\boldsymbol{\kappa}^* = (m, b, \eta^*, \sigma^2)$.

6.4 Dados Elicitados do Especialista

Vamos pedir ao especialista probabilidades como $P_x = \int_{-\infty}^{x} f(\theta)d\theta$ de sua distribuição. Nesse caso, o vetor de dados será

$$\mathbf{d}^t = \left(\int_{-\infty}^{x_1} f(\theta)d\theta, \ldots, \int_{-\infty}^{x_n} f(\theta)d\theta\right), \tag{6.14}$$

em que $x_1, x_2, \ldots, x_n$ são n valores selecionados para os quais o especialista fornece as probabilidades $P_{x_1}, \ldots, P_{xn}$. A restrição $\int_{\mathbb{R}} f(\theta) = 1$ tem que ser informada ao especialista.

Segundo Moala(2006) como $f(.)$ segue um processo Gaussiano e $P_x = \int_{-\infty}^{x} f(\theta)d\theta$ é um funcional [1] linear de $f(.)$ então P_x segue um processo Gaussiano.

Assim, o vetor $\mathbf{d}$ condicionado em $\boldsymbol{\kappa}^*$ é normalmente distribuído (por definição de processo Gaussiano),

$$\mathbf{d}|\boldsymbol{\kappa}^* \sim N_n(\mathbf{w}, \sigma^2 A), \tag{6.15}$$

[1] Intuitivamente, podemos dizer que um funcional é uma função de uma função.

em que os elementos do vetor de média **w** são dados por

$$E(P_x|\boldsymbol{\kappa}^*) = \int_{-\infty}^{x} E(f(\theta)|\boldsymbol{\kappa}^*)d\theta = \int_{-\infty}^{x} g(\theta)d\theta = \int_{-\infty}^{x} N(m,v)d\theta = \Phi\left(\frac{x-m}{\sqrt{b}}\right). \quad (6.16)$$

E os componentes da matriz de covariância $\sigma^2 A$ são dados por

$$Cov(P_x, P_y|\boldsymbol{\kappa}^*) = \int_{-\infty}^{x}\int_{-\infty}^{y} Cov(f(\theta), f(\theta')|\boldsymbol{\kappa}^*))d\theta d\theta' = \sigma^2 \int_{-\infty}^{x}\int_{-\infty}^{y} g(\theta)g(\theta')c(\theta,\theta')d\theta d\theta'$$

$$= \sigma^2 \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{2\pi b} exp\left\{-\frac{1}{2b}(\theta-m)^2\right\} exp\left\{-\frac{1}{2b}(\theta'-m)^2\right\} exp\left\{-\frac{1}{2b\eta^*}(\theta-\theta')^2\right\} d\theta d\theta'$$

$$= \sigma^2 \frac{1}{2\pi b}\int_{-\infty}^{x}\int_{-\infty}^{y} exp\left\{-\frac{1}{2\eta^*}\left[\frac{\eta^*(\theta-m)^2}{b} + \frac{\eta^*(\theta'-m)^2}{b} + \frac{(\theta-\theta')^2}{b}\right]\right\} d\theta d\theta'.$$

Mas,

$$\frac{\eta^*(\theta-m)^2}{b} + \frac{\eta^*(\theta'-m)^2}{b} + \frac{(\theta-\theta')^2}{b} = \frac{(\eta^*+1)(\theta-m)^2}{b} + \frac{(\eta^*+1)(\theta'-m)^2}{b} + \underbrace{\frac{(\theta-\theta')^2}{b} - \frac{(\theta'-m)^2}{b} - \frac{(\theta-m)^2}{b}}_{-2\frac{(\theta-m)(\theta'-m)}{b}}.$$

Assim,

$$Cov(P_x, P_y|\boldsymbol{\kappa}^*) = \sigma^2 \frac{1}{2\pi b}\int_{-\infty}^{x}\int_{-\infty}^{y} exp\left\{-\frac{(\eta^*+1)^2}{2\eta^*}\left[\frac{(\theta-m)^2}{b(\eta^*+1)} + \frac{(\theta'-m)^2}{b(\eta^*+1)} - 2\frac{(\theta-m)(\theta'-m)}{b(\eta^*+1)^2}\right]\right\} d\theta d\theta'$$

$$= \sigma^2 \sqrt{\frac{\eta^*}{\eta^*+2}}\int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{2\pi b}\sqrt{\frac{\eta^*+2}{\eta^*}} exp\left\{-\frac{(\eta^*+1)^2}{2\eta^*(\eta^*+2)}\left[\frac{(\eta^*+2)(\theta-m)^2}{b(\eta^*+1)} + \frac{(\eta^*+2)(\theta'-m)^2}{b(\eta^*+1)} - 2\frac{(\eta^*+2)}{b(\eta^*+1)^2}(\theta-m)(\theta'-m)\right]\right\} d\theta d\theta'.$$

Portanto

$$Cov(P_x, P_y|\boldsymbol{\kappa}^*) = \sigma^2 \sqrt{\frac{\eta^*}{\eta^*+2}}\int_{-\infty}^{x}\int_{-\infty}^{y} N_2((\theta,\theta')|(m,m),S)d\theta d\theta', \quad (6.17)$$

em que

$$S = b \begin{bmatrix} \frac{\eta^*+1}{\eta^*+2} & \frac{1}{\eta^*+2} \\ \frac{1}{\eta^*+2} & \frac{\eta^*+1}{\eta^*+2} \end{bmatrix}.$$

Assim, a função de verossimilhança é dada por

$$L(\boldsymbol{\kappa}^*|\mathbf{d}) = \frac{|A|^{-\frac{1}{2}}}{\sigma^n} exp\left\{-\frac{1}{2\sigma^2}(\mathbf{d}-\mathbf{w})^t A^{-1}(\mathbf{d}-\mathbf{w})\right\}. \tag{6.18}$$

6.5 Posteriori como atualização da priori

A distribuição a posteriori de $f(\theta)$ requer o conhecimento da covariância de $f(\theta)$ e os dados elicitados do especialista. Segundo Oakley e O'Hagan (2007) a covariância a priori entre $f(\theta)$ e $\mathbf{d}$ é dada por

$$\begin{aligned} Cov(f(\theta), P_x|\boldsymbol{\kappa}^*) &= Cov(f(\theta), \int_{-\infty}^{x} f(\theta')d\theta'|\boldsymbol{\kappa}^*) = \int_{-\infty}^{x} Cov(f(\theta), f(\theta')|\boldsymbol{\kappa}^*)d\theta' \\ &= \sigma^2 g(\theta)\sqrt{\frac{\eta^*}{\eta^*+1}} exp\left\{-\frac{(\theta-m)^2}{2b(\eta^*+1)}\right\} \Phi\left[x - \frac{\theta+m\eta^*}{\eta^*+1}\sqrt{\frac{\eta^*+1}{b\eta^*}}\right]. \end{aligned} \tag{6.19}$$

e a distribuição conjunta de $f(\theta)$ e $\mathbf{d}$ é então

$$\begin{pmatrix} \mathbf{y} \\ f(\theta) \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{w} \\ g(\theta) \end{pmatrix}, \begin{pmatrix} \sigma^2 A & \sigma^2\mathbf{k}(\theta) \\ \sigma^2\mathbf{k}(\theta)^T & C(\theta,\theta) \end{pmatrix}\right]$$

em que $\mathbf{k}(\theta)$ é o vetor de covariância a priori entre $f(\theta)$ e $\mathbf{d}$, e seus elementos são dados em (6.19).

Assim, a distribuição a posteriori $f(.)|\mathbf{d}, \boldsymbol{\kappa}^*$ é um processo Gaussiano com função média

$$E(f(\theta)|\mathbf{d}, \boldsymbol{\kappa}^*) = g^*(\theta) = g(\theta) + \mathbf{k}(\theta)^t A^{-1}(\mathbf{d}-\mathbf{w}) \tag{6.20}$$

e função de covariância

$$Cov(f(\theta), f(\theta')|\mathbf{d}, \boldsymbol{\kappa}^*) = \sigma^2 C^*(\theta,\theta') = \sigma^2\Big(g(\theta)g(\theta')c(\theta,\theta') - \mathbf{k}(\theta)^t A^{-1}\mathbf{k}(\theta')\Big). \tag{6.21}$$

6.5.1 Estimadores a posteriori para os hiperparâmetros

Utilizando (6.10) e (6.18) temos a distribuição a posteriori para os hiperparâmetros dada por

$$\begin{aligned} p(\boldsymbol{\kappa}^*|\mathbf{d}) &\propto \pi(\eta^*)\frac{|A|^{-\frac{1}{2}}}{b\sigma^{n+2}}exp\left\{-\frac{1}{2\sigma^2}(\mathbf{d}-\mathbf{w})^tA^{-1}(\mathbf{d}-\mathbf{w})\right\} \\ &\propto \frac{|A|^{-\frac{1}{2}}}{b\sigma^{n+2}}exp\left\{-\frac{1}{2\sigma^2}(\mathbf{d}-\mathbf{w})^tA^{-1}(\mathbf{d}-\mathbf{w}) - 8log(\eta^*-0.65)^2\right\}. \quad (6.22) \end{aligned}$$

O condicionamento em σ^2 pode ser removido da seguinte forma

$$p(m,b,\eta^*|\mathbf{d}) \propto \pi(\eta^*)\frac{|A|^{-\frac{1}{2}}}{b}\int_0^\infty \frac{1}{\sigma^{n+2}}exp\left\{-\frac{1}{2\sigma^2}(\mathbf{d}-\mathbf{w})^tA^{-1}(\mathbf{d}-\mathbf{w})\right\}d\sigma^2 \quad (6.23)$$

e sabendo que

$$\int_0^\infty t^{-(p+1)}e^{-zt^{-q}}dt = \frac{1}{q}z^{-\frac{p}{q}}\Gamma\left(\frac{p}{q}\right), \quad (6.24)$$

temos que

$$p(m,b,\eta^*|\mathbf{d}) \propto \pi(\eta^*)\frac{|A|^{-\frac{1}{2}}}{b}(\hat{\sigma}^2)^{-\frac{n}{2}}, \quad (6.25)$$

onde

$$\hat{\sigma}^2 = \frac{1}{n-2}(\mathbf{d}-\mathbf{w})^tA^{-1}(\mathbf{d}-\mathbf{w}). \quad (6.26)$$

Segundo Moala(2006) após a remoção do condicionamento de σ^2, temos que a distribuição a posteriori de $f(.)|m,b,\eta^*$ é um processo t-Student, dado por

$$f(\theta)|\mathbf{d},m,b,\eta^* \sim t_n\Big(g^*,\hat{\sigma}^2C^*(\theta,\theta),n\Big), \quad (6.27)$$

ou equivalentemente

$$\frac{f(\theta)-g^*(\theta)}{\hat{\sigma}\sqrt{C^*(\theta,\theta)}} \sim t_n(n) \quad (6.28)$$

A variância a posteriori estimada é dada por

$$\hat{var}\Big(f(\theta)|d,m,b,\eta^*\Big) = \frac{n}{n-2}\hat{\sigma}^2C^*(\theta,\theta). \quad (6.29)$$

Para cada valor de m,b e η^* gerado da posteriori $p(m,b,\eta^*|\mathbf{d})$ via MCMC, podemos amostrar uma função densidade $f(.)$ da distribuição a posteriori $p(f(.)|,\mathbf{d},m,b,\eta^*)$, dada em (6.27), para um número finito de valores de θ. Então, teremos uma amostra de funções $f(.)$ e plotamos a média dos pontos das funções amostrais, obtendo assim a estimativa de $f(.)$.

Exemplo 11: Para ilustrar o método, considere que a densidade verdadeira do especia-

lista é dada por

$$f(\theta) = \frac{0,4}{\sqrt{2\pi}} exp\left\{-\frac{1}{2}(\theta - 3)^2\right\} + \frac{0,6}{\sqrt{4\pi}} exp\left\{-\frac{1}{4}(\theta - 6,5)^2\right\}. \tag{6.30}$$

A opinião a priori do estatístico é

$$f(\theta)|\boldsymbol{\kappa}^* \sim PG\Big(g(\theta), \sigma^2 c(\theta, \theta)\Big), \tag{6.31}$$

em que a função média $g(\theta)$ é dada em (6.3) e $c(\theta, \theta)$ em (6.11).

Os valores de $P(\theta < x)$ elicitados do especialista foram

$$\begin{aligned}
P(\theta < 2) &= 0,06,\\
P(\theta < 3) &= 0,20,\\
P(\theta < 4) &= 0,36,\\
P(\theta < 5) &= 0,48,\\
P(\theta < 6) &= 0,62,\\
P(\theta < 7) &= 0,78,\\
P(\theta < 8) &= 0,91,\\
P(\theta < 9) &= 0,98.
\end{aligned}$$

Assim, o conjunto de dados utilizado na análise possui oito elementos.

Na estimação dos valores dos hiperparâmetros, m, b, η^* e σ^2, a distribuição conjunta a priori considerada é dada em (6.13). Utilizando o MCMC, algoritmo de Metropolis Hastings, foram geradas cadeias com 20000 iterações para cada hiperparâmetro. A Figura 26 mostra as cadeias geradas e respectivos gráficos de autocorrelação que sugerem convergência. Na Tabela 10 encontra-se as estimativas dos hiperparâmetros, variâncias e intervalos de credibilidade de 95%.

Tabela 9: Estimativas da média, variância e intervalos de credibilidade de 95% para os hiperparâmetros.

	$\hat{m}$	$\hat{b}$	$\hat{\eta}^*$	$\hat{\sigma}^2$
Média	4,784	4,153	0,675	0,955
Variância	0,158	0,429	0,048	0,359
I. Cred.	[3,863 ; 5,496]	[2,959 ; 5,346]	[0,273 ; 1,138]	[0,249 ; 2,440]

Utilizando os últimos 2000 valores das cadeias geradas para m, b, η^* e σ^2, foram geradas funções amostrais de $f(.)$ a partir da distribuição a posteriori apresentada em (6.27). A Figura 27 mostra que a função estimada está muito próxima da função verdadeira dada em (6.30).

Na fase *feedback*, o estatístico pode apresentar a função estimada, exibida na Figura 27, para que o especialista avalie se sua opinião está sendo bem representada. Porém, segundo O'Hagan e Oakley (2007), é mais fácil para o especialista avaliar a estimativa da distribuição acumulada, pois ele poderá observar se os valores de $P(\theta) < x$ para novos valores de x correspondem a sua opinião. A Figura 28 mostra a distribuição acumulada estimada.

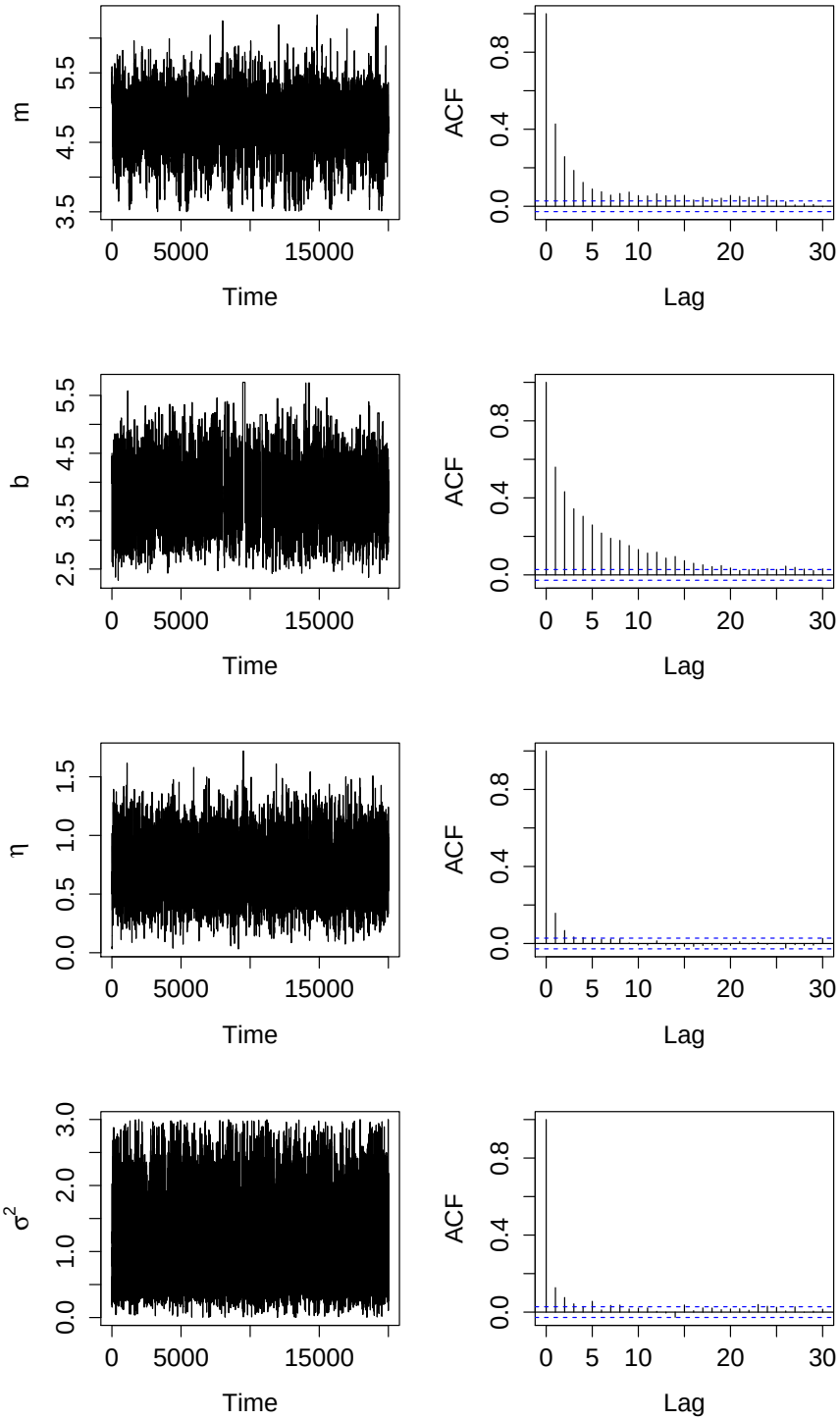


Figura 26: Trajetória da cadeia e respectivos correlogramas dos hiperparâmetros.

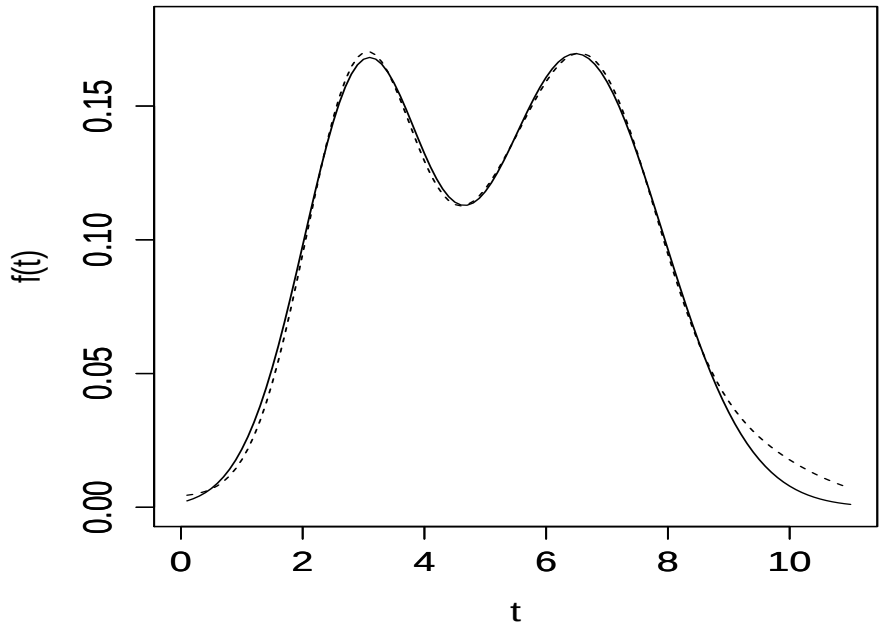


Figura 27: Função verdadeira (preta) e função estimada (tracejada).

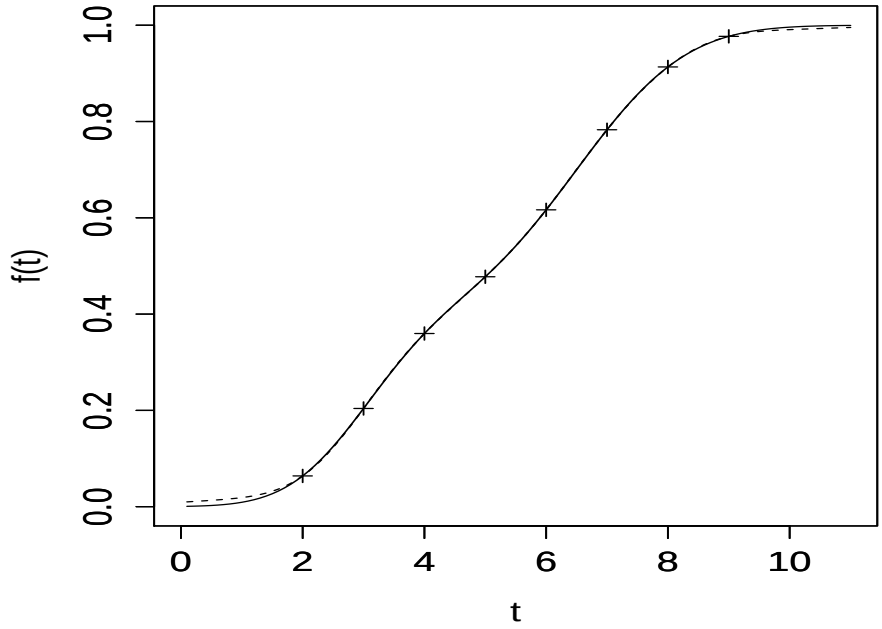


Figura 28: Distribuição acumulada verdadeira (preta), distribuição acumulada estimada (tracejada) e valores elicitados (×).

7 *Inferência Bayesiana Não-Paramétrica para Elicitação da Função de Confiabilidade*

Estudos já propuseram diversas metodologias para a análise do tempo de vida. Porém, a maior parte das pesquisas ignora o conhecimento do especialista e assume uma determinada família paramétrica para o tempo de vida, como Weibull, Exponencial, entre outras.

Nesse capítulo, discutiremos uma forma mais flexível de análise do tempo de falha, em que estimaremos uma distribuição a priori para a função de confiabilidade através da elicitação do conhecimento do especialista. A função de confiabilidade estimada representará então a opinião do especialista sobre o tempo de vida de determinado item em estudo.

No Capítulo 6, discutimos o método Bayesiano não-paramétrico desenvolvido por Oakley e O'Hagan (2007) para estimar a densidade desconhecida $f(.)$, em que $f(.)$ representa a opinião do especialista, utilizando processos Gaussianos e a elicitação. Nesse capítulo, aplicaremos esse método na análise de confiabilidade, assumindo que $f(.)$ é a função de confiabilidade, denotada por $R(.)$, e elicitando valores da confiabilidade.

7.1 Representação a priori para $R(.)$

Para o caso em que não assumimos nenhuma forma paramétrica para a função de confiabilidade, uma representação a priori adequada é dada por um processo Gaussiano. Assim, a distribuição a priori de $R(.)$ é dada por

$$R(.)|\boldsymbol{\kappa}_2 \sim PG\Big(\psi(.), \sigma^2 c(.,.)\Big), \tag{7.1}$$

em que $\boldsymbol{\kappa}_2$ é o vetor de hiperparâmetros, ψ é a função média e $\sigma^2 c(.,.)$ é a função de covariância.

Uma possível escolha para a esperança a priori de $R(.)$ é a função de confiabilidade

da distribuição Exponencial dada por

$$\psi(t) = exp\{-\lambda t\}. \tag{7.2}$$

Nesse caso, o estatístico acredita que a função de confiabilidade do especialista será similar a função de confiabilidade da distribuição Exponencial. Se essas condições não forem razoáveis, outra distribuição de probabilidade pode ser selecionada. A vantagem em se utilizar a função de confiabilidade da Exponencial como função média é que ela possui uma forma fechada e apenas um parâmetro.

Embora o estatístico acredite que a função de confiabilidade do especialista, $R(.)$, se aproxime da função de confiabilidade da Exponencial, o processo Gaussiano permite que o modelo se desvie dessa convicção de acordo com a quantidade de informações do especialista disponível. Isso foi ilustrado na Figura 25, que mostra a função média sendo alterada não-parametricamente para espelhar os dados.

Como visto na Seção 6.2, a função de covariância entre $R(t)$ e $R(z)$ pode ser escrita da seguinte forma

$$\begin{aligned} Cov(R(t), R(t')|\boldsymbol{\kappa}_2) &= \sigma^2\psi(t)\psi(t')c(t,t') \\ &= \sigma^2 exp\{-\lambda t\} exp\{-\lambda t'\} c(t,t'). \end{aligned} \tag{7.3}$$

Assumindo que a função de correlação é dada por

$$c(t,t') = exp\left\{-\frac{1}{2\eta}(t-t')^2\right\}, \tag{7.4}$$

temos que

$$Cov(R(t), R(t')|\boldsymbol{\kappa}_2) = \sigma^2 exp\left\{-\left(\lambda(t+t') + \frac{1}{2\eta}(t-t')^2\right)\right\}. \tag{7.5}$$

A distribuição a priori conjunta para os hiperparâmetros é dada por

$$\pi(\lambda, \eta, \sigma^2) = \frac{1}{\lambda\sigma^2}\pi(\eta) \tag{7.6}$$

em que $\pi(\eta)$ é a distribuição a priori proposta por Oakley e O'Hagan(2007) dada em (6.12).

7.2 Elicitando Valores da Confiabilidade

Vamos pedir ao especialista valores da confiabilidade, denotada por R_x, para determinados valores de x. Nesse caso, o vetor de dados será

$$\mathbf{d}^t = (R_{x_1}, \ldots, R_{x_n}). \tag{7.7}$$

em que $x_1, x_2, \ldots, x_n$ são n valores selecionados para os quais o especialista fornece a probabilidade do tempo de vida do objeto em estudo ser maior do que x_i.

Como $R(.)$ segue um processo Gaussiano, então R_x segue um processo Gaussiano. Assim, o vetor $\mathbf{d}$ condicionado em $\boldsymbol{\kappa}_2$ é normalmente distribuído (por definição de processo Gaussiano),

$$\mathbf{d}|\boldsymbol{\kappa}_2 \sim N_n(\mathbf{w}_2, \sigma^2 A_2), \tag{7.8}$$

em que os elementos do vetor de média $\mathbf{w}_2$ são dados por

$$E(R_x|\boldsymbol{\kappa}_2) = exp\{-\lambda x\}. \tag{7.9}$$

Já os elementos da matriz de covariância $\sigma^2 A_2$ são dados por

$$Cov\left(R_x, R_y|\,\boldsymbol{\kappa}_2\right) = \sigma^2 exp\left\{-\left(\lambda(x+y) + \frac{1}{2\eta}(x-y)^2\right)\right\}, \tag{7.10}$$

onde $\boldsymbol{\kappa}_2 = (\lambda, \eta, \sigma^2)$.

Logo, a função de verossimilhança é dada por

$$L(\lambda, \eta, \sigma^2|\mathbf{d}) = \frac{|A_2|^{-\frac{1}{2}}}{\sigma^n} exp\left\{-\frac{1}{2\sigma^2}(\mathbf{d}-\mathbf{w}_2)^t A_2^{-1}(\mathbf{d}-\mathbf{w}_2)\right\}. \tag{7.11}$$

7.3 Atualização da priori

A covariância a priori entre $R(t)$ e $\mathbf{d}$ é dada por

$$Cov\Big(R(t), \mathbf{d}|\boldsymbol{\kappa}_2\Big) = \begin{bmatrix} Cov\Big(R(t), R_{x_1}|\boldsymbol{\kappa}_2\Big) \\ Cov\Big(R(t), R_{x_2}|\boldsymbol{\kappa}_2\Big) \\ \vdots \\ Cov\Big(R(t), R_{x_n}|\boldsymbol{\kappa}_2\Big) \end{bmatrix}$$

em que

$$Cov\Big(R(t), R_x|\boldsymbol{\kappa}_2\Big) = \sigma^2 exp\left\{-\left(\lambda(t+x) + \frac{1}{2\eta}(t-x)^2\right)\right\}. \tag{7.12}$$

A distribuição a posteriori $R(.)|\boldsymbol{\kappa}_2, \mathbf{d}$ é um processo Gaussiano com função média

$$E(R(t)|\boldsymbol{\kappa}_2, \mathbf{d}) = \psi(t) + \mathbf{k}_2(t)^t A_2^{-1}(\mathbf{d} - \mathbf{w}_2) \tag{7.13}$$

e função de covariância

$$Cov(R(t), R(t')|\boldsymbol{\kappa}_2, \mathbf{d}) = \sigma^2\Big(C(t,t') - \mathbf{k}_2(t)^t A_2^{-1}\mathbf{k}_2(t')\Big). \tag{7.14}$$

em que $\mathbf{k}_2(.)$ é a covariância entre $R(t)$ e $\mathbf{d}$ - os elementos desse vetor são dados em (7.12)- e $C(t,t')$ é a covariância a priori entre $R(t)$ e $R(t')$ dada em (7.5).

A Seção 7.4 ilustra uma aplicação do método apresentado nesse capítulo. Como o treinamento de um especialista para a participação do processo de elicitação demanda muito tempo, os dados utilizados nesse exemplo foram gerados no pacote estatístico R.

7.4 Ilustração Numérica

Suponha que estamos interessados em estudar o tempo de vida de um componente de um nova aeronave. Por se tratar de um componente com altíssimo custo de produção, poucas unidades foram testadas. Resolveram então utilizar a elicitação do conhecimento do especialista para a resolução do problema.

A opinião a priori do estatístico é

$$R(t) \sim PG\Big(\psi(t), \sigma^2 c(t,z)\Big),$$

em que $\psi(t)$ é dada em (7.2) e $c(t,t') = exp\left\{-\frac{1}{2\eta}(t-t')^2\right\}$.

O projetista, que desenvolveu esse componente durante anos e que tem bastante conhecimento sobre o seu funcionamento, foi o especialista selecionado. Ele foi submetido a um treinamento, com o intuito de familiarizá-lo com o processo de elicitação.

Assumindo que $R(.)$ verdadeira é a função de confiabilidade da distribuição Weibull com parâmetros $\alpha = 4$ e $\beta = 1,3$, temos que

$$R(t) = exp\left\{-\left(\frac{t}{4}\right)^{1,3}\right\}. \tag{7.15}$$

O especialista foi solicitado a fornecer os valores de x_i que satisfazem:

$$\begin{aligned} q_1 &= P(X > x_1) = 0,10, \\ q_2 &= P(X > x_2) = 0,25, \\ q_3 &= P(X > x_3) = 0,50, \\ q_4 &= P(X > x_4) = 0,75, \\ q_5 &= P(X > x_5) = 0,90. \end{aligned}$$

Em um primeiro momento, admitimos que o especialista seria capaz de fornecer as especificações probabilísticas exatas de sua distribuição. Assim, os valores elicitados foram: $x_1 = 7,60$, $x_2 = 5,14$, $x_3 = 3,02$, $x_4 = 1,53$ e $x_5 = 0,71$. Com a informação do conhecimento do especialista transformada em percentis elicitados, o estatístico trata-os como um conjunto de dados $\mathbf{d}$ que serão utilizados para estimar os hiperparâmetros λ, η e σ^2.

Cadeias do MCMC, com 20000 iterações, foram geradas para λ, η e σ^2. A Figura 29 mostra as cadeias geradas pelo método MCMC e respectivos gráficos de autocorrelação que sugerem convergência. As estimativas dos hiperparâmetros, suas variâncias e intervalos de credibilidade de 95% são apresentadas na Tabela 11.

Tabela 10: Estimativas da média, variância e intervalos de credibilidade de 95% para os hiperparâmetros.

	$\hat{\lambda}$	$\hat{\eta}$	$\hat{\sigma}^2$
Média	0,283	0,704	0,046
Variância	0,002	0,058	0,005
I. Cred.	[0,186 ; 0,378]	[0,238 ; 1,179]	[0,025 ; 0,102]

Utilizando os últimos 2000 valores das cadeias geradas para λ, η e σ^2, geramos funções amostrais de $R(.)$ a partir da posteriori apresentada em (7.13) e (7.14). A Figura 30 mostra que a função de confiabilidade estimada está próxima da verdadeira, principalmente nas áreas onde existem informações provenientes dos dados.

Como é difícil para o especialista fornecer as especificações probabilísticas exatas, consideramos um erro $N(0,1)$ para cada um dos valores elicitados. Assim, os valores elicitados foram: $x_1 = 7,13$, $x_2 = 4,89$, $x_3 = 2,43$, $x_4 = 1,37$ e $x_5 = 0,63$.

Realizando o mesmo procedimento descrito anteriormente, obtivemos a função de confiabilidade estimada considerando um erro nos valores elicitados. A Figura 31 mostra que a função de confiabilidade estimada com erro não está tão próxima das demais, porém ainda pode ser considerada uma boa estimativa da função de confiabilidade verdadeira.

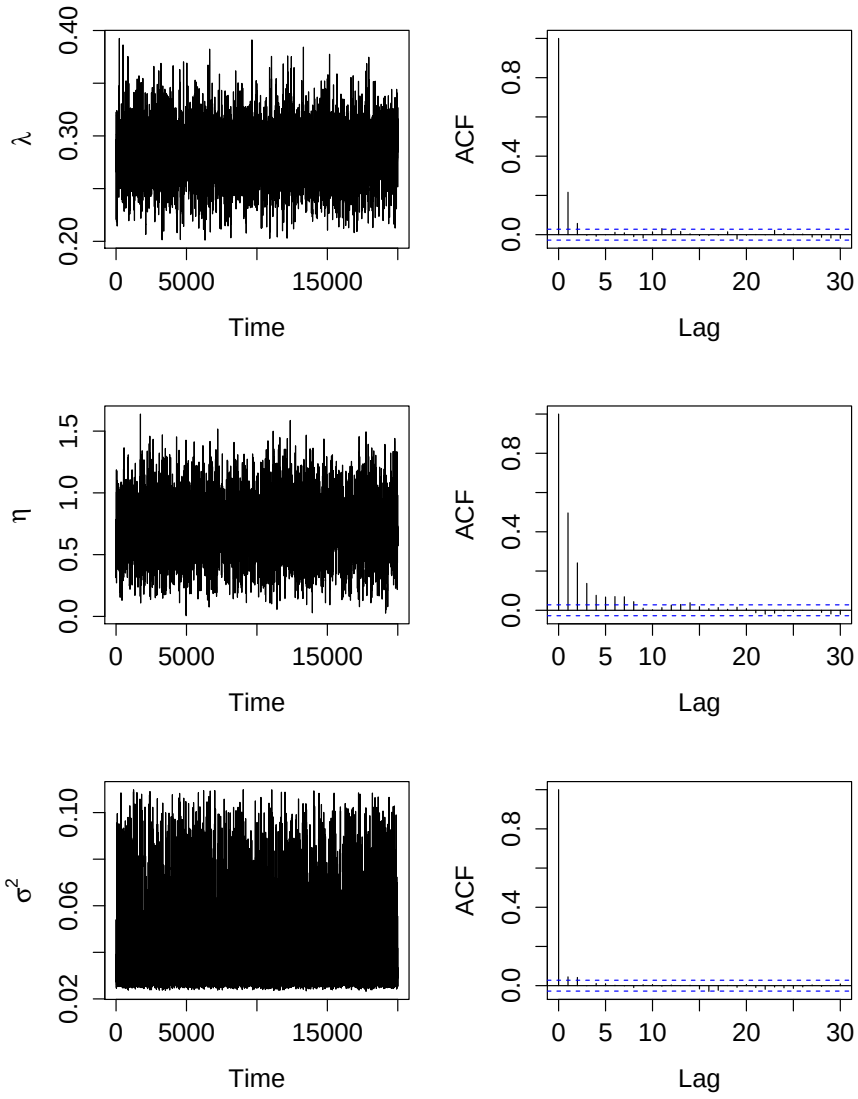


Figura 29: Cadeias geradas pelo MCMC para os hiperparâmetros λ, η e σ^2 e correspondentes correlogramas.

É importante ressaltar que a função de confiabilidade estimada deve ser apresentada ao especialista, para que ele avalie se o resultado obtido representa bem a sua opinião.

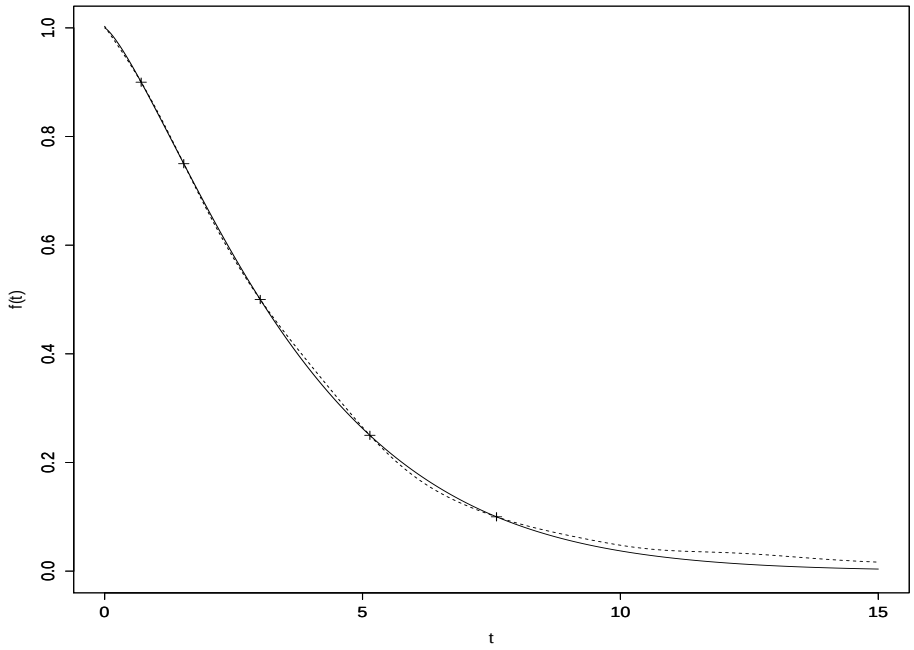


Figura 30: Função de confiabilidade verdadeira (sólida), função de confiabilidade estimada (tracejada) e valores elicitados (×).

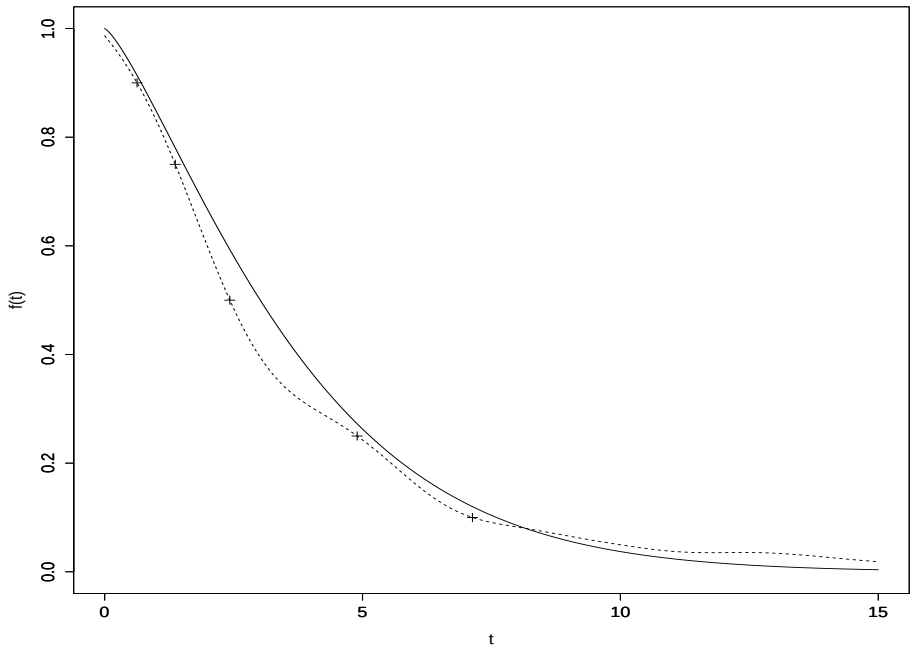


Figura 31: Função de confiabilidade verdadeira (sólida), função de confiabilidade estimada com erro (tracejada) e valores elicitados (×).

8 Conclusões

A análise de confiabilidade é um dos ramos da estatística mais pesquisados nos últimos anos. As mais diversas distribuições de probabilidade já foram consideradas para descrever o tempo de falha de determinado item, dentre elas a distribuição Weibull. É comum, sob o enfoque Bayesiano paramétrico, a análise de confiabilidade ser realizada utilizando distribuições a priori não-informativas, ignorando-se assim possíveis informações a priori existentes. No Capítulo 2, considerando dados de tempo de vida - com e sem censura - provenientes de uma distribuição Weibull, foram avaliadas algumas prioris não-informativas, como priori de Jeffreys, referência e cópula, que apresentaram resultados similares.

A elicitação é o processo de se extrair o conhecimento de um especialista sobre alguma quantidade desconhecida na forma de uma distribuição de probabilidade. Diversos métodos de elicitação podem ser encontrados na literatura, porém, a maioria requer a elicitação de momentos de primeira e segunda ordem. Dessa forma, a aplicação desses métodos torna-se inviável. No Capítulo 3, alguns métodos de elicitação paramétrica - que não empregam a média e a variância para a representação da opinião do especialista na forma de uma distribuição - foram discutidos.

A utilização da elicitação na análise de confiabilidade pode proporcionar conclusões mais realísticas, particularmente quando o tamanho amostral é pequeno. Foi proposto, no Capítulo 4, um método aplicado à distribuição Weibull com parâmetros α e β que utiliza algumas probabilidades elicitadas do especialista, $P(T \geq x)$, para a determinação dos hiperparâmetros da distribuição conjunta a priori, considerando a distribuição a priori de α uma Gama(a, b) e para β uma Gama(c, d). O Exemplo 7, que ilustra o método, mostra que o emprego da opinião do especialista produziu uma melhora significativa nos resultados, principalmente quando os dados não são numerosos.

Os métodos de elicitação paramétrica podem restringir muito a representação da opinião do especialista sobre o tempo de vida de um item em pesquisa, por dependerem de especificações restritivas, ajustamento de uma distribuição de probabilidade conhecida. A metodologia não-paramétrica é mais flexível e pode ser utilizada na análise de confiabilidade. Nessa dissertação, aplicamos o método Bayesiano não-paramétrico de Oa-

kley e O'Hagan e mostramos que é possível utilizar a elicitação para estimar a função de confiabilidade.

A função de confiabilidade estimada na Seção 7.4 mostrou que a metodologia Bayesiana não-paramétrica de elicitação produz bons resultados. Essa função de confiabilidade estimada representa a opinião do especialista sobre o tempo de vida de determinado item em estudo e pode ser usada como distribuição a priori em uma análise Bayesiana. Assim, a informação elicitada do especialista pode ser usada para complementar a informação dos dados observados.

O objetivo do presente trabalho foi alcançado, utilizando o método proposto por Oakley e O'Hagan (2007) foi possível estimar uma distribuição a priori para a confiabilidade a partir das informações fornecidas por um especialista, de uma forma flexível que pode ser aplicada a inúmeros estudos sobre o tempo de vida.

Em trabalhos futuros, a aplicação do método Bayesiano não-paramétrico utilizado para estimar a função de confiabilidade, poderia considerar outras funções de confiabilidade como função média do processo Gaussiano, como a função de confiabilidade da distribuição Weibull. O desempenho de outras funções de covariância que podem ser utilizadas no processo Gaussiano, como a Laplaciana (apresentada na Seção 5.3), poderia também ser avaliado.

Esperamos que esse trabalho provoque o interesse da comunidade científica pela elicitação, de forma que mais estatísticos possam aplicá-la em muitos problemas práticos.

Referências

ANG, M.L.; BUTTERY, N.E. An approach to the application of subjective probabilities in level 2 PSAs. **Reliability Engineering and System Safety**, v.58, p. 145-156, 1997.

BARTHOLOMEW, D. J. A Problem in Life Testing. **J. Am. Stat. Assoc.**, p. 350-355, 1957.

BEACH, L. R.; SWENSON, R. G. Intuitive estimation of means. **Psychonomic Science**, v. 5, p. 161-162, 1966.

BEDFORD, T.; COOKE, R. **Probabilistic risk analysis**: foundations and methods. Cambridge: Cambridge University Press, 2001. 393 p.

BERGER, J. O, BERNARDO, J. M. On the Development of the Reference Prior Method. In: Fourth Valencia International Meeting on Bayesian Statistics. Espanha, 1992.

BERNARDO, J. M. Reference posterior distributions for Bayesian inference. **J. Roy. Statist. Soc. Edward Elgar**, Brookfield, 1979.

BERNARDO, J. M. The concept of exchangeability and its applications. **Far East J. Mathematical Sciences**, v. 4, p. 111-121, 1996.

BERNARDO, J. M. Reference analysis. **Elsevier**, p. 17-90, 2005.

CAMPOSILVAN, D. M. Uma Contribuição à Avaliação de Contratos Bilaterais de Suprimento de Energia Elétrica. 2003. 146 p. Dissertação (Mestrado em Ciências em Engenharia Elétrica)- Universidade Federal de Itajubá, Itajubá.

COOKE, R. **Experts in Uncertainty: Opinion and Subjective Probability in Science**. Oxford University Press, Nova York, 1991.

COOLEN, F.P.A. et al. A Bayes-competing risk model for the use of expert judgement in reliability estimation.**Reliability Engineering and System Safety**, v. 35, p. 23-30, 1992.

DANESHKHAH, A. Psychological aspects influencing elicitation of subjective probability. BEEP research report, Universidade de Sheffield, 2004 .

DOMINITZ, J. Earning expectations, revisions, and realisations. The Review of Economics and Statistics, v. 80, n.3, p. 374-388, 1998.

FOX, B.L. **A Bayesian approach to reliability assessment**. Memorandum RM-5084-NASA, The Rand Corporation, Santa Monica, USA. 1966.

GARTHWAITE, P.H.; KADANE, J.B.; O'HAGAN, A. Statistical methods for eliciting probability distributions. **Journal of the American Statistical Association**, v. 100, p. 680-701, 2005.

GOSLING, J.P. **Elicitation: A Nonparametric View**. 2005. Tese (Doutorado em Estatística). Departmento de Estatística. Universidade de Sheffield. Inglaterra.

GRISLEY, W.; KELLOGG, E.D. Farmers subjective probabilities in Northern Thailand: an elicitation analysis. **American Journal of Agricultural Economics**, v. 65, p. 74-82, 1982.

GROSS, A.J. The application of exponential smoothing to reliability assessment. **Technometrics**, v. 65, p. 877-883, 1971.

GUSTAFSON, D.H. et al. Developing and testing a model to predict outcomes of organizational change. **Health Services Research**, v.38, p. 751-776, 2003.

HUGHES, G.; MADDEN, L. V. Some methods for eliciting expert knowledge of plant disease epidemics and their application in cluster sampling for disease incidence. **Crop Protection**, v. 21, p. 203-215, 2002.

JEFFREYS, H. **Theory of Probability**, 3rd ed. Oxford Univ. Press, 1967.

KADANE, J.B.; TIERNEY, L. Accurate Approximations for Posterior Moments and Marginal Densities. **Journal of the American Statistical Association**, v. 81, p. 82-86, 1986.

KADANE, J.B.; WOLFSON, L.J. Experiences in elicitation. **Journal of the Royal Statistical Society**, v. 47, p. 3-19, 1998.

KENNEDY, M.C.; O'Hagan, A. Iterative rescaling for Bayesian quadrature. In Bayesian Statistics, Oxford:Oxford University Press, p. 639?645, 1996.

KAMINSKIY, M.P.; KRIVTSOV, V.V. A Simple Procedure for Bayesian Estimation of the Weibull Distribution, **IEEE Trans. Reliability**, v. 54, p. 612-616, 2005.

LAWLESS, J.F. **Statistical Models and Methods for Lifetime Data**. Nova York: Wiley, 1982. 580 p.

LAWS, D.J.; O'HAGAN, A. A hierarchical Bayes model for multiloca-tional auditing. **Journal of the Royal Statistical Society**, v.51, p. 431-450, 2002.

LIEBLEIN, J.; ZELEN, M. Statistical investigation of the fatigue life of deep groobe ball bearings. **J. Res. Nat. Bur. Stand.**, p. 273-316, 1956.

MANN, N.R. Point and Interval Estimation for the Two-parameter Weibull and Extreme-value distribution. **Technometrics**, v.10, p. 231-256, 1968.

MANN, N.R.; FERTIG, K.W. Tables for obtaining confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme value distribution. **Technometrics**, p. 87-101, 1973.

MARTZ, H.F. ; Waller, R.A. **Bayesian Reliability Analysis**. Nova York: Wiley, 1982.

MOALA, F.A. **Elicitation of Multivariate Prior Distributions**. 2006. Tese (Doutorado em Estatística). Department of Statistics. University of Sheffield. Inglaterra.

MOALA, F. A. Elicitação da distribuição a priori em uma pesquisa médica. **Revista Brasileira de Estatística**, v. 70, p. 57-74, 2009.

MOALA, F.A.; RODRIGUES, J.; TOMAZELLA, V.L.D. A note on the prior distributions of Weibull parameters for the reliability function. **Communications in Statistics - Theory and Methods**. v. 38, p. 1041-1054, 2009.

MOALA, F. A.; O'HAGAN, A. Elicitation of Multivariate Prior Distributions: A nonparametric Bayesian. Journal of Statistical Planning and Inference (Print), p. 1-42, 2010.

MORGENSTERN, D. Einfache Beispiele Zweidimensionaler Verteilungen. **Mitteilingsblatt fur Matgematische Statistk**, v. 8, p. 234-253, 1956.

OAKLEY, J. E.; O'HAGAN, A. Bayesian inference for the uncertainty distribution of computer model outputs. **Biometrika**, v. 89, p. 769-784, 2002.

OAKLEY, J. E.; O'HAGAN, A. Uncertainty in prior elicitations: a nonparametric approach. **Biometrika**, v. 94, p. 427-441, 2007.

O'HAGAN, A. Curve Fitting and Optimal Design Theory. **Journal of Royal Statistical Society**, v. 40, p. 1-42, 1978.

O'HAGAN, A. Eliciting expert beliefs in substantial practical applications. **Journal of Royal Statistical Society**, v. 47, p. 21-35, 1998.

PAPOULIS, A. **Probability, Random Variables and Stochastic Processes**. Nova York: McGraw-Hill, 1965. 582 p.

PETERSON, C. R.; MILLER, A. Mode, median and mean as optimal strategies. **Journal of Experimental Psychology**, v.68, p. 363-367, 1964.

PRESS, S.J. Applied Multivariate Analysis. Nova York: Holt, Rinehart and Winston Inc. 1972. 521 p.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0.

Disponível em: <http://www.R-project.org/>.

RASMUSSEN, C, E; Williams, C, K, I. **Gaussian Processes and Machine Learning**. 1. ed. Massaschusetts Institute of Tecnology Press, 2006.

RASMUSSEN, C, E. Gaussian Processes in Machine Learning. **Springer**, v. 3176, p. 63-71, 2004.

SINGPURWALLA, N.D. An Interactive PC-Based Procedure for Reliability Assessment Incorporating Expert Opinion and Survival Data. **Journal of the American Statistical Association**, v.83, p. 43-51, 1988.

SPENCER, J. Estimating averages.**Ergonomics**, v.4, p. 317-328, 1961.

SPRINGER, M. D.; THOMPSON, W. E. Bayesian Confidence Limits for Reliability of Redundant Systems when Tests are Terminated at First Failure. **Technometrics**, v. 10, p. 29-36, 1965.

SPRINGER, M. D.; THOMPSON, W. E. Bayesian Confidence Limits for the Reliability of Cascade Exponential Subsystems. **IEEE Transactions on Reliability**, v. R-16, p. 86-89, 1967.

WEILER, H. The Use of Incomplete Beta Functions for Prior Distributions in Binomial Sampling. **Technometrics**, v.7, p. 335-347, 1965.

WINKLER, R.L. The assessment of prior distributions in Bayesian analysis. **Journal of the American Statistical association**, v. 62, p. 776-800, 1967.