

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

Dissertação de Mestrado

**Intervalos de Confiança para a
Função de Distribuição na Presença de
Censura Intervalar, Caso 1**

por

Erik Rodarte Lopes

Orientador: Prof. Dr. Antônio Eduardo Gomes

Abril de 2005

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Erik Rodarte Lopes

**Intervalos de Confiança para a
Função de Distribuição na Presença de
Censura Intervalar, Caso 1**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Mestre em Estatística.

**Universidade Federal de Minas Gerais
Belo Horizonte, Abril de 2005**

*Aos meus avós maternos, Ulisses
Vilela Rodarte (in memorian) e
Claudina Moreira Rodarte.*

Agradecimentos

- Primeiramente a DEUS, por ter me permitido alcançar mais essa vitória e também por estar sempre ao meu lado tanto nos bons quanto nos momentos mais difíceis da minha vida.
- Aos meus pais, Maria Luzia Rodarte Lopes e Maximino José Lopes, pelo amor e carinho incondicional dedicado à minha pessoa.
- Aos meus irmãos, Thaís, Fábio, Fernanda e Ana Cláudia, pela sincera amizade oferecida à minha pessoa.
- Aos meus avós paternos, vó Lia e vô Maximino (in memorian), por tudo que fizeram pelo meu pai e, consequentemente, pela minha família.
- Aos meus tios Wilson e Odília e aos meus primos Mirella, Jamille e Thales, pela imensa ajuda oferecida à mim e a toda minha família.
- À família Vilela Rey (Vanderlan, Regina, Ulisses e Thamires).
- À família Rodarte Parreira (Ney, Maria Aparecida, Marco Flávio, Cínthia, Alexandre e Amanda).
- Ao Prof. Antônio Eduardo Gomes, por ter sido mais do que orientador, foi um grande amigo que me apoiou em todos os momentos dessa minha jornada acadêmica.
- Aos professores Gregório e Sueli do Departamento de Estatística da UFMG e aos professores Olímpio, Maria Ignêz e Maria Teresa do Departamento de Matemática da UFU-MG, pela confiança que depositaram em mim.

- Aos professores Aniura (UFMG) e Mário Vivanco (UFLA) pelas valiosas sugestões que contribuíram para a melhora desse texto.
- À familia Cardoso Reis (Luiz Carlos, Ana Maria, Febiane, Fábio e Amanda), pelas caronas para Campo Belo-MG e pelo carinho que me recebem em sua casa.
- Aos meus amigos de Graduação em Matemática da UFU-MG (Alessandro, Cleiton, Edmilson, Ercílio, Márcio e Renato) e aos da graduação em Estatística da UFMG (Edmilson, Jeislân, Joel e Gonzaga), meu muito obrigado pelos momentos de muita alegria que passamos juntos.
- Aos meus amigos de mestrado, Anderson, Cristiano, Fábio, Francisco, Paula e Ricardo, pela valiosa ajuda e amizade ao longo desse trabalho.
- Aos meus amigos de Campo Belo-MG, Luciano (tijão), Marcelo, Nelison, Oswaldo (bil), Júnior, Fred e Dênis, por me proporcionarem momentos de grande alegria na minha terra natal.
- Enfim, agradeço à todos que de uma forma ou de outra contribuíram para que esse trabalho chegasse ao fim.

Sumário

Lista de Figuras	3
Lista de Tabelas	5
Resumo	6
Abstract	7
Motivação	8
Revisão da Literatura	9
1 Conceitos Básicos em Análise de Sobrevida	11
1.1 Introdução	11
1.2 Censura	11
1.3 Tempo de Falha	12
1.4 Métodos de Estimação da Função de Sobrevida na Presença de Censura à Direita	14
1.4.1 Abordagem Não-Paramétrica	14
1.4.2 Abordagem Paramétrica	15
2 Intervalos de Confiança para a Função de Distribuição	17
2.1 Introdução	17
2.2 O Estimador Não-Paramétrico de Máxima Verossimilhança	18
2.2.1 Regressão Isotônica	18
2.2.2 Cálculo do Intervalo de Confiança para a Função de Distribuição	20

2.3	Método Bootstrap	21
2.3.1	Aspectos Gerais	21
2.3.2	A Estimativa Bootstrap do Erro Padrão e Intervalos de Confiança	23
2.3.3	O Método Bootstrap na Presença de Censura Intervalar, Caso 1	24
2.4	Imputação	25
2.4.1	Aspectos Gerais da Imputação	25
2.4.2	Procedimentos Baseados em Imputação na Presença de Censura Intervalar, Caso 1	26
3	Núcleo Estimadores	27
3.1	Introdução	27
3.2	Medidas de Desempenho do Núcleo Estimador	28
3.3	Escolha da Janela Ótima	30
3.3.1	Escolha da Janela ótima na Ausência de Censura: Método Plug-in Modificado	30
3.3.2	Escolha da Janela Ótima na Presença de Censura Intervalar .	31
4	Simulação e Aplicação	33
4.1	Simulação	33
4.2	Aplicação	69
5	Conclusões e Trabalhos Futuros	72
5.1	Conclusões	72
5.2	Trabalhos Futuros	74
	Referências Bibliográficas	75

Lista de Figuras

4.1	Comparação das probabilidades de cobertura para o caso 1 e $n = 60$.	58
4.2	Comparação das probabilidades de cobertura para o caso 1 e $n = 100$	59
4.3	Comparação das probabilidades de cobertura para o caso 1 e $n = 200$	60
4.4	Comparação das probabilidades de cobertura para o caso 1 e $n = 500$	61
4.5	Comparação das probabilidades de cobertura para o caso 2 e $n = 60$.	62
4.6	Comparação das probabilidades de cobertura para o caso 2 e $n = 100$	63
4.7	Comparação das probabilidades de cobertura para o caso 2 e $n = 200$	64
4.8	Comparação das probabilidades de cobertura para o caso 2 e $n = 500$	65
4.9	Comparação das probabilidades de cobertura para o caso 3 e $n = 100$	66
4.10	Comparação das probabilidades de cobertura para o caso 3 e $n = 200$	67
4.11	Comparação das probabilidades de cobertura para o caso 3 e $n = 500$	68
4.12	Estimativas por nucleo estimador das funções densidade f e g.	70
4.13	F estimada e intervalo de 90%, 95% e 99% de confiança.	71

Lista de Tabelas

2.1	Alguns quantis exatos da distribuição de Chernoff	20
4.1	Caso 1, $n = 60$, ENPMV	35
4.2	Caso 1, $n = 100$, ENPMV	35
4.3	Caso 1, $n = 200$, ENPMV	36
4.4	Caso 1, $n = 500$, ENPMV	36
4.5	Caso 2, $n = 60$, ENPMV	37
4.6	Caso 2, $n = 100$, ENPMV	37
4.7	Caso 2, $n = 200$, ENPMV	38
4.8	Caso 2, $n = 500$, ENPMV	38
4.9	Caso 3, $n = 100$, ENPMV	39
4.10	Caso 3, $n = 200$, ENPMV	39
4.11	Caso 3, $n = 500$, ENPMV	40
4.12	Caso 1, $n = 60$, Kernel	40
4.13	Caso 1, $n = 100$, Kernel	41
4.14	Caso 1, $n = 200$, Kernel	41
4.15	Caso 1, $n = 500$, Kernel	42
4.16	Caso 2, $n = 60$, Kernel	42
4.17	Caso 2, $n = 100$, Kernel	43
4.18	Caso 2, $n = 200$, Kernel	43
4.19	Caso 2, $n = 500$, Kernel	44
4.20	Caso 3, $n = 100$, Kernel	44
4.21	Caso 3, $n = 200$, Kernel	45
4.22	Caso 3, $n = 500$, Kernel	45
4.23	Caso 1, $n = 60$, Bootstrap	46

4.24 Caso 1, $n = 100$, Bootstrap	46
4.25 Caso 1, $n = 200$, Bootstrap	47
4.26 Caso 1, $n = 500$, Bootstrap	47
4.27 Caso 2, $n = 60$, Bootstrap	48
4.28 Caso 2, $n = 100$, Bootstrap	48
4.29 Caso 2, $n = 200$, Bootstrap	49
4.30 Caso 2, $n = 500$, Bootstrap	49
4.31 Caso 3, $n = 100$, Bootstrap	50
4.32 Caso 3, $n = 200$, Bootstrap	50
4.33 Caso 3, $n = 500$, Bootstrap	51
4.34 Caso 1, $n = 60$, Imputação	51
4.35 Caso 1, $n = 100$, Imputação	52
4.36 Caso 1, $n = 200$, Imputação	52
4.37 Caso 1, $n = 500$, Imputação	53
4.38 Caso 2, $n = 60$, Imputação	53
4.39 Caso 2, $n = 100$, Imputação	54
4.40 Caso 2, $n = 200$, Imputação	54
4.41 Caso 2, $n = 500$, Imputação	55
4.42 Caso 3, $n = 100$, Imputação	55
4.43 Caso 3, $n = 200$, Imputação	56
4.44 Caso 3, $n = 500$, Imputação	56
4.45 Comportamento do h_{opt} para f e g	57
4.46 Idade da morte de ratos RFM machos	69

Resumo

A análise de sobrevivência é uma das áreas da estatística que mais tem crescido nos últimos anos. Geralmente, a variável resposta é o tempo até a ocorrência de um evento de interesse (denominado “tempo de falha”) e a sua principal característica é a presença de censura, que é a observação parcial da variável resposta. Entre as formas de censura existentes, destaca-se nesse trabalho a censura intervalar, que caracteriza-se pelo fato de não conhecermos o tempo exato de ocorrência do evento de interesse, mas apenas que seu valor pertence a um certo intervalo.

O objetivo desse trabalho é estudar a estimação e a construção de intervalos de confiança para a função de distribuição da variável resposta na presença de censura intervalar, caso 1 (tipo de censura intervalar onde apenas uma observação é feita para cada “indivíduo”). Para isso, quatro métodos distintos foram utilizados: (1) o ENPMV (estimador não-paramétrico de máxima verossimilhança) e sua distribuição assintótica, (2) uma variação do método anterior através da suavização do ENPMV, (3) o método bootstrap e (4) o método da imputação múltipla com cálculo do estimador de Kaplan-Meier e respectivo intervalo de confiança. Serão feitas comparações entre esses métodos via simulação. Porém, para aplicação de alguns desses métodos, necessitamos das estimativas (suavizadas) da densidade tanto da variável resposta quanto da variável que representa o instante de observação de cada indivíduo. Portanto, o método não paramétrico do núcleo estimador será utilizado bem como os métodos automáticos para a escolha da janela ótima: o método *plug-in* modificado proposto por Chiu (1991) no caso da variável não censurada e o método proposto por Pan (2000a) no caso da variável censurada.

Palavras Chave: *Censura Intervalar, Intervalos de Confiança, Função de Distribuição, Estimador Não-paramétrico de Máxima Verossimilhança, Bootstrap e Imputação.*

Abstract

The interest in the development of survival data analysis techniques has been growing in the past few years. In this area of statistics the response variable is generally the time until occurrence of an event of interest (failure time) which is quite often censored (i.e., partially observed). In this work we will be considering interval censoring (we do not observe the failure time but just an interval containing it).

The goal of this work is to study the estimation and the construction of confidence intervals for the distribution function of the failure time in the presence of interval censoring, case 1 (when only one observation time is available for each individual). For that purpose four different methods have been used and compared: the NPMLE (nonparametric maximum likelihood estimator) and its asymptotic distribution, an smoothed version of the NPMLE via kernel estimator with the NPMLE asymptotic distribution, bootstrap, and multiple imputation of the failure time for left censored observations with estimation of the survival function of the failure time via Kaplan-Meier estimator with confidence intervals given by the Greenwood formula. The methods involving the asymptotic distribution of the NPMLE require the estimation of the probability density functions of both the failure time and the censoring time. Kernel smoothing was used with bandwidth calculated as proposed by Chiu (1991) (for the censoring variable) and by Pan (2000a) (for the failure time).

key words: *interval censoring, confidence intervals, distribution function, non-parametric maximum likelihood estimation, bootstrap, imputation.*

Motivação

Este trabalho tem como objetivo o estudo de diferentes métodos para construção de intervalos de confiança para a função de distribuição do tempo de falha na presença de censura intervalar.

Uma motivação para se construir tais intervalos de confiança se encontra, por exemplo, em problemas reais de determinação do prazo de validade de produtos alimentícios, pois a partir desses intervalos de confiança para a função de distribuição do tempo de falha, pode-se obter intervalos de confiança para seus quantis, principalmente os de baixa ordem, já que o prazo de validade de um produto é determinado de tal forma que a probabilidade dele falhar (isto é, tornar-se impróprio para o consumo) dentro do prazo de validade seja muito pequena.

Uma outra motivação encontra-se em estudos clínicos sobre doenças assintomáticas com animais de laboratório que são sacrificados e submetidos a autópsia para detecção da presença da doença. Geralmente, se está interessado na estimação da função de distribuição do início da doença. Estimadores com vício pequeno e intervalos de confiança com amplitude menor e probabilidade de cobertura real igual ou superior ao valor nominal fornecem melhores estimativas das quantidades de interesse, como o tempo mediano de início da doença.

Revisão da Literatura

Estudos envolvendo dados de sobrevivência com censura intervalar foram introduzidos na literatura por Ayer *et al.* (1955) e Van Eeden (1956). Tais autores desenvolveram o algoritmo de agrupamento de violadores adjacentes para calcular o ENPMV (estimador não paramétrico de máxima verossimilhança) da função de distribuição F do tempo de falha sob a forma mais simples de censura intervalar: caso 1 de censura intervalar ou dados de estatus corrente.

Groeneboom e Wellner (1992) mostraram que a distribuição assintótica desse estimador envolve a distribuição de Chernoff. Eles também introduziram um procedimento que usa um algoritmo para o cálculo de regressões isotônicas para obtenção do ENPMV de F . Algumas modificações desse algoritmo podem ser encontradas em Jongbloed (1995) e Zhan e Wellner (1995) que o adaptaram para situações onde ocorre o modelo de censura dupla.

Groeneboom e Wellner (1992) também mostraram a consistência do ENPMV. Esses mesmos autores, em 2001, apresentaram um algoritmo (iterativo) que permitia calcular quantis da distribuição assintótica desse estimador, permitindo assim a construção de intervalos de confiança para a função de distribuição.

A normalidade assintótica do ENPMV de funcionais lineares de F para o caso 1 de censura intervalar foi demonstrada em Huang e Wellner (1995). Turnbull (1976) derivou equações de auto-consistência para situações mais gerais de censura que incluem a censura intervalar como caso especial.

Para o caso 2 de censura intervalar (quando há pelo menos dois instantes de observação por indivíduo), sabe-se que ainda não há uma forma fechada para o ENPMV. Nesse caso, Groeneboom e Wellner (1992) propuseram um algoritmo iterativo para o cálculo do ENPMV e que também, sob certas condições de regularidade, tem dis-

tribuição assintótica relacionada com a distribuição de Chernoff. Wellner (1995) fez um estudo mais geral onde algumas dessas condições são eliminadas.

Outro problema bastante interessante é a estimação da função de distribuição via núcleo estimadores na presença de censura. O estimador de Kaplan-Meier é um estimador não paramétrico da função de sobrevivência na presença de censura à direita. Padgett (1986) propôs um núcleo estimador da função quantil, baseado na suavização do estimador de Kaplan-Meier. Entretanto, como não há resultados sobre o erro quadrático médio integrado do estimador proposto e, consequentemente, não se pode obter o valor da janela que minimiza o MISE neste caso, o autor sugere o uso do método bootstrap para a janela em algumas situações.

Já para o caso de censura intervalar, Li *et al.* (1997) propõem um método de suavização do estimador da função de sobrevivência baseado no algoritmo EM. Duchesne e Satafford (2001) também apresentam um método iterativo de suavização para a função de distribuição. Essa estimativa resulta de um algoritmo onde esperanças condicionais de um núcleo são calculadas para cada iteração. Essas esperanças condicionais são calculadas com respeito à densidade estimada de iterações anteriores, permitindo ao estimador extraír mais informação dos dados para cada etapa. A escolha da janela ótima para essa proposta é similar ao método de validação cruzada para dados completos.

Um outro método de suavização da função de sobrevivência é o método proposto por Pan (2000a) (método que será abordado nesse texto) que também utiliza um procedimento similar ao da validação cruzada para a escolha da janela ótima.

Capítulo 1

Conceitos Básicos em Análise de Sobrevidência

1.1 Introdução

A análise de sobrevidência pode ser definida como o conjunto de técnicas e modelos estatísticos usados na análise do comportamento de variáveis aleatórias positivas cuja principal característica é a presença de **censura** (observação parcial da variável resposta). Geralmente, a variável resposta é o tempo até a ocorrência de um evento de interesse (denominado “tempo de falha”) que pode ser, por exemplo, o tempo até a morte de um paciente ou até a cura ou recidiva de uma doença.

1.2 Censura

A presença de censura é a característica marcante em estudos de dados em sobrevidência. Porém, ressaltamos o fato de que, mesmo censurados, todos os resultados obtidos através de um estudo de sobrevidência devem ser incorporados na análise estatística. A justificativa para tal procedimento se baseia nas seguintes questões:

- Mesmo sendo incompletas, as observações censuradas nos fornecem alguma informação à respeito da variável resposta;
- A omissão das censuras no cálculo das estatísticas de interesse certamente levará a conclusões viciadas (Colosimo, 2001).

Há três formas de censura. São elas:

- **Censura à direita** : caracteriza-se pelo fato de que o tempo de ocorrência do evento de interesse está à direita do tempo registrado;
- **Censura à esquerda** : ocorre quando o tempo registrado é maior que o tempo de falha, ou seja, o evento de interesse já aconteceu quando o indivíduo foi observado.
- **Censura intervalar** : essa forma de censura caracteriza-se pelo fato de que o valor exato da variável resposta não é observado, sendo sabido apenas que seu valor pertence a um certo intervalo. Tal situação ocorre, por exemplo, quando um indivíduo é submetido a um exame de laboratório para detectar a presença de uma doença assintomática. Neste caso não sabemos o instante de seu início, mas tão somente o momento de realização do exame e se a doença já havia tido início até aquele instante ou se isso ocorrerá no futuro. Se apenas um exame é realizado para cada indivíduo, temos uma estrutura de dados conhecida como caso 1 de censura intervalar (o caso 2 designa a situação em que pelo menos dois exames são realizados por indivíduo e não será abordada neste texto).

1.3 Tempo de Falha

A variável aleatória positiva T que representa o tempo de falha é normalmente caracterizada em análise de sobrevivência pela sua função de sobrevivência, pela função de taxa de falha (risco) apresentadas a seguir, e pela sua função densidade de probabilidade.

Definição 1.3.1. A função de Sobrevivência $S(t)$ é definida como a probabilidade de uma observação não falhar até o tempo t , ou seja:

$$S(t) = P(T \geq t). \quad (1.1)$$

Definição 1.3.2. A função de taxa de falha é definida por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}. \quad (1.2)$$

Observe que $h(t)$ é sempre uma função positiva e que se fizermos Δt tender a zero, $h(t)$ representa a taxa de falha instantânea no tempo t condicionada à sobrevivência até o tempo t . Se $h(t)$ for uma função crescente, isto será uma indicação de que a taxa de falha aumenta com o transcorrer do tempo, mostrando um efeito de envelhecimento gradativo. Se for constante, indica que a taxa de falha não se altera com o passar do tempo e, finalmente, se for decrescente, mostrará que a taxa de falha diminui à medida que o tempo passa. Além disso, essas funções estão assim relacionadas:

1. $h(t) = \frac{f(t)}{S(t)}$ onde $f(t)$ é a função densidade de probabilidade de T .

Prova:

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t P(T \geq t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)} \end{aligned}$$

2. $f(t) = -\frac{dS(t)}{dt}$

Prova: Basta lembrar que $S(t) = 1 - F(t)$

3. $h(t) = -\frac{d \log S(t)}{dt}$

Prova:

Sabemos que $S(t) = P(T \geq t) = 1 - F(t)$. Aplicando logaritmo a ambos os membros, temos:

$$\begin{aligned} \log S(t) &= \log(1 - F(t)) \text{ logo :} \\ \frac{d \log S(t)}{dt} &= \frac{-f(t)}{S(t)} = -h(t), \text{ ou seja,} \\ h(t) &= -\frac{d \log S(t)}{dt} \end{aligned}$$

Portanto, o problema básico em sobrevivência é estimar uma ou mais dessas funções para que possamos inferir sobre a população que originou os dados.

1.4 Métodos de Estimação da Função de Sobrevivência na Presença de Censura à Direita

1.4.1 Abordagem Não-Paramétrica

Essa abordagem consiste em estimar a função de sobrevivência sem nenhuma suposição sobre a forma matemática da distribuição da variável aleatória T (o espaço paramétrico não tem dimensão finita). Mesmo sendo menos eficiente do que a abordagem paramétrica, onde atribui-se uma distribuição para T , sugerimos utilizar essa abordagem antes de tentar encontrar um modelo teórico para T uma vez que métodos não-paramétricos são bastante úteis na escolha desse modelo, a menos que a suposição de uma determinada distribuição seja razoável numa determinada situação.

O estimador não-paramétrico para a função de sobrevivência mais conhecido, e também mais utilizado, é o estimador de Kaplan-Meier (1958) ou estimador limite-produto, definido por:

$$\hat{S}(t) = \prod_{i|t_i \leq t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i|t_i \leq t} \left(1 - \frac{d_i}{n_i} \right) \quad (1.3)$$

em que:

- d_i : número de falhas no tempo t_i , para $0 < t_1, \dots, < t_k \quad k \leq n$;
- n_i : número de observações sob risco (não falhou e não foi censurado) até o tempo t_i (exclusive).

Por outro lado, Breslow e Crowley (1974) e Meier (1975) mostraram, sob certas condições de regularidade, algumas propriedades assintóticas de $\hat{S}(t)$, tais como normalidade e consistência, permitindo assim, a construção de intervalos de confiança aproximados para $S(t)$ em um instante de tempo t . Por exemplo, um intervalo de 95% de confiança é dado por $\hat{S}(t)^{\exp(\pm 1.96\sqrt{\text{Var}(\hat{U}(t))})}$ em que:

- $\hat{U}(t) = \log(-\log(\hat{S}(t)))$
- $\text{Var}(\hat{U}(t)) = \frac{\sum_{i|t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}}{\left[\sum_{i|t_i \leq t} \log\left(\frac{n_i - d_i}{n_i}\right) \right]^2}.$

Além do estimador de Kaplan-Meier, existem outros dois estimadores para a função de sobrevivência: o estimador da tabela de vida e o estimador de Nelson-Aalen. Esses estimadores não serão discutidos nesse texto e maiores detalhes sobre os mesmos poderão ser encontrados em Lee (1992) e Colosimo (2001).

1.4.2 Abordagem Paramétrica

Essa abordagem, como dissemos anteriormente, consiste em atribuir uma distribuição de probabilidade à variável aleatória T e a partir daí, estimar qualquer parâmetro de interesse. É importante ressaltar que cada distribuição de probabilidade gera estimadores diferentes para a mesma quantidade desconhecida. Logo, uma escolha inadequada para essa distribuição pode acarretar erros grosseiros nas estimativas dessas quantidades. Deve-se, então, tomar bastante cuidado na escolha de tal distribuição.

A escolha da distribuição mais adequada para determinada situação é baseada em métodos empíricos como, por exemplo, o método da linearização da função de sobrevivência. Um estudo mais detalhado dessa abordagem bem como dos métodos de seleção do melhor modelo a ser utilizado pela variável T encontra-se em Lee (1992) e Colosimo (2001).

Após a escolha do modelo mais adequado para T , nosso objetivo passa a ser a estimativa dos parâmetros desse modelo. Entre vários métodos de estimação encontrados na literatura, o mais utilizado em dados de sobrevivência é o método de máxima verossimilhança, uma vez que as censuras são incorporadas de maneira relativamente fácil e também por apresentar propriedades ótimas para grandes amostras.

Considerando que a contribuição das observações censuradas é dada pela função de sobrevivência e das observações não censuradas é dada pela função densidade de

probabilidade, temos que a função de verossimilhança é dada por:

$$L(\theta) = \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i}. \quad (1.4)$$

Os estimadores de máxima verossimilhança de θ são os valores que maximizam $L(\theta)$ ou, equivalentemente, $\log L(\theta)$. Eles são encontrados resolvendo o seguinte sistema de equações:

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0. \quad (1.5)$$

Em várias situações, o sistema acima não pode ser resolvido analiticamente, tendo que ser resolvido por métodos numéricos tais como o método de Newton-Raphson (mais detalhes sobre esse método podem ser encontrados em Lee, 1992).

Capítulo 2

Intervalos de Confiança para a Função de Distribuição

2.1 Introdução

Como mencionado anteriormente, a característica marcante em dados de sobrevivência é a presença de censura, que se apresenta de três formas diferentes: censura à direita, censura à esquerda e censura intervalar.

O objetivo desse capítulo é descrever métodos para a estimação e a construção de intervalos de confiança da função de distribuição F da variável de interesse X (tempo de falha) na presença de censura intervalar, caso 1. Para isso, quatro diferentes métodos foram utilizados. São eles:

1. O ENPMV (estimador não-paramétrico de máxima verossimilhança) estudado por Groeneboom e Wellner (1992) e que usa um algoritmo para cálculo de regressões isotônicas para as estimativas de F e a distribuição assintótica desse estimador no cálculo do intervalo de confiança.
2. Uma variação do método anterior através da suavização do ENPMV via núcleo estimadores. Nesse caso, a distribuição assintótica do ENPMV também será usada para a construção de intervalos de confiança.

3. O método bootstrap que também usa o ENPMV para as estimativas de F . Para o cálculo do intervalo de confiança será utilizada a seguinte expressão:
 $\hat{F}(t) \pm z_{\alpha/2} e.p (\hat{F}(t))$.(maiores detalhes sobre essa expressão serão vistos na Seção 2.3).
4. O método da imputação múltipla com a finalidade de “transformar” uma estrutura de dados com censura intervalar em uma situação onde a estrutura é de censura à direita e, a partir daí, utilizar o estimador de Kaplan-Meier e a fórmula de Greenwood para a construção do intervalo de confiança.

2.2 O Estimador Não-Paramétrico de Máxima Verossimilhança

2.2.1 Regressão Isotônica

Definição 2.2.1. *Seja $(X_1, T_1), (X_2, T_2), \dots, (X_n, T_n)$ uma amostra aleatória onde X_i (tempo de sobrevivência) e T_i (instante de observação) são variáveis aleatórias independentes, não negativas com funções de distribuição F e G , respectivamente. Suponha também que apenas observamos as variáveis T_i e $\delta_i = 1_A$ onde $A = \{X_i \leq T_i\}$. Tal estrutura de dados é conhecida como **caso 1 de censura intervalar**.*

Logo, em situações onde ocorre caso 1 de censura intervalar, o estimador não-paramétrico de máxima verossimilhança de F é o valor $\hat{F}(T_i)$, $i = 1, \dots, n$, que maximiza a expressão:

$$\log L(F) = \sum_{i=1}^n \{\delta_i \log F(T_i) + (1 - \delta_i) \log(1 - F(T_i))\}. \quad (2.1)$$

Vamos assumir, sem perda de generalidade, que $0 < T_1 < T_2 < \dots < T_n$

Para encontrarmos tais estimativas, utilizaremos um procedimento proposto por Groeneboom e Wellner(1992) que baseia-se em um algoritmo para cálculo de regressões isotônicas.

Definição 2.2.2. *Uma função real h em D é isotônica se $x \leq y$ implica $h(x) \leq h(y)$ $\forall x, y \in D$.*

Definição 2.2.3. Seja s uma função qualquer em D e w uma função positiva em D . Uma função isotônica s^* em D é uma regressão isotônica de s com pesos w se, e somente se, minimiza, na classe de funções isotônicas h em D , a soma

$$\sum_{x \in D} [s(x) - h(x)]^2 w(x).$$

Teorema 2.2.1. Se h é isotônica em D , a imagem de h está em $I \subset \mathbb{R}$ e se ϕ é uma função estritamente convexa, então s^* é a única função que maximiza

$$\sum_x \{\phi(h(x)) + (s(x) - h(x))\varphi(h(x))\}w(x) \quad \text{em que} \quad \varphi(x) = \frac{d \phi(x)}{dx}.$$

Prova: Ver Barlow *et al.* (1972) página 42.

Por outro lado, podemos observar que (2.1) pode ser escrita na forma acima com $g(T_i) = \delta_i$, $w(T_i) = 1$, $\phi(x) = x \log x + (1-x) \log(1-x)$, $\varphi(x) = \phi'(x) = \log(x) - \log(1-x)$.

Portanto, pelo Teorema 2.2.1, o estimador não-paramétrico de máxima verossimilhança de F é dado pela regressão isotônica g^* e calculado do seguinte modo:

$$y_m = \max_{i \leq m} \min_{k \geq m} \frac{\sum_{j \leq i} \delta_j}{k - i + 1}$$

onde $y_m = \hat{F}(T_m)$, $m = 1, \dots, n$.

Essa solução também pode ser encontrada graficamente plotando os pontos $(0, 0)$ e $\left(i, \sum_{j \leq i} \delta_j\right)$, $i = 1, \dots, n$, no plano cartesiano (diagrama de somas acumuladas) e calculando o minorante convexo máximo dos pontos em $[0, n]$.

Definição 2.2.4. O minorante convexo máximo é definido como sendo a função $H^* : [0, n] \rightarrow \mathbb{R}$ tal que

$$H^*(t) = \sup\{H(t); H(i) \leq \sum_{j \leq i} \delta_j, 0 \leq i \leq n, H(0) = 0, H \text{ convexa}\}.$$

É importante ressaltar também que, se $\delta_i = 0$, $1 \leq i \leq k_1$ e $\delta_i = 1$, $k_2 \leq i \leq n$ para $0 < k_1 < k_2 < n$, então $y_i = 0$, $1 \leq i \leq k_1$ e $y_i = 1$, $k_2 \leq i \leq n$.

Obs: O valor y_i é dado pela derivada à esquerda de H^* em i para $i = 1, \dots, n$.

2.2.2 Cálculo do Intervalo de Confiança para a Função de Distribuição

Groeneboom e Wellner(1992) mostraram que $n^{1/3}(\hat{F}(t) - F(t))/K(t) \xrightarrow{d} 2C$ quando $n \rightarrow \infty$ onde:

$C = \arg \max\{B(t) - t^2\}$, sendo $B(t)$ um movimento browniano padrão em \mathbb{R} e $K(t) = \sqrt[3]{F(t)(1 - F(t))f(t)/2g(t)}$ em que f e g são as funções densidades de probabilidade das variáveis X e T , respectivamente.

A distribuição de probabilidade da v.a. C é conhecida como distribuição de Chernoff. Como os quantis dessa distribuição já foram tabelados (Groeneboom e Wellner, 2001), podemos calcular um intervalo de confiança para $F(t)$. Tal intervalo é dado por:

$$\left[\hat{F}(t) - 2 \frac{K(t)}{n^{1/3}} c_{\alpha/2}, \hat{F}(t) + 2 \frac{K(t)}{n^{1/3}} c_{\alpha/2} \right], \quad (2.2)$$

em que $c_{\alpha/2}$ é tal que $P(C > c_{\alpha/2}) = \alpha/2$

Tabela 2.1: Alguns quantis exatos da distribuição de Chernoff

$1 - \alpha/2$	$c_{\alpha/2}$
0,9	0,664235
0,95	0,845081
0,975	0,998181
0,99	1,171530
0,995	1,286659
0,999	1,516664

Como a expressão $K(t)$ envolve as quantidades desconhecidas $F(t)$, $f(t)$, e $g(t)$, podemos estimar $K(t)$ por:

$$\hat{K}(t) = \sqrt[3]{\hat{F}(t)(1 - \hat{F}(t))\hat{f}(t)/2\hat{g}(t)}.$$

As quantidades desconhecidas $\hat{f}(t)$ e $\hat{g}(t)$ podem ser estimadas via núcleo-estimadores, que será o tema do próximo capítulo.

Ou seja,

$$\hat{f}(t) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{t - T_i}{h}\right) (\hat{F}(T_i) - \hat{F}(T_{i-1})) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{t - u}{h}\right) d\hat{F}(u)$$

$$\hat{g}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - T_i}{h}\right)$$

onde $K(\cdot)$ é chamado de núcleo e h é o parâmetro de suavidade ou janela (*bandwidth*).

2.3 Método Bootstrap

2.3.1 Aspectos Gerais

Uma das principais finalidades da inferência estatística, além de encontrar estimadores para os parâmetros desconhecidos de uma determinada distribuição, é medir a precisão desses estimadores. Para isso, é necessário calcular algumas medidas da variabilidade desses estimadores, o que na maioria das vezes é feito utilizando o erro padrão dos mesmos. Em algumas situações (por exemplo, situações onde não conhecemos a distribuição exata do estimador), o cálculo dessas medidas torna-se muito complicado ou mesmo impossível. Nesse contexto, o método bootstrap, criado por Efron (1979), surge como uma solução para esse tipo de problema, uma vez que ele é um método puramente computacional e, portanto, pode ser utilizado independentemente de dificuldades teóricas ou matemáticas no cálculo da distribuição do estimador do parâmetro em questão.

Esse método consiste em uma técnica de reamostragem que permite aproximar a distribuição de uma função das observações a partir da distribuição empírica dos dados, baseado em uma amostra de tamanho finito. A amostragem é feita com reposição da amostra original (bootstrap não-paramétrico). Nessa situação, supomos que as observações são obtidas da função de distribuição empírica, \hat{F} , que designa uma massa de probabilidade igual a $1/n$ para cada ponto amostral. Já para o caso de bootstrap paramétrico, a amostragem é feita usando a distribuição da qual os dados são obtidos.

Atualmente, devido aos recursos computacionais disponíveis, o método bootstrap, além de ser usado para medir a precisão de estimadores, é aplicado também na cons-

trução de intervalos de confiança, testes de hipóteses, estimação do vício, etc. Vamos definir a amostra bootstrap seguindo a notação de Efron e Tibshirani (1993).

Definição 2.3.1. *Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes com função de distribuição comum F .*

Caso paramétrico: *Vamos supor F conhecida e pertencente a uma família paramétrica F_θ onde θ é um vetor de parâmetros desconhecidos. Nesse caso, primeiramente estimamos θ através de um estimador $\hat{\theta}$ e, em seguida, obtemos a distribuição estimada de F , $F_{\hat{\theta}}$. A amostra bootstrap, que denotaremos por $X^* = X_1^*, \dots, X_n^*$ é obtida diretamente por amostragem em $F_{\hat{\theta}}$.*

Caso não-paramétrico: *Se F é desconhecida, assumimos que a função geradora dos dados seja tal que, a cada um dos n valores da amostra seja designada uma probabilidade $1/n$ para sua ocorrência. Deste modo, a amostra bootstrap é construída retirando-se, com reposição, um conjunto de n observações da amostra original, que igualmente denotamos por $X^* = X_1^*, \dots, X_n^*$.*

Suponha agora que nossa estatística de interesse é dada por $T(X_1, \dots, X_n)$. O procedimento bootstrap básico consiste em gerar repetidas amostras bootstrap (paramétrico ou não-paramétrico), cada uma delas de tamanho n , um número suficientemente grande de vezes. Obtemos assim, as amostras bootstrap:

$$X^{*1}, \dots, X^{*B}$$

onde B é o número de replicações bootstrap.

O valor ideal para B depende da finalidade para qual a teoria está sendo utilizada. Por exemplo, se desejamos somente estimar o erro padrão , Efron e Tibshirani (1993) sugerem valores de B entre 50 a 200; se o interesse é construir intervalos de confiança, Hall (1986) fez um estudo do número de replicações necessárias para se obter os níveis de significância desejados e sugere tomar a probabilidade de cobertura nominal como múltiplo de $(B + 1)^{-1}$.

Para cada amostra bootstrap, calculamos o valor da estatística de interesse a qual denotaremos por $T^*(X^{*b})$, $b = 1, \dots, B$. Assim, podemos usar a distribuição empírica de $T^*(X^*)$ como uma aproximação para a verdadeira distribuição da estatística $T(X)$, e portanto fazer inferências sobre o parâmetro θ . O procedimento

acima é referido como princípio “*plug-in*” por Efron e Tibshirani (1993).

2.3.2 A Estimativa Bootstrap do Erro Padrão e Intervalos de Confiança

Como vimos anteriormente, o método bootstrap é um método que tem por finalidade principalmente determinar a precisão de estimadores para os quais uma forma analítica de erro padrão é muito complicada ou até mesmo inexistente.

Portanto, vamos supor que estamos interessados em estimar o erro padrão de uma estatística de interesse. O primeiro passo é construir as amostras bootstrap através de amostragem com reposição da amostra original. Assim, se dispomos de n observações X_1, X_2, \dots, X_n , a primeira amostra bootstrap, que denotamos por X^{*1} , é obtida retirando-se, com reposição, n elementos desse conjunto de dados. Para essa amostra, calculamos o valor da estatística de interesse T , que nesse caso, será denotada por $T^*(X^{*1})$. Repetimos esse procedimento um número grande B de vezes e, para cada amostra bootstrap, calculamos $T^*(X^{*b})$, $b = 1, \dots, B$. A estimativa bootstrap do erro padrão de $T(X)$ é, então, definida como o desvio padrão amostral das B replicações bootstrap, e é dada por :

$$\hat{e.p}(T^*) = \left[\frac{\sum_{b=1}^B [T^*(X^{*b}) - T^*(.)]^2}{B - 1} \right]^{1/2} \quad (2.3)$$

$$\text{onde } T^*(.) = \frac{\sum_{b=1}^B T^*(X^{*b})}{B}.$$

Obs.: O procedimento descrito acima é aplicado ao caso do bootstrap não-paramétrico. Se optarmos por utilizar o bootstrap paramétrico, procedemos da mesma forma, com a diferença de que cada amostra bootstrap é obtida da distribuição que originou o conjunto de dados.

Logo, utilizando-se o Teorema Central do Limite e a expressão (2.3), conclui-se que um intervalo de $100(1 - \alpha)\%$ de confiança (bootstrap) para θ (parâmetro de interesse) é dado por

$$T^* \pm z_{\alpha/2} e.p(T^*)$$

onde $T^*(.)$, como definido acima, é um estimador para θ e $z_{\alpha/2}$ é o quantil de ordem $1 - \alpha/2$ da distribuição normal padrão.

2.3.3 O Método Bootstrap na Presença de Censura Intervalar, Caso 1

Como vimos anteriormente, o método bootstrap, além de ser usado para medir a precisão de estimadores para um determinado parâmetro, é aplicado também, entre outras situações, na construção de intervalos de confiança.

Nesse trabalho, utilizaremos tal método para construir intervalos de confiança para $F(t)$ em situações onde ocorre censura intervalar, caso 1. Para isso, utilizaremos o seguinte procedimento:

1. A partir da amostra original, $(T_1, \delta_1), \dots, (T_n, \delta_n)$, uma amostra bootstrap é construída retirando-se, com reposição, um conjunto de n observações da amostra original, que denotamos por $(T_1^*, \delta_1^*), \dots, (T_n^*, \delta_n^*)$. Esse procedimento é repetido um número suficientemente grande B de vezes.
2. Para cada amostra bootstrap, calcula-se uma estimativa da $F(t)$ utilizando o ENPMV. A seguir, utiliza-se as B estimativas de $F(t)$ das amostras bootstrap para construir um intervalo de confiança para $F(t)$ através da expressão $\hat{F}(t) \pm z_{\alpha/2} e.p(\hat{F}(t))$ onde $\hat{F}(t)$ é a média das B estimativas obtidas através dos B conjuntos de dados avaliados em cada decil (para que tenhamos informação sobre a estimação da função de distribuição na maior parte do suporte da distribuição da variável aleatória X).

É importante ressaltar que, nesse caso, o nosso espaço paramétrico é o espaço constituído pelas funções de distribuição.

2.4 Imputação

2.4.1 Aspectos Gerais da Imputação

A ocorrência de dados ausentes é um problema bastante comum em diversas áreas de pesquisa, tais como estudos epidemiológicos, pesquisas amostrais, experimentos clínicos, etc. Embora o ideal nessa situação seja tentar recuperar os dados reais, isto raramente é possível na prática. Portanto, um procedimento muito utilizado no estudo de dados ausentes é a imputação. Essa técnica consiste em substituir os valores ausentes a fim de completar os dados e, então, analisar o conjunto de dados obtido através dos métodos estatísticos convencionais.

Há dois tipos de imputação: simples e múltipla. A imputação simples consiste em substituir cada valor ausente por um único valor imputado. Sabe-se que esse tipo de imputação tende a subestimar a variabilidade dos estimadores uma vez que ele trata os valores imputados como valores observados. Já a imputação múltipla, proposta por Rubin (1987), consiste em substituir cada valor ausente por dois ou mais valores imputados. O processo de imputação múltipla consiste basicamente de três etapas. São elas:

1. Para cada valor ausente, são gerados M valores ($M \geq 2$).
2. Os M valores são organizados de modo que o primeiro valor imputado para cada dado ausente produz o primeiro conjunto de dados completo, o segundo valor imputado para cada dado ausente produz o segundo conjunto de dados completo, e assim sucessivamente. Cada conjunto de dados completo é analisado segundo uma técnica padrão.
3. Os resultados das M análises são combinados permitindo, assim, que a incerteza da imputação seja considerada.

Maiores detalhes de como é feita essa combinação encontram-se em Rubin (1987).

2.4.2 Procedimentos Baseados em Imputação na Presença de Censura Intervalar, Caso 1

Mesmo sendo uma técnica muito comum em Estatística, a imputação não vem sendo utilizada com frequência em situações onde há presença de censura (Pan, 2000c).

Nesse trabalho, propomos utilizar a técnica da imputação com a finalidade de “transformar” uma estrutura de dados com censura intervalar em uma situação onde a estrutura é de censura à direita. Essa transformação tem como objetivo, principalmente, a utilização de técnicas padrão já existentes para dados com censura à direita como, por exemplo, o estimador de Kaplan-Meier para a função de sobrevivência, o teste de logrank, etc. O procedimento para tal “transformação” segue abaixo:

1. Para todo instante T_i onde $\delta_i = 1$, $i = 1, \dots, n$, seleciona-se aleatoriamente M valores segundo uma distribuição uniforme $[0, T_i]$. Esses valores serão considerados como os instantes exatos de falha. Aqueles instantes T_i onde $\delta_i = 0$ serão mantidos, obtendo assim, uma estrutura de dados com censura à direita.
2. Os M valores imputados são organizados de tal modo que o primeiro valor imputado de cada intervalo onde a imputação foi utilizada produz o primeiro conjunto de dados com a estrutura de censura à direita; o segundo valor de cada um desses intervalos produz o segundo conjunto de dados com a mesma estrutura, e assim sucessivamente. Para cada conjunto de dados com a estrutura de censura à direita, será obtido o valor do estimador de Kaplan-Meier, e consequentemente, uma estimativa da função de distribuição. Logo, utilizando-se de propriedades assintóticas como normalidade e a fórmula da variância de Greenwood, um intervalo de confiança aproximado para $F(t)$ pode ser construído.
3. A avaliação será feita em cada decil e a estimativa final de $F(t)$ será dada pela média das M estimativas obtidas para cada um dos conjuntos de dados gerados. O intervalo de confiança será dado pela média dos M limites de confiança.

Capítulo 3

Núcleo Estimadores

3.1 Introdução

O método do núcleo estimador é um método não paramétrico largamente utilizado na estimação de funções densidade de probabilidade. Tal método foi introduzido na literatura por Rosenblatt (1956) e Parzen (1962), embora suas idéias básicas tenham sido independentemente introduzidas por Fix e Hodges (1951) e Akaike (1954).

Definição 3.1.1. *Seja X_1, \dots, X_n uma amostra aleatória de uma distribuição com função densidade de probabilidade f . O núcleo estimador de f avaliado no ponto x é dado por:*

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.1)$$

onde h é denominado parâmetro de suavização ou janela (bandwidth) e K é uma função denominada núcleo.

Geralmente, a escolha do parâmetro de suavização h é muito mais importante para um melhor desempenho desse método do que a escolha da função núcleo K (Silverman, 1986). Nesse trabalho, utilizamos o núcleo de Epanechnikov, isto é:

$$K(t) = \begin{cases} 3/4(1 - t^2), & \text{se } -1 \leq t \leq 1 \\ 0, & \text{se } t < -1 \text{ ou } t > 1 \end{cases}$$

3.2 Medidas de Desempenho do Núcleo Estimador

Várias medidas têm sido estudadas para avaliar a qualidade do núcleo estimador. Nesta seção, estudaremos algumas dessas medidas e também utilizaremos tais medidas como um critério para encontrar o h_{opt} .

Considerando um ponto x fixo, uma medida do desempenho de \hat{f} é o ERRO QUADRÁTICO MÉDIO (EQM) definido por

$$EQM[\hat{f}(x)] = E \left[(\hat{f}(x) - f(x))^2 \right]. \quad (3.2)$$

Uma observação interessante é que podemos reescrever a expressão (3.2) em função do vício e da variância de $\hat{f}(x)$:

$$EQM[\hat{f}(x)] = Var[\hat{f}(x)] + \left\{ E[\hat{f}(x)] - f(x) \right\}^2. \quad (3.3)$$

Por outro lado, a medida mais utilizada para uma avaliação global do desempenho de $\hat{f}(x)$ é o ERRO QUADRÁTICO INTEGRADO MÉDIO (EQIM) cuja definição é dada por :

$$EQIM[\hat{f}] = E \int \{\hat{f}(x) - f(x)\}^2 dx. \quad (3.4)$$

Considerando que o integrando em (3.4) é não-negativo, podemos inverter a ordem da integração com a esperança e então encontramos uma forma alternativa para o EQIM dada por:

$$\begin{aligned} EQIM[\hat{f}] &= E \int \{\hat{f}(x) - f(x)\}^2 dx = \int E\{\hat{f}(x) - f(x)\}^2 dx = \int EQM\{\hat{f}(x)\} dx \\ &= \int \left\{ E[\hat{f}(x)] - f(x) \right\}^2 dx + \int Var[\hat{f}(x)] dx. \end{aligned} \quad (3.5)$$

Examinaremos agora o vício e a variância do estimador. Silverman (1986) mostra que:

$$E[\hat{f}(x)] = \int \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y) dy \quad (3.6)$$

e

$$Var[\hat{f}(x)] = \frac{1}{n} \left\{ \int \frac{1}{h^2} K\left(\frac{x-y}{h}\right)^2 f(y) dy - \left[\frac{1}{h} \int K\left(\frac{x-y}{h}\right) f(y) dy \right]^2 \right\}. \quad (3.7)$$

Sabe-se que, em geral, o vício e a variância não podem ser facilmente calculados. Porém, sob certas suposições, podemos encontrar expressões aproximadas para (3.6) e (3.7). Dessa forma, suponha que:

1. A função núcleo K é simétrica;
2. $\int K(t)dt = 1$;
3. $\int tK(t)dt = 0$;
4. $\int t^2K(t)dt = k_2 > 0$;
5. A função desconhecida f tem derivadas contínuas de todas as ordens necessárias.

Sabe-se que

$$Vicio[\hat{f}(x)] = \frac{1}{h} \int K\left(\frac{x-y}{h}\right) f(y)dy - f(x).$$

Portanto, fazendo a mudança de variável $y = x - ht$ e utilizando a suposição (1) tem-se:

$$Vicio[\hat{f}(x)] = \int K(t)f(x - ht)dt - f(x) = \int K(t)\{f(x - ht) - f(x)\}dt.$$

Por outro lado, usando expansão em série de Taylor, temos:

$$f(x - ht) = f(x) - ht f'(x) + \frac{1}{2}h^2 t^2 f''(x) + \dots$$

e então, utilizando novamente as suposições acima, temos que:

$$Vicio[\hat{f}(x)] \approx \frac{1}{2}h^2 f''(x)k_2. \quad (3.8)$$

Quanto à variância, usando a mesma substituição $y = x - ht$ e utilizando (3.6) e (3.7) concluimos que:

$$Var[\hat{f}(x)] = \frac{1}{nh} \int f(x - ht)K(t)^2dt - \frac{1}{n}\{f(x) + O(h^2)\}^2$$

Finalmente, utilizando expansão em série de Taylor, chegamos a seguinte expressão:

$$Var[\hat{f}(x)] \approx \frac{f(x) \int K(t)^2 dt}{nh}. \quad (3.9)$$

Portanto, através do EQIM e das aproximações (3.8) e (3.9), pode-se mostrar que o valor ótimo de h é dado por:

$$h_{opt} = k_2^{-2/5} \left\{ \int K(t)^2 dt \right\}^{1/5} \left\{ \int f''(x)^2 dx \right\}^{-1/5} n^{-1/5}. \quad (3.10)$$

3.3 Escolha da Janela Ótima

Nessa seção, estudaremos os métodos de seleção automática para a escolha da janela ótima tanto na ausência de censura (método *Plug-in* modificado) quanto na presença de censura intervalar (método proposto por Pan, 2000a).

3.3.1 Escolha da Janela ótima na Ausência de Censura: Método Plug-in Modificado

Definição 3.3.1. Seja X uma variável aleatória. Sua função característica é a função $\phi : \mathbb{R} \rightarrow \mathbb{C}$ definida por $\phi(\lambda) = E[e^{i\lambda X}]$, sendo,

$$E[e^{i\lambda X}] = E[\cos(\lambda X)] + iE[\sin(\lambda X)], \quad \lambda \in \mathbb{R}.$$

Teorema 3.3.1 (Identidade de Parseval). Seja ϕ a função característica de f . Então:

$$\int (f(x))^2 dx = \frac{1}{2\pi} \int |\phi(\lambda)|^2 d\lambda.$$

Prova: Ver Rudin (1987) página 187.

Teorema 3.3.2 (Teorema da Inversão). Se $\phi \in L^1(-\infty, \infty)$, então F é continuamente diferenciável e tem-se que:

$$F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix\lambda} \phi(\lambda) d\lambda.$$

Prova: Ver Chung (1974) página 155.

Por outro lado, observa-se que $G = \int f''(x)^2 dx$ é a única quantidade desconhecida em (3.10). A idéia do método plug-in é substituí-la por uma estimativa \hat{G} .

Pelo teorema da Inversão e identidade de Parseval, podemos afirmar que

$$G = \int \{f''(x)\}^2 dx = (2\pi)^{-1} \int \lambda^4 |\phi(\lambda)|^2 d\lambda.$$

Então, Chiu (1991) propõe o seguinte estimador para G :

$$\hat{G} = \frac{1}{\pi} \int_0^\Lambda \lambda^4 \left\{ |\tilde{\phi}(\lambda)|^2 - \frac{1}{n} \right\} d\lambda, \quad \text{onde } \Lambda = \min\{\lambda; |\tilde{\phi}(\lambda)| \leq c/n\}$$

para alguma constante $c > 1$ e $\tilde{\phi}(\lambda) = \frac{1}{n} \sum_{j=1}^n \exp(i\lambda X_j)$, $-\infty < \lambda < \infty$, é a função característica amostral.

Logo, o estimador pelo método plug-in da janela ótima é dado por:

$$\hat{h}_p = \left\{ \int K^2(x) dx \right\}^{1/5} \left\{ \int x^2 K(x) dx \right\}^{-2/5} \hat{G}^{-1/5}.$$

No entanto, \hat{h}_p não é um estimador \sqrt{n} consistente de h_{opt} (Hall *et al.*, 1991).

Uma maneira de ajustar essa diferença foi sugerida por Chiu(1991) surgindo assim, o seguinte estimador:

$$\hat{h}_{pm} = \left\{ \hat{h}_p + \frac{\tilde{R}'_n(\hat{h}_p)}{\tilde{A}''_n(\hat{h}_p)} \right\} n^{-1/5}$$

onde $\tilde{R}_n(h) = (24\pi)^{-1} n^{-2/5} h^6 \int_0^\Lambda \lambda^6 \left\{ |\tilde{\phi}(\lambda)|^2 - \frac{1}{n} \right\} d\lambda \int x^2 K(x) dx \int x^4 K(x) dx$

e

$$\tilde{A}_n(h) = 4^{-1} h^4 \hat{G} \{ \int x^2 K(x) dx \}^2 + h^{-1} \int K^2(x) dx - \tilde{R}_n(h).$$

3.3.2 Escolha da Janela Ótima na Presença de Censura Intervalar

Em situações onde ocorre a presença de censura (intervalar) não é recomendável a utilização do método anterior para a escolha da janela ótima. Um procedimento alternativo foi proposto por Pan (2000a) para essa escolha. Nessa situação, Pan propôs a seguinte expressão para o cálculo da janela ótima:

$$h_{opt} = \arg \max_h \sum_{v=1}^V L(\tilde{S}^{(-v)}(:, h) | \mathcal{D}^{(v)})$$

onde $L(S|\mathcal{D}) = \sum_{i=1}^n [\delta_i \log(1 - S(T_i)) + (1 - \delta_i) \log S(T_i)]$ e $\tilde{S}^{(-v)}(\cdot; h)$ é a estimativa suavizada via núcleo estimadores da função de sobrevivência S a partir das observações em $\mathcal{D} - \mathcal{D}^{(v)}$, $v = 1, \dots, V$, onde \mathcal{D} representa o conjunto de dados subdividido, aleatoriamente, em V subconjuntos de mesmo tamanho, isto é, $\{\mathcal{D}^{(1)}, \dots, \mathcal{D}^{(V)}\}$ formam uma partição de \mathcal{D} .

Capítulo 4

Simulação e Aplicação

4.1 Simulação

Em todas as situações onde técnicas de suavização via núcleo estimadores foram utilizadas consideramos o núcleo de Epanechnikov, isto é,

$$K(t) = \begin{cases} 3/4(1 - t^2), & \text{se } -1 \leq t \leq 1 \\ 0, & \text{se } t < -1 \text{ ou } t > 1 \end{cases}$$

Para as simulações, foram utilizadas amostras provenientes das seguintes distribuições:

- **Caso 1:** $X \sim \exp(1)$ e $T \sim U(0, 4.6)$.
- **Caso 2:** $X \sim N(4, 1)$ e $T \sim N(4, 1)$.
- **Caso 3:** $X \sim U(0, 1)$ e $T \sim U(0, 1)$.

Embora há outras distribuições mais adequadas para descrição do tempo de falha, a escolha das distribuições acima teve por motivação a investigação de problemas técnicos na implementação dos métodos.

As Tabelas 4.1 a 4.44 mostram as probabilidades de cobertura para cada caso simulado em cada método de obtenção dos intervalos de confiança para valores nominais de confiança de 90%, 95% e 99%, além da variância e do vício para cada método e da amplitude média dos intervalos de confiança. Todas estas quantidades foram

calculadas nos decis da verdadeira distribuição do tempo de falha em cada caso simulado, para vários tamanhos de amostra ($n=60, 100, 200$ e 500). Para cada combinação de caso, método e tamanho de amostra, foram geradas de 2000 a 5000 amostras e os métodos foram implementados em linguagem C.

Nestas tabelas, as colunas denotadas por pc90, pc95 e pc99 fornecem a estimativa da probabilidade real de cobertura dos intervalos de 90% , 95% e 99% de confiança.

As colunas denotadas por amp90, amp95 e amp99 fornecem os valores das amplitudes médias dos intervalos de 90% , 95% e 99% de confiança e a coluna denotada por k representa a ordem dos decis.

A tabela 4.45 mostra as variâncias das estimativas das janelas ótimas para a estimação das funções densidade de probabilidade das variáveis X (instante de falha) e T (instante de censura ou observação).

As informações contidas nas Tabelas 4.1 a 4.44 podem ser visualizadas através das Figuras 4.1 a 4.11. Cada figura mostra (para cada combinação de caso simulado, tamanho de amostra e probabilidade nominal de cobertura) a probabilidade real de cobertura dos intervalos de confiança obtidos pela aplicação de cada um dos quatro métodos aos dados simulados.

Na aplicação do método proposto por Pan(2000a) para o cálculo da janela ótima na estimativa da função densidade de probabilidade f do tempo de falha, cada amostra gerada foi dividida em $v = 3$ grupos quando $n = 60$, $v = 4$ grupos quando $n = 100$ ou 200 e $v = 5$ grupos quando $n = 500$.

Na aplicação do método de imputação foram utilizadas $M = 10$ imputações para cada observação com censura à esquerda.

Tabela 4.1: Caso 1, $n = 60$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0574	0.0325197	0.0598	0.0368769	0.0640	0.0436896	0.0047520	-0.0834443
2	0.2314	0.1295345	0.2460	0.1482871	0.2578	0.1778274	0.0219761	-0.1215602
3	0.4760	0.2558018	0.4986	0.2948631	0.5208	0.3578420	0.0418910	-0.1161547
4	0.6270	0.3729865	0.6850	0.4324339	0.7422	0.5301540	0.0503341	-0.0912673
5	0.7318	0.4511368	0.7966	0.5252094	0.8604	0.6492555	0.0450752	-0.0584485
6	0.7832	0.4618640	0.8482	0.5378263	0.9170	0.6654673	0.0350563	-0.0357799
7	0.8222	0.4259643	0.8772	0.4929782	0.9348	0.6039604	0.0252038	-0.0165775
8	0.8220	0.3363851	0.8672	0.3846184	0.9126	0.4633502	0.0162042	-0.0026980
9	0.6932	0.1771534	0.7060	0.1999438	0.7120	0.2378891	0.0067417	0.0127827

Tabela 4.2: Caso 1, $n = 100$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0944	0.0409688	0.1016	0.0465784	0.1074	0.0555528	0.0057748	-0.0758587
2	0.3516	0.1572734	0.3648	0.1808599	0.3822	0.2192606	0.0215294	-0.0985532
3	0.6046	0.2883286	0.6436	0.3343786	0.6852	0.4118490	0.0336655	-0.0818884
4	0.7170	0.3811776	0.7854	0.4454763	0.8544	0.5573600	0.0336903	-0.0531910
5	0.7864	0.4173073	0.8500	0.4899645	0.9252	0.6188958	0.0274521	-0.0331523
6	0.8184	0.4102886	0.8778	0.4814778	0.9492	0.6074302	0.0221552	-0.0181650
7	0.8522	0.3778836	0.9100	0.4410723	0.9612	0.5485823	0.0165383	-0.0079443
8	0.8554	0.3068130	0.9088	0.3536706	0.9578	0.4303028	0.0103121	0.0015083
9	0.8156	0.1731313	0.8538	0.1963466	0.8666	0.2344833	0.0043259	0.0094515

Tabela 4.3: Caso 1, $n = 200$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.1840	0.0606870	0.1922	0.0691858	0.2026	0.0830032	0.0060881	-0.0647988
2	0.5592	0.1942717	0.5952	0.2243232	0.6186	0.2740326	0.0171800	-0.0647142
3	0.7222	0.3049332	0.7890	0.3563484	0.8620	0.4459453	0.0191853	-0.0382892
4	0.7838	0.3386001	0.8430	0.3987816	0.9206	0.5082477	0.0177184	-0.0273444
5	0.8232	0.3453678	0.8866	0.4077779	0.9516	0.5243422	0.0155732	-0.0161699
6	0.8588	0.3382087	0.9108	0.3992713	0.9712	0.5127572	0.0125701	-0.0084694
7	0.8804	0.3134538	0.9342	0.3691948	0.9824	0.4696000	0.0093378	-0.0051515
8	0.8938	0.2547631	0.9374	0.2975910	0.9804	0.3703725	0.0060372	0.0001116
9	0.8598	0.1530639	0.9148	0.1756484	0.9546	0.2118409	0.0025046	0.0057516

Tabela 4.4: Caso 1, $n = 500$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.4084	0.0910176	0.4212	0.1040181	0.4306	0.1247176	0.0051605	-0.0467869
2	0.7204	0.2112424	0.7832	0.2463591	0.8552	0.3065935	0.0090364	-0.0300903
3	0.7886	0.2500391	0.8524	0.2947837	0.9268	0.3770566	0.0091074	-0.0163454
4	0.8192	0.2609751	0.8826	0.3081970	0.9582	0.3969768	0.0087727	-0.0136968
5	0.8596	0.2649010	0.9156	0.3128921	0.9730	0.4032984	0.0077500	-0.0090607
6	0.8862	0.2580324	0.9382	0.3047792	0.9800	0.3928622	0.0065916	-0.0053808
7	0.8932	0.2346200	0.9402	0.2771244	0.9850	0.3570697	0.0049917	-0.0009661
8	0.8920	0.1867770	0.9462	0.2205007	0.9838	0.2828747	0.0031849	0.0005374
9	0.8786	0.1183212	0.9280	0.1384838	0.9762	0.1731570	0.0013089	0.0031672

Tabela 4.5: Caso 2, $n = 60$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.2520	0.0816313	0.2594	0.0921537	0.2632	0.1091794	0.0063289	-0.0582105
2	0.6476	0.2357171	0.6738	0.2699246	0.6956	0.3252133	0.0159868	-0.0550582
3	0.7876	0.3477798	0.8412	0.4041479	0.9008	0.4987307	0.0192184	-0.0343028
4	0.8252	0.3982822	0.8848	0.4674483	0.9506	0.5894751	0.0197572	-0.0178557
5	0.8400	0.4112450	0.8968	0.4844876	0.9586	0.6173344	0.0198744	-0.0134983
6	0.8272	0.4028706	0.8854	0.4733959	0.9526	0.5991138	0.0203985	-0.0012303
7	0.8024	0.3634368	0.8624	0.4232565	0.9216	0.5254222	0.0194634	0.0140300
8	0.7180	0.2711662	0.7522	0.3114852	0.7776	0.3772208	0.0169567	0.0292572
9	0.3404	0.1173161	0.3528	0.1323544	0.3594	0.1567784	0.0085504	0.0401127

Tabela 4.6: Caso 2, $n = 100$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.3764	0.1027192	0.3854	0.1159441	0.3886	0.1375793	0.0056019	-0.0489323
2	0.7336	0.2474584	0.7884	0.2852957	0.8462	0.3476499	0.0119315	-0.0347466
3	0.8132	0.3188433	0.8770	0.3734386	0.9394	0.4693533	0.0129074	-0.0174628
4	0.8536	0.3441589	0.9074	0.4059441	0.9632	0.5200128	0.0132036	-0.0108383
5	0.8618	0.3508012	0.9168	0.4142517	0.9718	0.5330980	0.0133505	-0.0082226
6	0.8538	0.3455727	0.9068	0.4077399	0.9656	0.5227293	0.0134730	0.0002999
7	0.8250	0.3224501	0.8870	0.3782602	0.9460	0.4760413	0.0128443	0.0113314
8	0.7752	0.2639479	0.8268	0.3049399	0.8828	0.3724879	0.0115408	0.0219557
9	0.4542	0.1220428	0.4638	0.1381823	0.4732	0.1641609	0.0068625	0.0346112

Tabela 4.7: Caso 2, $n = 200$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.6033	0.1213087	0.6223	0.1378447	0.6306	0.1639910	0.0044883	-0.0323805
2	0.8126	0.2278160	0.8653	0.2658421	0.9230	0.3306973	0.0068046	-0.0188040
3	0.8553	0.2621069	0.9146	0.3092987	0.9670	0.3966183	0.0073898	-0.0086596
4	0.8703	0.2748672	0.9260	0.3246639	0.9806	0.4183614	0.0076249	-0.0062739
5	0.8646	0.2783429	0.9216	0.3287594	0.9753	0.4237367	0.0082865	-0.0039154
6	0.8576	0.2744775	0.9183	0.3241968	0.9766	0.4177645	0.0081980	0.0016105
7	0.8540	0.2617087	0.9126	0.3087645	0.9710	0.3955410	0.0074933	0.0078960
8	0.8393	0.2320138	0.8943	0.2713677	0.9463	0.3382477	0.0063557	0.0119179
9	0.6613	0.1357118	0.6803	0.1544844	0.6876	0.1845954	0.0047583	0.0234295

Tabela 4.8: Caso 2, $n = 500$, ENPMV

k	c90	amp90	pc95	amp95	pc99	amp99	var	vício
1	0.7983	0.1302991	0.8430	0.1504143	0.8980	0.1827277	0.0023119	-0.0148201
2	0.8580	0.1776735	0.9163	0.2096892	0.9670	0.2689262	0.0032870	-0.0073384
3	0.8733	0.1942868	0.9283	0.2294850	0.9813	0.2958029	0.0038200	-0.0059378
4	0.8786	0.2017338	0.9260	0.2382811	0.9770	0.3071458	0.0042677	-0.0026083
5	0.8813	0.2041752	0.9343	0.2411649	0.9823	0.3108630	0.0042258	-0.0019199
6	0.8746	0.2013598	0.9330	0.2378394	0.9796	0.3065764	0.0042962	0.0007017
7	0.8740	0.1934482	0.9313	0.2284945	0.9806	0.2945211	0.0038190	0.0041759
8	0.8730	0.1774304	0.9250	0.2094227	0.9770	0.2688148	0.0031738	0.0058170
9	0.7996	0.1328723	0.8453	0.1534481	0.9003	0.1869444	0.0024722	0.0125200

Tabela 4.9: Caso 3, $n = 100$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vicio
1	0.3766	0.0967016	0.3862	0.1098388	0.3924	0.1309720	0.0058201	-0.0473592
2	0.7424	0.2412468	0.7980	0.2787878	0.8486	0.3418207	0.0114759	-0.0317760
3	0.8142	0.3188638	0.8762	0.3732419	0.9428	0.4691180	0.0129091	-0.0208488
4	0.8584	0.3514065	0.9144	0.4143192	0.9684	0.5298300	0.0133489	-0.0124637
5	0.8694	0.3596135	0.9242	0.4245660	0.9780	0.5459388	0.0134596	-0.0080297
6	0.8652	0.3548891	0.9174	0.4185815	0.9698	0.5358342	0.0137356	-0.0016930
7	0.8364	0.3273188	0.8922	0.3838756	0.9516	0.4837275	0.0130725	0.0058907
8	0.7520	0.2486037	0.8096	0.2878911	0.8666	0.3537305	0.0119057	0.0210512
9	0.4296	0.1083271	0.4442	0.1232656	0.4546	0.1473416	0.0068436	0.0365731

Tabela 4.10: Caso 3, $n = 200$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.6028	0.1128387	0.6262	0.1292002	0.6398	0.1554361	0.0043661	-0.0314224
2	0.8096	0.2217438	0.8602	0.2592725	0.9280	0.3233854	0.0062856	-0.0201094
3	0.8586	0.2678570	0.9178	0.3159831	0.9726	0.4040580	0.0075119	-0.0113082
4	0.8856	0.2832282	0.9356	0.3345168	0.9802	0.4308930	0.0077565	-0.0068412
5	0.8790	0.2861813	0.9344	0.3380255	0.9814	0.4356936	0.0082166	-0.0021098
6	0.8768	0.2835819	0.9304	0.3349478	0.9770	0.4315867	0.0080636	0.0002725
7	0.8578	0.2684497	0.9158	0.3166613	0.9712	0.4055300	0.0077145	0.0054138
8	0.8198	0.2250397	0.8756	0.2631229	0.9310	0.3287876	0.0065762	0.0135391
9	0.6176	0.1193765	0.6454	0.1365764	0.6612	0.1643716	0.0045941	0.0275760

Tabela 4.11: Caso 3, $n = 500$, ENPMV

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.7420	0.1171516	0.8006	0.1359662	0.8660	0.1673393	0.0024306	-0.0150278
2	0.8558	0.1750763	0.9096	0.2065804	0.9660	0.2648462	0.0032737	-0.0086494
3	0.8840	0.2003814	0.9352	0.2366837	0.9802	0.3050532	0.0039569	-0.0055701
4	0.8902	0.2090492	0.9382	0.2469219	0.9824	0.3182838	0.0042766	-0.0028825
5	0.8972	0.2126964	0.9470	0.2512298	0.9868	0.3238368	0.0042191	-0.0008313
6	0.8872	0.2091025	0.9372	0.2469848	0.9820	0.3183649	0.0041757	0.0015337
7	0.8824	0.2011467	0.9394	0.2375827	0.9790	0.3061888	0.0040403	0.0016692
8	0.8564	0.1765995	0.9060	0.2084300	0.9676	0.2674009	0.0032965	0.0065281
9	0.7584	0.1206677	0.8200	0.1402764	0.8854	0.1729738	0.0024207	0.0128446

Tabela 4.12: Caso 1, $n = 60$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9430	0.4517493	0.9780	0.5083909	0.9885	0.5988998	0.0066044	0.1258667
2	0.9870	0.4750471	0.9890	0.5391922	0.9915	0.6377871	0.0063669	0.0583764
3	0.9885	0.4952163	0.9925	0.5689092	0.9945	0.6802164	0.0058087	-0.0011158
4	0.9840	0.5077203	0.9910	0.5912947	0.9925	0.7220680	0.0054615	-0.0515908
5	0.9765	0.5125473	0.9845	0.6021712	0.9925	0.7550863	0.0053328	-0.0916649
6	0.9680	0.5057789	0.9865	0.5963557	0.9940	0.7590747	0.0052354	-0.1169214
7	0.9460	0.4836902	0.9835	0.5692765	0.9935	0.7199594	0.0061019	-0.1218622
8	0.9310	0.4245682	0.9745	0.4944707	0.9910	0.6075860	0.0071646	-0.1014140
9	0.9140	0.2823162	0.9490	0.3199885	0.9595	0.3781834	0.0054100	-0.0540377

Tabela 4.13: Caso 1, $n = 100$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9330	0.4041099	0.9860	0.4589033	0.9950	0.5413489	0.0044035	0.1191837
2	0.9925	0.4205571	0.9940	0.4841724	0.9975	0.5792557	0.0042637	0.0546959
3	0.9950	0.4312292	0.9960	0.5038513	0.9980	0.6178351	0.0039920	-0.0016560
4	0.9855	0.4359263	0.9940	0.5133809	0.9960	0.6478670	0.0037890	-0.0489514
5	0.9775	0.4366608	0.9915	0.5154275	0.9965	0.6606938	0.0037057	-0.0842962
6	0.9590	0.4295327	0.9875	0.5072951	0.9970	0.6530068	0.0040399	-0.1038450
7	0.9480	0.4099256	0.9825	0.4839686	0.9970	0.6200961	0.0047099	-0.1035563
8	0.9495	0.3553018	0.9845	0.4174945	0.9970	0.5229366	0.0048544	-0.0804281
9	0.9620	0.2375297	0.9795	0.2722891	0.9860	0.3223969	0.0028246	-0.0387187

Tabela 4.14: Caso 1, $n = 200$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9005	0.3362243	0.9790	0.3894907	0.9975	0.4682151	0.0021863	0.1119068
2	0.9970	0.3432295	0.9985	0.4035557	0.9990	0.5007612	0.0020574	0.0507338
3	0.9995	0.3466332	1.0000	0.4092356	1.0000	0.5226404	0.0018620	-0.0017329
4	0.9965	0.3488528	0.9985	0.4119826	0.9985	0.5305305	0.0017796	-0.0441118
5	0.9895	0.3503521	0.9980	0.4138240	0.9995	0.5334217	0.0018652	-0.0737865
6	0.9720	0.3480362	0.9965	0.4110887	0.9990	0.5298013	0.0022301	-0.0869410
7	0.9570	0.3298248	0.9920	0.3895666	0.9980	0.5020489	0.0026872	-0.0805923
8	0.9720	0.2777659	0.9920	0.3280344	0.9975	0.4210298	0.0025129	-0.0568507
9	0.9810	0.1820145	0.9900	0.2132594	0.9940	0.2610822	0.0012642	-0.0251870

Tabela 4.15: Caso 1, $n = 500$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.8633	0.2504077	0.9596	0.2952742	0.9960	0.3754881	0.0011327	0.0974532
2	0.9973	0.2544043	0.9980	0.3004442	0.9983	0.3862389	0.0010018	0.0413396
3	0.9966	0.2578775	0.9966	0.3045861	0.9973	0.3925093	0.0008618	-0.0054423
4	0.9946	0.2616548	0.9970	0.3090578	0.9973	0.3983621	0.0008981	-0.0406829
5	0.9873	0.2660510	0.9973	0.3142505	0.9983	0.4050708	0.0010365	-0.0616431
6	0.9723	0.2629769	0.9936	0.3106195	0.9963	0.4003904	0.0011959	-0.0663495
7	0.9733	0.2432538	0.9940	0.2873232	0.9970	0.3703478	0.0012716	-0.0539217
8	0.9860	0.1960882	0.9950	0.2316128	0.9956	0.2985399	0.0009829	-0.0339408
9	0.9906	0.1277002	0.9940	0.1508144	0.9956	0.1936328	0.0004789	-0.0147288

Tabela 4.16: Caso 2, $n = 60$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9600	0.2852249	0.9740	0.3192216	0.9845	0.3763403	0.0045187	0.0337543
2	0.9795	0.3671055	0.9905	0.4231939	0.9930	0.5080831	0.0049407	0.0303247
3	0.9915	0.4013456	0.9945	0.4718320	0.9965	0.5915748	0.0044776	0.0183517
4	0.9940	0.4171666	0.9950	0.4926044	0.9960	0.6318039	0.0039103	0.0044204
5	0.9950	0.4224432	0.9955	0.4989249	0.9970	0.6426132	0.0037667	-0.0115022
6	0.9925	0.4203698	0.9955	0.4963745	0.9965	0.6380415	0.0039768	-0.0270721
7	0.9865	0.4073426	0.9910	0.4797445	0.9955	0.6060259	0.0046590	-0.0402795
8	0.9780	0.3802807	0.9895	0.4407048	0.9960	0.5333478	0.0051800	-0.0506414
9	0.9600	0.3062487	0.9790	0.3436725	0.9905	0.4050634	0.0049949	-0.0503452

Tabela 4.17: Caso 2, $n = 100$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9805	0.2585269	0.9895	0.2902350	0.9915	0.3406101	0.0027308	0.0304309
2	0.9900	0.3193851	0.9945	0.3745054	0.9950	0.4622227	0.0029908	0.0302037
3	0.9935	0.3408129	0.9950	0.4024895	0.9955	0.5166529	0.0027490	0.0210176
4	0.9950	0.3522808	0.9960	0.4161022	0.9960	0.5363103	0.0024619	0.0086433
5	0.9940	0.3565423	0.9950	0.4211358	0.9955	0.5428465	0.0024071	-0.0060230
6	0.9925	0.3535351	0.9935	0.4175838	0.9940	0.5382155	0.0025375	-0.0198317
7	0.9900	0.3431241	0.9930	0.4052691	0.9940	0.5209353	0.0027385	-0.0323024
8	0.9855	0.3245793	0.9935	0.3812087	0.9955	0.4730891	0.0029834	-0.0413529
9	0.9745	0.2678524	0.9865	0.3018585	0.9900	0.3549004	0.0028201	-0.0393860

Tabela 4.18: Caso 2, $n = 200$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9920	0.2209215	0.9960	0.2537311	0.9980	0.2997051	0.0012199	0.0300172
2	0.9960	0.2558131	0.9985	0.3021323	0.9990	0.3880395	0.0013298	0.0319075
3	0.9995	0.2709768	0.9995	0.3200687	1.0000	0.4125705	0.0012857	0.0236126
4	0.9995	0.2788815	0.9995	0.3294054	0.9995	0.4246056	0.0012356	0.0118135
5	1.0000	0.2811509	1.0000	0.3320860	1.0000	0.4280260	0.0012419	-0.0021875
6	0.9980	0.2787368	0.9985	0.3292345	0.9985	0.4243853	0.0013070	-0.0163154
7	0.9980	0.2718406	0.9985	0.3210822	0.9990	0.4138106	0.0013685	-0.0285039
8	0.9955	0.2567004	0.9985	0.3031678	0.9995	0.3895856	0.0014368	-0.0371407
9	0.9890	0.2238918	0.9960	0.2579007	0.9980	0.3061132	0.0013113	-0.0351302

Tabela 4.19: Caso 2, $n = 500$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9905	0.1671586	0.9970	0.1969544	0.9975	0.2500102	0.0007223	0.0331180
2	0.9925	0.1887396	0.9980	0.2229328	0.9995	0.2873429	0.0007344	0.0350010
3	0.9995	0.1990707	0.9995	0.2351356	1.0000	0.3030911	0.0006323	0.0264476
4	0.9985	0.2037561	0.9990	0.2406699	0.9990	0.3102249	0.0005388	0.0142726
5	0.9990	0.2052108	0.9995	0.2423881	0.9995	0.3124397	0.0005199	-0.0004755
6	0.9985	0.2033196	0.9990	0.2401543	0.9990	0.3095603	0.0005737	-0.0155167
7	0.9980	0.1985089	1.0000	0.2344720	1.0000	0.3022358	0.0006643	-0.0283296
8	0.9940	0.1886429	0.9985	0.2228187	0.9990	0.2872033	0.0007368	-0.0374077
9	0.9910	0.1681257	0.9985	0.1981839	0.9985	0.2518097	0.0007132	-0.0354649

Tabela 4.20: Caso 3, $n = 100$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9560	0.2348451	0.9715	0.2654965	0.9780	0.3134067	0.0028179	0.0229511
2	0.9715	0.2932203	0.9840	0.3416071	0.9905	0.4166577	0.0031002	-0.0011308
3	0.9820	0.3357912	0.9920	0.3960892	0.9950	0.5038428	0.0032320	-0.0093954
4	0.9855	0.3586087	0.9900	0.4235074	0.9940	0.5453728	0.0034036	-0.0091792
5	0.9855	0.3649767	0.9905	0.4310982	0.9950	0.5556232	0.0035681	-0.0063068
6	0.9890	0.3614687	0.9950	0.4269291	0.9970	0.5499852	0.0035265	-0.0032650
7	0.9885	0.3413751	0.9925	0.4027892	0.9955	0.5136775	0.0032417	-0.0015043
8	0.9795	0.2987918	0.9900	0.3486259	0.9940	0.4272089	0.0030599	-0.0083584
9	0.9605	0.2423803	0.9775	0.2747092	0.9875	0.3246007	0.0028883	-0.0304305

Tabela 4.21: Caso 3, $n = 200$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9835	0.1892714	0.9915	0.2173557	0.9955	0.2586204	0.0012854	0.0159257
2	0.9855	0.2368100	0.9920	0.2793276	0.9960	0.3537351	0.0014834	-0.0032563
3	0.9870	0.2724647	0.9895	0.3218261	0.9935	0.4147109	0.0017619	-0.0075378
4	0.9910	0.2857833	0.9930	0.3375576	0.9950	0.4351138	0.0019669	-0.0053834
5	0.9935	0.2886137	0.9960	0.3409007	0.9970	0.4394231	0.0020472	-0.0028760
6	0.9915	0.2862602	0.9935	0.3381209	0.9955	0.4358398	0.0020004	0.0000867
7	0.9925	0.2730069	0.9955	0.3224552	0.9960	0.4155275	0.0017759	0.0030925
8	0.9840	0.2397427	0.9900	0.2828420	0.9930	0.3593308	0.0015906	0.0000005
9	0.9810	0.1936648	0.9895	0.2228292	0.9930	0.2655462	0.0013815	-0.0185349

Tabela 4.22: Caso 3, $n = 500$, Kernel

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9945	0.1408856	0.9965	0.1662388	0.9980	0.2093888	0.0004746	0.0086525
2	0.9945	0.1780760	0.9965	0.2103269	0.9975	0.2710840	0.0006634	-0.0059394
3	0.9935	0.2016370	0.9965	0.2381668	0.9970	0.3069985	0.0008581	-0.0058165
4	0.9960	0.2102798	0.9980	0.2483754	0.9980	0.3201574	0.0009552	-0.0028179
5	0.9965	0.2136381	0.9980	0.2523421	0.9985	0.3252705	0.0009952	-0.0010877
6	0.9960	0.2102637	0.9975	0.2483564	0.9975	0.3201329	0.0009597	0.0003365
7	0.9970	0.2023444	0.9985	0.2390024	0.9985	0.3080722	0.0008549	0.0028645
8	0.9960	0.1793592	0.9975	0.2118531	0.9980	0.2730661	0.0006618	0.0029258
9	0.9975	0.1424718	0.9985	0.1681340	0.9985	0.2123292	0.0004556	-0.0111494

Tabela 4.23: Caso 1, $n = 60$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0656	0.0340556	0.0656	0.0378912	0.0656	0.0451108	0.0024913	-0.0875396
2	0.2618	0.1445217	0.2634	0.1614351	0.2638	0.1907546	0.0117800	-0.1408067
3	0.5120	0.3083508	0.5230	0.3459833	0.5306	0.4042968	0.0248891	-0.1555379
4	0.6982	0.4692935	0.7258	0.5307347	0.7536	0.6196246	0.0351143	-0.1394048
5	0.8138	0.5730112	0.8496	0.6549732	0.8882	0.7734470	0.0360277	-0.1032137
6	0.8688	0.5670949	0.9088	0.6558813	0.9522	0.7889888	0.0294669	-0.0657090
7	0.8848	0.4849683	0.9236	0.5613507	0.9620	0.6831366	0.0209638	-0.0297361
8	0.8468	0.3583360	0.8894	0.4108080	0.9320	0.4967883	0.0125850	-0.0032453
9	0.7272	0.1842286	0.7542	0.2083219	0.7904	0.2504570	0.0049823	0.0164474

Tabela 4.24: Caso 1, $n = 100$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.1102	0.0428353	0.1104	0.0477856	0.1104	0.0573453	0.0029222	-0.0822061
2	0.3818	0.1789973	0.3866	0.2006921	0.3882	0.2395422	0.0120761	-0.1221318
3	0.6416	0.3496546	0.6662	0.3965191	0.6884	0.4742032	0.0220995	-0.1208942
4	0.7944	0.4752104	0.8332	0.5477846	0.8720	0.6646309	0.0256102	-0.0932675
5	0.8676	0.5034698	0.9078	0.5895360	0.9498	0.7338328	0.0222625	-0.0612579
6	0.8878	0.4615429	0.9298	0.5448130	0.9730	0.6903091	0.0175923	-0.0332760
7	0.8926	0.3934566	0.9326	0.4636641	0.9680	0.5848556	0.0125871	-0.0134849
8	0.8604	0.2997467	0.9010	0.3492576	0.9466	0.4301489	0.0078765	0.0025346
9	0.7774	0.1666664	0.8118	0.1896951	0.8490	0.2278587	0.0032397	0.0145489

Tabela 4.25: Caso 1, $n = 200$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.2058	0.0637208	0.2060	0.0712294	0.2060	0.0853369	0.0030970	-0.0739645
2	0.5844	0.2227236	0.6020	0.2512426	0.6204	0.3008792	0.0108709	-0.0907302
3	0.7874	0.3707253	0.8322	0.4271382	0.8790	0.5217722	0.0143288	-0.0687534
4	0.8600	0.4023482	0.9040	0.4748232	0.9498	0.6028749	0.0137668	-0.0486373
5	0.8884	0.3833837	0.9308	0.4562549	0.9734	0.5948744	0.0116709	-0.0290124
6	0.9042	0.3487318	0.9440	0.4152774	0.9836	0.5431213	0.0092130	-0.0154100
7	0.9034	0.2993170	0.9434	0.3561385	0.9788	0.4632339	0.0067655	-0.0065050
8	0.8894	0.2348870	0.9292	0.2784320	0.9658	0.3554964	0.0043784	0.0014904
9	0.8310	0.1429881	0.8668	0.1661555	0.9174	0.2030254	0.0018757	0.0092186

Tabela 4.26: Caso 1, $n = 500$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.4204	0.0952627	0.4282	0.1065932	0.4314	0.1276058	0.0029599	-0.0590676
2	0.7652	0.2465006	0.8128	0.2839030	0.8618	0.3464597	0.0066284	-0.0505684
3	0.8686	0.2896442	0.9150	0.3428889	0.9538	0.4400017	0.0067820	-0.0305425
4	0.8954	0.2866528	0.9366	0.3414928	0.9768	0.4477199	0.0062809	-0.0220560
5	0.9142	0.2721412	0.9486	0.3242485	0.9826	0.4258867	0.0055485	-0.0142295
6	0.9144	0.2505657	0.9550	0.2985463	0.9880	0.3922125	0.0045905	-0.0078062
7	0.9036	0.2153455	0.9440	0.2565818	0.9812	0.3370141	0.0035002	-0.0017505
8	0.8956	0.1701397	0.9410	0.2026679	0.9792	0.2656276	0.0022497	0.0015940
9	0.8656	0.1076998	0.9056	0.1278894	0.9508	0.1644830	0.0009356	0.0046862

Tabela 4.27: Caso 2, $n = 60$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.2630	0.0759689	0.2636	0.0849220	0.2636	0.1018134	0.0034461	-0.0681245
2	0.6410	0.2427683	0.6646	0.2745159	0.6910	0.3292659	0.0107936	-0.0788698
3	0.8026	0.3747551	0.8436	0.4331809	0.8882	0.5299055	0.0149069	-0.0589227
4	0.8708	0.4325368	0.9128	0.5095652	0.9556	0.6430204	0.0154665	-0.0353309
5	0.8866	0.4425195	0.9298	0.5252027	0.9718	0.6750602	0.0154261	-0.0196313
6	0.8902	0.4389775	0.9300	0.5181907	0.9730	0.6594512	0.0156519	0.0001327
7	0.8608	0.4068117	0.9064	0.4728145	0.9522	0.5848777	0.0147294	0.0229986
8	0.7812	0.3159948	0.8184	0.3601298	0.8640	0.4356745	0.0116337	0.0412877
9	0.5336	0.1583157	0.5632	0.1789022	0.6046	0.4356745	0.0049940	0.0435149

Tabela 4.28: Caso 2, $n = 100$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.3840	0.0957525	0.3872	0.1071775	0.3892	0.1283872	0.0032006	-0.0603392
2	0.7320	0.2543433	0.7698	0.2909614	0.8170	0.3521745	0.0086062	-0.0561088
3	0.8452	0.3336844	0.8882	0.3921694	0.9340	0.4936883	0.0100055	-0.0343897
4	0.8898	0.3577598	0.9328	0.4255052	0.9704	0.5524604	0.0099565	-0.0200167
5	0.9040	0.3634667	0.9454	0.4328059	0.9812	0.5661989	0.0098670	-0.0114906
6	0.9042	0.3623523	0.9456	0.4309539	0.9790	0.5599734	0.0100224	0.0020601
7	0.8774	0.3434417	0.9204	0.4043666	0.9638	0.5120257	0.0096192	0.0179613
8	0.8044	0.2851886	0.8504	0.3276164	0.8998	0.3990860	0.0083552	0.0345633
9	0.5702	0.1426550	0.5908	0.1610876	0.6180	0.1948897	0.0040164	0.0426189

Tabela 4.29: Caso 2, $n = 200$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.5813	0.1145708	0.5996	0.1288966	0.6216	0.1542058	0.0028183	-0.0453852
2	0.8113	0.2313622	0.8603	0.2703469	0.9106	0.3365063	0.0051469	-0.0326675
3	0.8886	0.2649878	0.9320	0.3153610	0.9696	0.4106753	0.0055174	-0.0171272
4	0.9120	0.2765791	0.9473	0.3295206	0.9830	0.4326750	0.0055138	-0.0105677
5	0.9150	0.2824224	0.9550	0.3365032	0.9886	0.4420342	0.0058528	-0.0061389
6	0.9010	0.2804817	0.9443	0.3341519	0.9843	0.4385512	0.0059139	0.0029377
7	0.8956	0.2668082	0.9373	0.3173574	0.9753	0.4130074	0.0054236	0.0122107
8	0.8523	0.2366081	0.8983	0.2773882	0.9500	0.3468286	0.0048349	0.0216477
9	0.6870	0.1409731	0.7153	0.1594673	0.7493	0.1919382	0.0030043	0.0341634

Tabela 4.30: Caso 2, $n = 500$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.7606	0.1257044	0.8083	0.1446057	0.8586	0.1755933	0.0017187	-0.0247498
2	0.8703	0.1752843	0.9170	0.2086421	0.9656	0.2719054	0.0023813	-0.0136638
3	0.9080	0.1924130	0.9453	0.2292581	0.9816	0.3011395	0.0027091	-0.0097549
4	0.9083	0.2004052	0.9560	0.2387806	0.9826	0.3137042	0.0029859	-0.0048872
5	0.9213	0.2041349	0.9610	0.2432246	0.9893	0.3195425	0.0029808	-0.0027229
6	0.9166	0.2030393	0.9563	0.2419192	0.9890	0.3178275	0.0030103	0.0015606
7	0.9110	0.1919933	0.9486	0.2287580	0.9866	0.3005151	0.0026736	0.0065267
8	0.8900	0.1742514	0.9336	0.2074607	0.9746	0.2706993	0.0022933	0.0104029
9	0.7903	0.1314045	0.8280	0.1515511	0.8796	0.1846542	0.0017709	0.0208326

Tabela 4.31: Caso 3, $n = 100$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.3904	0.0971551	0.3928	0.1086466	0.3944	0.1300850	0.0032539	-0.0594915
2	0.7454	0.2583614	0.7866	0.2952978	0.8314	0.3574686	0.0083211	-0.0538323
3	0.8346	0.3289117	0.8840	0.3867602	0.9352	0.4875703	0.0099889	-0.0363600
4	0.8842	0.3567523	0.9290	0.4241273	0.9714	0.5505264	0.0101253	-0.0220763
5	0.8996	0.3640085	0.9370	0.4336085	0.9806	0.5676872	0.0101420	-0.0120601
6	0.8964	0.3606846	0.9418	0.4288483	0.9768	0.5572763	0.0102908	-0.0006825
7	0.8844	0.3459636	0.9262	0.4072488	0.9648	0.5158133	0.0098511	0.0137125
8	0.8034	0.2815904	0.8476	0.3240109	0.8980	0.3950984	0.0086376	0.0341821
9	0.5552	0.1394169	0.5792	0.1574372	0.6066	0.1907339	0.0040559	0.0440501

Tabela 4.32: Caso 3, $n = 200$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.5972	0.1169640	0.6184	0.1314014	0.6354	0.1572897	0.0027713	-0.0445994
2	0.8260	0.2332633	0.8664	0.2725558	0.9206	0.3384623	0.0047785	-0.0331168
3	0.8806	0.2640588	0.9256	0.3141204	0.9698	0.4089367	0.0054583	-0.0197276
4	0.9092	0.2790854	0.9490	0.3325201	0.9822	0.4365768	0.0055732	-0.0115902
5	0.9132	0.2810649	0.9512	0.3348858	0.9852	0.4399269	0.0058815	-0.0044901
6	0.9104	0.2796240	0.9482	0.3331277	0.9834	0.4371329	0.0057225	0.0016004
7	0.8908	0.2674368	0.9382	0.3181669	0.9782	0.4145420	0.0055811	0.0101439
8	0.8514	0.2383426	0.8954	0.2792362	0.9442	0.3492740	0.0048919	0.0230618
9	0.6584	0.1358082	0.6884	0.1535603	0.7246	0.1849308	0.0029130	0.0375013

Tabela 4.33: Caso 3, $n = 500$, Bootstrap

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.7482	0.1250852	0.7948	0.1438374	0.8510	0.1747758	0.0017622	-0.0251937
2	0.8754	0.1743912	0.9212	0.2075253	0.9620	0.2703651	0.0023684	-0.0149455
3	0.9048	0.1908853	0.9426	0.2274378	0.9794	0.2987585	0.0028148	-0.0093372
4	0.9108	0.2013461	0.9526	0.2399017	0.9902	0.3151770	0.0029811	-0.0049594
5	0.9210	0.2052821	0.9610	0.2445914	0.9896	0.3213382	0.0029752	-0.0017024
6	0.9170	0.2008728	0.9536	0.2393378	0.9870	0.3144361	0.0029578	0.0019763
7	0.9076	0.1929040	0.9468	0.2298430	0.9850	0.3019029	0.0028462	0.0047569
8	0.8808	0.1756009	0.9274	0.2090695	0.9690	0.2728056	0.0023822	0.0111306
9	0.7832	0.1316219	0.8284	0.1519758	0.8856	0.1851489	0.0017650	0.0214538

Tabela 4.34: Caso 1, $n = 60$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.8462	0.0988049	0.9766	0.1248463	0.9996	0.1880098	0.0001553	-0.0523361
2	0.2650	0.1333712	0.6478	0.1627990	0.9984	0.2263657	0.0003829	-0.1004719
3	0.0620	0.1589199	0.2204	0.1915076	0.8412	0.2583204	0.0006729	-0.1440510
4	0.0146	0.1805480	0.0726	0.2161875	0.5040	0.2872394	0.0009798	-0.1811448
5	0.0062	0.1985324	0.0256	0.2368009	0.2734	0.3117404	0.0012884	-0.2103702
6	0.0028	0.2129061	0.0116	0.2532423	0.1564	0.3311922	0.0015666	-0.2292993
7	0.0020	0.2223720	0.0114	0.2639180	0.1162	0.3433438	0.0018020	-0.2327066
8	0.0058	0.2233642	0.0212	0.2645726	0.1524	0.3425913	0.0019347	-0.2117744
9	0.0516	0.2016176	0.1262	0.2384421	0.4316	0.3076744	0.0016026	-0.1437761

Tabela 4.35: Caso 1, $n = 100$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.4148	0.0747383	0.8536	0.0924540	0.9998	0.1328326	0.0000928	-0.0525331
2	0.0242	0.1015811	0.1182	0.1227488	0.7352	0.1667678	0.0002411	-0.1006892
3	0.0018	0.1224919	0.0128	0.1469237	0.1952	0.1961062	0.0004110	-0.1440243
4	0.0000	0.1397451	0.0016	0.1669996	0.0412	0.2209340	0.0005961	-0.1812636
5	0.0000	0.1542032	0.0000	0.1838508	0.0100	0.2418745	0.0007915	-0.2105332
6	0.0000	0.1657453	0.0000	0.1972843	0.0040	0.2585110	0.0009798	-0.2292091
7	0.0000	0.1735382	0.0002	0.2062825	0.0040	0.2694267	0.0011402	-0.2324861
8	0.0004	0.1746997	0.0006	0.2074121	0.0074	0.2701159	0.0011866	-0.2113315
9	0.0026	0.1582636	0.0108	0.1877028	0.0866	0.2438526	0.0009563	-0.1432728

Tabela 4.36: Caso 1, $n = 200$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0090	0.0513566	0.0770	0.0622924	0.7810	0.0854358	0.0000432	-0.0524087
2	0.0000	0.0712140	0.0002	0.0854352	0.0158	0.1141211	0.0001141	-0.1006682
3	0.0000	0.0864371	0.0000	0.1033284	0.0004	0.1368338	0.0002006	-0.1440249
4	0.0000	0.0989024	0.0000	0.1180150	0.0000	0.1555951	0.0002943	-0.1814087
5	0.0000	0.1093598	0.0000	0.1303431	0.0000	0.1713688	0.0003875	-0.2106653
6	0.0000	0.1177121	0.0000	0.1401821	0.0000	0.1839340	0.0004801	-0.2295164
7	0.0000	0.1234202	0.0000	0.1468807	0.0000	0.1924072	0.0005504	-0.2328241
8	0.0000	0.1244248	0.0000	0.1479854	0.0000	0.1935653	0.0005671	-0.2114355
9	0.0000	0.1129396	0.0000	0.1342508	0.0000	0.1753661	0.0004587	-0.1434407

Tabela 4.37: Caso 1, $n = 500$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0000	0.0318286	0.0000	0.0381892	0.0008	0.0510328	0.0000177	-0.0525447
2	0.0000	0.0446801	0.0000	0.0533820	0.0000	0.0706027	0.0000451	-0.1008915
3	0.0000	0.0544696	0.0000	0.0649857	0.0000	0.0856521	0.0000796	-0.1443431
4	0.0000	0.0625314	0.0000	0.0745494	0.0000	0.0980819	0.0001179	-0.1815597
5	0.0000	0.0692300	0.0000	0.0824976	0.0000	0.1084173	0.0001587	-0.2109363
6	0.0000	0.0746226	0.0000	0.0888941	0.0000	0.1167285	0.0001941	-0.2295006
7	0.0000	0.0783059	0.0000	0.0932567	0.0000	0.1223762	0.0002269	-0.2328441
8	0.0000	0.0790171	0.0000	0.0940804	0.0000	0.1233826	0.0002346	-0.2116835
9	0.0000	0.0718034	0.0000	0.0854720	0.0000	0.1120306	0.0001939	-0.1436720

Tabela 4.38: Caso 2, $n = 60$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0006	0.1984367	0.0010	0.2364554	0.0038	0.3105768	0.0019430	0.2151658
2	0.0386	0.2107269	0.0750	0.2506410	0.2084	0.3277681	0.0025377	0.1746719
3	0.3500	0.2212849	0.4854	0.2628546	0.7296	0.3426578	0.0030524	0.1223116
4	0.7704	0.2330814	0.8648	0.2764972	0.9594	0.3592790	0.0035530	0.0684487
5	0.9398	0.2475973	0.9756	0.2932263	0.9968	0.3794843	0.0041413	0.0157333
6	0.9350	0.2662304	0.9750	0.3145179	0.9976	0.4046678	0.0047963	-0.0314262
7	0.8602	0.2894557	0.9270	0.3405995	0.9886	0.4342763	0.0056313	-0.0688876
8	0.7734	0.3043082	0.8684	0.3562521	0.9620	0.4493808	0.0068030	-0.0926133
9	0.6438	0.2674366	0.7520	0.3118501	0.8982	0.3905612	0.0073714	-0.0998642

Tabela 4.39: Caso 2, $n = 100$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0000	0.1543554	0.0000	0.1839204	0.0000	0.2416161	0.0011890	0.2156464
2	0.0038	0.1642130	0.0082	0.1954514	0.0362	0.2560825	0.0015655	0.1753627
3	0.1556	0.1727259	0.2458	0.2054219	0.4914	0.2686315	0.0018998	0.1231296
4	0.6564	0.1822234	0.7662	0.2165442	0.9198	0.2826264	0.0021951	0.0691356
5	0.9320	0.1939336	0.9692	0.2302340	0.9958	0.2997770	0.0025187	0.0165027
6	0.9210	0.2096488	0.9660	0.2485298	0.9968	0.3224643	0.0029079	-0.0306500
7	0.8056	0.2315618	0.8920	0.2738094	0.9758	0.3531219	0.0033636	-0.0684042
8	0.7082	0.2574544	0.8180	0.3031133	0.9510	0.3871329	0.0041489	-0.0909090
9	0.6660	0.2466823	0.7760	0.2892536	0.9144	0.3665478	0.0048242	-0.0877623

Tabela 4.40: Caso 2, $n = 200$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0000	0.1095063	0.0000	0.1304778	0.0000	0.1714186	0.0006089	0.2156231
2	0.0000	0.1165952	0.0000	0.1388484	0.0003	0.1821729	0.0007986	0.1751454
3	0.0156	0.1227644	0.0333	0.1461376	0.1230	0.1915524	0.0009505	0.1229300
4	0.4230	0.1296550	0.5596	0.1542791	0.8003	0.2020286	0.0011081	0.0687351
5	0.9186	0.1382220	0.9606	0.1643937	0.9956	0.2150193	0.0012660	0.0158401
6	0.8813	0.1499931	0.9406	0.1782656	0.9893	0.2327547	0.0014746	-0.0314790
7	0.6403	0.1668855	0.7736	0.1980933	0.9360	0.2578559	0.0017010	-0.0691982
8	0.5083	0.1906640	0.6626	0.2257627	0.8686	0.2921583	0.0021047	-0.0915870
9	0.6056	0.2106355	0.7353	0.2485606	0.8916	0.3192937	0.0025783	-0.0830719

Tabela 4.41: Caso 2, $n = 500$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.0000	0.0693668	0.0000	0.0826505	0.0000	0.1085858	0.0002322	0.2150069
2	0.0000	0.0739034	0.0000	0.0880367	0.0000	0.1156004	0.0003030	0.1742503
3	0.0000	0.0778578	0.0000	0.0927328	0.0000	0.1217201	0.0003584	0.1219138
4	0.0836	0.0823087	0.1610	0.0980186	0.3930	0.1286081	0.0004194	0.0677988
5	0.8973	0.0878728	0.9490	0.1046244	0.9893	0.1372105	0.0004817	0.0153062
6	0.7583	0.0954936	0.8660	0.1136658	0.9743	0.1489645	0.0005512	-0.0318713
7	0.2580	0.1065049	0.4086	0.1267106	0.7056	0.1658618	0.0006459	-0.0694262
8	0.1200	0.1228184	0.2250	0.1459737	0.5213	0.1906150	0.0007730	-0.0915640
9	0.3020	0.1459566	0.4506	0.1731110	0.7220	0.2249351	0.0009881	-0.0815935

Tabela 4.42: Caso 3, $n = 100$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9974	0.1034279	0.9996	0.1250201	1.0000	0.1699781	0.0002824	-0.0000427
2	0.9906	0.1396876	0.9974	0.1671212	0.9998	0.2216926	0.0007229	-0.0002476
3	0.9848	0.1651016	0.9954	0.1968017	1.0000	0.2587643	0.0012127	-0.0006217
4	0.9758	0.1840638	0.9926	0.2189228	0.9994	0.2863247	0.0017023	-0.0008355
5	0.9692	0.1983583	0.9910	0.2355046	0.9990	0.3066890	0.0021634	-0.0009172
6	0.9596	0.2089093	0.9872	0.2475813	0.9996	0.3066890	0.0025859	-0.0013341
7	0.9626	0.2158549	0.9874	0.2552629	0.9990	0.3293026	0.0028785	-0.0014246
8	0.9540	0.2157946	0.9830	0.2547164	0.9964	0.3272827	0.0030736	-0.0012871
9	0.8574	0.1734489	0.9072	0.2050769	0.9538	0.2647802	0.0029009	-0.0048072

Tabela 4.43: Caso 3, $n = 200$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9976	0.0722437	0.9996	0.0866909	1.0000	0.1158631	0.0001424	-0.0000506
2	0.9922	0.0986950	0.9980	0.1178357	1.0000	0.1555760	0.0003515	-0.0000534
3	0.9850	0.1171204	0.9964	0.1395789	1.0000	0.1834646	0.0005780	-0.0002332
4	0.9810	0.1309612	0.9962	0.1559016	0.9998	0.2043690	0.0008075	-0.0002600
5	0.9736	0.1414713	0.9928	0.1682625	0.9998	0.2200921	0.0010283	-0.0003156
6	0.9686	0.1493340	0.9906	0.1774518	0.9996	0.2315967	0.0012336	-0.0004462
7	0.9620	0.1549914	0.9878	0.1839659	0.9996	0.2394435	0.0014094	-0.0000759
8	0.9646	0.1578299	0.9894	0.1870883	0.9992	0.2427612	0.0014854	0.0002043
9	0.9292	0.1460807	0.9636	0.1736702	0.9884	0.2269402	0.0014626	-0.0004051

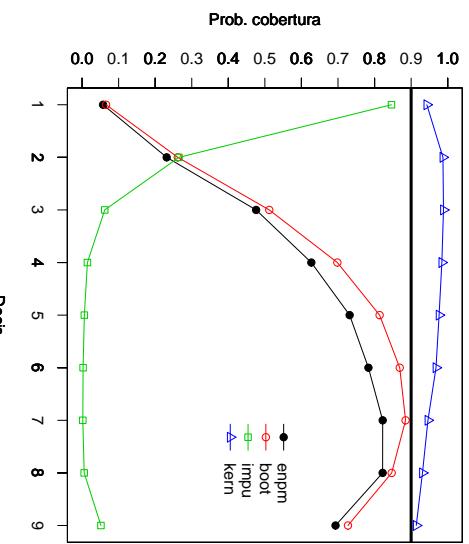
Tabela 4.44: Caso 3, $n = 500$, Imputação

k	pc90	amp90	pc95	amp95	pc99	amp99	variância	vício
1	0.9970	0.0455083	0.9992	0.0543744	1.0000	0.0719246	0.0000558	-0.0000044
2	0.9942	0.0624379	0.9986	0.0744548	1.0000	0.0980111	0.0001410	0.0000054
3	0.9868	0.0742176	0.9964	0.0884373	1.0000	0.1162109	0.0002308	-0.0000723
4	0.9792	0.0830920	0.9944	0.0989684	0.9998	0.1299093	0.0003251	-0.0001777
5	0.9726	0.0898971	0.9914	0.1070354	1.0000	0.1403749	0.0004192	-0.0000820
6	0.9694	0.0950356	0.9884	0.1131120	0.9996	0.1482108	0.0004965	-0.0000605
7	0.9626	0.0987594	0.9888	0.1174908	0.9988	0.1537785	0.0005640	-0.0002009
8	0.9650	0.1011816	0.9884	0.1202994	0.9994	0.1572243	0.0006068	-0.0001643
9	0.9610	0.1016686	0.9852	0.1210129	0.9988	0.1585805	0.0006088	-0.0002599

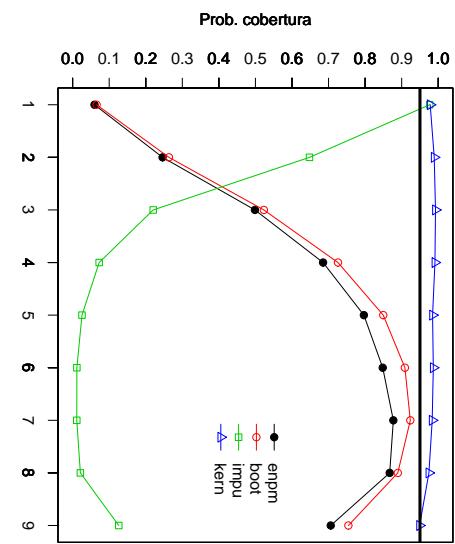
Tabela 4.45: Comportamento do h_{opt} para f e g

caso	n	$\text{var}(h_f)$	$\text{var}(h_g)$
1	60	0.5610325	0.2025051
	100	0.3128659	0.1001783
	200	0.1565486	0.0365909
	500	0.0512011	0.0115125
2	60	0.3977849	0.0366412
	100	0.1811114	0.0211418
	200	0.0673056	0.0123736
	500	0.0746262	0.0047299
3	100	0.0282473	0.0069342
	200	0.0122252	0.0027337
	500	0.0030071	0.0006412

Caso 1, prob.de cob.=0.9 n=60



Caso 1, prob.de cob.=0.95, n=60



Caso 1, prob.de cob.=0.99, n=60

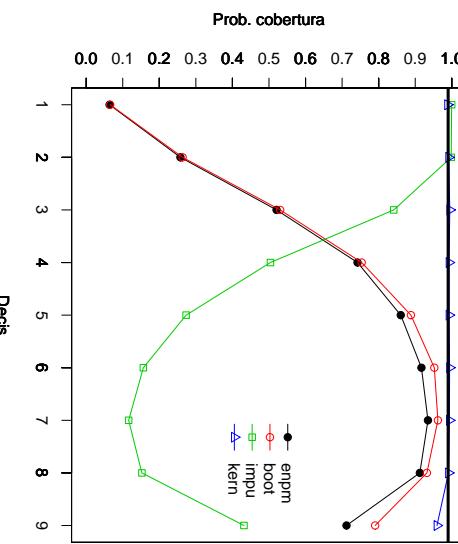
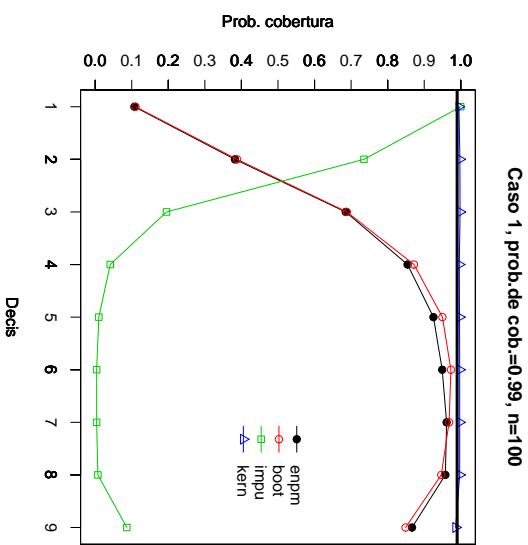


Figura 4.1: Comparação das probabilidades de cobertura para o caso 1 e $n = 60$

Caso 1, prob.de cob.=0.99, n=100



Caso 1, prob.de cob.=0.95, n=100

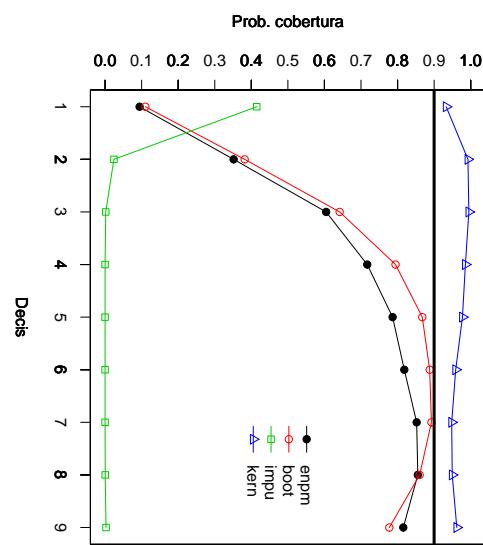
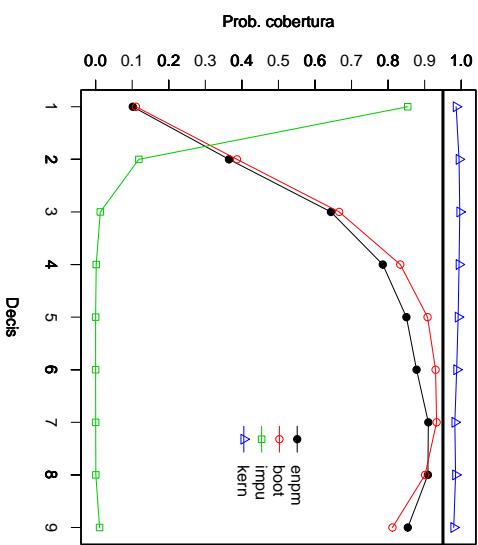
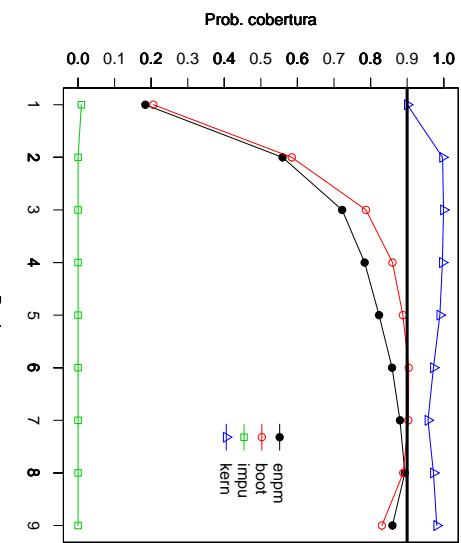
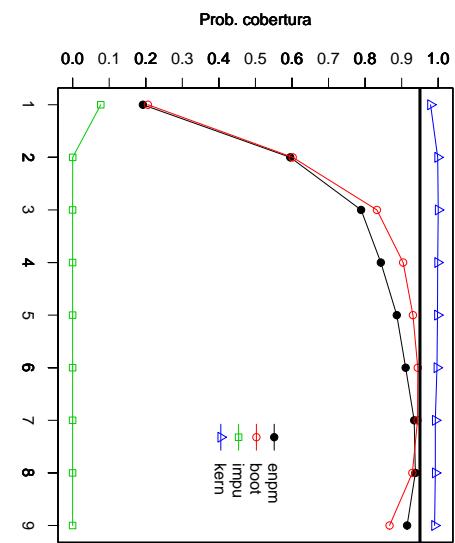


Figura 4.2: Comparação das probabilidades de cobertura para o caso 1 e $n = 100$

Caso 1, prob.de cob.=0.90, n=200



Caso 1, prob.de cob.=0.95, n=200



Caso 1, prob.de cob.=0.99, n=200

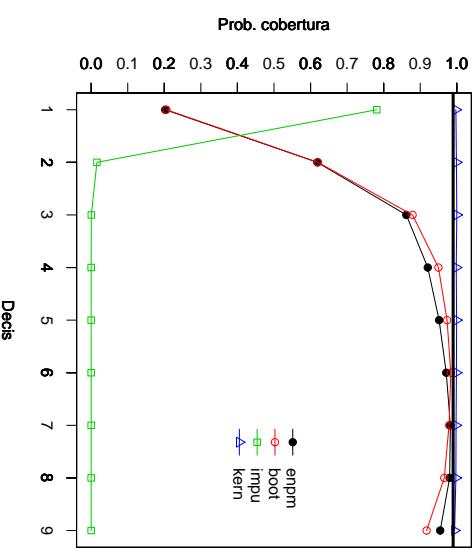
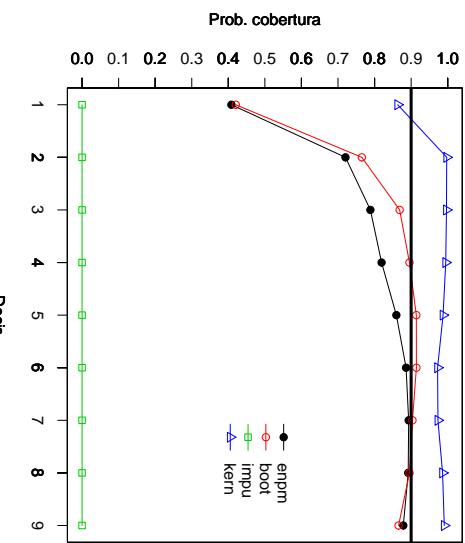
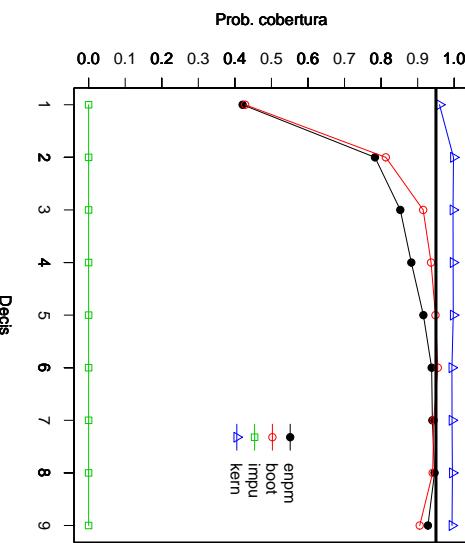


Figura 4.3: Comparação das probabilidades de cobertura para o caso 1 e $n = 200$

Caso 1, prob.de cob.=0.90, n=500



Caso 1, prob.de cob.=0.95, n=500



Caso 1, prob.de cob.=0.99, n=500

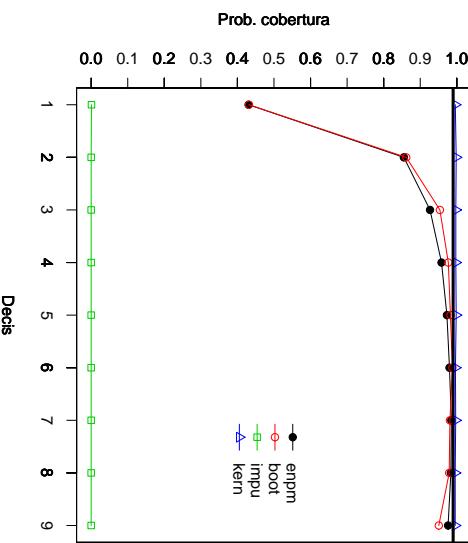
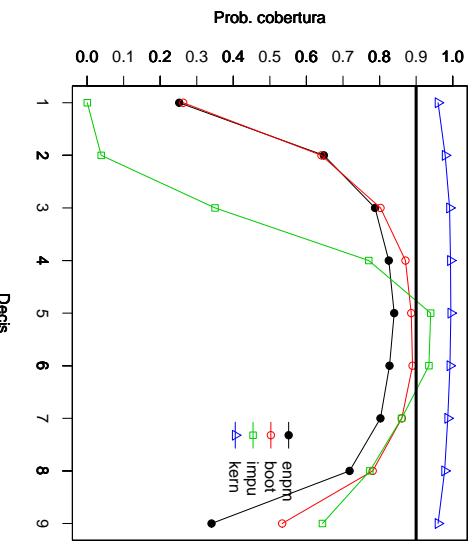
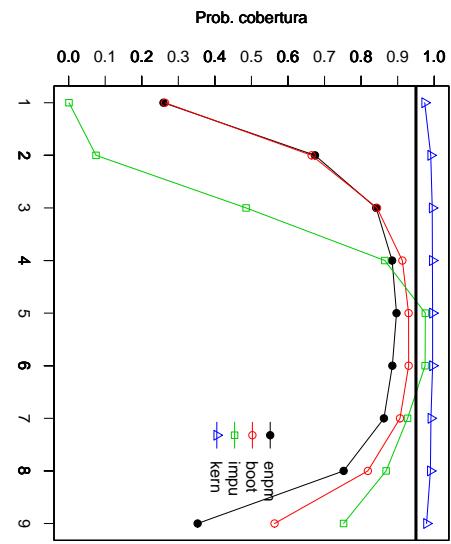


Figura 4.4: Comparação das probabilidades de cobertura para o caso 1 e $n = 500$

Caso 2, prob.de cob.=0.90, n=60



Caso 2, prob.de cob.=0.95, n=60



Caso 2, prob.de cob.=0.99, n=60

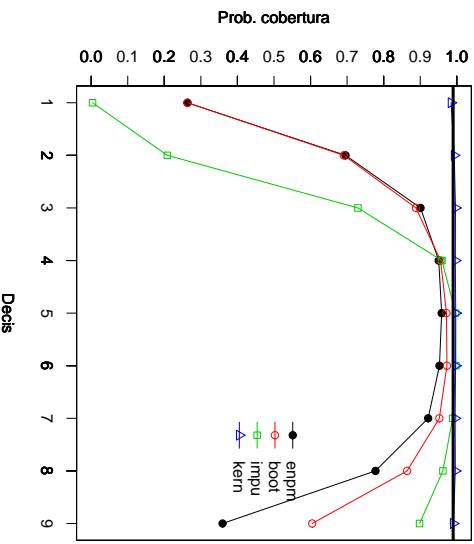
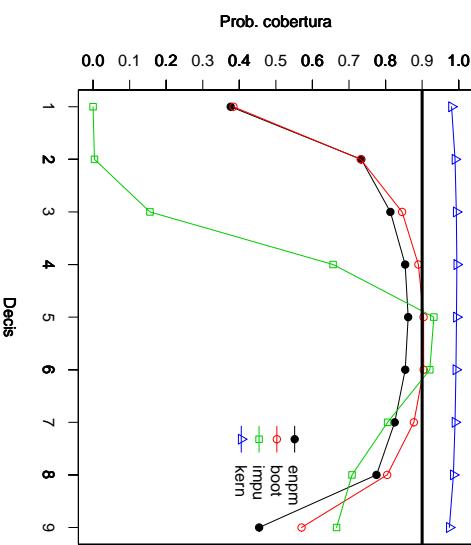
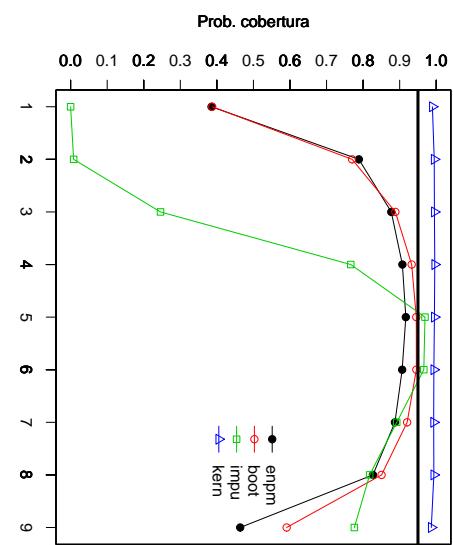


Figura 4.5: Comparação das probabilidades de cobertura para o caso 2 e $n = 60$

Caso 2, prob.de cob.=0.90, n=100



Caso 2, prob.de cob.=0.95, n=100



Caso 2, prob.de cob.=0.99, n=100

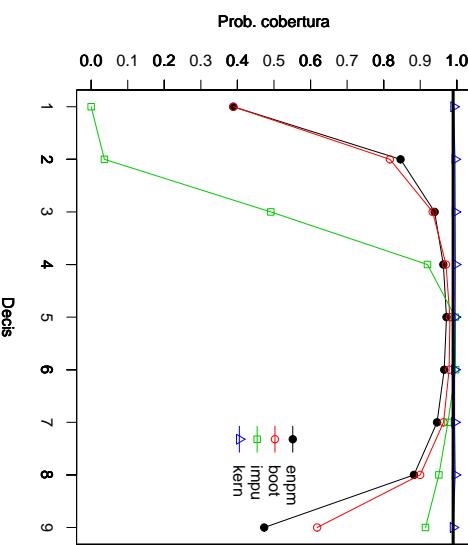
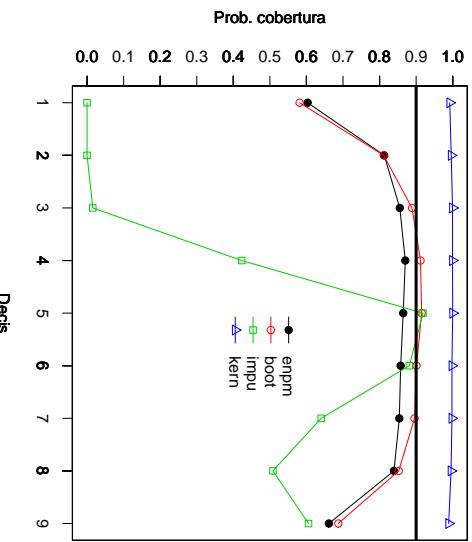
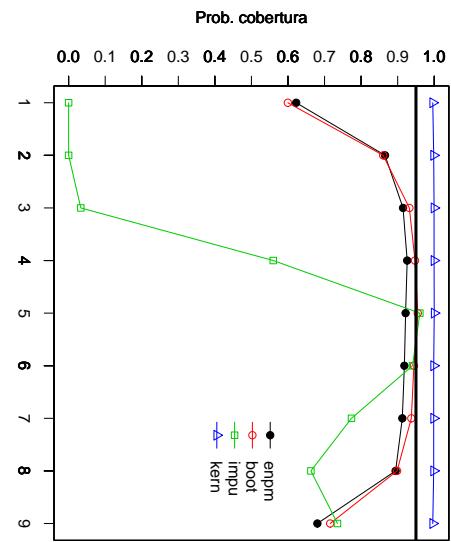


Figura 4.6: Comparação das probabilidades de cobertura para o caso 2 e $n = 100$

Caso 2, prob.de cob.=0.90, n=200



Caso 2, prob.de cob.=0.95, n=200



Caso 2, prob.de cob.=0.99, n=200

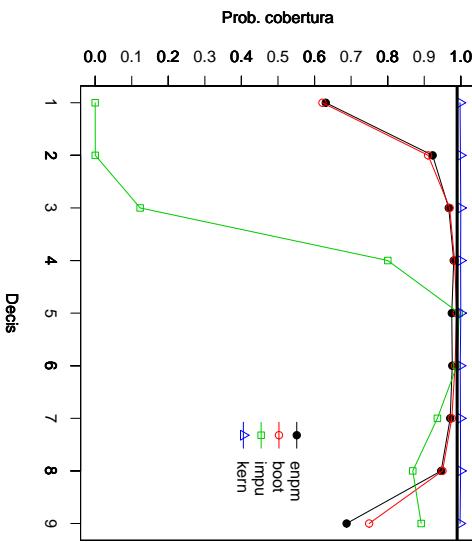
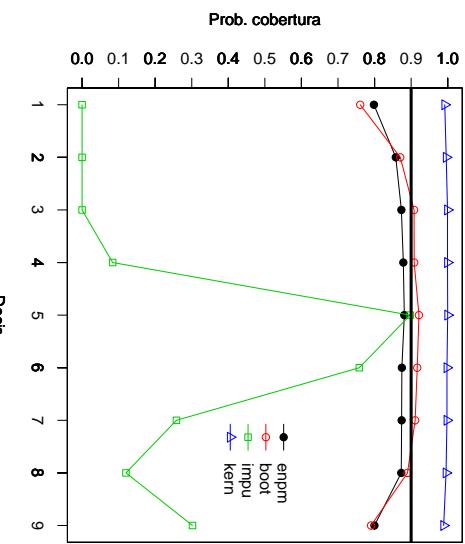
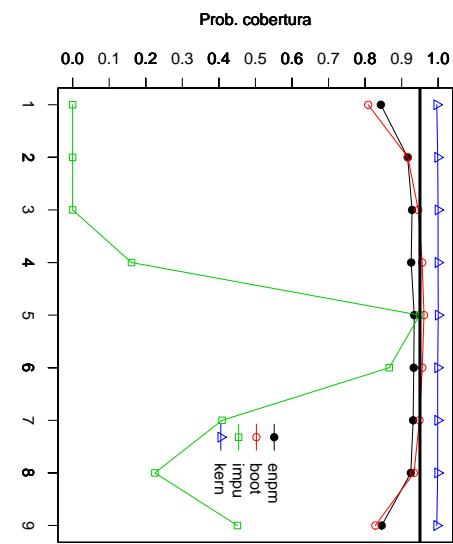


Figura 4.7: Comparação das probabilidades de cobertura para o caso 2 e $n = 200$

Caso 2, prob.de cob.=0.90, n=500



Caso 2, prob.de cob.=0.95, n=500



Caso 2, prob.de cob.=0.99, n=500

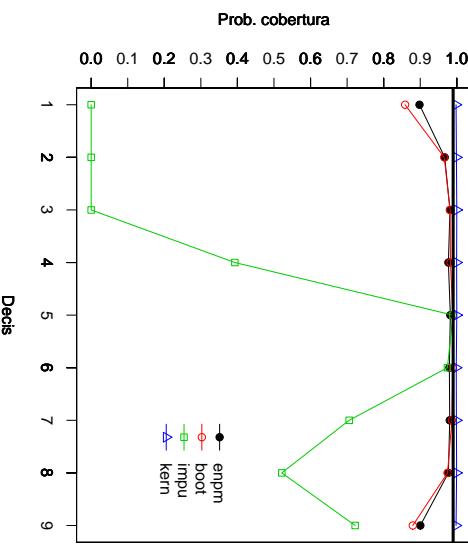
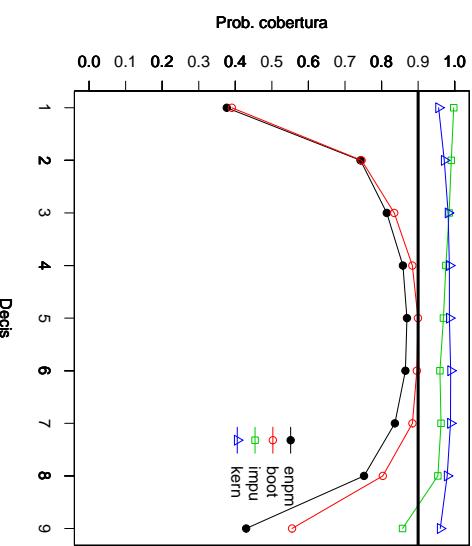
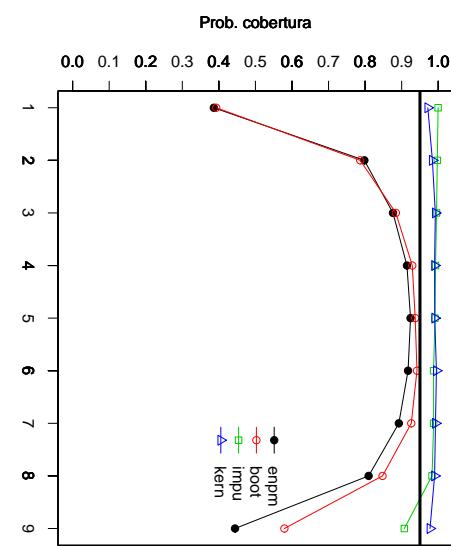


Figura 4.8: Comparação das probabilidades de cobertura para o caso 2 e $n = 500$

Caso 3, prob.de cob.=0.90, n=100



Caso 3, prob.de cob.=0.95, n=100



Caso 3, prob.de cob.=0.99, n=100

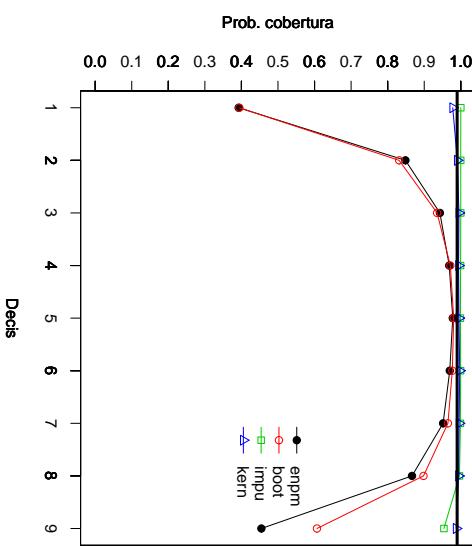
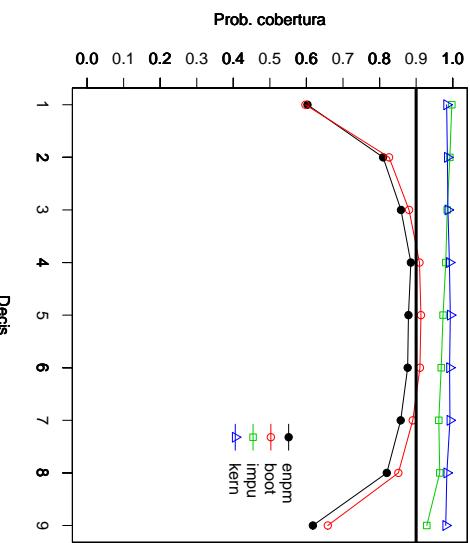
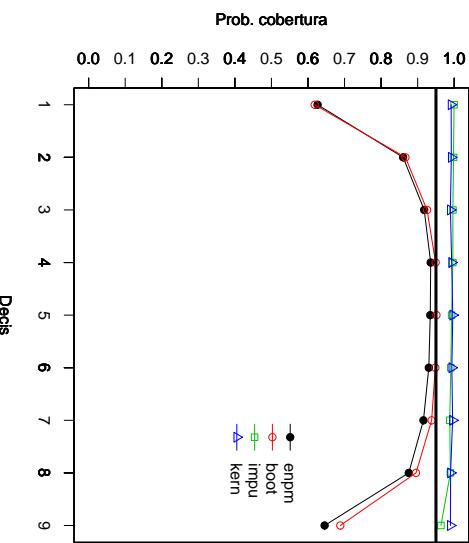


Figura 4.9: Comparação das probabilidades de cobertura para o caso 3 e $n = 100$

Caso 3, prob.de cob.=0.90, n=200



Caso 3, prob.de cob.=0.95, n=200



Caso 3, prob.de cob.=0.99, n=200

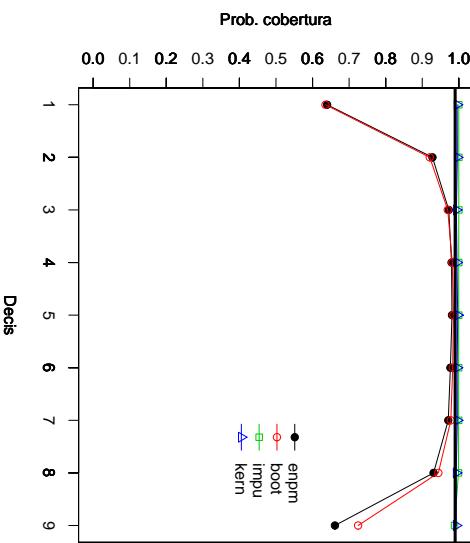
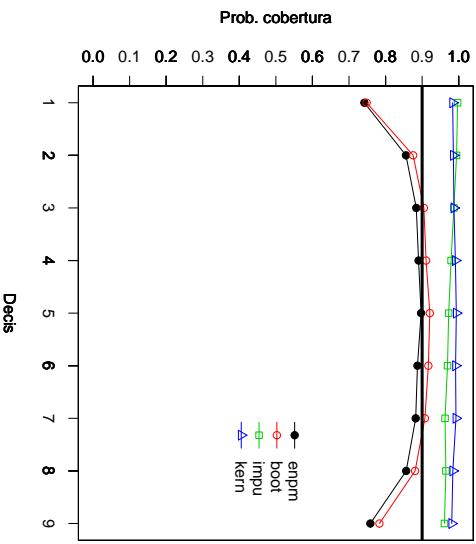
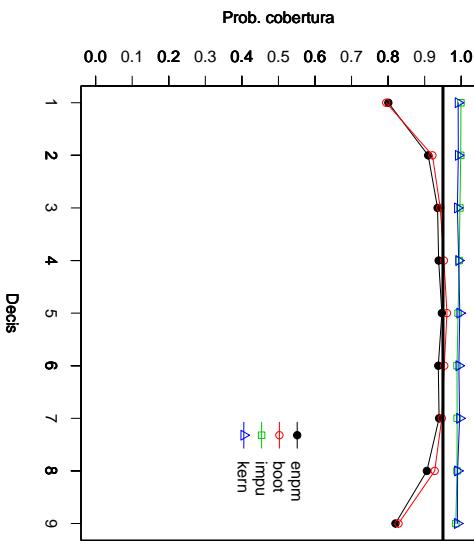


Figura 4.10: Comparação das probabilidades de cobertura para o caso 3 e $n = 200$

Caso 3, prob.de cob.=90, n=500



Caso 3, prob.de cob.=95, n=500



Caso 3, prob.de cob.=0.99, n=500

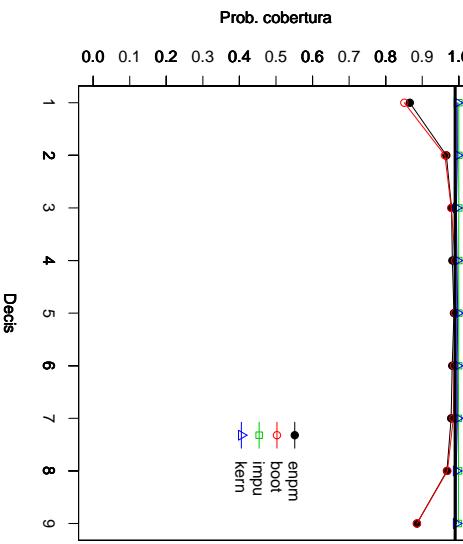


Figura 4.11: Comparação das probabilidades de cobertura para o caso 3 e $n = 500$

4.2 Aplicação

Os dados da tabela abaixo referem-se a um estudo feito por Hoel e Walburg(1972) sobre ratos machos da raça RFM e seu objetivo era estudar o tempo até a ocorrência de um tumor no pulmão a partir da aplicação de uma substância cancerígena. A primeira coluna diz respeito à presença ($\delta = 1$) ou não ($\delta = 0$) do tumor, enquanto que a segunda coluna é a idade de cada rato no instante da morte (por sacrifício ou outra causa).

Tabela 4.46: Idade da morte de ratos RFM machos

Considerando que f é a densidade do tempo até a ocorrência do tumor e g a densidade do instante de observação (morte) dos ratos, temos as seguintes estimativas: ($h_f = 463,970390625$ e $h_g = 134,079048910847$)

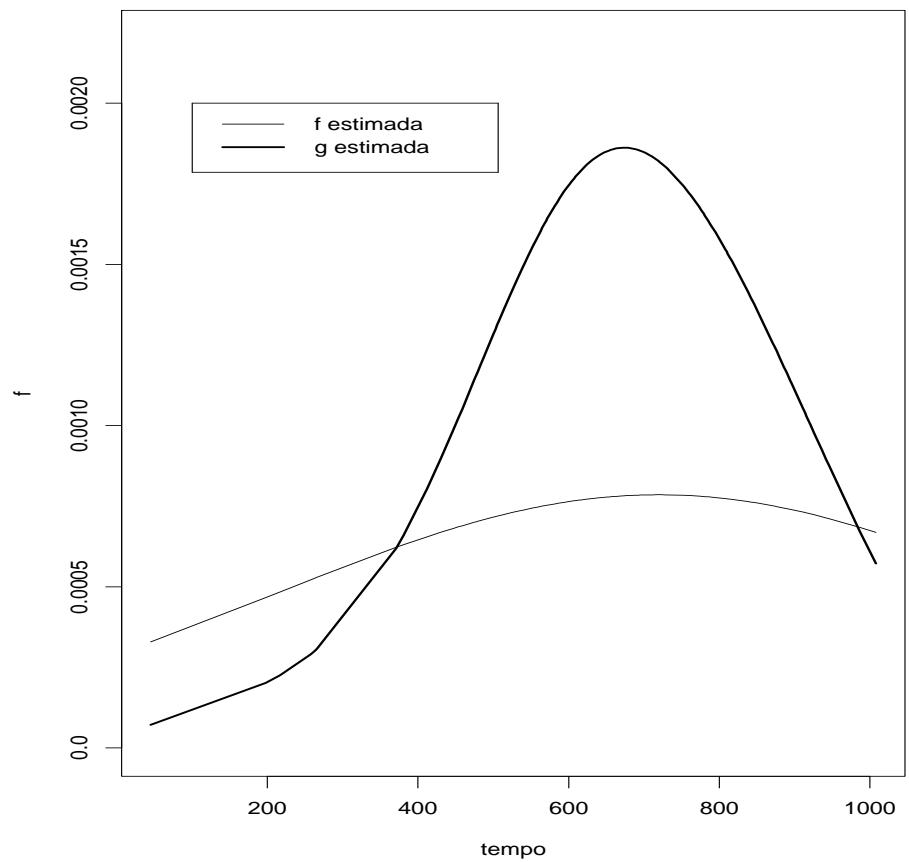


Figura 4.12: Estimativas por nucleo estimador das funções densidade f e g .

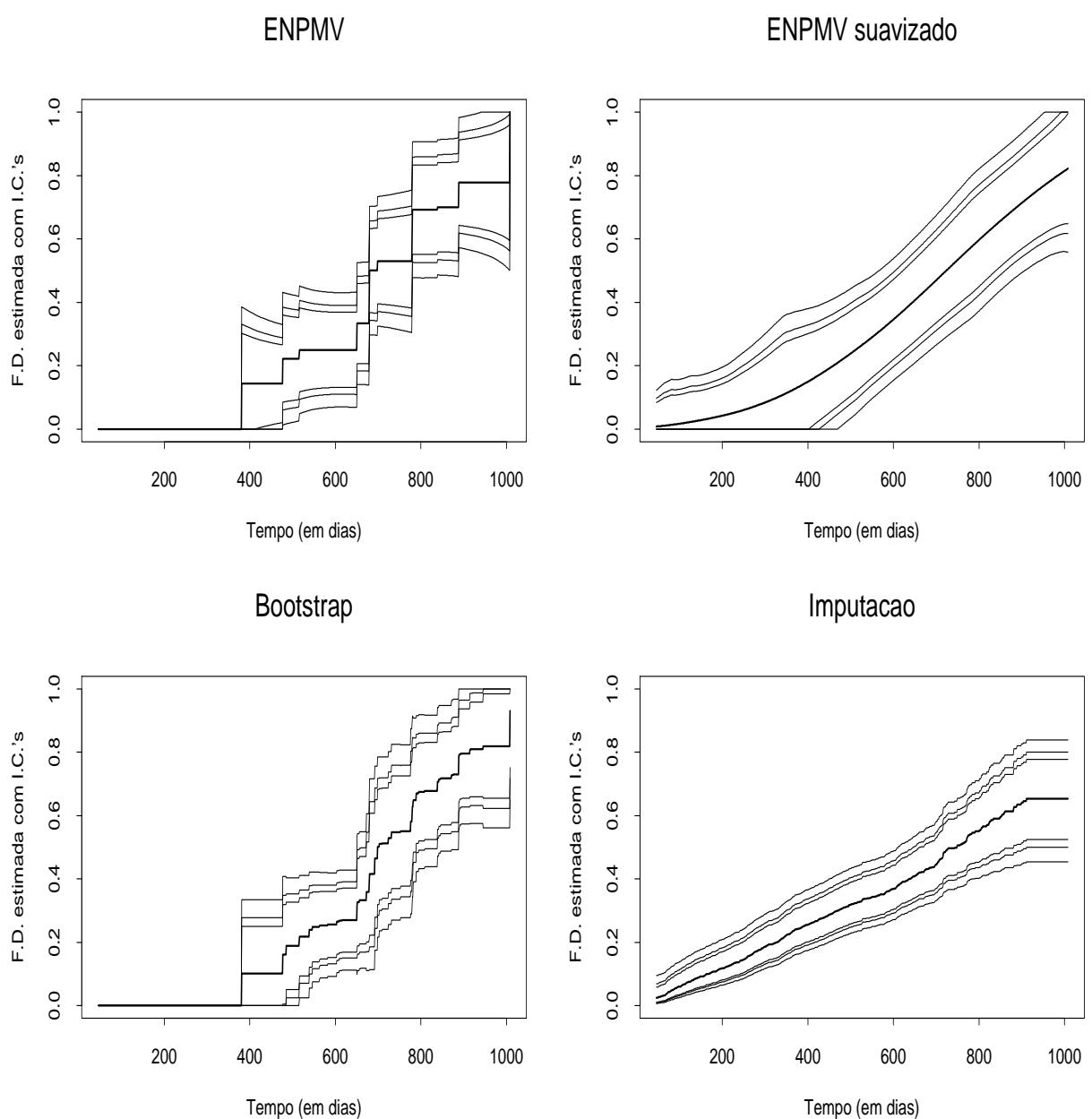


Figura 4.13: F estimada e intervalo de 90%, 95% e 99% de confiança.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

As figuras do capítulo anterior mostram que, de maneira geral, a construção dos intervalos de confiança pelo ENPMV suavizado e a distribuição assintótica do ENPMV apresenta probabilidade de cobertura superior aos demais métodos, e bastante superior ao valor nominal, o que indica intervalos de confiança com amplitude maior do que o necessário. No entanto, a Figura 4.13 mostra que, para os dados reais utilizados como exemplo, a amplitude dos intervalos obtidos a partir do ENPMV suavizado é comparável aos demais métodos, exceto o método de imputação uniforme que fornece intervalos com amplitude menor do que os demais. Porém, as simulações mostram que este método apresenta vício muito grande quando a distribuição do tempo de falha não é uniforme. O ENPMV suavizado apresenta variância menor que os demais métodos (exceto o de imputação).

O método bootstrap apresenta, em geral, a segunda maior probabilidade de cobertura, seguido pelo ENPMV e sua distribuição assintótica.

O método da imputação uniforme com cálculo do intervalo de confiança via estimador de Kaplan-Meier e fórmula de Greenwood apresenta probabilidade de cobertura tendendo a zero quando o tamanho da amostra cresce, exceto para o caso em que a distribuição do tempo de falha é uniforme. Isto fica claro, também, pelo valor do vício em cada decil. Este fenômeno parece ser decorrência do fato de ter-se utilizado a imputação uniforme, e parece indicar que, caso fosse possível adotar o processo

de imputação utilizando uma distribuição próxima da distribuição real do tempo de falha, este método poderia ter um bom desempenho na construção de intervalos de confiança. Uma tal distribuição poderia ser aquela fornecida pela versão suavizada do ENPMV, provavelmente, porque a imputação é feita utilizando uma distribuição (uniforme) que tem uma forma distante da verdadeira distribuição do tempo de falha.

De maneira geral, todos os métodos parecem ter desempenho ruim (probabilidade de cobertura muito abaixo do valor nominal) em pontos próximos dos limites do suporte da distribuição do tempo de falha. A exceção parece ser o método que utiliza o ENPMV suavizado e a distribuição assintótica do ENPMV, além do método de imputação uniforme para o caso em que o tempo de falha também tem distribuição uniforme e para o primeiro decil no caso 1 (tempo de falha com distribuição exponencial). Neste último caso, o processo de imputação para observações com censura à esquerda faz com que a estimativa obtida para a função de distribuição detecte mais rapidamente do que o ENPMV o crescimento forte da função de distribuição próximo de zero no caso da distribuição exponencial.

Com base nas observações acima, a construção de intervalos de confiança para a função de distribuição do tempo de falha para o caso 1 de censura intervalar pelo ENPMV suavizado parece ser o caminho que fornece as maiores probabilidades de cobertura com amplitude comparável aos demais métodos e com vício de magnitude igual ou inferior.

Com exceção do caso 1, com pequenas amostras ($n = 60$), o método bootstrap fornece intervalos de confiança com probabilidade de cobertura para $F(md)$ bem próximo do valor nominal, onde md é o tempo mediano de falha. Isto coloca tal método como uma boa opção para a construção do intervalo de confiança para o tempo mediano de falha, medida de posição mais utilizada em estudo de sobrevivência em função da presença de censura.

5.2 Trabalhos Futuros

Os estudos de simulação mostraram ser provável que o método de imputação possa ter um bom desempenho quando se utiliza uma distribuição próxima da real para o processo de imputação. Uma boa estimativa de tal distribuição é fornecida pelo ENPMV suavizado. Logo, uma alternativa seria utilizá-lo em conjunto com o método de imputação. É possível que se obtenha intervalos de confiança com probabilidade de cobertura alta, porém com amplitude menor do que os demais métodos aqui estudados, já que o método de imputação forneceu intervalos com menor amplitude que os demais.

Outra método possível para construção de intervalos de confiança seria utilizar a distribuição assintótica da estatística da razão de verossimilhança para o teste das hipóteses $H_0 : F(t_0) = \theta_0$ versus $H_1 : F(t_0) \neq \theta_0$, estudada e tabelada em Banerjee e Wellner (2001).

Um estudo semelhante pode ser desenvolvido para o caso 2 de censura intervalar. No entanto, isto envolve dificuldades computacionais num grau maior do que as já enfrentadas na implementação dos métodos para o caso 1.

Referências Bibliográficas

- [1] Akaike, H., 1954. “An Approximation to the Density Function.” *Ann. Inst. Statistics. Math.* **6**, 127-32.
- [2] Ayer, M., Brunk, H. D., Ewing, G. M, Reid, W. T., Silverman, E., 1955. “An Empirical Distribution Function for Sampling with Incomplete Information.” *Ann. Math. Statist.* **26**, 641-647
- [3] Barlow, R. E., Bartholomew, D. J., Bremmer, J. M. e Brunk, H. D. 1972. “Statistical Inference Under Order Restrictions”. *John Wiley & Sons, New York.*
- [4] Banerjee, M. e Wellner J. A. 2001. “ Likelihood Ratio Tests for Monotone Functions.” *Annals of Statistics*, **29**, 1699-1731.
- [5] Bessegato, L. F., 2001. “Escolha do Parâmetro de Suavidade na Estimativa da Função de Distribuição”. *Dissertação de Mestrado, Departamento de Estatística da UFMG.*
- [6] Breslow, N. E. e Crowley, J. 1974. “ A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship”. *The Annals of Statistics*, **2**, 437-53.
- [7] Chiu, S. T. 1991. “ Bandwidth Selection for Kernel Density Estimation”. *The Annals of Statistics*, **19**, 1883-1905.
- [8] Chung, K. L. 1974. “ A Course in Probability Theory”. *Ed. Academic Press, New York.*
- [9] Colosimo, E. A. 2001.“ Análise de Sobrevivência Aplicada”. *46ª Reunião Anual da RBRAS e 9º SEAGRO. Departamento de Estatística da UFMG.*

- [10] Damasceno, E. C. 2000. “ Escolha do Parâmetro de Suavidade em Estimação Funcional”. *Dissertação de Mestrado, Departamento de Estatística da UFMG.*
- [11] Duchesne, T., Stafford, J. E. 2001 “ A Kernel Density Estimate for Interval Censored Data”. *Technical Report nº 0106, University of Toronto.*
- [12] Efron, B. 1979. “ Bootstrap methods: Another Look at the Jackknife”. *Ann. Statist.* **7**, 1-26.
- [13] Efron, B., Tibshirani, R. 1993. “An Introduction to the Bootstrap”. *Chapman & Hall, New York.*
- [14] Fix, E. and Hodges, J. L. 1951. “ Discriminatory Analysis, Nonparametric Estimation: Consistency Properties”. *Report Nº 4, Project nº 21-49-004. USAF School of Aviation Medicine, Randolph Field, Texas.*
- [15] Groeneboom, P. 1996. “Lectures on Inverse Problems”. In *Lecture Notes in Mathematics, 1948, Springer-Verlag*, 67-164.
- [16] Groeneboom, P. e Wellner, J. A. 1992. “ Information Bounds and Nonparametric Maximum Likelihood Estimation”. *Birkhauser Verlag, Berlin.*
- [17] Groeneboom, P. e Wellner, J. A. 2001. “Computing Chernoff’s Distribution”. *Journal of Computational and Graphical Statistics*, **10**, 388-400.
- [18] Hall, P. 1986. “On the Number of Bootstrap Simulations Required to Construct a Confidence Interval”. *Ann. Statist.*, **14**, 1453-1462.
- [19] Hall, P., Sheather, S. J., Jones, M. C. e Marron, J. S. 1991. “ On Optimal Data-based Bandwidth Selection in Kernel Density Estimation”. *Biometrika* **78**, 263-269.
- [20] Hoel, D. G. e Walburg, H. E. Jr. 1972. “ Statistical Analysis of Survival Experiments”. *Journal of the National Cancer Institute*, **49**, 361-362.
- [21] Huang, J. e Wellner, J. A. 1995. “ Asymptotic Normality of the NPMLE of Linear Functionals of Interval Censored Data, Caso 1”. *Statistica Neerlandica*, **49**, 153-163.

- [22] Jongbloed, G. 1995. “ The Iterative Convex Minorant Algorithm for Nonparametric Estimation”. *Submitted to J. of Computational and Graphical Stat.*
- [23] Lee, E. T. 1992. “ Statistical Methods for Survival Data Analysis”. *John Wiley & Sons, New York.*
- [24] Lemos, S. M. N. B. 2003. “ Tratamento de Dados Ausentes em Estudos Longitudinais”. *Dissertação de Mestrado, Departamento de Estatística da UFMG.*
- [25] Li, L., Watkins, T., Yu, Q. 1997. “An EM Algorithm for Smoothing the Self-Consistent Estimator of Survival Functions with Interval Censored Data”. *Scandinavian Journal of Statistics*, **24**, 531-542.
- [26] Meier, P. 1975. “ Estimation of a Distribution Function Incomplete Observations”. *Perspectives in Probability and Statistics, J. Gari. Ed. Sheffield, England.*
- [27] Padgett, W. J. 1986. “A Kernel-type Estimator of a Quantile Function from Right-Censored Data”. *Journal of the American Statistical Association*, **39**, 215-222.
- [28] Pan, W. 2000a. “ Smooth Estimation of the Survival Functions for Interval Censored Data”. *Statistics in Medicine*, **19**, 2611-2624.
- [29] Pan, W. 2000b. “ A Two-Sample Test With for Interval Censored Data via Multiple Imputation”. *Statistics in Medicine*, **19**, 1-11.
- [30] Pan, W. 2000c. “ A Multiple Imputation Approach to Cox Regression with Interval- Censored Data”. *Biometrics*, **56**, 199-203.
- [31] Parzen, E. 1962. “ On the Estimation of a Probability Density Function and the Mode”. *Ann. Math. Statist.* **33**, 1065-76.
- [32] Rosenblatt, M. 1956. “ Remarks on Some Nonparametric Estimates of a Density Function”. *Ann. Math. Statist.* **27**, 832-7.
- [33] Rubin, D. B. 1987. “Multiple Imputation for Nonresponse in Surveys”. *Wiley, New York.*

- [34] Rudin, W. 1987. “ Real and Complex Analysis”. *McGraw-Hill, New York*.
- [35] Silverman, B. W. 1986. “ Density Estimation for Statistics and Data Analysis”. *Chapman & Hall, London*.
- [36] Simonoff, J. S. 1996. “ Smoothing Methods in Statistics”. *Springer Series in Statistics*.
- [37] Sun, J. 1998. “Interval Censoring”. *Encyclopedia of Biostatistics. First Edition, John Wiley & Sons Ltda*, 2090-2095.
- [38] Travassos, A. P. A. 2003. “ Problemas de Fronteira dos Núcleo-Estimadores e suas abordagens”. *Dissertação de Mestrado, Departamento de Estatística da UFMG*.
- [39] Turnbull, B. W. 1976. “The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data”. *Journal of the Royal of the InterStatistical Society, Ser. B*, **38**, 290-295.
- [40] Van Eeden, C. 1956. “Maximum Likelihood Estimation of Ordered Probabilities”. *Indagationes Mathematicae*, **18**, 444-455.
- [41] Wand, M. P. e Jones, M. C. 1995 “ Kernel Smoothing”. *Chapman & Hall, London*.
- [42] Wellner, J. A. 1995. “Interval Censoring Case 2: An Exploration of Alternative Hypotheses”. *Proceedings of the International Conference on Censored Survival Data, IMS Lecture Notes Monograph Series*, **27**, 271-291.
- [43] Zhan, Y., Wellner, J. A. 1995. “ Double Censoring: Characterization and Computation of the Nonparametric Maximum Likelihood Estimator”. *Technical Report 292, University of Washington, Dept. of Statistics*.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)

[Baixar livros de Literatura de Cordel](#)

[Baixar livros de Literatura Infantil](#)

[Baixar livros de Matemática](#)

[Baixar livros de Medicina](#)

[Baixar livros de Medicina Veterinária](#)

[Baixar livros de Meio Ambiente](#)

[Baixar livros de Meteorologia](#)

[Baixar Monografias e TCC](#)

[Baixar livros Multidisciplinar](#)

[Baixar livros de Música](#)

[Baixar livros de Psicologia](#)

[Baixar livros de Química](#)

[Baixar livros de Saúde Coletiva](#)

[Baixar livros de Serviço Social](#)

[Baixar livros de Sociologia](#)

[Baixar livros de Teologia](#)

[Baixar livros de Trabalho](#)

[Baixar livros de Turismo](#)