

Universidade Federal de São Carlos

Departamento de Estatística

**Estimadores não paramétricos
para dados com censura**

Paulo Ricardo Simioni

São Carlos 25 de junho de 2013

Universidade Federal de São Carlos

Departamento de Estatística

**Estimadores não paramétricos
para dados com censura**

Dissertação apresentada ao Programa de
Pós-Graduação em Estatística
da Universidade Federal de São Carlos,
como parte dos requisitos para a obtenção
do título de Mestre em Estatística.

Paulo Ricardo Simioni

Orientador: Adriano Polpo de Campos

São Carlos 25 de junho de 2013

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S589ep Simioni, Paulo Ricardo.
Estimadores não paramétricos para dados com censura /
Paulo Ricardo Simioni. -- São Carlos : UFSCar, 2013.
66 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2013.

1. Estatística. 2. Estimadores de Bayes. 3. Sistema em
paralelo. 4. Sistema em série. I. Título.

CDD: 519.5 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40

FONE: (016) 3351-8292 – Email: ppgest@ufscar.br

13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : Paulo Ricardo Simioni

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 19/04/2013

PELA COMISSÃO JULGADORA:

Presidente

Prof. Dr. Adriano Polpo de Campos (DEs-UFSCar/Orientador)

1º Examinador

Prof. Dr. Luis Ernesto Bueno Salasar (DEs-UFSCar)

2º Examinador

Prof. Dr. Marcelo de Souza Laretto (EACH-USP)

Resumo

Neste trabalho estudamos a confiabilidade de sistemas com componentes ligados em série e sistemas ligados em paralelo. Para sistemas em série, o dispositivo falha quando o primeiro componente falhar, já no sistema em paralelo isto acontece quando o último componente falhar. Definimos as funções de distribuição e sub-distribuição, bem como suas propriedades. Apresentamos o estimador Bayesiano não-paramétrico para os componentes de ambos os sistemas, ilustrando através de exemplos. Além disso, para ambos os casos, realizamos um estudo comparativo entre o estimador Bayesiano não paramétrico e o estimador de Kaplan-Meier.

Abstract

In this paper we study the reliability of systems with connected components in series and parallel. For systems in series, the device fails when the first component fails. Although in the parallel systems this happens when the last component fails. We define the distribution functions, sub-distribution functions and their respective properties. Here we present the Bayesian nonparametric estimator for the components of both systems, illustrating by examples. For both cases we performed a comparative study between the Bayesian nonparametric estimator and the Kaplan-Meier method.

Sumário

1	Introdução	p. 7
2	Conceitos básicos	p. 9
2.1	Introdução	p. 9
2.2	A Taxa de Falha e a Taxa de Falha Reversa	p. 10
2.3	Funções de Distribuição e Sub-distribuição	p. 11
3	Estimador Bayesiano não paramétrico	p. 17
3.1	Introdução	p. 17
3.2	Sistema em Série	p. 18
3.3	Sistema em Paralelo	p. 20
4	Generalização	p. 22
4.1	Introdução	p. 22
4.2	Sistema em Série	p. 23
4.3	Sistema em Paralelo	p. 36
5	Estudo de Simulação	p. 50
5.1	Sistema com Componentes em Série	p. 51
5.2	Sistema com Componentes em Paralelo	p. 51
6	Comentários Finais	p. 58
	Apêndice A – Código Computacional	p. 59

1 *Introdução*

Em análise de sobrevivência, o problema de dados censurados à direita foi estudado por vários autores. Para uma revisão, veja Ibrahim *et al.* [4], sob uma perspectiva Bayesiana e Lawless [7], para o ponto de vista frequentista. Para o problema de dados censurados à esquerda, até onde sabemos, existem poucos trabalhos. Podemos citar Polpo e Pereira [10] e Polpo *et al.* [9]. Além destes, também citamos Barlow e Prochan [1], como uma boa referência para o problema da confiabilidade.

Neste trabalho seguimos principalmente: Peterson [8], que descreve uma relação funcional para o problema de dados censurados a direita; Salinas *et al.* [12] que apresenta o estimador Bayesiano não-paramétrico para o problema de sistemas em série e Polpo e Sinha [11] que posteriormente apresentaram uma correção do trabalho de Salinas *et al.* [12]; e Polpo e Pereira [10] que apresentaram um estimador Bayesiano não-paramétrico para o problema da confiabilidade de sistemas em paralelo.

Em um sistema em série, todos os componentes devem estar funcionando para que o sistema funcione. Denotando por X_j o tempo de falha de do j -ésimo componente, $j = 1, \dots, k$, o tempo de falha do sistema será dado por $T = \min(X_1, \dots, X_k)$. Equivalentemente, um sistema de componentes em série falhará no momento em que o primeiro componente falhar.

No caso do sistema de componentes em paralelo, basta que um dos componentes esteja funcionando para o sistema funcionar. Então, o tempo de falha do sistema será dado por $T = \max(X_1, \dots, X_k)$. Um sistema em paralelo falha quando o seu último componente falhar.

Para a estimação da confiabilidade de um sistema o problema é mais simples uma vez que não temos censura no seu tempo de falha. Neste trabalho, consideramos que censuras provocadas por efeitos externos podem ser representadas pela adição de uma ou mais componentes ao sistema. Nosso objetivo é estudar o comportamento do tempo de falha dos componentes envolvidos no sistema, em que cada observação de um componente

pode ser censurada à direita (no caso de um sistema em série), à esquerda (no caso de um sistema em paralelo) ou não censurada.

Este trabalho está organizado da seguinte forma: no Capítulo 2, discutimos a teoria base para o desenvolvimento dos estimadores não paramétricos; no Capítulo 3 desenvolvemos os estimadores Bayesianos não paramétricos para os dois tipos de sistemas; no Capítulo 4 apresentamos a generalização do Capítulo anterior, além de exemplo para os sistemas em série e em paralelo; no Capítulo 5 apresentamos um estudo de simulação comparando o estimador Bayesiano não paramétrico com o estimador de Kaplan-Meier [5]; por fim, no Capítulo 6, discutimos os principais aspectos do trabalho. No apêndice A temos o código computacional, na linguagem R, com as funções utilizadas no trabalho.

Além disso, neste trabalho adotamos uma única numeração sequencial para capítulos, seções, equações, definições, propriedades, teoremas, tabelas e figuras, facilitando para o leitor encontrar as referências citadas ao longo do texto.

2 *Conceitos básicos*

2.1 Introdução

Neste capítulo, apresentamos a base teórica para o desenvolvimento dos estimadores não-paramétricos para o problema de confiabilidade. Seguimos aqui os passos apresentados por Peterson [8] e Polpo e Pereira [10]. Para simplificar a notação adotamos o uso das (sub-)distribuições, reescrevendo apropriadamente, quando necessário, os resultados apresentados por Peterson [8], que descreve seu trabalho em termos de (sub-)sobrevivências.

Antes, porém, abaixo damos duas definições que são importantes para melhor compreensão do texto.

2.1.1 Definição. *Ponto de salto de uma função g é o ponto s em que $g(s) \neq g(s^+)$ ou $g(s) \neq g(s^-)$.*

2.1.2 Definição. *Seja I um intervalo da reta real \mathbb{R} . Uma função $f : I \rightarrow \mathbb{R}$ é absolutamente contínua em I se, para todo número positivo ε , existe um número positivo δ tal que qualquer sequência finita disjunta de sub-intervalos (a_k, b_k) de I satisfaçam*

$$\sum_k |b_k - a_k| < \delta$$

então

$$\sum_k |f(b_k) - f(a_k)| < \varepsilon.$$

A função $f(x) = 0$ se $x = 0$ e $f(x) = x \sin(1/x)$ se $x \neq 0$, em um intervalo finito contendo a origem, é um exemplo de função contínua em toda parte mas não absolutamente contínua. Para mais detalhes veja Kolmogorov e Fomin [6].

2.2 A Taxa de Falha e a Taxa de Falha Reversa

A taxa de falha é vista como a chance de que a falha ocorra num instante de tempo logo após t dado que não ocorreu antes de t . A taxa de falha reversa é a chance de ele falhar num instante de tempo t dado que funcionará até no máximo t .

A taxa de falha reversa relaciona-se com a função distribuição assim como a taxa de falha relaciona-se com a função de sobrevivência.

Denotando por f a densidade e por F a função distribuição, temos a taxa de falha (contínua e discreta)

$$(2.2.1) \quad \lambda_c(t) = \frac{f(t)}{1 - F(t^-)} \quad \text{e} \quad \lambda_d(t) = \frac{\Pr(T = t)}{1 - F(t^-)} = \frac{\Pr(T = t)}{\Pr(T \geq t)}$$

e a taxa de falha reversa (contínua e discreta)

$$(2.2.2) \quad \mu_c(t) = \frac{f(t)}{F(t)} \quad \text{e} \quad \mu_d(t) = \frac{\Pr(T = t)}{F(t)} = \frac{\Pr(T = t)}{\Pr(T \leq t)}.$$

i) Para funções de distribuição absolutamente contínuas

$$(2.2.3) \quad F(t) = 1 - \exp \left[\int_0^t \lambda_c(y) dy \right] = \exp \left[- \int_t^\infty \mu_c(y) dy \right].$$

ii) Para funções de distribuição discretas

$$(2.2.4) \quad F(t) = 1 - \prod_{y \leq t} [1 - \lambda_d(y)] = \prod_{y > t} [1 - \mu_d(y)].$$

iii) Para funções de distribuição com parte contínua e parte discreta

$$(2.2.5) \quad F(t) = 1 - \exp[-\Lambda(t)] = \exp[-M(t)].$$

Aqui, $\Lambda(\cdot)$ é a taxa de falha acumulada e $M(\cdot)$ é a taxa de falha reversa acumulada. Tais funções são definidas da seguinte maneira:

$$(2.2.6) \quad \Lambda(t) = \int_0^t \lambda_c(y) dy + \sum_{y \leq t} [-\ln(1 - \lambda_d(y))]$$

e

$$(2.2.7) \quad M(t) = \int_t^\infty \mu_c(y) dy + \sum_{y > t} [-\ln(1 - \mu_d(y))],$$

em que a integral é sobre intervalos abertos e disjuntos que não incluem os pontos de saltos de F .

2.3 Funções de Distribuição e Sub-distribuição

Considerando que temos apenas dois componentes no sistema e que a informação disponível é o tempo da falha do sistema e a componente que “produziu” a falha. Isto é, observamos T e δ , em que $T = \min(X_1, X_2)$ ou $T = \max(X_1, X_2)$ (se o sistema tem seus componentes em série ou paralelo) e $\delta = 1$ se o componente 1 produziu a falha ou $\delta = 2$ se foi o componente 2 o responsável. Assumimos que X_1 e X_2 são independentes.

Agora, definimos para a j -ésima componente, $j = 1, 2$, a sua função de distribuição por $F_j(t) = \Pr(X_j \leq t)$ e a sua função de sub-distribuição por $F_j^*(t) = \Pr(T \leq t, \delta = j)$, que nada mais é do que uma função de distribuição conjunta do tempo de falha do sistema (ser menor que um tempo $t > 0$) e da falha ocorrer pela j -ésima componente. O nosso principal interesse consiste em estimar a $F_j(\cdot)$. Porém, tanto no sistema em série, quanto no sistema em paralelo, estamos sujeitos a censura na observação dos dados. Desta forma nem sempre é possível observar diretamente as realizações das funções de distribuição de interesse. Por outro lado, a observação de realizações da função de sub-distribuição é facilmente obtida, que é equivalente a observação das realizações do par (T, δ) .

Algumas propriedades das funções de sub-distribuição $F_1^*(\cdot)$ e $F_2^*(\cdot)$ são dadas por:

2.3.1 Propriedade. *A função de sub-distribuição $F_1^*(\cdot)$ pode ser expressa em termos das funções de distribuição F_1 e F_2 como*

$$(2.3.2) \quad F_1^*(t) = \int_0^t [1 - F_2(s)] [dF_1(s)],$$

para o sistema em série e

$$(2.3.3) \quad F_1^*(t) = \int_0^t F_2(s) dF_1(s),$$

para o sistema em paralelo.

Demonstração. Para o Sistema em Série, temos:

$$\begin{aligned} F_1^*(t) &= \Pr(T \leq t, \delta = 1) \\ &= \Pr(T \leq t, X_1 < X_2) \\ &= \int_0^\infty \Pr(T \leq t, X_2 > s | X_1 = s) dF_1(s) \end{aligned}$$

Observemos que

$$\Pr(T \leq t, X_2 > s | X_1 = s) = \begin{cases} 0, & \text{se } s > t, \\ \Pr(X_2 > s | X_1 = s), & \text{se } s \leq t. \end{cases}$$

Assim,

$$\begin{aligned} F_1^*(t) &= \int_0^t \Pr(X_2 > s | X_1 = s) dF_1(s) \\ &= \int_0^t \Pr(X_2 > s) dF_1(s) \\ &= \int_0^t (1 - F_2(s)) dF_1(s). \end{aligned}$$

Para o Sistema em Paralelo, temos:

$$\begin{aligned} F_1^*(t) &= \Pr(T \leq t, \delta = 1) \\ &= \Pr(T \leq t, X_1 > X_2) \\ &= \int_0^\infty \Pr(T \leq t, X_2 < s | X_1 = s) dF_1(s) \end{aligned}$$

Observemos que

$$\Pr(T \leq t, X_2 < s | X_1 = s) = \begin{cases} 0, & \text{se } s > t, \\ \Pr(X_2 < s | X_1 = s), & \text{se } s \leq t. \end{cases}$$

Assim,

$$\begin{aligned} F_1^*(t) &= \int_0^t \Pr(X_2 < s | X_1 = s) dF_1(s) \\ &= \int_0^t \Pr(X_2 < s) dF_1(s) \\ &= \int_0^t F_2(s) dF_1(s). \end{aligned}$$

□

2.3.4 Definição. A confiabilidade de uma componente é a probabilidade desta provocar a falha do sistema, isto é, a confiabilidade de X_j é dada por $\Pr(\delta = j)$

Desta forma, temos a seguinte propriedade

2.3.5 Propriedade. As funções de sub-distribuição da componente j , quando $t \rightarrow \infty$, convergem para a sua confiabilidade e a soma das sub-distribuições é igual a 1.

1. $F_1^*(+\infty) = \Pr(\delta = 1)$;

2. $F_2^*(+\infty) = \Pr(\delta = 2)$;
3. $F_1^*(+\infty) + F_2^*(+\infty) = 1$.

Demonstração.

1.
$$\begin{aligned} F_1^*(+\infty) &= \Pr(T < \infty, \delta = 1) \\ &= \Pr(T < \infty | \delta = 1) \Pr(\delta = 1) = \Pr(\delta = 1). \end{aligned}$$
2.
$$\begin{aligned} F_2^*(+\infty) &= \Pr(T < \infty, \delta = 2) \\ &= \Pr(T < \infty | \delta = 2) \Pr(\delta = 2) = \Pr(\delta = 2). \end{aligned}$$
3.
$$F_1^*(+\infty) + F_2^*(+\infty) = \Pr(\delta = 1) + \Pr(\delta = 2) = 1.$$

□

2.3.6 Propriedade. *A soma das funções de sub-distribuição, para $t > 0$, é igual a função de distribuição do sistema ($F(t)$).*

$$\begin{aligned} F_1^*(t) + F_2^*(t) &= \Pr[(T \leq t, \delta = 1) \cup (T \leq t, \delta = 2)] \\ &= \Pr(T \leq t) = F(t). \end{aligned}$$

Note que, a nomenclatura “sub-distribuição” é motivada pela Propriedade 2.3.6, uma vez que as sub-distribuições podem ser vistas como uma decomposição da distribuição do sistema.

2.3.7 Propriedade.

1. *Os pontos de salto de $F_j(\cdot)$ são pontos de salto de $F_j^*(\cdot)$ e vice-versa, $j = 1, 2$;*
2. *Se $F_1(\cdot)$ e $F_2(\cdot)$ não têm pontos de salto em comum, então $F_1^*(\cdot)$ e $F_2^*(\cdot)$ também não tem;*

2.3.8 Propriedade. *Ao menos uma das duas funções de sub-distribuição $F_1^*(\cdot)$ e $F_2^*(\cdot)$ é positiva para $t < t^*$, e ambas são um para $t \geq t^*$, em que t^* define o suporte da função de distribuição de F_j , $j = 1, 2$.*

Segundo Breslow e Crowley [2], para sistema de componentes em série, se as funções $F_1^*(\cdot)$ e $F_2^*(\cdot)$ forem contínuas, a função de distribuição $F_1(\cdot)$ pode ser expressa em termos

das funções de sub-distribuição $F_1^*(\cdot)$ e $F_2^*(\cdot)$ por,

$$(2.3.9) \quad F_1(t) = \int_t^\infty \frac{dF_1^*(s)}{F_1^*(s) + F_2^*(s)}.$$

Porém, tal equação não será válida quando as funções de sub-distribuição $F_1^*(\cdot)$ e $F_2^*(\cdot)$ não forem absolutamente contínuas.

O funcional proposto por Peterson [8] é uma generalização da Equação (2.3.9), para o sistema em série, e retira a restrição de $F_1^*(\cdot)$ e $F_2^*(\cdot)$ serem absolutamente contínuas, ficando apenas com a restrição de que $F_1^*(\cdot)$ e $F_2^*(\cdot)$ não podem ter pontos de descontinuidade em comum. Polpo e Pereira [10], apresentaram a versão dual de Peterson [8], para o sistema em paralelo. Estes funcionais são dados no teorema abaixo.

2.3.10 Teorema. *As funções sub-distribuição F_1^* e F_2^* , determinam (unicamente) a função distribuição F_1 para $t < t^*$. Isto é, para o sistema em série temos*

$$(2.3.11) \quad F_1(t) = \Phi_s(F_1^*, F_2^*, t) \\ = 1 - \exp \left[\int_0^t \frac{-dF_1^*(s)}{1 - (F_1^*(s) + F_2^*(s))} + \sum_{s \leq t} \ln \left(\frac{1 - (F_1^*(s^+) + F_2^*(s^+))}{1 - (F_1^*(s^-) + F_2^*(s^-))} \right) \right]$$

e para o sistema em paralelo, temos

$$(2.3.12) \quad F_1(t) = \Phi_p(F_1^*, F_2^*, t) \\ = \exp \left[\int_t^\infty \frac{-dF_1^*(s)}{F_1^*(s) + F_2^*(s)} + \sum_{s > t} \ln \left(\frac{F_1^*(s^-) + F_2^*(s^-)}{F_1^*(s^+) + F_2^*(s^+)} \right) \right].$$

Demonstração. Veja que, se $F_1(t^*) = 0$, então $F_1(t) = 0$ para todo $t > t^*$, isto é, o maior tempo de falha possível é do componente 1 é t^* .

Tomando a \sum sob todos os pontos de salto de F_1^* , das equações (2.2.5), (2.2.6) e (2.2.7) é suficiente provar que, para o sistema em série,

$$(2.3.13) \quad \int_0^t \frac{dF_1^*(s)}{1 - (F_1^*(s) + F_2^*(s))} = \int_0^t \lambda_c^1(s) ds,$$

e

$$(2.3.14) \quad \sum_{s \leq t} \ln \left(\frac{1 - (F_1^*(s^+) + F_2^*(s^+))}{1 - (F_1^*(s^-) + F_2^*(s^-))} \right) = \sum_{s \leq t} \ln[1 - \lambda_d^1(s)].$$

em que λ_c^1 e λ_d^1 são respectivamente as taxas de falha contínua e discreta do componente

X_1 . Para o sistema em paralelo, basta provar que

$$(2.3.15) \quad \int_t^\infty \frac{dF_1^*(s)}{F_1^*(s) + F_2^*(s)} = \int_t^\infty \mu_c^1(s) ds,$$

e

$$(2.3.16) \quad \sum_{s>t} \ln \left(\frac{F_1^*(s^-) + F_2^*(s^-)}{F_1^*(s^+) + F_2^*(s^+)} \right) = \sum_{s>t} \ln[1 - \mu_d^1(s)].$$

em que μ_c^1 e μ_d^1 são respectivamente as taxas de falha contínua e discreta do componente X_1 .

A Equação (2.3.13) é consequência das propriedades 2.3.1 e 2.3.6 e de

$$\begin{aligned} \int_0^t \frac{dF_1^*(s)}{1 - (F_1^*(s) + F_2^*(s))} &= \int_0^t \frac{(1 - F_2(s))dF_1(s)}{1 - F(s)} \\ &= \int_0^t \frac{(1 - F_2(s))dF_1(s)}{(1 - F_1(s))(1 - F_2(s))} \\ &= \int_0^t \frac{dF_1(s)}{1 - F_1(s)} \\ &= \int_0^t \lambda_c^1(s) ds. \end{aligned}$$

A Equação (2.3.16) é consequência da Propriedade 2.3.6, do fato de que $F_2(s^-) = F_2(s^+)$ quando s é ponto de salto de $F_1(\cdot)$ e de

$$\begin{aligned} \sum_{s>t} \ln \left(\frac{1 - (F_1^*(s^-) + F_2^*(s^-))}{1 - (F_1^*(s^+) + F_2^*(s^+))} \right) &= \sum_{s>t} \ln \left(\frac{1 - F(s^-)}{1 - F(s^+)} \right) \\ &= \sum_{s>t} \ln \left(\frac{(1 - F_1(s^-))(1 - F_2(s^-))}{(1 - F_1(s^+))(1 - F_2(s^+))} \right) \\ &= \sum_{s>t} \ln \left(\frac{(1 - F_1(s^-))}{(1 - F_1(s^+))} \right) \\ &= \sum_{s>t} \ln [1 - \lambda_d^1(s)]. \end{aligned}$$

A Equação (2.3.15) é consequência das propriedades 2.3.1 e 2.3.6 e de

$$\begin{aligned} \int_t^\infty \frac{dF_1^*(s)}{F_1^*(s) + F_2^*(s)} &= \int_t^\infty \frac{F_2(s)dF_1(s)}{F_1(s)F_2(s)} \\ &= \int_t^\infty \frac{dF_1(s)}{F_1(s)} \\ &= \int_t^\infty \mu_c^1(s) ds. \end{aligned}$$

A Equação (2.3.16) é consequência da Propriedade 2.3.6, do fato de que $F_2(s^-) = F_2(s^+)$

quando s é ponto de salto de $F_1(\cdot)$ e de

$$\begin{aligned} \sum_{s>t} \ln \left(\frac{F_1^*(s^-) + F_2^*(s^-)}{F_1^*(s^+) + F_2^*(s^+)} \right) &= \sum_{s>t} \ln \left(\frac{F_1(s^-)F_2(s^-)}{F_1(s^+)F_2(s^+)} \right) \\ &= \sum_{s>t} \ln \left(\frac{F_1(s^-)}{F_1(s^+)} \right) \\ &= \sum_{s>t} \ln [1 - \mu_d^1(s)]. \end{aligned}$$

□

Desta forma, os funcionais $\Phi_s(\cdot)$ e $\Phi_p(\cdot)$ são as inversas das equações (2.3.2) e (2.3.3). Com isso temos as funções de distribuição F_1 e F_2 expressas em termos das sub-distribuições F_1^* e F_2^* . Obviamente, para obter resultado semelhante para o componente 2, basta substituir apropriadamente os índices 1 por 2 (e vice-versa) no Teorema 2.3.10.

3 Estimador Bayesiano não paramétrico

3.1 Introdução

Consideramos agora uma amostra de tamanho n do par (T, δ) . Observar os dados $(T_1, \delta_1), \dots, (T_n, \delta_n)$ equivale a observar, para cada $t > 0$, o vetor de contagem aleatória $n\mathbf{F}_n^*(t) = (nF_{1n}^*(t), nF_{2n}^*(t), n(1 - F_n(t)))$ que assumimos ter distribuição trinomial com amostra de tamanho n e parâmetros $(F_1^*(t), F_2^*(t); 1 - F(t))$, em que

$$(3.1.1) \quad F_n^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t),$$

$$(3.1.2) \quad F_{jn}^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, \delta_i = j), \quad j = 1, 2,$$

em que $F_n^*(t)$ é a função distribuição empírica do sistema, $F_{jn}^*(t)$ é a função sub-distribuição empírica do j -ésimo componente, $I(A)$ é a função indicadora do conjunto A e pela Propriedade 2.3.6, $F_1^*(t) + F_2^*(t) + (1 - F(t)) = 1$. Note que, $nF_{jn}^*(t)$ é a quantidade de falhas do sistema até o tempo t devido ao componente j e $n(1 - F_n(t))$ é o total de amostras do sistema que não falharam até o tempo t . Desta forma, a informação referente aos tempos de falha (do sistema e componentes) contida em $n\mathbf{F}_n^*(t)$ é equivalente à informação contida em $(T_1, \delta_1), \dots, (T_n, \delta_n)$. Além disso, $nF_{jn}^*(t)$ ocorre com probabilidade dada pela verdadeira sub-distribuição $F_j^*(t)$, desconhecida.

3.1.3 Definição. *Seja $(\mathcal{X}; A)$ um espaço mensurável e α_1, α_2 medidas finitas, não-negativas e não-nulas em $(\mathcal{X}; A)$. Seja $\rho = (\rho_1, \rho_2)$; P_1, P_2 elementos aleatórios mutuamente independentes definidos em $(\Omega; \mathcal{F}; \mathcal{Q})$. Suponha que ρ tenha uma distribuição de Dirichlet $D(\alpha_1(\mathcal{X}), \alpha_2(\mathcal{X}))$ e P_j é um processo de Dirichlet com parâmetro α_j , ou seja, $P_j \sim D(\alpha_j)$, $j = 1, 2$. Definimos $P^* = (P_1^*, P_2^*) = (\rho_1 P_1, \rho_2 P_2)$. Então P^* é um processo de Dirichlet bivariado com parâmetro $\alpha = (\alpha_1, \alpha_2)$, ou seja, $\mathbf{P}^* \sim DM_2(\alpha)$ (cf., Polpo e Pereira [10] e Salinas [13]).*

No contexto de confiabilidade, sejam α_1, α_2 medidas finitas, não-negativas e não-nulas em $((0, \infty), \mathcal{B}_{(0, \infty)})$. Sejam também $\rho = (\rho_1, \rho_2) = (\Pr(\delta = 1), \Pr(\delta = 2)) \sim D_s(\alpha_1(0, \infty), \alpha_2(0, \infty))$, $P_j^* = \Pr(T \leq t | \delta = j)$, $F^* \sim D(\alpha_j)$, $j=1,2$. Suponha que $\rho, \mathbf{P}_1^*, \mathbf{P}_2^*$ são mutuamente independentes. Então, a priori para \mathbf{F}^* é dada por $\mathbf{F}^* = (\rho_1 T_1^*, \rho_2 T_2^*) \sim DM_2(\alpha_1, \alpha_2)$.

A priori induzida para F_1^* é dada por:

$$(3.1.4) \quad F_1^*(t) \sim \text{Beta}(c_2 F_{1,0}^*(t), c_2(1 - F_{1,0}^*(t))), \quad t > 0,$$

em que $c_2 = \sum_{j=1}^2 \alpha_j(0, \infty)$ e $F_{1,0}^*(t) = \alpha_1(0, t)/c_2$ é a média a priori de F_1^* . Também, $F_0(t) = F_{1,0}(t) + F_{2,0}(t)$ é a média priori de $F(\cdot)$.

Considerando o processo de Dirichlet como priori para $F^*(t)$, a distribuição a posteriori de $\mathbf{F}^*(t)$ é um processo de Dirichlet bivariado dado por:

$$(3.1.5) \quad \mathbf{F}^*(t) | n\mathbf{F}_n^*(t) \sim PD(\alpha_1(t, \infty) + nF_{1n}^*(t), \alpha_2(t, \infty) + nF_{2n}^*(t); \\ \alpha_1(0, t] + \alpha_2(0, t] + n(1 - F_n(t))),$$

em que F_{jn}^* é como definido na Equação (3.1.2).

Seja $p_n = c_2/(c_2 + n)$, os estimadores Bayesianos de $F_1^*(\cdot)$ e $F(\cdot)$ são dados por

$$(3.1.6) \quad \hat{F}_1^*(t) = p_n F_{1,0}^*(t) + (1 - p_n) F_{1n}^*(t)$$

e

$$(3.1.7) \quad \hat{F}(t) = \hat{F}_1^*(t) + \hat{F}_2^*(t)$$

Note que, para estimação de F_2^* basta alterar apropriadamente a Equação (3.1.6). No restante do texto nos referimos apenas à estimação do componente 1, uma vez que para o componente 2 os resultados são análogos.

3.2 Sistema em Série

Sejam m ($m \leq n$) estatísticas de ordem distintas, de $T, T_{(1)} < \dots < T_{(m)}$, $n_i = \sum_{j=1}^n I(T_j \geq T_{(i)})$ e $d_i = \sum_{j=1}^n I(T_j = T_{(i)}, \delta_j = 1)$, $i = 1, \dots, m$.

Definimos

$$(3.2.1) \quad i(t) = \exp \left\{ \frac{-1}{c_2 + n} \int_0^t \frac{d\alpha_1(0, s]}{1 - \hat{F}(s)} \right\}$$

e

$$(3.2.2) \quad \pi(t) = \prod_{i:T(i) \leq t} \frac{\alpha_1(T(i), \infty) + \alpha_2(T(i), \infty) + n_i - d_i}{\alpha_1(T(i), \infty) + \alpha_2(T(i), \infty) + n_i},$$

3.2.3 Teorema. *Suponha que a função $f(s) = (\alpha_1(0, s], \dots, \alpha_r(0, s])$ seja contínua em $(0, t)$, para cada $t > 0$, e F_1 e F_2 não tenham pontos de salto em comum, então, para cada $t \leq T_{(m)}$, temos*

$$(3.2.4) \quad \hat{F}_1(t) = \Phi_s(\hat{F}_1^*(\cdot), \hat{F}_2^*(\cdot), t) = 1 - i(t)\pi(t).$$

Demonstração. Fazendo de forma similar à Peterson [8], substituímos os estimadores Bayesianos de F_1^* e F_2^* na Equação (2.3.11) e temos

$$(3.2.5) \quad \hat{F}_1(t) = 1 - \exp \left[\int_0^t \frac{-d\hat{F}_1^*(s)}{1 - \hat{F}(s)} \right] \prod_{s \leq t} \frac{1 - \sum_{j=1}^2 \hat{F}_j^*(s^+)}{1 - \sum_{j=1}^2 \hat{F}_j^*(s^-)}, \quad t < T_{(m)},$$

em que $\prod_{s \leq t}$ é o produto sobre todos os pontos de salto s de \hat{F}_1^* e a integral é sob os intervalos disjuntos que não incluem os pontos de salto de $\hat{F}_1^*(s)$.

Da Equação (3.1.6) temos

$$\begin{aligned} \hat{F}_1^* &= \frac{c_2 F_{1,0}^*(t)}{c_2 + n} + \left(1 - \frac{c_2}{c_2 + n}\right) F_{1n}^*(t) \\ &= \frac{c_2 F_{1,0}^*(t)}{c_2 + n} + \frac{n F_{1n}^*(t)}{c_2 + n} \\ &= \frac{\alpha_1(0, t]}{c_2 + n} + \frac{\sum_{i=1}^n I(T_i \leq t, \delta_i = 1)}{c_2 + n}. \end{aligned}$$

A função $\sum_{i=1}^n I(T_i \leq t, \delta_i = 1)$ é constante, exceto nos pontos de salto, assim sua derivada é zero para intervalos contínuos. Então,

$$d\hat{F}_1^*(t) = \frac{d\alpha_1(0, t]}{c_2 + n}$$

e o primeiro termo da Equação (3.2.5) torna-se $i(t)$. Para cada $t > 0$ fixado, $\alpha_j(0, \cdot]$, $j = 1, 2$ são funções monótonas e contínuas em $(0, t)$, e $(1/\hat{F}(\cdot))$ é monótona em $(0, t)$. Portanto, $\alpha_j(0, \cdot]$, $j = 1, \dots, k$, pode ser decomposto unicamente como a diferença de funções monótona contínua e $(1/\hat{F}(\cdot))$, função monótona. Assim, a integral $\int_0^t [d\alpha_1(0, s]/(1 - \hat{F}(s))]$

é bem definida. Além disso, o segundo fator da Equação (3.2.5) é

$$\begin{aligned} & \prod_{s \leq t} \frac{1 - \sum_{j=1}^2 \alpha_j(0, s] + \sum_{i=1}^n I(T_i \geq s^+)}{1 - \sum_{j=1}^2 \alpha_j(0, s] + \sum_{i=1}^n I(T_i \geq s^-)} \\ &= \prod_{s \leq t} \frac{\sum_{j=1}^2 \alpha_j(s, \infty) + \sum_{i=1}^n I(T_i \geq s^+)}{\sum_{j=1}^2 \alpha_j(s, \infty) + \sum_{i=1}^n I(T_i \geq s^-)} = \pi(t), \end{aligned}$$

em que s é um ponto de salto de \hat{F}_1^* , por fim $t < T_{(m)}$ segue da Propriedade 2.3.8. \square

3.3 Sistema em Paralelo

Sejam $m (\leq n)$ estatísticas de ordem distintas de T , $T_{(1)} < \dots < T_{(m)}$. $N_i = \sum_{j=1}^n I(T_j < T_{(i)})$, e $D_i = \sum_{j=1}^n I(T_j = T_{(i)}, \delta_j = 1)$, $i = 1, \dots, m$. Definimos

$$(3.3.1) \quad I(t) = \exp \left[\frac{-1}{n + c_2} \int_t^\infty \frac{d\alpha_1(0, s]}{\hat{F}(s)} \right]$$

e

$$(3.3.2) \quad \Pi(t) = \prod_{i: T_{(i)} > t} \frac{\sum_{j=1}^2 \alpha_j(0, T_{(i)}] + N_i}{\sum_{j=1}^2 \alpha_j(0, T_{(i)}] + N_i + D_i},$$

3.3.3 Teorema. *Suponha que $\alpha_1(0, \cdot), \alpha_2(0, \cdot)$ são contínuos em (t, ∞) , para cada $t > 0$, e F_1 e F_2 não têm pontos de descontinuidade em comum. Então, para $t < T_{(m)}$,*

$$(3.3.4) \quad \hat{F}_1(t) = \Phi_p(\hat{F}_1^*, \hat{F}_2^*, t) = I(t)\Pi(t),$$

é o estimador Bayesiano não paramétrico de $F_1(t)$ baseado na média posteriori.

Demonstração. Fazendo de forma similar à Polpo e Pereira [10], substituímos os estimadores Bayesianos de F_1^* e F_2^* na Equação (2.3.12) e temos

$$(3.3.5) \quad \hat{F}_1(t) = \exp \left[\int_t^\infty \frac{-d\hat{F}_1^*(s)}{\hat{F}(s)} \right] \prod_{s > t} \frac{\sum_{j=1}^2 \hat{F}_j^*(s^-)}{\sum_{j=1}^2 \hat{F}_j^*(s^+)}, \quad t < T_{(m)},$$

em que $\prod_{s > t}$ é o produto sobre todos os pontos de salto s de \hat{F}_1^* e a integral é sob os intervalos disjuntos que não incluem os pontos de salto de $\hat{F}_1^*(s)$.

Note que, como visto na demonstração do Teorema 3.2.3, $d\hat{F}_1^*(s) = d\alpha_1(0, s]/(c_2 + n)$ e o primeiro termo da Equação (3.3.5) torna-se $I(t)$. Para cada $t > 0$ fixado, $\alpha_j(0, \cdot)$, $j = 1, 2$ são funções monótonas e contínuas em (t, ∞) , e $(1/\hat{F}(\cdot))$ é monótona em (t, ∞) . Por-

tanto, $\alpha_j(0, \cdot)$, $j = 1, \dots, k$, pode ser decomposto unicamente como a diferença de funções monótona contínua e $(1/\hat{F}(\cdot))$, função monótona. Assim, a integral $\int_t^\infty [d\alpha_1(0, s)/\hat{F}(s)]$ é bem definida. Além disso, o segundo fator da Equação (3.3.5) é

$$\prod_{s>t} \frac{\sum_{j=1}^2 \alpha_j(0, s] + \sum_{i=1}^n I(T_i \leq s^-)}{\sum_{j=1}^2 \alpha_j(0, s] + \sum_{i=1}^n I(T_i \leq s^+)} = \Pi(t).$$

em que s é um ponto de salto de \hat{F}_1^* . Por fim, $t < T_{(m)}$ segue da Propriedade 2.3.8. \square

4 Generalização

4.1 Introdução

A partir de agora, trabalharemos com sistemas em série e em paralelo com k componentes, $k \geq 2$. Para isso dividimos os k componentes em dois grupos, Δ e Δ^c , de tal forma que $n(\Delta) + n(\Delta^c) = k$. De agora em diante, substituiremos as funções F_1 por F_Δ e F_2 por F_{Δ^c} .

Observar os dados $(T_1, \delta_1), \dots, (T_n, \delta_n)$ equivale a observar, para cada $t > 0$, o vetor de contagem aleatória $n\mathbf{F}_n^*(t) = (nF_{\Delta n}^*(t), nF_{\Delta^c n}^*(t), n(1 - F_n(t)))$ que tem distribuição trinomial com amostra de tamanho n e parâmetros $(F_\Delta^*(t), F_{\Delta^c}^*(t); 1 - F(t))$, em que

$$(4.1.1) \quad F_n^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t),$$

$$(4.1.2) \quad F_{jn}^*(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, \delta_i = j), \quad j = 1, \dots, r,$$

em que $F_n^*(t)$ função distribuição empírica do sistema, $F_{jn}^*(t)$ é a função sub-distribuição empírica do j -ésimo componente, $I(A)$ é a função indicadora do conjunto A e pela Propriedade 2.3.6, $F_\Delta^*(t) + F_{\Delta^c}^*(t) + (1 - F(t)) = 1$. Note que, $nF_{jn}^*(t)$ é a quantidade de falhas do sistema até o tempo t devido ao componente j e $n(1 - F_n(t))$ é o total de amostras do sistema que não falharam até o tempo t . Desta forma, para todos os t , a informação referente aos tempos de falha (do sistema e componentes) contida em $n\mathbf{F}_n^*(t)$ é equivalente a informação contida em $(T_1, \delta_1), \dots, (T_n, \delta_n)$. Além disso, $nF_{jn}^*(t)$ ocorre com probabilidade dada pela verdadeira sub-distribuição $F_j^*(t)$, desconhecida.

4.1.3 Definição. *Seja $(\mathcal{X}; A)$ um espaço mensurável e $\alpha_1, \dots, \alpha_r$ medidas finitas, não-negativas e não-nulas em $(\mathcal{X}; A)$. Seja $\rho = (\rho_1, \dots, \rho_r)$; P_1, \dots, P_r elementos aleatórios mutuamente independentes definidos em $(\Omega; \mathcal{F}; \mathcal{Q})$. Suponha que ρ tenha uma distribuição de Dirichlet $D(\alpha_1(\mathcal{X}), \dots, \alpha_r(\mathcal{X}))$ e P_j é um processo de Dirichlet com parâmetro α_j , ou seja, $P_j \sim D(\alpha_j)$, $j = 1, \dots, r$. Definimos $P^* = (P_1^*, \dots, P_r^*) = (\rho_1 P_1, \dots, \rho_r P_r)$.*

Então P^* é um processo de Dirichlet multivariado com parâmetro $\alpha = (\alpha_1, \dots, \alpha_r)$, ou seja, $\mathbf{P}^* \sim DM_r(\alpha)$ (cf., Polpo e Pereira [10] e Salinas [13]).

No contexto de confiabilidade, sejam $\alpha_1, \dots, \alpha_r$ medidas finitas, não-negativas e não-nulas em $((0, \infty), \mathcal{B}_{(0, \infty)})$. Sejam também $\rho = (\rho_1, \dots, \rho_r) = (\Pr(\delta = 1), \dots, \Pr(\delta = r)) \sim D_s(\alpha_1(0, \infty), \dots, \alpha_r(0, \infty))$, $P_j^* = \Pr(T \leq t | \delta = j)$, $P^* \sim D(\alpha_j)$, $j = 1, \dots, r$. Suponha que $\rho, \mathbf{P}_1^*, \dots, \mathbf{P}_r^*$ são mutuamente independentes. Então, a priori para \mathbf{F}^* é dada por $\mathbf{F}^* = (\rho_1 T_1^*, \dots, \rho_r T_r^*) \sim DM(\alpha_1, \dots, \alpha_r)$.

A priori induzida para F_{Δ}^* é dada por:

$$(4.1.4) \quad F_{\Delta}^*(t) \sim \text{Beta}(c_r F_{\Delta,0}^*(t), c_r(1 - F_{\Delta,0}^*(t))), \quad t > 0,$$

em que $c_r = \sum_{j=1}^r \alpha_j(0, \infty)$ e $F_{\Delta,0}^*(t) = \sum_{i \in \Delta} \alpha_i(0, t)/c_r$, $i = 1, \dots, r$ é a média a priori de F_{Δ}^* . Também, $F_0(t) = F_{\Delta,0}(t) + F_{\Delta^c,0}(t)$ é a média priori de $F(\cdot)$.

Considerando o processo de Dirichlet como priori para $F^*(t)$, a distribuição a posteriori de $\mathbf{F}^*(t)$ é um processo de Dirichlet multivariado dado por:

$$(4.1.5) \quad \mathbf{F}^*(t) | n\mathbf{F}_n^*(t) \sim PD(\alpha_1(t, \infty) + nF_{1n}^*(t), \dots, \alpha_r(t, \infty) + nF_{rn}^*(t); \\ \alpha_1(0, t] + \dots + \alpha_r(0, t] + n(1 - F_n(t))),$$

em que F_{jn}^* é como definido na Equação (4.1.2).

Seja $p_n = c_r/(c_r + n)$, os estimadores Bayesianos de $F_{\Delta}^*(\cdot)$ e $F(\cdot)$ são dados por

$$(4.1.6) \quad \hat{F}_{\Delta}^* = p_n F_{\Delta,0}^* + (1 - p_n) F_{\Delta n}^*(t)$$

e

$$(4.1.7) \quad \hat{F}(t) = \hat{F}_{\Delta}^*(t) + \hat{F}_{\Delta^c}^*(t)$$

Note que, para estimação de F_{Δ}^* basta alterar apropriadamente a Equação (4.1.6). No restante do texto nos referimos apenas a estimação do componente Δ , uma vez que para o componente Δ^c os resultados são análogos.

4.2 Sistema em Série

Sejam m ($m \leq n$) estatísticas de ordem distintas, de $T, T_{(1)} < \dots < T_{(m)}$, $n_i = \sum_{j=1}^n I(T_j \geq T_{(i)})$ e $d_{\Delta i} = \sum_{j=1}^n I(T_j = T_{(i)}, \delta_j \in \Delta)$, $i = 1, \dots, m$.

Definimos

$$(4.2.1) \quad i_{\Delta}(t) = \exp \left\{ \frac{-1}{c_r + n} \sum_{j \in \Delta} \int_0^t \frac{d\alpha_j(0, s]}{1 - \hat{F}(s)} \right\}$$

e

$$(4.2.2) \quad \pi_{\Delta}(t) = \prod_{i: T(i) \leq t} \frac{\sum_{j=1}^r \alpha_j(T(i), \infty) + n_i - d_{\Delta i}}{\sum_{j=1}^r \alpha_j(T(i), \infty) + n_i},$$

4.2.3 Teorema. *Suponha que a função $f(s) = (\alpha_1(0, s], \dots, \alpha_r(0, s])$ seja contínua em $(0, t)$, para cada $t > 0$, e F_{Δ} e F_{Δ^c} não tenham pontos de salto em comum, então, para cada $t \leq T_{(m)}$, temos*

$$(4.2.4) \quad \hat{F}_{\Delta}(t) = \Phi_s(\hat{F}_{\Delta}^*(\cdot), \hat{F}_{\Delta^c}^*(\cdot), t) = 1 - i_{\Delta}(t)\pi_{\Delta}(t).$$

Demonstração. A demonstração é análoga a do Teorema 3.2.3, substituindo F_1 por F_{Δ^c} e F_2 por F_{Δ} , e fazendo as modificações necessárias. \square

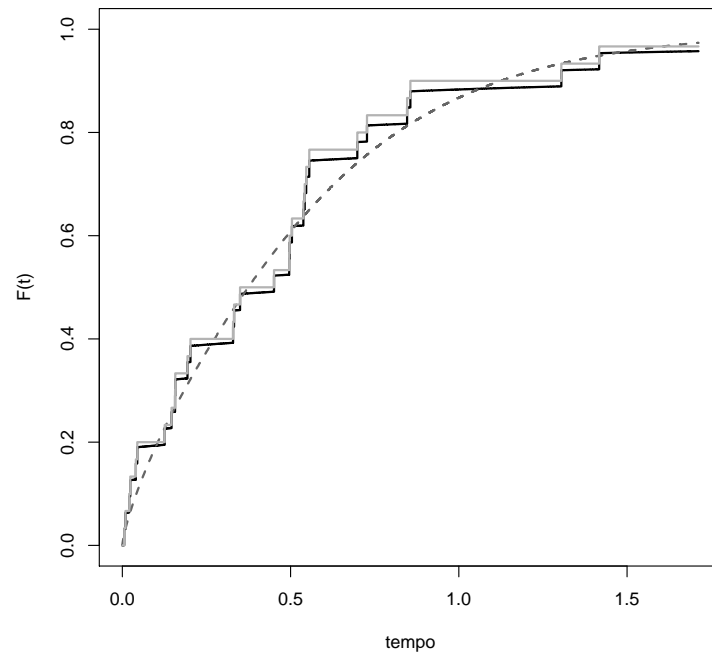
4.2.5 Exemplo. Temos um sistema com três componentes ligados em série, X_1 tem distribuição exponencial com média 1.4 e variância 0.5, X_2 tem distribuição gama com média 1.4 e variância 3.2 e X_3 tem distribuição log-normal com média 1.5 e variância 1.3.

Simulamos observações do sistema considerando três casos distintos em que os tamanhos amostrais foram $n = 1000$, $n = 100$ e $n = 30$, respectivamente. Como o sistema está ligado em série, o primeiro componente a falhar é o responsável pela falha do sistema, e os outros são censurados. Com isso obtemos o seguinte resultado:

Para $n = 30$

- O componente 1 foi responsável pela falha do sistema por 15 vezes.
- O componente 2 foi responsável pela falha do sistema por 11 vezes.
- O Componente 3 foi responsável pela falha do sistema por 4 vezes.

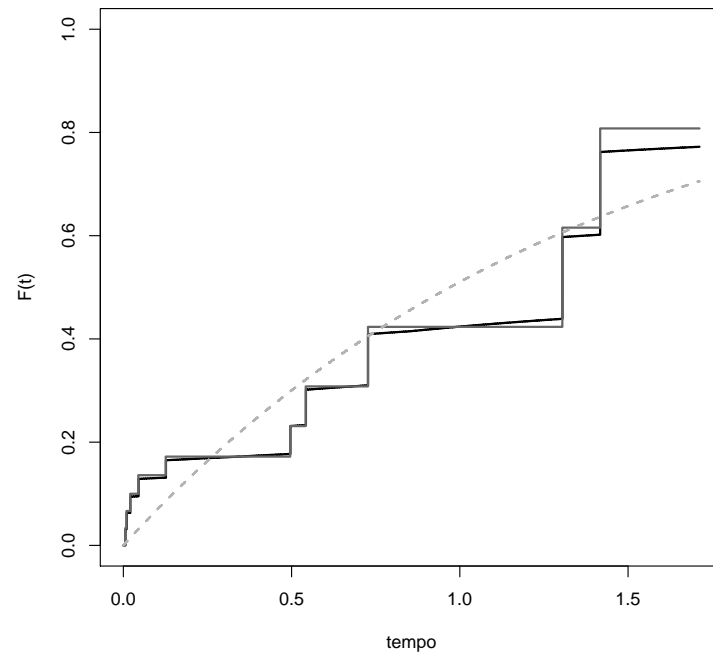
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) calculamos a estimativa da função distribuição do sistema, dado na Figura 4.2.6. Para solução da integral foi utilizado um método numérico baseado na regra de Simpson, para este e outros métodos numéricos para solução de integrais ver Davis e Rabinowitz [3].



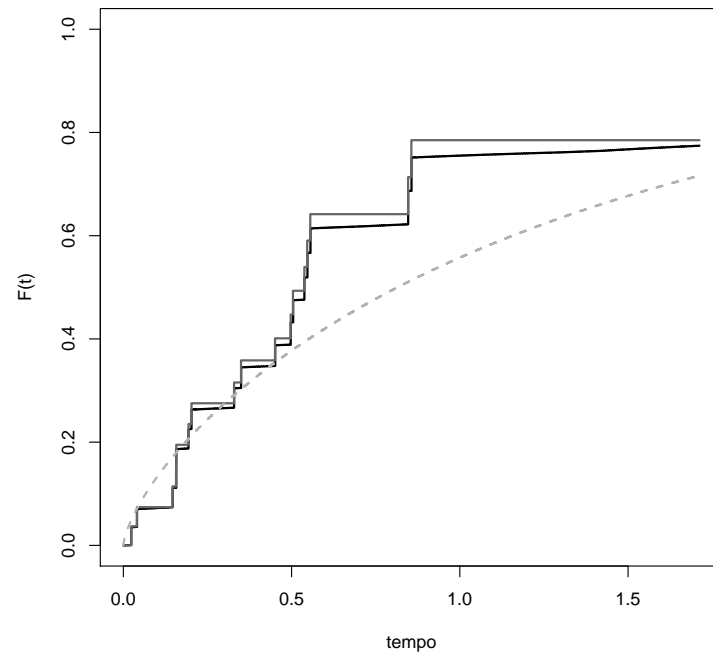
4.2.6 Figura. *Função distribuição do sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está muito próxima da função de distribuição geradora dos dados.

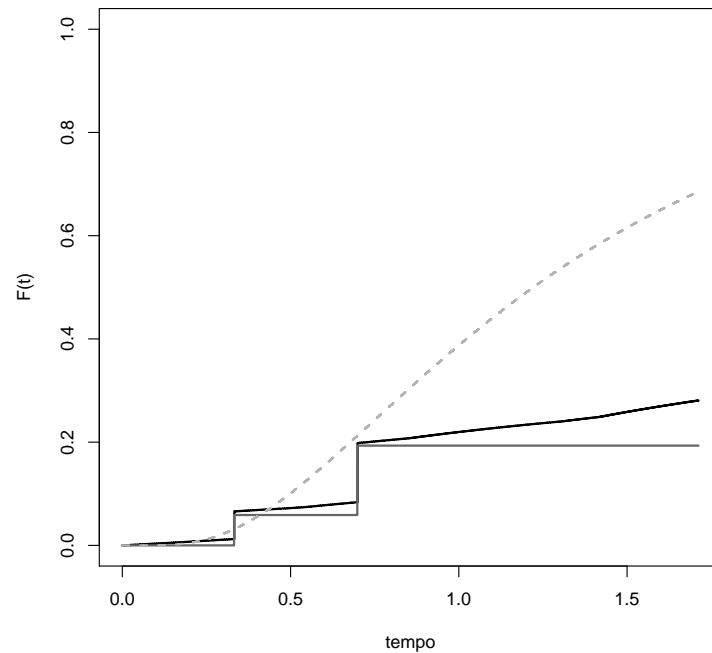
Agora, utilizando a Equação (4.2.4), obtemos a estimativa da função distribuição do j -ésimo componente, dado nas figuras 4.2.7-4.2.9.



4.2.7 Figura. *Função distribuição do componente 1 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.2.8 Figura. *Função distribuição do componente 2 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*



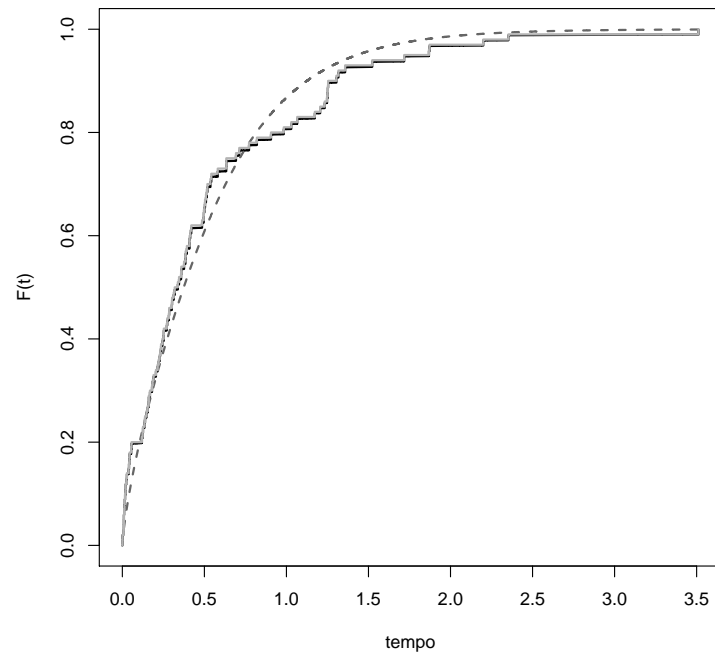
4.2.9 Figura. *Função distribuição do componente 3 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Como podemos observar pelas figuras, as estimativas ficaram próximas das verdadeiras funções de distribuição dos dados.

Para $n = 100$

- O componente 1 foi responsável pela falha do sistema por 49 vezes.
- O componente 2 foi responsável pela falha do sistema por 34 vezes.
- O Componente 3 foi responsável pela falha do sistema por 17 vezes.

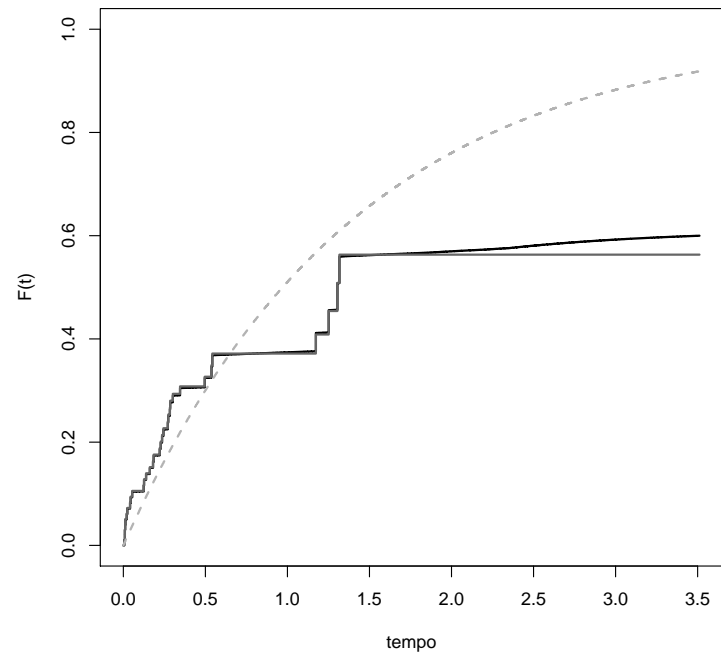
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) calculamos a estimativa da função distribuição do sistema, dado na Figura 4.2.10. Para solução da integral novamente foi utilizado o método numérico baseado na regra de Simpson.



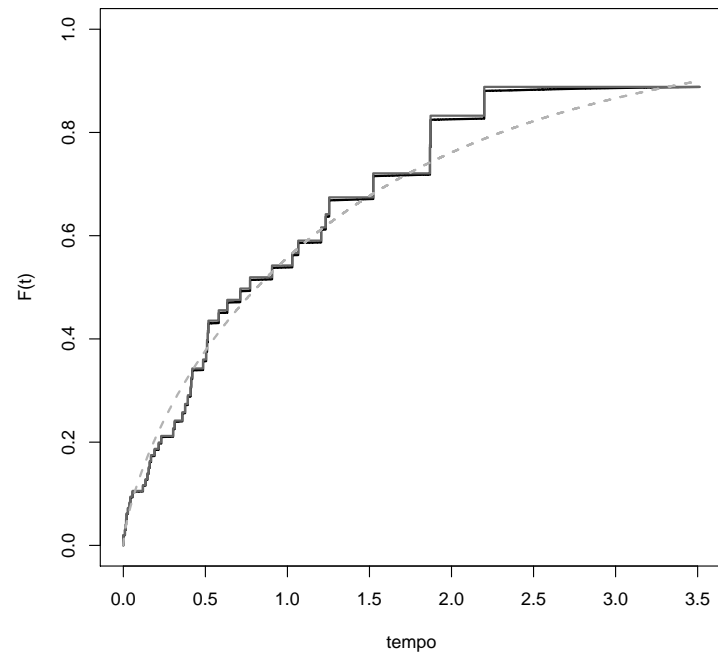
4.2.10 Figura. *Função distribuição do sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está muito próxima da função de distribuição geradora dos dados.

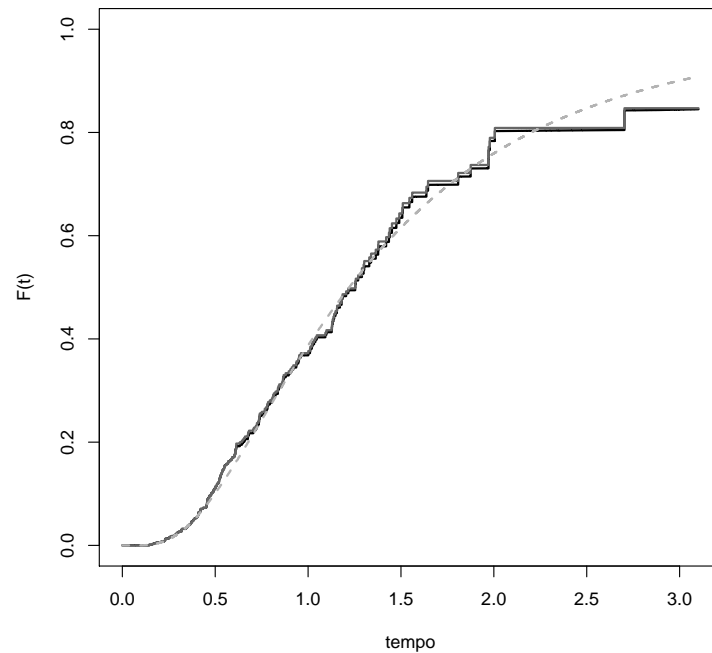
Agora, utilizando a Equação (4.2.4), obtemos a estimativa da função distribuição do j -ésimo componente, dado nas figuras 4.2.11-4.2.13.



4.2.11 Figura. *Função distribuição do componente 1 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.2.12 Figura. *Função distribuição do componente 2 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*



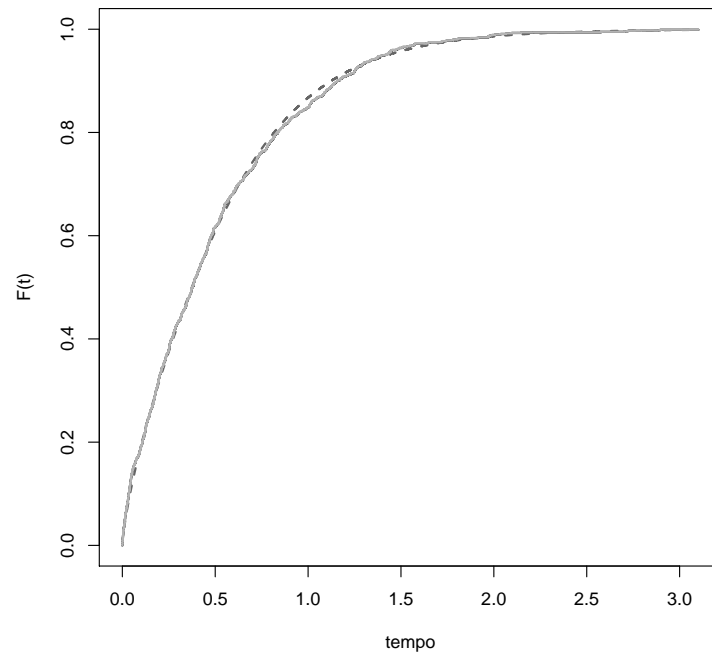
4.2.13 Figura. *Função distribuição do componente 3 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Como podemos observar pelas figuras, as estimativas ficaram próximas das verdadeiras funções de distribuição dos dados.

Para $n = 1000$

- O componente 1 foi responsável pela falha do sistema por 527 vezes.
- O componente 2 foi responsável pela falha do sistema por 328 vezes.
- O Componente 3 foi responsável pela falha do sistema por 145 vezes.

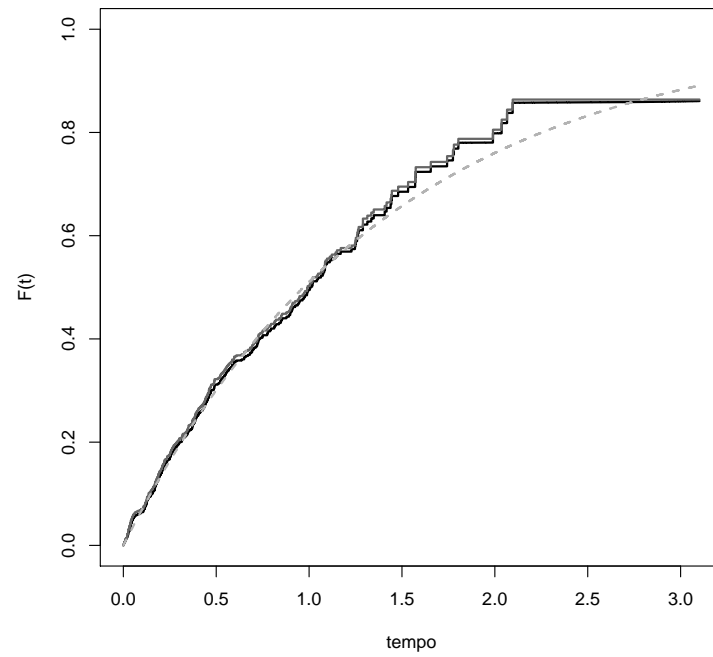
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) calculamos a estimativa da função distribuição do sistema, dado na Figura 4.2.14. Para solução da integral novamente foi utilizado o método numérico baseado na regra de Simpson.



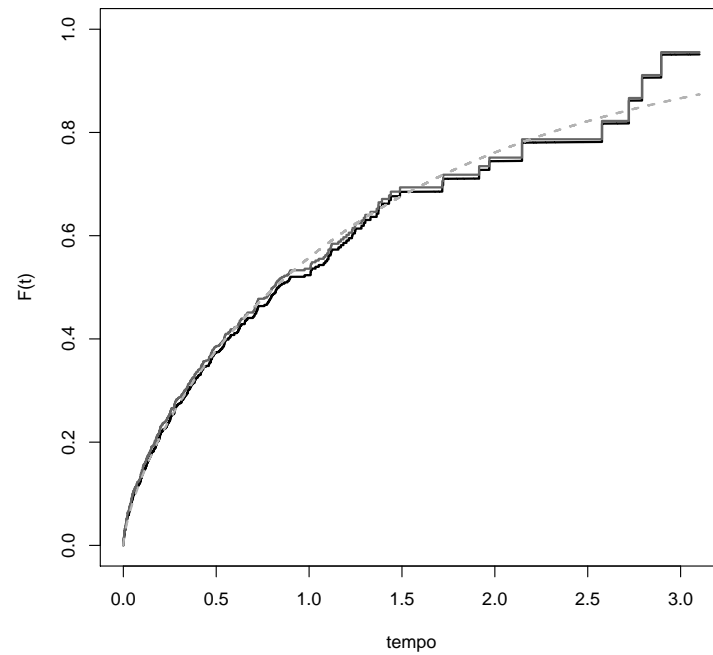
4.2.14 Figura. *Função distribuição do sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está muito próxima da função de distribuição geradora dos dados.

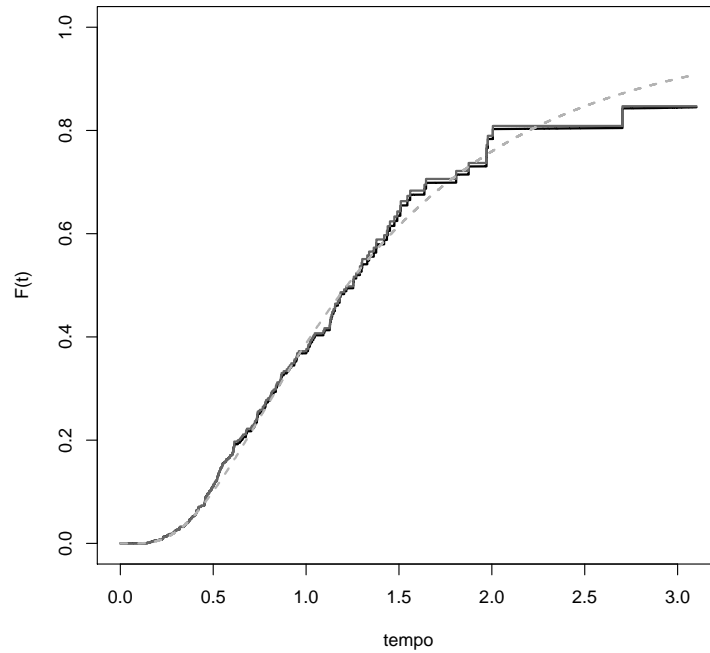
Agora, utilizando a Equação (4.2.4), obtemos a estimativa da função distribuição do j -ésimo componente, dado nas figuras 4.2.15-4.2.17.



4.2.15 Figura. *Função distribuição do componente 1 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.2.16 Figura. *Função distribuição do componente 2 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*



4.2.17 Figura. *Função distribuição do componente 3 no sistema em série (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Como podemos observar pelas figuras, as estimativas ficaram próximas das verdadeiras funções de distribuição dos dados, sendo pior conforme se afasta da origem.

4.3 Sistema em Paralelo

Sejam m ($\leq n$) estatísticas de ordem distintas de T , $T_{(1)} < \dots < T_{(m)}$. $N_i = \sum_{j=1}^n I(T_j < T_{(i)})$, e $D_{\Delta i} = \sum_{j=1}^n I(T_j = T_{(i)}, \delta_j \in \Delta)$, $i = 1, \dots, m$. Definimos

$$(4.3.1) \quad I_{\Delta}(t) = \exp \left[\frac{-1}{n + c_r} \sum_{j \in \Delta} \int_t^{\infty} \frac{d\alpha_j(0, s]}{\hat{F}(s)} \right]$$

e

$$(4.3.2) \quad \Pi_{\Delta}(t) = \prod_{i: T_{(i)} > t} \frac{\sum_{j=1}^r \alpha_j(0, T_{(i)}] + N_i}{\sum_{j=1}^r \alpha_j(0, T_{(i)}] + N_i + D_{\Delta i}}$$

4.3.3 Teorema. *Suponha que $\alpha_1(0, \cdot), \dots, \alpha_r(0, \cdot)$ são contínuos em (t, ∞) , para cada*

$t > 0$, e F_Δ e F_{Δ^c} não têm pontos de descontinuidade em comum. Então, para $t < T_{(m)}$,

$$(4.3.4) \quad \hat{F}_\Delta(t) = \Phi_p(\hat{F}_\Delta^*, \hat{F}_{\Delta^c}^*, t) = I_\Delta(t)\Pi_{\Delta^c}(t),$$

é o estimador Bayesiano não paramétrico de $F_\Delta(t)$ baseado na média posteriori.

Demonstração. A demonstração é análoga a do Teorema 3.3.3, substituindo F_1 por F_Δ e F_2 por F_{Δ^c} , e fazendo as modificações necessárias. \square

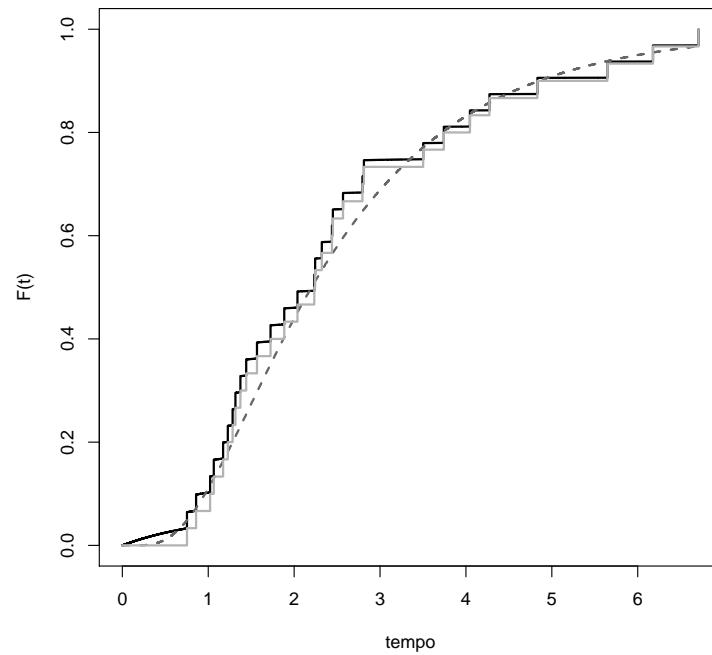
4.3.5 Exemplo. Temos um sistema com três componentes ligados em paralelo, X_1 tem distribuição exponencial com média 1.4 e variância 0.5, X_2 tem distribuição gama com média 1.4 e variância 3.2 e X_3 tem distribuição log-normal com média 1.5 e variância 1.3.

Simulamos observações do sistema utilizando três tamanhos de amostras, $n = 1000$, $n = 100$ e $n = 30$ observações do tempo de vida do sistema. Como o sistema está ligado em paralelo, o último componente a falhar é o responsável pela falha do sistema, e os outros são censurados. Com isso obtemos o seguinte resultado para cada tamanho de amostra:

Para $n = 30$

- O componente 1 foi responsável pela falha do sistema por 11 vezes.
- O componente 2 foi responsável pela falha do sistema por 4 vezes.
- O componente 3 foi responsável pela falha do sistema por 15 vezes.

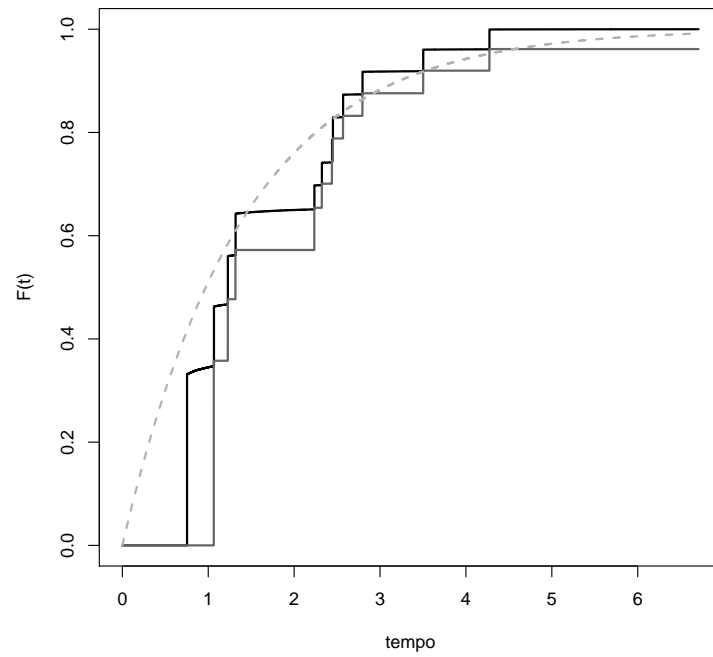
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) encontramos a função distribuição do sistema, dado na Figura 4.3.6. Novamente, para solução da integral foi utilizado um método numérico baseado na regra de Simpson.



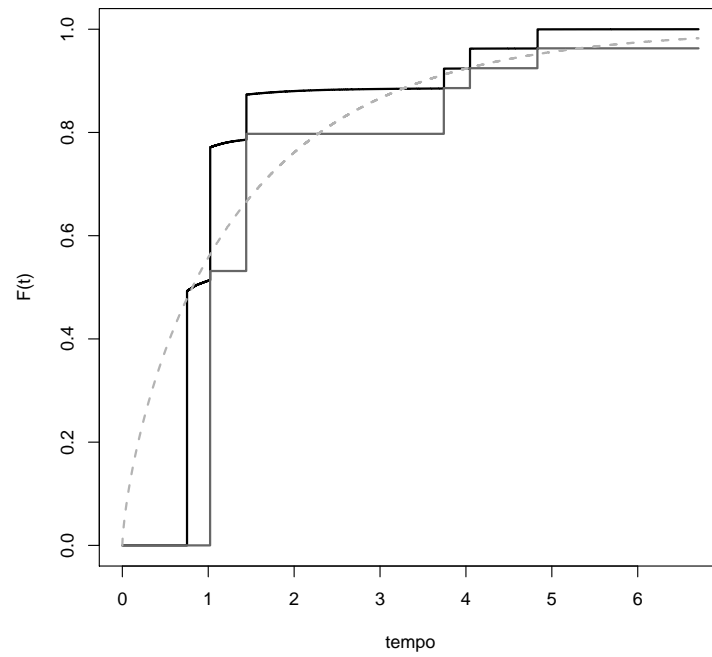
4.3.6 Figura. *Função distribuição do sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está próxima da verdadeira função de distribuição dos dados.

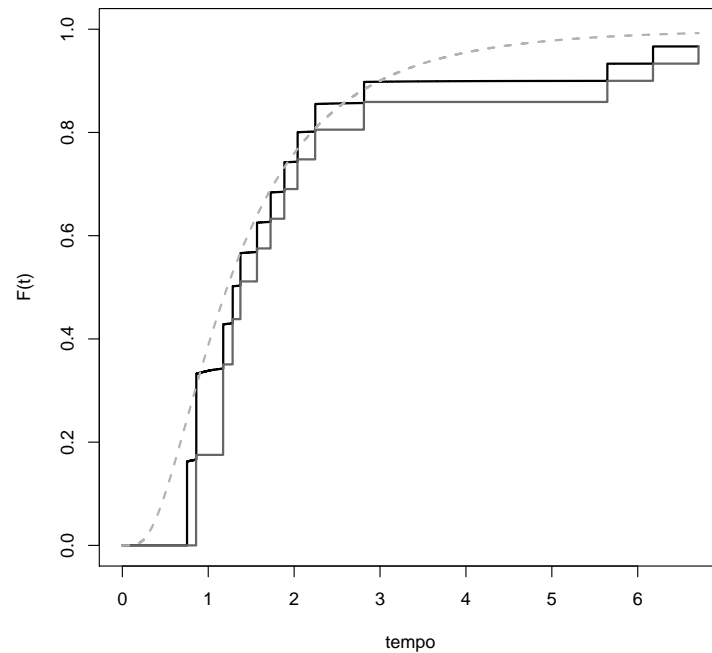
Agora, utilizando a Equação (4.3.4), obtemos a função distribuição do estimador do j -ésimo componente, dado nas figuras 4.3.7-4.3.9.



4.3.7 Figura. *Função distribuição do componente 1 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.3.8 Figura. *Função distribuição do componente 2 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*

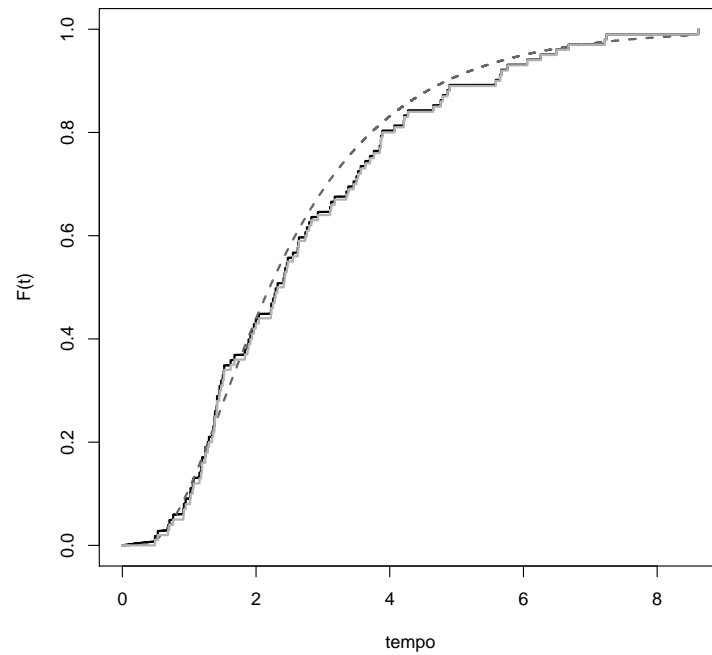


4.3.9 Figura. *Função distribuição do componente 3 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Para $n = 100$

- O componente 1 foi responsável pela falha do sistema por 27 vezes.
- O componente 2 foi responsável pela falha do sistema por 27 vezes.
- O componente 3 foi responsável pela falha do sistema por 46 vezes.

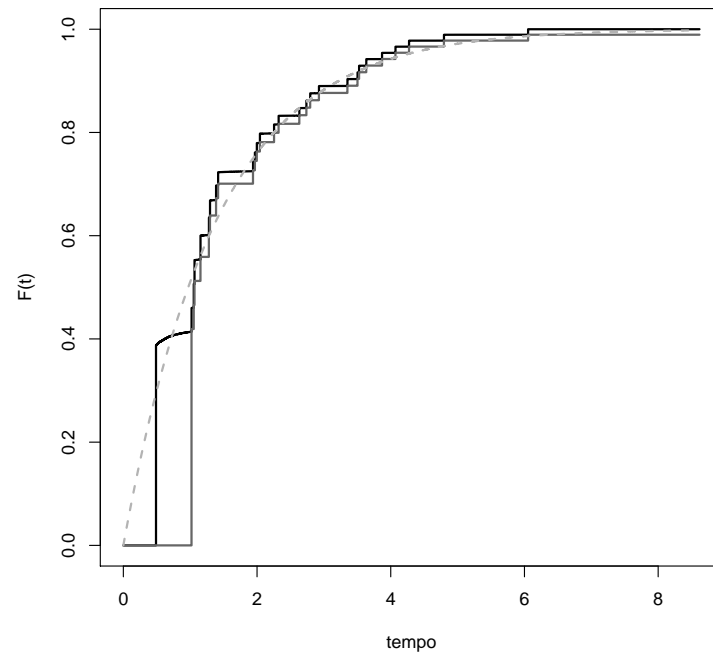
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) encontramos a função distribuição do sistema, dado na Figura 4.3.10. Novamente, para solução da integral foi utilizado um método numérico baseado na regra de Simpson.



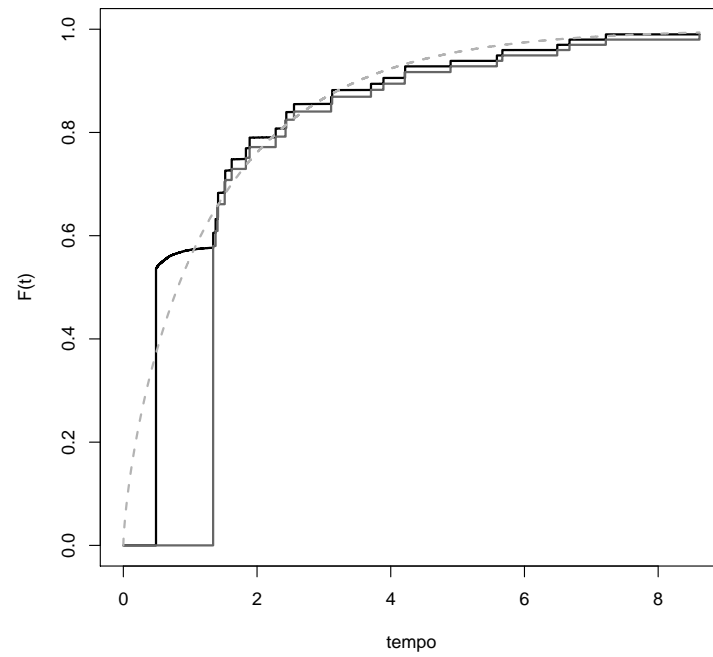
4.3.10 Figura. *Função distribuição do sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está próxima da verdadeira função de distribuição dos dados.

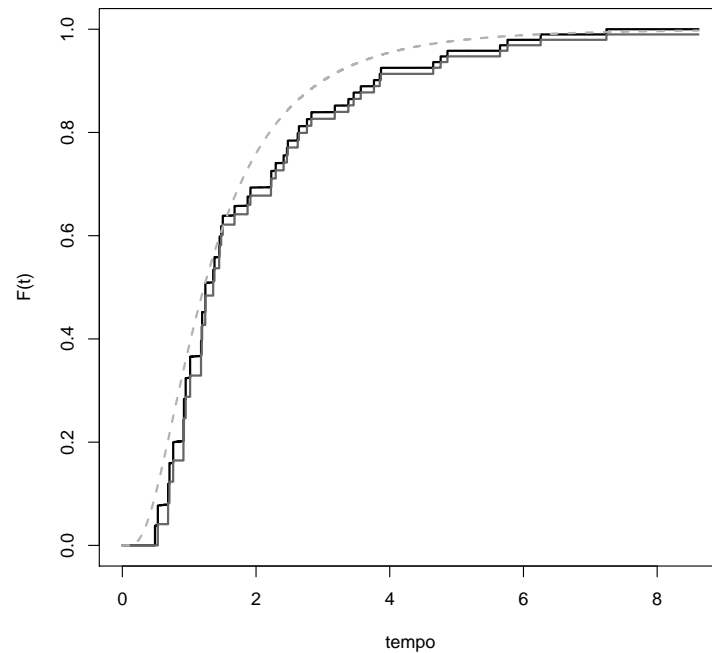
Agora, utilizando a Equação (4.3.4), obtemos a função distribuição do estimador do j -ésimo componente, dado nas figuras 4.3.11-4.3.13.



4.3.11 Figura. *Função distribuição do componente 1 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.3.12 Figura. *Função distribuição do componente 2 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*

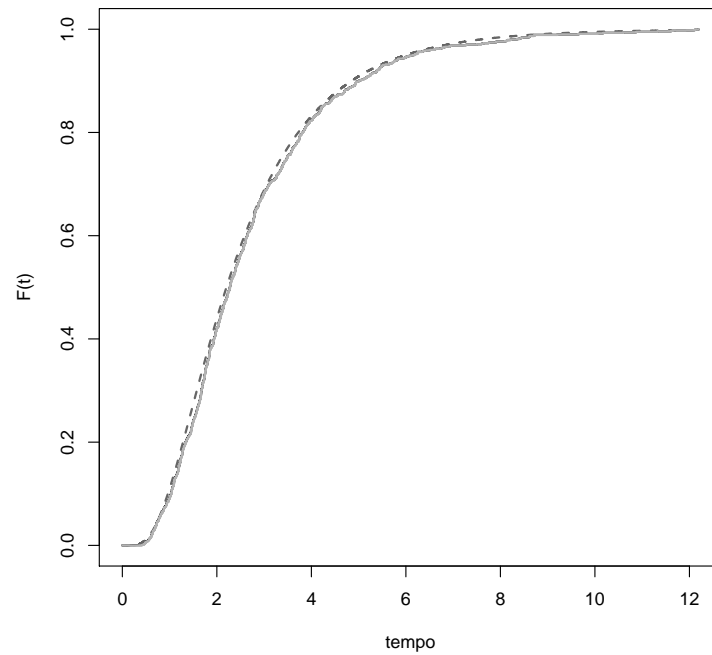


4.3.13 Figura. *Função distribuição do componente 3 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Para $n = 1000$

- O componente 1 foi responsável pela falha do sistema por 201 vezes.
- O componente 2 foi responsável pela falha do sistema por 232 vezes.
- O componente 3 foi responsável pela falha do sistema por 567 vezes.

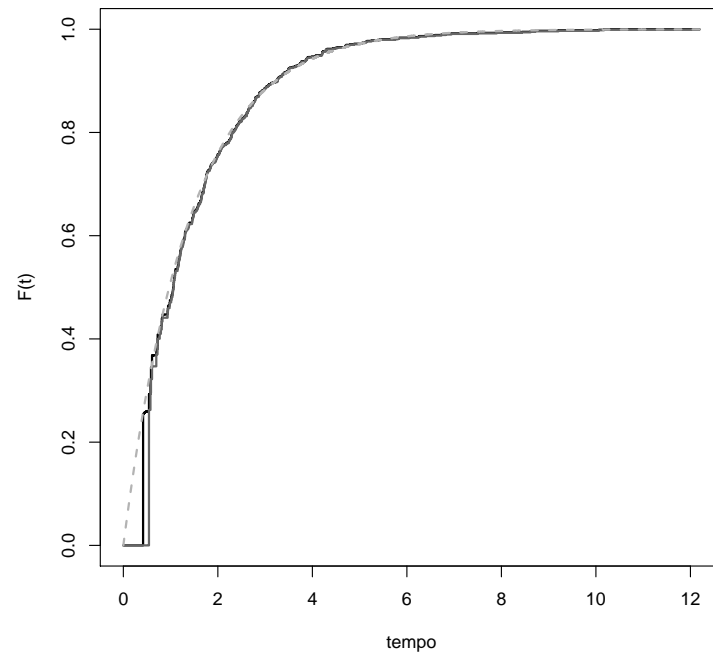
Cada um desses componentes tem como medida α a priori uma distribuição exponencial com média 1. Utilizando a Equação (4.1.7) encontramos a função distribuição do sistema, dado na Figura 4.3.14. Novamente, para solução da integral foi utilizado um método numérico baseado na regra de Simpson.



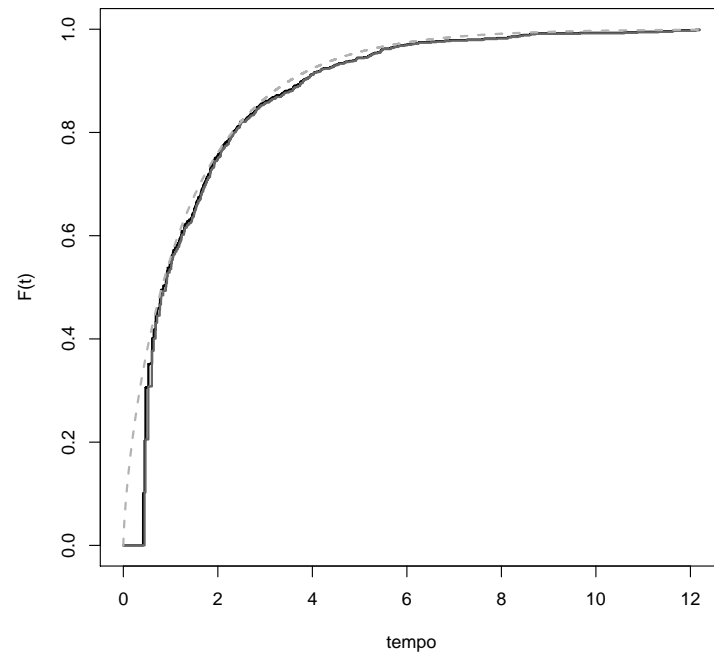
4.3.14 Figura. *Função distribuição do sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do sistema.)*

Observando o gráfico, percebemos que a função distribuição estimada está próxima da verdadeira função de distribuição dos dados.

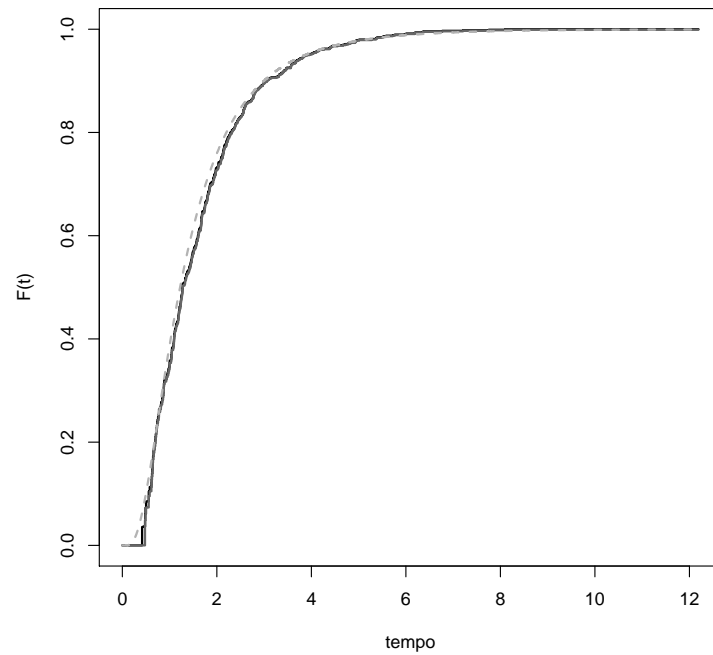
Agora, utilizando a Equação (4.3.4), obtemos a função distribuição do estimador do j -ésimo componente, dado nas figuras 4.3.15-4.3.17.



4.3.15 Figura. *Função distribuição do componente 1 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 1.)*



4.3.16 Figura. *Função distribuição do componente 2 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 2.)*



4.3.17 Figura. *Função distribuição do componente 3 no sistema em paralelo (curva contínua escura: estimativa Bayesiana não-paramétrica; curva contínua clara: estimativa de Kaplan-Meier; curva tracejada: verdadeira distribuição do componente 3.)*

Novamente, as estimativas ficaram muito próximas das verdadeiras curvas, exceto próximo da origem.

5 *Estudo de Simulação*

Neste capítulo apresentamos um estudo comparativo do estimador Bayesiano não paramétrico com o estimador de Kaplan-Meier. Utilizamos as estatísticas Norma- L_2 , Máximo Erro Absoluto (MEA) e Erro Quadrático Médio (EQM) para comparar os estimadores.

A medida Norma- L_2 é definida como

$$(5.0.1) \quad \|F_1 - A\|_2 = \left\{ \int_0^{T^*(m)} (F_1(t) - A(t))^2 dt \right\}^{1/2},$$

em que F_1 é a verdadeira distribuição, e $A(t)$ é um dos dois possíveis estimadores: $\widehat{F}_1(t)$ para o estimador Bayesiano não paramétrico e $KM(t)$ para o estimador de Kaplan-Meier.

O Erro Máximo Absoluto (EMA) é definido como

$$(5.0.2) \quad EMA = \max_t |F_1(t) - A(t)|$$

em que F_1 é a verdadeira distribuição, e $A(t)$ é um dos dois possíveis estimadores: $\widehat{F}_1(t)$ para o estimador Bayesiano não paramétrico e $KM(t)$ para o estimador de Kaplan-Meier.

O Erro Quadrático Médio (EQM) é definido como

$$(5.0.3) \quad EQM = \frac{\sum_{i=1}^n (F_1(t_i) - A(T_i))^2}{n}$$

em que F_1 é a verdadeira distribuição, e $A(t)$ é um dos dois possíveis estimadores: $\widehat{F}_1(t)$ para o estimador Bayesiano não paramétrico e $KM(t)$ para o estimador de Kaplan-Meier.

5.1 Sistema com Componentes em Série

O estimador de Kaplan-Meier pode ser obtido aplicando o estimador empírico das sub-distribuições no funcional Φ_s (cf. Peterson [8]), obtendo

$$(5.1.1) \quad \begin{aligned} \hat{F}_1(t) &= \Phi_s(\hat{F}_{1n}^*, \hat{F}_{2n}^*, t) \\ &= 1 - \exp \left[\sum_{s \leq t} \ln \left(\frac{1 - (\hat{F}_{1n}^*(s^+) + \hat{F}_{2n}^*(s^+))}{1 - (\hat{F}_{1n}^*(s^-) + \hat{F}_{2n}^*(s^-))} \right) \right]. \end{aligned}$$

Note que, a parte integral da Equação (2.3.11) é igual a zero, uma vez que $d\hat{F}_{1n}^* = 0$

Simulamos 1000 amostras aleatórias de tamanho n de um sistema com três componentes ligados em série. Especificamos a distribuição de X_1 como uma distribuição exponencial com média 1.4 e variância 0.5, X_2 como uma distribuição gama com média 1.4 e variância 3.2 e X_3 como uma distribuição Log-Normal com média 1.5 e variância 1.3. Consideramos três diferentes valores para n , $\{30, 100, 1000\}$. Assim temos 9 diferentes cenários no nosso estudo de simulação, 3 para cada componente.

As Tabelas 5.1.2-5.1.4 trazem a comparação dos estimadores de Kaplan-Meier e o Bayesiano não paramétrico para cada componente, com 30, 100 e 1000 repetições.

Ao observarmos as tabelas, vemos que o desempenho de ambos os estimadores são similares para o Erro Quadrático Médio (EQM) e Norma- L_2 . No caso do Erro Máximo Absoluto (EMA), o estimador Bayesiano apresentou melhor desempenho (menores valores de EMA).

5.2 Sistema com Componentes em Paralelo

O estimador de Kaplan-Meier pode ser obtido aplicando o estimador empírico das sub-distribuições no funcional Φ_p (cf. Peterson [8]), obtendo

$$(5.2.1) \quad \begin{aligned} \hat{F}_1(t) &= \Phi_p(\hat{F}_{1n}^*, \hat{F}_{2n}^*, t) \\ &= \exp \left[\sum_{s > t} \ln \left(\frac{\hat{F}_{1n}^*(s^-) + \hat{F}_{2n}^*(s^-)}{\hat{F}_{1n}^*(s^+) + \hat{F}_{2n}^*(s^+)} \right) \right]. \end{aligned}$$

Note que, a parte integral da Equação (2.3.12) é igual a zero, uma vez que $d\hat{F}_{1n}^* = 0$

Simulamos 1000 amostras aleatórias de tamanho n de um sistema com três componentes ligados em paralelo. Especificamos a distribuição de X_1 como uma distribuição exponencial com média 1.4 e variância 0.5, X_2 como uma distribuição gama com média 1.4

5.1.2 Tabela. $n = 1000$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.07540	0.02861	0.02686	0.05486	0.06904	0.08881	0.20570
KM	0.08536	0.02894	0.02994	0.06502	0.08147	0.10220	0.19880
EQM							
\widehat{F}_1	0.00030	0.00023	0.00004	0.00015	0.00023	0.00037	0.00221
KM	0.00034	0.00026	0.00004	0.00017	0.00025	0.00043	0.00249
<i>Norma - L₂</i>							
\widehat{F}_1	0.00427	0.00160	0.00165	0.00311	0.00393	0.00506	0.01332
KM	0.00452	0.00169	0.00143	0.00329	0.00415	0.00542	0.01197
Componente 2							
EMA							
\widehat{F}_1	0.07139	0.02750	0.02344	0.05220	0.06612	0.08513	0.21014
KM	0.08748	0.03220	0.02920	0.06191	0.08230	0.10707	0.21452
EQM							
\widehat{F}_1	0.00029	0.00024	0.00003	0.00014	0.00022	0.00366	0.00217
KM	0.00037	0.00032	0.00004	0.00017	0.00027	0.00045	0.00307
<i>Norma - L₂</i>							
\widehat{F}_1	0.00421	0.00164	0.00139	0.00299	0.00388	0.00507	0.01358
KM	0.00470	0.00192	0.00155	0.00334	0.00427	0.00556	0.01497
Componente 3							
EMA							
\widehat{F}_1	0.07896	0.02658	0.02863	0.06014	0.07446	0.09263	0.19590
KM	0.08620	0.02616	0.03169	0.06703	0.08228	0.10146	0.20171
EQM							
\widehat{F}_1	0.00029	0.00023	0.00003	0.00014	0.00022	0.00034	0.00233
KM	0.00030	0.00023	0.00004	0.00016	0.00024	0.00037	0.00212
<i>Norma - L₂</i>							
\widehat{F}_1	0.00420	0.00156	0.00141	0.00310	0.00385	0.00505	0.01310
KM	0.00432	0.00153	0.00159	0.00321	0.00401	0.00508	0.01249

e variância 3.2 e X_3 como uma distribuição Log-Normal com média 1.5 e variância 1.3. Consideramos três diferentes valores para n , $\{30, 100, 1000\}$. Assim temos 9 diferentes cenários no nosso estudo de simulação, 3 para cada componente.

As Tabelas 5.2.2-5.2.4 trazem a comparação dos estimadores de Kaplan-Meier e o Bayesiano não paramétrico para cada componente, com 30, 100 e 1000 repetições.

Semelhante ao exemplo de simulação do sistema em série, observamos pelas tabelas que o desempenho de ambos os estimadores são similares para o Erro Quadrático Médio (EQM) e Norma- L_2 , e o Bayesiano é melhor no caso do Erro Máximo Absoluto (EMA). Entretanto, neste caso, aparentemente, o comportamento de ambos os estimadores é insatisfatório uma vez que as medidas de erros são muito altas, indicando que o estimador está longe da verdadeira distribuição.

5.1.3 Tabela. $n = 100$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.14325	0.05696	0.04899	0.10284	0.13223	0.17203	0.44796
KM	0.17609	0.06497	0.05379	0.12850	0.16433	0.21405	0.45057
EQM							
\widehat{F}_1	0.00274	0.00227	0.00031	0.00120	0.00200	0.00352	0.02009
KM	0.00307	0.00229	0.00042	0.00152	0.00238	0.00377	0.01812
<i>Norma - L₂</i>							
\widehat{F}_1	0.00333	0.00141	0.00102	0.00231	0.00305	0.00403	0.01169
KM	0.00352	0.00134	0.00126	0.00254	0.00323	0.00416	0.01137
Componente 2							
EMA							
\widehat{F}_1	0.13069	0.04975	0.04951	0.09425	0.12033	0.15707	0.33420
KM	0.16609	0.05960	0.04561	0.11988	0.15653	0.20503	0.39027
EQM							
\widehat{F}_1	0.00275	0.00267	0.00036	0.00118	0.00190	0.00332	0.02489
KM	0.00311	0.00273	0.00035	0.00143	0.00228	0.00380	0.02356
<i>Norma - L₂</i>							
\widehat{F}_1	0.00329	0.00146	0.00119	0.00229	0.00296	0.00396	0.01040
KM	0.00350	0.00140	0.00111	0.00248	0.00323	0.00414	0.01063
Componente 3							
EMA							
\widehat{F}_1	0.17441	0.06727	0.05923	0.12238	0.16359	0.20677	0.49859
KM	0.20480	0.07117	0.06234	0.15347	0.19592	0.24851	0.52000
EQM							
\widehat{F}_1	0.00262	0.00208	0.00033	0.00117	0.00199	0.00341	0.01523
KM	0.00303	0.00217	0.00042	0.00154	0.00240	0.00389	0.01791
<i>Norma - L₂</i>							
\widehat{F}_1	0.00327	0.00138	0.00122	0.00226	0.00296	0.00394	0.01038
KM	0.00349	0.00130	0.00112	0.00254	0.00322	0.00416	0.00988

5.1.4 Tabela. $n = 30$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.19049	0.07522	0.04788	0.13584	0.17542	0.23414	0.51635
KM	0.25166	0.09289	0.07065	0.18005	0.23621	0.31583	0.63349
EQM							
\widehat{F}_1	0.00789	0.00726	0.00051	0.00315	0.00556	0.01011	0.06268
KM	0.01030	0.00797	0.00074	0.00523	0.00813	0.01266	0.07158
<i>Norma - L₂</i>							
\widehat{F}_1	0.00266	0.00126	0.00056	0.00176	0.00234	0.00331	0.00897
KM	0.00301	0.00117	0.00079	0.00220	0.00280	0.00361	0.00954
Componente 2							
EMA							
\widehat{F}_1	0.17220	0.06496	0.05470	0.12420	0.16090	0.20800	0.45050
KM	0.22660	0.08304	0.05887	0.16246	0.21678	0.27859	0.58793
EQM							
\widehat{F}_1	0.00756	0.00686	0.00074	0.00305	0.00531	0.00948	0.05493
KM	0.00967	0.00730	0.00079	0.00477	0.00763	0.01227	0.05637
<i>Norma - L₂</i>							
\widehat{F}_1	0.00262	0.00126	0.00074	0.00169	0.00234	0.00320	0.00803
KM	0.00294	0.00118	0.00076	0.00212	0.00272	0.00356	0.00929
Componente 3							
EMA							
\widehat{F}_1	0.23026	0.09602	0.05308	0.15828	0.21144	0.28387	0.63375
KM	0.30460	0.11707	0.08250	0.21550	0.28590	0.37740	0.75640
EQM							
\widehat{F}_1	0.00691	0.00599	0.00060	0.00298	0.00517	0.00860	0.04768
KM	0.01027	0.00746	0.00097	0.00523	0.00838	0.01307	0.06624
<i>Norma - L₂</i>							
\widehat{F}_1	0.00246	0.00112	0.00063	0.00164	0.00221	0.00299	0.00796
KM	0.00295	0.00104	0.00098	0.00221	0.00279	0.00345	0.00928

5.2.2 Tabela. $n = 1000$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.17476	0.03477	0.07305	0.15133	0.17595	0.19954	0.27318
KM	0.23586	0.04950	0.08884	0.20212	0.23699	0.26822	0.38152
EQM							
\widehat{F}_1	0.00033	0.00022	0.00007	0.00018	0.00026	0.00041	0.00170
KM	0.00061	0.00039	0.00011	0.00033	0.00050	0.00075	0.00281
<i>Norma - L₂</i>							
\widehat{F}_1	0.00905	0.00315	0.00376	0.00681	0.00843	0.01056	0.02484
KM	0.01248	0.00413	0.00493	0.00941	0.01185	0.01464	0.02974
Componente 2							
EMA							
\widehat{F}_1	0.25400	0.03775	0.13800	0.22990	0.25700	0.28120	0.35230
KM	0.33160	0.04991	0.17550	0.29750	0.33650	0.36850	0.45270
EQM							
\widehat{F}_1	0.00038	0.00025	0.00007	0.00022	0.00030	0.00044	0.00313
KM	0.00110	0.00077	0.00020	0.00057	0.00086	0.00139	0.00652
<i>Norma - L₂</i>							
\widehat{F}_1	0.00944	0.00317	0.00347	0.00718	0.00877	0.01094	0.02923
KM	0.01661	0.00578	0.00687	0.01229	0.01532	0.01986	0.04923
Componente 3							
EMA							
\widehat{F}_1	0.04697	0.01354	0.02030	0.03710	0.04520	0.05468	0.09726
KM	0.04864	0.01395	0.02122	0.03854	0.04722	0.05621	0.10160
EQM							
\widehat{F}_1	0.00025	0.00019	0.00003	0.00012	0.00032	0.00161	0.00019
KM	0.00026	0.00019	0.00003	0.00012	0.00036	0.00149	0.00026
<i>Norma - L₂</i>							
\widehat{F}_1	0.00798	0.00291	0.00259	0.00579	0.00961	0.02168	0.00749
KM	0.00811	0.00297	0.00254	0.00591	0.00975	0.02088	0.00756

5.2.3 Tabela. $n = 100$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.27386	0.06091	0.09768	0.23224	0.27312	0.31435	0.45160
KM	0.37530	0.00713	0.15330	0.31130	0.37630	0.43580	0.61180
EQM							
\widehat{F}_1	0.00364	0.00232	0.00065	0.00206	0.00296	0.00442	0.01972
KM	0.00928	0.00713	0.00120	0.00458	0.00732	0.01128	0.05508
<i>Norma - L₂</i>							
\widehat{F}_1	0.00760	0.00263	0.00270	0.00571	0.00701	0.00893	0.01958
KM	0.01241	0.00480	0.00416	0.00416	0.01150	0.03512	0.03512
Componente 2							
EMA							
\widehat{F}_1	0.34970	0.05892	0.16790	0.31050	0.35140	0.38960	0.50870
KM	0.46560	0.08406	0.21200	0.40850	0.47260	0.52300	0.70800
EQM							
\widehat{F}_1	0.00415	0.00255	0.00105	0.00248	0.00337	0.00502	0.02461
KM	0.01540	0.01238	0.00197	0.00721	0.01174	0.01919	0.11120
<i>Norma - L₂</i>							
\widehat{F}_1	0.00793	0.00273	0.00350	0.00592	0.00736	0.00934	0.02350
KM	0.01594	0.00653	0.00579	0.01111	0.01463	0.01913	0.04955
Componente 3							
EMA							
\widehat{F}_1	0.13856	0.04133	0.04823	0.10791	0.13432	0.16494	0.29701
KM	0.15443	0.05111	0.05328	0.11856	0.14709	0.18226	0.39016
EQM							
\widehat{F}_1	0.00250	0.00178	0.00029	0.00128	0.00192	0.00330	0.01287
KM	0.00280	0.00200	0.00043	0.00139	0.00217	0.00363	0.01528
<i>Norma - L₂</i>							
\widehat{F}_1	0.00659	0.00239	0.00216	0.00488	0.00607	0.00789	0.01673
KM	0.00697	0.00253	0.00240	0.00508	0.00641	0.00843	0.01672

5.2.4 Tabela. $n = 30$.

	Média	DP	Min	P25	Mediana	P75	Max
Componente 1							
EMA							
\widehat{F}_1	0.35510	0.08110	0.14040	0.29740	0.35560	0.41210	0.62320
KM	0.49090	0.12212	0.20610	0.40310	0.48550	0.56970	0.85490
EQM							
\widehat{F}_1	0.01302	0.00771	0.00227	0.00787	0.01082	0.01590	0.06883
KM	0.04111	0.03591	0.00361	0.01972	0.03071	0.04764	0.28294
<i>Norma - L_2</i>							
\widehat{F}_1	0.00674	0.00231	0.00227	0.00504	0.00637	0.00801	0.01850
KM	0.01243	0.00533	0.00350	0.00885	0.01139	0.01422	0.04236
Componente 2							
EMA							
\widehat{F}_1	0.41910	0.07653	0.20350	0.36750	0.41980	0.47380	0.64930
KM	0.56680	0.11351	0.23630	0.48660	0.56630	0.64460	0.90510
EQM							
\widehat{F}_1	0.01375	0.00710	0.00311	0.00924	0.01207	0.01615	0.07708
KM	0.06103	0.05239	0.00762	0.02728	0.04357	0.07469	0.44342
<i>Norma - L_2</i>							
\widehat{F}_1	0.00678	0.00222	0.00316	0.00527	0.00631	0.00780	0.02128
KM	0.01516	0.00667	0.00466	0.01028	0.01366	0.01849	0.04878
Componente 3							
EMA							
\widehat{F}_1	0.23425	0.07944	0.07122	0.17616	0.22690	0.28354	0.56363
KM	0.29234	0.11789	0.06403	0.20727	0.26851	0.36088	0.78174
EQM							
\widehat{F}_1	0.00933	0.00649	0.00109	0.00474	0.00745	0.01204	0.04807
KM	0.01465	0.01352	0.00094	0.00631	0.01123	0.01825	0.15182
<i>Norma - L_2</i>							
\widehat{F}_1	0.00599	0.00225	0.00195	0.00431	0.00553	0.00729	0.01646
KM	0.00747	0.00328	0.00193	0.00511	0.00689	0.00907	0.02940

6 *Comentários Finais*

Neste trabalho apresentamos as principais características do problema de confiabilidade de sistemas em série e paralelo com dois componentes. Também apresentamos o estimador Bayesiano não paramétrico, além da generalização de sistemas com n , $n > 2$ componentes e seguindo passos similares a Polpo e Pereira [10], generalizamos os estimadores frequentistas para o caso com mais de dois componentes. Além disso, foram apresentados dois exemplos simples, com 3 componentes para ilustrar os estimadores, tanto para o caso dos sistemas em série quanto em paralelo.

Desenvolvemos um estimador frequentista para o sistema em paralelo, seguindo as ideias apresentadas por Peterson [8] na derivação do estimador de Kaplan-Meier como função das sub-distribuições.

Para Verificar a qualidade dos estimadores, realizamos um estudo de simulação com diferentes tamanhos de amostras, comparando os estimadores frequentistas (e de Kaplan-Meier [5]) e Bayesianos para os sistemas em série e em paralelo.

Nos exemplos de simulação, notamos que, aumentando o tamanho da amostra diminuía-se as medidas de erro dos estimadores, evidenciando a consistência dos mesmos. Além disso, o estimador Bayesiano foi superior ao frequentista. Entretanto, para o sistema em paralelo ambas as abordagens (frequentista e Bayesiana) ficaram a quem do que nós consideramos um bom estimador. Observando os gráficos do exemplo 4.3.5 notamos que o maior erro dá-se para os menores tempos observados. Supomos que isto ocorreu devido a natureza da censura à esquerda, em que geralmente não observa-se falhas do sistema em tempos próximos de zero.

Possíveis sequências deste trabalho poderiam abordar comparações na variância dos estimadores, estimativa do tempo mediano de falha e considerar outros estimadores na comparação.

APÊNDICE A – Código Computacional

```
#####
dist.emp <- function(x) {
  aux.x <- sort(x)
  aux.y <- cumsum(table(aux.x)/length(aux.x))
  row.names(aux.y) <- NULL
  out <- stepfun(unique(aux.x),c(0,aux.y))
  return(out)
}

#####
subdist.emp <- function(x,d=1) {
  aux.x <- sort(x[x[,2]==d,1])
  aux.y <- cumsum(table(aux.x)/length(x[,1]))
  row.names(aux.y) <- NULL
  out <- stepfun(unique(aux.x),c(0,aux.y))
  return(out)
}

#####
alpha <- function(x,d=1) {
  if (d == 1) {
    out <- 1-exp(-x)
  } else if (d == 2) {
    out <- 1-exp(-x)
  }
  return(out)
}

```

```
#####
d.alpha <- function(x,d=1) {
  if (d == 1) {
    out <- exp(-x)
  } else if (d == 2) {
    out <- exp(-x)
  }
  return(out)
}

#####
est.subdist <- function(t,x,d=1) {
  c.2 <- alpha(Inf,1)+alpha(Inf,2)
  n <- length(x[,1])
  p.n <- c.2/(c.2+n)
  aux <- subdist.emp(x,d)

  out <- p.n*(alpha(t,1)/c.2)+(1-p.n)*aux(t)
  return(out)
}

#####
est.distsys <- function(t,x) {
  out <- est.subdist(t,x,1)+est.subdist(t,x,2)
  return(out)
}

#####
est.distcomp.serie <- function(x,d=1,h=(length(x[,1])*100),
                              x.max=(max(x[,1])+1)) {
  # integral part
  integral <- function(x,d,h,x.max,h2=3) {
    fn.aux <- function(s,const,d) {
      out <- const*d.alpha(s,d)/(1-est.distsys(s,x))
    }
  }
}
```

```

    return(out)
  }
  c.2 <- alpha(Inf,1)+alpha(Inf,2)
  n <- length(x[,1])
  h.aux <- x.max/(h*h2-1)
  t.aux <- seq(0,x.max,h.aux)
  aux <- fn.aux(t.aux,const=-1/(c.2+n),d)

  out <- stepfun(t.aux[-1],exp(c(0,
                                cumsum(h.aux*(aux[-length(aux)]+aux[-1])/2))))
  return(out)
}
# product part
product <- function(x,d) {
  aux.x <- sort(unique(x[,1]))
  ni <- rep(NA,length(aux.x))
  for (i in 1:length(aux.x))
    ni[i] <- length(x[x[,1] >= aux.x[i],1])
  di <- rep(NA,length(aux.x))
  for (i in 1:length(aux.x))
    di[i] <- length(x[(x[,1] == aux.x[i]) & (x[,2] == d),1])

  c.2 <- alpha(Inf,1)+alpha(Inf,2)
  aux <- cumprod((c.2-alpha(aux.x,1)-alpha(aux.x,2)+ni-di)/
                (c.2-alpha(aux.x,1)-alpha(aux.x,2)+ni))
  out <- stepfun(aux.x,c(1,aux))
  return(out)
}

aux.int <- integral(x,d,h,x.max,3)
aux.prod <- product(x,d)

h.aux <- x.max/(h-1)
t.aux <- seq(0,x.max,h.aux)
aux <- 1-aux.int(t.aux)*aux.prod(t.aux)

```

```

out  <- stepfun(t.aux,c(0,aux))
return(out)
}

#####
est.distcomp.parallel <- function(x,d=1,h=(length(x[,1])*100),
                                x.max=(max(x[,1])+1)) {
# integral part
integral <- function(x,d,h,x.max,h2=3) {
  fn.aux <- function(s,const,d) {
    out <- const*d.alpha(s,d)/est.distsys(s,x)
    return(out)
  }
  fn.aux2 <- function(s,a=t.aux[length(t.aux)],const,d) {
    out <- fn.aux(a+s/(1-s),const,d)/(1-s^2)
    return(out)
  }
  c.2 <- alpha(Inf,1)+alpha(Inf,2)
  n <- length(x[,1])
  h.aux <- x.max/(h*h2-1)
  t.aux <- seq(0,x.max,h.aux)
  aux <- fn.aux(t.aux,const=-1/(c.2+n),d)

  aux2 <- fn.aux2(seq(0,0.99,0.01),const=-1/(c.2+n),d=d)
  int.ainf <- sum(0.01*(aux2[-length(aux2)]+aux2[-1])/2)

  aux3 <- c(h.aux*(aux[-length(aux)]+aux[-1])/2,int.ainf,0)
  aux3 <- cumsum(aux3[length(aux3):1])[length(aux3):1]

  out <- stepfun(t.aux,exp(aux3))
  return(out)
}
# product part
product <- function(x,d) {
  aux.x <- sort(unique(x[,1]))

```



```

ni <- rep(NA,length(aux.x))
for (i in 1:length(aux.x))
  ni[i] <- length(x[x[,1] < aux.x[i],1])
di <- rep(NA,length(aux.x))
for (i in 1:length(aux.x))
  di[i] <- length(x[(x[,1] == aux.x[i]) & (x[,2] == d),1])

aux <- (alpha(aux.x,1)-alpha(aux.x,2)+ni)/(alpha(aux.x,1)
                                             -alpha(aux.x,2)+ni+di)
aux <- cumprod(aux[length(aux):1])[length(aux):1]
out <- stepfun(aux.x,c(aux,1))
return(out)
}

aux.int <- integral(x,d,h,x.max,3)
aux.prod <- product(x,d)

h.aux <- x.max/(h-1)
t.aux <- seq(0,x.max,h.aux)
aux <- aux.int(t.aux)*aux.prod(t.aux)
out <- stepfun(t.aux,c(0,aux))
return(out)
}

#####
est.distcomp <- function(x,type="serie",d=1,h=(length(x[,1])*100),
                        x.max=(max(x[,1])+1)) {
  if (type == "serie") {
    out <- est.distcomp.serie(x,d,h,x.max)
  } else if (type == "parallel") {
    out <- est.distcomp.parallel(x,d,h,x.max)
  }
  return(out)
}
}

```

```
#####
km.serie <- function(x,d=1,dif=10^(-5)) {
  aux1 <- subdist.emp(x,d=1)
  aux2 <- subdist.emp(x,d=2)
  aux.x <- sort(unique(x[x[,2] == d,1]))

  out <- stepfun(aux.x,1-cumprod(c(1,
                                (1-(aux1(aux.x+dif)+aux2(aux.x+dif)))/
                                (1-(aux1(aux.x-dif)+aux2(aux.x-dif))))))
  return(out)
}

#####
km.parallel <- function(x,d=1,dif=10^(-5)) {
  aux1 <- subdist.emp(x,d=1)
  aux2 <- subdist.emp(x,d=2)
  aux.x <- sort(unique(x[x[,2] == d,1]))

  aux3 <- c(0,(aux1(aux.x-dif)+aux2(aux.x-dif))/
            (aux1(aux.x+dif)+aux2(aux.x+dif)))
  aux3 <- cumprod(aux3[length(aux3):1])[length(aux3):1]

  out <- stepfun(aux.x,aux3)
  return(out)
}

#####
km <- function(x,type="serie",d=1,dif=10^(-5)) {
  if (type == "serie") {
    out <- km.serie(x,d,dif)
  } else if (type == "parallel") {
    out <- km.parallel(x,d,dif)
  }
  return(out)
}

```

```
#####
gen.data <- function(n,seed=1234,type="serie") {
  set.seed(seed)
  x1 = rexp(n,1/1.4)
  x2 = rgamma(n,0.7,1/2)
  x3 = rlnorm(n,0.2,0.7)

  if (type == "serie") {
    t <- apply(cbind(x1,x2,x3),1,min)
  } else if (type == "parallel") {
    t <- apply(cbind(x1,x2,x3),1,max)
  }
  d <- ifelse(t == x1,1,ifelse(t == x2,2,3))

  out <- cbind(t,d)
  out <- out[order(out[,1]),]
  return(out)
}

#####
choice <- function(x,j="d1") {
  out <- switch(j,
    "d1" = pexp(x,1/1.4),
    "d2" = pgamma(x,0.7,1/2),
    "d3" = plnorm(x,0.2,0.7))
  return(out)
}
```

Referências

- [1] Barlow, R.E. e F. Proschan: *Statistical Theory of Reliability and Life Testing: Probability Models*. To Begin With, Silver Spring, 1981.
- [2] Breslow, N.E. e J. Crowley: *A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship*. The Annals of Statistics, 2:437–453, 1974.
- [3] Davis, P. e P. Rabinowitz: *Methods of Numerical Integration*. Academic Press, New York, 1984. 2a. ed.
- [4] Ibrahim, J. G., M H. Chen e D. Sinha: *Bayesian Survival Analysis*. Springer, 2001.
- [5] Kaplan, E.L. e P. Meier: *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, 53:457–481, 1958.
- [6] Komolgorov, A.N. e S.V. Fomin: *Introductory Real Analysis*. Dover Publications, Inc., New York, 1970. Translated and Edited by R.A. Silverman.
- [7] Lawless, J. F.: *Statistical Models and Methods for Lifetime Date*. Wiley-Interscience, 2003.
- [8] Peterson, A.V.: *Expressing the Kaplan-Meier Estimator as a Function of Empirical Subsurvival Functions*. Journal of the American Statistical Association, 72(360):854–858, 1977.
- [9] Polpo, A., M. Coque e C. A. B. Pereira: *Statistical analysis for Weibull distributions in presence of right and left censoring*. The proceedings of 8th international conference on reliability, maintainability and safety, 1:213–223, 2009.
- [10] Polpo, A. e C. A. B. Pereira: *Reliability Nonparametric Bayesian Estimation in Parallel Systems*. IEEE Transactions on Reliability, 58(2):364–373, 2009.
- [11] Polpo, A. e D. Sinha: *Correction in Bayesian nonparametric estimation in a series system or a competing-risks model*. Statistics & Probability Letters, 81(12):1756–1759, 2011.
- [12] Salinas, V., C.A.B. Pereira e R. Tiwari: *Bayesian Nonparametric Estimation in a Series System or a Competing Risks Model*. Journal of Nonparametric Statistics, 14:449–458, 2002.
- [13] Salinas, V. H., J. S. Romeo e A. Peña: *On Bayesian estimation of a survival curve: comparative study and examples*. Computational Statistics, 25:375–389, 2010.