Arlaine Aparecida da Silva

Análise de Sobrevivência com Erros de Classificação

Belo Horizonte, março de 2012

Arlaine Aparecida da Silva

Análise de Sobrevivência com Erros de Classificação

Dissertação apresentada ao Programa de Pós Graduação
em Estatística da UFMG como requisito parcial
para obtenção de grau de Mestre em Estatística
pela Universidade Federal de Minas Gerais.

Orientadora: Prof.ª Dr.ª Magda Carvalho Pires
Co-Orientador: Prof. Dr. Enrico Antônio Colosimo

Belo Horizonte, março de 2012

Agradecimentos

À professora Magda C. Pires, minha orientadora, pelo comprometimento e dedicação a este trabalho, pelo incentivo, paciência, por todos os ensinamentos passados e por ter me dado a oportunidade de trabalharmos juntas.

Ao professor Enrico A. Colosimo, meu co-orientador, por todo incentivo, pelos conhecimentos transmitidos e pela oportunidade de desenvolver com ele esta dissertação.

Aos professores e funcionários do Departamento de Estatística da UFMG pela oportunidade e por terem contribuído na minha formação.

Aos meus colegas e amigos da Pós-Graduação em Estatística, em especial ao Fernando Henrique, que me encorajou a participar do programa, por sua amizade e companheirismo desde a graduação.

À minha família, ao meu irmão Arlei que me incentivou a voltar a estudar e a minha filha por toda força e apoio nos momentos difíceis.

Acima de tudo, agradeço a Deus por mais esta conquista.

Resumo

A análise de dados de sobrevivência consiste no estudo do tempo até a ocorrência de um evento de interesse como, por exemplo, a morte de um paciente, cura ou recidiva de uma doença. Quando o tempo exato de ocorrência não é conhecido, mas sabe-se que ele aconteceu no intervalo entre duas avaliações consecutivas do indivíduo, estamos diante de um estudo com censura intervalar. Entretanto, a detecção do evento depende da qualidade dos testes aplicados, pois estes podem estar sujeitos a erros de classificação: um indivíduo pode ser diagnosticado como doente quando na verdade ele está sadio ou um indivíduo doente pode ser diagnosticado como sadio. Nesses casos, ao utilizar métodos tradicionais de Análise de Sobrevivência, estimativas viciadas para os parâmetros da distribuição do tempo de falha são obtidas [Paggiaro e Torelli (2004)]. Apresentamos, então, uma proposta que incorpora a sensibilidade e a especificidade do teste ao modelo de análise de sobrevivência com dados grupados (caso especial de censura intervalar em que todas as unidades são avaliadas nos mesmos instantes). Estudos de simulação Monte Carlo demonstraram que, quando a sensibilidade e a especificidade do teste são conhecidas, o método proposto é bastante eficiente, pois suas estimativas apresentam menor vício relativo do que aquelas fornecidas pelo método tradicional.

Abstract

Survival data analysis is concerned with the study of time until the occurrence of an event of interest, such as the death of a patient, the cure or the recurrence of a disease. If the exact time of the event of interest is not known, but instead, the event is known to have occurred during a particular interval of time, the data are known as interval-censored survival data. However, if the diagnostic tool used to detect failure is not perfectly sensitive and specific, the subjects may be misclassified: a healthy one may be diagnosed as sick and a sick individual may be diagnosed as healthy as well. In such cases, the traditional survival analysis methods produce biased estimates for the failure time distribution parameters [Paggiaro and Torelli (2004)]. So we developed a model that incorporates sensitivity and specificity in grouped survival data analysis (a special case of interval- censored data in which all subjects are tested at predetermined time points). Monte Carlo simulation studies have shown that, if sensitivity and specificity are known, the proposed method is very efficient, since its estimates percent bias are lower than those provided by traditional method.

Sumário

Capítulo 1

Introdução

1.1 Análise de sobrevivência

Estudos de sobrevivência analisam o tempo até a ocorrência de um evento de interesse (tempo de falha) podendo ser, por exemplo, o tempo até a morte, o tempo até a cura ou até a recidiva de uma doença. Os dados são caracterizados pela presença de observação parcial da resposta (censura) que ocorre quando o acompanhamento do paciente foi interrompido por alguma razão. Considere, por exemplo, o caso em que se deseja comparar o tempo médio do surgimento de determinada doença entre g grupos. Quando não existem censuras, a comparação pode ser realizada com técnicas usuais de análise de variância. No caso em que existem censuras, faz-se necessário o uso de métodos de análise de sobrevivência que possibilitam incorporar na análise estatística a informação contida nos dados censurados (Colosimo e Giolo, 2006).

A presença de censura à direita indica que o tempo de falha é superior ao tempo observado. Existem três tipos de censura à direita:

- Censura do tipo I - Ocorre quando o estudo termina após um período de tempo pré-estabelecido.
- Censura do tipo II - Ocorre quando o estudo termina após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos.
- Censura aleatória - Ocorre quando um indivíduo sai do estudo sem ter ocorrido a falha por razões aleatórias.

Quando não se sabe o tempo exato de ocorrência do evento de interesse, mas sabe-se que ele ocorreu dentro de um intervalo, a censura é denominada censura intervalar. A censura intervalar acontece, por exemplo, em estudos nos quais pacientes são acompanhados em visitas periódicas e sabe-se somente que o evento de interesse ocorreu entre uma visita e outra. Por este fato, os dados são denominados dados de censura intervalar. No capítulo 2, trataremos desse tipo de censura com mais detalhes.

1.1.1 Função de sobrevivência e função de taxa de falha

Seja T a variável aleatória contínua que representa o tempo de falha. Em análise de sobrevivência, T pode ser especificada pela sua função de sobrevivência S(t), definida como a probabilidade de uma observação não falhar até o instante de tempo t, ou seja, a probabilidade de uma observação sobreviver ao tempo t:

$$S(t) = P(T > t).$$

Como descrito em Colosimo e Giolo (2006), a taxa de falha no intervalo [t, t+ Δ t) é definida como a probabilidade da falha ocorrer em um intervalo de tempo [t, t+ Δ t) dado que não ocorreu antes de t, dividida pelo comprimento do intervalo, podendo ser expressa como:

$$\begin{aligned} \lambda(t) &= \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} \\ \lambda(t) &= \frac{P(t < T < t + \Delta t)}{P(T > t)\Delta t} \\ \lambda(t) &= \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}. \end{aligned}$$

Se fizermos Δt bem pequeno, $\lambda(t)$ será a função de taxa de falha instantânea no tempo t condicionada à sobrevivência até t. Assim, a função de taxa de falha é dada por:

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

Para estimarmos a função de sobrevivência, podemos utilizar métodos paramétricos e métodos não paramétricos. Os estimadores paramétricos são construídos a partir de modelos probabilísticos através da função de distribuição acumulada, $F(t)$, pois

$$S(t) = 1 - F(t).$$

Nas próximas seções apresentaremos os principais modelos probabilísticos utilizados para descrever o tempo até a falha.

1.1.2 Modelos paramétricos

Distribuição Exponencial

É o modelo mais simples, matematicamente, para descrever o tempo de falha. Apresenta um único parâmetro e tem função de taxa de falha constante. As funções de sobrevivência e de taxa de falha são dadas, respectivamente, por:

$$S(t) = e^{-\alpha t}$$

e

$$\lambda(t) = \alpha,$$

em que $\alpha \geq 0$. Podemos obter a média $(E(T))$, a mediana $(t_{0,5})$ e a variância do tempo de falha $(V(T))$:

$$E(T) = \frac{1}{\alpha},$$

$$t_{0,5} = -\frac{log(0,5)}{\alpha}$$

e

$$Var(T) = \frac{1}{\alpha^2}.$$

Distribuição Weibull

Esta modelagem vem sendo frequentemente utilizada em estudos biomédicos e industriais. Sua vasta aplicação prática se deve ao fato de apresentar uma grande variedade de formas com função de taxa de falha monótona, ou seja, ela é crescente, decrescente ou constante.

As funções de sobrevivência e de taxa de falha são dadas por:

$$S(t) = exp\left(-t^{\gamma}\left(\frac{1}{\alpha}\right)^{\gamma}\right)$$

e

$$\lambda(t) = \left(\frac{\gamma}{\alpha^{\gamma}}\right)t^{\gamma-1},$$

para $t \geq 0$, α e $\gamma \geq 0$. Note que a distribuição Exponencial é um caso particular da distribuição Weibull quando $\gamma = 1$.

Podemos obter a média $(E(T))$, a mediana $(t_{0,5})$ e a variância do tempo de falha $(V(T))$, sendo que, para o modelo ajustado pela distribuição Weibull, temos:

$$E(T) = \alpha\Gamma(1 + (1/\gamma)),$$

$$t_{0,5} = \alpha[-log(0,5)]^{\frac{1}{\gamma}}$$

e

$$Var(T) = \alpha^2[\Gamma(1 + 2/\gamma)) - \Gamma((1 + (1/\gamma)))^2.$$

1.1.3 Estimador de Máxima Verossimilhança

Considere uma amostra de tamanho N e seja ρ o vetor de parâmetros da distribuição do tempo de falha. A função de verossimilhança, quando não houver censura, é dada por:

$$L(t) = \prod_{i=1}^{N} f(t_i; \rho).$$

Percebe-se, portanto, que a função de verossimilhança é construída a partir da contribuição de cada observação da amostra não censurada, sendo esta contribuição a sua função de densidade.

Quando há censura não informativa e estamos interessados em estimar os parâmetros da função de sobrevivência, além da contribuição dos dados não censurados, temos a contribuição de cada observação censurada. Estas observações informam que o tempo de falha é maior que o tempo de censura observado caracterizando como censura à direita e, portanto, sua contribuição será a sua função de sobrevivência. Suponnha que em uma amostra de tamanho N tivemos que as primeiras r observações falharam e as $N - r$ observações restantes não apresentaram falha (foram censuradas). Desta forma a amostra pode ser dividida em duas partes onde as r primeiras ordenadas são as não censuradas e as $N - r$ restantes serão as censuradas. Assim, a função de verossimilhança é dada por:

$$L(\rho) \propto \prod_{i=1}^{r} f(t_i; \rho) \prod_{r+1}^{N} S(t_i; \rho),$$

em que r é o número de falhas. A função de verossimilhança acima pode ser escrita da seguinte maneira:

$$L(\rho) \propto \prod_{i=1}^{N} [f(t_i; \rho)]^{d_i} [S(t_i; \rho)]^{1-d_i},$$

em que d_i é o indicador de falha igual a 1 se ocorreu falha no tempo t_i e igual a 0 se ocorreu censura no tempo t_i.

1.2 Erros de classificação na detecção do Evento

Quando o interesse é estudar a prevalência de determinada doença é preciso avaliar sua presença através de testes de diagnósticos que podem estar sujeitos a erros de classificação. Os erros de classificação mais comuns, neste contexto, são classificar um indivíduo como doente quando na verdade ele está sadio ou classificar um indivíduo como sadio quando na verdade ele está doente. Podemos então quantificar a qualidade dos testes através de duas medidas: a sensibilidade e a especificidade, θ e ϕ, respectivamente.

A sensibilidade é a probabilidade de o teste identificar corretamente a doença entre aqueles que a possuem, ou seja, avalia o quão sensível é o teste. A especificidade é a probabilidade de o teste excluir corretamente aqueles que não possuem a doença, ou seja, avalia o quão específico o teste é. Desta forma, podemos escrever a sensibilidade e a especificidade, respectivamente, como:

$$\theta = P(d_i^o = 1|d_i = 1)$$

e

$$\phi = P(d_i^o = 0|d_i = 0),$$

em que d_i^o é a indicadora para o resultado observado do teste ($d_i^o = 0$ se o teste foi negativo e $d_i^o = 1$ se o teste foi positivo) e d_i é a indicadora para a ocorrência verdadeira do evento ($d_i = 0$ se o evento não ocorreu e $d_i = 1$ se o evento ocorreu). Um bom teste possui um alto valor para a sensibilidade e para a especificidade, pois ele identifica corretamente aqueles que têm a doença e aqueles que não têm.

Quando estudamos a prevalência, p, de uma doença em uma amostra de tamanho N e os testes de diagnósticos não estão sujeitos a erros, ou seja, a sensibilidade e a especificidade são iguais a 1, o número de doentes, X, na amostra tem distribuição binomial com parâmetros N e p, em que $p = P(d_i = 1)$. Deste modo,

$$P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}.$$

Nos casos em que os testes de diagnósticos estão sujeitos a erros de classificação, a distribuição do número de doentes será *Bin(n, q)* em que q é a probabilidade do teste ser positivo dada por:

$$\begin{aligned} q &= P(d_i^o = 1) = P(d_i^o = 1|d_i = 1)P(d_i = 1) + P(d_i^o = 1|d_i = 0)P(d_i = 0) \\ q &= \theta p + (1-\phi)(1-p). \end{aligned} \tag{1.1}$$

Quando a sensibilidade e a especificidade são iguais a 1, temos:

$$q = P(d_i^o = 1) = p = P(d_i = 1).$$

Esse problema de erros de classificação foi primeiramente abordado por Bross (1954). Johnson $et.al.$ (1991) fazem uma revisão dos métodos clássicos propostos. Alguns métodos bayesianos foram propostos por Gaba e Winkler (1990) e Joseph $et.al.$(1995), por exemplo.

No contexto de análise de sobrevivência, estamos interessados em analisar o tempo até a ocorrência da doença, cuja detecção é realizada através de exames clínicos e laboratoriais que estão sujeitos a erros de classificação. Ignorar estes erros pode acarretar em inferências viesadas (Paggiaro e Torelli, 2004) e, por isso, é necessário utilizar um modelo de análise de sobrevivência que incorpore os erros de classificação em sua estrutura e forneça estimativas não viciadas para os parâmetros de interesse.

1.3 Objetivos

No contexto de Análise de Sobrevivência para dados com censura intervalar e eventos não catastróficos que ocorrem frequentemente mas sem efeitos muito graves (exemplo: doença passageira), pretende-se:

- ilustrar o efeito provocado pelos erros de classificação na estimação dos parâmetros da distribuição do tempo de falha e revisar modelos propostos na literatura para lidar com esse problema;

- propor e avaliar um modelo geral capaz de incorporar os erros de classificação na análise univariada de dados;

- estender a proposta para o Modelo de Riscos Proporcionais Discretos.

1.4 Organização do trabalho

No Capítulo 2 revisamos a análise de dados de sobrevivência grupados e o Modelo de Riscos Proporcionais Discretos (Collett, 2003). No Capítulo 3 ilustramos uma situação em que estimativas viciadas para o parâmetro da distribuição do tempo de falha são obtidas devido à presença de erros de classificação. Além disso, fazemos uma breve revisão dos modelos presentes na literatura para lidar com o problema, em especial o modelo proposto por Meier et al. (2003). No Capítulo 4 propomos um novo modelo univariado para casos mais gerais que os tratados por Meier et al. (2003) e avaliamos seu desempenho através de simulações Monte Carlo. No Capítulo 5, apresentamos e avaliamos o modelo proposto na presença de covariáveis . Uma aplicação desse modelo é apresentada no Capítulo 6. Conclusões e discussões são apresentadas no Capítulo 7.

Capítulo 2

Censura Intervalar e Dados Grupados

Quando os dados de sobrevivência são registrados em intervalos de tempo tem-se respostas com censura intervalar, como citado no capítulo anterior. Esta situação aparece com frequência em estudos clínicos longitudinais onde a ocorrência do evento de interesse é monitorada em visitas médicas de rotina (Colosimo e Giolo, 2006). Nestes estudos o tempo de falha não é conhecido, sabe-se somente que o evento ocorreu em algum momento dentro do intervalo $(t_{i-1}, t_i]$.

A análise de modelos paramétricos (ou probabilísticos) para dados de sobrevivência intervalar também é de interesse. Sendo assim, após ter especificado um modelo, é preciso estimar seus parâmetros, o que pode ser realizado pelo método de máxima verossimilhança. Na construção da função de verossimilhança, a natureza intervalar dos dados deve ser levada em consideração. A contribuição de um indivíduo que apresente um tempo de falha em um certo intervalo é dada pela probabilidade de que o tempo de ocorrência do evento pertença a este intervalo e a contribuição de um indivíduo censurado é dada pela função de sobrevivência em t_i. Assim, a função de verossimilhança é escrita como:

$$L(\rho) = \prod_{i=1}^{N} [S(t_{i-1};\rho) - S(t_i;\rho)]^{d_i} [S(t_i;\rho)]^{1-d_i}, \tag{2.1}$$

em que ρ é o parâmetro (ou vetor de parâmetros) da distribuição de T que queremos estimar, d_i é a indicadora de ocorrência do evento, N é o tamanho da amostra e $S(t_i)$ é a função de sobrevivência em t_i.

Dados grupados surgem quando todas as unidades amostrais são avaliadas nos mesmos instantes e, por isso, são muitas vezes associados a situações com excesso de empates caracterizando-se, assim, em um caso particular de censura intervalar, ou seja, quando tratamos de dados grupados os tempos de observações são fixos e iguais para todas as unidades. Um procedimento utilizado para análise de dados grupados é considerar a variável aleatória tempo como contínua e assumir que o evento ocorreu no final ou no ponto médio do intervalo. Alguns autores, resaltam que assumir os tempos intervalares como tempos exatos de falha pode conduzir a inferências viciadas. Neste trabalho abordaremos somente casos de dados grupados, assunto que será apresentado na próxima seção.

2.1 Análise de dados grupados

Considere que os tempos de vida são grupados em k intervalos denotados por $I_j = (t_{j-1}, t_j]$, $j = 1, ..., k$, em que $0 = t_0 < t_1 < t_2 < ... < t_k$, e assuma que as censuras ocorrem no final do intervalo. Seja d_i uma variável indicadora para a ocorrência do evento do i-ésimo indivíduo no j-ésimo intervalo de tempo ($d_i = 0$, se for censurado, $d_i = 1$, se ocorreu a falha). Para dados grupados, a função de verossimilhança é frequentemente escrita em termos da probabilidade de falha do i-ésimo indivíduo em I_j, dado que ele estava vivo (ou não doente) em t_{j-1} (Colosimo e Giolo, 2006), ou seja,

$$\begin{aligned}
p_j &= P[T_i < t_j | T_i > t_{j-1}] \\
&= \frac{P[t_{j-1} < T_i < t_j]}{P[T_i > t_{j-1}]} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} \\
&= 1 - \frac{S(t_j)}{S(t_{j-1})}.
\end{aligned}$$

Podemos escrever, então:

$$S(t_j) = S(t_{j-1})[1 - p_j].$$

Utilizando essa notação e supondo que um indivíduo falhou no intervalo $(t_{j-1}, t_j]$, sua contribuição na função de verossimilhança será dada por:

$$\begin{aligned}
S(t_{j-1}) - S(t_j) &= S(t_{j-1}) - S(t_{j-1})[1 - p_j] = S(t_{j-1})[1 - (1 - p_j)] \\
&= S(t_{j-1})p_j = S(t_{j-2})[1 - p_{j-1}]p_j \\
&= ... = S(t_0)[1 - p_1][1 - p_2]...[1 - p_j]p_j.
\end{aligned}$$

Como $S(t_0) = 1$, temos que

$$S(t_{j-1}) - S(t_j) = p_j \prod_{l=1}^{j-1} [1 - p_l].$$

Da mesma forma, a contribuição de uma observação censurada em $(t_{j-1}, t_j]$ é dada por:

$$\begin{aligned} S(t_j) &= S(t_{j-1})[1-p_j] = S(t_{j-2})[1-p_{j-1}][1-p_j] \\ &= ... = S(t_0)[1-p_1][1-p_2]...[1-p_j]. \end{aligned}$$

Como $S(t_0) = 1$, temos que

$$S(t_j) = \prod_{l=1}^{j}[1-p_l].$$

De maneira geral, a contribuição para a função de verossimilhança de um indivíduo i que foi observado no j-ésimo intervalo de tempo (I_j) pode ser expressa como:

$$\begin{aligned} & [[\{1-p_1\}...\{1-p_{j-1}\}]p_j]^{d_i}[\{1-p_1\}...\{1-p_j\}]^{(1-d_i)} \\ &= \left(\prod_{l=1}^{j-1}(1-p_l)p_j\right)^{d_i}\left(\prod_{l=1}^{j}(1-p_l)\right)^{(1-d_i)} \\ &= \left(\prod_{l=1}^{j-1}(1-p_l)p_j\right)^{d_i}\left(\prod_{l=1}^{j-1}(1-p_l)(1-p_j)\right)^{(1-d_i)} \\ &= \left(\prod_{l=1}^{j-1}(1-p_l)\right)^{d_i} {p_j}^{d_i}(1-p_l)^{(1-d_i)}(1-p_j)^{(1-d_i)} \\ &= \left(\prod_{l=1}^{j-1}(1-p_l)\right) {p_j}^{d_i}(1-p_j)^{(1-d_i)}. \end{aligned} \quad (2.2)$$

Se assumirmos que os indivíduos foram avaliados nos tempos $1, 2, 3, ..., k$; $(t_j - t_{j-1} = 1, \forall j)$, então:

$$p_j = P[T_i < t_j | T_i > t_{j-1}] = \lambda(t_j) = \lambda_j,$$

ou seja, p_j pode ser visto como a função de taxa de falha no intervalo j. Neste caso, a contribuição do i-ésimo indivíduo, em (2.2), pode ser reescrita como:

$$\prod_{l=1}^{j-1}(1-\lambda_l){\lambda_j}^{d_i}(1-\lambda_j)^{1-d_i}. \quad (2.3)$$

Utilizando a reparametrização $\xi_j = log(\frac{\lambda_j}{1-\lambda_j})$, $j = 1, ..., k$, podemos fazer:

$$e^{\xi_j} = \frac{\lambda_j}{1-\lambda_j}$$

e

$$\lambda_j = \frac{1}{1+e^{\xi_j}} = (1+e^{\xi_j})^{-1}.$$

Assim, a contribuição de um indivíduo i que foi observado até o j-ésimo intervalo de tempo para a função de verossimilhança será dada por:

$$\left(\prod_{l=1}^{j-1}(1-(1+e^{\xi_l})^{-1})\right)((1+e^{\xi_j})^{-1})^{d_i}(1-(1+e^{\xi_j})^{-1})^{1-d_i}.$$

Se assumirmos que o tempo até a falha possui distribuição exponencial com parâmetro α, então:

$$\lambda_j = p_j = \frac{1}{\alpha}$$

e, com base em (2.3), temos que a contribuição de um indivíduo será dada por:

$$\left(\left(1-\frac{1}{\alpha}\right)\left(\frac{1}{\alpha}\right)^{d_i}\left(1-\frac{1}{\alpha}\right)^{(1-d_i)}\right)^{j-1}.$$

Se assumirmos, agora, que o tempo até a falha possui distribuição Weibull com parâmetros α e γ com ambos maiores que zero, então:

$$\lambda_j = p_j = \left(\frac{\gamma}{\alpha^\gamma}\right)^{t_j{}^{\gamma-1}}.$$

A contribuição de um indivíduo para o modelo será dada por:

$$\prod_{l=1}^{j-1}\left(1-\left(\frac{\gamma}{\alpha^\gamma}\right)^{t_l{}^{\gamma-1}}\right)\left(\left(\frac{\gamma}{\alpha^\gamma}\right)^{t_j{}^{\gamma-1}}\right)^{d_i}\left(1-\left(\frac{\gamma}{\alpha^\gamma}\right)^{t_j{}^{\gamma-1}}\right)^{(1-d_i)}.$$

2.2 Modelos de regressão discretos

Como na seção 2.1, considere que os tempos de vida são grupados em k intervalos denotados por $I_j = (t_{j-1}, t_j]$, $j = 1, ..., k$, em que $0 = t_0 < t_1 < t_2 < ... < t_k$, e assuma que as censuras ocorrem no final do intervalo. Agora, considerando a presença de covariáveis no estudo, temos que a função de verossimilhança é escrita em termos de probabilidade de falha do i-ésimo indivíduo em I_j, dado que ele estava vivo ou não doente em I_{j-1} e o vetor de covariáveis $\mathbf{x}_i$, ou seja,

$$\begin{aligned} p_j(\mathbf{x}_i) &= P[T_i < t_j | T_i > t_{j-1}, \mathbf{x}_i] \\ &= \frac{P[t_{j-1} < T_i < t_j | \mathbf{x}_i]}{P[T_i > t_{j-1} | \mathbf{x}_i]} \\ &= \frac{S(t_{j-1}|\mathbf{x}_i) - S(t_j|\mathbf{x}_i)}{S(t_{j-1}|\mathbf{x}_i)} \\ &= 1 - \frac{S(t_j|\mathbf{x}_i)}{S(t_{j-1}|\mathbf{x}_i)}. \end{aligned}$$

Então, como na seção 2.1, podemos escrever a contribuição do indivíduo i com tempo de censura em $(t_{j-1}, t_j]$ como:

$$\begin{aligned} S(t_j|\mathbf{x}_i) &= S(t_{j-1}|\mathbf{x}_i)[1 - p_j(\mathbf{x}_i)] = S(t_{j-2}|\mathbf{x}_i)[1 - p_{j-1}(\mathbf{x}_i)][1 - p_j(\mathbf{x}_i)] \\ &= ... = S(t_0|\mathbf{x}_i)[1 - p_1(\mathbf{x}_i)][1 - p_2(\mathbf{x}_i)]...[1 - p_j(\mathbf{x}_i)]. \end{aligned}$$

Como $S(t_0|\mathbf{x}_i) = 1$, temos que

$$S(t_j|\mathbf{x}_i) = \prod_{l=1}^{j} [1 - p_l(\mathbf{x}_i)]. \tag{2.4}$$

Da mesma forma, a contribuição do indivíduo i com tempo de falha em $(t_{j-1}, t_j]$ é:

$$\begin{aligned} S(t_{j-1}|\mathbf{x}_i) - S(t_j|\mathbf{x}_i) &= S(t_{j-1}|\mathbf{x}_i) - S(t_{j-1}|\mathbf{x}_i)[1 - p_j(\mathbf{x}_i)] = S(t_{j-1}|\mathbf{x}_i)[1 - (1 - p_j(\mathbf{x}_i))] \\ &= S(t_{j-1}|\mathbf{x}_i)p_j(\mathbf{x}_i) = S(t_{j-2}|\mathbf{x}_i)[1 - p_{j-1}(\mathbf{x}_i)]p_j(\mathbf{x}_i) \\ &= ... = S(t_0|\mathbf{x}_i)[1 - p_1(\mathbf{x}_i)][1 - p_2(\mathbf{x}_i)]...[1 - p_j(\mathbf{x}_i)]p_{j-1}(\mathbf{x}_i). \end{aligned}$$

Como $S(t_0|\mathbf{x}_i) = 1$, temos que

$$S(t_{j-1}|\mathbf{x}_i) - S(t_j|\mathbf{x}_i) = p_j(\mathbf{x}_i) \prod_{l=1}^{j-1} [1 - p_l(\mathbf{x}_i)]. \tag{2.5}$$

Podemos então escrever a função de verossimilhança como (Colosimo e Giolo, 2006):

$$\prod_{j=1}^{k} \prod_{i \in R_j} (p_j(\mathbf{x}_i))^{d_i} (1 - p_j(\mathbf{x}_i))^{(1-d_i)},$$

em que R_j é o conjunto de todas as observações sob risco em $(t_{j-1}, t_j]$.

Como observaram Colosimo e Giolo (2006), a expressão 2.5 corresponde à função de verossimilhança de uma variável aleatória com distribuição Bernoulli com variável resposta d_i e probabilidade de sucesso $p_j(\mathbf{x}_i)$. Além disso, a estrutura de regressão representada por $p_j(\mathbf{x}_i)$ pode ser modelada por meio de um modelo de riscos proporcionais (Collett, 2006).

Se o modelo de riscos proporcionais de Cox para o tempo de vida T for assumido, a função de sobrevivência tem a seguinte forma:

$$S(t_j|\mathbf{x}_i) = exp[-\int_0^t \lambda(u|\mathbf{x})du] = [S_0(t_j)]^{exp(\beta'\mathbf{x}_i)},$$

em que $S_0(t_j)$ é a função de sobrevivência básica e β é o vetor de coeficientes da regressão.

Então, podemos escrever:

$$p_j(\mathbf{x}_i) = 1 - \frac{S(t_j|\mathbf{x}_i)}{S(t_{j-1}|\mathbf{x}_i)} = 1 - \frac{[S_0(t_j)]^{exp(\beta'\mathbf{x}_i)}}{[S_0(t_{j-1})]^{exp(\beta'\mathbf{x}_i)}} = 1 - \frac{S_0(t_j)}{S_0(t_{j-1})}^{exp(\beta'\mathbf{x}_i)}$$

$$p_j(\mathbf{x}_i) = 1 - \gamma_j^{exp(\beta'\mathbf{x}_i)}, \tag{2.6}$$

em que

$$\gamma_j = \frac{S_0(t_j)}{S_0(t_{j-1})}.$$

Assim, podemos escrever a função de verossimilhança como:

$$L = \prod_{j=1}^{k} \prod_{i \in R_j} \left(1 - \gamma_j^{exp(\beta'\mathbf{x}_i)}\right)^{d_j} \left(\gamma_j^{exp(\beta'\mathbf{x}_i)}\right)^{(1-d_j)}$$

e o logaritmo da função de verossimilhança é:

$$l = \sum_{j=1}^{k} \sum_{i \in R_j} \left(d_j log \left(1 - {\gamma_j}^{exp(\beta' \mathbf{x}_i)} \right) + (1 - d_j) log \left({\gamma_j}^{exp(\beta' \mathbf{x}_i)} \right) \right).$$

Utilizando a reparametrização sugerida por Prentice e Gloeckler (1978) ${\gamma_j}^* = log(-log(\gamma_j))$ que torna ${\gamma_j}^*$ irrestrito, temos:

$$l = \sum_{j=1}^{k} \sum_{i \in R_j} \left(-(1 - d_j) exp({\gamma_j}^* + \beta' \mathbf{x}_i) + d_j log \left(1 - exp(-exp {\gamma_j}^* + \beta' \mathbf{x}_i) \right) \right).$$

Além disso, substituindo (2.6) nas expressões (2.4) e (2.5), temos os seguintes resultados:

$$S(t_j|\mathbf{x}_i) = \prod_{l=1}^{j} [1 - (1 - {\gamma_l}^{exp(\beta' \mathbf{x}_i)})] = \prod_{l=1}^{j} {\gamma_l}^{exp(\beta' \mathbf{x}_i)} \tag{2.7}$$

e

$$\begin{aligned} S(t_{j-1}|\mathbf{x}_i) - S(t_j|\mathbf{x}_i) &= [1 - {\gamma_j}^{exp(\beta' \mathbf{x}_i)}] \prod_{l=1}^{j-1} [1 - (1 - {\gamma_l}^{exp(\beta' \mathbf{x}_i)})] \\ &= [1 - {\gamma_j}^{exp(\beta' \mathbf{x}_i)}] \prod_{l=1}^{j-1} {\gamma_l}^{exp(\beta' \mathbf{x}_i)}, \end{aligned} \tag{2.8}$$

respectivamente. Esses resultados serão utilizados no capítulo 5.

Capítulo 3

Modelos de Análise de Sobrevivência com Erros de Classificação

Considere um estudo em que N indivíduos são avaliados individualmente sobre a ocorrência de uma doença durante um período pré-estipulado de tempo ou até a ocorrência do evento. Assim, vamos supor que periodicamente os indivíduos são avaliados através de um teste de diagnóstico. Se o exame for positivo, registra-se o tempo da ocorrência e o indivíduo sai do estudo, caso contrário, o individuo continuará a ser examinado até que se registre a ocorrência ou até o fim do estudo (tempo de censura). Temos, portanto, dados de sobrevivência com censura intervalar.

Neste contexto, os exames de diagnóstico para o registro da ocorrência podem estar sujeitos a erros de classificação que podem influenciar no processo inferencial. Na seção 3.1, ilustramos o efeito desses erros nas estimativas dos parâmetros da distribuição. A seção 3.2 cita algumas abordagens presentes na literatura para lidar com o problema dos erros de classificação nos dados de sobrevivência. A seção 3.3 trata especificamente do modelo de Meier *et. al.*(2003).

3.1 Ilustração do efeito dos erros de classificação no processo inferencial

Considere que 1000 indivíduos são avaliados periodicamente. Se fosse detectada a doença, registrava-se a ocorrência, e se até 30ª avaliação não fosse detectada a ocorrência, o tempo do indivíduo seria considerado como tempo de censura. Supondo que o tempo até a ocorrência tem distribuição exponencial com parâmetro igual a 1, foi gerada uma amostra de tamanho igual a 1000 sujeita a erros de classificação tal que a sensibilidade e a especificidade do teste realizado são, respectivamente, iguais a 0,5 e 1.

Utilizando a função *Survival* do software *R*, obteve-se a estimativa de 1,21 para o parâmetro com erro padrão igual a 0,03, o que equivale a um vício relativo de 21%. Este vício reflete na estimação de algumas quantidades de interesse, como podemos observar na Tabela 3.1.

Tabela 3.1: Comparação entre a distribuição estimada e a distribuição real.

α	Média	Mediana	Variância	1º Quartil	3º Quartil
1,00	1,000	0,693	1,000	0,288	1,386
1,21	1,210	0,838	1,464	0,463348	1,677

Pela análise acima, percebemos que utilizando testes com sensibilidade igual a 0,5 e especificidade igual a 1, o tempo médio até a ocorrência do evento ficou superestimado. Portanto, percebemos a necessidade de incorporar estes erros no modelo para que as estimativas dos parâmetros não sejam viciadas.

Um estudo mais completo do efeito dos erros de classificação é apresentado por Paggiaro e Torelli (2004).

3.2 Alguns modelos propostos na literatura

Considerando o problema de dados de sobrevivência com erros de classificação, são encontrados na literatura vários trabalhos realizados. Dentre eles, destacaremos alguns.

No estudo de Paggiaro e Torelli (2004), os indivíduos eram avaliados em dois momentos e assim foi explorado o efeito da classificação errada tanto no primeiro quanto no segundo momento na estimativa dos parâmetros do modelo estatístico. Os autores ampliaram o modelo em que não se leva em consideração a possibilidade de erros e incorporam os valores da sensibilidade e da especificidade. Além disso, propuseram uma abordagem Bayesiana para a estimativa dos parâmetros.

Mckeown e Jewell (2010) incorporam a sensibilidade e a especificidade, através de um método não paramétrico, na função de verossimilhança para casos em que os indivíduos são analisados apenas uma vez. Em um exame para detectar o vírus HPV, se o diagnóstico era positivo registrava-se o tempo de falha, se negativo registrava-se o tempo de censura.

Para o modelo de Regressão de Cox com erros de classificação, Snapinn (1998) propôs o uso de variáveis auxiliares que ajudam na definição da relação entre o resultado verdadeiro e o observado do teste. O autor utilizou uma abordagem bayesiana, sendo necessário definir distribuições a *priori* condicionadas nos resultados verdadeiros para as variáveis auxiliares. Balasubramanian e Lagakos (2001) desenvolveram um procedimento para estimar o risco de transmissão perinatal do HIV durante os estágios finais da gravidez supondo que a sensibilidade dos testes variavam com o tempo.

Meier *et.al.* (2003) estenderam o modelo discreto de riscos proporcionais de Kalbfleish e Prentice (1980) para incorporar os resultados com erros. No entanto, além de estimar a sobrevida cumulativa, estimou-se os efeitos das covariáveis condicional em conhecer a sensibilidade e a especificidade.

Hunsberg, Albert e Dodd (2010), baseados no trabalho de Meier *et.al.* (2003) propuseram um modelo no qual as medidas de erros variam para as covariáveis do modelo.

Como o modelo proposto por Meier *et.al.* (2003) tem um contexto semelhante ao desenvolvido neste trabalho, a metodologia do artigo será detalhada na próxima seção.

3.3 O Modelo de Meier *et.al.*(2003)

Considere o contexto descrito a seguir. Para $i = 1, \ldots, N$, seja t_i (t_i^o) o tempo verdadeiro (observado) que o i-ésimo indivíduo teve o evento e d_i (d_i^o) o indicador de status verdadeiro (observado) (1=falha, 0=censura). Denota-se por θ e ϕ, respectivamente, a sensibilidade e a especificidade do teste. Seja o vetor $I_i = (I_{i1}, \ldots, I_{ik})$ indicador de visitas faltantes (1=perdidas, 0=não perdidas) para o sujeito i nos momentos $1, \ldots, k$ em que k é o último momento observado. Visitas perdidas são resumidas pelo vetor $C_i = (C_{i1}, \ldots, C_{it_i^o})$ em que $C_{ik} = \sum_{l=1}^{k-1} I_{il}$, para $(j = 2, \ldots, t_i^o)$ e $C_{i1} = 0$, isto é, C_{ik} é o número de visitas faltantes do sujeito i antes do tempo k.

Suponha que sujeitos são testados em $t_1 = 1, t_2 = 2, t_3 = 3, ...$ até o momento em que a falha é observada pela primeira vez. Indivíduos com indicador de falha positivo são excluídos mesmo que a falha seja falsa.

Para um sujeito que perdeu uma visita anterior à falha no tempo t_i, e perdeu uma segunda visita depois de t_i, mas antes do teste positivo t_i^o, os eventos seriam ordenados da seguinte maneira:

$$\underbrace{1, 2, ..., \overbrace{t_i - 4}^{perdida}, ..., t_i - 1}_{t_i - 2 \ verdadeiros \ negativos}, \underbrace{\overbrace{t_i}^{falha}, ..., \overbrace{t_i + 3}^{perdida}, ..., t_i^o - 1}_{t_i^o - t_i - 1 \ falsos \ negativos} \qquad \overbrace{t_i^o}^{falha \ observada} \qquad .$$

O indicador de faltosos inclui $I_{i(t_i-4)} = I_{i(t_i+3)} = 1$, então $C_{it_i} = \sum_{l=1}^{t_i-1} I_{il} = 1$ e $C_{it_i^o} = 2$. A probabilidade destes resultados dada a falha em t_i é $\phi^{t_i-2}(1-\theta)^{t_i^o - t_i - 1}\theta$. Em geral,

$$f(t_i^o, d_i^o | t_i = t_i^o, d_i = 0, \theta, \phi) = \phi^{t_i^o - C_{it_i^o} - 1} \phi^{1-d_i^o} (1-\phi)^{d_i^o} = \Gamma_i$$

e

$$f(t_i^o, d_i^o | t_i \leq t_i^o, d_i = 1, \theta, \phi) = \phi^{t_i - C_{it_i} - 1} (1-\theta)^{t_i^o - t_i - C_{it_i^o} + C_{it_i}} (1-\theta)^{1-d_i^o} \theta^{d_i^o} = \Delta_{it_i}.$$

Observe que, tanto para Γ_i quanto para Δ_{it_i}, a contribuição do último teste em t_i^o depende se ocorreu censura ou falha. Assumiu-se que o seguimento termina quando o evento é observado, de modo que eventos ocorridos após t_i^o são censurados em t_i^o. Assim, tem-se $t_i = t_i^o$, $d_i = 0$ ou $t_i \leq t_i^o$, $d_i = 1$ mas não $t_i > t_i^o$.

Seja ρ o vetor de parâmetros da distribuição do tempo até a falha. A densidade marginal dos dados observados foi calculada sobre os possíveis valores de t_i e d_i:

$$\begin{aligned}
f(t_i^o, d_i^o, \rho, \theta, \phi) &= \sum_{\forall t_i} f(t_i^o, d_i^o, \rho, \theta, \phi, t_i, d_i = 0) + \sum_{\forall t_i} f(t_i^o, d_i^o, \rho, \theta, \phi, t_i, d_i = 1) \\
&= f(t_i = t_i^o, d_i = 0, t_i^o, d_i^o, \rho, \theta, \phi) + \sum_{l=1}^{t_i^o} f(t_i = l, d_i = 1, t_i^o, d_i^o, \rho, \theta, \phi) \\
&= f(t_i = t_i^o, d_i = 0, \rho, \theta, \phi) f(t_i^o, d_i^o | t_i = t_i^o, d_i = 0, \rho, \theta, \phi) + \\
&+ \sum_{l=1}^{t_i^o} f(t_i = l, d_i = 1, \rho, \theta, \phi) f(f_i^o, d_i^o | t_i = l, d_i = 1, \rho, \theta, \phi) \\
&= f(t_i = t_i^o, d_i = 0, \rho, \theta, \phi)\Gamma_i + \sum_{l=1}^{t_i^o} f(t_i = l, d_i = 1, \rho, \theta, \phi)\Delta_{il}. \qquad (3.1)
\end{aligned}$$

As estimativas dos parâmetros podem ser obtidas por maximização da função de verossimilhança com base em (3.1). A variância do estimador de máxima verossimilhança é aproximada por $I(\hat{\rho})^{-1}$, em que $I(\hat{\rho})$ é a matrix de informação observada.

No caso em que não se tem visitas perdidas, a função de densidade conjunta é dada por:

$$f(t_i, d_i) = [S(t_{i-1}) - S(t_i)]^{d_i} [S(t_i)]^{1-d_i}$$

como em (2.1). Então, utilizando (3.1), temos:

$$f(t_i^o, d_i^o) = S(t_i^o)\phi^{t_i^o - 1}\phi^{1-d_i^o}(1-\phi)^{d_i^o} + \sum_{l=1}^{t_i^o} [S(l-1) - S(l)]\phi^{l-1}(1-\theta)^{t_i^o - l}(1-\theta)^{1-d_i^o}\theta^{d_i^o}$$

$$\begin{aligned}
&= \left(S(t_i^o)\phi^{t_i^o - 1}(1-\phi) + \sum_{l=1}^{t_i^o} [S(l-1) - S(l)]\phi^{l-1}(1-\theta)^{t_i^o - l}\theta \right)^{(d_i^o)} \times \\
&\quad \left(S(t_i^o)\phi^{t_i^o - 1}\phi + \sum_{l=1}^{t_i^o} [S(l-1) - S(l)]\phi^{l-1}(1-\theta)^{t_i^o - l}(1-\theta) \right)^{(1-d_i^o)}
\end{aligned}$$

$$\begin{aligned}
&= \left(S(t_i^o)\phi^{t_i^o - 1}(1-\phi) + \theta \sum_{l=1}^{t_i^o} [S(l-1) - S(l)]\phi^{l-1}(1-\theta)^{t_i^o - l} \right)^{(d_i^o)} \times \\
&\quad \left(S(t_i^o)\phi^{t_i^o} + (1-\theta) \sum_{l=1}^{t_i^o} [S(l-1) - S(l)]\phi^{l-1}(1-\theta)^{t_i^o - l} \right)^{(1-d_i^o)}. \qquad (3.2)
\end{aligned}$$

3.3.1 Modelo de Riscos Proporcionais

Considere agora a presença de covariáveis no estudo. O vetor de variáveis explicativas para o i-ésimo sujeito é dado por $\mathbf{x}_i$ com comprimento p. Utilizando $\gamma_{0j} = log(\frac{\lambda_{0j}}{1-\lambda_{0j}})$, em que λ_{0j} é a função de taxa de falha no intervalo j , temos que

$$\lambda_{0j} = \frac{exp(\gamma_{0j})}{1 + exp(\gamma_{0j})}$$

e, como demonstra Kalbfleisch e Prentice (2002):

$$f(t_i, d_i;\ \mathbf{x}_i, \beta, \lambda_0) = \prod_{j=1}^{t_i-1}((1+e^{\gamma_{0j}})^{-e^{\beta'\mathbf{x}_i}})(1-(1+e^{\gamma_{0t_i}})^{-e^{\beta'\mathbf{x}_i}})^{d_i}((1+e^{\gamma_{0t_i}})^{-e^{\beta'\mathbf{x}_i}})^{(1-d_i)}.$$

Quando o teste de diagnóstico está sujeito a erros de classificação, podemos utilizar as definições de Γ_i e Δ_{it_i} (seção 3.1), obtendo:

$$f(t_i^o, d_i^o,\ \mathbf{x}_i, \beta, \gamma_0, \theta, \phi) = f(t_i = t_i^o, d_i = 0, t_i^o, d_i^o,\ \mathbf{x}_i, \beta, \gamma_0, \theta, \phi) + \sum_{l=1}^{t_i^o} f(t_i = l, d_i = 1, t_i^o, d_i^o,\ \mathbf{x}_i, \beta, \gamma_0, \theta, \phi)$$

$$= \left(\prod_{j=1}^{t_{i-1}^o}(1+e^{\gamma_0 j})^{-e^{\beta'\mathbf{x}_i}}\right)((1+e^{\gamma_0 t_i})^{-e^{\beta'\mathbf{x}_i}})\Gamma_i + \sum_{l=1}^{t_i^o}\prod_{j=1}^{l-1}((1+e^{\gamma_0 j})^{-e^{\beta'\mathbf{x}_i}})(1-(1+e^{\gamma_0 l})^{-e^{\beta'\mathbf{x}_i}})\Delta_{il}.$$

Como no primeiro termo da expressão temos $t_i = t_i^o$, a contribuição do indivíduo i é:

$$f(t_i^o, d_i^o,\ \mathbf{x}_i, \beta, \gamma_0, \theta, \phi) = \prod_{j=1}^{t_i^o}(1+e^{\gamma_{0j}})^{-e^{(\beta'\mathbf{x}_i)}}\Gamma_i + \sum_{l=1}^{t_i^o}\left(\prod_{j=1}^{l-1}((1+e^{\gamma_{0j}})^{-e^{(\beta'\mathbf{x}_i)}})(1-(1+e^{\gamma_{0l}})^{-e^{(\beta'\mathbf{x}_i)}})\Delta_{il}\right). \tag{3.3}$$

As estimativas dos parâmetros $\Omega = \binom{\lambda_0}{\beta}$ foram obtidas por maximização da função de verossimilhança com base em (3.3). A variância do estimador de máxima verossimilhança é aproximada por $I(\hat{\Omega})^{-1}$, em que $I(\hat{\Omega})$ é a matrix de informação observada.

Capítulo 4

Uma Proposta para Análise de Dados de Sobrevivência com Erros de Classificação

No modelo de Meier *et.al.* (2003) apresentado no capítulo anterior, os indivíduos são avaliados nos instantes de tempos $1, 2, ..., k$, em que k é o último tempo de avaliação. Suponha agora que os indivíduos são avaliados em t_1, t_2, ..., t_k (não necessariamente $t_1 = 1, t_2 = 2, t_k = k$). Nossa proposta é, portanto, generalizar os resultados de Meier *et.al.* (2003) para quaisquer intervalos de tempo. Novamente, como não se sabe o momento exato da ocorrência do evento e ocorrem empates de tempos, temos censura do tipo intervalar, caso especial com dados grupados.

4.1 Modelo proposto

Suponha que a falha de um indivíduo seja detectada em t_1, então essa falha (1) ocorreu entre t_0 e t_1 e foi detectada em t_1 ou (2) ocorreu depois de t_1 e foi erroneamente detectada em t_1. Então a contribuição deste indivíduo na verossimilhança é:

$$(S(t_0) - S(t_1))(\theta) + (1 - \phi)S(t_1).$$

Supondo que a falha de um indivíduo foi detectada em t_2, então a falha (1) pode ter ocorrido entre t_0 e t_1 e não ter sido detectada em t_1, (2) ocorreu entre t_1 e t_2 e foi detectada em t_2 ou (3) pode não ter ocorrido em nenhum dos tempos e sido detectada erroneamente em t_2. Então, a contribuição de um indivíduo que teve a falha detectada no t_2 na verossimilhança é dada por:

$$(S(t_0) - S(t_1))(1 - \theta)\theta + \theta\phi(S(t_1) - S(t_2)) + \phi(1 - \phi)S(t_2).$$

Se a falha de um indivíduo foi detectada em t_3, a falha pode (1) ter ocorrido entre t_0 e t_1 e não ter sido detectada em t_1, nem em t_2 e ter sido detectada em t_3, (2) pode ter ocorrido entre t_1 e t_2, não ter sido detectada em t_2 e ter sido detectada em t_3, (3) pode ter realmente ocorrido em t_3 e, (4) pode não ter ocorrido nem t_1, nem em t_2 e nem em t_3 e ter sido detectada erroneamente em t_3. Então, a contribuição de um indivíduo que teve a falha detectada no t_3 na verossimilhança é dada por:

$$(S(t_0) - S(t_1))(1-\theta)^2\theta + (1-\theta)\theta\phi(S(t_1) - S(t_2)) + (\theta)\phi^2(S(t_2) - S(t_3)) + \phi^2(1-\phi)S(t_3).$$

Seguindo o mesmo raciocínio, a contribuição para a verossimilhança de uma unidade que foi observada a falha em t_j; $j = 1, \ldots, k$ é:

$$\phi^{(j-1)}(1-\phi)S(t_j) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(S(t_{l-1}) - S(t_l)),$$

em que $S(t) = P(T \geq t)$ e $S(t_0) = 1$.

Para o indivíduo em que ocorreu a censura, ele pode realmente ter sido censurado ou a falha pode ter ocorrido em qualquer tempo e não ter sido detectada em nenhum dos tempos posteriores.

Suponha que o tempo observado da censura seja o t_3. Deste modo, a falha pode ter ocorrido antes do tempo t_1 e não ter sido detectada nos testes um, dois e três, pode ter ocorrido entre o t_1 e o t_2 e não ter sido detectada nos testes 2 e 3, pode ter ocorrido entre o t_2 e o t_3 e não ter sido detectada no terceiro teste e pode realmente não ter falhado até o t_3, ou seja, ele sobreviveu ao terceiro teste. Assim, a contribuição na verossimilhança para este indivíduo pode ser dada por:

$$(S(t_0) - S(t_1))(1-\theta)^3 + (S(t_1) - S(t_2))\phi(1-\theta)^2 + (S(t_2) - S(t_3))\phi^2(1-\theta) + S(t_3)\phi^3.$$

Deste modo, a contribuição para a verossimilhança de uma unidade que foi censurada em t_j; $j = 1, \ldots, k$ é:

$$\phi^j S(t_j) + (1-\theta)\sum_{l=1}^{j}(S(t_{l-1}) - S(t_l))\phi^{(l-1)}(1-\theta)^{(j-l)}).$$

De modo geral, a contribuição de um indivíduo para a função de verossimilhança é dada por:

$$\begin{aligned}\left(\phi^{(j-1)}(1-\phi)S(t_j) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(S(t_{l-1}) - S(t_l))\right)^{d_j} \times \\ \times\left(\phi^j S(t_j) + (1-\theta)\sum_{l=1}^{j}(S(t_{l-1}) - S(t_l))\phi^{(l-1)}(1-\theta)^{(j-l)}\right)^{1-d_j}.\end{aligned} \quad (4.1)$$

Se em (4.1) considerarmos que $t_1 = 1, t_2 = 2, ..., t_l = l$ e utilizando a mesma notação que a usada por Meier *et.al.* (2003), teremos que esta expressão é igual a expressão (3.2).

Considerando novamente uma amostra de tamanho N tal que n_j é o número de falhas e m_j é o número de censuras em t_j, então a função de verossimilhança é dada por:

$$L = \prod_{j=1}^{k} \left(\phi^{(j-1)}(1-\phi)S(t_j) + \theta \sum_{l=1}^{j} \phi^{(l-1)}(1-\theta)^{(j-l)}(S(t_{l-1}) - S(t_l)) \right)^{n_j} \times$$
$$\times \left(\phi^j S(t_j) + (1-\theta) \sum_{l=1}^{j} (S(t_{l-1}) - S(t_l))\phi^{(l-1)}(1-\theta)^{(j-l)} \right)^{m_j}.$$

Os estimadores de máxima verossimilhança para os parâmetros da distribuição podem ser obtidos maximizando o logaritmo da função de verossimilhança que é dado por:

$$l = \sum_{j=1}^{k} n_j log \left(\phi^{(j-1)}(1-\phi)S(t_j) + \theta \sum_{l=1}^{j} (S(t_{l-1}) - S(t_l))\phi^{(l-1)}(1-\theta)^{(j-l)} \right)$$
$$+ m_j log \left(\phi^j S(t_l) + (1-\theta) \sum_{l=1}^{j} (S(t_{l-1}) - S(t_j))\phi^{(l-1)}(1-\theta)^{(j-l)} \right).$$

4.1.1 Caso particular: especificidade igual a 1

Considerando o caso em que temos especificidade igual a 1, sensibilidade menor que 1 e amostras de tamanho N e supondo que foi detectada falha em t_1, então a falha ocorreu entre t_0 e t_1 e foi detectada em t_1. Portanto, a contribuição deste indivíduo na verossimilhança é dada por:

$$(S(t_0) - S(t_1))(\theta).$$

Supondo que foi detectada falha no t_2, então a falha (1) ocorreu antes de t_1, não foi detectada em t_1 mas foi detectada em t_2 ou (2) ocorreu entre t_1 e t_2 e foi detectada em t_2. Desta forma, a contribuição na verossimilhança de um indivíduo que teve a falha detectada no t_2 é dada por:

$$(S(t_0) - S(t_1))(1-\theta)(\theta) + (S(t_1) - S(t_2))(\theta).$$

Se foi detectada falha em t_3, a falha pode ter (1) ocorrido antes de t_1 e não ter sido detectada nem em t_1 nem em t_2 e foi detectada em t_3, pode ter (2) ocorrido entre t_1 e t_2, não ter sido detectada em t_2 e ter sido detectada em t_3 e pode ter (3) realmente ocorrido entre t_2 e t_3 e ter sido detectada em t_3. Assim, a contribuição na verossimilhança de um indivíduo que teve a falha detectada no t_3 é dada por:

$$(S(t_0) - S(t_1))(1-\theta)^2(\theta) + (S(t_1) - S(t_2))(1-\theta)(\theta) + (S(t_2) - S(t_3))(\theta).$$

Seguindo o mesmo raciocínio, a contribuição para a verossimilhança de uma unidade que foi observada a falha em t_j, $j = 1, \ldots, k$, é:

$$\theta\sum_{l=1}^{j}(S(t_{l-1})-S(t_l))(1-\theta)^{(j-l)},$$

em que $S(t)=P(T\geq t)$ e $S(t_0)=1$.

Para o indivíduo em que a censura ocorreu em t_j, ele pode realmente ter sido censurado ou a falha pode ter ocorrido em qualquer tempo anterior e não ter sido detectada em nenhum dos tempos posteriores.

Suponha que o tempo da censura seja t_3. Deste modo, a falha pode ter (1) ocorrido antes do tempo t_1 e não ter sido detectada nos testes um, dois e três, pode ter (2) ocorrido entre o t_1 e o t_2 e não ter sido detectada nos testes 2 e 3, pode ter (3) ocorrido entre o t_2 e o t_3 e não ter sido detectada no 3º teste e pode (4) realmente não ter falhado até o t_3, ou seja, ele sobreviveu ao terceiro teste. Assim, a contribuição na verossimilhança para este indivíduo pode ser dada por:

$$(S(t_0)-S(t_1))(1-\theta)^3+(S(t_1)-S(t_2))(1-\theta)^2+(S(t_2)-S(t_3))(1-\theta)+S(t_3).$$

Deste modo, a contribuição para a verossimilhança de uma unidade que foi censurada em t_j; $j=1,\ldots,k$ é:

$$S(t_j)+(1-\theta)\sum_{l=1}^{j}(S(t_{l-1})-S(t_l))(1-\theta)^{(j-l)}.$$

Considere uma amostra de tamanho N tal que n_j é o número de falhas e m_j é o número de censuras em t_j. Assim:

$$N=\sum_{j=1}^{k}(n_j+m_j).$$

Então, a função de verossimilhança é dada por

$$\prod_{j=1}^{k}[\theta\sum_{l=1}^{j}(S(t_{l-1})-S(t_l))(1-\theta)^{(j-l)}]^{n_j}[S(t_j)+(1-\theta)\sum_{l=1}^{j}(S(t_{l-1})-S(t_l))(1-\theta)^{(j-l)}]^{m_j},$$

e o logaritmo da verossimilhança é dado por:

$$\sum_{j=1}^{k}(n_j log(\theta\sum_{l=1}^{j}(S(t_{l-1})-S(t_l))(1-\theta)^{(j-l)})+m_j log(S(t_l)+(1-\theta)\sum_{l=1}^{j}(S(t_{l-1})-S(t_j))(1-\theta)^{(j-l)})).$$

4.1.2 Modelo utilizando a Distribuição Exponencial

Para o caso em que o tempo de falha tem distribuição Exponencial, a função de verossimilhança é dada por:

$$\prod_{j=1}^{k}[\phi^{(j-1)}(1-\phi)exp(-\alpha t_j)+\theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(exp(-\alpha t_{l-1}-exp(-\alpha t_l)))^{n_j}\times$$
$$\times\phi^j exp(-\alpha t_j)+(1-\theta)\sum_{l=1}^{j}(exp(-\alpha t_{l-1})-exp(-\alpha t_l))\phi^{(l-1)}(1-\theta)^{(j-l)}]^{m_j}.$$

O logaritmo da função de verossimilhança é:

$$l=\sum_{j=1}^{k}(\mathrm{n}_j\ \log\ (\phi^{(j-1)}(1-\phi)exp(-t_j\alpha)+\theta\textstyle\sum_{l=1}^{j}(exp(-t_{l-1}\alpha)-exp(-t_l\alpha))\phi^{(l-1)}(1-\theta)^{(j-l)})$$
$$+m_j log(\phi^j exp(-t_j\alpha)+(1-\theta)\sum_{l=1}^{j}(exp(-t_{l-1}\alpha)-exp(-t_l\alpha))\phi^{(l-1)}(1-\theta)^{(j-l)})$$

Derivando a expressão acima, obtemos:

$$\frac{d(l)}{d(\alpha)}=\sum_{j=1}^{k}\left(n_j\frac{\phi^{(j-1)}(1-\phi)(-t_j)exp^{-t_j\alpha}+\theta A_j}{\phi^{(j-1)}(1-\phi)exp^{-t_j\alpha}+\theta B_j}+m_j\frac{-t_j\phi^j exp^{-t_j\alpha}+(1-\theta)A_j}{\phi^j exp^{-t_j\alpha}+(1-\theta)B_j}\right),$$

em que

$$A_j=\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(t_l exp^{-t_l\alpha}-t_{l-1}exp^{-t_{l-1}\alpha})$$

e

$$B_j=\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(exp^{-t_{l-1}\alpha}-exp^{-t_l\alpha}).$$

A derivada segunda de l é dada por:

$$\frac{d^2l}{d\alpha^2}=\sum_{j=1}^{k}nj\frac{((1-\phi)\phi^{(j-1)}t_j^2 exp^{-t_j\alpha}+\theta C_j)((1-\phi)\phi^{(j-1)}exp^{-t_j\alpha}+\theta B_j)-(\phi^{(j-1)}(1-\phi)(-t_j)exp^{-t_j\alpha}+\theta A_j)^2}{((1-\phi)\phi^{(j-1)}exp^{-t_j\alpha}+\theta B_j)^2}+$$

$$+m_j\frac{(\phi^j t_j^2 exp^{-t_j\alpha}+(1-\theta)C_j)(\phi^j exp^{-t_j\alpha}+(1-\theta)B_j)-(-t_j exp^{-t_j\alpha}\phi^j+(1-\theta)A_j)^2}{(\phi^j exp^{-t_j\alpha}+(1-\theta)B_j)^2},$$

em que

$$C_j = \sum_{l=1}^{j}(\phi^{(l-1)}(1-\theta)^{(j-l)}(t_{l-1}^2 exp^{-t_{l-1}\alpha} - t_l^2 exp^{-t_l\alpha}).$$

Desta forma, a variância do estimador de máxima verossimilhança pode ser aproximada por :

$$var(\hat{\alpha}) = \frac{1}{(-\frac{d^2l}{d\alpha^2})}. \tag{4.2}$$

4.1.3 Modelo utilizando a Distribuição Weibull

Assumindo que o tempo tem distribuição Weibull(α, γ), temos que a função de verossimilhança é dada por:

$$L = \prod_{j=1}^{k}\left(\phi^{(j-1)}(1-\phi)exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)}\right)\right)^{n_j} \times$$

$$\times\left(\phi^j exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + (1-\theta)\sum_{l=1}^{j}\left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)}\right)\phi^{(l-1)}(1-\theta)^{(j-l)}\right)^{m_j}.$$

O logaritmo da função de verossimilhança é dado por:

$$l = \sum_{j=1}^{k} n_j log\left(\phi^{(j-1)}(1-\phi)exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + \theta\sum_{l=1}^{j}\left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)}\right)\phi^{(l-1)}(1-\theta)^{(j-l)}\right) +$$

$$+m_j log\left(\phi^j exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + (1-\theta)\sum_{l=1}^{j}\left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)}\right)\phi^{(l-1)}(1-\theta)^{(j-l)}\right).$$

O cálculo da variância observada para os estimadores de máxima verossimilhança é apresentado no Apêndice A.

4.2 Avaliação do modelo via Simulação Monte Carlo

Todos os programas utilizados para a avaliação dos modelos foram implemantados no $software$ estatístico R onde utilizou-se as funções *optim* e *optimize* para a maximização da função de verossimilhança.

4.2.1 Desempenho geral do modelo

Foram realizadas simulações em que o tempo de falha seguia distribuições exponenciais com parâmetros iguais a 1 e a 0,5. Os tempos de avaliação foram iguais a $[0, 0.5, 1, 1.6, 2, 2.7, 3]$ e $[0, 0.5, 1, 1.5, 2, 2.7, 3.5, 4, 4.6]$ respectivamente. Os tamanhos das amostras foram iguais a 120 e 1200 e foram realizadas 1000 simulações em cada cenário. O valores de especificidade e sensibilidade variaram entre 0,8 e 1,0.

Realizou-se, também, simulações em que o tempo da falha seguia distribuição Weibull com parâmetros de escala e de forma iguais a 2 e tempos de avaliação iguais a $[0, 0.5, 1.0, 1.6, 2.0, 2.7, 3.0]$. Novamente os tamanhos das amostras foram iguais a 120 e 1200 e um total de 1000 simulações em cada cenário.

Para avaliar o desempenho do modelo proposto utilizamos o vício relativo (% vício), o erro padrão das estimativas, o erro quadrático médio (EQM) e a variância assintótica média cálculada na seção anterior:

$$\%\text{ vício}= \tfrac{100(\overline{\alpha}-\alpha)}{\alpha},$$

em que $\overline{\alpha}$ é a média das estimativas e α é o valor especificado para o parâmetro.

$$\text{Erro padrão}= \left(\tfrac{1}{s}\textstyle\sum_{k=1}^{s}\left(\hat{\alpha}_k-\overline{\alpha}\right)^2\right)^{1/2},$$

em que s é o número de simulações, e

$$\text{EQM}= \tfrac{1}{s}\textstyle\sum_{k=1}^{s}\left(\hat{\alpha}_k-\alpha\right)^2.$$

Essas quantidades obtidas pelo modelo proposto (Prop.) foram também comparadas com aquelas obtidas pelo modelo de análise de sobrevivência convenvional (Conv.) que não considera a possibilidade de erros de classificação no teste de diagnóstico aplicado (sensibilidade e especificidade iguais a 1). Os resultados obtidos para a distribuição exponencial, utilizando percentual de falha igual a 80%, são apresentados nas Tabelas (4.1) a (4.4). Resultados para outros percentuais são apresentados no Apêndice B.

Tabela 4.1: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	4,589	1,023	0,018	0,018	0,002	0,000	0,001	0,001
1,00	0,90	10,054	2,441	0,013	0,012	0,010	0,001	0,001	0,001
1,00	0,80	12,875	-2,965	0,013	0,010	0,017	0,001	0,001	0,001
0,95	1,00	15,957	26,214	0,014	0,014	0,026	0,069	0,001	0,001
0,95	0,95	16,412	22,250	0,014	0,013	0,027	0,050	0,001	0,001
0,95	0,90	16,597	17,922	0,016	0,014	0,028	0,032	0,001	0,001
0,95	0,80	16,481	8,541	0,021	0,015	0,028	0,008	0,002	0,001
0,90	1,00	15,478	36,481	0,020	0,020	0,024	0,133	0,002	0,002
0,90	0,95	15,943	32,130	0,022	0,020	0,026	0,104	0,002	0,002
0,90	0,90	15,856	27,155	0,024	0,020	0,026	0,074	0,002	0,001
0,90	0,80	15,931	17,046	0,029	0,021	0,026	0,029	0,002	0,001
0,80	1,00	15,206	59,419	0,032	0,032	0,024	0,354	0,002	0,002
0,80	0,95	15,621	54,080	0,033	0,030	0,025	0,293	0,002	0,002
0,80	0,90	15,539	48,061	0,037	0,031	0,025	0,232	0,003	0,002
0,80	0,80	15,478	35,705	0,041	0,028	0,026	0,128	0,003	0,002

Tabela 4.2: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=120

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	4,589	0,925	0,025	0,024	0,003	0,001	0,010	0,009
1,00	0,90	9,371	1,684	0,026	0,022	0,009	0,001	0,011	0,009
1,00	0,80	13,557	-2,640	0,041	0,030	0,020	0,002	0,013	0,008
0,95	1,00	15,408	25,761	0,044	0,044	0,026	0,068	0,014	0,014
0,95	0,95	15,683	21,511	0,044	0,041	0,027	0,048	0,014	0,013
0,95	0,90	16,432	17,594	0,049	0,042	0,029	0,033	0,015	0,012
0,95	0,80	16,252	8,074	0,064	0,047	0,030	0,009	0,016	0,010
0,90	1,00	14,528	35,735	0,060	0,059	0,025	0,131	0,016	0,016
0,90	0,95	15,011	31,248	0,071	0,065	0,028	0,102	0,017	0,015
0,90	0,90	15,778	26,850	0,071	0,060	0,030	0,076	0,018	0,014
0,90	0,80	14,701	15,783	0,085	0,061	0,029	0,029	0,020	0,012
0,80	1,00	13,140	57,713	0,100	0,098	0,027	0,343	0,022	0,022
0,80	0,95	14,268	52,886	0,101	0,091	0,031	0,288	0,024	0,021
0,80	0,90	14,971	47,418	0,121	0,100	0,037	0,235	0,026	0,019
0,80	0,80	15,448	35,157	0,140	0,094	0,044	0,132	0,031	0,016

Independente do tamanho amostral, o vício relativo absoluto do método proposto foi menor em todos os cenários, com exceção dos casos em que (i) ϕ=1 e =0,80, 0,90 ou 0,95 e a proporção de falha está entre 0,60 e 0,8 ou (ii) Exp(1), ϕ=0,95 e θ=0,80.

Na grande maioria dos cenários simulados ocorreu superestimação do parâmetro (vício relativo maior que zero) no método proposto e no convencional. A subestimação é observada com maior freqüência apenas nos casos em que ϕ=1 e θ <1.

Para o método convencional, na medida em que a qualidade do teste diminui, o vício relativo aumenta, enquanto que no método proposto essa quantidade apresenta certa estabilidade.

Nas Tabelas 4.3 e 4.4 abaixo, apresentamos os resultados para a distribuição exponencial com parâmetro igual a 0,5 onde os tempos de avaliação foram iguais a $[0, 0.5, 1, 1.5, 2, 2.7, 3.5, 4, 4.6]$

Tabela 4.3: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(0.5) com n=1200

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	14,828	13,087	0,005	0,005	0,006	0,004	0,000	0,000
1,00	0,90	17,348	13,251	0,006	0,005	0,008	0,004	0,000	0,000
1,00	0,80	18,742	9,275	0,005	0,005	0,009	0,002	0,000	0,000
0,95	1,00	21,472	40,640	0,009	0,009	0,012	0,041	0,000	0,000
0,95	0,95	21,150	37,892	0,009	0,009	0,011	0,036	0,000	0,000
0,95	0,90	20,320	34,378	0,010	0,009	0,010	0,030	0,000	0,000
0,95	0,80	19,774	27,697	0,011	0,009	0,010	0,019	0,000	0,000
0,90	1,00	20,143	59,960	0,014	0,014	0,010	0,090	0,001	0,001
0,90	0,95	19,516	56,584	0,014	0,014	0,010	0,080	0,001	0,001
0,90	0,90	19,102	53,113	0,014	0,013	0,009	0,071	0,001	0,000
0,90	0,80	18,526	45,642	0,015	0,013	0,009	0,052	0,001	0,000
0,80	1,00	18,129	104,005	0,021	0,022	0,009	0,271	0,001	0,001
0,80	0,95	17,700	100,006	0,024	0,024	0,008	0,251	0,001	0,001
0,80	0,90	17,375	95,727	0,024	0,023	0,008	0,230	0,001	0,001
0,80	0,80	17,301	86,688	0,027	0,022	0,008	0,188	0,001	0,001

Tabela 4.4: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(0.5) com n=120

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	10,187	8,599	0,013	0,013	0,003	0,002	0,003	0,003
1,00	0,90	13,050	9,289	0,010	0,010	0,004	0,002	0,003	0,003
1,00	0,80	13,315	4,654	0,014	0,012	0,005	0,001	0,003	0,002
0,95	1,00	19,153	38,322	0,029	0,029	0,010	0,038	0,004	0,004
0,95	0,95	19,502	36,395	0,030	0,029	0,010	0,034	0,004	0,004
0,95	0,90	18,047	32,461	0,033	0,031	0,009	0,027	0,004	0,004
0,95	0,80	16,724	25,470	0,033	0,029	0,008	0,017	0,004	0,003
0,90	1,00	18,875	58,677	0,043	0,044	0,011	0,088	0,005	0,005
0,90	0,95	18,857	56,084	0,047	0,046	0,011	0,081	0,005	0,005
0,90	0,90	18,338	52,673	0,047	0,045	0,011	0,071	0,006	0,005
0,90	0,80	16,850	44,714	0,053	0,045	0,010	0,052	0,006	0,005
0,80	1,00	19,741	105,578	0,077	0,080	0,016	0,285	0,009	0,009
0,80	0,95	20,241	102,559	0,078	0,076	0,016	0,269	0,009	0,009
0,80	0,90	19,316	97,823	0,081	0,075	0,016	0,245	0,010	0,008
0,80	0,80	17,950	87,695	0,093	0,076	0,017	0,198	0,011	0,008

No caso da distribuição exponencial com parâmetro igual a 0,5 e independente do tamanho amostral e do percentual de falhas, o vício relativo absoluto foi menor em todos os cenários, com exceção dos casos em que ϕ=1 e θ <1. Observa-se que ocorreu superestimação do parâmetro (vício relativo maior que zero) no método proposto e no convencional em todos os cenários simulados. Para o método convencional, na medida em que a qualidade do teste diminui, o vício relativo aumenta, enquanto que no método proposto essa quantidade apresenta certa estabilidade.

Quanto à variabilidade, as duas distribuições simuladas (Exp(1) e Exp(0.5)) apresentaram resultados semelhantes. O erro padrão das estimativas em geral se mostrou ligeiramente maior no método proposto e sofreu um pequeno acréscimo nos dois métodos na medida em que sensibilidade e/ou especificidade diminuíam. O erro quadrático médio foi, em geral, menor no método proposto e sofreu pouca variação na medida em que a qualidade do teste diagnóstico diminui, enquanto no método convencional teve seu valor aumentado. Variância observada média dos estimadores apresentou valores bem pequenos em ambos os métodos.

Nas Tabelas 4.5 a 4.8 são apresentados os resultados para a distribuição de tempo de falha Weibull com parâmetros de forma e escala iguais a 2, amostras com tamanho iguais a 120 e 1200 e percentual de falha igual a 80%. Resultados para outros percentuais são apresentados no Apêndice *C*.

Tabela 4.5: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=1200

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-1,959	-1,194	0,017	0,016	0,002	0,001	0,006	0,006
1,00	0,90	-1,156	0,507	0,018	0,017	0,001	0,000	0,006	0,006
1,00	0,80	-1,278	2,346	0,027	0,025	0,001	0,003	0,006	0,006
0,95	1,00	-0,117	-12,037	0,063	0,041	0,004	0,060	0,007	0,004
0,95	0,95	-4,768	-15,033	0,034	0,022	0,010	0,091	0,007	0,004
0,95	0,90	-5,613	-15,285	0,038	0,025	0,014	0,094	0,007	0,004
0,95	0,80	-7,482	-15,615	0,043	0,030	0,024	0,098	0,009	0,005
0,90	1,00	-8,069	-23,698	0,051	0,027	0,029	0,225	0,009	0,004
0,90	0,95	-6,597	-22,901	0,055	0,031	0,020	0,211	0,010	0,004
0,90	0,90	-7,773	-23,379	0,060	0,031	0,028	0,220	0,011	0,004
0,90	0,80	-7,987	-23,146	0,055	0,027	0,029	0,215	0,013	0,005
0,80	1,00	-11,398	-32,663	0,070	0,023	0,057	0,427	0,020	0,003
0,80	0,95	-10,362	-32,528	0,076	0,026	0,049	0,424	0,022	0,003
0,80	0,90	-9,797	-32,662	0,094	0,025	0,047	0,427	0,026	0,003
0,80	0,80	-10,622	-32,803	0,114	0,031	0,058	0,431	0,033	0,004

Tabela 4.6: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=1200

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	1,610	2,939	0,015	0,014	0,001	0,004	0,002	0,002
1,00	0,90	-0,543	2,595	0,011	0,011	0,000	0,003	0,002	0,002
1,00	0,80	-2,407	4,815	0,013	0,012	0,002	0,009	0,002	0,002
0,95	1,00	-9,291	-18,927	0,051	0,043	0,037	0,145	0,002	0,001
0,95	0,95	-12,717	-20,335	0,018	0,016	0,065	0,166	0,002	0,002
0,95	0,90	-13,563	-19,373	0,019	0,016	0,074	0,150	0,002	0,002
0,95	0,80	-13,770	-15,866	0,023	0,019	0,076	0,101	0,003	0,003
0,90	1,00	-12,597	-30,096	0,025	0,018	0,064	0,363	0,003	0,001
0,90	0,95	-12,890	-28,779	0,024	0,018	0,067	0,332	0,003	0,002
0,90	0,90	-13,495	-27,796	0,029	0,022	0,074	0,310	0,003	0,002
0,90	0,80	-13,368	-24,475	0,028	0,020	0,072	0,240	0,004	0,003
0,80	1,00	-12,671	-43,642	0,051	0,024	0,067	0,762	0,006	0,001
0,80	0,95	-12,724	-42,593	0,042	0,022	0,067	0,726	0,006	0,001
0,80	0,90	-12,483	-41,473	0,047	0,024	0,065	0,689	0,007	0,001
0,80	0,80	-12,852	-39,124	0,054	0,028	0,069	0,613	0,010	0,002

Tabela 4.7: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=120

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	18,849	18,335	0,056	0,053	0,145	0,137	0,059	0,056
1,00	0,90	18,543	18,594	0,076	0,069	0,143	0,143	0,063	0,059
1,00	0,80	13,181	15,413	0,117	0,107	0,083	0,107	0,068	0,065
0,95	1,00	13,907	-4,829	0,157	0,092	0,102	0,018	0,092	0,044
0,95	0,95	14,045	-5,059	0,171	0,100	0,108	0,020	0,098	0,047
0,95	0,90	13,553	-5,338	0,194	0,109	0,111	0,023	0,105	0,051
0,95	0,80	12,692	-5,049	0,221	0,123	0,113	0,025	0,123	0,048
0,90	1,00	13,388	-15,543	0,252	0,103	0,135	0,107	0,140	0,040
0,90	0,95	14,252	-15,753	0,262	0,109	0,150	0,111	0,154	0,044
0,90	0,90	14,701	-16,081	0,297	0,110	0,175	0,116	0,168	0,049
0,90	0,80	14,542	-15,753	0,307	0,115	0,179	0,113	0,192	0,046
0,80	1,00	14,294	-28,673	1,326	0,097	1,840	0,338	0,781	0,034
0,80	0,95	12,542	-28,831	0,452	0,099	0,267	0,342	0,300	0,035
0,80	0,90	15,571	-29,292	1,226	0,100	1,601	0,353	1,004	0,042
0,80	0,80	19,383	-29,452	1,666	0,108	2,926	0,359	2,517	0,052

Tabela 4.8: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=120

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-1,228	0,657	0,015	0,015	0,001	0,000	0,012	0,012
1,00	0,90	-1,508	2,370	0,022	0,021	0,001	0,003	0,013	0,013
1,00	0,80	-2,390	5,549	0,036	0,034	0,004	0,013	0,016	0,018
0,95	1,00	-1,043	-12,443	0,063	0,051	0,004	0,065	0,017	0,015
0,95	0,95	-1,961	-11,383	0,065	0,054	0,006	0,055	0,019	0,017
0,95	0,90	-2,211	-9,809	0,067	0,057	0,006	0,042	0,021	0,021
0,95	0,80	-2,381	-6,138	0,076	0,064	0,008	0,019	0,027	0,016
0,90	1,00	-0,663	-21,732	0,097	0,068	0,010	0,193	0,025	0,016
0,90	0,95	-0,983	-20,417	0,097	0,067	0,010	0,171	0,028	0,019
0,90	0,90	-1,204	-19,076	0,103	0,073	0,011	0,151	0,032	0,023
0,90	0,80	-1,385	-15,705	0,108	0,076	0,012	0,104	0,034	0,020
0,80	1,00	0,326	-37,366	0,161	0,076	0,026	0,564	0,073	0,013
0,80	0,95	-0,999	-36,751	0,160	0,081	0,026	0,547	0,034	0,015
0,80	0,90	-0,521	-35,570	0,167	0,082	0,028	0,513	0,018	0,020
0,80	0,80	-0,321	-32,944	0,180	0,086	0,032	0,442	0,034	0,029

Para a distribuição Weibull com tamanho de amostra igual a 1200, com cenários realizados nos quais ϕ e θ variam entre 0,8 e 1,0 e percentuais de falha variando entre 0,6 e 0,9, o modelo proposto apresentou menor vício relativo absoluto para o parâmetro de forma em 69% dos cenários e, para o parâmetro de escala, em 81% dos cenários avaliados. Para tamanhos de amostras iguais a 120, estes percentuais foram próximos aos percentuais para amostras maiores (67% e 80%). Observou-se que, independente do tamanho da amostra, os cenários em que o modelo proposto apresentou maior percentual de vício relativo absoluto maior que o modelo convencional para o parâmetro de forma foi quando tínhamos percentual de falha igual a 90% (quando N=1200, tivemos apenas 5 cenários em 16 e $N = 120$ apenas 4 cenários em 16).

Em praticamente todos os cenários avaliados, o parâmetro de escala foi subestimado em ambos os modelos (vício relativo menor que zero) com exceções dos casos em que $\phi = \theta = 1$ onde o parâmetro de escala foi superestimado (exceto quando N=1200 e percentual de falha igual a 90%). Em se tratando do parâmetro de forma, também houve uma subestimação do parâmetro na maioria dos cenários mas observa-se que quando (i) percentual de falha igual 90% para N=120 e $N = 1200$ e (ii) percentual de falha é igual a 80% e N=120 houve uma superestimação do parâmetro de forma apenas pelo modelo proposto.

Ambos os modelos apresentaram medidas de variabilidade pequenas tanto para o parâmetro de forma quanto para o parâmetro de escala mas, nota-se que, como ocorreu para o modelo com a distribuição Exponencial, o modelo proposto apresentou valores de erro padrão e variância assintótica levemente maiores que o modelo convencional.

4.2.2 Avaliação do efeito da má especificação dos erros de classificação

Para avaliar o desempenho do modelo proposto quando os valores de sensibilidade e especificidades não são especificados corretamente, foram realizadas simulações via método de Monte Carlo.

Considere amostras de tamanho 1200 de uma distribuição Exp(1). Os indivíduos são avaliados nos tempos $[0, 0.5, 1, 1.6, 2, 2.7, 3]$ por um teste diagnóstico com sensibilidade conhecida e especificidade desconhecida. Nesse caso, avaliamos o modelo proposto especificando alguns possíveis valores para a especificidade (ϕ^*). Os resultados são apresentados nas Tabelas 4.9 a 4.12. Vale ressaltar que, para o método convencional, temos $\phi^* = 1$ e $\theta^* = 1$ em todos os cenários.

Tabela 4.9: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ é bem especificado e ϕ=0,95

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,90	0,95	3,865	21,014	0,015	0,014	0,002	0,044	0,001	0,001
0,85	0,95	-7,916	20,917	0,014	0,013	0,006	0,044	0,001	0,001
0,75	0,95	-33,272	21,035	0,015	0,014	0,111	0,044	0,001	0,001
0,90	0,90	3,597	16,760	0,017	0,014	0,002	0,028	0,001	0,001
0,85	0,90	-8,461	16,735	0,016	0,014	0,007	0,028	0,001	0,001
0,75	0,90	-34,866	16,633	0,016	0,015	0,122	0,028	0,001	0,001
0,90	0,85	3,141	12,227	0,018	0,014	0,001	0,015	0,002	0,001
0,85	0,85	-9,213	12,259	0,018	0,015	0,009	0,015	0,001	0,001
0,75	0,85	-36,671	12,042	0,017	0,014	0,135	0,015	0,001	0,001
0,90	0,75	2,191	2,681	0,021	0,015	0,001	0,001	0,002	0,001
0,85	0,75	-11,042	2,724	0,022	0,015	0,013	0,001	0,002	0,001
0,75	0,75	-40,418	2,681	0,020	0,014	0,164	0,001	0,002	0,001

Tabela 4.10: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ é bem especificado e ϕ=0,90

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	19,849	25,517	0,036	0,034	0,041	0,066	0,002	0,001
0,85	0,95	-2,654	25,946	0,021	0,020	0,001	0,068	0,002	0,001
0,75	0,95	-28,340	25,969	0,021	0,019	0,081	0,068	0,002	0,001
0,95	0,90	20,703	21,597	0,022	0,019	0,043	0,047	0,002	0,001
0,85	0,90	-3,134	21,387	0,021	0,018	0,001	0,046	0,002	0,001
0,75	0,90	-29,416	21,615	0,022	0,019	0,087	0,047	0,002	0,001
0,95	0,85	21,043	17,051	0,023	0,018	0,045	0,029	0,002	0,001
0,85	0,85	-3,457	16,909	0,024	0,019	0,002	0,029	0,002	0,001
0,75	0,85	-30,760	17,030	0,023	0,019	0,095	0,029	0,002	0,001
0,95	0,75	21,905	7,392	0,028	0,018	0,049	0,006	0,002	0,001
0,85	0,75	-4,354	7,275	0,029	0,019	0,003	0,006	0,002	0,001
0,75	0,75	-33,893	7,365	0,028	0,019	0,116	0,006	0,002	0,001

Tabela 4.11: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ é bem especificado e ϕ=0,85

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	31,334	36,168	0,037	0,035	0,100	0,132	0,002	0,002
0,90	0,95	20,454	36,432	0,027	0,024	0,043	0,133	0,002	0,002
0,75	0,95	-16,963	36,498	0,027	0,025	0,030	0,134	0,002	0,002
0,95	0,90	32,382	31,629	0,028	0,023	0,106	0,101	0,002	0,002
0,90	0,90	20,729	31,492	0,028	0,024	0,044	0,100	0,002	0,002
0,75	0,90	-17,757	31,721	0,027	0,023	0,032	0,101	0,002	0,002
0,95	0,85	33,259	26,704	0,028	0,022	0,111	0,072	0,002	0,001
0,90	0,85	21,322	26,622	0,031	0,024	0,046	0,071	0,002	0,001
0,75	0,85	-18,803	26,691	0,029	0,023	0,036	0,072	0,002	0,001
0,95	0,75	35,233	16,039	0,037	0,023	0,126	0,026	0,002	0,001
0,90	0,75	22,454	16,057	0,035	0,022	0,052	0,026	0,002	0,001
0,75	0,75	-21,116	16,137	0,035	0,023	0,046	0,027	0,002	0,001

Tabela 4.12: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ é bem especificado e ϕ=0,75

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	57,076	59,805	0,043	0,040	0,328	0,359	0,003	0,002
0,90	0,95	45,910	59,789	0,037	0,034	0,212	0,359	0,003	0,002
0,85	0,95	34,262	53,857	0,040	0,037	0,119	0,360	0,003	0,002
0,95	0,90	59,035	53,990	0,043	0,035	0,350	0,293	0,003	0,002
0,90	0,90	47,357	53,875	0,040	0,033	0,226	0,291	0,003	0,002
0,85	0,90	35,155	47,752	0,042	0,035	0,125	0,291	0,003	0,002
0,95	0,85	61,002	47,800	0,041	0,030	0,374	0,229	0,003	0,002
0,90	0,85	49,067	47,788	0,044	0,032	0,243	0,229	0,003	0,002
0,85	0,85	36,272	35,080	0,044	0,033	0,133	0,229	0,003	0,002
0,95	0,75	66,863	34,982	0,057	0,031	0,450	0,123	0,004	0,002
0,90	0,75	54,042	59,901	0,050	0,029	0,295	0,124	0,004	0,002
0,85	0,75	39,905	59,895	0,053	0,030	0,162	0,124	0,004	0,002

Em apenas 54% dos cenários simulados (26/48), o método proposto apresentou vício relativo absoluto menor que o método convencional. No método proposto, ocorreu subestimação do parâmetro da distribuição quando (i) ϕ é igual a 0,90 ou 0,95 e ϕ^* igual a 0,75 ou 0,85 e (ii) ϕ é igual a 0,85 e ϕ^* é igual a 0,75. Como esperado, quanto maior a diferença $\phi - \phi^*$, maior o vício relativo absoluto em ambos os métodos. Para o método proposto, nos casos em que $\phi - \phi^* = 0,05$, o vício relativo absoluto é menor que do que nos casos em que $\phi - \phi^* = -0,05$. No método convencional, fixando-se ϕ (e consequentemente a diferença $\phi - \phi^*$, em que $\phi^* = 1$), temos que o vício relativo absoluto diminui na medida em que θ aumenta. Em todos os cenários a variabilidade de ambos os modelos foi pequena e, em geral, o modelo proposto apresentou valores de erro quadrático, EQM e variância assintótica levemente maior que o modelo convencional.

Considere novamente amostras de tamanho 1200 de uma distribuição Exp(1) em que os indivíduos também são avaliados nos tempos $[0, 0.5, 1, 1.6, 2, 2.7, 3]$, mas suponha agora que o teste diagnóstico tenha especificidade conhecida e sensibilidade desconhecida. Nesse caso, avaliamos o modelo proposto especificando alguns possíveis valores para a sensibilidade (θ^*). Os resultados são apresentados nas Tabelas 4.13 a 4.16.

Tabela 4.13: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando ϕ é bem especificado e θ=0,95

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,90	18,945	20,171	0,015	0,013	0,036	0,041	0,002	0,001
0,95	0,85	24,674	20,137	0,017	0,014	0,061	0,041	0,002	0,001
0,95	0,75	41,197	20,189	0,022	0,013	0,170	0,041	0,003	0,001
0,90	0,90	19,121	30,245	0,024	0,020	0,037	0,092	0,002	0,001
0,90	0,85	25,335	30,086	0,025	0,020	0,065	0,091	0,002	0,001
0,90	0,75	44,027	29,967	0,034	0,020	0,195	0,090	0,003	0,001
0,85	0,90	19,021	40,550	0,028	0,024	0,037	0,165	0,002	0,002
0,85	0,85	26,221	40,590	0,032	0,024	0,070	0,165	0,003	0,002
0,85	0,75	48,199	40,594	0,043	0,024	0,234	0,165	0,005	0,002
0,75	0,90	20,597	64,343	0,043	0,035	0,044	0,415	0,003	0,002
0,75	0,85	29,689	64,196	0,052	0,037	0,091	0,413	0,004	0,002
0,75	0,75	61,584	64,299	0,078	0,036	0,385	0,415	0,009	0,002

Tabela 4.14: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando ϕ é bem especificado e θ=0,90

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	10,698	17,006	0,041	0,040	0,013	0,030	0,001	0,001
0,95	0,85	21,821	18,066	0,019	0,015	0,048	0,033	0,002	0,001
0,95	0,75	37,292	18,022	0,023	0,015	0,140	0,033	0,003	0,001
0,90	0,95	10,935	27,624	0,021	0,019	0,012	0,077	0,002	0,001
0,90	0,85	21,808	27,480	0,025	0,020	0,048	0,076	0,002	0,001
0,90	0,75	39,736	37,957	0,034	0,020	0,159	0,077	0,003	0,001
0,85	0,95	10,370	37,772	0,029	0,026	0,012	0,145	0,002	0,002
0,85	0,85	22,477	37,917	0,033	0,025	0,052	0,144	0,003	0,002
0,85	0,75	43,134	60,556	0,046	0,026	0,188	0,144	0,004	0,002
0,75	0,95	9,256	60,813	0,040	0,036	0,010	0,368	0,003	0,002
0,75	0,85	24,611	27,648	0,049	0,036	0,063	0,367	0,004	0,002
0,75	0,75	54,028	60,743	0,071	0,034	0,297	0,370	0,008	0,002

Tabela 4.15: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando ϕ é bem especificado e θ=0,85

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	5,887	12,447	0,038	0,037	0,005	0,017	0,001	0,001
0,95	0,90	11,156	13,404	0,018	0,015	0,013	0,018	0,001	0,001
0,95	0,75	30,378	13,437	0,023	0,015	0,093	0,018	0,002	0,001
0,90	0,95	5,602	22,575	0,021	0,020	0,004	0,051	0,001	0,001
0,90	0,90	10,128	22,441	0,024	0,021	0,011	0,051	0,002	0,001
0,90	0,75	32,031	22,712	0,032	0,020	0,104	0,052	0,003	0,001
0,85	0,95	4,453	32,365	0,028	0,026	0,003	0,105	0,002	0,002
0,85	0,90	9,376	32,181	0,030	0,026	0,010	0,104	0,002	0,002
0,85	0,75	34,006	32,333	0,043	0,025	0,118	0,105	0,004	0,002
0,75	0,95	2,084	53,810	0,038	0,035	0,002	0,291	0,002	0,002
0,75	0,90	8,440	53,938	0,041	0,034	0,009	0,292	0,003	0,002
0,75	0,75	41,136	53,925	0,064	0,033	0,173	0,292	0,007	0,002

Tabela 4.16: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando ϕ é bem especificado e θ=0,75

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-4,189	2,907	0,035	0,034	0,003	0,002	0,001	0,001
0,95	0,90	0,247	3,678	0,018	0,016	0,000	0,002	0,001	0,001
0,95	0,85	4,549	3,733	0,020	0,016	0,002	0,002	0,001	0,001
0,90	0,95	-5,611	11,961	0,021	0,019	0,004	0,015	0,001	0,001
0,90	0,90	-1,718	11,916	0,024	0,021	0,001	0,015	0,001	0,001
0,90	0,85	2,806	11,842	0,023	0,019	0,001	0,014	0,001	0,001
0,85	0,95	-7,770	20,817	0,026	0,024	0,007	0,044	0,001	0,001
0,85	0,90	-3,724	20,618	0,029	0,025	0,002	0,043	0,002	0,001
0,85	0,85	1,459	20,750	0,029	0,023	0,001	0,044	0,002	0,001
0,75	0,95	-12,871	39,759	0,036	0,034	0,018	0,159	0,002	0,002
0,75	0,90	-7,867	39,790	0,039	0,033	0,008	0,159	0,002	0,002
0,75	0,85	-2,089	39,601	0,043	0,033	0,002	0,158	0,003	0,002

Ao especificarmos θ incorretamente, obtemos vício relativo absoluto menor no método proposto em 36 dos 48 cenários simulados, sendo observado o oposto em 10 casos nos quais θ^* ou θ são iguais a 0,75.

Quando os valores de θ^* são menores que θ, obtemos vícios relativos absolutos maiores que aqueles obtidos quando especificamos valores maiores que o simulado.

Observa-se também que houve superestimação do parâmetro da distribuição do tempo de falha (vício relativo maior que zero), com exceção para os casos em que a sensibilidade simulada foi igual a 0,75 e a especificada igual a 0,90 ou 0,95.

Como quando os valores de θ e ϕ eram bem especificados, no geral, observa-se que o modelo proposto apresentou maiores valores de erro padrão e de variância assintótica que o modelo proposto. Já os valores de EQM foram menores para o modelo proposto.

Supondo agora que não conhecemos os valores reais de θ e ϕ, alguns valores serão especificados (mal especificados) para ambos, novamente para distribuição do tempo de falha exponencial(1) com tamanho de amostra igual a 1200. Nas Tabelas 4.17 a 4.18 apresentamos alguns resultados. No Apêndice *D* são apresentados os resultados de outros cenários nos quais os valores de θ e ϕ são ambos mal especificados.

Tabela 4.17: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ=0,95 e ϕ=0,75

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,90	67,536	59,924	0,037	0,034	0,458	0,360	0,003	0,002
0,95	0,85	77,626	60,593	0,042	0,034	0,604	0,368	0,003	0,002
0,95	0,75	103,954	59,779	0,061	0,036	1,084	0,359	0,005	0,002
0,90	0,90	56,834	59,608	0,035	0,032	0,324	0,356	0,003	0,002
0,90	0,85	66,085	59,940	0,042	0,034	0,438	0,360	0,003	0,002
0,90	0,75	92,082	59,866	0,056	0,033	0,851	0,359	0,005	0,002
0,85	0,90	45,587	59,517	0,036	0,033	0,209	0,355	0,003	0,002
0,85	0,85	54,116	59,608	0,038	0,031	0,294	0,356	0,003	0,002
0,85	0,75	79,787	60,127	0,058	0,035	0,640	0,363	0,005	0,002

Tabela 4.18: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ=0,85 e ϕ=0,75

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	47,234	47,306	0,042	0,042	0,225	0,226	0,002	0,002
0,95	0,90	54,446	47,844	0,035	0,033	0,298	0,230	0,002	0,002
0,95	0,75	84,852	47,774	0,053	0,033	0,723	0,229	0,004	0,002
0,90	0,95	37,833	47,937	0,033	0,033	0,144	0,231	0,002	0,002
0,90	0,90	43,926	47,657	0,031	0,029	0,194	0,228	0,002	0,002
0,90	0,75	73,305	47,914	0,051	0,032	0,540	0,231	0,004	0,002
0,85	0,95	27,046	47,913	0,033	0,033	0,074	0,231	0,002	0,002
0,85	0,90	32,778	47,610	0,036	0,033	0,109	0,228	0,002	0,002
0,85	0,75	61,077	48,091	0,049	0,031	0,375	0,232	0,004	0,002

Tabela 4.19: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando θ=0,75 e ϕ=0,75

		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ^*	θ^*	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	34,472	34,564	0,037	0,037	0,120	0,121	0,002	0,002
0,95	0,90	40,631	35,086	0,033	0,031	0,166	0,124	0,002	0,002
0,95	0,85	47,821	35,521	0,039	0,033	0,230	0,127	0,002	0,002
0,90	0,95	24,952	35,069	0,030	0,030	0,063	0,124	0,002	0,002
0,90	0,90	30,103	34,857	0,029	0,027	0,091	0,122	0,002	0,002
0,90	0,85	36,628	35,039	0,036	0,030	0,135	0,124	0,002	0,002
0,85	0,95	14,264	35,133	0,032	0,032	0,021	0,124	0,002	0,002
0,85	0,90	19,029	34,832	0,034	0,032	0,037	0,122	0,002	0,002
0,85	0,85	24,991	34,844	0,035	0,030	0,064	0,122	0,002	0,002

Ao especificarmos θ^* e ϕ^* incorretamente, obtemos vício relativo absoluto menor no método proposto em apenas 37 dos 81 cenários simulados. Analisando as Tabelas acima e as Tabelas do Apêndice D, observou-se que a maior proporção de vício relativo absoluto maior para o modelo proposto é obtida quando $\theta - \theta^* > 0$ e $\phi - \phi^* < 0$ e a maior proporção de vício relativo absoluto menor para o modelo proposto acontece quando temos $\theta - \theta^* > 0$ e $\phi - \phi^* > 0$.

Em quase todos os cenários, o modelo proposto apresentou medidas de variabilidade maiores que o modelo convencional mas ambos apresentaram medidas pequenas de EQM, erro padrão e variância assintótica.

Capítulo 5

Extensões do Modelo Proposto

5.1 Modelo Weibull de Riscos Proporcionais

Considere novamente uma amostra de N indivíduos avaliados em intervalos de tempos através de um teste diagnóstico sujeito a erros de classificação. Suponha agora que, além do registro do tempo de falha/censura de cada indivíduo, temos também o vetor $\mathbf{x}_i$ de p covariáveis e considere que o tempo de falha tem distribuição Weibull(α, γ), podemos escrever:

$$S(t) = exp(-\alpha t^{\gamma})$$

e

$$\lambda(t) = \alpha\gamma t^{\gamma-1}.$$

Assim, temos que

$$\lambda(t|\mathbf{x}_i) = exp(\beta_1 x_{1i} + ... + \beta_p x_{pi})\lambda_0(t) = exp(\beta_1 x_{1i} + ... + \beta_p x_{pi})\alpha\gamma t^{\gamma-1}$$

Portanto,

$$T|\mathbf{x}_i \sim Weibull[exp(\beta_0 + \beta_1 x_i + ... + \beta_p x_{p_i}), \gamma]$$

em que $\beta_0 = log(\alpha)$.

Então

$$S(t|\mathbf{x}_i) = exp(-exp(\beta'\mathbf{x}_i)t^{\gamma}).$$

A função de verossimilhança é dada por:

$$L = \prod_{j=1}^{k}\prod_{\mu\in S_j}\left(\phi^{(j-1)}(1-\phi)S(t_j|\mathbf{x}_\mu) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}(S(t_{l-1}|\mathbf{x}_\mu) - S(t_l|\mathbf{x}_\mu))\right)^{n_{j,\mu}} \times$$
$$\times\left(\phi^{j}S(t_j|\mathbf{x}_\mu) + (1-\theta)\sum_{l=1}^{j}(S(t_{l-1}|\mathbf{x}_\mu) - S(t_l|\mathbf{x}_i))\phi^{(l-1)}(1-\theta)^{(j-l)}\right)^{m_{j,\mu}}$$

em que S_j é conjunto de observações com tempo de falha ou censura registrado em $(tj-1, tj]$.

Fazendo

$$a_{j,\mu} = \sum_{l=1}^{j} (S(t_{l-1}|\mathbf{x}_\mu) - S(t_l|\mathbf{x}_\mu))\phi^{(l-1)}(1-\theta)^{(j-l)}$$

e

$$b_{j,\mu} = S(t_j|\mathbf{x}_\mu)$$

podemos reescrever a função de verossimilhança como:

$$L = \prod_{j=1}^{k} \prod_{\mu \in S_j} \left(\phi^{(j-1)}(1-\phi)b_{j,\mu} + \theta a_{j,\mu}\right)^{n_{j,\mu}} \left(\phi^j b_{j,\mu} + (1-\theta)a_{j,\mu}\right)^{m_{j,\mu}}.$$

O logaritmo da função de verossimilhança é dado por:

$$l = \sum_{j=1}^{k} \sum_{\mu \in S_j} l1 + l2$$

em que

$$l1 = n_{j,\mu} log\left(\phi^{(j-1)}(1-\phi)b_{j,\mu} + \theta a_{j,\mu}\right)$$

e

$$l2 = m_{j,\mu} log\left(\phi^j b_{j,\mu} + (1-\theta)a_{j,\mu}\right).$$

Através do logaritmo da função de verossimilhança, utilizamos maximização para encontrarmos as estimativas dos coeficientes da regressão e, através das derivadas da função, calculamos a variância observada do estimador de β_1. Os cálculos para a variância dos $\hat{\beta}_s$ é demonstrado no apêndice E.

5.1.1 Avaliação do modelo proposto via simulação Monte Carlo

Para avaliar o desempenho do método proposto e compará-lo com método convencional, em que θ=1 e ϕ=1, foram realizadas simulações Monte Carlo para tempo de falha com distribuição Weibull(2,2) com amostras de 1200 indivíduos com uma única covariável binária (X=0 ou 1) e que foram avaliados nos tempos [0, 0.5, 1.0, 1.6, 2.0, 2.7, 3.0]. São apresentados os vícios relativos e as medidas de variabilidade para o coeficiente de regressão β_1 a seguir (Tabelas 5.7 a 5.11), em que β_1=1,5, ϕ e θ variam entre 0,7 e 0,95 e o percentual de falhas varia entre 0,6 e 0,95. Em cada cenário foram geradas 1000 amostras.

Tabela 5.1: Estimação no modelo com n=1200 e percentual de falha igual a 95%

		% Vicio γ		% Vicio β_0		% Vicio β_1		Erro padrão médio		Desvio padrão		EQM	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	1,647	-6,861	4,087	-16,969	5,144	-12,051	0,047	0,041	0,053	0,043	0,009	0,035
0,95	0,90	1,768	-7,005	4,118	-15,323	5,194	-15,037	0,046	0,039	0,058	0,043	0,009	0,053
0,95	0,80	1,937	-7,830	3,953	-11,834	4,705	-22,059	0,029	0,023	0,067	0,046	0,009	0,112
0,95	0,70	2,091	-9,441	4,200	-7,796	4,944	-29,487	0,010	0,006	0,088	0,054	0,013	0,198
0,90	0,95	1,924	-13,564	3,997	-36,646	5,240	-24,037	0,050	0,039	0,069	0,050	0,011	0,132
0,90	0,90	2,118	-13,808	4,195	-35,150	5,159	-27,042	0,051	0,039	0,074	0,048	0,012	0,167
0,90	0,80	2,221	-15,027	4,270	-32,180	5,022	-33,619	0,037	0,026	0,100	0,055	0,016	0,257
0,90	0,70	2,387	-16,947	3,794	-29,314	4,842	-41,053	0,009	0,005	0,112	0,053	0,018	0,382
0,80	0,95	2,338	-23,366	4,272	-68,197	5,324	-42,728	0,058	0,037	0,096	0,052	0,016	0,413
0,80	0,90	3,043	-23,696	4,425	-67,368	5,041	-45,994	0,061	0,037	0,119	0,058	0,020	0,479
0,80	0,80	2,529	-25,445	4,025	-65,065	4,575	-51,934	0,047	0,027	0,139	0,057	0,024	0,610
0,80	0,70	1,887	-27,760	2,200	-63,042	3,119	-58,954	0,015	0,008	0,188	0,060	0,037	0,786
0,70	0,95	2,572	-30,188	3,570	-93,960	4,205	-57,194	0,067	0,035	0,165	0,061	0,031	0,740
0,70	0,90	2,829	-30,781	3,691	-92,941	4,245	-59,710	0,072	0,035	0,150	0,051	0,027	0,805
0,70	0,80	4,058	-32,329	4,032	-90,855	5,619	-64,974	0,067	0,030	0,190	0,059	0,043	0,953
0,70	0,70	3,340	-34,927	4,791	-88,373	5,508	-71,316	0,023	0,010	0,255	0,059	0,072	1,148

Tabela 5.2: Estimação no modelo com n=1200 e percentual de falha igual a 90%

		% Vicio γ		% Vicio β_0		% Vicio β_1		Erro padrão médio		Desvio padrão		EQM	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	4,984	-5,499	10,009	-14,053	5,759	-13,403	0,027	0,024	0,046	0,036	0,010	0,042
0,95	0,90	4,822	-5,912	9,655	-12,666	5,817	-16,227	0,032	0,026	0,063	0,046	0,012	0,061
0,95	0,80	4,718	-6,833	9,199	-9,213	5,020	-23,023	0,031	0,023	0,075	0,048	0,011	0,122
0,95	0,70	4,922	-8,561	9,304	-5,545	5,803	-30,148	0,007	0,004	0,101	0,059	0,018	0,208
0,90	0,95	4,737	-13,320	9,593	-35,233	5,369	-26,436	0,030	0,025	0,064	0,044	0,011	0,159
0,90	0,90	4,751	-13,704	9,803	-33,641	5,948	-28,752	0,034	0,026	0,078	0,049	0,014	0,188
0,90	0,80	4,803	-14,863	9,688	-30,659	6,024	-34,723	0,032	0,021	0,100	0,053	0,018	0,274
0,90	0,70	5,708	-16,455	10,187	-27,333	5,730	-42,201	0,008	0,004	0,115	0,052	0,021	0,403
0,80	0,95	5,490	-23,791	11,196	-67,458	6,474	-45,278	0,035	0,025	0,112	0,053	0,022	0,464
0,80	0,90	5,721	-23,954	10,281	-66,314	5,913	-47,578	0,044	0,028	0,113	0,051	0,021	0,512
0,80	0,80	5,193	-25,812	9,828	-64,322	5,886	-53,345	0,038	0,021	0,146	0,054	0,029	0,643
0,80	0,70	6,063	-27,831	10,604	-61,907	7,272	-59,511	0,014	0,007	0,186	0,054	0,046	0,800
0,70	0,95	4,917	-30,944	9,356	-93,946	5,166	-59,653	0,047	0,026	0,152	0,053	0,029	0,803
0,70	0,90	5,208	-31,537	11,101	-92,289	7,441	-61,049	0,053	0,027	0,175	0,053	0,043	0,841
0,70	0,80	5,316	-33,008	9,860	-90,303	5,830	-66,598	0,051	0,023	0,221	0,059	0,056	1,001
0,70	0,70	7,973	-35,029	11,955	-87,988	8,117	-72,396	0,024	0,009	0,292	0,058	0,100	1,183

Tabela 5.3: Estimação no modelo com n=1200 e percentual de falha igual a 80%

		% Vicio γ		% Vicio β_0		% Vicio β_1		Erro padrão médio		Desvio padrão		EQM	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	1,648	-8,467	16,648	-9,048	3,212	-16,141	0,007	0,007	0,059	0,047	0,006	0,061
0,95	0,90	1,976	-8,242	16,850	-7,081	3,604	-18,164	0,008	0,008	0,059	0,044	0,006	0,076
0,95	0,80	2,105	-8,660	16,468	-3,625	3,474	-23,641	0,009	0,007	0,083	0,056	0,010	0,129
0,95	0,70	2,675	-9,666	17,130	0,552	4,302	-29,828	0,005	0,002	0,105	0,058	0,015	0,204
0,90	0,95	2,012	-15,734	16,997	-30,969	3,613	-29,098	0,007	0,008	0,070	0,048	0,008	0,193
0,90	0,90	1,942	-15,935	16,726	-29,556	3,564	-31,285	0,008	0,009	0,078	0,048	0,009	0,223
0,90	0,80	2,406	-16,705	17,566	-26,330	4,451	-36,247	0,011	0,008	0,098	0,052	0,014	0,298
0,90	0,70	2,630	-17,986	16,335	-23,807	3,725	-42,821	0,006	0,003	0,121	0,059	0,018	0,416
0,80	0,95	2,453	-25,903	17,630	-65,177	4,204	-48,461	0,007	0,008	0,126	0,057	0,020	0,532
0,80	0,90	2,122	-26,376	16,750	-64,319	3,706	-50,613	0,010	0,009	0,123	0,053	0,018	0,579
0,80	0,80	2,425	-27,413	16,712	-62,081	4,565	-55,102	0,015	0,009	0,151	0,056	0,027	0,686
0,80	0,70	3,484	-28,996	17,513	-59,538	5,631	-60,560	0,008	0,004	0,195	0,060	0,045	0,829
0,70	0,95	2,492	-32,646	19,134	-91,248	5,310	-61,773	0,007	0,007	0,159	0,057	0,032	0,862
0,70	0,90	2,779	-33,111	18,583	-90,739	5,465	-63,936	0,011	0,008	0,182	0,060	0,040	0,923
0,70	0,80	4,464	-34,074	19,203	-88,513	6,282	-68,170	0,018	0,009	0,238	0,060	0,065	1,049
0,70	0,70	4,002	-36,231	19,213	-86,640	7,927	-73,105	0,013	0,005	0,271	0,059	0,087	1,206

Tabela 5.4: Estimação no modelo com n=1200 e percentual de falha igual a 70%

		% Vicio γ		% Vicio β_0		% Vicio β_1		Erro padrão médio		Desvio padrão		EQM	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	4,136	-6,664	21,478	-5,716	3,565	-16,928	0,006	0,007	0,053	0,044	0,006	0,066
0,95	0,90	4,257	-6,493	21,308	-3,915	4,028	-18,577	0,008	0,008	0,064	0,048	0,008	0,080
0,95	0,80	4,448	-6,797	20,321	-0,801	3,299	-24,137	0,011	0,009	0,074	0,051	0,008	0,134
0,95	0,70	4,834	-7,553	20,492	3,674	3,226	-29,937	0,007	0,004	0,096	0,057	0,011	0,205
0,90	0,95	4,480	-14,572	21,210	-29,252	3,671	-31,022	0,006	0,008	0,080	0,055	0,009	0,220
0,90	0,90	4,693	-14,751	21,612	-27,760	4,489	-32,773	0,009	0,010	0,092	0,056	0,013	0,245
0,90	0,80	4,883	-15,201	21,073	-24,646	3,496	-37,886	0,011	0,010	0,110	0,058	0,015	0,326
0,90	0,70	4,863	-16,680	19,657	-22,080	2,730	-43,949	0,007	0,004	0,136	0,063	0,020	0,439
0,80	0,95	6,326	-24,795	22,779	-64,224	4,715	-50,957	0,007	0,009	0,129	0,059	0,021	0,588
0,80	0,90	5,694	-25,437	22,799	-63,088	5,399	-52,540	0,011	0,011	0,142	0,057	0,027	0,624
0,80	0,80	6,025	-26,343	20,802	-61,375	3,857	-57,387	0,017	0,012	0,168	0,062	0,031	0,745
0,80	0,70	5,547	-28,234	19,198	-59,221	2,124	-62,980	0,010	0,005	0,190	0,058	0,037	0,896
0,70	0,95	7,818	-31,487	23,309	-91,231	5,228	-64,534	0,011	0,010	0,180	0,055	0,038	0,940
0,70	0,90	7,430	-32,322	24,474	-90,133	6,982	-65,969	0,016	0,011	0,219	0,066	0,059	0,983
0,70	0,80	8,815	-33,222	21,983	-88,753	5,508	-70,436	0,024	0,012	0,262	0,064	0,075	1,120
0,70	0,70	7,755	-35,532	20,857	-86,763	4,975	-75,208	0,020	0,008	0,342	0,063	0,122	1,277

Tabela 5.5: Estimação no modelo com n=1200 e percentual de falha igual a 60%

		% Vicio γ		% Vicio β_0		% Vicio β_1		Erro padrão médio		Desvio padrão		EQM	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	11,481	-3,895	37,172	2,307	-0,697	-24,878	0,011	0,014	0,059	0,047	0,004	0,141
0,95	0,90	11,642	-3,621	36,849	4,048	-0,822	-26,518	0,012	0,014	0,070	0,054	0,005	0,161
0,95	0,80	11,350	-3,853	35,494	7,438	-1,126	-30,117	0,014	0,012	0,086	0,056	0,008	0,207
0,95	0,70	11,443	-4,862	34,629	10,951	-1,951	-35,505	0,008	0,004	0,096	0,059	0,010	0,287
0,90	0,95	12,465	-13,393	38,763	-23,404	0,310	-39,512	0,011	0,014	0,086	0,054	0,007	0,354
0,90	0,90	12,042	-13,618	36,995	-22,806	0,177	-40,861	0,012	0,014	0,105	0,060	0,011	0,379
0,90	0,80	13,134	-13,515	36,411	-19,774	-0,704	-44,719	0,015	0,014	0,111	0,062	0,012	0,454
0,90	0,70	13,204	-14,991	36,526	-16,539	-0,116	-48,955	0,008	0,005	0,127	0,061	0,016	0,543
0,80	0,95	12,803	-25,663	38,652	-61,652	0,060	-59,289	0,012	0,013	0,148	0,058	0,022	0,794
0,80	0,90	13,653	-25,829	38,410	-61,000	0,514	-60,697	0,013	0,014	0,154	0,059	0,024	0,832
0,80	0,80	13,826	-26,566	38,173	-58,652	1,310	-63,457	0,018	0,013	0,191	0,065	0,037	0,910
0,80	0,70	14,580	-27,939	36,034	-56,715	-0,300	-67,680	0,014	0,008	0,217	0,065	0,047	1,035
0,70	0,95	15,411	-32,673	40,024	-89,560	1,430	-71,427	0,013	0,012	0,198	0,060	0,039	1,152
0,70	0,90	14,684	-33,043	37,223	-89,263	-1,175	-73,331	0,016	0,013	0,219	0,060	0,048	1,213
0,70	0,80	17,290	-33,936	40,877	-86,881	3,149	-75,541	0,026	0,013	0,253	0,062	0,066	1,288
0,70	0,70	16,291	-35,819	37,068	-85,338	-0,236	-79,480	0,022	0,008	0,322	0,062	0,103	1,425

Da análise dos resultados, percebemos a prevalência de superestimação pelo modelo proposto e subestimação pelo modelo convencional. Percebe-se que quanto maior o percentual de falhas, menor é o vício do modelo proposto. Analisando apenas o vício relativo absoluto para o coeficiente de regressão β_1, percebemos que em todos os cenários simulados o modelo proposto apresentou menor vício relativo e que este vício sempre superestimou o coeficiente. Para o coeficiente β_0 , conforme o percentual de falha diminuía, ocorreram cenários em que o modelo proposto apresentou maior vício relativo absoluto que o modelo convencional, mas nota-se que o vício do modelo proposto manteve-se estável enquanto que para o modelo convencional o vício crescia conforme o teste aplicado tinha pior qualidade (sensibilidade e especificidade pequenas). Sobre as medidas de variabilidade de β_1, ambos os modelos apresentaram baixos valores.

5.2 Modelo de Riscos Proporcionais de Cox

Seguindo o mesmo raciocínio apresentado na seção 4.1, a contribuição do i-ésimo indivíduo que tem vetor de covariáveis $\mathbf{x}_i$ e registro de falha em $(t_{j-1}, t_j]$ pode ser escrita como:

$$\phi^{(j-1)}(1-\phi)S(t_j|\mathbf{x}_i) + \theta \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[S(t_{l-1}|\ \mathbf{x}_i) - S(t_l|\ \mathbf{x}_i)]$$

e, utilizando os resultados (2.7) e (2.8), temos:

$$\phi^{(j-1)}(1-\phi)\prod_{q=1}^{j} {\gamma_q}^{exp(\beta'\ \mathbf{x}_i)} + \theta \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[1-{\gamma_l}^{exp(\beta'\ \mathbf{x}_i)}] \prod_{q=1}^{l-1} {\gamma_q}^{exp(\beta'\ \mathbf{x}_i)}.$$

Da mesma forma, a contribuição do i-ésimo indivíduo que tem vetor de covariáveis $\mathbf{x}_i$ e registro de censura em $(t_{j-1}, t_j]$ é:

$$\phi^j S(t_j|x_i) + (1-\theta) \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[S(t_{l-1}|x_i) - S(t_l|\ \mathbf{x}_i)],$$

podendo ser reescrita como:

$$\phi^j \prod_{q=1}^{j} {\gamma_q}^{exp(\beta'\ \mathbf{x}_i)} + (1-\theta) \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[1-{\gamma_l}^{exp(\beta'\ \mathbf{x}_i)}] \prod_{q=1}^{l-1} {\gamma_q}^{exp(\beta'\ \mathbf{x}_i)}.$$

Fazendo $\prod_{q=1}^{j} {\gamma_q}^{exp(\beta'\ \mathbf{x}_\mu)} = b_{j\mu}$ e $\sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[1-{\gamma_l}^{exp(\beta'\ \mathbf{x}_\mu)}] \prod_{q=1}^{l-1} {\gamma_q}^{exp(\beta'\ \mathbf{x}_\mu)} = a_{j\mu}$, a função de verossimilhança pode ser escrita como:

$$L = \prod_{j=1}^{k} \prod_{\mu \in S_j} \left\{ \phi^{(j-1)}(1-\phi) b_{j\mu} + \theta a_{j\mu} \right\}^{n_{j\mu}} \left\{ \phi^{j} b_{j\mu} + (1-\theta) a_{j\mu} \right\}^{m_{j\mu}},$$

em que $n_{j\mu}$ ($m_{j\mu}$) é o número de falhas (censuras) observadas em t_j dentre os indivíduos com covariáveis $\mathbf{x}_\mu$.

O logaritmo da função de verossimilhança é dado por:

$$l = \sum_{j=1}^{k} \sum_{\mu \in S_j} \left\{ n_{j\mu} log[\phi^{(j-1)}(1-\phi) b_{j\mu} + \theta a_{j\mu}] + q_{j\mu} log[\phi^{j} b_{j\mu} + (1-\theta) a_{j\mu}] \right\}. \tag{5.1}$$

As estimativas dos parâmetros $\Omega = \binom{\lambda_0}{\beta}$ podem ser obtidas maximizando o logaritmo da função de verossimilhança (5.1). A variância do estimador de máxima verossimilhança obtido pode ser aproximada por $I(\hat{\Omega})^{-1}$, em que $I(\hat{\Omega})$ é a matriz de informação observada calculada no Apêndice D.

Por seu caráter semi-paramétrico, essa extensão proposta utiliza apenas a quantidade de avaliações dos indivíduos independente do tamanho dos intervalos entre avaliações. Trata-se, portanto, de um caso particular do modelo de Meier *et.at.* (2003), quando não há perda de visitas e não apresenta vantagens em relação a este. Apesar disso, o modelo será avaliado na próxima seção.

5.2.1 Avaliação do modelo via simulação Monte Carlo

Para avaliar o desempenho do método proposto e compará-lo com método convencional, em que θ=1 e ϕ=1, foram realizadas simulações Monte Carlo para vários valores de λ_0, β, θ e ϕ . As amostras eram constituídas por 120 ou 1200 indivíduos com uma única covariável binária (X=0 ou 1) e que foram avaliados nos tempos [0, 0.5, 1.0, 1.6, 2.0, 2.7, 3.0]. Alguns resultados são apresentados a seguir (Tabelas 5.1 a 5.6), em que β=1,5, $\lambda_0 = 1$ para todos os intervalos de tempo, ϕ e θ variam entre 0,7 e 0,95 e o percentual de falhas varia entre 0,6 e 0,95. Em cada cenário foram geradas 1000 amostras.

Tabela 5.6: Estimação no modelo de Riscos Proporcionais com n=1200 e percentual de falha igual a 95%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-3,317	-13,137	0,007	0,042	0,066	0,055	0,003	0,002
0,95	0,90	-3,913	-18,303	0,007	0,078	0,063	0,046	0,005	0,002
0,95	0,80	-5,555	-29,933	0,012	0,205	0,073	0,054	0,008	0,002
0,95	0,70	-7,768	-37,849	0,020	0,324	0,080	0,046	0,014	0,002
0,90	0,95	-7,832	-23,262	0,019	0,125	0,069	0,056	0,004	0,002
0,90	0,90	-7,616	-28,291	0,017	0,182	0,059	0,039	0,005	0,002
0,90	0,80	-9,618	-36,973	0,031	0,309	0,100	0,041	0,007	0,002
0,90	0,70	-3,780	-44,578	0,015	0,450	0,110	0,056	0,016	0,002
0,80	0,95	-4,617	-30,313	0,012	0,210	0,085	0,060	0,004	0,002
0,80	0,90	-3,479	-35,059	0,010	0,278	0,086	0,040	0,007	0,002
0,80	0,80	-1,415	-44,365	0,010	0,444	0,096	0,033	0,015	0,002
0,80	0,70	-8,098	-56,489	0,042	0,720	0,165	0,044	0,031	0,002
0,70	0,95	-1,034	-36,401	0,009	0,301	0,091	0,056	0,006	0,003
0,70	0,90	-7,677	-45,099	0,028	0,461	0,120	0,055	0,008	0,002
0,70	0,80	-8,991	-56,561	0,042	0,723	0,154	0,059	0,022	0,002
0,70	0,70	-10,342	-65,830	0,060	0,979	0,188	0,062	0,021	0,002

Tabela 5.7: Estimação no modelo de Riscos Proporcionais com n=1200 e percentual de falha igual a 80%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-14,688	-26,643	0,051	0,163	0,047	0,055	0,003	0,002
0,95	0,90	-22,800	-31,896	0,122	0,231	0,073	0,047	0,004	0,002
0,95	0,80	-24,341	-40,136	0,137	0,365	0,061	0,048	0,005	0,003
0,95	0,70	-16,200	-42,785	0,062	0,419	0,053	0,084	0,009	0,002
0,90	0,95	-17,757	-31,497	0,075	0,225	0,062	0,041	0,003	0,002
0,90	0,90	-19,947	-36,001	0,093	0,294	0,059	0,048	0,004	0,002
0,90	0,80	-16,290	-40,306	0,064	0,367	0,064	0,044	0,006	0,002
0,90	0,70	-14,996	-44,846	0,062	0,457	0,109	0,068	0,008	0,002
0,80	0,95	-12,422	-38,548	0,037	0,335	0,047	0,034	0,004	0,002
0,80	0,90	-19,898	-45,177	0,095	0,461	0,075	0,045	0,005	0,002
0,80	0,80	-10,556	-46,841	0,040	0,498	0,124	0,067	0,008	0,002
0,80	0,70	-19,873	-59,194	0,121	0,793	0,178	0,067	0,011	0,002
0,70	0,95	-18,695	-48,061	0,098	0,526	0,139	0,078	0,005	0,002
0,70	0,90	-19,963	-54,724	0,101	0,677	0,106	0,058	0,006	0,002
0,70	0,80	-16,513	-59,010	0,089	0,788	0,166	0,068	0,014	0,002
0,70	0,70	-13,016	-65,045	0,065	0,957	0,164	0,071	0,019	0,002

Tabela 5.8: Estimação no modelo de Riscos Proporcionais com n=1200 e percentual de falha igual a 60%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-25,904	-33,534	0,172	0,268	0,144	0,124	0,003	0,003
0,95	0,90	-28,727	-41,323	0,233	0,407	0,217	0,153	0,004	0,003
0,95	0,80	-21,696	-37,195	0,147	0,326	0,202	0,123	0,005	0,003
0,95	0,70	-25,367	-47,915	0,170	0,526	0,159	0,097	0,007	0,004
0,90	0,95	-28,864	-39,705	0,225	0,360	0,193	0,075	0,005	0,003
0,90	0,90	-25,596	-40,747	0,176	0,379	0,169	0,074	0,005	0,003
0,90	0,80	-23,568	-44,490	0,150	0,454	0,159	0,092	0,006	0,003
0,90	0,70	-19,841	-53,591	0,116	0,660	0,165	0,116	0,008	0,003
0,80	0,95	-21,167	-46,936	0,123	0,498	0,149	0,046	0,005	0,003
0,80	0,90	-30,842	-54,473	0,224	0,672	0,100	0,063	0,005	0,003
0,80	0,80	-18,720	-52,639	0,118	0,628	0,198	0,070	0,007	0,002
0,80	0,70	-15,911	-57,175	0,081	0,739	0,154	0,055	0,011	0,003
0,70	0,95	-21,170	-57,381	0,119	0,744	0,133	0,059	0,006	0,002
0,70	0,90	-20,228	-59,240	0,113	0,793	0,143	0,058	0,009	0,003
0,70	0,80	-20,550	-64,092	0,128	0,929	0,180	0,071	0,010	0,003
0,70	0,70	-21,427	-69,944	0,140	1,105	0,190	0,063	0,018	0,002

Analisando os resultados, podemos notar que o houve subestimação do coeficiente β (vício relativo negativo) em todos os cenários para os dois métodos, mas a subestimação foi menor no método proposto (vício relativo absoluto menor). Além disso, novamente percebemos que quanto maior o percentual de falhas e pior a qualidade do teste diagnóstico, maior o vício relativo, principalmente no modelo proposto; no modelo convencional a influência foi menor.

Sobre a variabilidade, o erro padrão das estimativas e a variância do estimador se mostraram ligeiramente melhores no método convencional, sofreu um pequeno acréscimo nos dois métodos na medida em que sensibilidade e/ou especificidade diminuíam. Para essas quantidades, não foi observada influência da variação do percentual de falhas. O erro quadrático médio foi menor no método proposto em todos os cenários simulados, sofrendo pouca variação na medida em que a qualidade do teste diagnóstico diminui, enquanto no método convencional teve seu valor aumentado.

Nas Tabelas 5.4 a 5.6 a seguir, temos os resultados para N=120.

Tabela 5.9: Estimação no modelo de Riscos Proporcionais com n=120 e percentual de falha igual a 95%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-5,658	-15,095	0,024	0,060	0,131	0,095	0,050	0,020
0,95	0,90	1,652	-16,290	0,017	0,071	0,127	0,105	0,443	0,018
0,95	0,80	-1,205	-27,824	0,066	0,188	0,257	0,119	0,081	0,016
0,95	0,70	-14,008	-43,425	0,158	0,453	0,338	0,169	0,093	0,019
0,90	0,95	1,193	-15,714	0,016	0,066	0,127	0,102	0,041	0,021
0,90	0,90	-3,868	-25,967	0,024	0,162	0,144	0,102	0,070	0,019
0,90	0,80	5,810	-30,610	0,058	0,237	0,226	0,162	0,289	0,018
0,90	0,70	6,553	-41,895	0,251	0,422	0,492	0,165	0,071	0,020
0,80	0,95	1,449	-27,059	0,055	0,186	0,232	0,144	0,063	0,021
0,80	0,90	-3,417	-36,642	0,093	0,329	0,301	0,164	0,071	0,022
0,80	0,80	-15,593	-48,050	0,139	0,536	0,290	0,128	0,059	0,020
0,80	0,70	74,984	-45,065	1,993	0,481	0,853	0,156	2,879	0,018
0,70	0,95	-2,284	-37,289	0,087	0,339	0,293	0,161	0,100	0,021
0,70	0,90	10,283	-35,146	0,241	0,304	0,466	0,162	0,104	0,023
0,70	0,80	29,822	-46,956	0,695	0,525	0,704	0,170	0,222	0,020
0,70	0,70	56,152	-54,246	1,175	0,699	0,682	0,192	1,512	0,021

Tabela 5.10: Estimação no modelo de Riscos Proporcionais com n=120 e percentual de falha igual a 80%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-27,664	-37,696	0,218	0,347	0,215	0,164	0,048	0,028
0,95	0,90	-34,152	-40,062	0,294	0,372	0,178	0,103	0,044	0,025
0,95	0,80	-24,188	-38,052	0,205	0,357	0,271	0,177	0,059	0,034
0,95	0,70	-31,685	-48,978	0,285	0,555	0,243	0,122	0,106	0,028
0,90	0,95	-33,008	-42,891	0,267	0,430	0,148	0,128	0,036	0,022
0,90	0,90	-28,543	-43,935	0,243	0,461	0,244	0,163	0,061	0,037
0,90	0,80	-24,686	-48,429	0,200	0,546	0,252	0,135	0,188	0,025
0,90	0,70	-25,540	-49,751	0,203	0,578	0,237	0,145	0,119	0,028
0,80	0,95	-34,655	-51,073	0,344	0,607	0,271	0,140	0,059	0,031
0,80	0,90	-30,023	-50,507	0,302	0,602	0,314	0,168	0,065	0,026
0,80	0,80	-16,855	-51,403	0,194	0,633	0,360	0,197	0,116	0,028
0,80	0,70	-9,871	-54,889	0,301	0,708	0,528	0,173	0,166	0,027
0,70	0,95	-31,608	-56,429	0,331	0,746	0,326	0,172	0,077	0,033
0,70	0,90	-40,380	-62,203	0,486	0,915	0,345	0,211	0,116	0,036
0,70	0,80	-23,663	-59,135	0,258	0,824	0,364	0,192	0,239	0,021
0,70	0,70	-30,471	-67,886	0,425	1,088	0,465	0,227	0,257	0,025

Tabela 5.11: Estimação no modelo de Riscos Proporcionais com n=120 e percentual de falha igual a 60%

		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	-1,859	-15,205	0,125	0,133	0,352	0,285	0,040	0,031
0,95	0,90	-19,303	-25,972	0,208	0,218	0,352	0,257	0,048	0,022
0,95	0,80	-7,520	-21,762	0,267	0,200	0,504	0,305	0,059	0,040
0,95	0,70	-13,817	-33,980	0,172	0,329	0,360	0,263	0,077	0,029
0,90	0,95	-12,071	-33,136	0,175	0,339	0,377	0,304	0,054	0,043
0,90	0,90	-11,733	-29,483	0,084	0,222	0,230	0,162	0,051	0,021
0,90	0,80	16,736	-26,735	0,298	0,204	0,484	0,208	0,239	0,023
0,90	0,70	-7,947	-41,343	0,256	0,433	0,491	0,221	0,122	0,030
0,80	0,95	9,524	-34,857	0,214	0,302	0,440	0,170	0,112	0,032
0,80	0,90	-8,726	-40,831	0,266	0,407	0,499	0,178	0,129	0,030
0,80	0,80	24,133	-37,155	0,364	0,340	0,483	0,172	0,286	0,029
0,80	0,70	17,764	-46,175	0,339	0,507	0,518	0,166	0,164	0,027
0,70	0,95	8,383	-43,499	0,312	0,462	0,544	0,190	0,069	0,024
0,70	0,90	1,374	-45,504	0,165	0,494	0,405	0,168	0,077	0,029
0,70	0,80	105,668	-42,263	2,755	0,425	0,493	0,153	0,432	0,024
0,70	0,70	18,761	-57,474	0,478	0,787	0,631	0,210	0,222	0,018

Podemos notar que houve subestimação do coeficiente β (vício relativo negativo) em todos os cenários para o modelo convencional e superestimação do coeficiente pelo modelo proposto em apenas 35,5% dos cenários avaliados. O modelo proposto apresentou menores valores de vício relativo absoluto que o modelo convencional em todos os cenários com exceção dos caso em que tínhamos (i) ϕ=0,8, θ=0,7 e percentual de falhas igual a 0,95 e (ii) ϕ=0,7, θ=0,8 e percentual de falha igual a 0,6. Novamente, percebe-se que o erro padrão das estimativas e a variância do estimador se mostraram melhores no método convencional.

Capítulo 6

Aplicação: Análise da Recorrência de Úlcera

Collett (2003) apresenta um estudo para comparar dois tratamentos para úlcera, um novo (A) e um tratamento padrão (B). O ensaio clínico aleatório foi do tipo duplo cego. Após o tratamento, os pacientes deveriam retornar ao consultório depois de 6 e 12 meses para detectar por meio de endoscopia se houve recorrência da úlcera. O ensaio clínico foi realizado em vários países mas, assim como Collett (2003), analisaremos apenas os pacientes da Bélgica.

Um resultado positivo da endoscopia indica que a recorrência da úlcera ocorreu entre o último resultado negativo e o exame atual. Neste caso, registrava-se o tempo de falha do paciente e o mesmo deixava o estudo. Se o paciente não apresentasse recorrência até o 12 º mês, ele entraria no estudo com tempo de censura igual a 12. Dos 43 pacientes, 6 apresentaram sintomas entre os períodos de avaliação e por isso foram submetidos à endoscopia. Após 12 meses de acompanhamento, 32 pacientes não tiveram recorrência de úlcera, resultando em 25,6% de falha.

Para o estudo foram registradas as covariáveis idade (em anos), duração da doença (categorizada em 1 se o tempo em que o paciente teve úlcera é inferior a 5 anos e em 2 se o tempo em que o paciente teve úlcera é igual ou maior a 5 anos), tratamento (A=0 e B=1) e resultado (1=recorrência da úlcera, 0=não recorrência).

Collett (2003) ajustou modelos que consideravam apenas as covariáveis idade, duração e tratamento, cada uma separadamente e, em seguida, os modelos com as combinações destas covariáveis. Através da análise de *Deviance* concluiu-se que o melhor ajuste foi obtido pelo modelo que continha apenas a covariável Tratamento. O coeficiente estimado para essa covariável foi igual a 0,378 com erro padrão de 0,629. Desta forma, a razão entre o risco de recorrência antes de 12 meses em um paciente do tratamento B em relação a um paciente do tratamento A é dada por $exp(0,378) = 1,46$, ou seja, o risco de recorrência no ano seguinte ao tratamento é maior para um paciente do tratamento B se comparado com um paciente do tratamento A.

6.1 Ajuste do modelo

Primeiramente ajustamos o Modelo de Riscos Proporcionais de Cox para este banco de dados e, em seguida, o modelo Weibull de Riscos Proporcionais proposto e observamos que o mesmo não foi bem ajustados a este banco de dados. A falta de eficiência do modelo Weibull de Riscos Proporcionais ao conjunto de dados pode ser devida ao fato do pequeno tamanho amostral e ao fato de termos apenas dois tempos avaliados. Assim, apresentaremos apenas os resultados do ajuste do

Modelo de Riscos Proporcionais de Cox. Para a aplicação do Modelo de Riscos Proporcionais de Cox neste banco de dados, utilizaremos apenas a covariável tratamento uma vez que foi considerada a mais significativa e os pacientes que apresentaram sintomas fora dos períodos de avaliação (6 e 12) foram excluídos pois o modelo proposto neste trabalho supõe que todos os indivíduos sejam avaliados nos mesmos tempos (dados grupados). Ao se excluir os pacientes que tiveram sintomas antes dos tempos estipulados de visitas, ficamos com um total de 37 observações onde apenas 5 apresentaram recorrência de úlcera, ou seja, temos um percentual de falha igual a 13,5%.

Os dados utilizados para a análise encontram-se na Tabela 6.1 a seguir:

Tabela 6.1: Dados de recorrência de úlcera após tratamento primário

Paciente	Tratamento	Tempo	Recorrência
2	1	12	0
3	1	12	0
4	1	12	0
5	0	12	0
6	0	12	0
7	1	12	0
8	0	12	0
9	1	12	0
11	1	12	0
12	1	12	0
13	0	12	0
14	0	6	1
15	1	6	1
16	0	6	0
19	0	12	0
20	1	12	0
21	0	12	0
22	0	12	0
23	1	12	0
24	0	12	0
25	1	12	0
28	0	6	0
29	0	12	0
30	1	12	1
31	1	12	0
32	1	12	1
33	0	12	0
34	0	6	0
35	1	12	0
36	1	12	0
37	0	12	1
38	1	12	0
39	1	12	0
40	0	12	0
41	1	12	0
42	0	12	0
43	1	12	0

O modelo foi ajustado através de função desenvolvida no *software R* fornecendo a estimativa de β (coeficiente da regressão para variável Tratamento), seu erro padrão e os valores de $exp(\hat{\beta})$ que é a estimativa da razão do risco do grupo que recebeu o tratamento B em relação ao grupo que recebeu o tratamento A. Os resultados obtidos para diferentes valores de sensibilidade e especificidade são apresentados nas Tabelas 6.2 e 6.3 a seguir.

Tabela 6.2: Estimativa do coeficiente da regressão do Modelo de Riscos Proporcionais para alguns valores de θ e ϕ supondo apenas um deles diferente de 1

ϕ	θ	$\hat{\beta}$	$ep(\hat{\beta})$	$exp(\hat{\beta})$
1,00	1,00	0,123	0,577	1,130
1,00	0,95	0,120	0,591	1,128
1,00	0,90	0,121	0,607	1,129
1,00	0,85	0,121	0,626	1,129
1,00	0,80	0,122	0,648	1,130
1,00	0,70	0,122	0,708	1,130
0,95	1,00	0,698	1,280	2,010
0,90	1,00	0,155	0,736	1,168
0,85	1,00	0,006	0,249	1,006
0,80	1,00	-0,007	0,042	0,993
0,70	1,00	-0,008	0,140	0,992

Tabela 6.3: Estimativa do coeficiente da regressão do Modelo de Riscos Proporcionais para alguns valores de θ e ϕ supondo ambos diferentes de 1

ϕ	θ	$\hat{\beta}$	$ep(\hat{\beta})$	$exp(\hat{\beta})$
0,95	0,95	0,697	1,311	2,007
0,95	0,90	0,700	1,347	2,013
0,95	0,80	0,705	1,441	2,024
0,95	0,70	0,706	1,594	2,026
0,90	0,95	0,156	0,768	1,168
0,90	0,90	0,153	0,812	1,165
0,90	0,80	0,155	0,941	1,168
0,90	0,70	0,155	1,311	1,168
0,80	0,95	-0,007	0,040	0,993
0,80	0,90	-0,007	0,037	0,993
0,80	0,80	-0,005	0,030	0,995
0,80	0,70	-0,004	0,019	0,996
0,70	0,95	-0,008	0,146	0,992
0,70	0,90	-0,007	0,153	0,993
0,70	0,80	-0,006	0,175	0,994
0,70	0,70	-0,005	0,225	0,995

Supondo que o exame de endoscopia não está sujeito a erros de diagnóstico, obtemos um coeficiente igual a 0,123 com um erro padrão de 0,577. Pode-se afirmar, então, que o risco de um paciente submetido ao tratamento padrão (B) ter recorrência de úlcera antes de 12 meses é 1,130 (=exp(0,123)) maior que o risco de um paciente submetido ao novo tratamento (A). Esse valor é menor que aquele obtido por Collett (2003) ao considerar o conjunto de dados com 43 observações.

Suponha agora que o exame de endoscopia não tenha sensibilidade perfeita, ou seja, há uma chance de diagnosticar um indivíduo com úlcera como sadio. Nesse caso, os resultados da Tabela 6.2 demonstram que a estimativa de β (e consequentemente de $\exp(\beta\ )$) sofre pouca variação em relação ao caso em que a sensibilidade é 100%. Por outro lado, quando o exame não tem especificidade igual a 1, ou seja, quando existe uma probabilidade não nula de que um indivíduo sadio seja diagnosticado com úlcera, a estimativa do coeficiente diminui na medida em que a especificidade diminui.

Se considerarmos que tanto a especificidade quanto a sensibilidade são menores que 1, ou seja, a endoscopia pode diagnosticar erroneamente tanto os pacientes sadios quanto aqueles com úlcera, resultados semelhantes são obtidos. Nos casos em que a especificidade é igual a 0,7 ou 0,8, temos que o risco do grupo B é aproximadamente 0,99 o risco do grupo A, ou seja, pacientes do grupo B tem menos risco de recorrência de úlcera.

A significativa influência da especificidade nessa aplicação pode ser explicada pelo pequeno tamanho amostral e pelo alto percentual de censuras presente nos dados.

Capítulo 7

Conclusões

A estimação da distribuição do tempo até a ocorrência de determinado evento ou a avaliação do risco do evento ocorrer na presença de certos fatores é de grande importância na área médica. Sabendo-se que a detecção dos eventos é realizada por exames que podem estar sujeitos a erros de classificação, é de grande interesse adaptar os métodos de análise de sobrevivência para os casos em que estes erros estão presentes. Nessa dissertação, apresentamos propostas já elaboradas para esse contexto e propomos uma nova metodologia.

Através de simulações Monte Carlo, comparamos o método proposto com o método convencional de análise, em que se considera que os testes diagnósticos têm sensibilidade e especificidade iguais a 1. Demonstramos que o modelo proposto neste trabalho é mais eficiente (menos viciado) que o método convencional na estimação dos parâmetros da distribuição do tempo de falha, tanto para a distribuição Exponencial quanto para a distribuição Weibull. Observou-se também que, na medida em que o percentual de censura aleatória aumenta, aumenta-se também o vício das estimativas, mas o método proposto é ainda superior ao convencional.

O método proposto foi adaptado para o Modelo Weibull de riscos proporcionais. Novamente percebe-se a maior eficiência do modelo proposto em relação ao convencional, independente do percentual de falhas.

Os modelos desenvolvidos nesse trabalho supõe que a sensibilidade e a especificidade do teste diagnóstico sejam conhecidas. Quando esses valores são mal especificados, demonstramos que estimativas bastante viciadas são obtidas pelo modelo proposto, sendo por vezes mais viciadas que aquelas produzidas pelo método convencional. Nesses casos, faz-se necessária a utilização de modelos que incorporem tais parâmetros para serem estimados.

No contexto analisado por Meier *et.al* (2003), Margaret (2008) desenvolveu uma abordagem para situações em que os erros são desconhecidos, mas tem-se disponível um conjunto de validação, ou seja, um subconjunto em que o estado verdadeiro dos indivíduos (falha ou censura) é conhecido. Quando esse subconjunto não está disponível, estudos preliminares demonstraram a ocorrência do problema de não identificabilidade, inviabilizando a derivação de estimadores de máxima verossimilhança. Soluções alternativas para esse problema são, portanto, de grande relevância e serão abordadas nos trabalhos futuros.

Além disso, outras extensões desse trabalho podem ser estudadas, como a abordagem bayesiana, o contexto em que os indivíduos são avaliados em diferentes tempos e o caso em que sensibilidade e especificidade variam com o tempo e/ou com outras covariáveis.

Referências

BALASUBRAMANIAN, R.; LAGAKOS, S.W. Estimation of Timing of Perinatal Transmission of HIV. *Biometrics*, V. 57, p. 1048-1058, 2001.

BENDER, R.; AUGUSTIN, T.; BLETTNER, M. Generating Survival Times to Simulate Cox Proportional Hazards Models. *Statist Med* 2005; 24:1713-1723.

BROSS, I. Misclassification in 2 X 2 Tables. *Biometrics*, p. 478-486, 1954.

COLLETT, D. *Modelling Survival Data in Medical Research*. London: Chapman Hall, 2003.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. São Paulo: Edgard Blucher, 2006.

GABA, A.; WINKLER, R. L. Implications of Errors in Survey Data: A Bayesian Model. *Management Science*, V. 38, p. 913-925, 1992.

HUNSBERG, S.; ALBERT, P. S.; DODD, L. Analysis of Progression-free Survival Data Using a Discrete Time Survival Model that Incorporates Measurements with and without Diagnostic Error. *Clinic Trials* 2010; 7:634-642.

JOHNSON, N. L., KOTZ, S., Wu, X. *Inspection Errors for Attributes in Quality Control*. London: Chapman Hall, 1991.

JOSEPH, L., GYORKOS, T. W., COUPAL, L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 141(3), 263-272, 1995.

KALBFLEISCH, D. J.; PRENTICE, L. R. *The Statistical Analysis of Failure Time Data*. New Jersey: Wiley & Sons, 2002.

MARGARET A. S. Incorporating Validation Subsets into Discrete Proportional Hazard Models for Mismeasured Outcomes. *Stat Med* 2008; 27:5456-70.

MCKEOWN, K.; JEWELL, N. P. Misclassification of Current Status Data. *Life Data Anal*, V. 16, p. 215-230,2010.

MEIER, A. S.; RICHARDSON, B. A.; HUGHES, J. P. Discrete Proportional Hazards Models for Mismeasured Outcomes. *Biometrics*,V. 59, p. 947-954, 2003.

PAGGIARO, A.; TORELLI, N. The Effect of Classification Errors in Survival Data Analysis. *SMA*, V. 13, p. 213-225, 2004.

PRENTICE, P.L.; GLOECKLER, L.A. Regression Analysis of Grouped Survival Data with Applications to Breast Cancer Data. *Biometrics*, V. 34, p. 57 -67, 1978.

SNAPINN, S. M. Survival Analysis with uncertain endpoints. *Biometrics*, v. 54, p. 209-218, 1998.

STRAPASSON, E. Comparação de Modelos com Censura Intervalar em Análise de Sobrevivência. *USP*, 2007.

Apêndice A

Cálculo da Variância do Estimador de Máxima Verossimilhança no Caso da Distribuição Weibull

Dado que o logaritmo da função de verossimilhança é:

$$l = \sum_{j=1}^{k} n_j log \left(\phi^{(j-1)}(1-\phi) exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + \theta \sum_{l=1}^{j} \left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)} \right) \phi^{(l-1)} (1-\theta)^{(j-l)} \right) +$$

$$+ m_j log \left(\phi^{j} exp^{\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)} + (1-\theta) \sum_{l=1}^{j} \left(exp^{\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)} - exp^{\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)} \right) \phi^{(l-1)} (1-\theta)^{(j-l)} \right).$$

A derivada primeira do logaritmo da função de verossimilhança é dada por:

$$\sum_{j=1}^{k} \frac{n_j \left(-\phi^{(j-1)}(1-\phi) exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) \left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + \theta \sum_{l=1}^{j} \phi^{(l-1)} (1-\theta)^{(j-l)} \left(exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) \left(\frac{t_l}{\alpha}\right)^{\gamma} ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) \left(\frac{t_{l-1}}{\alpha}\right)^{\gamma} ln\left(\frac{t_{l-1}}{\alpha}\right) \right) \right)}{\phi^{(j-1)}(1-\phi) exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta \sum_{l=1}^{j} \phi^{(l-1)} (1-\theta)^{(j-l)} \left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) \right)}$$

$$+ \frac{m_j \left(-\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) \left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + (1-\theta) \sum_{l=1}^{j} \phi^{(l-1)} (1-\theta)^{(j-l)} \left(exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) \left(\frac{t_l}{\alpha}\right)^{\gamma} ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right) \left(\frac{t_{l-1}}{\alpha}\right)^{\gamma} ln\left(\frac{t_{l-1}}{\alpha}\right) \right) \right)}{\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta) \sum_{l=1}^{j} \phi^{(l-1)} (1-\theta)^{(j-l)} \left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) \right)}.$$

Para o cálculo da variância assintótica, utilizou-se a matriz de Informação de Fisher obtida através das derivadas:

$$\frac{d(l)^2}{d(\gamma^2)} = \sum_{j=1}^{k} n_j \frac{d(g1)}{d(\gamma)} + m_j \frac{d(g2)}{d(\gamma)},$$

$$\frac{d(l)^2}{d(\alpha^2)} = \sum_{j=1}^{k} n_j \frac{d(g3)}{d(\alpha)} + m_j \frac{d(g4)}{d(\alpha)} e$$

$$\frac{d(l)^2}{d(\gamma)d(\alpha)} = \sum_{j=1}^{k} n_j \frac{d(g1)}{d(\alpha)} + m_j \frac{d(g2)}{d(\alpha)}.$$

Em que

$$g1 = \frac{\left(-\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\left(\frac{t_l}{\alpha}\right)^{\gamma} ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right)\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma} ln\left(\frac{t_{l-1}}{\alpha}\right)\right)\right)}{\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\right)},$$

$$g2 = \frac{\left(-\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + (1-\theta)\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\left(\frac{t_l}{\alpha}\right)^{\gamma} ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_{(l-1)}}{\alpha}\right)^{\gamma}\right)\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma} ln\left(\frac{t_{l-1}}{\alpha}\right)\right)\right)}{\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\right)},$$

$$g3 = \frac{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) {t_j}^{\gamma}\gamma\alpha^{-(\gamma+1)} + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) {t_{l-1}}^{\gamma}\gamma\alpha^{-(\gamma+1)} - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) {t_l}^{\gamma}\gamma\alpha^{-(\gamma+1)}\right)\right)}{\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\right)},$$

e

$$g4 = \frac{\left(\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) {t_j}^{\gamma}\gamma\alpha^{-(\gamma+1)} + (1-\theta)\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) {t_{l-1}}^{\gamma}\gamma\alpha^{-(\gamma+1)} - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right) {t_l}^{\gamma}\gamma\alpha^{-(\gamma+1)}\right)\right)}{\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)\sum_{l=1}^{j}\phi^{(l-1)}(1-\theta)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^{\gamma}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^{\gamma}\right)\right)}.$$

Assim, teremos:

$$\frac{d(g1)}{d(\gamma)} = \frac{\left(-\phi^{(j-1)}(1-\phi)\left(ln\left(\frac{t_j}{\alpha}\right)\right)^2\left(exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} - exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{2\gamma}\right) + \theta A1\right)\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B1\right)}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B1\right)^2} -$$

$$-\frac{\left(-\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + \theta C1\right)^2}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B1\right)^2},$$

$$\frac{d(g2)}{d(\gamma)} = \frac{\left(-\phi^{j}\left(ln\left(\frac{t_j}{\alpha}\right)\right)^2\left(exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} - exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{2\gamma}\right) + (1-\theta)A1\right)\left(\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B1\right)}{\left(\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B1\right)^2} -$$

$$-\frac{\left(-\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma} ln\left(\frac{t_j}{\alpha}\right) + (1-\theta)C1\right)^2}{\left(\phi^{j} exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B1\right)^2},$$

$$\frac{d(g3)}{d(\alpha)} = \frac{\left(\phi^{(j-1)}(1-\phi)(t_j)^{\gamma}\gamma\left(exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)(t_j)^{\gamma}\gamma\alpha^{-2(\gamma+1)} - (\gamma+1)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\alpha^{-(\gamma+2)}\right) + \theta A2\right)\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B2\right)}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B2\right)^2} -$$

$$-\frac{\left(\phi^{j-1}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)t_j{}^{\gamma}\gamma\alpha^{-(\gamma+1)} + \theta C2\right)^2}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B2\right)^2},$$

$$\frac{d(g4)}{d(\alpha)} = \frac{\left(\phi^{j}(t_j)^{\gamma}\gamma\left(exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)(t_j)^{\gamma}\gamma\alpha^{-2(\gamma+1)} - (\gamma+1)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\alpha^{-(\gamma+2)}\right) + (1-\theta)A2\right)\left(\phi^{j}exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B2\right)}{\left(\phi^{j}exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B2\right)^2} -$$

$$-\frac{\left(\phi^{j}exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)t_j{}^{\gamma}\gamma\alpha^{-(\gamma+1)} + (1-\theta)C2\right)^2}{\left(\phi^{j}exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + (1-\theta)B2\right)^2},$$

$$\frac{d(g1)}{d(\alpha)} = \frac{\left(-\phi^{(j-1)}(1-\phi)\left(exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)(t_j)^{\gamma}\gamma\alpha^{-(\gamma+1)}\left(\frac{t_j}{\alpha}\right)^{\gamma}ln\left(\frac{t_j}{\alpha}\right) - exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\gamma t_j{}^{\gamma}\alpha^{-(\gamma+1)}ln\left(\frac{t_j}{\alpha}\right) + exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma}\frac{\alpha}{t_j}\right) + \theta A3\right)}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B3\right)^2} \times$$

$$\times\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B3\right) -$$

$$-\frac{\left(-\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)\left(\frac{t_j}{\alpha}\right)^{\gamma}ln\left(\frac{t_j}{\alpha}\right) + \theta C3\right)\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right)t_j^{\gamma}\gamma\alpha^{-(\gamma+1)} - \theta C3\right)}{\left(\phi^{(j-1)}(1-\phi)exp\left(-\left(\frac{t_j}{\alpha}\right)^{\gamma}\right) + \theta B3\right)^2},$$

e

$$\frac{d(g2)}{d(\alpha)} = \frac{\left(-\phi^j \left(exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right)(t_j)^\gamma \gamma\alpha^{-(\gamma+1)}\left(\frac{t_j}{\alpha}\right)^\gamma ln\left(\frac{t_j}{\alpha}\right) - exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right)\gamma(t_j)^\gamma\alpha^{-(\gamma+1)} ln\left(\frac{t_j}{\alpha}\right) + exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right)\left(\frac{t_j}{\alpha}\right)^\gamma \frac{\alpha}{t_j}\right) + (1-\theta)A3\right)}{\left(\phi^j exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right) + (1-\theta)B3\right)^2} \times$$

$$\times \left(\phi^j exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right) + (1-\theta)B3\right) -$$

$$-\frac{\left(-\phi^j exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right)\left(\frac{t_j}{\alpha}\right)^\gamma ln\left(\frac{t_j}{\alpha}\right) + (1-\theta)C3\right)\left(\phi^j exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right)(t_j)^\gamma\gamma\alpha^{-(\gamma+1)} - \theta C3\right)}{\left(\phi^j exp\left(-\left(\frac{t_j}{\alpha}\right)^\gamma\right) + (1-\theta)B3\right)^2}.$$

Em que:

$$A1 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\{\left[\left(ln\left(\frac{t_l}{\alpha}\right)\right)^2\left(exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\left(\frac{t_l}{\alpha}\right)^\gamma - exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\left(\frac{t_l}{\alpha}\right)^{2\gamma}\right)\right] -$$
$$-\left[\left(ln\left(\frac{t_{l-1}}{\alpha}\right)\right)^2\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\left(\frac{t_{l-1}}{\alpha}\right)^\gamma - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\left(\frac{t_{l-1}}{\alpha}\right)^{2\gamma}\right)\right]\}$$

$$B1 = B2 = B3 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\left[exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\right]$$

$$C1 = C3 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\left[exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\left(\frac{t_l}{\alpha}\right)^\gamma ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\left(\frac{t_{l-1}}{\alpha}\right)^\gamma ln\left(\frac{t_{l-1}}{\alpha}\right)\right]$$

$$A2 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\{\left[(t_{l-1})^\gamma\gamma\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\gamma\alpha^{-2(\gamma+1)} - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)(\gamma+1)\alpha^{-(\gamma+2)}\right)\right] -$$
$$\left[(t_l)^\gamma\gamma\left(exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\left(\frac{t_l}{\alpha}\right)^\gamma\gamma\alpha^{-2(\gamma+1)} - exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)(\gamma+1)\alpha^{-(\gamma+2)}\right)\right]\}$$

$$C2 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)(t_{l-1})^\gamma\gamma\alpha^{-(\gamma+1)} - exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)(t_l)^\gamma\gamma\alpha^{-(\gamma+1)}\right)$$

$$A3 = \sum_{l=1}^{k} \phi^{(l-1)}(1-\phi)^{(j-l)}\{\left[\left(exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)(t_l)^\gamma\gamma\alpha^{-(\gamma+1)}\left(\frac{t_l}{\alpha}\right)^\gamma ln\left(\frac{t_l}{\alpha}\right) - exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\gamma(t_l)^\gamma\alpha^{-(\gamma+1)} ln\left(\frac{t_l}{\alpha}\right) + exp\left(-\left(\frac{t_l}{\alpha}\right)^\gamma\right)\left(\frac{t_l}{\alpha}\right)^\gamma\frac{\alpha}{t_l}\right)\right] -$$
$$\left[\left(exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)(t_{l-1})^\gamma\gamma\alpha^{-(\gamma+1)}\left(\frac{t_{l-1}}{\alpha}\right)^\gamma ln\left(\frac{t_{l-1}}{\alpha}\right) - exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\gamma(t_{l-1})^\gamma\alpha^{-(\gamma+1)} ln\left(\frac{t_{l-1}}{\alpha}\right) + exp\left(-\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\right)\left(\frac{t_{l-1}}{\alpha}\right)^\gamma\frac{\alpha}{t_{l-1}}\right)\right]\}$$

Apêndice B

Análise de Efeito do Percentual de Falhas

Resultados para a eficiência dos modelos com diferentes percentuais de falha para a distribuição Exponencial.

Tabela B.1: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 95%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-3,311	-6,106	0,010	0,009	0,001	0,004	0,001	0,001
1,00	0,90	-1,092	-7,259	0,008	0,007	0,000	0,005	0,001	0,001
1,00	0,80	0,207	-12,844	0,010	0,007	0,000	0,017	0,001	0,001
0,95	1,00	2,239	12,599	0,016	0,016	0,001	0,016	0,001	0,001
0,95	0,95	3,395	10,051	0,012	0,011	0,001	0,010	0,001	0,001
0,95	0,90	3,805	6,669	0,013	0,011	0,002	0,005	0,001	0,001
0,95	0,80	3,696	-1,232	0,017	0,013	0,002	0,000	0,001	0,001
0,90	1,00	2,174	23,313	0,017	0,017	0,001	0,055	0,001	0,001
0,90	0,95	2,776	19,852	0,019	0,018	0,001	0,040	0,001	0,001
0,90	0,90	2,700	15,673	0,023	0,020	0,001	0,025	0,001	0,001
0,90	0,80	2,726	7,167	0,025	0,018	0,001	0,005	0,002	0,001
0,80	1,00	0,352	44,689	0,029	0,028	0,001	0,201	0,002	0,002
0,80	0,95	1,026	40,608	0,031	0,028	0,001	0,166	0,002	0,002
0,80	0,90	0,792	35,484	0,029	0,025	0,001	0,127	0,002	0,002
0,80	0,80	0,722	25,275	0,036	0,025	0,001	0,065	0,002	0,001

Tabela B.2: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 90%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-6,710	-9,459	0,011	0,010	0,005	0,009	0,001	0,001
1,00	0,90	-3,826	-9,871	0,010	0,009	0,002	0,010	0,001	0,001
1,00	0,80	-1,705	-14,529	0,012	0,009	0,000	0,021	0,001	0,001
0,95	1,00	3,001	13,348	0,018	0,018	0,001	0,018	0,001	0,001
0,95	0,95	3,906	10,555	0,013	0,012	0,002	0,011	0,001	0,001
0,95	0,90	4,231	7,083	0,014	0,012	0,002	0,005	0,001	0,001
0,95	0,80	4,523	-0,534	0,016	0,012	0,002	0,000	0,001	0,001
0,90	1,00	2,756	23,884	0,018	0,018	0,001	0,057	0,001	0,001
0,90	0,95	3,379	20,458	0,019	0,018	0,002	0,042	0,001	0,001
0,90	0,90	3,268	16,227	0,022	0,019	0,002	0,027	0,001	0,001
0,90	0,80	3,795	8,042	0,023	0,017	0,002	0,007	0,002	0,001
0,80	1,00	1,507	45,845	0,029	0,029	0,001	0,211	0,002	0,002
0,80	0,95	1,506	41,101	0,029	0,027	0,001	0,170	0,002	0,002
0,80	0,90	2,167	36,763	0,029	0,025	0,001	0,136	0,002	0,002
0,80	0,80	2,800	26,835	0,039	0,027	0,002	0,073	0,002	0,001

Tabela B.3: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 85%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-10,462	-13,151	0,012	0,011	0,011	0,017	0,001	0,001
1,00	0,90	-6,376	-12,372	0,011	0,010	0,004	0,015	0,001	0,001
1,00	0,80	-3,570	-16,425	0,010	0,008	0,001	0,027	0,001	0,001
0,95	1,00	5,727	16,073	0,033	0,033	0,004	0,027	0,001	0,001
0,95	0,95	8,222	14,628	0,012	0,012	0,007	0,022	0,001	0,001
0,95	0,90	8,664	11,030	0,015	0,013	0,008	0,012	0,001	0,001
0,95	0,80	8,550	2,606	0,018	0,013	0,008	0,001	0,001	0,001
0,90	1,00	7,409	28,537	0,018	0,018	0,006	0,082	0,001	0,001
0,90	0,95	7,702	24,528	0,021	0,020	0,006	0,061	0,001	0,001
0,90	0,90	7,995	20,385	0,022	0,019	0,007	0,042	0,002	0,001
0,90	0,80	8,053	11,320	0,024	0,018	0,007	0,013	0,002	0,001
0,80	1,00	6,794	51,154	0,030	0,030	0,006	0,263	0,002	0,002
0,80	0,95	7,156	46,378	0,032	0,029	0,006	0,216	0,002	0,002
0,80	0,90	7,091	41,038	0,034	0,029	0,006	0,169	0,002	0,002
0,80	0,80	7,331	30,138	0,041	0,029	0,007	0,092	0,003	0,001

Tabela B.4: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 65%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	0,631	-2,789	0,032	0,031	0,001	0,002	0,001	0,001
1,00	0,90	9,760	2,200	0,021	0,019	0,010	0,001	0,001	0,001
1,00	0,80	14,927	-1,313	0,016	0,013	0,023	0,000	0,001	0,001
0,95	1,00	19,669	29,908	0,015	0,015	0,039	0,090	0,001	0,001
0,95	0,95	20,001	25,772	0,015	0,014	0,040	0,067	0,002	0,001
0,95	0,90	19,932	21,093	0,018	0,015	0,040	0,045	0,002	0,001
0,95	0,80	19,779	11,290	0,021	0,016	0,040	0,013	0,002	0,001
0,90	1,00	19,611	40,589	0,021	0,021	0,039	0,165	0,002	0,002
0,90	0,95	19,771	35,884	0,022	0,021	0,040	0,129	0,002	0,002
0,90	0,90	19,538	30,636	0,025	0,021	0,039	0,094	0,002	0,001
0,90	0,80	19,624	20,061	0,029	0,020	0,039	0,041	0,002	0,001
0,80	1,00	19,848	64,042	0,034	0,033	0,041	0,411	0,002	0,002
0,80	0,95	19,821	58,213	0,036	0,033	0,041	0,340	0,003	0,002
0,80	0,90	19,593	51,906	0,036	0,030	0,040	0,270	0,003	0,002
0,80	0,80	19,431	38,894	0,042	0,028	0,040	0,152	0,003	0,002

Tabela B.5: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 50%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-31,308	-33,725	0,025	0,023	0,099	0,114	0,001	0,001
1,00	0,90	-22,967	-28,718	0,018	0,016	0,053	0,083	0,001	0,001
1,00	0,80	-14,077	-26,992	0,021	0,018	0,020	0,073	0,001	0,001
0,95	1,00	22,074	32,327	0,119	0,119	0,063	0,119	0,002	0,002
0,95	0,95	31,564	36,584	0,018	0,017	0,100	0,134	0,002	0,002
0,95	0,90	31,654	31,269	0,020	0,018	0,101	0,098	0,002	0,001
0,95	0,80	31,600	19,986	0,026	0,019	0,101	0,040	0,002	0,001
0,90	1,00	32,053	53,065	0,023	0,023	0,103	0,282	0,002	0,002
0,90	0,95	31,168	46,534	0,026	0,024	0,098	0,217	0,002	0,002
0,90	0,90	31,591	41,027	0,030	0,025	0,101	0,169	0,002	0,002
0,90	0,80	30,468	27,894	0,034	0,024	0,094	0,078	0,003	0,001
0,80	1,00	33,191	77,513	0,037	0,036	0,112	0,602	0,003	0,003
0,80	0,95	32,670	70,119	0,035	0,032	0,108	0,493	0,003	0,003
0,80	0,90	31,767	62,186	0,047	0,038	0,103	0,388	0,003	0,002
0,80	0,80	31,533	47,119	0,048	0,031	0,102	0,223	0,004	0,002

Tabela B.6: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) com n=1200 percentual de falha de 40%

Erros		% Vicio		EQM		Erro padrão		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-40,364	-42,371	0,036	0,034	0,164	0,181	0,001	0,001
1,00	0,90	-29,266	-34,430	0,022	0,021	0,086	0,119	0,001	0,001
1,00	0,80	-20,898	-32,645	0,020	0,016	0,044	0,107	0,001	0,001
0,95	1,00	21,739	32,105	0,138	0,137	0,066	0,122	0,002	0,002
0,95	0,95	31,817	36,890	0,017	0,016	0,102	0,136	0,002	0,002
0,95	0,90	31,608	31,349	0,021	0,018	0,100	0,099	0,002	0,002
0,95	0,80	30,749	19,569	0,025	0,018	0,095	0,039	0,002	0,001
0,90	1,00	32,287	53,477	0,026	0,026	0,105	0,287	0,002	0,002
0,90	0,95	32,288	47,698	0,024	0,022	0,105	0,228	0,002	0,002
0,90	0,90	31,753	41,323	0,025	0,021	0,101	0,171	0,002	0,002
0,90	0,80	31,728	29,067	0,036	0,026	0,102	0,085	0,003	0,001
0,80	1,00	34,242	78,739	0,037	0,036	0,119	0,621	0,003	0,003
0,80	0,95	34,131	71,675	0,041	0,037	0,118	0,515	0,003	0,003
0,80	0,90	33,506	63,902	0,046	0,038	0,114	0,410	0,003	0,002
0,80	0,80	32,074	47,775	0,054	0,035	0,106	0,229	0,004	0,002

Resultados para a eficiência dos modelos com diferentes percentuais de falha para a distribuição Weibull.

Tabela B.7: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=1200 percentual de falha de 90%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	20,000	19,428	0,020	0,020	0,160	0,151	0,006	0,005
1,00	0,90	18,904	18,691	0,025	0,023	0,144	0,140	0,006	0,005
1,00	0,80	15,106	16,720	0,037	0,033	0,093	0,113	0,006	0,006
0,95	1,00	14,936	-2,493	0,047	0,031	0,091	0,003	0,008	0,004
0,95	0,95	15,680	-2,398	0,050	0,032	0,101	0,003	0,009	0,004
0,95	0,90	15,767	-2,335	0,059	0,035	0,103	0,003	0,010	0,004
0,95	0,80	14,350	-2,667	0,066	0,039	0,087	0,004	0,011	0,005
0,90	1,00	14,867	-12,989	0,068	0,034	0,093	0,069	0,012	0,004
0,90	0,95	15,934	-13,148	0,078	0,035	0,108	0,070	0,013	0,004
0,90	0,90	15,767	-13,304	0,083	0,038	0,106	0,072	0,014	0,004
0,90	0,80	14,335	-14,069	0,095	0,037	0,091	0,081	0,016	0,005
0,80	1,00	12,232	-26,535	0,106	0,030	0,071	0,283	0,024	0,003
0,80	0,95	13,249	-26,682	0,124	0,032	0,086	0,286	0,026	0,003
0,80	0,90	13,281	-27,021	0,128	0,033	0,087	0,293	0,029	0,003
0,80	0,80	12,467	-27,861	0,139	0,034	0,081	0,312	0,036	0,004

Tabela B.8: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=1200 percentual de falha de 90%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-4,832	-3,142	0,012	0,012	0,009	0,004	0,001	0,001
1,00	0,90	-5,899	-2,317	0,007	0,007	0,014	0,002	0,001	0,001
1,00	0,80	-6,728	0,836	0,011	0,010	0,018	0,000	0,001	0,001
0,95	1,00	-6,202	-16,038	0,018	0,016	0,016	0,103	0,001	0,001
0,95	0,95	-6,475	-14,626	0,019	0,017	0,017	0,086	0,001	0,001
0,95	0,90	-6,645	-13,083	0,020	0,017	0,018	0,069	0,002	0,001
0,95	0,80	-6,766	-9,562	0,023	0,019	0,019	0,037	0,002	0,002
0,90	1,00	-5,797	-24,250	0,027	0,020	0,014	0,236	0,002	0,001
0,90	0,95	-5,934	-22,977	0,026	0,019	0,015	0,212	0,002	0,001
0,90	0,90	-6,221	-21,685	0,028	0,022	0,016	0,189	0,002	0,002
0,90	0,80	-6,251	-18,554	0,032	0,024	0,017	0,138	0,003	0,002
0,80	1,00	-5,332	-38,662	0,040	0,022	0,013	0,598	0,004	0,001
0,80	0,95	-5,491	-37,682	0,045	0,025	0,014	0,569	0,004	0,001
0,80	0,90	-5,839	-36,804	0,044	0,025	0,016	0,542	0,004	0,001
0,80	0,80	-6,117	-34,627	0,049	0,026	0,017	0,480	0,005	0,002

Tabela B.9: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=1200 percentual de falha de 70%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	13,117	12,879	0,031	0,032	0,070	0,067	0,005	0,005
1,00	0,90	7,926	8,156	0,034	0,033	0,026	0,028	0,005	0,005
1,00	0,80	3,008	5,170	0,038	0,034	0,005	0,012	0,005	0,005
0,95	1,00	-4,761	-14,922	0,045	0,031	0,011	0,090	0,006	0,004
0,95	0,95	-4,565	-14,900	0,042	0,028	0,010	0,090	0,007	0,004
0,95	0,90	-4,358	-14,497	0,043	0,028	0,009	0,085	0,007	0,004
0,95	0,80	-5,010	-14,029	0,051	0,032	0,013	0,080	0,008	0,005
0,90	1,00	-5,662	-22,266	0,050	0,025	0,015	0,199	0,009	0,004
0,90	0,95	-5,306	-22,239	0,057	0,028	0,015	0,199	0,010	0,004
0,90	0,90	-4,958	-22,009	0,059	0,028	0,013	0,195	0,010	0,004
0,90	0,80	-5,387	-21,756	0,073	0,034	0,017	0,190	0,012	0,005
0,80	1,00	-8,484	-31,807	0,078	0,027	0,035	0,405	0,019	0,003
0,80	0,95	-7,365	-31,699	0,093	0,030	0,030	0,403	0,021	0,003
0,80	0,90	-7,662	-31,873	0,094	0,029	0,032	0,407	0,024	0,003
0,80	0,80	-6,993	-31,754	0,104	0,030	0,030	0,404	0,030	0,004

Tabela B.10: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=1200 percentual de falha de 70%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-5,260	-3,589	0,030	0,029	0,012	0,006	0,001	0,001
1,00	0,90	-9,382	-5,780	0,018	0,018	0,036	0,014	0,001	0,001
1,00	0,80	-11,380	-3,714	0,013	0,013	0,052	0,006	0,001	0,002
0,95	1,00	-13,722	-22,747	0,021	0,018	0,076	0,207	0,002	0,001
0,95	0,95	-14,075	-21,460	0,018	0,016	0,080	0,184	0,002	0,002
0,95	0,90	-14,238	-19,899	0,019	0,016	0,081	0,159	0,002	0,002
0,95	0,80	-14,225	-16,197	0,022	0,018	0,081	0,105	0,002	0,002
0,90	1,00	-13,268	-30,341	0,027	0,019	0,071	0,369	0,002	0,001
0,90	0,95	-13,693	-29,250	0,028	0,020	0,076	0,343	0,003	0,002
0,90	0,90	-13,663	-27,747	0,028	0,020	0,075	0,308	0,003	0,002
0,90	0,80	-13,770	-24,516	0,031	0,023	0,077	0,241	0,003	0,002
0,80	1,00	-12,881	-43,519	0,043	0,021	0,068	0,758	0,005	0,001
0,80	0,95	-12,984	-42,505	0,043	0,023	0,069	0,723	0,005	0,001
0,80	0,90	-13,234	-41,558	0,047	0,023	0,072	0,691	0,006	0,001
0,80	0,80	-13,186	-38,946	0,049	0,024	0,072	0,607	0,007	0,002

Tabela B.11: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=1200 percentual de falha de 60%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	1,931	1,929	0,049	0,049	0,004	0,004	0,004	0,004
1,00	0,90	-4,678	-4,002	0,040	0,039	0,010	0,008	0,004	0,004
1,00	0,80	-10,628	-7,850	0,034	0,032	0,046	0,026	0,005	0,005
0,95	1,00	-19,016	-24,863	0,034	0,025	0,146	0,248	0,006	0,004
0,95	0,95	-18,975	-24,676	0,027	0,019	0,145	0,244	0,007	0,004
0,95	0,90	-18,674	-24,051	0,032	0,023	0,141	0,232	0,007	0,005
0,95	0,80	-18,912	-23,028	0,039	0,026	0,145	0,213	0,008	0,006
0,90	1,00	-20,528	-30,090	0,038	0,022	0,170	0,363	0,009	0,004
0,90	0,95	-20,272	-29,828	0,040	0,023	0,166	0,356	0,010	0,004
0,90	0,90	-19,610	-29,266	0,045	0,025	0,156	0,343	0,010	0,004
0,90	0,80	-19,690	-28,504	0,057	0,030	0,158	0,326	0,012	0,005
0,80	1,00	-23,682	-37,007	0,057	0,024	0,228	0,548	0,035	0,003
0,80	0,95	-23,036	-36,732	0,066	0,027	0,217	0,540	0,038	0,003
0,80	0,90	-22,885	-36,491	0,075	0,029	0,215	0,533	0,083	0,003
0,80	0,80	-22,111	-35,920	0,083	0,029	0,202	0,517	0,081	0,003

Tabela B.12: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=1200 percentual de falha de 60%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-7,558	-5,869	0,040	0,040	0,024	0,015	0,001	0,002
1,00	0,90	-12,892	-9,260	0,023	0,022	0,067	0,035	0,002	0,002
1,00	0,80	-15,889	-8,078	0,015	0,015	0,101	0,026	0,002	0,002
0,95	1,00	-19,945	-28,319	0,023	0,019	0,160	0,321	0,003	0,002
0,95	0,95	-20,265	-27,003	0,018	0,015	0,165	0,292	0,003	0,002
0,95	0,90	-20,297	-25,326	0,020	0,017	0,165	0,257	0,003	0,002
0,95	0,80	-20,220	-21,460	0,022	0,018	0,164	0,185	0,004	0,003
0,90	1,00	-19,542	-35,443	0,027	0,019	0,153	0,503	0,004	0,002
0,90	0,95	-19,899	-34,287	0,028	0,020	0,159	0,471	0,004	0,002
0,90	0,90	-19,720	-32,663	0,027	0,019	0,156	0,427	0,004	0,002
0,90	0,80	-19,812	-29,330	0,030	0,021	0,158	0,345	0,005	0,003
0,80	1,00	-19,044	-47,696	0,041	0,020	0,147	0,910	0,015	0,001
0,80	0,95	-18,983	-46,575	0,044	0,022	0,146	0,868	0,016	0,001
0,80	0,90	-19,345	-45,556	0,048	0,024	0,152	0,831	0,037	0,001
0,80	0,80	-19,040	-42,709	0,049	0,023	0,147	0,730	0,034	0,002

Tabela B.13: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=120 percentual de falha de 90%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	19,925	19,217	0,057	0,054	0,162	0,151	0,058	0,055
1,00	0,90	19,105	18,832	0,075	0,067	0,152	0,146	0,059	0,055
1,00	0,80	17,912	19,345	0,107	0,094	0,140	0,158	0,066	0,061
0,95	1,00	15,344	-1,938	0,143	0,097	0,115	0,011	0,087	0,041
0,95	0,95	15,075	-2,646	0,157	0,099	0,116	0,013	0,092	0,042
0,95	0,90	15,432	-2,556	0,180	0,107	0,128	0,014	0,100	0,045
0,95	0,80	13,765	-3,091	0,214	0,125	0,122	0,019	0,113	0,055
0,90	1,00	14,276	-12,895	0,207	0,103	0,124	0,077	0,125	0,037
0,90	0,95	14,102	-13,658	0,224	0,101	0,130	0,085	0,134	0,038
0,90	0,90	15,402	-13,652	0,240	0,108	0,153	0,086	0,150	0,042
0,90	0,80	15,505	-13,834	0,301	0,123	0,187	0,092	0,178	0,054
0,80	1,00	12,004	-26,568	0,348	0,099	0,179	0,292	0,248	0,031
0,80	0,95	14,667	-26,678	0,383	0,101	0,233	0,295	0,315	0,033
0,80	0,90	13,482	-27,364	0,394	0,102	0,228	0,310	0,308	0,035
0,80	0,80	14,321	-28,076	0,517	0,112	0,349	0,328	0,391	0,043

Tabela B.14: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=120 percentual de falha de 90%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-3,631	-1,843	0,019	0,019	0,006	0,002	0,011	0,011
1,00	0,90	-4,481	-0,800	0,021	0,020	0,008	0,001	0,011	0,012
1,00	0,80	-4,731	2,962	0,031	0,029	0,010	0,004	0,013	0,014
0,95	1,00	-5,302	-15,212	0,058	0,050	0,015	0,095	0,013	0,012
0,95	0,95	-6,312	-14,415	0,062	0,054	0,020	0,086	0,015	0,013
0,95	0,90	-6,265	-12,667	0,059	0,052	0,019	0,067	0,016	0,015
0,95	0,80	-6,388	-9,125	0,070	0,060	0,021	0,037	0,020	0,022
0,90	1,00	-5,136	-23,731	0,079	0,060	0,017	0,229	0,019	0,012
0,90	0,95	-6,058	-23,009	0,090	0,066	0,023	0,216	0,020	0,014
0,90	0,90	-5,602	-21,257	0,085	0,064	0,020	0,185	0,023	0,017
0,90	0,80	-5,739	-18,082	0,098	0,074	0,023	0,136	0,028	0,026
0,80	1,00	-5,093	-38,589	0,128	0,072	0,027	0,601	0,039	0,011
0,80	0,95	-4,744	-37,389	0,135	0,073	0,027	0,564	0,054	0,013
0,80	0,90	-5,342	-36,718	0,144	0,074	0,032	0,545	0,039	0,015
0,80	0,80	-5,561	-34,482	0,161	0,085	0,038	0,483	0,043	0,022

Tabela B.15: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=120 percentual de falha de 70%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	0,968	1,283	0,055	0,052	0,003	0,003	0,049	0,048
1,00	0,90	1,273	1,970	0,073	0,069	0,006	0,006	0,050	0,048
1,00	0,80	0,255	2,635	0,080	0,070	0,006	0,008	0,054	0,052
0,95	1,00	-7,626	-15,709	0,092	0,070	0,032	0,104	0,059	0,036
0,95	0,95	-8,654	-16,469	0,097	0,069	0,039	0,113	0,063	0,038
0,95	0,90	-7,751	-15,526	0,121	0,085	0,039	0,104	0,068	0,040
0,95	0,80	-8,858	-15,311	0,143	0,098	0,052	0,103	0,081	0,050
0,90	1,00	-9,388	-22,542	0,123	0,077	0,050	0,209	0,080	0,033
0,90	0,95	-7,909	-21,887	0,154	0,088	0,049	0,199	0,089	0,035
0,90	0,90	-7,744	-22,031	0,148	0,085	0,046	0,201	0,098	0,037
0,90	0,80	-7,787	-21,394	0,189	0,091	0,060	0,191	0,120	0,046
0,80	1,00	-10,797	-31,195	0,192	0,084	0,084	0,396	0,138	0,028
0,80	0,95	-8,528	-30,581	0,257	0,096	0,095	0,383	0,222	0,028
0,80	0,90	-9,827	-31,158	0,264	0,097	0,108	0,398	0,186	0,030
0,80	0,80	-7,671	-30,580	0,300	0,105	0,113	0,385	0,166	0,036

Tabela B.16: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=120 percentual de falha de 70%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-5,645	-4,069	0,027	0,025	0,013	0,007	0,016	0,016
1,00	0,90	-9,506	-5,939	0,031	0,030	0,037	0,015	0,014	0,015
1,00	0,80	-11,736	-4,019	0,037	0,036	0,056	0,008	0,015	0,017
0,95	1,00	-16,943	-25,095	0,052	0,045	0,118	0,254	0,016	0,012
0,95	0,95	-18,180	-24,591	0,059	0,050	0,136	0,244	0,018	0,014
0,95	0,90	-18,130	-22,825	0,059	0,052	0,135	0,211	0,019	0,016
0,95	0,80	-18,211	-19,183	0,064	0,055	0,137	0,150	0,024	0,023
0,90	1,00	-17,797	-33,027	0,075	0,054	0,132	0,439	0,022	0,011
0,90	0,95	-17,557	-31,421	0,087	0,063	0,131	0,399	0,024	0,013
0,90	0,90	-17,689	-30,148	0,080	0,058	0,132	0,367	0,027	0,015
0,90	0,80	-17,638	-26,714	0,094	0,065	0,133	0,290	0,034	0,023
0,80	1,00	-17,468	-45,345	0,121	0,060	0,137	0,826	0,031	0,009
0,80	0,95	-17,519	-44,092	0,136	0,077	0,141	0,784	0,064	0,010
0,80	0,90	-17,784	-43,313	0,129	0,065	0,143	0,755	0,041	0,012
0,80	0,80	-16,794	-40,102	0,148	0,071	0,135	0,648	0,011	0,017

Tabela B.17: Desempenho do método proposto e do convencional na estimação do parâmetro de forma (α) da distribuição Weibull(2,2) com n=120 percentual de falha de 60%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	5,632	5,431	0,063	0,061	0,017	0,016	0,045	0,044
1,00	0,90	2,585	2,854	0,083	0,077	0,010	0,009	0,047	0,045
1,00	0,80	-1,549	0,740	0,093	0,084	0,010	0,007	0,053	0,053
0,95	1,00	-6,142	-16,705	0,115	0,074	0,028	0,117	0,062	0,039
0,95	0,95	-5,906	-16,588	0,109	0,071	0,026	0,115	0,068	0,042
0,95	0,90	-5,691	-16,224	0,133	0,083	0,031	0,112	0,076	0,048
0,95	0,80	-6,081	-15,777	0,163	0,099	0,041	0,109	0,096	0,076
0,90	1,00	-7,861	-24,393	0,164	0,079	0,052	0,244	0,091	0,037
0,90	0,95	-6,156	-23,893	0,199	0,091	0,055	0,237	0,104	0,040
0,90	0,90	-5,196	-23,567	0,218	0,096	0,058	0,231	0,122	0,048
0,90	0,80	-6,209	-23,415	0,236	0,102	0,071	0,230	0,187	0,070
0,80	1,00	-9,089	-33,556	0,316	0,085	0,133	0,458	0,097	0,032
0,80	0,95	-7,126	-33,037	0,333	0,090	0,132	0,445	0,678	0,032
0,80	0,90	-7,610	-33,450	0,356	0,092	0,150	0,456	0,297	0,035
0,80	0,80	-4,929	-32,838	0,449	0,107	0,211	0,443	0,355	0,041

Tabela B.18: Desempenho do método proposto e do convencional na estimação do parâmetro de escala (γ) da distribuição Weibull(2,2) com n=120 percentual de falha de 60%

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
ϕ	θ	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
1,00	0,95	-7,390	-5,705	0,031	0,031	0,023	0,014	0,014	0,015
1,00	0,90	-10,411	-6,782	0,027	0,027	0,044	0,019	0,015	0,016
1,00	0,80	-11,850	-4,114	0,032	0,031	0,057	0,008	0,019	0,022
0,95	1,00	-13,092	-22,536	0,056	0,046	0,072	0,205	0,021	0,017
0,95	0,95	-13,709	-21,421	0,057	0,047	0,078	0,186	0,022	0,019
0,95	0,90	-13,916	-19,896	0,063	0,053	0,081	0,161	0,025	0,024
0,95	0,80	-13,957	-16,254	0,069	0,057	0,083	0,109	0,035	0,046
0,90	1,00	-13,733	-31,130	0,088	0,059	0,083	0,391	0,030	0,016
0,90	0,95	-13,320	-29,454	0,088	0,062	0,079	0,351	0,033	0,018
0,90	0,90	-13,120	-27,860	0,094	0,067	0,078	0,315	0,040	0,025
0,90	0,80	-14,091	-25,207	0,106	0,071	0,091	0,259	0,078	0,043
0,80	1,00	-12,606	-44,101	0,153	0,075	0,087	0,784	0,003	0,012
0,80	0,95	-13,120	-43,135	0,139	0,071	0,088	0,749	0,284	0,013
0,80	0,90	-13,356	-42,312	0,143	0,070	0,092	0,721	0,076	0,015
0,80	0,80	-13,425	-39,627	0,151	0,075	0,095	0,634	0,085	0,021

Apêndice C

Má especificação dos valores de θ e de ϕ

Tabela C.1: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,95 e phi=0,95

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,90	0,90	3,964	16,864	0,015	0,013	0,002	0,029	0,001	0,001
0,90	0,85	8,986	16,779	0,016	0,013	0,008	0,028	0,002	0,001
0,90	0,75	23,269	16,762	0,020	0,013	0,055	0,028	0,002	0,001
0,85	0,90	-8,212	16,751	0,015	0,013	0,007	0,028	0,001	0,001
0,85	0,85	-3,421	16,818	0,017	0,013	0,001	0,028	0,002	0,001
0,85	0,75	9,572	16,752	0,021	0,014	0,010	0,028	0,002	0,001
0,75	0,90	-34,410	16,874	0,015	0,013	0,119	0,029	0,001	0,001
0,75	0,85	-30,879	16,759	0,016	0,013	0,096	0,028	0,002	0,001
0,75	0,75	-20,651	16,757	0,020	0,013	0,043	0,028	0,002	0,001

Tabela C.2: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,85 e phi=0,95

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,90	0,95	-4,089	13,519	0,015	0,014	0,002	0,018	0,001	0,001
0,90	0,90	-0,084	13,470	0,018	0,015	0,000	0,018	0,001	0,001
0,90	0,75	17,848	13,511	0,022	0,014	0,032	0,018	0,002	0,001
0,85	0,95	-15,858	13,476	0,016	0,015	0,025	0,018	0,001	0,001
0,85	0,90	-12,202	13,433	0,017	0,015	0,015	0,018	0,001	0,001
0,85	0,75	4,182	13,536	0,023	0,015	0,002	0,019	0,002	0,001
0,75	0,95	-41,447	13,449	0,016	0,015	0,172	0,018	0,001	0,001
0,75	0,90	-38,655	13,401	0,019	0,016	0,150	0,018	0,001	0,001
0,75	0,75	-26,088	13,523	0,023	0,015	0,069	0,019	0,002	0,001

Tabela C.3: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,75 e phi=0,95

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,90	0,95	-14,322	3,799	0,016	0,015	0,021	0,002	0,001	0,001
0,90	0,90	-10,934	3,757	0,018	0,016	0,012	0,002	0,001	0,001
0,90	0,85	-6,953	3,771	0,018	0,015	0,005	0,002	0,001	0,001
0,85	0,95	-26,010	3,778	0,016	0,015	0,068	0,002	0,001	0,001
0,85	0,90	-22,964	3,700	0,018	0,016	0,053	0,002	0,001	0,001
0,85	0,85	-19,326	3,753	0,019	0,016	0,038	0,002	0,001	0,001
0,75	0,95	-51,402	3,767	0,017	0,016	0,265	0,002	0,001	0,001
0,75	0,90	-49,150	3,670	0,019	0,017	0,242	0,002	0,001	0,001
0,75	0,85	-46,434	3,699	0,018	0,015	0,216	0,002	0,001	0,001

Tabela C.4: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,95 e phi=0,85

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,90	49,051	42,909	0,031	0,029	0,242	0,185	0,002	0,002
0,95	0,85	55,925	42,540	0,031	0,026	0,314	0,182	0,002	0,002
0,95	0,75	76,395	42,611	0,041	0,027	0,585	0,182	0,004	0,002
0,90	0,90	38,297	42,531	0,028	0,026	0,147	0,182	0,002	0,002
0,90	0,85	45,107	42,402	0,032	0,027	0,205	0,181	0,002	0,002
0,90	0,75	64,634	42,531	0,039	0,025	0,419	0,182	0,004	0,002
0,75	0,90	3,072	42,551	0,025	0,023	0,002	0,182	0,002	0,002
0,75	0,85	8,616	42,454	0,029	0,024	0,008	0,181	0,002	0,002
0,75	0,75	24,568	42,604	0,036	0,024	0,062	0,182	0,003	0,002

Tabela C.5: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,85 e phi=0,85

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	25,700	25,806	0,042	0,041	0,068	0,068	0,001	0,001
0,95	0,90	31,488	26,637	0,025	0,024	0,100	0,072	0,002	0,001
0,95	0,75	52,789	26,608	0,033	0,023	0,280	0,071	0,002	0,001
0,90	0,95	16,510	26,712	0,022	0,022	0,028	0,072	0,001	0,001
0,90	0,90	21,011	26,508	0,023	0,021	0,045	0,071	0,002	0,001
0,90	0,75	41,719	26,895	0,032	0,022	0,175	0,073	0,002	0,001
0,75	0,95	-17,777	26,661	0,025	0,025	0,032	0,072	0,001	0,001
0,75	0,90	-14,432	26,300	0,026	0,024	0,022	0,070	0,002	0,001
0,75	0,75	1,784	26,629	0,036	0,026	0,002	0,072	0,002	0,001

Tabela C.6: Desempenho do método proposto e do convencional na estimação do parâmetro da distribuição exp(1) quando teta=0,75 e phi=0,85

Erros		% Vicio		Erro Padrão		EQM		Média (Var ass.)	
$\phi*$	$\theta*$	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.	Prop.	Conv.
0,95	0,95	15,262	15,381	0,037	0,037	0,025	0,025	0,001	0,001
0,95	0,90	20,223	16,107	0,024	0,022	0,041	0,026	0,001	0,001
0,95	0,85	25,386	16,319	0,028	0,024	0,065	0,027	0,001	0,001
0,90	0,95	5,909	16,114	0,021	0,021	0,004	0,026	0,001	0,001
0,90	0,90	9,745	15,945	0,022	0,021	0,010	0,026	0,001	0,001
0,90	0,85	14,565	16,112	0,026	0,022	0,022	0,026	0,001	0,001
0,75	0,95	-28,347	16,022	0,023	0,023	0,081	0,026	0,001	0,001
0,75	0,90	-25,550	15,717	0,025	0,024	0,066	0,025	0,001	0,001
0,75	0,85	-21,571	16,059	0,025	0,022	0,047	0,026	0,001	0,001

Apêndice D

Cálculo da Variância do Coeficiente $\hat{\beta}$ no Modelo de Cox

Seja

$$L = \prod_{j=1}^{k} \prod_{\mu \in R_j} \left\{ \phi^{(j-1)}(1-\phi) b_{j,\mu} + \theta a_{j,\mu} \right\}^{n_{j,\mu}} \left\{ \phi^j b_{j,\mu} + (1-\theta) a_{j,\mu} \right\}^{m_{j,\mu}}$$

Sejam $l1 = n_{j,\mu} ln(\phi^{j-1}(1-\phi)b_{j,\mu} + \theta a_{j,\mu})$ e $l2 = n_{j,\mu} ln(\phi^j b_{j,\mu} + (1-\theta)a_{j,\mu})$.
Então

$$l = \sum_{j=1}^{k} \sum_{\mu \in S_j} (l1 + l2).$$

Na definição acima, $a_{j,\mu} = \sum_{l=1}^{j} (\phi^{l-1}(1-\theta)^{j-l} c_{l,\mu} b_{l-1,\mu})$, $b_{j,\mu} = \left(\prod q = 1^j \gamma_q\right)^{exp(\beta' \mathbf{x}_\mu)}$
e

$$c_{j,\mu} = 1 - \gamma_j{}^{exp(\beta' \mathbf{x}_\mu)}.$$

Cálculo da variância do estimador de $\beta_s (s = 1, ..., p)$.

$$\frac{\partial l}{\partial \beta_s} = \sum_j \sum \mu \left(\frac{\partial l1}{\partial \beta_s} + \frac{\partial l2}{\partial \beta_s} \right).$$

Temos que:

$$\frac{\partial l1}{\partial \beta_s} = n_{j,\mu} \frac{1}{\phi^{j-1}(1-\phi) b_{j,\mu} + \theta a_{j,\mu}} \left(\phi^{j-1}(1-\phi) \frac{\partial b_j, \mu}{\partial \beta_s} + \theta \frac{\partial a_{j,\mu}}{\partial \beta_s} \right)$$

e

$$\frac{\partial l2}{\partial \beta_s} = m_{j,\mu} \frac{1}{\phi^j b_{j,\mu} + (1-\theta) a_{j,\mu}} \left(\phi^j \frac{\partial b_j, \mu}{\partial \beta_s} + (1-\theta) \frac{\partial a_{j,\mu}}{\partial \beta_s} \right)$$

em que

$$\frac{\partial a_{j,\mu}}{\partial \beta_s} = \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{j-l} \left(\frac{\partial c_{l,\mu}}{\partial \beta_s} b_{l-1,\mu} + c_{(l,\mu)} \frac{\partial b_{l-1,\mu}}{\partial \beta_s} \right)$$

$$\frac{\partial b_{j,\mu}}{\partial \beta_s} = b_{j,\mu} ln \left(\prod_{q=1}^{j} \gamma_q \right) exp(\beta' \mathbf{x}_\mu) \mathbf{x}_s,$$

e

$$\frac{\partial c_{j,\mu}}{\partial \beta_s} = -\gamma_j{}^{exp(\beta' \mathbf{x}_\mu)} exp(\beta' \mathbf{x}_\mu) \mathbf{x}_s ln(\gamma_j).$$

Cálculo da derivada segunda:

$$\frac{\partial^2 l}{\partial \beta_s{}^2} = \sum_j \sum_\mu \left(\frac{\partial^2 l1}{\partial \beta_s^2} + \frac{\partial^2 l2}{\partial \beta_s^2} \right)$$

$$\frac{\partial^2 l1}{\partial \beta_s{}^2} = n_{j,\mu} \left(- \frac{\phi^{j-1}(1-\theta) \frac{\partial b_{j,\mu}}{\partial \beta'_s} + \theta \frac{\partial a_{j,\mu}}{\partial \beta'_s}^2}{(\phi(1-\phi) b_{j,\mu} + \theta a_{j,\mu})^2} + \frac{\phi^{j-1}(1-\theta) \frac{\partial^2 b_{j,\mu}}{\partial \beta'^2_s} + \theta \frac{\partial^2 a_{j,\mu}}{\partial \beta'^2_s}}{(\phi^{j-1}(1-\phi) b_{j,\mu} + \theta a_{j,\mu})} \right)$$

$$\frac{\partial^2 l2}{\partial \beta_s{}^2} = m_{j,\mu} \left(- \frac{\phi^j \frac{\partial b_{j,\mu}}{\partial \beta'_s} + (1-\theta) \frac{\partial a_{j,\mu}}{\partial \beta'_s}^2}{(\phi^j b_{j,\mu} + (1-\theta) a_{j,\mu})^2} + \frac{\phi^j \frac{\partial^2 b_{j,\mu}}{\partial \beta'^2_s} + (1-\theta) \frac{\partial^2 a_{j,\mu}}{\partial \beta'^2_s}}{(\phi^j b_{j,\mu} + (1-\theta) a_{j,\mu})} \right)$$

em que

$$\frac{\partial^2 a_{j,\mu}}{\partial \beta_s^2} = \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{j-l} \left(\frac{\partial^2 c_{l,\mu}}{\partial \beta_s^2} b_{l-1,\mu} + 2 \frac{\partial c_{l,\mu}}{\partial \beta_s} \frac{\partial b_{l-1,\mu}}{\partial \beta_s} + c_{l,\mu} \frac{\partial^2 b_{l-1,\mu}}{\partial \beta_s^2} \right)$$

$$\frac{\partial^2 b_{j,\mu}}{\partial \beta_s{}^2} = ln \left(\prod_{q=1}^{j} \gamma_q \right) \mathbf{x}_s \left(exp(\beta' \mathbf{x}_\mu) \frac{\partial b_{j,\mu}}{\partial \beta_s} + b_{j,\mu} exp(\beta' \mathbf{x}_\mu) \mathbf{x}_s \right)$$

$$\frac{\partial^2 c_{j,\mu}}{\partial \beta_s^2} = ln(\gamma_j) exp(\beta' \mathbf{x}_\mu) \left(-\gamma_j{}^{exp(\beta' \mathbf{x}_\mu)} \right) (\mathbf{x}_s)^2 \left(ln(\gamma_j) exp(\beta' \mathbf{x}_\mu) + 1 \right).$$

Assim, temos:

$$Var(\hat{\beta}_s) = \left(- \frac{\partial^2 l}{\partial \beta^2{}_s} \right)^{-1}.$$

Apêndice E

Cálculo da Variância dos Estimadores de β_s

Para o cálculo da variância dos estimadores, utilizamos a matriz da variância observada dada pelas segundas derivadas:

$$\frac{\partial^2 l}{\partial^2 \beta_s} = \sum_{j=1}^{k} \sum_{\mu \in S_j} \left(\frac{\partial^2 l1}{\partial^2 \beta_s} + \frac{\partial^2 l2}{\partial^2 \beta_s} \right),$$

em que

$$\frac{\partial^2 l1}{\partial^2 \beta_s} = n_{j,\mu} \left(- \left(\frac{\phi^{\gamma-1}(1-\phi)\frac{\partial b_{j,\mu}}{\partial \beta_s} + \theta \frac{\partial a_{j,\mu}}{\partial \beta_s}}{(\phi^{j-1}(1-\phi)b_{j,\mu} + \theta a_{j,\mu})} \right)^2 + \left(\frac{\phi^{\gamma-1}(1-\phi)\frac{\partial^2 b_{j,\mu}}{\partial^2 \beta_s} + \theta \frac{\partial^2 a_{j,\mu}}{\partial^2 \beta_s}}{(\phi^{j-1}(1-\phi)b_{j,\mu} + \theta a_{j,\mu})} \right) \right)$$

e

$$\frac{\partial^2 l2}{\partial^2 \beta_s} = m_{j,\mu} \left(- \left(\frac{\phi^{j}\frac{\partial b_{j,\mu}}{\partial \beta_s} + (1-\theta) \frac{\partial a_{j,\mu}}{\partial \beta_s}}{(\phi^{j} b_{j,\mu} + (1-\theta) a_{j,\mu})^2} \right)^2 + \frac{\phi^{j}\frac{\partial^2 b_{j,\mu}}{\partial^2 \beta_s} + (1-\theta) \frac{\partial^2 a_{j,\mu}}{\partial^2 \beta_s}}{\phi^{j} b_{j,\mu} + (1-\theta) a_{j,\mu}} \right)$$

Como

$$b_{j,\mu} = S(t_j|\mathbf{x}_\mu) = exp(-exp(\beta'\mathbf{x}_\mu){t^\gamma}_j)$$

temos que

$$\frac{\partial b_{j,\mu}}{\partial \beta_s} = exp(-exp(\beta'\mathbf{x}_\mu){t_j}^\gamma)(-{t_j}^\gamma)exp(\beta'\mathbf{x}_\mu)\mathbf{x}_\mu = b_{j,\mu}(-{t_j}^\gamma \mathbf{x}_s)exp(\beta'\mathbf{x} - \mu)$$

e

$$\frac{\partial^2 b_{j,\mu}}{\partial^2 \beta_s} = -{t_j}^\gamma {\mathbf{x}_s}^2 b_{j,\mu} exp(\beta'\mathbf{x}) \left(-{t_j}^\gamma exp(\beta'\mathbf{x}) + 1 \right).$$

E, como

$$a_{j,\mu} = \sum_{l=1}^{j} \phi^{(l-1)}(1-\theta)^{(j-l)}(b_{l-1,\mu} - b_{l,\mu}),$$

$$\frac{\partial a_{j,\mu}}{\partial \beta_s} = \sum_{l=1}^{j} \phi^{(l-1)}(1-\theta)^{(j-l)} \left(\frac{\partial b_{l-1,\mu}}{\partial \beta_s} - \frac{\partial b_{l,\mu}}{\partial \beta_s} \right),$$

e

$$\frac{\partial^2 a_{j,\mu}}{\partial \beta_s} = \sum_{l=1}^{j} \phi^{(l-1)}(1-\theta)^{(j-l)} \left(\frac{\partial^2 b_{l-1,\mu}}{\partial^2 \beta_s} - \frac{\partial^2 b_{l,\mu}}{\partial^2 \beta_s} \right).$$