

Modelo de Regressão de Cox com Verossimilhança Monótona

Frederico Machado Almeida

Departamento de Estatística - ICEX - UFMG

17 de Fevereiro de 2017

Modelo de Regressão de Cox com Verossimilhança Monótona

Frederico Machado Almeida

Orientador: Prof. Dr. Enrico Antônio Colosimo

Coorientador: Prof. Dr. Vinícius Diniz Mayrink

Dissertação submetida ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Departamento de Estatística
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Belo Horizonte, MG - Brasil

17 de Fevereiro de 2017

Aos meus pais

”Não durma perto da estrada que as poeiras irão sujar seus sonhos. E aconteceu. Mas eu..., eu sempre gostei de poeira porque me traz ilusão dos caminhos que não conheço”.

Mia Couto

Agradecimentos

Em primeiro lugar agradeço a Deus pelo dom de vida e, por me fortalecer dia após dia.

Aos meus pais, Machado Almeida e Matilde Marcelino pelos ensinamentos e valores morais que sempre implantaram em mim.

A toda minha família, em especial ao meu irmão José Machado que incansavelmente continua me ajudando e me encorajando em todos os momentos.

Ao professor Enrico A. Colosimo meu orientador, por ter me facultado a carta de aceite, pela paciência e, acima de tudo pelos conhecimentos que incansavelmente vem transmitindo.

Ao professor Vinícius D. Mayrink meu co-orientador, pela paciência e pelas valiosas ideias e sugestões que, de forma incansável vem transmitido.

A todos os professores do Departamento de Estatística do ICEX pelos conhecimentos transmitido, acima de tudo pela paciência.

A todos os professores do Departamento de Matemática e Informática da Universidade Eduardo Mondlane, em especial ao MSc. Alberto Mulenga e MSc. Osvaldo André Loquiha, vai um especial agradecimento pelos conhecimentos transmitidos durante a graduação.

A todos os meus amigos e colegas, em especial ao Leonardo, Arthur, Jussiane, Erick, Rumenik e Bruno pelos momentos de convívio que juntos passamos.

A Rose, Rogéria e a Stepfany, secretarias da pós-graduação vai um agradecimento pela ajuda incondicional que me proporcionaram.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro concebido durante os dois anos que frequentei o mestrado.

Resumo

Quando se modela estudos clínicos ou epidemiológicos envolvendo eventos raros é comum obter amostras com alto percentual de censura. A presença de covariáveis binárias nas análises pode criar problemas durante o processo de estimação fazendo com que a função de verossimilhança não seja maximizada para determinados parâmetros. Este fenômeno é conhecido na literatura por verossimilhança monótona, e ocorre quando um dos níveis da covariável binária não está associado ao evento de interesse. Uma solução para o problema baseada na inferência Clássica, foi sugerida por Heinze e Schemper (2001) sendo uma adaptação do procedimento de Firth (1993), originalmente desenvolvido para reduzir o viés dos estimadores de máxima verossimilhança. Este procedimento garante a obtenção de estimativas finitas para os parâmetros. No entanto, a solução tem a desvantagem de fornecer estimativas viesadas associadas a erros-padrão elevados. Neste trabalho, propomos corrigir o problema usando a abordagem Bayesiana, onde diferentes funções de penalidade (distribuições *a priori*) para os parâmetros foram testadas, como por exemplo a $N(m, v)$, $\text{Cauchy}(l, s)$ e $\log-F(a, b)$. Um estudo de simulação foi desenvolvido para avaliar o comportamento assintótico das estimativas. Finalmente, realiza-se uma aplicação destes procedimentos a um banco de dados real referente a pacientes com melanoma.

Palavras-chave: Análise de Sobrevivência, Verossimilhança Monótona, Correção de Firth, Abordagem Bayesiana, Melanoma.

Abstract

When modeling clinical and epidemiological studies involving rare events, it is common to obtain strongly censored samples. The presence of binary covariates in the analysis may cause problems during the estimation process so that the likelihood function does not maximize for certain parameters. This phenomenon is known in literature as monotone likelihood and occurs when at least one level of the binary covariate has not experienced any failure. A solution suggested by Heinze e Schemper (2001) is an adaptation of the procedure in Firth (1993) originally developed to reduce the bias of maximum likelihood estimation. This method leads to finite parameter estimates; however, it has some drawbacks, especially biased estimators and high standard errors. In this study, we explore the Bayesian approach as a solution to the monotone likelihood problem and investigate other penalty functions (*prior* distributions); for example, $N(m, v)$, $\text{Cauchy}(l, s)$ and the $\log-F(a, b)$. A simulation study was developed to evaluate the asymptotic behavior of the estimates. Finally, a real application is developed to study a melanoma data set.

Keywords: Survival Analysis, Monotone Likelihood, Firth Correction, Bayesian Approach, Melanoma.

Conteúdo

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Contribuição do trabalho | 3 |
| 1.2 | Organização da dissertação | 3 |
| 2 | Análise de sobrevivência | 5 |
| 2.1 | Dados Censurados | 5 |
| 2.2 | Função de sobrevivência e função taxa de falha | 6 |
| 2.3 | Modelo de regressão de Cox | 7 |
| 2.3.1 | Função de Verossimilhança | 8 |
| 2.3.2 | Função de verossimilhança parcial | 10 |
| 2.4 | Distribuição Weibull | 10 |
| 2.4.1 | Modelo de regressão Weibull | 12 |
| 2.4.2 | Resumo do capítulo | 13 |
| 3 | Verossimilhança monótona | 14 |
| 3.1 | Métodos de redução do viés dos EsMV | 15 |
| 3.1.1 | Métodos explícitos | 15 |
| 3.1.2 | Métodos implícitos | 16 |
| 3.2 | Resumo do capítulo | 19 |
| 4 | Inferência Bayesiana | 20 |
| 4.1 | Distribuições a <i>priori</i> a serem estudadas | 21 |
| 4.1.1 | <i>Priori</i> de Jeffreys | 21 |

| | | |
|----------|--|-----------|
| 4.1.2 | Outras distribuições a <i>priori</i> | 22 |
| 4.2 | Amostrador de Gibbs | 23 |
| 4.3 | Algoritmo de Metropolis-Hastings | 25 |
| 4.3.1 | Resumo do capítulo | 27 |
| 5 | Estudo simulado | 28 |
| 5.1 | Resultados para a regressão Weibull | 30 |
| 5.2 | Resultados para o modelo de Cox | 33 |
| 5.2.1 | Abordagem clássica | 33 |
| 5.2.2 | Abordagem Bayesiana | 34 |
| 5.3 | Conclusão do capítulo | 35 |
| 6 | Aplicação a dados de melanoma | 36 |
| 7 | Considerações finais | 41 |
| 7.1 | Trabalhos futuros | 43 |
| | Apêndice | 44 |

Capítulo 1

Introdução

Por mais de três décadas, o modelo de riscos proporcionais de Cox (1972) tem sido o mais popular para modelar dados de sobrevivência. A flexibilidade bem como a facilidade na interpretação das estimativas dos parâmetros, como logaritmo do risco relativo, favorecem a sua popularidade em pesquisas médicas. Nesse tipo de pesquisa, principalmente quando envolvem eventos raros, é comum obter amostras com pequeno número de eventos, como consequência, as estimativas de máxima verossimilhança tendem a ser viesadas, podendo não assumir um valor finito. Este fenômeno é conhecido na literatura como problema da verossimilhança monótona (VM), e ocorre durante o processo de estimação no modelo de regressão de Cox quando as estimativas convergem para um valor finito, mas existe pelo menos uma estimativa que diverge para $\pm\infty$ Heinze e Schemper (2001).

O problema da VM pode ocorrer nas situações que envolvem amostras de tamanho pequeno e com alto percentual de censuras. Em geral, ela ocorre na presença de covariáveis binárias quando um dos seus níveis não está associado ao evento de interesse. Nesse sentido, quanto maior o número de regressores binários incluídos no modelo, maior são as chances de ocorrência do fenômeno. Este problema raramente aparece quando modelamos apenas covariáveis contínuas ou amostras não censuradas.

Este trabalho foi motivado a partir de um banco de dados de pacientes com melanoma, que é um tumor agressivo e de alta mortalidade quando diagnosticado em estágios avançados. A principal estratégia para a redução da mortalidade baseia-se na identificação precoce do melanoma primário e daqueles pacientes em risco de desenvolver a

“metástase” Damato et al. (2011). Com o objetivo de avaliar a influência de cinco fatores epidemiológicos e histopatológicos no desenvolvimento de “metástase” em pacientes com diagnóstico de melanoma cutâneo primário invasivo, um estudo foi desenvolvido entre 1995 e 2012 no Serviço de Dermatologia do Hospital das Clínicas da Universidade Federal de Minas Gerais e no serviço privado Oncologia Cirúrgica do Aparelho Digestivo (ONCAD). O modelo de regressão de Cox (1972) foi ajustado onde constatou-se que a estimativa associada a covariável “mitose” uma das mais importantes no modelo não assume um valor finito, isto significa a ocorrência da VM.

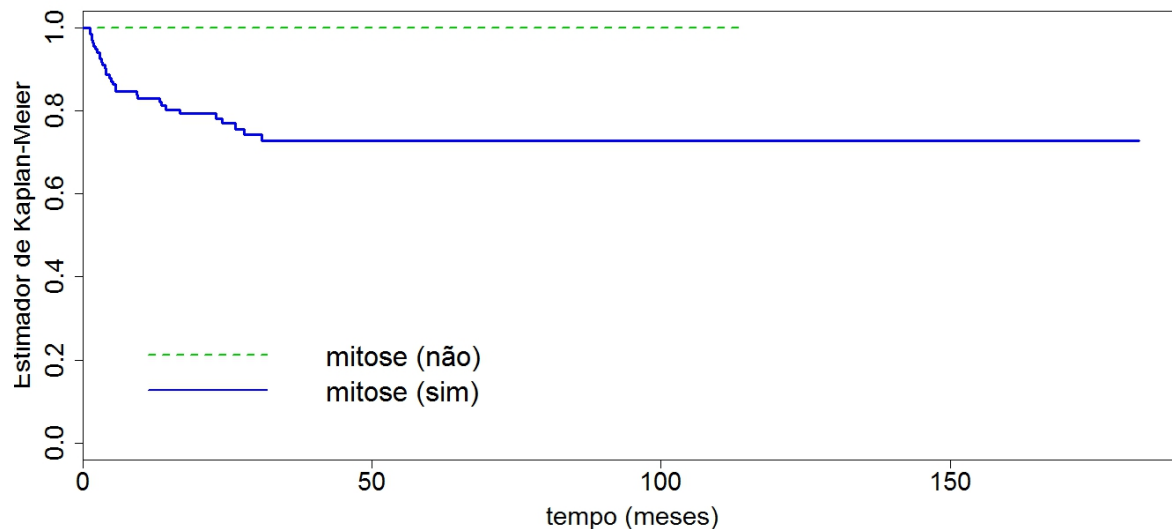


Figura 1.1: *Representação gráfica do estimador de Kaplan-Meier para a covariável mitose.*

Como ilustramos na Figura 1.1, um dos níveis da covariável “mitose” não está associado ao evento de interesse, pois a estimativa da função de sobrevivência em pacientes sem “mitose” é constante no tempo, assim a estimativa do parâmetro associado a essa covariável não vai ter um valor finito; no Apêndice A, apresentamos de forma resumida a fórmula para calcular o estimador de Kaplan-Meier.

Uma proposta para a solução do problema da VM foi sugerida por Heinze e Schemper (2001), sendo baseada no procedimento de Firth (1993), originalmente desenvolvido com

objetivo de reduzir o viés dos estimadores de máxima verossimilhança (EsMV) no contexto de modelos lineares generalizados. Este método de penalização garante a obtenção de estimativas finitas e com variância menor.

A penalização é um método muito geral de estabilização das estimativas, que pode ser vista sob duas abordagens de inferência, a Clássica e a Bayesiana. O método de Firth é um exemplo bem conhecido de penalização que pode ser derivado com base na *priori* de Jeffreys (1946). No entanto, este método tem algumas desvantagens, pois tende a produzir estimativas tendenciosos com erros-padrão elevados.

1.1 Contribuição do trabalho

Neste trabalho investigamos o problema da VM propondo uma metodologia de correção sob o ponto de vista Bayesiano. Uma análise de sensibilidade é feita usando diferentes estruturas de distribuições *a priori* para os parâmetros como a $\log-F(a, b)$, $N(m, v)$ e a $\text{Cauchy}(l, s)$. Nesta análise, são testados diferentes hiperparâmetros das especificações *a priori* fornecendo diferentes níveis de informação: o objetivo é avaliar seu impacto sobre os resultados.

A escolha da distribuição $\log-F$ foi motivada pelo trabalho de Greenland e Mansournia (2015), que investigaram a questão da VM tanto na modelagem logística quanto em dados de sobrevivência. Os autores comparam algumas configurações de distribuições *a priori* pouco informativas como forma de solucionar o problema. As principais conclusões do trabalho são tiradas de uma análise de um banco de dados real referente a mortalidade neonatal, entretanto, eles não exploram dados simulados para avaliar e comparar o comportamento das estimativas dos parâmetros, especialmente em termos de viés, ao aumentar o tamanho da amostra ou escolher diferentes níveis de informação *a priori*.

1.2 Organização da dissertação

A presente dissertação é composta por sete capítulos e segue a seguinte organização:

O Capítulo 2, que apresentamos a seguir, faz uma descrição de alguns conceitos

em análise de sobrevivência, com principal enfoque nos modelos de regressão de Cox e Weibull. Apresentamos também os métodos de construção das funções de verossimilhança e intervalos de confiança.

O Capítulo 3 descreve alguns procedimentos de correção do problema da VM; os métodos explícitos e implícitos são discutidos com principal enfoque sobre a proposta de correção de Firth.

O Capítulo 4 apresenta alguns conceitos básicos em inferência Bayesiana, onde descrevemos algumas distribuições *a priori* que são o principal foco deste trabalho. Apresentamos ainda alguns algoritmos de simulação, como por exemplo, o Metropolis-Hastings e o *Gibb Sampling*.

No Capítulo 5 fizemos uma aplicação envolvendo os modelos discutidos no Capítulo 2 com objetivo de corrigir o problema da VM no estudo simulado.

Uma aplicação baseada no banco de dados de melanoma é apresentada no Capítulo 6; aqui o ajuste é referente ao modelo de regressão de Cox considerando as funções de penalidade que descrevemos no Capítulos 4.

Por fim, o Capítulo 7 faz um resumo do material apresentado e discutido neste trabalho, além de apresentar algumas propostas para trabalhos futuros.

Capítulo 2

Análise de sobrevivência

A análise de sobrevivência engloba um conjunto de técnicas estatísticas que tem como objetivo principal modelar o tempo até a ocorrência de um determinado evento de interesse. Uma importante característica dos dados de sobrevivência é a presença de censuras, que são observações parciais (ou incompletas) da resposta. Por exemplo, na medicina isso acontece quando o acompanhamento de algum paciente é interrompido por alguma razão ou porque o estudo terminou antes do paciente experimentar o evento de interesse; Colosimo e Giolo (2006).

2.1 Dados Censurados

A seguir descrevemos alguns tipos de mecanismos de censura. Uma discussão mais detalhada sobre o assunto pode ser encontrada em Colosimo e Giolo (2006), Kleinbaum e Klein (2012), Collett (2003), Klein e Moeschberger (2003) ou Lawless (2003).

- a) **Censura à direita:** estamos perante este tipo de censura quando apenas sabemos que o tempo de ocorrência do evento está à direita do tempo registrado, ou seja, o tempo de falha é maior que o tempo registrado. Por exemplo, supomos que se pretende avaliar o efeito de um determinado tipo de vacina para um grupo de crianças. Se for conduzido um ensaio clínico aleatorizado, o último indivíduo no estudo pode registar o evento de interesse muitos anos após o término do estudo. Para essa situação naturalmente temos um mecanismo de censura à direita. Este tipo de censura subdivide-se em:

- **Censuras do tipo I:** ocorrem naqueles estudos que ao serem finalizados após um período pré-estabelecido de tempo, regista-se em seu término alguns indivíduos ou itens que ainda não experimentaram o evento de interesse. Neste caso, o estudo é iniciado com um número fixo de indivíduos ou itens.
 - **Censuras do tipo II:** resultam de estudos os quais são finalizados após a ocorrência de um número pré-estabelecido de indivíduos ou itens.
 - **Censuras do tipo aleatório:** o período de acompanhamento é fixado e os participantes entram no estudo em diferentes instantes de tempo. Alguns podem falhar antes do seu término e os tempos de sobrevivência exatos são conhecidos. Este mecanismo de censura é comum em estudos clínicos.
- b) **Censura à esquerda:** acontece quando o tempo de falha é inferior ao tempo registrado, ou seja, não conhecemos o momento da ocorrência do evento mas sabemos que ele ocorreu em algum instante de tempo anterior ao registrado. Este é o tipo de censura menos comum. Ele ocorre, por exemplo, quando temos apenas a informação da data em que uma determinada infecção foi diagnosticada, isto é, não sabemos ao certo quando o indivíduo foi infectado, apenas sabemos que a infecção ocorreu muito antes da data de diagnóstico.
- c) **Censura intervalar:** neste tipo de censura, sabemos apenas que o evento de interesse ocorreu em algum instante dentro de um intervalo de tempo. Um exemplo para esse tipo de censura seria considerar um paciente que consulta regularmente seu médico e em uma dessas consultas, ele descobre uma determinada infecção. Nesse caso, a única informação que temos é que a infecção ocorreu em um intervalo entre duas consultas consecutivas, mas não sabemos com exatidão quando isso ocorreu.

2.2 Função de sobrevivência e função taxa de falha

Para formalizar as ideias, consideremos T como sendo uma variável aleatória não negativa representando o tempo até a ocorrência do evento de interesse, e cuja função de distribuição e densidade são dadas por $F(t)$ e $f(t)$, respectivamente. Denotaremos por C uma

variável aleatória, independente de T , representando o tempo de censura, com função de distribuição $G(c)$ e densidade $g(c)$. Considere que $F(t)$ e $G(c)$ são absolutamente contínuas. Para um determinado indivíduo e assumindo um mecanismo de censura à direita, a função de sobrevivência $S(t)$ é definida pela probabilidade do indivíduo sobreviver até o tempo t , isto é:

$$S(t) = P(T > t) = 1 - F(t).$$

Aqui, $S(t)$ é uma função decrescente em t com as seguintes propriedades: $S(0) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

A taxa de falha no intervalo $[t, t + \Delta t)$ é definida como a probabilidade de que a falha ocorra nesse intervalo, dado que não ocorreu antes de t dividida pelo comprimento do intervalo.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

Assim, $\lambda(t)\Delta t$ será a probabilidade aproximada da falha ocorrer no intervalo $[t, t + \Delta t)$ dado que ela não ocorreu antes de t . A função de sobrevivência pode ser reescrita com dependência a função taxa de falha acumulada, $\Lambda(t)$ como:

$$S(t) = \exp \left\{ - \int_0^t \lambda(x) dx \right\} = \exp \{ -\Lambda(t) \} \quad \text{onde} \quad \Lambda(t) = \int_0^t \lambda(x) dx.$$

2.3 Modelo de regressão de Cox

Para modelar o tempo de sobrevivência de uma população homogênea, em geral, são utilizadas distribuições contínuas univariadas. Porém, a existência de heterogeneidade entre os indivíduos é influenciada por alguns fatores de risco. Para que possamos incorporar esses fatores, que supomos afetarem o tempo de vida, recorreremos a um modelo de regressão, onde o tempo de sobrevivência é a variável resposta que será especificada por um conjunto de variáveis independentes.

O modelo de regressão proposto por Cox (1972), é bastante popular para estudar a influência desses fatores de risco no tempo de sobrevivência. Para definir o modelo

assumimos que temos uma amostra de n indivíduos em que o i -ésimo apresenta a tripla $D = (t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \dots, n$, sendo t_i o tempo observado, $\mathbf{x}_i = (x_1, \dots, x_n)'$ o vetor das covariáveis ou fatores de risco e δ_i é a função indicadora de falha, que pode ser igual a 1 se ocorrer a falha ou 0 quando ocorre a censura. O modelo de regressão de Cox fica dado por:

$$\lambda(t_i|\mathbf{x}_i) = \lambda_0(t_i)g(\mathbf{x}_i'\boldsymbol{\beta}), \quad (2.1)$$

em que $\lambda_0(t_i)$ é a função taxa de falha base cuja forma não é especificada, $\boldsymbol{\beta}=(\beta_1, \dots, \beta_p)'$ é o vetor de coeficientes e $g(\mathbf{x}_i'\boldsymbol{\beta})$ é uma função positiva. Este modelo é conhecido como modelo semi-paramétrico porque o componente paramétrico é assumido apenas no efeito das covariáveis em $g(\mathbf{x}_i'\boldsymbol{\beta})$.

$$g(\mathbf{x}_i'\boldsymbol{\beta}) = \exp \{ \mathbf{x}_i'\boldsymbol{\beta} \}.$$

Sua função de sobrevivência é dada por:

$$S(t_i|\mathbf{x}_i) = \exp \left\{ -\Lambda_0(t_i) \exp \{ \mathbf{x}_i'\boldsymbol{\beta} \} \right\},$$

sendo $\Lambda_0(t_i)$ a função taxa de falha base acumulada.

A suposição básica para o uso do modelo de regressão de Cox é a proporcionalidade das taxas de falhas, ou seja, a razão de taxas de falha para dois indivíduos i e l deve ser constante no tempo:

$$\frac{\lambda(t|\mathbf{x}_i)}{\lambda(t|\mathbf{x}_l)} = \frac{\lambda_0(t) \exp \{ \mathbf{x}_i'\boldsymbol{\beta} \}}{\lambda_0(t) \exp \{ \mathbf{x}_l'\boldsymbol{\beta} \}} = \exp \{ \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_l'\boldsymbol{\beta} \}.$$

Essa razão de taxas de falha não depende de tempo. Como pode ser observado, as covariáveis afetam a razão das funções taxas de falha de modo multiplicativo, de acordo com o fator $\exp \{ \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_l'\boldsymbol{\beta} \}$.

2.3.1 Função de Verossimilhança

Como apresentamos na Seção 2.2, $G(c_i)$, $F(t_i)$ e $g(c_i)$, $f(t_i)$ são as funções de distribuição e densidade para os tempos de censura e sobrevivência, respectivamente. Assumindo

independência entre T_i e C_i para o i -ésimo indivíduo, temos:

- se for observada uma censura:

$$P(T_i = t_i, \delta_i = 0) = P(C_i = t_i, T_i > C_i) = P(C_i = t_i, T_i > t_i) = g(t_i)S(t_i; \boldsymbol{\beta});$$

- se for observada uma falha:

$$P(T_i = t_i, \delta_i = 1) = P(T_i = t_i, T_i \leq C_i) = P(T_i = t_i, C_i \geq t_i) = f(t_i; \boldsymbol{\beta})G(t_i).$$

Fazendo uma combinação dos dois resultados anteriores, temos:

$$L(\boldsymbol{\beta}|D) = \prod_{i=1}^r f(t_i; \boldsymbol{\beta})G(t_i) \prod_{i=r+1}^n g(t_i)S(t_i; \boldsymbol{\beta}), \quad D = \{(t_i, \delta_i) : i = 1, 2, \dots, n\},$$

onde r representa o número de observações não-censuradas e $n - r$ corresponde o número de observações censuradas.

Pela suposição de que o mecanismo de censura é não-informativo, isto é, a função de distribuição $G(\cdot)$ não carrega nenhuma informação sobre o parâmetro de interesse, os termos $G(t_i)$ e $g(t_i)$ podem ser desprezados na função de verossimilhança, o que resulta em:

$$L(\boldsymbol{\beta}|D) \propto \prod_{i=1}^n [f(t_i|\mathbf{x}_i, \boldsymbol{\beta})]^{\delta_i} [S(t_i|\mathbf{x}_i, \boldsymbol{\beta})]^{1-\delta_i} = \prod_{i=1}^n [\lambda(t_i|\mathbf{x}_i, \boldsymbol{\beta})^{\delta_i} S(t_i|\mathbf{x}_i, \boldsymbol{\beta})].$$

Para o modelo de regressão de Cox apresentado na Eq. (2.1), a função de verossimilhança completa é dada por:

$$L(\boldsymbol{\beta}|D) = \prod_{i=1}^n \left\{ \lambda_0(t_i) \exp \left\{ \mathbf{x}'_i \boldsymbol{\beta} \right\} \right\}^{\delta_i} \exp \left\{ -\Lambda_0(t_i) \exp \left\{ \mathbf{x}'_i \boldsymbol{\beta} \right\} \right\}. \quad (2.2)$$

Conforme Colosimo e Giolo (2006) a presença do componente não paramétrico $\lambda_0(t_i)$ na função de verossimilhança completa torna este método de estimação inapropriado. Entretanto, uma solução razoável é usar a função de verossimilhança parcial, que consiste em condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras para eliminar esta perturbação da verossimilhança.

2.3.2 Função de verossimilhança parcial

Para a construção da função de verossimilhança parcial, assume-se a existência de uma amostra de n indivíduos, entre os quais existem r tempos de falhas distintos e $n - r$ tempos de censuras. Assume-se também que apenas uma falha ocorreu em cada t_i , para r tempos de falha ordenados, $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, tal que $t_{(i)}$ é o i -ésimo tempo de falha ordenado. Denote por $R(t_i)$ o conjunto de todos os indivíduos que ainda não falharam em t_i , também conhecido como o conjunto de todos os indivíduos sob risco. Assim, a função de verossimilhança parcial será dada por:

$$L(\boldsymbol{\beta}|D) = \prod_{i=1}^n \left(\frac{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \boldsymbol{\beta}\}} \right)^{\delta_i}, \quad (2.3)$$

o denominador da função de verossimilhança parcial é a soma dos valores de $\exp\{\mathbf{x}_j \boldsymbol{\beta}\}$ para todos os indivíduos sob o risco de falhar em t_i .

A função de verossimilhança parcial que apresentamos nessa dissertação assume que os tempos de sobrevivência são contínuos e conseqüentemente, não pressupõe a possibilidade de empates nos valores observados. Maiores detalhes sobre a função de verossimilhança com empate nos tempos de sobrevivência podem ser encontrados em Colosimo e Giolo (2006) ou Klein e Moeschberger (2003).

O logaritmo da função de verossimilhança dada na Eq. 2.3 é:

$$l(\boldsymbol{\beta}|D) = \log L(\boldsymbol{\beta}|D) = \sum_{i=1}^n \delta_i \left\{ \mathbf{x}'_i \boldsymbol{\beta} - \log \left[\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \boldsymbol{\beta}\} \right] \right\}. \quad (2.4)$$

Os parâmetros podem ser estimados no caso clássico através do algoritmo de Newton-Rapson, ver Apêndice A.

2.4 Distribuição Weibull

A distribuição Weibull é uma das mais utilizadas em análise de sobrevivência e confiabilidade. Sua função densidade de probabilidade é dada pela expressão a seguir:

$$f(t) = \alpha \varphi t^{\alpha-1} \exp \{-\varphi t^\alpha\}, \quad t > 0, \quad (2.5)$$

em que $\alpha > 0$ e $\varphi > 0$ são os parâmetros de forma e escala, respectivamente. Para o i -ésimo indivíduo, a função de sobrevivência e taxa de falha são dadas por:

$$S(t_i) = \exp \{-\varphi t_i^\alpha\} \text{ e } \lambda(t_i) = \alpha \varphi t_i^{\alpha-1}.$$

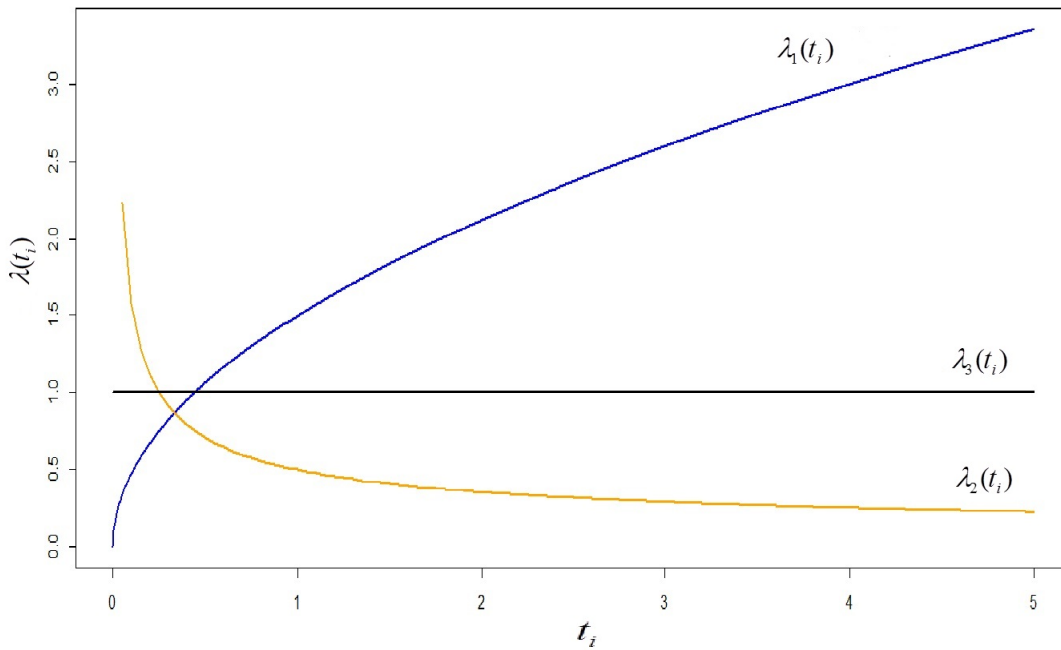


Figura 2.1: Funções taxas de falha para diferentes valores de α , crescente quando $\alpha > 1$, representada por $\lambda_1(t_i)$, constante quando $\alpha = 1$, representada por $\lambda_3(t_i)$, e decrescente quando $\alpha < 1$, representada por $\lambda_2(t_i)$.

A função taxa de falha $\lambda_1(t_i)$, que apresentamos na Figura 2.1, indica que o risco de falha aumenta em função do tempo; $\lambda_2(t_i)$ indica que o risco de falha decresce em função do tempo. Por fim, o função taxa de falha $\lambda_3(t_i)$, indica que o risco de falha é constante no tempo (distribuição exponencial).

2.4.1 Modelo de regressão Weibull

Um problema que frequentemente nos deparamos ao modelar dados de sobrevivência usando o modelo de riscos proporcionais é a especificação de uma forma paramétrica para a função taxa de falha basal $\lambda_0(t_i)$. Uma forma de contornar o problema de especificidade da função taxa de falha base é assumir que $T_i \sim \text{Weibull}(\alpha, \varphi)$, e que o vetor de covariáveis associado ao i -ésimo indivíduo é nulo da Eq. (2.1) a função taxa de falha base fica dada por:

$$\lambda(t_i | \mathbf{x}_i = \mathbf{0}) = \lambda_0(t_i) = \varphi \alpha t_i^{\alpha-1},$$

consequentemente,

$$\lambda(t_i | \mathbf{x}_i) = \varphi \alpha t_i^{\alpha-1} \exp \{ \mathbf{x}'_i \boldsymbol{\beta} \}. \quad (2.6)$$

Aqui, observa-se que o tempo de sobrevivência para o i -ésimo indivíduo no estudo tem a distribuição Weibull com parâmetro de escala $\varphi'_i = \varphi \exp \{ \mathbf{x}'_i \boldsymbol{\beta} \}$ e parâmetro de forma α . Essa estrutura mantém a proporcionalidade de riscos do modelo de regressão de Cox e mostra que o efeito das variáveis explicativas no modelo alteram a distribuição do parâmetro de escala, enquanto o parâmetro de forma permanece constante, ver Collett (2003).

A função de sobrevivência será:

$$S(t_i | \mathbf{x}_i) = \exp \left\{ -\varphi t_i^\alpha \exp \{ \mathbf{x}'_i \boldsymbol{\beta} \} \right\}. \quad (2.7)$$

Partindo da Eq. (2.6) e Eq. (2.7), a função de verossimilhança fica dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\varphi \alpha t_i^{\alpha-1} \exp \{ \mathbf{x}'_i \boldsymbol{\beta} \} \right)^{\delta_i} \exp \left\{ -\varphi \exp \{ \mathbf{x}'_i \boldsymbol{\beta} \} \right\} \quad (2.8)$$

A seguir apresentamos alguns métodos de construção de intervalos de confiança para os coeficientes da regressão. Neste trabalho, iremos focar nos método de Wald e da razão de verossimilhança para determinar intervalos com nível $100(1 - \gamma)\%$ de confiança. Considere as seguintes hipóteses:

$$H_0 : \beta_k = \beta_k^0 \quad \text{contra} \quad H_1 : \beta_k \neq \beta_k^0;$$

em que β_k^0 é uma constante qualquer, que em geral assume-se como sendo igual a zero e β_k o k -ésimo elemento de β . Assumindo que os EsMV satisfazem algumas propriedades matemáticas básicas, para grandes amostras, sob certas condições de regularidade estabelece-se que o EMV ($\hat{\beta}_k$) segue assintoticamente uma distribuição normal com média β_k e variância $Var(\hat{\beta}_k)$. Em que $Var(\hat{\beta}_k)$ o k -ésimo elemento da diagonal de $[i(\beta)]^{-1} \equiv -[E(I(\beta))]^{-1}$, $I(\beta)$ é a matriz de informação observada. Colosimo e Giolo (2006) acrescentam que em situações nas quais é impossível ou difícil de calcular a esperança matemática, usa-se simplesmente $[-I(\beta)]^{-1}$. Desta forma, o intervalo de confiança baseado no método de Wald é dado por:

$$\left(\hat{\beta}_k - z_{(\frac{\gamma}{2})} ep(\hat{\beta}_k); \hat{\beta}_k + z_{(1-\frac{\gamma}{2})} ep(\hat{\beta}_k) \right),$$

onde $z_{(\frac{\gamma}{2})}$ é o quantil da distribuição normal padrão; $ep(\hat{\beta}_k) = [Var(\hat{\beta}_k)]^{\frac{1}{2}}$.

O intervalo de confiança baseado no método de verossimilhança perfilada será:

$$\left\{ \hat{\beta}_k : 2 \left(l(\hat{\beta}_k) - l(\beta^0) \right) < \chi_{p,1-\gamma}^2 \quad \text{ou} \quad l(\beta^0) \geq l(\hat{\beta}_k) - \frac{1}{2} \chi_{p,1-\gamma}^2 \right\},$$

sendo $l(\cdot)$ o logaritmo da função de verossimilhança dada na Eq. (2.4), $\hat{\beta}_k$ é o EMV de β_k , e $\chi_{p,1-\gamma}^2$ o quantil da distribuição χ^2 com p graus de liberdade.

2.4.2 Resumo do capítulo

Neste capítulo, foram apresentado alguns conceitos básicos em análise de sobrevivência com principal enfoque nos modelos de regressão de Cox e Weibull. Apresentamos ainda as funções de sobrevivência e taxas de falhas para os dois modelos, bem como as respectivas funções de verossimilhanças. No próximo capítulo, descrevemos de forma resumida a ocorrência da VM e apresentamos algumas propostas de correção do viés dos EsMV.

Capítulo 3

Verossimilhança monótona

A VM é um problema de não existência de EsMV sob certas condições na amostra. Fatores como o tamanho de amostra, a porcentagem de censura e a presença de covariáveis binárias nas análises aumentam as chances de ocorrência deste fenômeno. A seguir apresentamos de forma detalhada a ocorrência da VM no banco de dados de melanoma mencionado no Capítulo 1, em que a covariável binária “mitose” e o evento em estudo “metástase” são confrontadas.

| | | Censura | Falha |
|--------|-----|---------|-------|
| Mitose | Não | 87 | 0 |
| | Sim | 157 | 48 |

Tabela 3.1: *Cruzamento entre a variável mitose e o evento em estudo (metástase).*

A Tabela 3.1 é baseada na mesma informação da Figura 1.1 do Capítulo 1. Nela observamos que um dos níveis da covariável “mitose” não está associado a ocorrência da “metástase”, ou seja, pacientes sem “mitose” não desenvolveram o evento de interesse, sendo esse um indicador da presença da VM. Isto porque o risco de um pacientes sem “mitose” falhar é igual a zero. Entretanto, a razão de riscos nos pacientes com e sem “mitose” $RR_{(sim)}/RR_{(não)}$ não vai ter um valor finito e como consequência, a estimativa associada a esse parâmetro no modelo de regressão de Cox não vai convergir. Maiores detalhes sobre o cálculo do risco relativo a partir de uma tabela de contingência podem

ser encontrados em Therneau e Grambsch (2000).

Quando a VM é detectada, Brayson e Johnson (1981) e Brayson et al. (1982) apresentam algumas propostas para contornar o problema, como por exemplo:

- (a) Testar diferentes tipos de modelos;
- (b) Omitir o fator que causa o problema;
- (c) Fazer uma análise estratificada no fator;
- (d) Continuar com o modelo de Cox na presença da VM.

Em geral, a desvantagem associada as opções (b), (c) e (d) residem no fato de não conseguirmos avaliar o real efeito da covariável no modelo. A opção (a) não é apropriada, pois testar diferentes modelos implica em obter estimativas de parâmetros com diferentes interpretações. Portanto, a proposta mais notável, que é baseada na abordagem Clássica de inferência, foi apresentada por Firth (1993) e objetiva reduzir o vício dos EsMV pela subtração do viés de primeira ordem na função escore. A seguir apresentamos algumas propostas ou métodos de redução do viés dos EsMV.

3.1 Métodos de redução do viés dos EsMV

O viés de um estimador é definido como a diferença entre o seu valor esperado e o parâmetro a ser estimado. Kosmidis (2014) apresenta duas classificações dos métodos de redução do viés: os explícitos e os implícitos.

3.1.1 Métodos explícitos

Os métodos explícitos são baseados em uma etapa, onde o viés $b(\beta)$ é estimado e então subtraído de $\hat{\beta}$, resultando em um novo estimador $\hat{\beta}^*$. Essa classe é composta pelos métodos: *Jackknife*, *Bootstrap* e o método de correção assintótica que apresentamos a seguir.

Seja β o vetor de coeficientes que apresentamos na Seção 2.3. Cox e Hinkley (1974) mostraram que o viés assintótico para $\hat{\beta}_k$ pode ser dado pela seguinte expressão:

$$b(\beta_k) = E(\hat{\beta}_k - \beta_k) = \frac{b_1(\beta_k)}{n} + O(n^{-1}), \quad (3.1)$$

sendo n o número total de observações na amostra e $b_1(\beta_k)/n$ o viés da primeira ordem obtido pela expansão da série de Taylor da Eq. (3.1).

Substituindo β_k por $\hat{\beta}_k$ na equação anterior, o EMV corrigido é dado por:

$$\hat{\beta}_k^{(BC)} = \hat{\beta}_k - \frac{b_1(\hat{\beta}_k)}{n},$$

que remove o primeiro termo do viés assintótico de $\hat{\beta}_k$.

A maior vantagem dos métodos explícitos é a simplicidade na sua aplicação. Entretanto, uma vez que o viés é subtraído de $\hat{\beta}_k$, esses métodos são mais corretivos que preventivos no sentido de que o EMV é primeiramente calculado e depois corrigido. Em amostras finitas com VM, essa abordagem acaba não sendo viável, pois requer a existência de $\hat{\beta}_k$.

Foi nesse sentido que Firth (1993) propôs uma abordagem “preventiva” que pertence a classe dos métodos implícitos em que o viés é reduzido pela modificação da função escore. O autor mostrou que este método remove o viés de ordem $O(n^{-1})$ dos EsmV.

3.1.2 Métodos implícitos

Segundo Kosmidis (2014), esses métodos são subdivididos em inferência indireta e o método de correção proposto por Firth (1993), que apresentamos a seguir.

Para apresentar a proposta de correção do viés baseada na abordagem preventiva, vamos considerar o método de estimação por máxima verossimilhança em que os estimadores são obtidos resolvendo a equação da função de estimação.

$$\frac{\partial l(\boldsymbol{\beta}|D)}{\partial \beta_k} = U(\boldsymbol{\beta}) = 0.$$

Para motivar um desenvolvimento mais geral, vamos considerar inicialmente um modelo da família exponencial com um único parâmetro β_k em que o logaritmo da função de verossimilhança é proporcional a $l(\beta_k|D) = y\beta_k - Z(\beta_k)$. A função escore fica dada por:

$$U(\beta_k) = y - Z'(\beta_k), \quad (3.2)$$

tal que a estatística suficiente y afeta apenas na locação de $U(\beta_k)$ e não a forma. Firth (1993) e Gamerman et al. (2014) mostram como o viés em $\hat{\beta}_k$ surge pela avaliação da curvatura da função escore.

Para uma formulação geral, consideremos novamente $I(\boldsymbol{\beta})$ a matriz de informação observada para os dados que apresentamos no Capítulo 2. A função escore modificada $U^*(\boldsymbol{\beta})$ é dada por:

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - I(\boldsymbol{\beta})b(\boldsymbol{\beta}). \quad (3.3)$$

Assim, o vetor dos EsMV corrigidos $\hat{\boldsymbol{\beta}}^*$ será a solução da equação $U^*(\boldsymbol{\beta}) = 0$.

Para formalizar o argumento acima e calcular a mudança necessária na função escore, o autor apresenta algumas notações para o cálculo de cumulantes cujo objetivo é especificar as derivadas dos logaritmos da função de verossimilhança; informações adicionais são encontradas em Cordeiro e Cribari-Neto (2014).

Para modelos da família exponencial com parametrização canônica, onde a matriz de informação observada obtida pela Eq. (3.2) não dependerá dos dados, isto é, coincide com a matriz de informação esperada. Firth (1993) mostrou que o termo de penalização na Eq. (3.3) é dado por:

$$a_k(\boldsymbol{\beta}) = \frac{1}{2} \text{tr} \left\{ i^{-1}(\boldsymbol{\beta}) \left(\frac{\partial i(\boldsymbol{\beta})}{\partial \beta_k} \right) \right\} = \frac{\partial}{\partial \beta_k} \left\{ \frac{1}{2} \log |i(\boldsymbol{\beta})| \right\},$$

sendo $i(\boldsymbol{\beta})$ a matriz de informação esperada. Assim, a função escore modificada é dada por:

$$U_k^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) + a_k(\boldsymbol{\beta}).$$

Entretanto, o logaritmo da função de verossimilhança penalizada será:

$$l^*(\boldsymbol{\beta}|D) = l(\boldsymbol{\beta}|D) + \frac{1}{2} \log |i(\boldsymbol{\beta})|.$$

Conseqüentemente, a função de verossimilhança penalizada é dada por:

$$L^*(\beta|D) = L(\beta|D) |i(\beta)|^{\frac{1}{2}}. \quad (3.4)$$

Como o modelo de regressão de Cox não pertence a família exponencial, e dada a dificuldade em obter a matriz de informação esperada, o termo $|i(\beta)|$ é substituído pela matriz de informação observada $|I(\beta)|$.

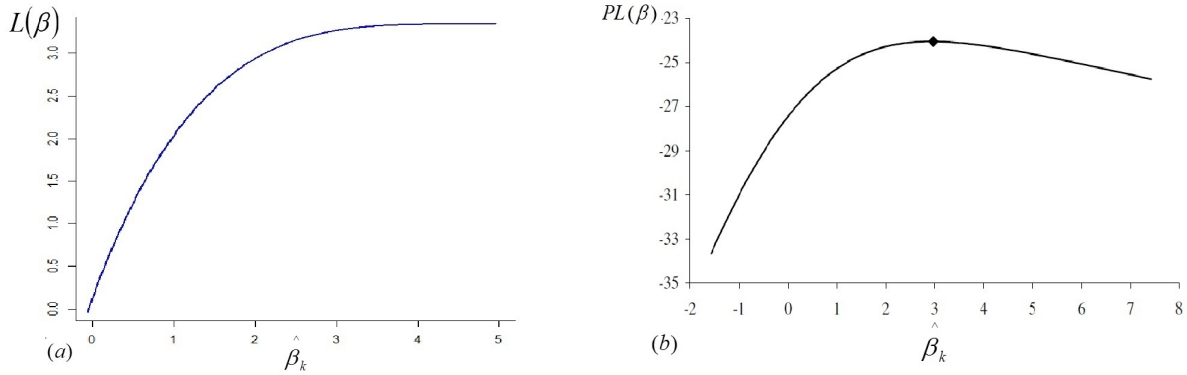


Figura 3.1: (a) Função de Verossimilhança usual $L(\beta)$ e, (b) Função de Verossimilhança Penalizada $PL(\beta)$.

A Figura 3.1 foi retirada no artigo de Heinze e Schemper (2001). Observe que a função de verossimilhança usual é monótona crescente. O erro-padrão e a estimativa de $\hat{\beta}_k$ tendem a infinito. Após subtrair uma pequena quantidade do viés na função escore, ela deixa de ser estritamente crescente e passa a ter um ponto de máximo em $\hat{\beta}_k$.

Segundo Kosmidis (2014) os métodos implícitos tem a vantagem de não dependerem diretamente do EMV e, portanto, são mais vantajosos em comparação com os métodos explícitos. Contudo, uma desvantagem associada a proposta de Firth é que sua aplicação requer a obtenção de soluções para as equações implícitas que apresentamos anteriormente.

Como mencionamos neste capítulo, a proposta de Firth (1993) é baseada na abordagem Clássica, em que o termo de penalização na Eq. (3.4) é a especificação a *priori* de Jeffreys para modelos que pertencem a família exponencial, neste caso, o viés de ordem $O(n^{-1})$ é removido calculando a moda a *posteriori*. Na abordagem que propomos, o termo de penalização $|i(\beta)|^{\frac{1}{2}}$ será substituído por uma distribuição a *priori* $p(\beta)$. Natu-

ralmente, as inferências serão baseadas nas amostras obtidas a partir da distribuição a *posteriori* de β .

No próximo capítulo apresentamos algumas especificações de distribuições a *priori*, como novas propostas de funções de penalidade.

3.2 Resumo do capítulo

Neste capítulo, abordamos o problema da VM, e descrevemos alguns métodos de correção do problema, com foco na proposta de Firth. No próximo capítulo, apresentamos alguns conceitos em inferência Bayesiana e indicamos algumas distribuições a *priori* que são considerada durante a análise dos dados.

Capítulo 4

Inferência Bayesiana

A informação que se tem sobre uma quantidade de interesse $\boldsymbol{\beta}=(\beta_1, \dots, \beta_p)'$ antes de observar os dados, pode ser importante para a análise estatística. O verdadeiro valor de $\boldsymbol{\beta}$ é desconhecido e desejamos expressar nossa incerteza sobre essa quantidade. Do ponto de vista Bayesiano os diferentes graus de incerteza são representados através de modelos probabilísticos para $\boldsymbol{\beta}$ Gelman et al. (2014).

A abordagem Bayesiana baseia-se no teorema de Bayes, o qual expressa a informação sobre o parâmetro desconhecido após observar os dados de interesse. Para formular a regra de Bayes, assumamos que D é o conjunto de dados a serem observados determinando a função de verossimilhança denotada por $L(\boldsymbol{\beta}|D)$; conforme a Eq. (2.3) ou Eq. (2.8). Assumamos que $p(\boldsymbol{\beta})$ é a distribuição a *priori* expressando nossa incerteza inicial sobre $\boldsymbol{\beta}$, antes de observar os dados. As análises serão baseadas na distribuição a *posteriori* que é na verdade uma atualização da informação a *priori*. A regra de Bayes determina a distribuição a *posteriori*:

$$p(\boldsymbol{\beta}|D) = \frac{L(\boldsymbol{\beta}|D)p(\boldsymbol{\beta})}{p(D)}, \quad (4.1)$$

em que $p(D)$ é a constante normalizadora também chamada de distribuição preditiva.

Novamente ressaltamos que toda inferência será feita com base na distribuição a *posteriori*, de onde obtêm-se as estimativas necessárias para resumir o comportamento de $\boldsymbol{\beta}$. Dentre as principais medidas a *posteriori* podemos citar:

- média a *posteriori*: $E(\beta_k|D) = \int_{\beta_k \in \mathcal{B}} \beta_k p(\beta_k|D) d\beta_k$;
- o quantil γ a *posteriori*: $Q(\gamma) = \left\{ \beta'_k \in \mathcal{B} : \int_{-\infty}^{\beta'_k} p(\beta_k|D) d\beta_k = \gamma \right\}, \gamma \in (0, 1)$;
para $\gamma = 0.5$ temos aqui a mediana a *posteriori*.
- O intervalo de $100(1 - \gamma)\%$ de credibilidade a *posteriori*:

$$(L, U) = \left\{ (L', U') \subset \mathbb{R}^2 : \int_{L'}^{U'} p(\beta_k|D) d\beta_k = 1 - \gamma \right\}, \gamma \in (0, 1).$$

Caso o intervalo seja simétrico, os limites para a mais alta densidade a *posteriori* vão ser dados por, $L=Q(\gamma/2)$ e $U=Q(1 - \gamma/2)$.

A quantidade $p(\beta_k|D)$ representa a distribuição marginal de β_k e β'_k uma quantidade qualquer que β_k pode assumir.

Uma justificativa para o uso da abordagem Bayesiana no modelo de regressão de Cox baseada na verossimilhança parcial é apresentada por Sinha et al. (2003) e Lin et al. (2013). Os autores propõem diferentes estruturas de distribuições a *priori* para os parâmetros. Outra referência importante que usa a abordagem Bayesiana no modelo de regressão de Cox é Ibrahim et al. (2001). A seguir descrevemos as distribuições a *priori* que iremos utilizar nos próximos Capítulos.

4.1 Distribuições a *priori* a serem estudadas

A distribuição a *priori* $p(\beta)$ pode ser vista como uma representação do conhecimento que se tem sobre o parâmetro de interesse antes de observar os dados. Ela constitui uma importante quantidade em inferência Bayesiana.

4.1.1 *Priori* de Jeffreys

O método de Jeffreys (1946) surge como uma crítica da inconsistência da distribuição uniforme na reparametrização formal que leva em conta a propriedade de invariância. Entre os procedimentos que asseguram essa invariância está aquela advogada por Jeffreys e que se baseia no uso da medida de informação de Fisher sobre $\beta \in \mathbb{R}^p$, ou seja,

$$i_{ks}(\boldsymbol{\beta}) = E_D \left[\left(\frac{\partial \log f(D|\boldsymbol{\beta})}{\partial \beta_k \beta_s} \right)^2 \middle| D \right].$$

Considerando uma reparametrização não linear e monótona em β_k tal que $\eta = \phi(\beta_k)$ temos:

$$i(\eta) = i(\beta_k(\phi)) \left(\frac{d\beta_k}{d\phi} \right)^2.$$

Isso mostra que a distribuição a *priori* proposta por Jeffreys para o caso uniparamétrico, dada por:

$$p(\beta_k) = [i(\beta_k)]^{\frac{1}{2}}.$$

possui a referida propriedade de invariância.

4.1.2 Outras distribuições a *priori*

Conforme já falamos anteriormente, propomos corrigir o problema da VM assumindo como distribuição a *priori* $N(m_k, v_k)$; $\text{Cauchy}(l_k, s_k)$ sendo m_k e l_k parâmetros de locação; v_k e s_k parâmetros de escala, respetivamente. E a distribuição a *priori* $\log-F(a_k, b_k)$, sendo a_k e b_k os parâmetros da distribuição.

Naturalmente, outras distribuições a *priori* poderiam ser avaliadas neste estudo, no entanto, escolhemos seguir os passos de Greenland e Mansournia (2015) e Lin et al. (2013) que investigaram a ocorrência da VM tanto no modelo de regressão logística como em modelos envolvendo dados de sobrevivência.

A distribuição a *priori* $N(m_k, v_k)$ é simétrica e unimodal. Temos que a média, moda e mediana são iguais. A quantidade de informação presente nessa distribuição é controlada pela sua variância v_k . Valores pequenos de v_k significam que a massa probabilística está mais concentrada em torno de m_k e, portanto, ela carrega mais informação sobre o parâmetro. A distribuição $\text{Cauchy}(l_k, s_k)$ é simétrica em torno de l_k , sua média e variância não são definidas, e sua moda e a mediana são iguais a l_k .

Conforme Ojo e Olapade (2003) a função densidade de probabilidade para a distribuição $\log-F(a_k, b_k)$ é dada por:

$$p(\beta_k | 2a_k, 2b_k) = \frac{a_k^{a_k} b_k^{b_k}}{B(a_k, b_k)} \frac{\exp\{a_k \beta_k\}}{(b_k + a_k \exp\{\beta_k\})^{(a_k + b_k)}},$$

em que $B(a_k, b_k) = \Gamma(a_k)\Gamma(b_k)/\Gamma(a_k + b_k)$ é a função beta, $a_k = \nu_{1k}/2$ e $b_k = \nu_{2k}/2$, sendo ν_{1k} e ν_{2k} graus de liberdade da distribuição F ; a_k e b_k são os graus de liberdade da distribuição log- F .

A distribuição log- $F(a_k, b_k)$ é unimodal podendo ser simétrica se $a_k = b_k = gl$, que se aproxima da distribuição normal centrada em 0 cuja variância diminui à medida em que se aumenta gl . A média, moda e a mediana a *priori* são iguais. Se gl for suficientemente pequeno, as caudas da distribuição a *priori* se tornam mais pesadas; para $a_k < b_k$, a distribuição a *priori* log- F é assimétrica positiva, e assimétrica negativa quando $a_k > b_k$. A média e a variância da distribuição a *priori* log- $F(a_k, b_k)$ são dadas por:

$$E(\beta_k) = \psi(a_k) - \psi(b_k) \quad \text{e} \quad \text{Var}(\beta_k) = \psi'(a_k) + \psi'(b_k),$$

sendo $\psi(a_k) = \frac{\partial}{\partial a_k} \log \Gamma(a_k)$ e $\psi'(a_k) = \frac{\partial^2}{\partial a_k^2} \log \Gamma(a_k)$ as funções digama e trigama, respectivamente; maiores detalhes são encontrados em Amos (1983). Mantendo fixo o valor do parâmetro b_k , a média aumenta a medida em que aumentamos o parâmetro a_k . Fixando o valor de a_k , a média a *priori* decresce a medida em que aumentamos o parâmetro b_k . Nos dois casos a variância decresce para $\pi^2/6$ quando $a_k \rightarrow \infty$ ou $b_k \rightarrow \infty$, respectivamente.

Em muitas situações a distribuição a *posteriori* não tem uma forma fechada, pois o denominador da Eq. (4.1) não possui uma forma analítica. Portanto, a distribuição a *posteriori* tem que ser obtida através de métodos numéricos.

Na seção a seguir apresentamos uma breve introdução aos métodos de simulação de Monte Carlo via Cadeias de Markov (MCMC); maiores detalhes são encontrados em Gelman et al. (2014) e Gamerman e Lopes (2006).

4.2 Amostrador de Gibbs

O *Gibbs sampling* foi a primeira classe de esquemas largamente empregados para simulação estocástica via cadeias de Markov. Ele foi originalmente proposto por Geman e Geman (1984) no contexto de reconstrução de imagens. Os autores propuseram um

esquema de amostragem da distribuição de *Gibbs* explorando juntamente suas condicionais completas, por meio de um algoritmo iterativo que define uma cadeia de Markov. Mais tarde, Gelfand e Smith (1990) compararam esse amostrador com outros esquemas de simulação estocástica e mostraram que o *Gibbs Sampling* poderia ser utilizado em muitos casos para amostrar de distribuições a *posteriori*. Gamerman e Lopes (2006) acrescentam que no *Gibbs sampling* os elementos do vetor de parâmetros $\boldsymbol{\beta}$ podem ser atualizados um de cada vez, ou em blocos, tomando a distribuição condicional completa como proposta com probabilidade de aceitação igual a 1. As transições de um estado para outro são feitas de acordo com as distribuições condicionais completas $p(\beta_k | \boldsymbol{\beta}_{-k}, D)$, sendo $\boldsymbol{\beta}_{-k} = (\beta_1, \dots, \beta_p)'$.

Suponha que a distribuição de interesse é $\pi(\boldsymbol{\beta}) \equiv p(\beta_k | \boldsymbol{\beta}_{-k}, D)$, cada um dos componentes β_k pode ser um escalar, vetor ou matriz. Considere também que as distribuições condicionais completas $\pi(\beta_k) = p(\beta_k | \boldsymbol{\beta}_{-k}, D)$ estão disponíveis, isto é, assume-se que elas são completamente conhecidas e podem ser amostradas. A distribuição condicional completa é uma distribuição do k -ésimo componente de $\boldsymbol{\beta}$ condicionada em todos outros componentes e nos dados observados.

Em muitas situações, a geração de uma amostra diretamente de $\pi(\beta_k)$ pode ser custosa, complicada ou simplesmente impossível. Se as distribuições condicionais completas forem conhecidas, o amostrador de *Gibbs*, segundo Gamerman e Lopes (2006), é definido pelo seguinte esquema:

Passo 1: Inicialize o contador de interações da cadeia $j = 0$;

Passo 2: Especifique valores iniciais $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)}, \dots, \beta_p^{(0)})$;

Passo 3: Obtenha um novo valor de $\beta_k^{(j)}$ através de sucessivas gerações:

gerar $\beta_1^{(j)}$ de $p(\beta_1 | \beta_2^{(j-1)}, \beta_3^{(j-1)}, \dots, \beta_p^{(j-1)})$;

gerar $\beta_2^{(j)}$ de $p(\beta_2 | \beta_1^{(j)}, \beta_3^{(j-1)}, \dots, \beta_p^{(j-1)})$;

⋮

gerar $\beta_p^{(j)}$ de $p(\beta_p | \beta_1^{(j)}, \beta_2^{(j)}, \dots, \beta_{p-1}^{(j)})$;

Passo 4: Incremente o contador de j para $j + 1$ e retorne ao passo 2 até obter convergência.

A medida em que j cresce, a cadeia aproxima-se de uma condição de equilíbrio. Quando a convergência for alcançada, os valores resultantes formam uma amostra de $\pi(\beta_k)$.

No caso em que a distribuição condicional completa a *posteriori* não tiver uma forma fechada, o algoritmo de Metropolis-Hastings (Metropolis et al. (1953); Hastings (1970)) é uma boa opção para gerar amostras da distribuição a *posteriori* como um passo dentro do *Gibbs Sampling*.

4.3 Algoritmo de Metropolis-Hastings

O algoritmo Metropolis-Hastings é segundo Gelman et al. (2014) e Gamerman e Lopes (2006) um método MCMC em que um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade. Este mecanismo de correção garante a convergência da cadeia para uma distribuição de equilíbrio que, neste caso, é a distribuição a *posteriori*. Em outras palavras, se desejamos gerar valores da distribuição a *posteriori* $\pi(\beta_k) \equiv p(\beta_k | \beta_{-k}, D)$, precisamos especificar uma distribuição geradora de candidatos $q(\beta_k^{\text{cand}} | \beta_k)$ válida no domínio de β_k e condicionada a $\beta_k^{(j-1)}$.

Suponha que a cadeia esteja no estado β_k e um valor β_k^{cand} é gerado de uma distribuição proposta $q(\cdot | \beta_k)$. Esta distribuição pode depender do estado atual da cadeia. Assim, o novo valor β_k^{cand} é aceito com probabilidade

$$\alpha(\beta_k^{(j-1)}, \beta_k^{\text{cand}}) = \min \left\{ 1, \frac{\pi(\beta_k^{\text{cand}})q(\beta_k^{(j-1)} | \beta_k^{\text{cand}})}{\pi(\beta_k^{(j-1)})q(\beta_k^{\text{cand}} | \beta_k^{(j-1)})} \right\}, \quad (4.2)$$

sendo $\pi(\beta_k^{(j-1)})$ a distribuição de interesse.

Uma característica importante é que só precisamos conhecer $\pi(\cdot)$ parcialmente, isto é, a menos de uma constante já que, neste caso, as constantes não alteram a probabilidade dada pela Eq. (4.2). Esse aspecto é fundamental em aplicações Bayesianas onde não conhecemos completamente a *posteriori*. Durante o processo de simulação, a cadeia pode

permanecer no mesmo estado por muitas interações e, na prática, costuma-se monitorar isto calculando a porcentagem média de interações para os quais novos valores são aceitos (Gamerman e Lopes, 2006). O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

Passo 1: Inicialize o contador de interações $j = 0$ e especifique um valor inicial $\beta_k^{(0)}$;

Passo 2: Gere um novo candidato β_k^{cand} da distribuição $q(\beta_k^{\text{cand}}|\beta_k^{(j-1)})$;

Passo 3: Calcule a probabilidade de aceitação $\alpha(\beta_k^{(j-1)}, \beta_k^{\text{cand}})$;

Passo 4: Gere U de uma Uniforme (0,1);

Passo 5: Se $U \leq \alpha(\beta_k^{(j-1)}, \beta_k^{\text{cand}})$, aceite o novo valor gerado e faça $\beta_k^{(j)} = \beta_k^{\text{cand}}$. Caso contrário, rejeite e faça $\beta_k^{(j)} = \beta_k^{(j-1)}$;

Passo 6: Incremente o contador de j para $j + 1$ e volte para passo 2;

Embora a distribuição proposta possa ser escolhida arbitrariamente, na prática deve-se tomar alguns cuidados para garantir a eficiência do algoritmo. Em aplicações Bayesianas a distribuição de interesse é $p(\beta_k|\beta_{-k}, D)$ e a probabilidade de aceitação assume uma forma particular,

$$\alpha(\beta_k^{(j-1)}, \beta_k^{\text{cand}}) = \min \left\{ 1, \frac{p(D|\beta_k^{\text{cand}}) p(\beta_k^{\text{cand}}) q(\beta_k^{(j-1)}|\beta_k^{\text{cand}})}{p(D|\beta_k^{(j-1)}) p(\beta_k^{(j-1)}) q(\beta_k^{\text{cand}}|\beta_k^{(j-1)})} \right\}.$$

Um caso particular do algoritmo de Metropolis-Hastings é chamado algoritmo de Metropolis, obtido quando a distribuição proposta é simétrica em seus argumentos, isto é, $q(\beta_k^{\text{cand}}|\beta_k^{(j-1)}) = q(\beta_k^{(j-1)}|\beta_k^{\text{cand}})$ para todos os valores de β_k e β_k^{cand} . Neste caso, a probabilidade de aceitação se reduz para:

$$\alpha(\beta_k^{(j-1)}, \beta_k^{\text{cand}}) = \min \left\{ 1, \frac{p(\beta_k^{\text{cand}}|\beta_{-k}, D)}{p(\beta_k^{(j-1)}|\beta_{-k}, D)} \right\}.$$

4.3.1 Resumo do capítulo

Neste capítulo descrevemos alguns elementos importantes em inferência Bayesiana e apresentamos algumas distribuições a *priori* que são abordadas mais adiante nesta dissertação. Descrevemos ainda os algoritmos de simulação como o *Gibbs Sampling* e Metropolis-Hastings.

A seguir, um estudo simulado foi desenvolvido para avaliar as propriedades assintóticas dos EsMV e explorar o banco de dados de melanoma.

Capítulo 5

Estudo simulado

Um estudo simulado foi realizado com o objetivo de comparar o desempenho das diferentes funções de penalidade usando a abordagem Clássica e a Bayesiana. A simulação foi desenvolvida para explorar as propriedades assintóticas dos estimadores. Os modelos que apresentamos no Capítulo 2 são avaliados. A performance em termos de estimação é avaliada usando as seguintes medidas: erro quadrático médio (*EQM*), viés relativo (*VR*), erro-padrão (*EP*) e taxa de cobertura (*Cob*). A taxa de cobertura representa a porcentagem de vezes que o intervalo de confiança ou credibilidade contém o verdadeiro valor do parâmetro nas N repetições de Monte Carlo (MC) que iremos explorar. O objetivo do uso de repetições MC é evitar que nossa análise seja inteiramente baseada em uma única amostra simulada que, eventualmente, poderia não estar configurada com o problema da VM, ou até mesmo ser uma amostra com valores atípicos dificultando a estimação.

O viés relativo e o erro quadrático médio são dados por:

- $VR_{\hat{\beta}_k} = 100 * E(\hat{\beta}_k - \beta_k) / |\beta_k|$;
- $EQM_{\hat{\beta}_k} = Var(\hat{\beta}_k) + [E(\hat{\beta}_k - \beta_k)]^2$;

sendo β_k o verdadeiro valor e $\hat{\beta}_k$ um estimador, podendo ser a média a *posteriori* ou o EMV penalizado para β_k .

Os dados gerados são compostos por duas variáveis explicativas, sendo x_1 de natureza qualitativa binária, que foi configurada de modo a ter apenas 5 valores observados iguais a 1 e os restantes $n - 5$ iguais a 0, para diferentes tamanhos amostrais n . Essa configuração objetiva induzir a ocorrência da VM, isto é, objetiva criar situações em que um dos níveis da covariável binária não esteja associado ao evento de interesse. A variável x_2 é de natureza quantitativa contínua, e foi gerada de uma distribuição normal padrão.

A análise dos dados referente a abordagem Clássica foi baseada na proposta de correção de Firth (1993) e usamos o pacote “coxphf” que foi implementado no Software R (R Core Team, 2016) por Heinze e Ploner (2015). Dois cenários são considerados assumindo aproximadamente 75% e 25% de amostras com VM entre as repetições de MC.

O ajuste do modelo de regressão de Cox usando a abordagem Clássica foi baseado em 1000 amostras MC para diferentes valores de n . O cálculo das medidas de performance foi baseado no vetor das 1000 estimativas de um dado parâmetro. Para garantir a comparabilidade das estimativas de β_1 (parâmetro associado a covariável binária) e forçar a ocorrência da VM, fixamos $\beta_1 = -2$ e o parâmetro de forma da distribuição Weibull em $\alpha = 1.5$; ver Eq. (2.5). A inclusão do β_0 poderia atrapalhar as análise neste sentido, assumimos que $\beta_0 = 0$. Consequentemente, o parâmetro de escala da distribuição Weibull será ($\varphi = 1$).

A seguir apresentamos diferentes configurações para o parâmetro associado a covariável contínua x_2 , e o parâmetro associado a distribuição dos tempos de censura (ω_c) que fornecem aproximadamente 75% e 25% de amostras com VM. A porcentagem de censura fixamos em torno dos 70% e 75% em todos os cenários.

Assumindo tempos de falha $T_i \sim \text{Weibull}(\alpha, \varphi_i)$, sendo $\varphi_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$, e de censuras $C_i \sim \text{Exp}(\omega_c)$, foi possível obter a seguinte configuração:

Cenário 1: Aproximadamente 75% de amostras com VM, $\beta_2 = -0.74$ e $\omega_c = 2.1$;

Cenário 2: Aproximadamente 25% de amostras com VM, $\beta_2 = 4.97$ e $\omega_c = 4.7$.

Na abordagem Bayesiana apenas o Cenário 1 foi considerado para ajustar o modelo. No entanto, foram geradas 1000 cadeias para cada parâmetro e as inferências foram

baseadas no vetor das médias a *posteriori*. Aqui, avaliamos o impacto de diferentes funções de penalidade nas estimativas a *posteriori*. Como as distribuições a *posteriori* conjunta não possuem formas fechada as amostras foram geradas do algoritmo *Gibbs Sampling* com passos de Metropolis-Hastings assumindo 7000 iterações com $lag=1$. As primeiras 1000 formam o período de aquecimento (*burn-in*). No entanto, a convergência foi avaliada a partir de uma única amostra gerada. Como sugerido na literatura, a taxa de aceitação foi fixada entre 40% a 50%; ver Roberts et al. (1997).

A seguir apresentamos algumas configurações das distribuições a *priori* para o vetor (β, α) assumindo diferentes hiperparâmetros.

- **Caso 1:** $\beta_k \sim N(0, 1)$ e $\alpha \sim \text{Gama}(1, 1)$;
- **Caso 2:** $\beta_k \sim N(0, 5)$ e $\alpha \sim \text{Gama}(0.2, 0.2)$;
- **Caso 3:** $\beta_k \sim \log-F(1, 1)$ e $\alpha \sim \text{Gama}(0.2, 0.2)$;
- **Caso 4:** $\beta_k \sim \log-F(2, 2)$ e $\alpha \sim \text{Gama}(1, 1)$.

Para avaliar o comportamento assintótico das estimativas, as simulações serão baseadas em diferentes tamanhos de amostra $n : 50, 200, 600$ e 1000 . Consideramos também 95% para o intervalo de confiança e de credibilidade.

Na próxima seção apresentamos os resultados das simulações para o modelo de regressão Weibull que foram baseados apenas na abordagem Bayesiana. Para o caso do modelo de regressão de Cox, comparamos os resultados obtidos com base nas duas abordagens (Clássica e Bayesiana).

5.1 Resultados para a regressão Weibull

A distribuição a *posteriori* conjunta foi obtida a partir da função de verossimilhança da Eq. (2.8) do Capítulo 2 e apresentamos no Apêndice B.

Os resultados das Tabelas 5.1 e 5.2 foram baseados nas distribuições a *priori* $N(m_k, v_k)$ e $\log-F(a_k, b_k)$ para diferentes configurações dos hiperparâmetros, tal como apresentamos

| n | Casos | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $VR_{\hat{\alpha}}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EQM_{\hat{\alpha}}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | $EP_{\hat{\alpha}}$ | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|-------|----------------------|----------------------|---------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|---------------------|-----------------|-----------------|----------------|
| 50 | 1 | 61.21 | 9.79 | 4.43 | 1.668 | 0.082 | 0.078 | 0.41 | 0.28 | 0.27 | 75.90 | 96.20 | 96.10 |
| | 2 | 11.60 | -0.18 | 7.85 | 0.696 | 0.121 | 0.123 | 0.80 | 0.35 | 0.33 | 96.70 | 94.10 | 94.90 |
| 200 | 1 | 58.95 | 1.83 | 1.46 | 1.579 | 0.022 | 0.018 | 0.44 | 0.15 | 0.13 | 77.20 | 94.40 | 95.20 |
| | 2 | 9.56 | -0.16 | 1.94 | 0.619 | 0.022 | 0.021 | 0.76 | 0.15 | 0.14 | 96.70 | 95.10 | 93.70 |
| 600 | 1 | 59.40 | 1.05 | 0.40 | 1.596 | 0.006 | 0.006 | 0.43 | 0.08 | 0.08 | 77.70 | 95.60 | 95.40 |
| | 2 | 7.73 | -0.13 | 0.64 | 0.585 | 0.007 | 0.006 | 0.75 | 0.08 | 0.08 | 96.80 | 93.50 | 95.00 |
| 1000 | 1 | 60.44 | 0.75 | 0.17 | 1.631 | 0.004 | 0.003 | 0.41 | 0.06 | 0.06 | 77.70 | 94.10 | 95.70 |
| | 2 | 11.06 | 0.06 | 0.25 | 0.650 | 0.004 | 0.004 | 0.78 | 0.06 | 0.06 | 95.90 | 95.70 | 94.90 |

Tabela 5.1: Resultado das simulações para a abordagem Bayesiana baseado no modelo de regressão Weibull assumindo distribuições a priori dos Casos 1 e 2.

| n | Casos | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $VR_{\hat{\alpha}}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EQM_{\hat{\alpha}}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | $EP_{\hat{\alpha}}$ | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|-------|----------------------|----------------------|---------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|---------------------|-----------------|-----------------|----------------|
| 50 | 3 | -14.17 | -3.09 | 8.82 | 1.385 | 0.119 | 0.128 | 1.14 | 0.34 | 0.33 | 96.90 | 93.90 | 94.00 |
| | 4 | 29.12 | -1.48 | 8.27 | 0.866 | 0.109 | 0.122 | 0.73 | 0.33 | 0.33 | 93.00 | 94.60 | 96.10 |
| 200 | 3 | -11.90 | 0.45 | 1.79 | 1.305 | 0.023 | 0.021 | 1.12 | 0.15 | 0.14 | 96.00 | 93.70 | 95.10 |
| | 4 | 27.91 | 1.11 | 1.55 | 0.795 | 0.022 | 0.020 | 0.70 | 0.15 | 0.14 | 93.10 | 94.90 | 94.20 |
| 600 | 3 | -18.13 | 0.31 | 0.41 | 1.212 | 0.006 | 0.006 | 1.04 | 0.08 | 0.08 | 97.50 | 95.10 | 94.30 |
| | 4 | 26.59 | 0.49 | 0.44 | 0.743 | 0.007 | 0.006 | 0.68 | 0.08 | 0.08 | 93.70 | 95.40 | 94.70 |
| 1000 | 3 | -13.78 | 0.05 | 0.25 | 1.24 | 0.004 | 0.003 | 1.08 | 0.06 | 0.06 | 96.60 | 94.40 | 94.20 |
| | 4 | 29.06 | 0.30 | 0.26 | 0.790 | 0.004 | 0.003 | 0.67 | 0.06 | 0.06 | 93.60 | 93.60 | 95.30 |

Tabela 5.2: Resultado das simulações para a abordagem Bayesiana baseado no modelo de regressão Weibull assumindo distribuições a priori dos Casos 3 e 4.

anteriormente. Como na prática pouco se sabe sobre o valor do hiperparâmetro de locação m_k , optamos em centrar a distribuição a priori em zero ($m_k=0$). O hiperparâmetro de escala foi assumido como: $v_k=1$ para o Caso 1 e $v_k=5$ para o Caso 2. Os hiperparâmetros da distribuição a priori $\log-F(a_k, b_k)$ foram configurados de tal forma que ela seja centrada em zero, isto é, $a_k=b_k$.

Os resultados mostram que assumindo uma distribuição a priori suficientemente informativa para os parâmetros (Casos 1 e 4), as estimativas de β_1 (que é o parâmetro associado a covariável binária), tendem a ser mais viesadas, com valores relativamente elevados do VR e EQM . Por exemplo, o Caso 1 está associado a um VR que varia entre 58.98% – 61.21%, com EQM variando de 1.598 – 1.668 e EP entre 0.41 – 0.44. O VR associado ao Caso 4 está em torno de 26.59% – 29.12% com um EQM e EP variando em torno de 0.743 – 0.866 e 0.67 – 0.73, respectivamente. As taxas de cobertura associadas

ao Caso 1 estão abaixo do valor nominal, enquanto o Caso 4 apresenta taxas de cobertura em torno do valor nominal.

Conforme já falamos anteriormente, o nosso objetivo é testar diferentes valores para os hiperparâmetros das distribuições *a priori* que forneçam estimativas com boas propriedades, isto é, em geral menos viesadas e com taxas de cobertura em torno do valor nominal. Os resultados mostram que uma escolha adequada dos hiperparâmetros que forneçam uma variabilidade *a priori* relativamente maior (Casos 2 e 3) permitiu obter estimativas pouco viesadas para β_1 . A título de exemplo, o *VR*, *EQM* e o *EP* para o Caso 2 variam entre 7.73% – 11.60%, 0.585 – 0.696 e 0.75 – 0.80, respectivamente. Para o Caso 3, o *VR* foi subestimado; variando entre 11.90% – 18.13%. O *EQM* e o *EP* estão compreendidos entre 1.212 – 1.385 e 1.04 – 1.14, respectivamente. Em ambos os casos o β_1 está associado a taxas de coberturas em torno do valor nominal.

Os resultados mostram ainda que o aumento do tamanho de amostra tem um impacto muito pequeno nas estimativas de β_1 (parâmetro associadas a covariável binária), pois elas não registaram praticamente nenhuma melhoria. Como forma de visualizar o comportamento das estimativas para diferentes valores de n , apresentamos nas Figuras C.5 e C.6 do Apêndice C algumas curvas de densidades para os parâmetros. Foi possível constatar que para qualquer valor do n a curva associada ao β_1 mantém-se estável, por contrário, as curvas associadas aos demais parâmetros que não são afetados diretamente pelo fenômeno, como é o caso do β_2 e α , tendem a ter caudas mais leves e concentram massa de probabilidade em torno do valor real quando aumentamos o tamanho de amostra.

Esses dois parâmetros (β_2 , α) apresentam estimativas com boas propriedades assintóticas, pois para os quatro casos temos *VR*, *EQM*, e *EP* com valores relativamente pequenos e que diminuem a medida em que aumentamos o tamanho de amostra, estando associados a taxas de cobertura em torno dos valor nominal. Os *EQM*'s e *EP*'s são relativamente iguais nos quatro casos. De forma geral, estes resultados dão uma ideia do real impacto do problema da VM nas estimativas de parâmetros associadas a covariável binária.

5.2 Resultados para o modelo de Cox

Nesta seção, ajustamos o modelo de regressão de Cox com riscos proporcionais baseado nas duas abordagens que apresentamos nas seções anteriores.

5.2.1 Abordagem clássica

O ajuste do modelo foi baseado na correção de Firth. As taxas de cobertura foram obtidas usando os métodos Wald (Cob_{wa}) e verossimilhança perfilada (Cob_{pl}).

| n | %VM | %VR $_{\hat{\beta}_1}$ | %VR $_{\hat{\beta}_2}$ | EQM $_{\hat{\beta}_1}$ | EQM $_{\hat{\beta}_2}$ | EP $_{\hat{\beta}_1}$ | EP $_{\hat{\beta}_2}$ | $Cob_{\beta_1}.wa$ | $Cob_{\beta_1}.pl$ | $Cob_{\beta_2}.wa$ | $Cob_{\beta_2}.pl$ |
|------|-----|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|--------------------|--------------------|--------------------|--------------------|
| 50 | 75 | 53.93 | -2.40 | 2.200 | 0.150 | 1.43 | 0.36 | 86.90 | 91.68 | 95.10 | 94.10 |
| | 25 | 46.91 | 2.750 | 3.840 | 2.650 | 1.35 | 1.26 | 85.90 | 91.78 | 93.80 | 94.68 |
| 200 | 75 | 52.28 | -1.66 | 1.860 | 0.030 | 1.28 | 0.16 | 85.10 | 81.67 | 94.50 | 94.10 |
| | 25 | 50.45 | 0.87 | 3.050 | 0.260 | 1.04 | 0.49 | 78.90 | 90.13 | 94.80 | 95.80 |
| 600 | 75 | 53.77 | -0.70 | 1.920 | 0.010 | 1.24 | 0.09 | 83.20 | 74.09 | 95.10 | 95.10 |
| | 25 | 52.10 | 0.28 | 3.130 | 0.080 | 1.00 | 0.27 | 76.40 | 84.20 | 93.60 | 93.80 |
| 1000 | 75 | 49.40 | -0.54 | 1.720 | 0.001 | 1.27 | 0.07 | 86.80 | 72.73 | 95.40 | 95.30 |
| | 25 | 48.98 | 0.48 | 2.880 | 0.040 | 1.00 | 0.20 | 77.90 | 82.81 | 95.60 | 95.70 |

Tabela 5.3: Resultado das simulações de MC para os Cenários 1 e 2 onde assumimos que 75% e 25% das amostras tem VM. Porcentagem de censura entre 70% e 75%.

Os resultados da Tabela 5.3 mostram mais uma vez o quão difícil é estimar β_1 nos dois cenários. Por exemplo, o Cenário 1 está associado a um VR , EQM e EP relativamente elevados, variando em torno de 49.40%–53.93%, 1.72–2.20 e 1.24–1.43, respectivamente. O Cenário 2 apresenta valores do VR , EQM e EP em torno dos 46.91% – 52.10%, 2.88 – 3.84 e 1.00 – 1.35, respectivamente. As taxas de cobertura estão abaixo do valor nominal nos dois cenários. Em geral, os resultados são praticamente idênticos, isto é, diminuir a porcentagem de amostras com VM mantendo constante a porcentagem de censuras, não registrou grandes melhorias quanto a qualidade das estimativas de β_1 .

Mais uma vez, o impacto da VM nas estimativas do β_2 é quase inexistente consequentemente, suas estimativas tem boas propriedades assintóticas com VR , EQM e o EP diminuindo a medida em que aumentamos o tamanho de amostra.

Na próxima seção apresentamos resultados referentes ao ajuste do modelo de regressão

de Cox usando abordagem Bayesiana, onde diferentes distribuições *a priori* são testadas.

5.2.2 Abordagem Bayesiana

As estimativas que apresentamos nesta seção foram geradas da distribuição *a posteriori* apresentada no Apêndice B. Como ela foi baseada na função de verossimilhança parcial não dependerá do parâmetro de forma α e assim a dimensão do espaço paramétrico reduz para 2.

| n | Casos | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | Cob_{β_1} | Cob_{β_2} |
|------|-------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------|-----------------|
| 50 | 1 | 66.21 | 1.76 | 1.900 | 0.100 | 0.39 | 0.32 | 77.50 | 96.20 |
| | 2 | 16.74 | -9.64 | 0.820 | 0.170 | 0.84 | 0.40 | 96.70 | 95.40 |
| 200 | 1 | 61.09 | 0.21 | 1.650 | 0.020 | 0.40 | 0.15 | 79.70 | 95.40 |
| | 2 | 14.43 | -0.96 | 0.670 | 0.030 | 0.77 | 0.17 | 96.40 | 94.00 |
| 600 | 1 | 60.89 | 0.20 | 1.650 | 0.012 | 0.41 | 0.09 | 78.00 | 94.50 |
| | 2 | 12.41 | -0.86 | 0.660 | 0.012 | 0.79 | 0.09 | 95.90 | 94.60 |
| 1000 | 1 | 60.40 | 0.18 | 1.630 | 0.004 | 0.41 | 0.06 | 79.00 | 95.10 |
| | 2 | 9.55 | -0.64 | 0.660 | 0.004 | 0.79 | 0.06 | 96.60 | 95.30 |

Tabela 5.4: Resultado das simulações usando o modelo de regressão de Cox baseado nas configurações da distribuição *a priori* dos Casos 1 e 2.

| n | Casos | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | Cob_{β_1} | Cob_{β_2} |
|------|-------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------|-----------------|
| 50 | 3 | -8.89 | -13.26 | 1.270 | 0.180 | 1.12 | 0.41 | 96.80 | 94.10 |
| | 4 | 34.56 | -8.11 | 0.970 | 0.140 | 0.70 | 0.38 | 93.50 | 95.40 |
| 200 | 3 | -14.89 | -2.04 | 1.170 | 0.020 | 1.04 | 0.16 | 97.00 | 94.10 |
| | 4 | 28.64 | -1.38 | 0.760 | 0.020 | 0.66 | 0.15 | 93.30 | 95.30 |
| 600 | 3 | -15.10 | -0.75 | 1.190 | 0.011 | 1.05 | 0.09 | 96.20 | 93.30 |
| | 4 | 28.55 | -0.57 | 0.770 | 0.010 | 0.67 | 0.09 | 93.70 | 94.30 |
| 1000 | 3 | -16.19 | 0.15 | 1.240 | 0.001 | 1.06 | 0.07 | 96.10 | 94.30 |
| | 4 | 27.73 | 0.31 | 0.770 | 0.001 | 0.68 | 0.06 | 92.80 | 95.30 |

Tabela 5.5: Resultado das simulações usando o modelo de regressão de Cox baseado nas configurações da distribuição *a priori* dos Casos 3 e 4.

Baseado nos resultados das Tabelas 5.4 e 5.5 observamos que, de forma geral, não há grandes diferenças entre as estimativas obtidas ajustando o modelo de regressão de Cox comparadas aos resultados da Seção 5.1, onde ajustamos o modelo de regressão Weibull.

Contudo, as estimativas obtidas para o parâmetro β_1 usando as distribuições *a priori* dos Casos 2 e 3 são menos viesadas que as obtidas com base no método Clássico. O VR , EQM e o EP são relativamente menores. De igual forma, as taxas de cobertura estão em torno do valor nominal. As estimativas obtidas para o Caso 1 são similares com as obtidas usando a abordagem Clássica. Mais uma vez, as estimativas do parâmetro associado a covariável contínua apresenta boas propriedades em termos assintóticos. O VR , EQM e o EP tem valores pequenos e que diminuem a medida em que aumentamos o tamanho de amostra.

No Apêndice C apresentamos os resultados das simulações, que foram obtidos centrando as distribuições *a priori* dos Casos 1 e 2 no valor real (β_k). Apresentamos também os resultados obtidos usando distribuições *a priori* Cauchy(0,1,2) e a log- $F(9, 9)$. Em geral estas funções de penalidade forneceram estimativas ruins, isto é, muito viesadas. A distribuição *a priori* centrada no valor real $N(\beta_k, v_k)$ apresentou estimativas pouco viesadas, associadas a taxas de cobertura muito acima do valor nominal.

Distribuições *a priori* como Cauchy(0,1,2) e log- $F(9, 9)$ produzirão estimativas muito viesadas. Distribuições *a priori* representadas pelos Casos 1 e 2 quando centradas no valor real, isto é, $N(\beta_k, v_k)$ apresentaram taxas de cobertura acima do valor nominal, principalmente quando ($v_k = 1$).

5.3 Conclusão do capítulo

Neste capítulo apresentamos alguns resultados de simulação referentes as abordagens Clássica e Bayesiana. O estudo objetivou avaliar as propriedades assintóticas dos estimadores. No próximo capítulo investigamos o impacto de diferentes funções de penalidade para corrigir o fenômeno da VM no banco de dados de melanoma.

Capítulo 6

Aplicação a dados de melanoma

Apresentamos aqui os resultados da análise do modelo de regressão de Cox sob riscos proporcionais para o banco de dados referente a pacientes com melanoma, que mencionamos no Capítulo 1. Fatores epidemiológicos e histopatológicos são considerados, como por exemplo: (i) β_2 Gênero (*Feminino*); (ii) Tipo Histológico (*TH*) com os seguintes níveis β_{31} (*Nodular*), β_{32} (*Lentiginoso Acral*), o nível *extensivo superficial+lentiginoso maligno* que foi assumido como sendo de referência; (iii) Índice de Breslow (*CB*) com os níveis β_{41} (*1-4mm*), β_{42} (*> 4mm*) e o nível (*< 1mm*) foi assumido como sendo de referência; (iv) Ulceração β_5 (*Sim*); (v) β_1 Mitose ou taxa mitótica, que é a covariável associada ao problema da VM. O fator é codificado em 1, se o paciente desenvolver a “mitose” e 0, caso contrário. Metástase é o evento em estudo e o tempo de acompanhamento (medido em semanas) vai desde o diagnóstico do melanoma até a data da última consulta ou a data de aparecimento dos primeiros sinais da “metástase”.

| | Abordagem Clássica | | Abordagem Bayesiana | | | | |
|--------------------|--------------------|---------|---------------------|-----------|----------------|----------------|----------------|
| | Cox | Firth | $N(0, 1)$ | $N(0, 5)$ | $\log-F(1, 1)$ | $\log-F(2, 2)$ | $\log-F(9, 9)$ |
| $\hat{\beta}_1$ | 18.52 | 2.2214 | 1.4033 | 2.7724 | 3.6909 | 2.4310 | 1.0801 |
| $\hat{\beta}_2$ | -0.651 | -0.6414 | -0.589 | -0.6327 | -0.6298 | -0.6203 | -0.531 |
| $\hat{\beta}_{31}$ | 0.984 | 0.9578 | 1.0376 | 1.0770 | 1.0487 | 1.0465 | 0.9940 |
| $\hat{\beta}_{32}$ | 0.236 | 0.2978 | 0.1721 | 0.2332 | 0.2152 | 0.2114 | 0.1488 |
| $\hat{\beta}_{41}$ | 0.110 | 1.0356 | 0.7312 | 1.0612 | 1.0602 | 0.9221 | 0.5263 |
| $\hat{\beta}_{42}$ | 0.184 | 1.7598 | 1.4229 | 1.7823 | 1.7983 | 1.6486 | 1.2067 |
| $\hat{\beta}_5$ | 0.699 | 0.6848 | 0.7925 | 0.7594 | 0.7332 | 0.7656 | 0.7844 |

Tabela 6.1: *Estimativas dos parâmetros baseadas na abordagem Clássica e Bayesiana para o banco de dados de melanoma.*

| | Abordagem Clássica | | Abordagem Bayesiana | | | | |
|---------------------|--------------------|--------|---------------------|-----------|----------------|----------------|----------------|
| | Cox | Firth | $N(0, 1)$ | $N(0, 5)$ | $\log-F(1, 1)$ | $\log-F(2, 2)$ | $\log-F(9, 9)$ |
| $\hat{\sigma}_1$ | 5131.0 | 1.5003 | 0.6956 | 1.4272 | 2.3215 | 1.4229 | 0.5599 |
| $\hat{\sigma}_2$ | 0.356 | 0.3554 | 0.3342 | 0.3537 | 0.3484 | 0.3479 | 0.3149 |
| $\hat{\sigma}_{31}$ | 0.449 | 0.4511 | 0.3872 | 0.4535 | 0.4436 | 0.4405 | 0.3667 |
| $\hat{\sigma}_{32}$ | 0.634 | 0.6183 | 0.5334 | 0.6292 | 0.6071 | 0.5871 | 0.4515 |
| $\hat{\sigma}_{41}$ | 0.692 | 0.6657 | 0.4875 | 0.6625 | 0.6135 | 0.5894 | 0.4021 |
| $\hat{\sigma}_{42}$ | 0.734 | 0.7147 | 0.5301 | 0.6953 | 0.6505 | 0.6318 | 0.4329 |
| $\hat{\sigma}_5$ | 0.385 | 0.3872 | 0.3574 | 0.3910 | 0.3859 | 0.3756 | 0.3390 |

Tabela 6.2: *Estimativas do erro-padrão baseadas na abordagem Clássica e Bayesiana para o banco de dados de melanoma.*

As Tabelas 6.1 e 6.2, juntamente com as Figuras 6.1 e 6.2, apresentam em geral os mesmos resultados. Com exceção das estimativas do parâmetro associado a covariável binária, todos os demais parâmetros apresentam estimativas similares para diferentes funções de penalidade tanto na abordagem Clássica como Bayesiana. A maior diferença entre as estimativas do β_1 aparece quando assumimos distribuições a priori $N(0, 1)$ e $\log-F(9, 9)$, que são provavelmente as mais informativas na análise e centradas no valor errado (em torno de zero) do espaço paramétrico. A estimativa (média a posteriori) do parâmetro associado ao fator binário “mitose” é menor usando a distribuição a priori $\log-F(9, 9)$, conseqüentemente, o risco é menor ($\exp \{1.0801\} = 2.944974$) compa-

rando com o risco obtido com base na estimativa de máxima verossimilhança penalizada ($\exp\{2.2214\} = 9.220$). Isto é, o risco de um paciente com “mitose” falhar é menor no caso Bayesiano, quando assumimos distribuições a *priori* $\log-F(9, 9)$ ou $N(0, 1)$, que no Clássico.

Entre os fatores analisados, o efeito da covariável gênero feminino é negativo, pois sua estimativa penalizada e/ou a média a *posteriori* para diferentes configurações das distribuições a *priori* tem um sinal negativo, entretanto, o risco de desenvolver a “metástase” é menor nos pacientes de gênero feminino que no masculino. Esse resultado é consistente com os referenciados em alguns trabalhos da área clínica, como por exemplo, Damato et al. (2011) e Arce et al. (2014) sustentam que pacientes do gênero feminino constitui o grupo com menor risco de desenvolver a “metástase” e melhor sobrevida em relação aos do gênero masculino.

Na Figura 6.1 comparamos os intervalos de confiança (abordagem Clássica) com os intervalos de credibilidade (abordagem Bayesiana) para um nível de 95%. Com exceção da distribuição a *priori* $\log-F(1, 1)$ que apresenta um intervalo de confiança com maior amplitude para o parâmetro β_1 , a abordagem Bayesiana apresentou em geral melhores resultados do que a Clássica. De acordo com as estimativas obtidas no caso Clássico, os fatores “mitose” e “ulceração” não são importantes no modelo, para o nível de significância que assumimos, com $z=2.221/1.500=1.481$ e $z=0.685/0.387=1.769$, respectivamente. No entanto, para a abordagem Bayesiana, os dois fatores são importantes no caso em que assumimos como distribuição a *priori* $N(0,1)$, $N(0,5)$ e $\log-F(9, 9)$. Os fatores TH (*Nodular*) e CB ($> 4mm$) são importantes no modelo para as duas abordagens. Para uma melhor visualização, apresentamos nas Tabelas 4 a 10 do Apêndice D os intervalos de credibilidade para os parâmetros. Em geral, as estimativas do erro-padrão são menores no método Bayesiano que no Clássico (Figura 6.2).

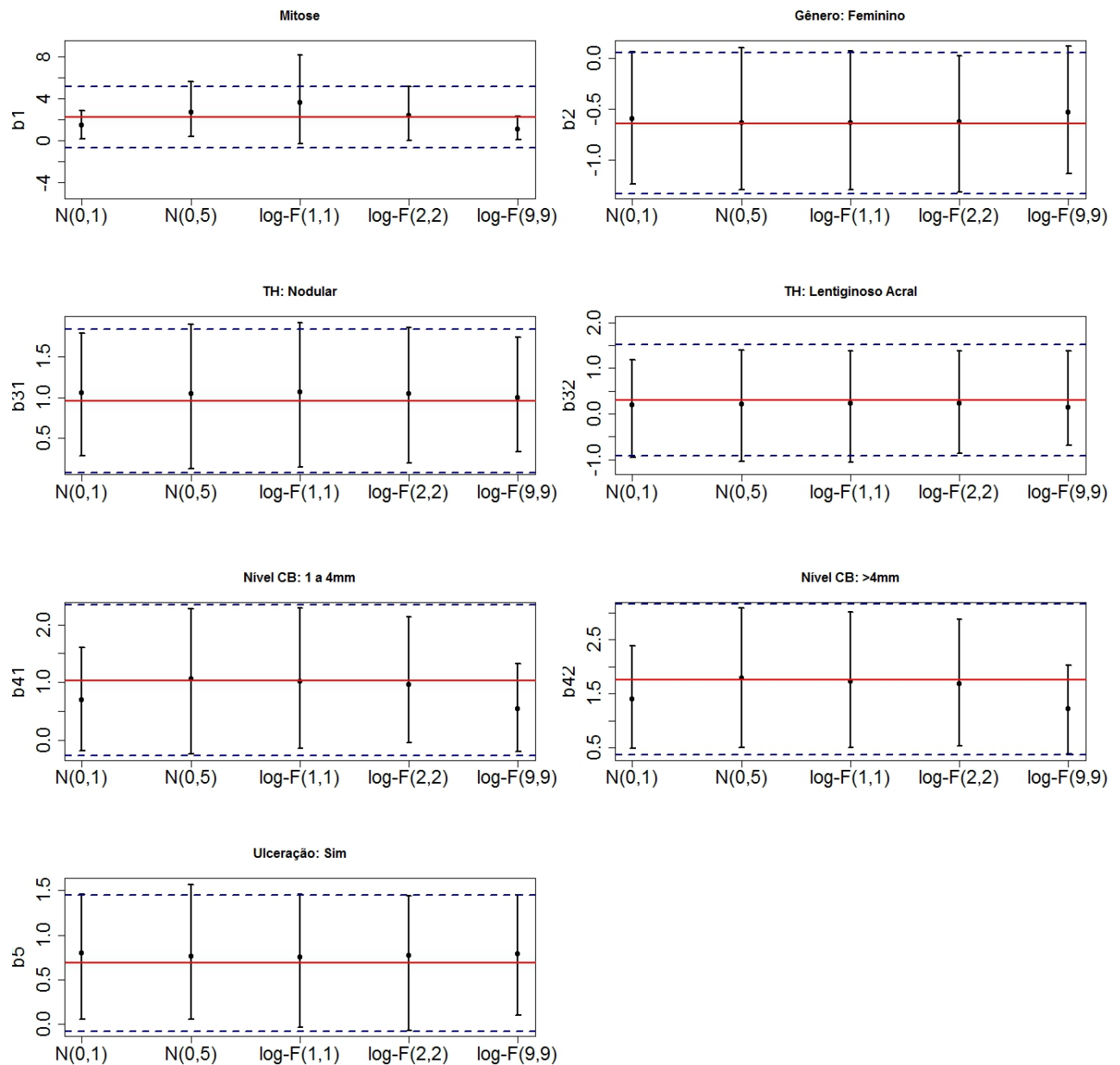


Figura 6.1: Comparação: intervalo de confiança de 95% (linhas tracejadas) e o intervalo de credibilidade de 95% (linhas sólidas e verticais). As linhas vermelhas e horizontais representam as estimativas de máxima verossimilhança penalizada (abordagem Clássica). Os pequenos pontos circulares indicam as médias a posteriori.

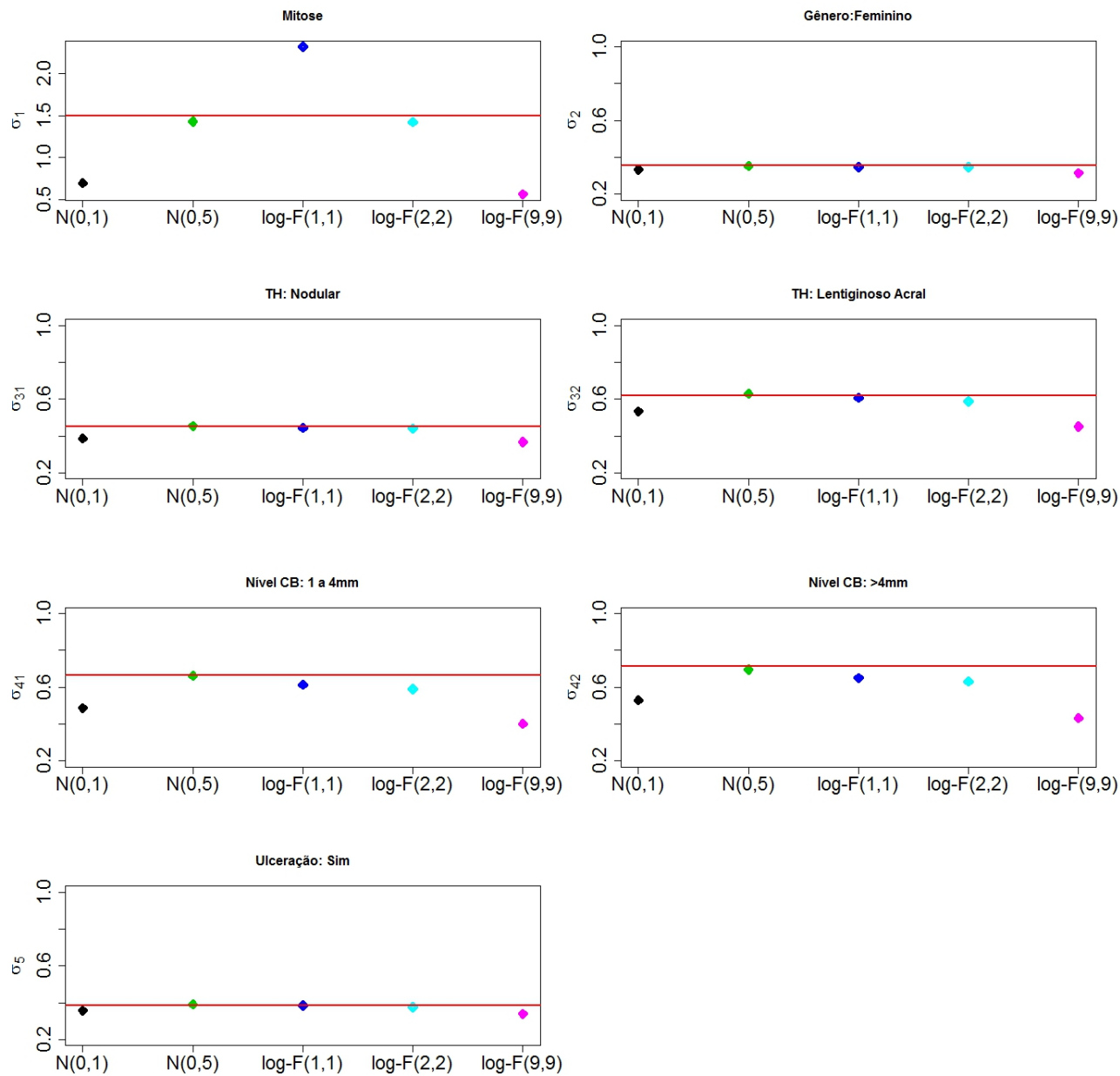


Figura 6.2: Comparação das estimativas do erro-padrão obtidas com base na abordagem Clássica (linha vermelha horizontal) e Bayesiana (pontos sólidos, para diferentes distribuições a priori).

Capítulo 7

Considerações finais

O problema da VM ocorre com frequência quando se analisamos dados envolvendo eventos raros e em pequenas amostras. Este trabalho teve uma contribuição positiva ao investigar outras funções de penalidade para corrigir este fenômeno. A proposta de análise é baseada na abordagem Bayesiana tendo como um dos focos centrais a comparação de desempenho com o caso Clássico que é baseado na proposta de penalização de Firth (1993). O ajuste foi referente aos modelos de regressão de Cox e Weibull com riscos proporcionais, onde desenvolvemos um estudo de simulação assumindo diferentes configurações das distribuições *a priori* para os coeficientes. O objetivo foi identificar a configuração entre as estudadas que fornece estimativas com o menor vício possível e taxas de cobertura em torno do valor nominal.

Consideramos dois cenários com aproximadamente 75% e 25% de amostras com VM, com porcentagem de censura variando entre 70% e 75%. Nos dois casos, o coeficiente associado a covariável binária apresenta taxas de cobertura abaixo do valor nominal, tanto para o método de Wald quanto para a verossimilhança perfilada, suas estimativas são muito viesadas para diferentes tamanhos de amostras. Em geral, a abordagem Bayesiana forneceu melhores estimativas nos dois modelos aqui considerados em comparação com a Clássica, pois permitiu reduzir o VR , EQM e EP associados a β_1 . A abordagem Bayesiana com o uso de diferentes funções de penalidade permitiu também estabilizar a taxa de cobertura em torno do valor nominal. Concluímos também que o aumento do tamanho de amostra não tem uma influência direta nas estimativas do parâmetro

associado a covariável binária.

As duas abordagens apresentadas neste trabalho, mostraram que as estimativas dos parâmetros que não são afetados diretamente pelo problema da VM como por exemplo β_2 e α são pouco viesadas e apresentam boas propriedades em termos assintóticos.

Os resultados da análise do banco de dados referente a pacientes com melanoma mudam de acordo com a escolha da distribuição a *priori*. Em geral, a abordagem Bayesiana teve um bom desempenho em relação a abordagem Clássica, pois foi possível obter estimativas finitas com erros-padrão relativamente menores. As funções de penalidade que investigamos neste trabalho permitiram identificar alguns fatores importantes no modelo que não foram identificados na abordagem Clássica, como é o caso da “mitose” e “ulceração”. Curiosamente, as distribuições a *priori* $N(0,1)$ e $\log-F(9,9)$ que apresentaram resultados ruins no estudo simulado, forneceram valores pequenos para a estimativa do EP na análise do banco de dados de melanoma. De forma geral, concluímos que as distribuições a *priori* $N(0,5)$ e $\log-F(1,1)$, indicando maior incerteza a *priori*, forneceram bons resultados tanto no estudo simulado quanto na análise do banco de dados de melanoma.

Por fim, parece ser mais sensato trabalhar com distribuições a *priori* pouco informativas (maior variância) para os coeficientes visando corrigir o problema da VM. Uma distribuição com variância pequena pode determinar maior vício se estiver centrada no valor errado.

7.1 Trabalhos futuros

Uma proposta para trabalhos futuros seria implementar a metodologia de correção do viés no contexto Clássico baseado no modelo de regressão Weibull. Outro aspecto que merece especial atenção e que pode ter causado o problema da VM é a presença dos *missings datas* no conjunto de dados de melanoma. Isto porque a VM ocorre quando descartamos tais indivíduos com observações faltantes nas análises. Como o principal foco neste trabalho foi estudar a correção do problema através do método de penalização conforme a proposta de Firth, não fizemos nenhum estudo sobre os métodos de imputação. Diante disso, uma proposta para trabalhos futuros seria usar os métodos de imputação na tentativa de resolver o problema da VM neste banco de dados. Propomos também:

- Comparar diferentes modelos usando os critérios AIC, BIC, LPML, entre outros;
- Especificar *prioris* para os parâmetros da distribuição log-F e;
- Comparar a razão dos vícios relativos de dois modelos diferentes para o mesmo banco de dados.

Apêndice

Apêndice A: Conceitos adicionais

Estimação não-paramétrica

Segundo Lee e Wang (2003) Os métodos não-paramétricos são menos eficientes que os paramétricos, quando os tempos de sobrevivência seguem uma distribuição teórica e mais eficientes quando nenhuma distribuição teórica é conhecida. Portanto, os autores sugerem o uso dos métodos não-paramétricos para a análise dos dados de sobrevivência antes de tentar encaixá-lo a uma distribuição teórica.

Caso o objetivo seja de encontrar um modelo adequado para os dados, as estimativas obtidas pelos métodos não-paramétricos e gráficos podem ajudar na escolha da distribuição dos T , que definimos no Capítulo 1.

Estimador de Kaplan-Meier

Kaplan e Meier (1958) propuseram um estimador não-paramétrico para a função de sobrevivência, que é conhecido por estimador de Kaplan-Meier ou estimador limite-produto. Na sua construção, o estimador de Kaplan-Meier incorpora informação de todas as observações disponíveis (censuradas e não censuradas). Entretanto, ela é uma adaptação da função de sobrevivência empírica que, na ausência de censuras é definida como,

$$\hat{S}(t) = \frac{\text{número de observações que não falharam até o tempo } t}{\text{número total de observações no estudo}}, \quad t \geq 0.$$

Na presença de censura à direita, o estimador de Kaplan-Meier para a função de sobrevivência é dado por:

$$\hat{S}(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right),$$

sendo n_i o número de indivíduos sob risco e d_i o número de falhas ocorridas em T_i detalhes são encontrados em Colosimo e Giolo (2006) e Collett (2003).

Algoritmo de Newton-Rapson

Seja $U(\boldsymbol{\beta})$ vetor $p \times 1$ das primeiras derivadas da Eq. 2.4 em relação a β_k , também conhecida como função escore é dada por:

$$U_s(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_s} = \sum_{i=1}^n \delta_i \left\{ x_{is} - \frac{\sum_{j \in R(t_{(i)})} x_{js} \exp \{ \mathbf{x}_j \boldsymbol{\beta}' \}}{\sum_{j \in R(t_{(i)})} \exp \{ \mathbf{x}_j \boldsymbol{\beta}' \}} \right\}.$$

Calculada a função escore, seja $I(\boldsymbol{\beta})$ uma matriz não negativa com dimensão $p \times p$, que é composta pelas segundas derivadas do logaritmo da função de verossimilhança, também conhecida como matriz de informação observada. Assim, o elemento da k -ésima linha e s -ésima coluna de $I(\boldsymbol{\beta})$ será dada por:

$$I_{ks}(\boldsymbol{\beta}) = - \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_s \partial \beta_k}$$

De acordo com o procedimento de Newton-Rapson, uma estimativa do parâmetro $\boldsymbol{\beta}$ do $s + 1$ ciclo do procedimento iterativo $\beta_k^{(s+1)}$ será dado por:

$$\hat{\beta}_k^{(s+1)} = \hat{\beta}_k^{(s)} + \mathbf{I}^{-1}(\hat{\beta}_k^{(s+1)}) U(\hat{\beta}_k^{(s)}).$$

Para um inteiro $s = 0, 1, 2, \dots$ sendo $U(\hat{\beta}_k)$ função escore ou de gradientes e $\mathbf{I}^{-1}(\hat{\beta}_k)$ a inversa da matriz de informação observada, ambas avaliadas em $\hat{\beta}_k$. Esse processo pode ser iniciado assumindo que $\hat{\beta}_k^{(0)} = 0$. Ele termina quando a mudança no logaritmo da verossimilhança é suficientemente pequena ou, para um determinado valor de $\epsilon > 0$, a diferença absoluta entre $\hat{\beta}_k^{(s+1)}$ e $\hat{\beta}_k^{(s)}$ for relativamente pequena.

$$\left\| \hat{\beta}_k^{(s+1)} - \hat{\beta}_k^{(s)} \right\| < \epsilon, \quad \text{e} \quad L(\hat{\beta}_k^{(s+1)}) - L(\hat{\beta}_k^{(s)}) < \epsilon.$$

Assim que o procedimento iterativo convergir, a matriz de variância-covariância dos parâmetros estimados será dada pela inversa da matriz de informação observada.

Apêndice B: Conceitos adicionais em inferência Bayesiana

Distribuições Conjugadas

A partir do conhecimento que se tem sobre β , pode se definir uma família paramétrica de densidades. Nesta caso, a distribuição a *priori* será representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Esses parâmetros indexados à família de distribuições a *priori* são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse.

Definição: Seja $F = \{p(D|\beta_k \in \beta)\}$ uma classe de distribuições amostrais, então uma classe de distribuições P é conjugada a F se,

$$\forall p(D|\beta_k) \in F \quad \text{e} \quad p(\beta) \in P \Rightarrow p(\beta|D) \in P.$$

Gamerman et al. (2014) alertam para os cuidados na utilização indiscriminada de distribuições conjugadas. Essencialmente, o problema é que tal utilização nem sempre é uma reparametrização adequada da incerteza a *priori*. A família conjugada pode ser obtida da seguinte forma:

- Identifique a classe P de distribuições para β tal que $l(\beta)$ seja proporcional a um membro desta classe.
- A classe P deve ser fechada por amostragem, i.e, $\forall P_1$ e $P_2 \in P$, $\exists \kappa$ tal que $\kappa P_1 P_2 \in P$.

Se além disso existe uma constante $\kappa^{-1} = \int l(\beta; D) d\beta < \infty$ e todo $P_1, P_2 \in P$ é definido como $p(\beta) = \kappa l(\beta; D)$ então P é uma família conjugada natural ao modelo amostral gerador de $l(\beta; D)$.

Priori não-informativa: Uma *priori* não informativa é usada quando não existe informação *priori* inicial palpável quer de natureza objectiva ou subjetiva, frequentemente chamado de estado de “ignorância a *priori*”, ou quando o conhecimento a *priori* é pouco significativo relativamente a informação amostral, o chamado estado de conhecimento “vago ou difuso”.

Método de Bayes-Laplace: O método de Bayes-Laplace ou princípio da razão insuficiente diz que na ausência da razão suficiente para privilegiar algumas possibilidades em detrimento de outras, decorrente da escassez informativa a *priori*, deve-se adotar a equiprobabilidade.

Para o espaço paramétrico finito e numerável $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$, a distribuição a *priori* não-informativa gerada por esse argumento é a distribuição uniforme discreta:

$$p(\beta_k) = \frac{1}{p}, \quad \beta_k \in \boldsymbol{\beta}$$

Distribuição a *posteriori* conjunta e marginais

No modelo de regressão Weibull temos que:

$$p(\alpha, \boldsymbol{\beta} | D) \propto \alpha^r \exp \left\{ r\lambda + \sum_{i=1}^n (\delta_i(\alpha - 1) \log(t_i)) - \exp \{ \lambda \} t_i^\alpha \right\} p(\boldsymbol{\beta}),$$

sendo $p(\boldsymbol{\beta})$ a distribuição a *priori* discutida no Capítulo 3, $r = \sum_{i=1}^n \delta_i$ e $\lambda = \exp \{ \boldsymbol{\beta}' \mathbf{x}_i \}$.

As distribuições marginais para os parâmetros quando assumimos que $\alpha \sim \text{Gama}(\phi, \theta)$ e $\beta_k \sim N(m_k, v_k)$ são dadas por:

$$p(\alpha | \boldsymbol{\beta}, D) \propto \alpha^{r+\phi-1} \exp \left\{ \alpha \sum_{i=1}^n \delta_i \log(t_i) - \alpha\theta - \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} t_i^\alpha \right\},$$

$$p(\beta_1 | \alpha, \beta_2, D) \propto \exp \left\{ r\beta_1 x_1 - \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} t_i^\alpha - \frac{1}{2v_1} (\beta_1^2 - 2\beta_1 m_1) \right\},$$

$$p(\beta_2 | \alpha, \beta_1, D) \propto \exp \left\{ r\beta_2 x_2 - \exp \{ \mathbf{x}_i \boldsymbol{\beta} \} t_i^\alpha - \frac{1}{2v_2} (\beta_2^2 - 2\beta_2 m_2) \right\},$$

No modelo de Cox teremos a seguinte distribuição conjunta para os coeficientes.

$$p(\boldsymbol{\beta} | D) \propto \prod_{i=1}^n \left(\frac{\exp \{ \mathbf{x}'_i \boldsymbol{\beta} \}}{\sum_{j \in R(t_{(i)})} \exp \{ \mathbf{x}'_j \boldsymbol{\beta} \}} \right)^{\delta_i} p(\boldsymbol{\beta}).$$

lembrando que $R(t_{(i)})$ representa o conjunto de todos indivíduos sob risco de falhar em $t_{(i)}$. As distribuições marginais são dadas por:

$$p(\beta_1|\beta_2, D) \propto \prod_{i=1}^n \left(\frac{\exp \{ \mathbf{x}_1 \beta_1 \}}{\sum_{j \in R(t_{(i)})} \exp \{ \mathbf{x}'_j \beta \}} \right)^{\delta_i} p(\beta_1).$$

$$p(\beta_2|\beta_1, D) \propto \prod_{i=1}^n \left(\frac{\exp \{ \mathbf{x}_2 \beta_2 \}}{\sum_{j \in R(t_{(i)})} \exp \{ \mathbf{x}'_j \beta \}} \right)^{\delta_i} p(\beta_2).$$

Apêndice C: Resultados adicionais do estudo simulado

Apresentamos a seguir algumas tabelas e gráficos relacionados ao modelo de regressão Weibull.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $VR_{\hat{\alpha}}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EQM_{\hat{\alpha}}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | $EP_{\hat{\alpha}}$ | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|----------------------|----------------------|---------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|---------------------|-----------------|-----------------|----------------|
| 50 | -3.89 | -3.97 | 6.50 | 0.16 | 0.10 | 0.11 | 0.39 | 0.31 | 0.31 | 100.00 | 95.50 | 95.80 |
| 200 | -3.90 | 0.06 | 0.99 | 0.17 | 0.02 | 0.02 | 0.40 | 0.15 | 0.13 | 99.80 | 94.00 | 95.10 |
| 600 | -4.38 | 0.25 | 0.39 | 0.15 | 0.01 | 0.01 | 0.38 | 0.09 | 0.08 | 99.90 | 94.00 | 93.70 |
| 1000 | -4.38 | 0.07 | 0.50 | 0.16 | 0.00 | 0.00 | 0.39 | 0.06 | 0.06 | 99.90 | 94.80 | 94.20 |

Tabela 1: Resultado das simulações para a abordagem Bayesiana baseados no modelo de Weibull com riscos proporcionais assumindo distribuições a priori $N(\beta_k, 1)$.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $VR_{\hat{\alpha}}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EQM_{\hat{\alpha}}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | $EP_{\hat{\alpha}}$ | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|----------------------|----------------------|---------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|---------------------|-----------------|-----------------|----------------|
| 500 | -32.43 | -3.44 | 9.72 | 1.38 | 0.12 | 0.14 | 0.98 | 0.34 | 0.35 | 98.80 | 94.60 | 94.20 |
| 200 | -32.94 | -0.87 | 1.78 | 1.37 | 0.02 | 0.02 | 0.97 | 0.15 | 0.14 | 98.40 | 95.20 | 94.20 |
| 600 | -29.04 | -0.57 | 0.69 | 1.24 | 0.01 | 0.01 | 0.95 | 0.08 | 0.08 | 99.00 | 94.30 | 94.60 |
| 1000 | -31.53 | 0.03 | 0.40 | 1.23 | 0.00 | 0.00 | 0.92 | 0.06 | 0.06 | 98.80 | 95.10 | 94.50 |

Tabela 2: Resultado das simulações para a abordagem Bayesiana baseados no modelo de Weibull com riscos proporcionais assumindo distribuições a priori $N(\beta_k, 5)$.

| n | VR_{β_1} | VR_{β_2} | VR_{α} | EQM_{β_1} | EQM_{β_2} | EQM_{α} | EP_{β_1} | EP_{β_2} | EP_{α} | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|----------------|----------------|---------------|-----------------|-----------------|----------------|----------------|----------------|---------------|-----------------|-----------------|----------------|
| 50 | 75.91 | 17.12 | 5.39 | 2.397 | 0.088 | 0.099 | 0.30 | 0.27 | 0.31 | 8.40 | 94.00 | 95.80 |
| 200 | 75.03 | 5.57 | 0.97 | 2.361 | 0.023 | 0.018 | 0.33 | 0.14 | 0.13 | 10.00 | 93.00 | 95.10 |
| 600 | 73.35 | 1.90 | 0.19 | 2.247 | 0.064 | 0.006 | 0.31 | 0.08 | 0.08 | 10.30 | 94.80 | 94.20 |
| 1000 | 74.14 | 1.00 | 0.11 | 2.295 | 0.004 | 0.003 | 0.31 | 0.06 | 0.06 | 9.30 | 94.80 | 94.80 |

Tabela 3: Resultado das simulações para a abordagem Bayesiana baseados no modelo de Weibull com riscos proporcionais assumindo distribuições a priori $\log-F(9, 9)$.

| n | VR_{β_1} | VR_{β_2} | VR_{α} | EQM_{β_1} | EQM_{β_2} | EQM_{α} | EP_{β_1} | EP_{β_2} | EP_{α} | Cob_{β_1} | Cob_{β_2} | Cob_{α} |
|------|----------------|----------------|---------------|-----------------|-----------------|----------------|----------------|----------------|---------------|-----------------|-----------------|----------------|
| 50 | -179.54 | 3.39 | 7.99 | 110.270 | 0.109 | 0.122 | 9.87 | 0.33 | 0.33 | 93.20 | 94.70 | 94.90 |
| 200 | -347.53 | 2.05 | 2.18 | 8891.812 | 0.020 | 0.019 | 94.09 | 0.14 | 0.14 | 91.30 | 95.30 | 95.20 |
| 600 | -203.13 | -0.16 | 0.71 | 396.233 | 0.007 | 0.006 | 19.50 | 0.08 | 0.08 | 90.50 | 93.80 | 94.80 |
| 1000 | -232.33 | 0.42 | 0.27 | 1611.595 | 0.004 | 0.003 | 39.89 | 0.06 | 0.06 | 93.00 | 93.70 | 95.70 |

Tabela 4: Resultado das simulações para a abordagem Bayesiana baseados no modelo de Weibull com riscos proporcionais assumindo distribuições a priori Cauchy(0, 1.2).

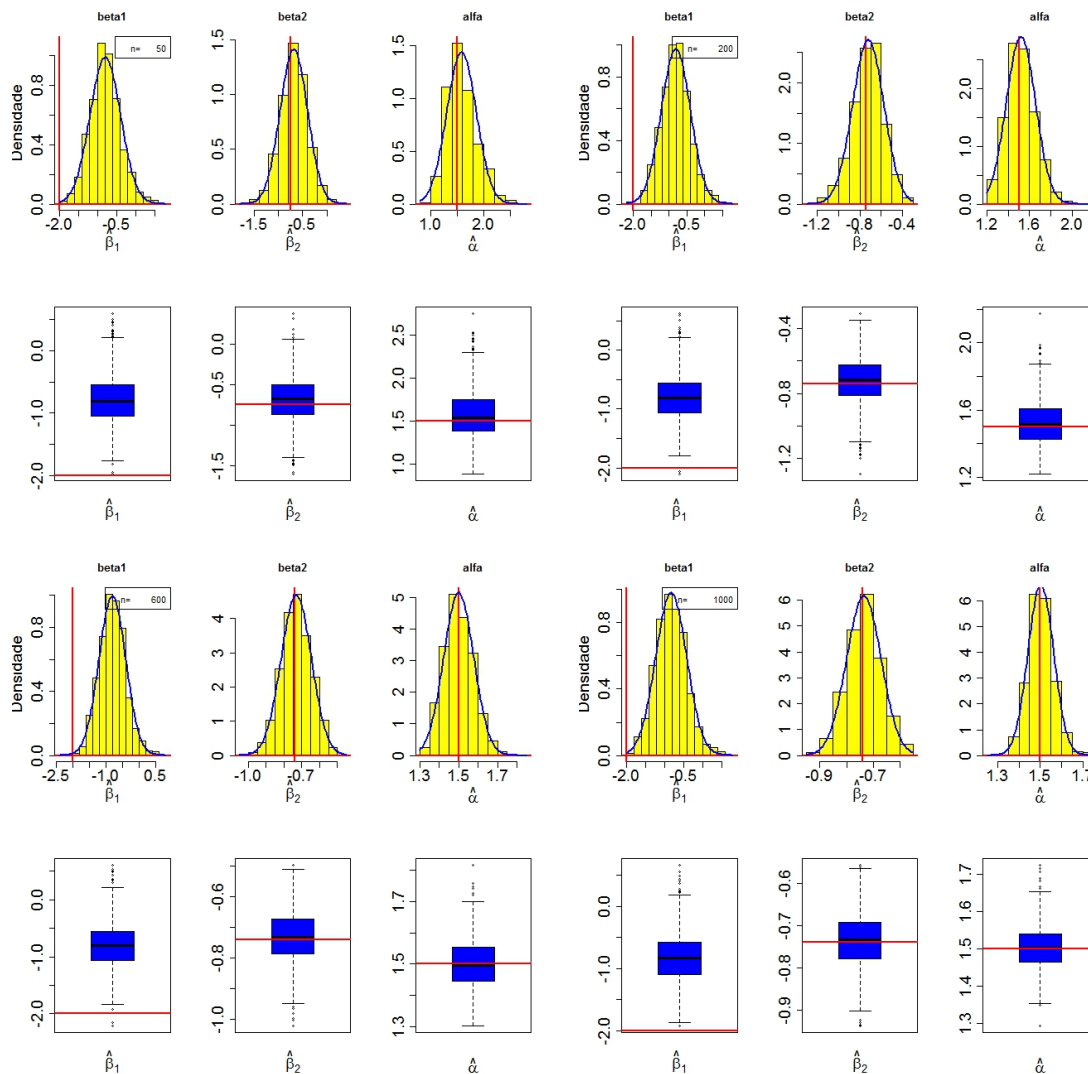


Figura C.1: Boxplots e histogramas para o Caso 1 do método Bayesiano.

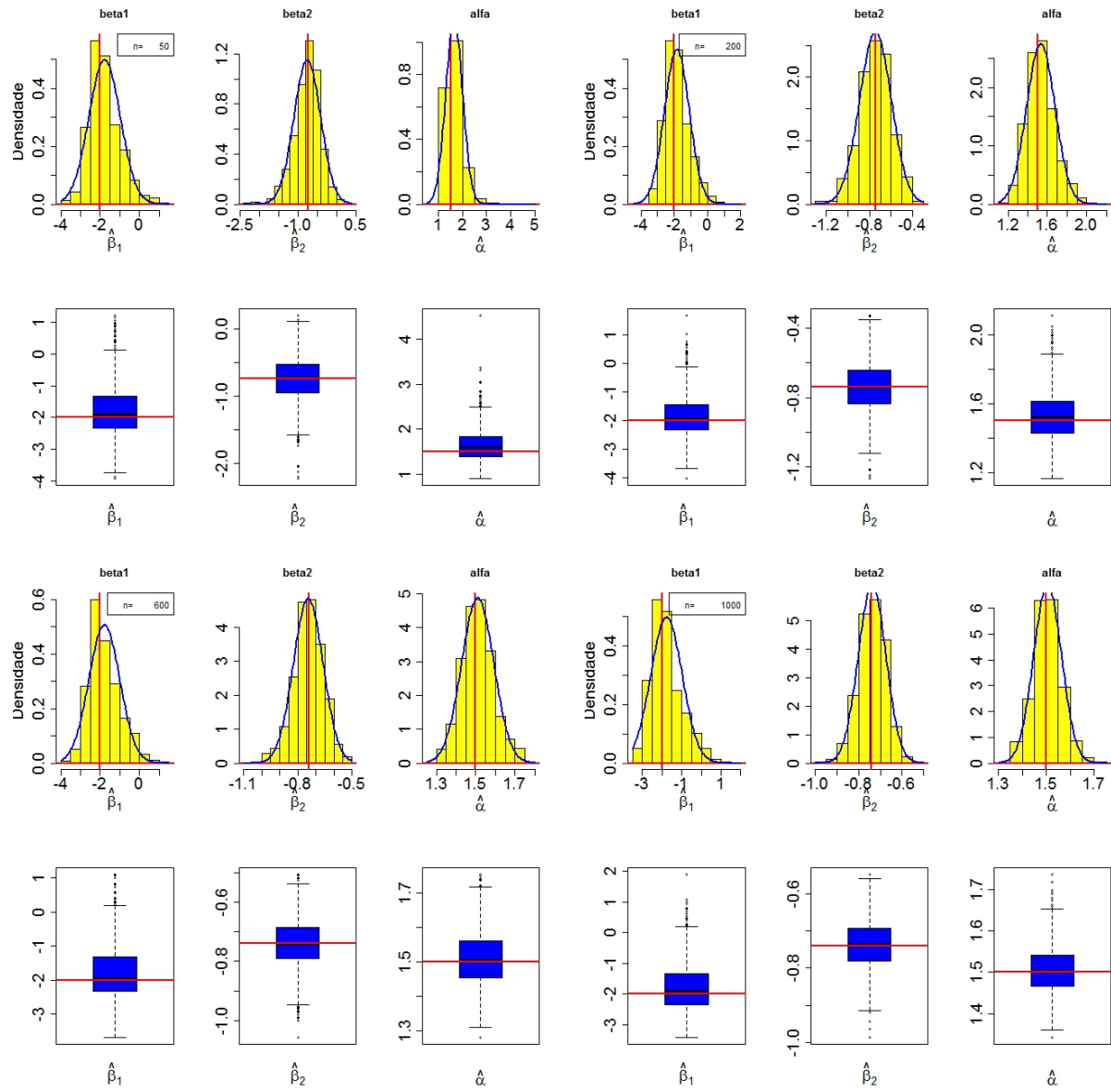


Figura C.2: *Boxplots e histogramas para o Caso 2 do método Bayesiano.*

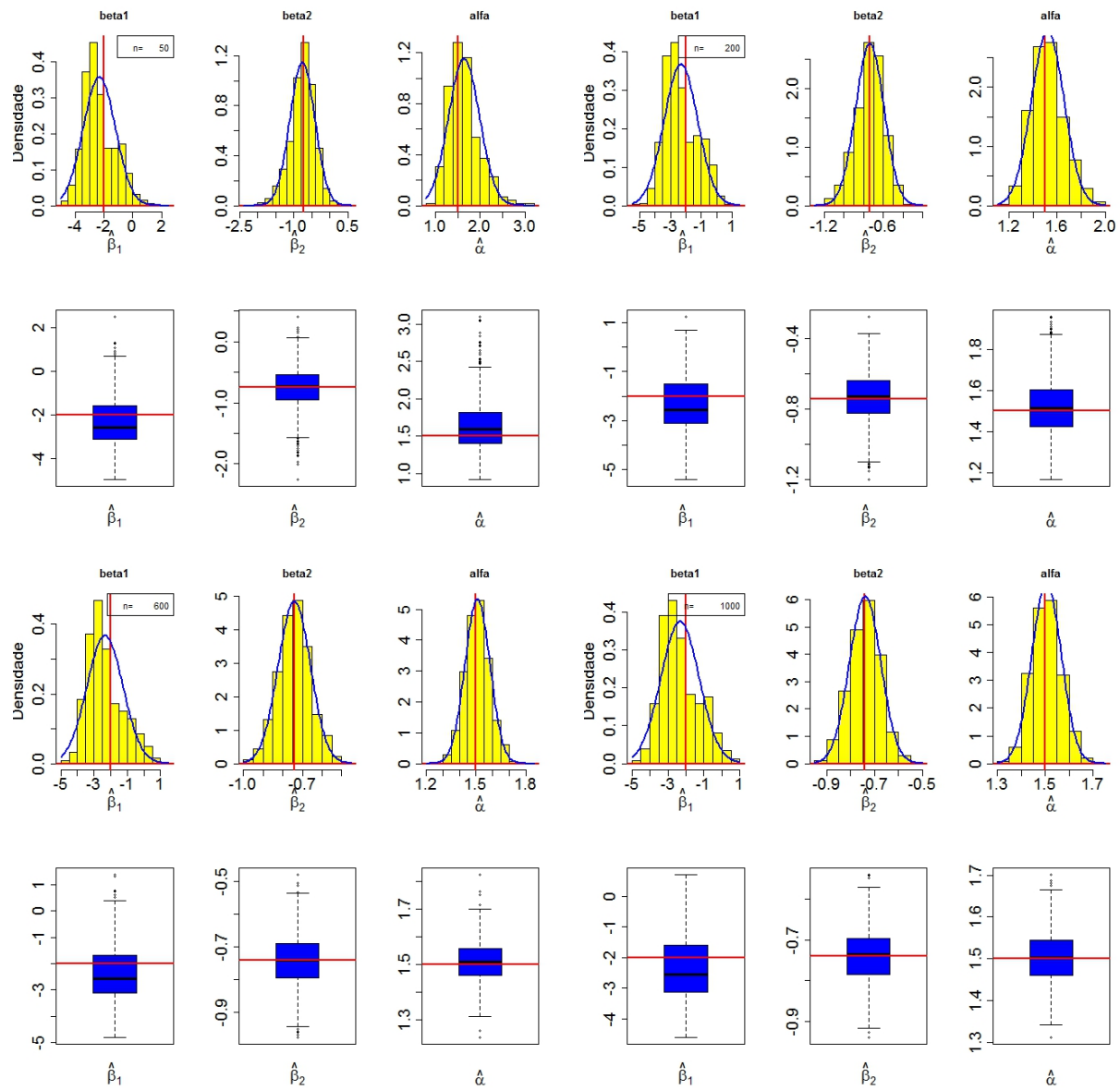


Figura C.3: *Boxplots e histogramas para o Caso 3 do método Bayesiano.*

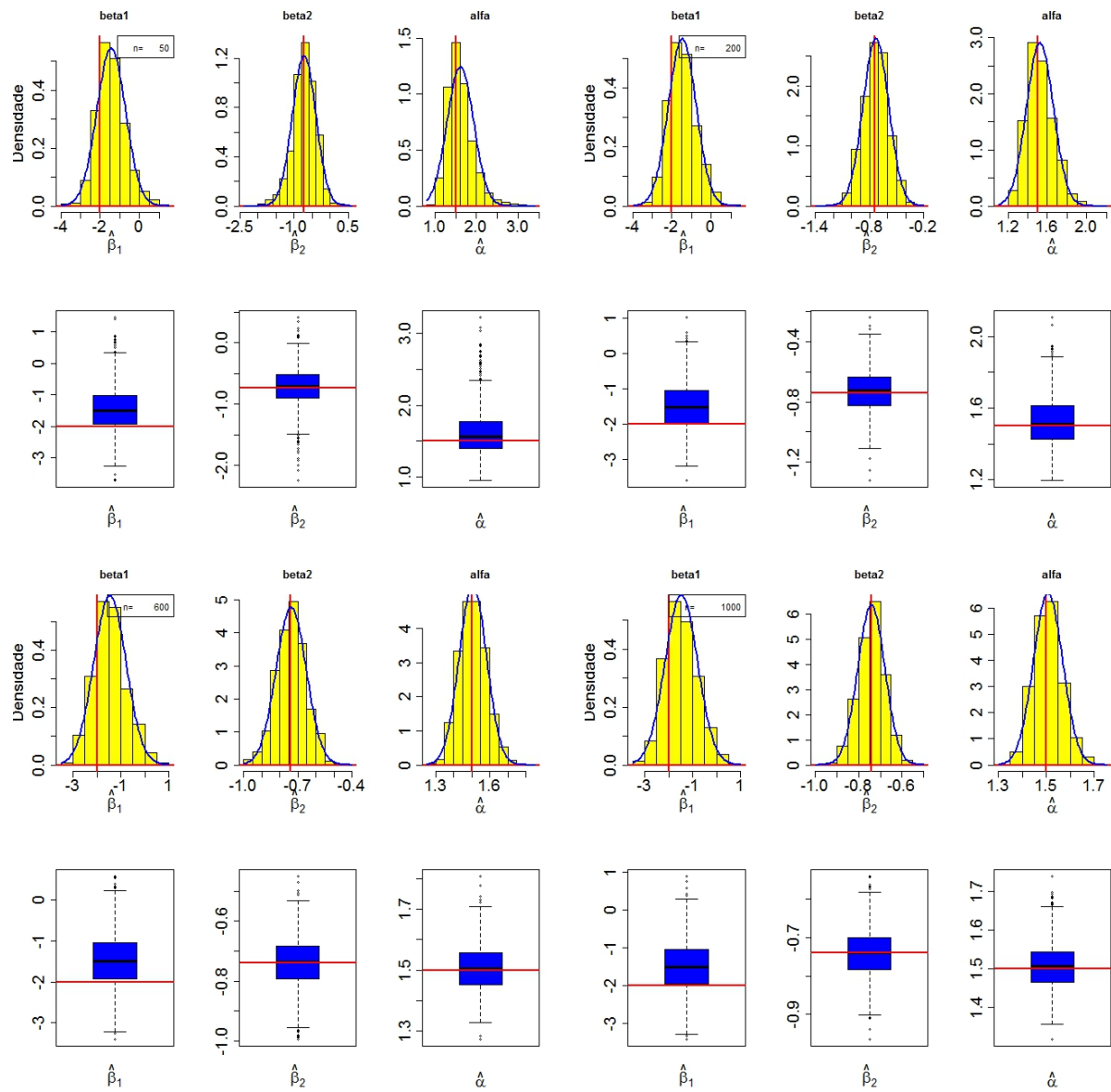


Figura C.4: *Boxplots e histogramas para o Caso 4 do método Bayesiano.*

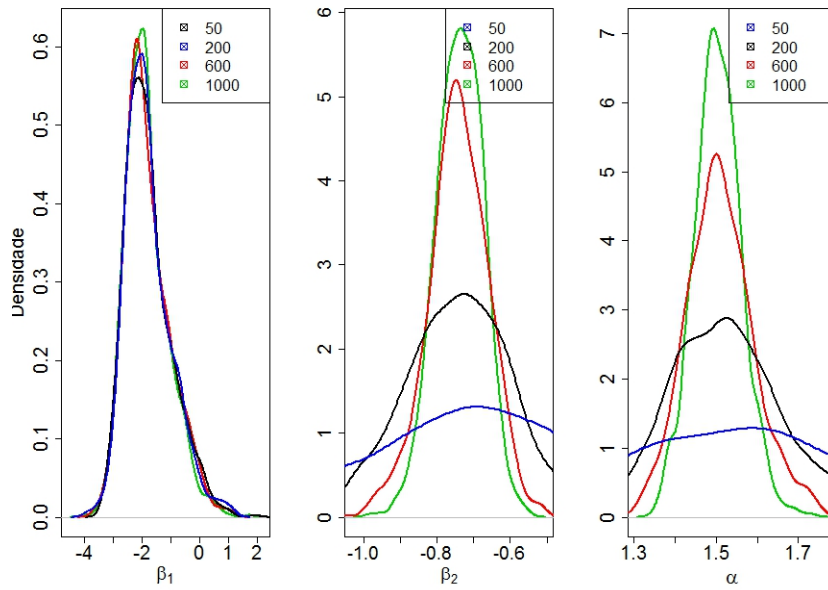


Figura C.5: *Curvas de densidades para o modelo Weibull assumindo distribuição a priori do Caso 2.*

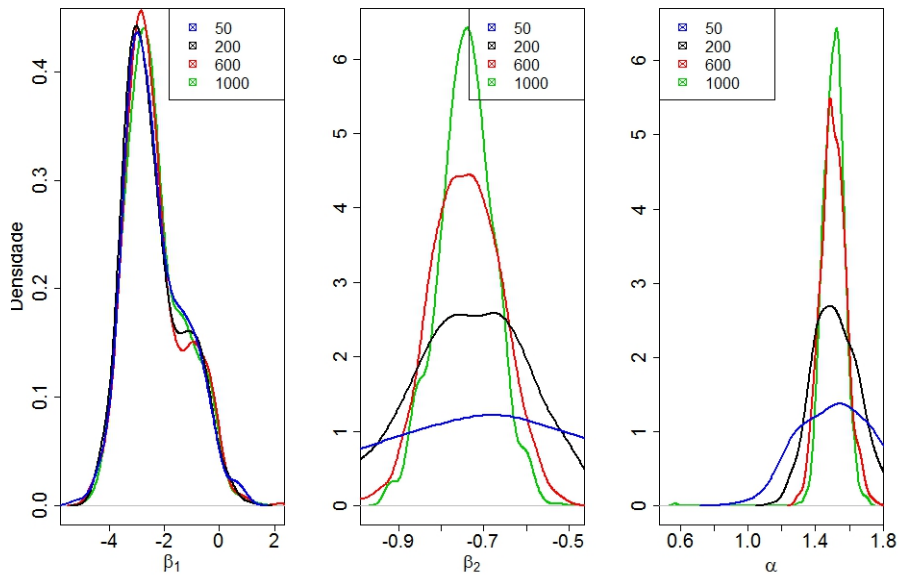


Figura C.6: *Curvas de densidades para o modelo Weibull assumindo distribuição a priori do Caso 3.*

As tabelas e gráficos mostrados a seguir são referentes ao modelo de regressão de Cox.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | Cob_{β_1} | Cob_{β_2} |
|------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------|-----------------|
| 50 | -3.39 | -12.06 | 0.15 | 0.12 | 0.38 | 0.34 | 99.80 | 95.70 |
| 200 | -4.35 | -2.48 | 0.16 | 0.02 | 0.38 | 0.15 | 99.00 | 95.30 |
| 600 | -4.13 | -0.90 | 0.16 | 0.00 | 0.39 | 0.09 | 99.00 | 94.30 |
| 1000 | -4.29 | 0.10 | 0.16 | 0.00 | 0.39 | 0.06 | 99.90 | 95.50 |

Tabela 5: Resultado das simulações para a abordagem Bayesiana baseados no modelo de regressão de Cox com distribuições a priori $N(\beta_k, 1)$ para os coeficientes.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | Cob_{β_1} | Cob_{β_2} |
|------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------|-----------------|
| 50 | -27.75 | -16.55 | 1.24 | 0.18 | 0.96 | 0.41 | 99.10 | 94.70 |
| 200 | -30.80 | -2.66 | 1.22 | 0.02 | 0.91 | 0.16 | 98.90 | 94.90 |
| 600 | -30.77 | -0.95 | 1.24 | 0.01 | 0.93 | 0.09 | 98.50 | 94.00 |
| 1000 | -31.44 | 0.07 | 1.26 | 0.00 | 0.94 | 0.06 | 98.60 | 95.20 |

Tabela 6: Resultado das simulações para a abordagem Bayesiana baseados no modelo de regressão de Cox com distribuição a priori $N(\beta_k, 5)$ para os coeficientes.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | Cob_{β_1} | Cob_{β_2} |
|------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------|-----------------|
| 50 | 79.24 | 14.16 | 257.75 | 8.65 | 0.26 | 0.27 | 4.40 | 95.00 |
| 200 | 75.01 | 3.07 | 233.35 | 2.14 | 0.29 | 0.14 | 8.10 | 94.70 |
| 600 | 74.66 | 0.83 | 232.11 | 0.75 | 0.30 | 0.09 | 10.20 | 94.50 |
| 1000 | 74.26 | 1.13 | 229.68 | 0.42 | 0.30 | 0.06 | 10.60 | 95.40 |

Tabela 7: Resultado das simulações para a abordagem Bayesiana baseados no modelo de regressão de Cox com distribuições a priori $\log-F(9, 9)$ para os coeficientes.

| n | $VR_{\hat{\beta}_1}$ | $VR_{\hat{\beta}_2}$ | $EQM_{\hat{\beta}_1}$ | $EQM_{\hat{\beta}_2}$ | $EP_{\hat{\beta}_1}$ | $EP_{\hat{\beta}_2}$ | $Cob_{\hat{\beta}_1}$ | $Cob_{\hat{\beta}_2}$ |
|------|----------------------|----------------------|-----------------------|-----------------------|----------------------|----------------------|-----------------------|-----------------------|
| 50 | -214.74 | -13.13 | 138.93 | 0.18 | 10.98 | 0.41 | 96.60 | 94.90 |
| 200 | -215.37 | -1.96 | 96.34 | 0.02 | 8.82 | 0.15 | 96.60 | 94.90 |
| 600 | -238.16 | -0.73 | 171.01 | 0.01 | 12.18 | 0.09 | 96.20 | 94.70 |
| 1000 | -235.81 | 0.17 | 141.00 | 0.00 | 10.90 | 0.07 | 96.20 | 94.30 |

Tabela 8: Resultado das simulações para a abordagem Bayesiana baseados no modelo de regressão de Cox com distribuição a priori Cauchy(0, 1.2) para os coeficientes.

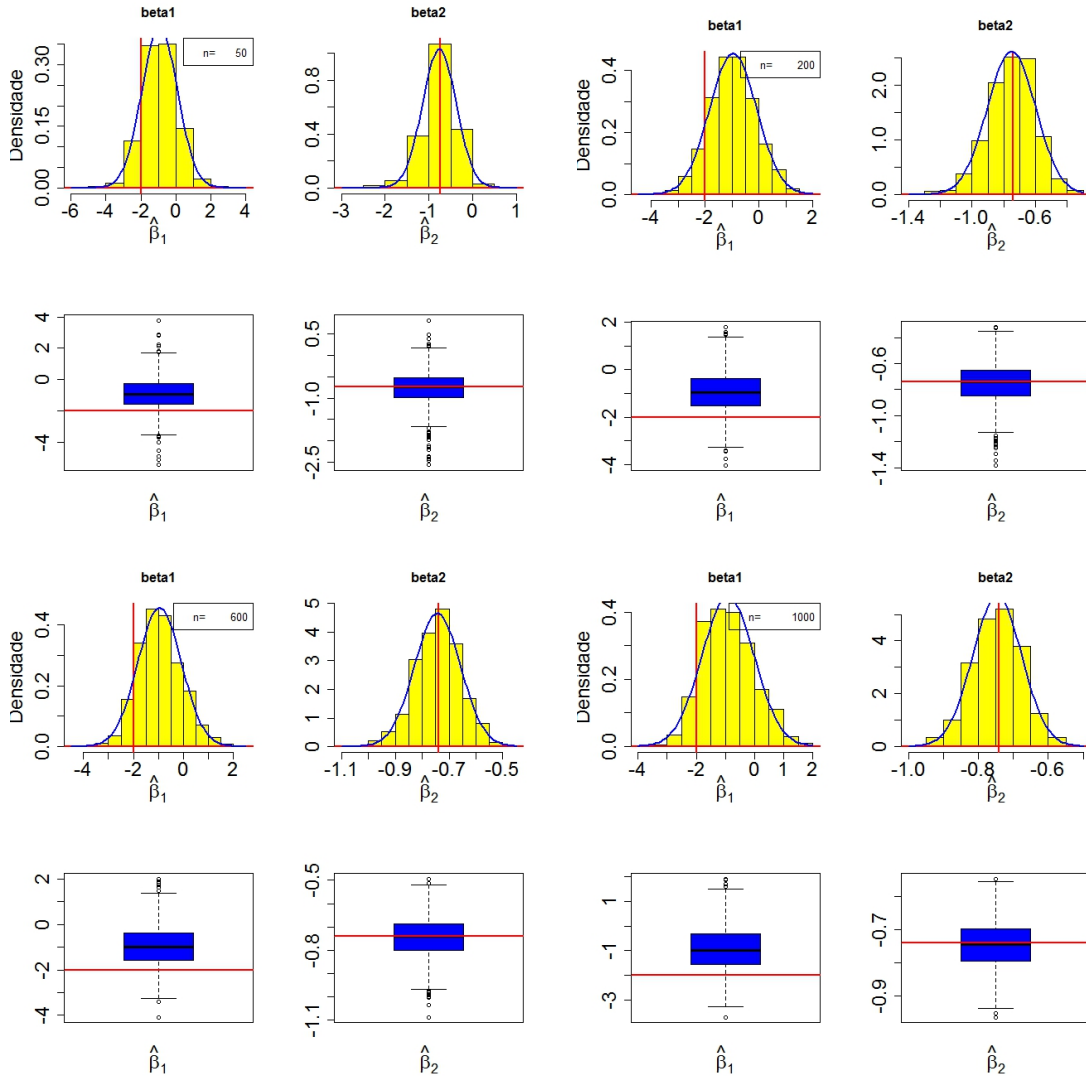


Figura C.7: Boxplots e histogramas para o Cenário 1 do método Clássico.

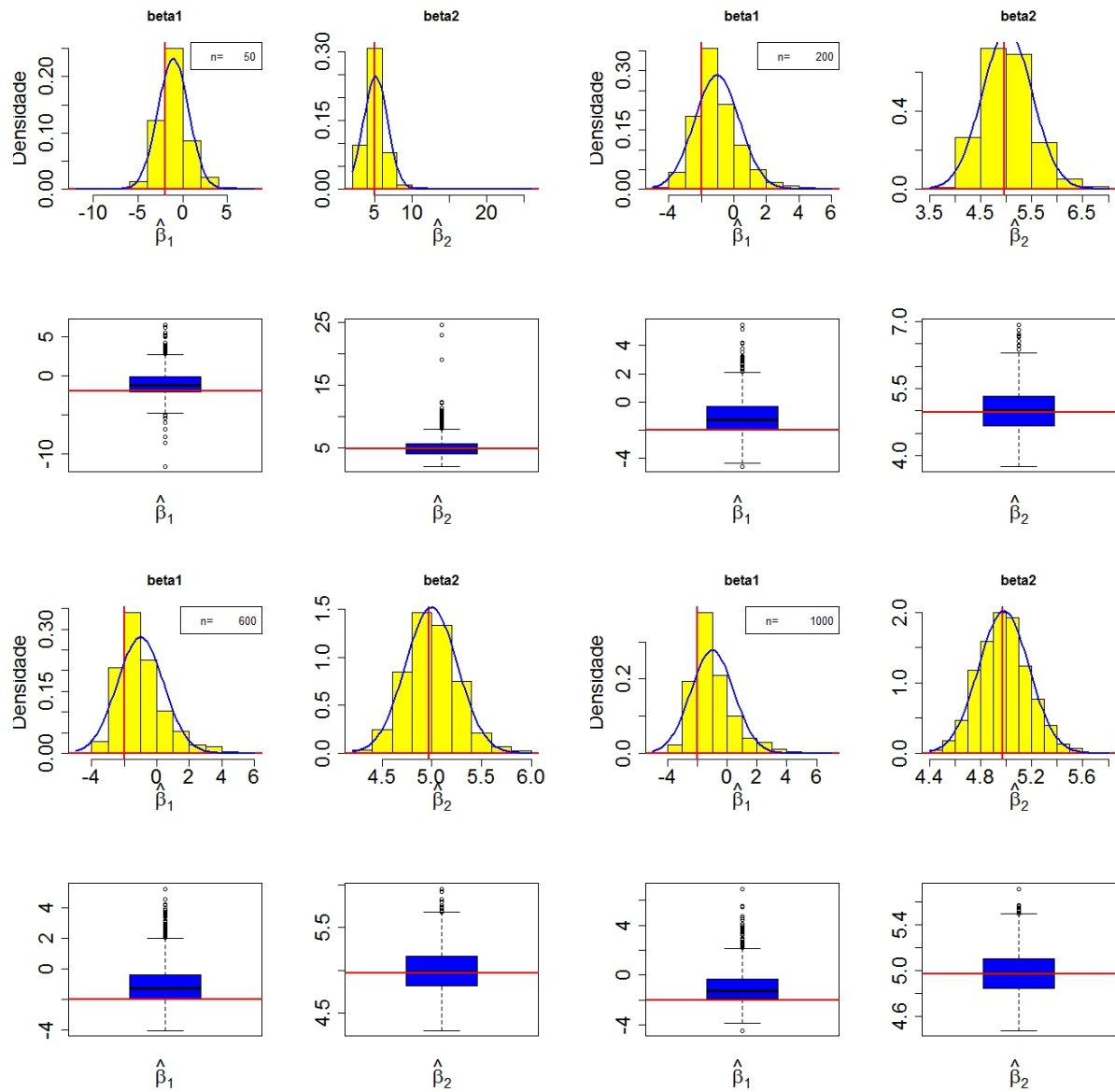


Figura C.8: *Boxplots e histogramas para o Cenário 2 do método Clássico.*

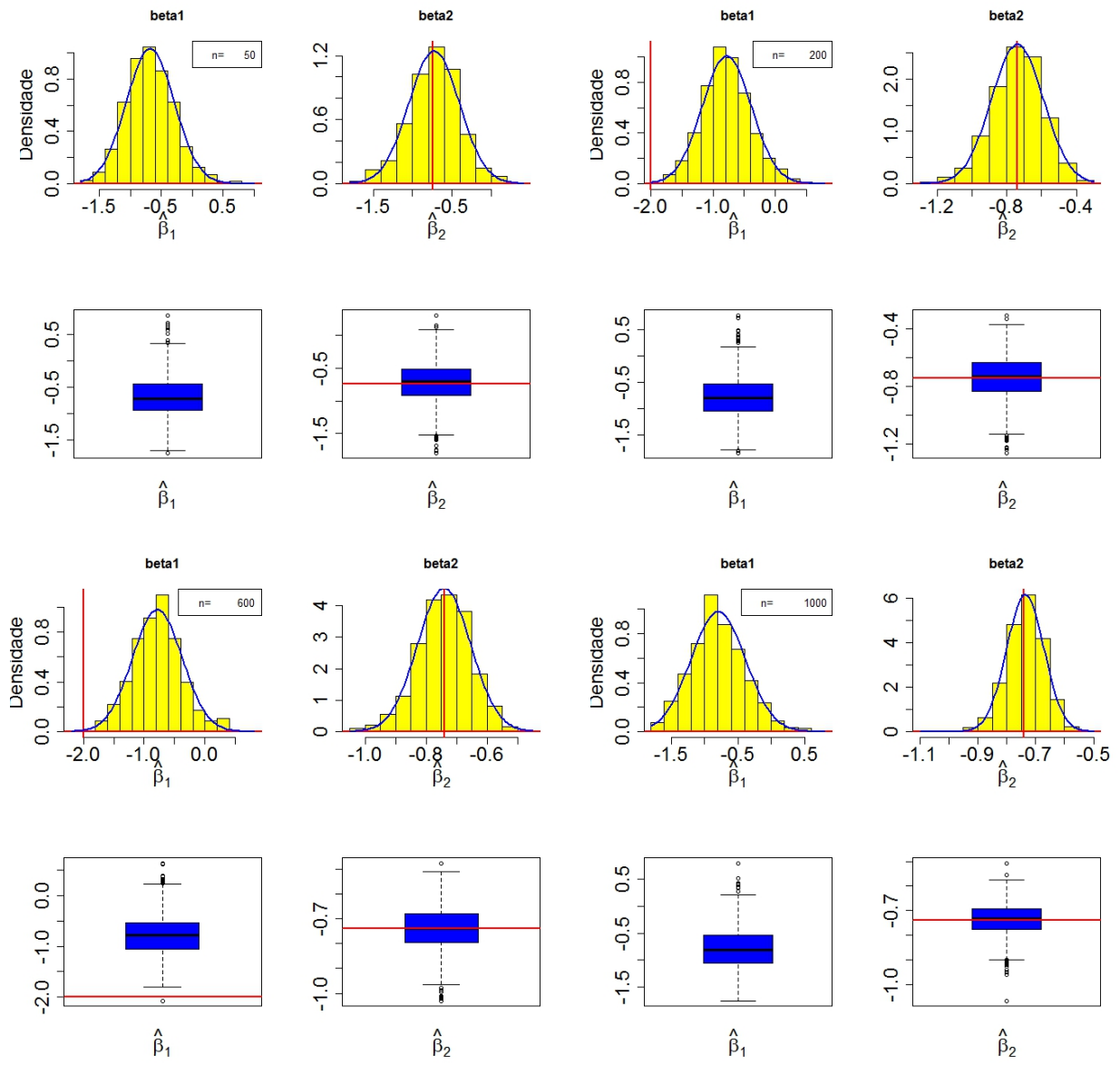


Figura C.9: *Boxplots e histogramas para o Caso 1 do método Bayesiano.*

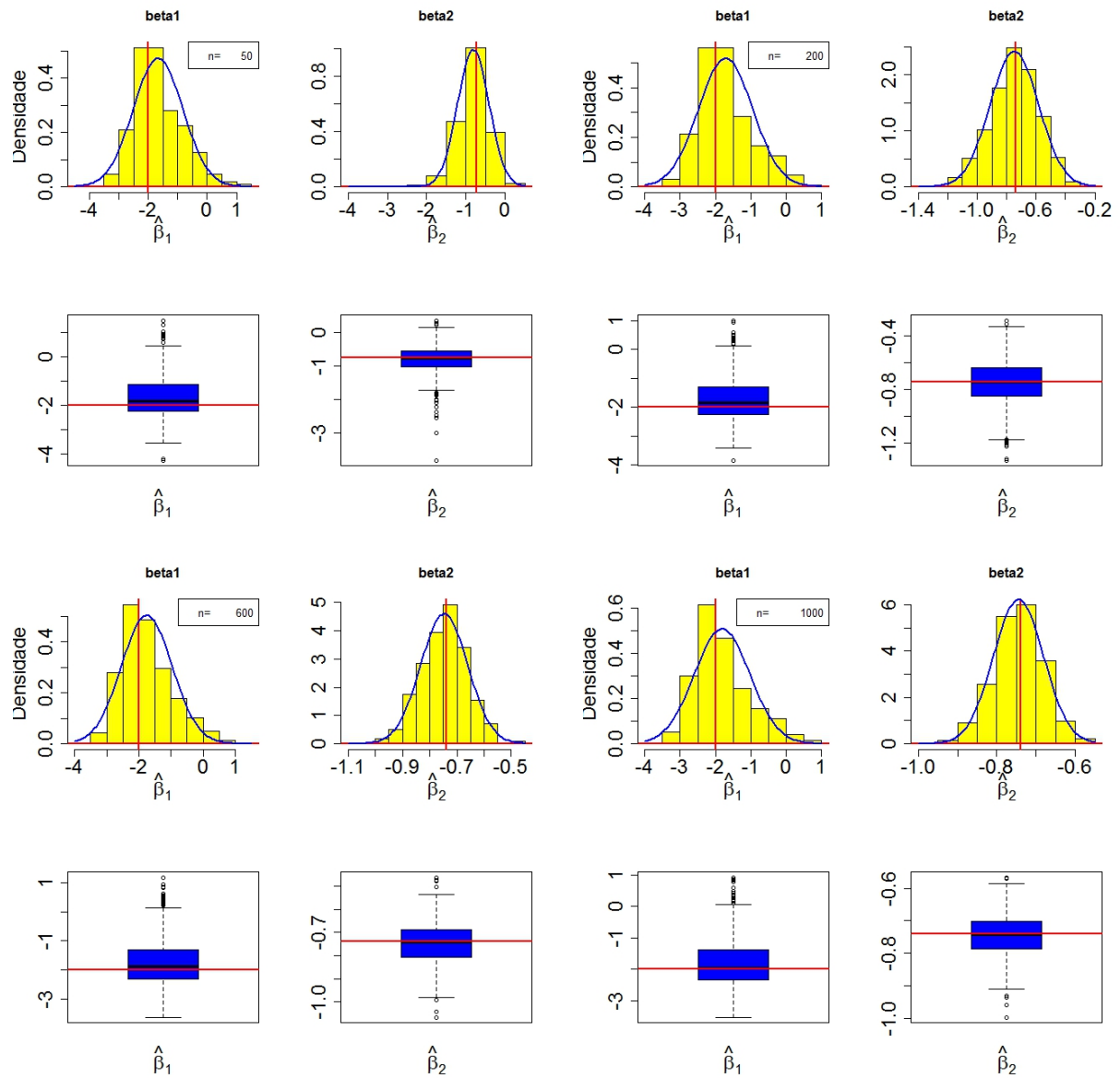


Figura C.10: *Boxplots e histogramas para o Caso 2 do método Bayesiano.*

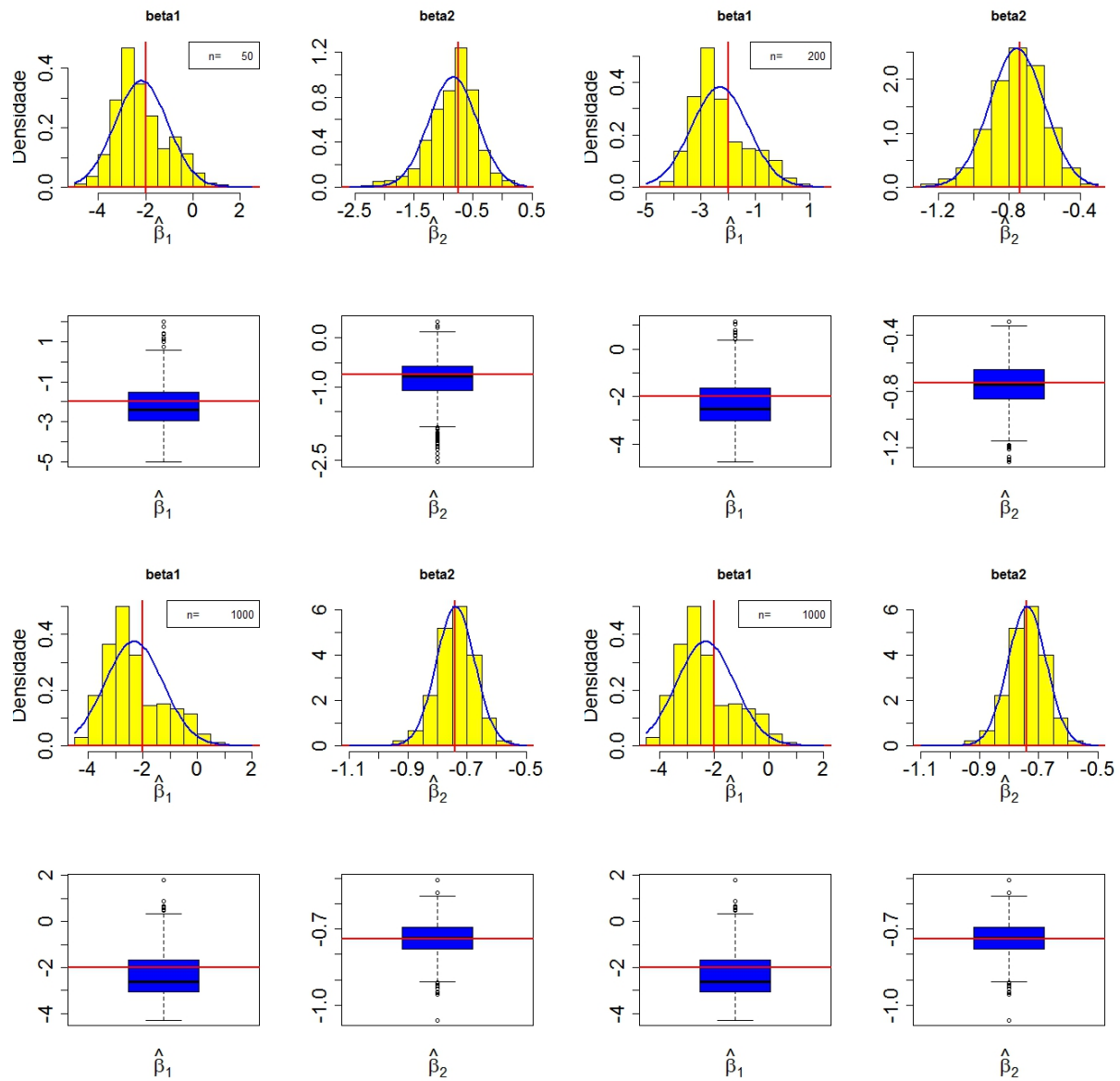


Figura C.11: *Boxplots e histogramas para o Caso 3 do método Bayesiano.*

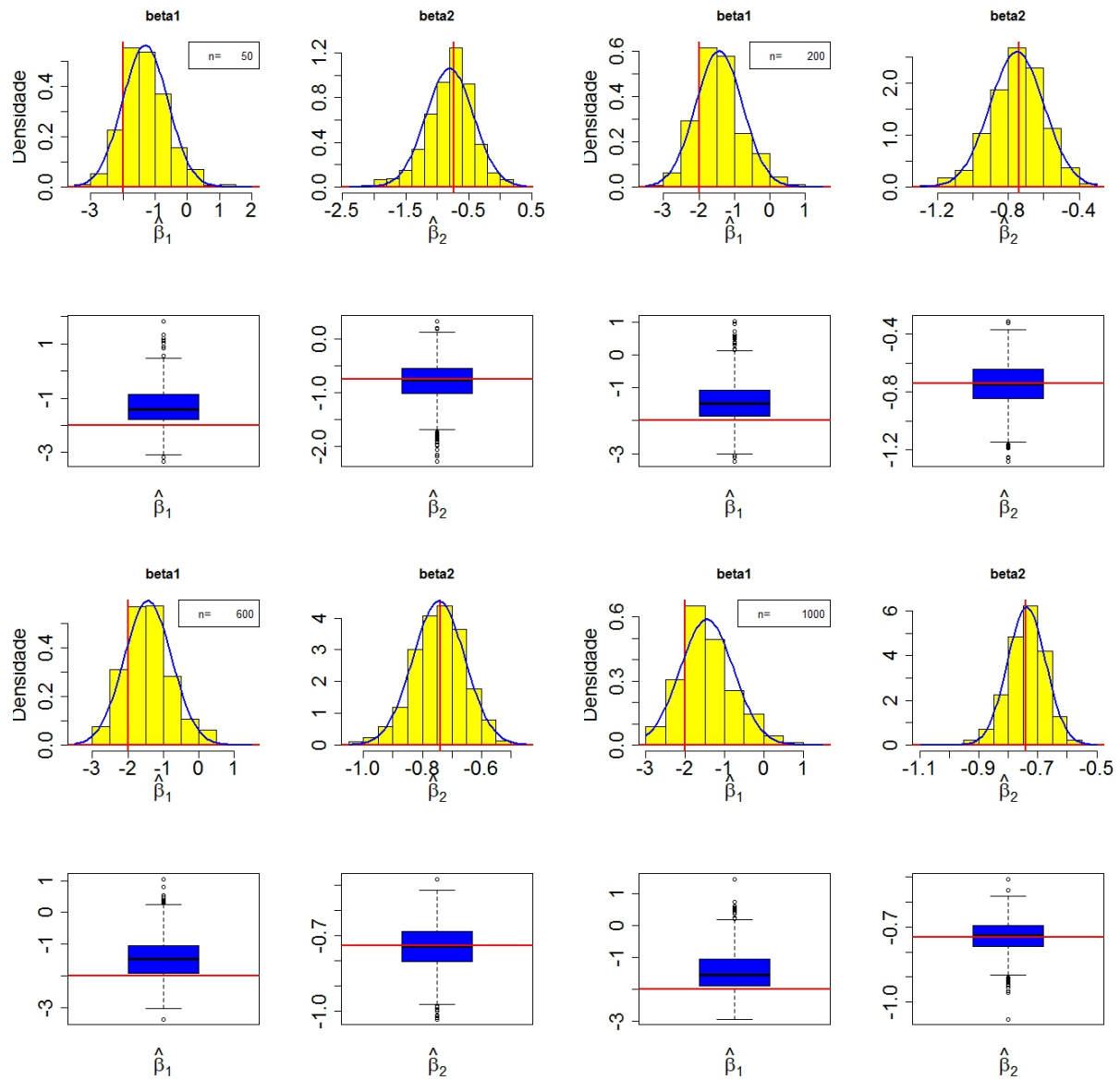


Figura C.12: *Boxplots e histogramas para o Caso 4 do método Bayesiano.*

Apêndice D: Resultados adicionais relacionados aos dados de melanoma

As tabelas a seguir investigam intervalos de credibilidade e apresentamos, logo depois, alguns gráficos exploratórios.

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|-------------|-----------|----------|
| N(0,1) | 0.12099261 | 2.840334 | 1.429305 |
| N(0,5) | 0.38586034 | 5.606414 | 2.706946 |
| log- $F(1, 1)$ | -0.32919201 | 8.196091 | 3.584699 |
| log- $F(2, 2)$ | -0.05645063 | 5.174841 | 2.369354 |
| log- $F(9, 9)$ | 0.02432665 | 2.257186 | 1.090222 |

Tabela 9: $IC_{95\%}$ para β_1 .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|------------|-----------|----------|
| N(0,1) | -1.2423462 | 0.0656280 | -0.59535 |
| N(0,5) | -1.2975625 | 0.0994444 | -0.63252 |
| log- $F(1, 1)$ | -1.2920473 | 0.0724753 | -0.63783 |
| log- $F(2, 2)$ | -1.3234233 | 0.0222132 | -0.62521 |
| log- $F(9, 9)$ | -1.1363448 | 0.1184297 | -0.53278 |

Tabela 10: $IC_{95\%}$ para β_2 .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|-----------|-----------|----------|
| N(0,1) | 0.2856954 | 1.7943540 | 1.061344 |
| N(0,5) | 0.1256627 | 1.8968989 | 1.048561 |
| log- $F(1, 1)$ | 0.1498070 | 1.9241898 | 1.069327 |
| log- $F(2, 2)$ | 0.1926840 | 1.8599387 | 1.044058 |
| log- $F(9, 9)$ | 0.3334411 | 1.7388682 | 0.998716 |

Tabela 11: $IC_{95\%}$ para β_{31} .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|------------|-----------|----------|
| N(0,1) | -0.9556892 | 1.1745972 | 0.189234 |
| N(0,5) | -1.0355896 | 1.3884391 | 0.213466 |
| log- $F(1, 1)$ | -1.0540135 | 1.3772816 | 0.230052 |
| log- $F(2, 2)$ | -0.8613425 | 1.3676267 | 0.225072 |
| log- $F(9, 9)$ | -0.6881769 | 1.3676267 | 0.141906 |

Tabela 12: $IC_{95\%}$ para β_{32} .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|-----------|-----------|----------|
| N(0,1) | -0.180988 | 1.6068529 | 0.695485 |
| N(0,5) | -0.248949 | 2.2794884 | 1.058796 |
| log- $F(1, 1)$ | -0.150315 | 2.2859193 | 1.008949 |
| log- $F(2, 2)$ | -0.049637 | 2.1393103 | 0.952975 |
| log- $F(9, 9)$ | -0.193598 | 1.3240684 | 0.540839 |

Tabela 13: $IC_{95\%}$ para β_{41} .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|----------------|------------|-----------|----------|
| N(0,1) | 0.48422781 | 2.3811598 | 1.396758 |
| N(0,5) | 0.49338881 | 3.0870034 | 1.785627 |
| log- $F(1, 1)$ | 0.49959656 | 3.0207637 | 1.733056 |
| log- $F(2, 2)$ | 0.52529098 | 2.8801831 | 1.677106 |
| log- $F(9, 9)$ | 0.37299237 | 2.0311825 | 1.217437 |

Tabela 14: $IC_{95\%}$ para β_{42} .

| <i>Priori</i> | L_{inf} | L_{sup} | Média |
|---------------|------------|-----------|-----------|
| N(0,1) | 0.0555745 | 1.454729 | 0.7989417 |
| N(0,5) | 0.0600994 | 1.567776 | 0.7615555 |
| log-F(1, 1) | -0.0339157 | 1.457956 | 0.7539663 |
| log-F(2, 2) | -0.0685803 | 1.441123 | 0.7690262 |
| log-F(9, 9) | 0.1030674 | 1.451114 | 0.7861393 |

Tabela 15: $IC_{95\%}$ para β_5 .

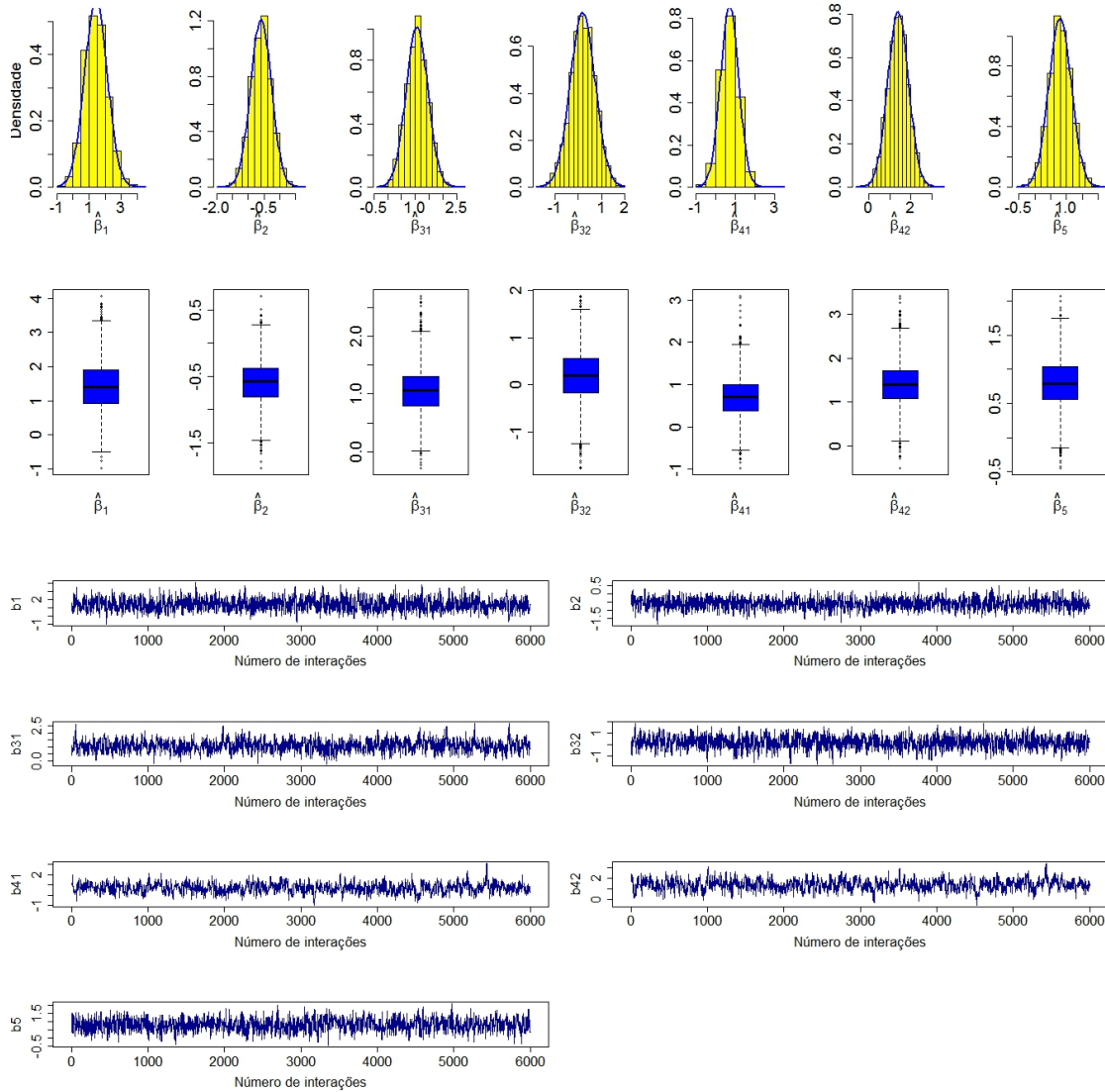


Figura D.1: Gráfico das cadeias e boxplots para a distribuição a priori do Caso 1.

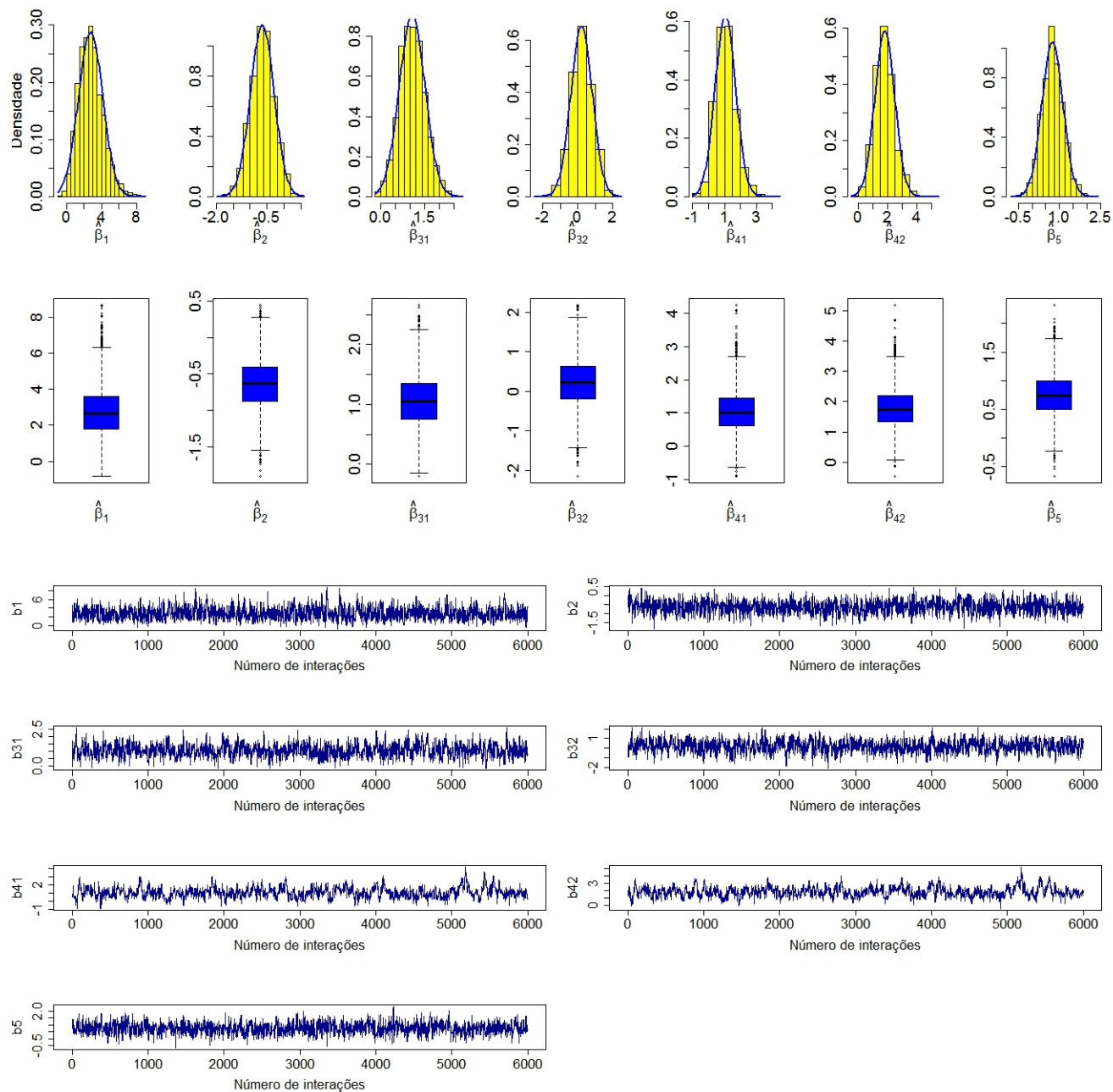


Figura D.2: Gráfico das cadeias e boxplots para a distribuição a priori do Caso 2.

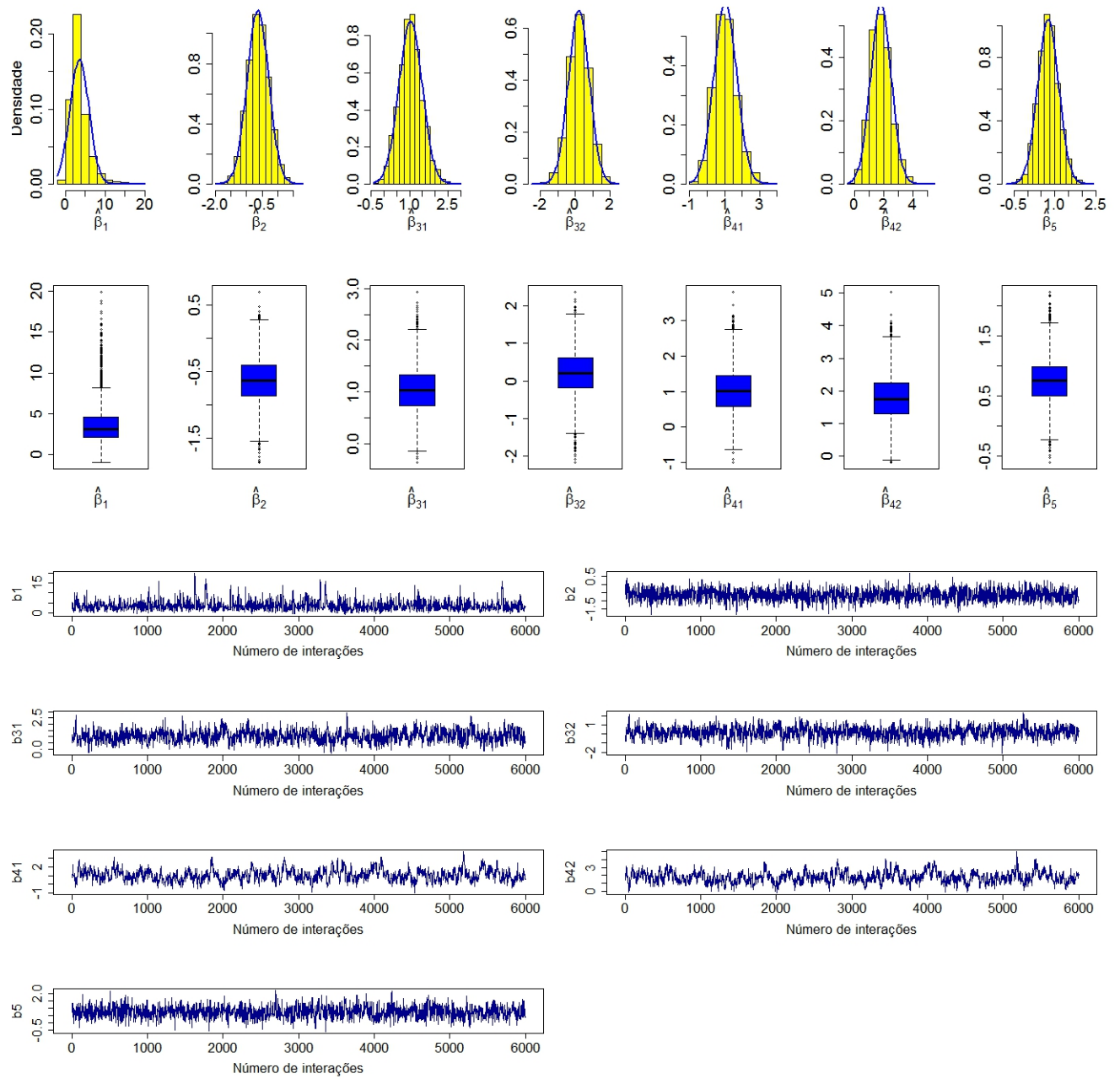


Figura D.3: Gráfico das cadeias e boxplots para a distribuição a priori do Caso 3.

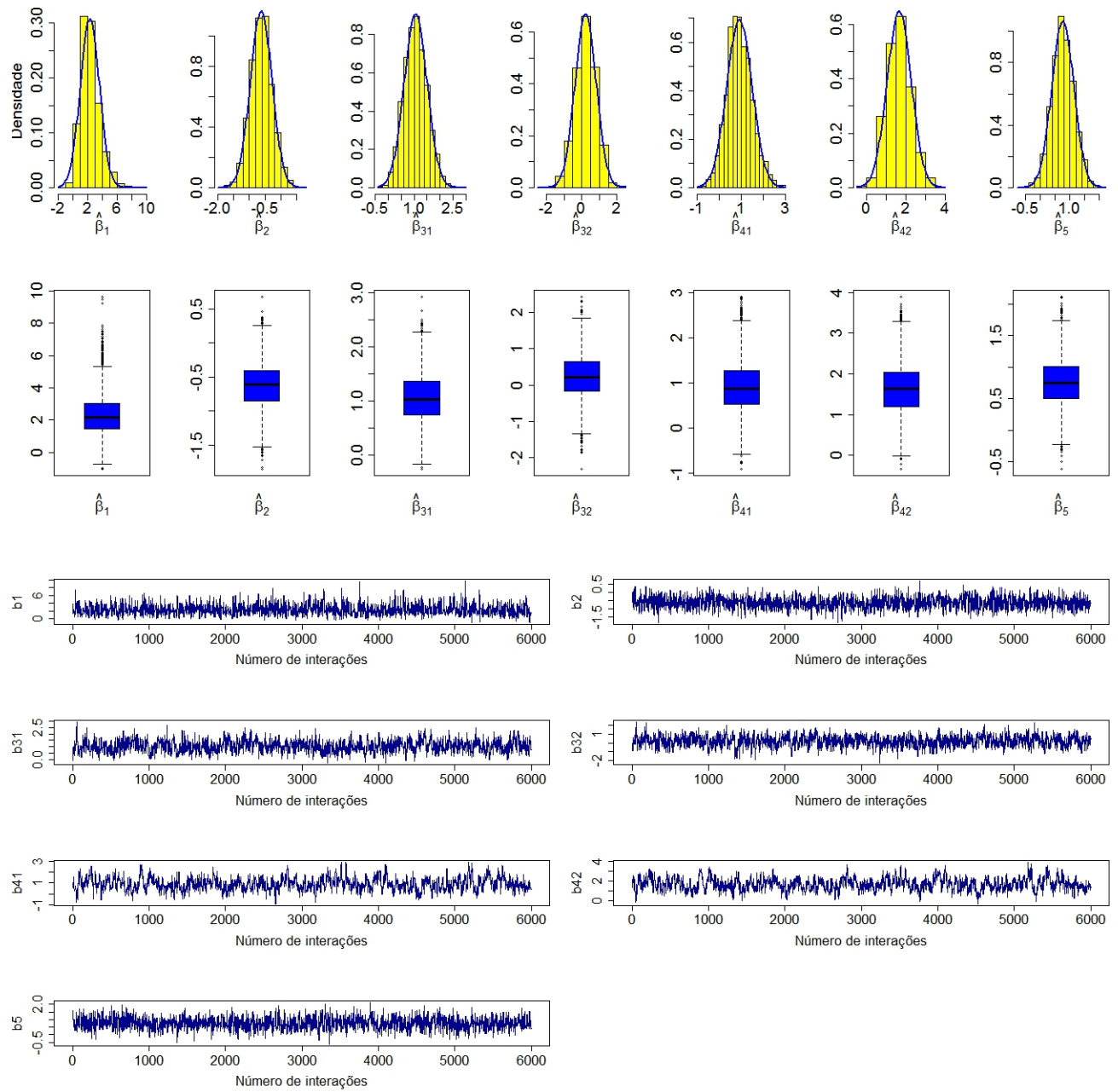


Figura D.4: Gráfico das cadeias e boxplots para a distribuição a priori do Caso 4.

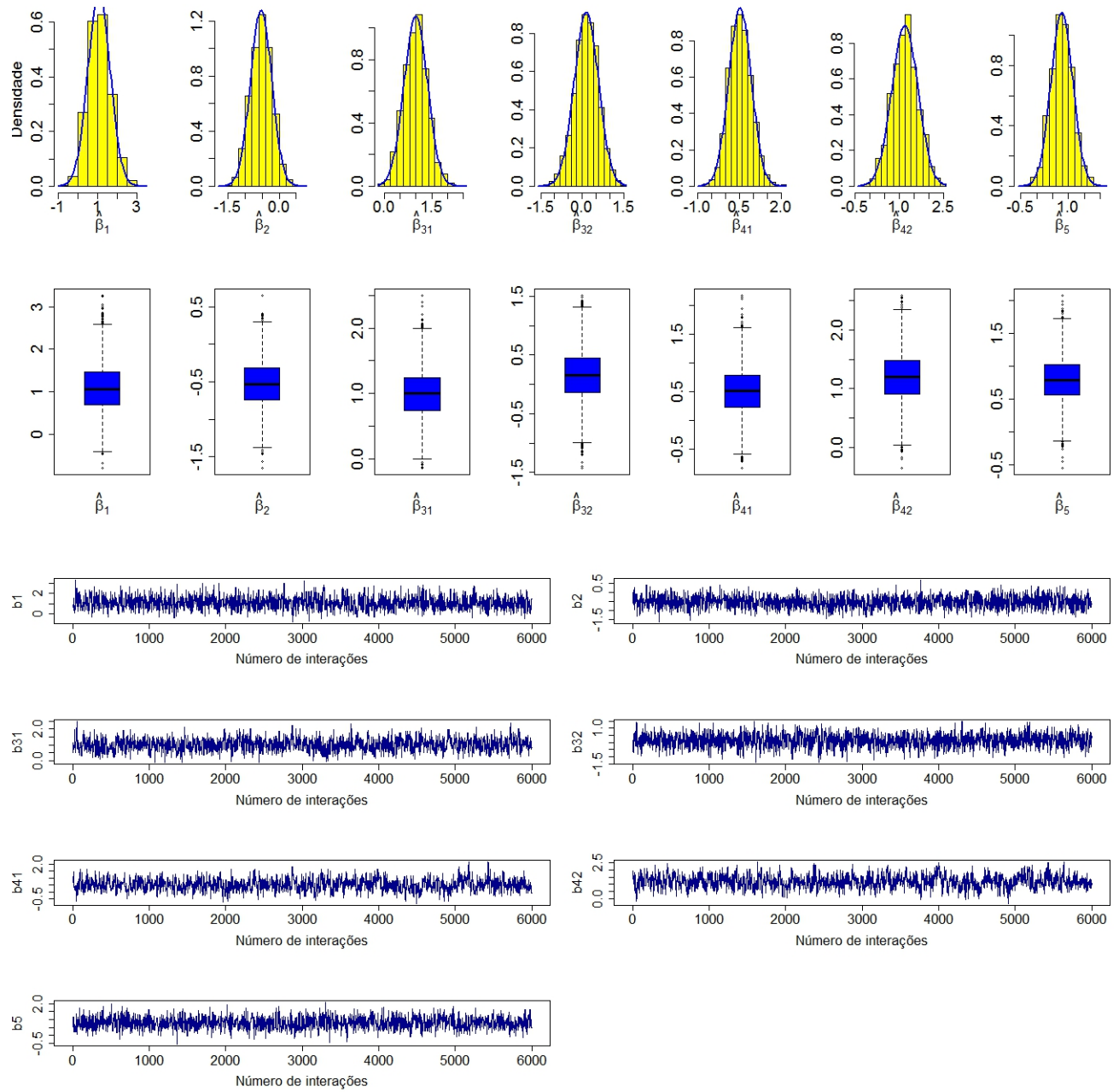


Figura D.5: Gráfico das cadeias e boxplots para a distribuição a priori $\log-F(9,9)$.

Bibliografia

- Amos, D. E. (1983), “A portable Fortran Subroutine for Derivatives of the Psi Function,” *Mathematics and Computing*, 9, 494–502.
- Arce, P. M., Camilon, P. R., Stokes, W. A., Nguyen, S. A., e Lentsch, E. J. (2014), “Is Sex an Independent Prognostic Factor in Cutaneous Head and Neck Melanoma,” *The American Laryngological, Rhinological and Otological Society*, 124, 1363–1367.
- Brayson, M. C. e Johnson, M. E. (1981), “The Incidence of Monotone Likelihood in the Cox Model,” *Technometrics*, 23, 381–383.
- Brayson, M. C., Johnson, M. E., e Tolley, H. D. (1982), “Covariate Analysis of Survival Data: A Small-Sample Study of Cox Model,” *Biometrics*, 38, 685–698.
- Collett, D. (2003), *Modelling Survival Data in Medical Research*, Chapman and Hall/CRC, London, 2 edn.
- Colosimo, E. A. e Giolo, S. R. (2006), *Análise de Sobrevida Aplicada*, ABE-Projeto Fisher, Brazil.
- Cordeiro, G. M. e Cribari-Neto, F. (2014), *An Introduction to Bartlett Correction and Bias Reduction*, Springer, New York, 1 edn.
- Cox, D. R. (1972), “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society, Series B*, 34, 187–220.
- Cox, D. R. e Hinkley, D. V. (1974), *Theoretical Statistics*, Springer, New York, 1 edn.

- Damato, B., Eleuteri, A., Taktak, A. F. G., e Coupland, S. E. (2011), “Estimating prognosis for survival after treatment of choroidal melanoma,” *Elsevier*, 30, 1–11.
- Firth, D. (1993), “Bias Reduction of Maximum Likelihood Estimates,” *Biometrika*, 80, 27–38.
- Gamerman, D. e Lopes, H. F. (2006), *Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference*, Texts in Statistical Science, Chapman and Hall/CRC, London, 2 edn.
- Gamerman, D., Mingon, H. S., e Louzada, F. (2014), *Statistical Inference, An Integrated Approach*, Chapman and Hall/CRC, New York, 2 edn.
- Gelfand, A. E. e Smith, A. F. M. (1990), “Sampling-Based Approaches to Calculating Marginal Densities,” *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., Calin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., e Rubin, D. B. (2014), *Bayesian Data Analysis*, Chapman and Hall/CRC, New York, 3 edn.
- Geman, S. e Geman, D. (1984), “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of imagens,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Greenland, S. e Mansournia, M. A. (2015), “Penalization, bias Reduction and default priors in logistic and related categorical and survival regression,” *Statistics in Medicine in Medicine*, 34, 3133–3143.
- Hastings, W. (1970), “Monte Carlo Sampling Using Markov Chains and Their Applications,” *Biometrika*, 57, 97–109.
- Heinze, G. e Ploner, M. (2015), *coxphf: Cox regression with Firth’s penalized likelihood*, R package version 1.11.
- Heinze, G. e Schemper, M. (2001), “A Solution to the Problem of Monotone Likelihood in Cox Regression,” *Biometrics*, 57, 114–119.

- Ibrahim, J. G., Sinha, D., e Chen, M. H. (2001), *Bayesian Survival Analysis*, Series in Statistics, Springer, New York, 1 edn.
- Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems,” *Royal Society of London*, 186, 453–461.
- Kaplan, E. L. e Meier, P. (1958), “Nonparametric Estimation from Incomplete Observations,” *Journal of the American Statistical Association*, 53, 457–481.
- Klein, J. P. e Moeschberger, M. L. (2003), *Survival Analysis, Techniques for Censored and Truncated Data*, Springer, New York, 2 edn.
- Kleinbaum, D. G. e Klein, M. (2012), *Survival Analysis*, Springer, New York, 3 edn.
- Kosmidis, I. (2014), “Bias in parametric estimation: reduction and useful side-effects,” *Wiley Interdisciplinary Reviews: Computational Statistics*, 6, 185–196.
- Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Probability and Statistics, Wiley Interscience, 2 edn.
- Lee, E. T. e Wang, J. W. (2003), *Statistical Methods for Survival Data Analysis*, Probability and Statistics, Wiley Interscience, New Jersey, 3 edn.
- Lin, I. F., Chang, W. P., e Liao, Y. N. (2013), “Shrinkage Methods enhanced the accuracy of parameter estimation using Cox models with small number of events,” *Clinical Epidemiology*, 66, 743–751.
- Metropolis, N., Rosenbluth, A., e Teller, M. (1953), “Equations of State Calculations by Fast Computing Machines,” *Journal of Chemistry and Physics*, 21, 1087–1091.
- Ojo, M. O. e Olapade, A. K. (2003), “On the Generalized Logistic and Log-Logistic Distribution,” *Kragujevac Journal of Math*, 25, 65–73.
- R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Roberts, G. O., Gelman, A., e Gilks, W. R. (1997), “Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms,” *The Annals of Applied Probability*, 7, 110–120.

Sinha, D., Ibrahim, J. G., e Chen, M. H. (2003), “A Bayesian justification of Cox partial likelihood,” *Biometrika*, 90, 629–641.

Therneau, T. e Grambsch, P. (2000), *Modeling Survival Data, Extending the Cox Model*, Statistics for Biology and Health, Springer, New York, 1 edn.