**ANÁLISE BAYESIANA NÃO-PARAMÉTRICA
NA TEORIA DE RISCOS COMPETITIVOS**

Victor Hugo Salinas Torres

**TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM
ESTATÍSTICA**

**Área de concentração: Estatística
Orientador: Prof. Dr. Carlos Alberto de Bragança Pereira**

– São Paulo, outubro de 1992 –

Á minha mãe.

Á minha companheira e ao meu filho.

Á minha irmã e aos meus irmãos.

AGRADECIMENTOS

Ao meu orientador, Prof. Carlos A.B. Pereira, que sugeriu o tema e que durante o curso de doutorado contribuiu de maneira fundamental na minha formação.

Ao Prof. Ram C. Tiwari (U. of North Carolina at Charlotte) que revisou a versão preliminar dando valiosas sugestões para a concretização deste trabalho.

Aos Profs. José G. Leite (IME-USP), Sérgio Wechsler (IME-USP), Jorge A. Achcar (ICMSC-USP), Mauro S.F. Marques (IMECC-UNICAMP), Nelson I. Tanaka (IME-USP) e Hélio Migon (IM-UFRJ), por terem aceitado participar da Comissão Julgadora na defesa desta tese.

À colega Mônica C. Sandoval pelo auxílio na revisão dos manuscritos.

Ao colega Rinaldo Artes pela ajuda na elaboração dos gráficos.

Ao Sr. Walter V. Fernandes pelo excelente trabalho datilográfico.

No início esta tese foi financiada pela CAPES.

RESUMO

Este trabalho apresenta uma abordagem bayesiana não-paramétrica para a estimação da função de sobrevivência de um sistema em série e seus componentes, conhecido como modelo de riscos competitivos. O problema inferencial inicia-se quando observa-se o tempo de sobrevivência do sistema e identifica-se o componente falhado à época da observação. Considerando-se um processo de Dirichlet multivariado - introduzido pela primeira vez aqui - como uma classe de distribuições a priori para o vetor formado pelas funções de subsobrevivência, estimadores bayesianos para a função de sobrevivência e para a confiabilidade do sistema são obtidos, sob funções de perda quadrática e 0-1. Utilizando-se a fórmula de Peterson, estimadores não-paramétricos para as funções de sobrevivência dos componentes são propostos e suas propriedades assintóticas são estudadas. O caso particular em que o sistema possui apenas dois componentes corresponde a um modelo com observações aleatoriamente censuradas. Um exemplo numérico também é apresentado.

ABSTRACT

This work presents a nonparametric Bayesian approach to the estimation of the survival function of a series system and its components or a competing risks model. The inference problem starts when one observes the life time of the machine and identifies the component failing at the time of observation. By considering a Dirichlet multivariate process - introduced here for the first time - as a class of prior distributions for the vector of components' subsurvival functions, nonparametric Bayes estimators for the survival function and the reliability of the system are derived, under quadratic and 0-1 loss functions. By using the Peterson's formula, nonparametric estimators for the components' survival functions are proposed, and their large sample properties are studied. The special case when the system has only two components corresponds randomly censored model. Also, a numerical example is presented.

ÍNDICE

CAPÍTULO 1	Introdução	1
<i>cap 2 e 3 introdução e introdução ao processo de Dirichlet</i>		
CAPÍTULO 2	O processo de Dirichlet na inferência bayesiana não-paramétrica	7
2.1	Introdução	7
2.2	Inferência bayesiana não-paramétrica	7
2.3	O processo de Dirichlet	9
2.4	Aplicações do processo de Dirichlet a alguns problemas não-paramétricos	17
CAPÍTULO 3	Um processo de Dirichlet multivariado	25
3.1	Introdução	25
3.2	Um processo de Dirichlet <u>bivariado</u> no contexto de Ferguson	26
3.3	Um processo de Dirichlet bivariado a partir de dois processos de Dirichlet e de uma variável aleatória Beta	28
3.4	A extensão <u>ao</u> processo de Dirichlet multivariado	36
<i>cap 4 e 5 introdução e introdução ao processo de Dirichlet multivariado</i>		
CAPÍTULO 4	Estimação bayesiana não-paramétrica em modelos de riscos competitivos	41
4.1	Introdução e preliminares	41
4.2	Uma abordagem bayesiana não-paramétrica para o problema com observações aleatoriamente censuradas	46
4.3	Estimação bayesiana não-paramétrica na teoria de riscos competitivos	63

CAPÍTULO 5	Propriedades assintóticas das soluções obtidas	72
5.1	Introdução	72
5.2	Propriedades assintóticas do estimador $\hat{S}_1$	73
APÊNDICE A	A distribuição de Dirichlet	80
APÊNDICE B	Os processos empíricos na estatística não-paramétrica	86
REFERÊNCIAS		95

CAPÍTULO 1

Introdução

Gratuito a
revisão do
v3

O problema que definiremos na sequência é relacionado com problemas de estimação quando as observações podem estar aleatoriamente censuradas. Em várias áreas de aplicação de métodos estatísticos, como em engenharia e medicina, este tipo de problema aparece com bastante frequência, e é comum que os dados sejam representados pelos tempos de vida de componentes, ou pelos tempos de vida de pacientes submetidos a um tratamento médico. A análise estatística de dados incompletos tem motivado um grande número de pesquisas, conforme podemos observar nos mais diversos periódicos em Estatística e Probabilidade dos últimos anos.

Consideramos um sistema com dois componentes em série. Representamos por X o tempo de vida do primeiro componente e por Y o tempo de vida do segundo componente. Suponhamos que $X \sim F_1$ e $Y \sim F_2$, que X e Y são independentes, e que só é possível observar o tempo de vida $Z = \min\{X, Y\}$. Além disso suponhamos que também observamos a variável aleatória $\delta = \begin{cases} 1 & , \text{ se } Z = X \\ 2 & , \text{ se } Z = Y \end{cases}$.

Agora consideramos uma amostra de n sistemas nas condições anteriores. Então o estatístico só tem disponível os dados $(Z_1, \delta_1) \dots, (Z_n, \delta_n)$.

Embora o problema seja simétrico nas variáveis X e Y , é usual dizer que X é a variável de interesse e Y a variável censora. A observação Z_i é dita não-censurada se $\delta_i = 1$ e censurada se $\delta_i = 2$.

Com base na amostra anterior, o objetivo é estimar as funções de distribuição $F_1(t) = \Pr(X \leq t)$ e $F_2(t) = \Pr(Y \leq t)$, o funcional $\rho(X, Y) = \Pr(X < Y)$, como também a função de distribuição do sistema $F(t) = \Pr(Z \leq t)$. Notamos que em muitas ocasiões

é mais conveniente colocar o problema em termos das funções de sobrevivência $S_1(t) := \Pr(X > t)$, $S_2(t) := \Pr(Y > t)$ e $S(t) := \Pr(Z > t)$. Vamos supor também que todas as variáveis aleatórias envolvidas assumem valores no intervalo $(0, \infty)$.

Na terminologia da teoria de riscos competitivos, o problema é formulado como segue. Um indivíduo está exposto a duas potenciais causas de morte X e Y . Se observa (Z, δ) , onde Z representa o tempo de vida do indivíduo e δ é o *label* que representa a causa de morte. Assim, neste caso, as funções de sobrevivência condicionais $T_1^*(t) := \Pr(Z > t/\delta = 1)$ e $T_2^*(t) := \Pr(Z > t/\delta = 2)$ estão associadas a cada causa de morte do indivíduo, respectivamente, e o funcional $\rho(X, Y)$ representa a escolha aleatória de uma causa ou outra. Maiores detalhes e exemplos podem ser encontrados em Berman (1963) e Basu (1984).

Uma generalização do modelo anterior é considerar um sistema com r componentes em série nas mesmas condições anteriores, mas nesta introdução, para uma maior simplicidade na exposição, consideramos só o caso $r = 2$. Notamos que no desenvolvimento deste trabalho o problema geral é considerado com bastante detalhe.

A seguir, apresentamos algumas das abordagens utilizadas na resolução do problema de estimação, quando as observações são aleatoriamente censuradas.

Do ponto de vista *paramétrico*, suponhamos que as variáveis aleatórias X e Y são absolutamente contínuas, tais que $X \sim \exp(\lambda)$ com função de densidade $f_1(t) = \lambda e^{-\lambda t}$, $t > 0$, e a função de densidade de Y é $f_2(t)$, $t > 0$.

A função de verossimilhança é, neste caso,

$$\text{Lik} = \left(\prod_{i:\delta_i=1} f_1(z_i) \right) \left(\prod_{i:\delta_i=2} S_1(z_i) \right) \left(\prod_{i:\delta_i=2} f_2(z_i) \right) \left(\prod_{i:\delta_i=1} S_2(z_i) \right),$$

onde as z_i 's são as realizações das variáveis aleatórias Z_i 's. Mas, sob o suposto que a variável censora Y não possui relação alguma com a variável de interesse X , os últimos dois produtos da função de verossimilhança $\prod_{i:\delta_i=2} f_2(z_i)$ e $\prod_{i:\delta_i=1} S_2(z_i)$ podem ser con-

siderados como constantes no processo de maximização de Lik. Logo, escrevemos

$$\begin{aligned}\text{Lik} &= \left(\prod_{i:\delta_i=1} f_1(z_i) \right) \left(\prod_{i:\delta_i=2} S_1(z_i) \right) = \lambda^{n_1} \exp \left\{ -\lambda \left(\sum_{i:\delta_i=1} z_i + \sum_{i:\delta_i=2} z_i \right) \right\} \\ &= \lambda^{n_1} \exp \left\{ -\lambda \sum_{i=1}^n z_i \right\} .\end{aligned}$$

onde n_1 é o número de observações não-censuradas.

É fácil ver então, que o *estimador de máxima verossimilhança* de λ é

$$\hat{\lambda} = \frac{n_1}{\sum_{i=1}^n z_i} .$$

Em geral, na maximização da função de verossimilhança, métodos iterativos são requeridos. Para maiores detalhes e exemplos, ver Miller (1981) e Bolfarine et al. (1991).

Consideramos agora uma *abordagem bayesiana*. Suponhamos como distribuição a priori para o parâmetro λ a distribuição Gamma($\alpha, 1$), cuja densidade é dada por $\pi(\lambda) = \frac{\lambda^\alpha e^{-\lambda}}{\Gamma(\alpha)}$, $\lambda > 0$. Logo, a densidade a posteriori de λ dadas as observações, $\pi(\lambda/\text{Dados})$, satisfaz:

$$\pi(\lambda/\text{Dados}) \propto \frac{\lambda^{n_1+\alpha-1} \exp \{ -\lambda(1 + \sum_1^n z_i) \}}{\Gamma(\alpha)} .$$

Portanto a distribuição a posteriori é tal que $\lambda/\text{Dados} \sim \text{Gamma}(n_1 + \alpha, 1 + \sum_1^n z_i)$.

Logo, o estimador bayesiano sob função de perda quadrática é dado por

$$\hat{\lambda}_B = \frac{n_1 + \alpha}{1 + \sum_1^n z_i} = \frac{1}{1 + \sum_1^n z_i} \alpha + \frac{\sum_1^n z_i}{1 + \sum_1^n z_i} \hat{\lambda} ,$$

onde $\hat{\lambda}$ é o estimador de máxima verossimilhança de λ , e notamos que α é a média da distribuição a priori. Nesta etapa, métodos bayesianos empíricos são frequentemente utilizados. Aplicações na teoria da confiabilidade podem ser encontradas em Martz e Waller (1982).

Do ponto de vista *não-paramétrico*, o parâmetro de interesse é a função de sobrevivência $S_1(t) = \Pr(X > t)$. Definamos $S_1^*(t) := \Pr(Z > t, \delta = 1)$ e $S_2^*(t) := \Pr(Z > t, \delta = 2)$, as chamadas funções de subsobrevivência. Notamos que observar

$(Z_1, \delta_1), \dots, (Z_n, \delta_n)$ é equivalente a observar $nS_{1n}^*(t) = \sum_1^n I(Z_i > t, \delta_i = 1)$, $nS_{2n}^*(t) = \sum_1^n I(Z_i > t, \delta_i = 2)$. As funções $S_{1n}^*(t)$ e $S_{2n}^*(t)$ são ditas funções de subsobrevivência empírica.

Neste contexto, para cada t , a função de verossimilhança do problema corresponde a um modelo multinomial $(n; p_1(t), p_2(t), p_3(t))$, onde $p_1(t) = S_1^*(t)$, $p_2(t) = S_2^*(t)$ e $p_3(t) = 1 - S(t)$, sendo $S(t) = S_1^*(t) + S_2^*(t) = \Pr(Z > t)$ a função de sobrevivência do sistema. Assim, para cada t ,

$$\begin{aligned} \text{Lik} &= \Pr(nS_{1n}^*(t) = j_1(t), nS_{2n}^*(t) = j_2(t)) \\ &\propto S_1^*(t)^{j_1(t)} S_2^*(t)^{j_2(t)} (1 - S(t))^{n-j_1(t)-j_2(t)}. \end{aligned}$$

É conhecido que os estimadores de máxima verossimilhança de $S_1^*(t)$ e $S_2^*(t)$ são dados por $S_{1n}^*(t)$ e $S_{2n}^*(t)$, respectivamente.

A representação de Peterson (1977), a qual expressa a função de sobrevivência $S_1(t)$ como um funcional das funções de subsobrevivência $S_1^*(t)$ e $S_2^*(t)$, determina o estimador de máxima verossimilhança de $S_1(t)$, o qual corresponde ao estimador produto-limite de Kaplan e Meier (1958). Este estimador não-paramétrico é dado por

$$\hat{S}_1^{PL}(t) = \begin{cases} 1 & t \leq Z_{(1)}^* \\ \prod_{j=1}^k (1 - d_j/n_j) & Z_{(k)}^* \leq t < Z_{(k+1)}^*, \quad k = 1, \dots, m-1 \\ 0 & t \geq Z_{(m)}^* \end{cases}$$

onde $Z_{(1)}^* < \dots < Z_{(m)}^*$ representam as m estatísticas de ordem diferentes entre $Z_1, \dots, Z_n$, $n_j = \sum_{i=1}^n I(Z_i \geq Z_{(j)}^*)$ e $d_j = \sum_{i=1}^n I(Z_i = Z_{(j)}^*, \delta_i = 1)$, $j = 1, 2, \dots, m$, e $I(A)$ é a função indicadora do conjunto A .

Para um estudo das propriedades assintóticas do estimador produto-limite, ver por exemplo, Breslow e Crowley (1974) e Tsai e Crowley (1986).

Considerando uma abordagem *bayesiana não-paramétrica*, existem algumas contribuições onde o processo de Dirichlet é frequentemente utilizado. Os primeiros avanços foram obtidos por Susarla e Van Ryzin (1976, 1978a).

* Fazemos a convenção que $Z_1, \dots, Z_k$ são as observações não-censuradas (i.é, $\delta_i = 1$, $i = 1, \dots, k$) e $Z_{k+1}, \dots, Z_n$ são as observações censuradas (i.é, $\delta_i = 2$, $i = k+1, \dots, n$).

* Denotamos por $Z_{(k+1)}^* < \dots < Z_{(m)}^*$ as estatísticas de ordem diferentes entre as observações censuradas $Z_{k+1}, \dots, Z_n$.

Sejam, $\lambda_j = \sum_{i=k+1}^n I(Z_i = Z_{(j)}^*, \delta_i = 2)$, i.é, o número de observações censuradas que são iguais a $Z_{(j)}^*$, $j = k+1, \dots, m$, $N(t) = \sum_{i=1}^n I(Z_i \geq t)$, i.é., o número de observações maiores ou iguais a t , e $N^+(t) = \sum_{i=1}^n I(Z_i > t)$.

Então, de acordo com Susarla e Van Ryzin (1976), o estimador bayesiano da função de sobrevivência $S_1(t) = \Pr(X > t)$, considerando como distribuição a priori para S_1 um processo de Dirichlet de parâmetro α e sob função de perda quadrática, é dado pela expressão:

$$\hat{S}_1^{SV}(t) = \frac{\alpha(t, \infty) + N^+(t)}{\alpha(0, \infty) + n} \prod_{j=k+1}^{\ell} \frac{\alpha[Z_{(j)}^*, \infty) + N(Z_{(j)}^*)}{\alpha[Z_{(j)}^*, \infty) + N(Z_{(j)}^*) - \lambda_j},$$

no intervalo $Z_{(\ell)}^* \leq t < Z_{(\ell+1)}^*$, $\ell = k, \dots, m$, onde $Z_{(k)}^* = 0$ e $Z_{(m+1)}^* = \infty$.

Propriedades assintóticas deste estimador são estudadas em Susarla e Van Ryzin (1978b) e uma abordagem bayesiana empírica é considerada em Susarla e Van Ryzin (1978a).

O presente trabalho é dedicado à *análise bayesiana não-paramétrica na teoria de riscos competitivos*. Assim, algumas contribuições nessa área pretendem ser dadas.

No capítulo 2, apresentamos, primeiramente, a formulação do problema bayesiano não-paramétrico. Sendo a classe de processos de Dirichlet fundamental na resolução desses problemas, as construções e propriedades de tais processos desenvolvidas por Ferguson (1973) e por Sethuraman (Tiwari, 1981; Sethuraman e Tiwari, 1982) são descritas com alguns detalhes. Aplicações do processo de Dirichlet em certos problemas de estimação são também consideradas, a título de ilustração.

Uma classe de processos de Dirichlet multivariado é introduzida no Capítulo 3. Consideramos primeiro a construção de um processo de Dirichlet bivariado. Duas provas de

existência do processo de Dirichlet bivariado são dadas. Uma é não-construtiva e é baseada no teorema de existência de Kolmogorov. A outra definição é dada em função de dois processos de Dirichlet e de uma certa variável aleatória Beta, mutuamente independentes. Mostramos também que o processo de Dirichlet bivariado satisfaz uma relação do tipo distribucional. Os resultados do caso bivariado são estendidos ao caso multivariado. Estabelecemos então o processo de Dirichlet multivariado.

Utilizando os resultados do capítulo anterior, no Capítulo 4, o processo de Dirichlet multivariado é considerado como distribuição a priori para o modelo de riscos competitivos. Em uma primeira etapa, o processo de Dirichlet bivariado é utilizado como distribuição a priori para o problema com dados censurados. Assim, são obtidos estimadores bayesianos de algumas quantidades de interesse. Em particular, tendo obtido um estimador bayesiano, sob perda quadrática, para o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, com base na fórmula de Peterson (1977), propomos um estimador não paramétrico da função de sobrevivência $S_1(t) = \Pr(X > t)$. Os resultados são estendidos ao problema de riscos competitivos, onde utilizamos o processo de Dirichlet multivariado como distribuição a priori.

Finalmente, o Capítulo 5 é dedicado a estudar propriedades assintóticas do estimador não-paramétrico da função de sobrevivência $S_1(t) = \Pr(X > t)$, digamos $\hat{S}_1$, obtido com base numa função de perda quadrática e na fórmula de Peterson. Utilizando argumentos de continuidade a consistência forte de $\hat{S}_1$ é provada. A convergência fraca e uma lei do logaritmo iterado para o estimador $\hat{S}_1$ são estabelecidas, usando uma abordagem baseada no cálculo diferencial em espaços vetoriais normados.

CAPÍTULO 2

O processo de Dirichlet na inferência bayesiana não-paramétrica

2.1 Introdução

Na Inferência Estatística, problemas com um espaço de parâmetros infinito-dimensional, usualmente um espaço de distribuições de probabilidade, são de muita importância, tanto do ponto de vista teórico quanto do aplicado. O enfoque bayesiano de tais problemas requer uma distribuição de probabilidade definida sobre esse espaço (ver por exemplo Ferguson, 1974). Muitos avanços têm sido feitos nas duas últimas décadas, conforme podemos observar nos mais diversos periódicos em Estatística e Probabilidade.

Neste capítulo, apresentamos, primeiramente, a formulação do problema bayesiano não-paramétrico. Sendo a classe de processos de Dirichlet fundamental na resolução desses problemas, as construções e propriedades de tais processos desenvolvidas por Ferguson (1973), e por Sethuraman (Tiwari, 1981; Sethuraman e Tiwari, 1982), são descritas com alguns detalhes. Também são consideradas, a título de ilustração, aplicações do processo de Dirichlet em alguns problemas estatísticos.

2.2 Inferência bayesiana não-paramétrica

Na Análise Bayesiana Não-Paramétrica considera-se uma sequência de variáveis aleatórias $X_1, X_2, \dots$ independentes e identicamente distribuídas com distribuição desconhecida P , isto corresponde a dizer que dada P , $X_1, X_2, \dots$ são identicamente distribuídas segundo P e são condicionalmente independentes.

Assim, P é o parâmetro de interesse, $P \in \mathcal{P}$, onde $\mathcal{P}$ é uma classe de medidas de probabilidade sobre um espaço mensurável $(\mathcal{X}, \mathcal{A})$. Uma priori para P é uma medida de probabilidade sobre $(\mathcal{P}, \sigma(\mathcal{P}))$, sendo $\sigma(\mathcal{P})$ a menor σ -álgebra tal que a função $P \in \mathcal{P} \mapsto P(A) \in [0, 1]$ é $\sigma(\mathcal{P})$ -mensurável para todo $A \in \mathcal{A}$. Para um estudo relativo a σ -álgebras sobre espaços de medidas de probabilidade, ver Gaudard e Hadwin (1989).

Esta priori pode ser vista como um processo estocástico definido num espaço de probabilidade $(\Omega, \mathcal{F}, \eta)$, onde o conjunto de índices é a σ -álgebra $\mathcal{A}$, ou seja, temos uma família de variáveis aleatórias $\{P(A) : A \in \mathcal{A}\}$, onde as trajetórias $P(\cdot, \omega) : \mathcal{A} \rightarrow [0, 1]$, $\omega \in \Omega$, são medidas de probabilidade sobre $(\mathcal{X}, \mathcal{A})$.

Podemos entender este tipo de processo como uma medida de probabilidade aleatória P sobre $(\mathcal{X}, \mathcal{A})$, ou seja, uma função de transição de $(\Omega, \mathcal{F})$ em $(\mathcal{X}, \mathcal{A})$, i.é.:

$P(\cdot, \cdot) : \Omega \times \mathcal{A} \rightarrow [0, 1]$ é uma função mensurável tal que:

- (i) para todo $\omega \in \Omega$, $P(\omega, \cdot) : \mathcal{A} \rightarrow [0, 1]$ é uma medida de probabilidade sobre $(\mathcal{X}, \mathcal{A})$.
- (ii) para todo $A \in \mathcal{A}$, $P(\cdot, A) : \Omega \rightarrow [0, 1]$ é uma variável aleatória a valores em $[0, 1]$.

Ver por exemplo Parthasarathy (1967). Em nosso trabalho outras condições são também consideradas, ver Doksum (1974).

Uma medida de probabilidade aleatória P pode também ser vista como uma função mensurável:

$$\begin{aligned} P : (\Omega, \mathcal{F}, \eta) &\rightarrow (P, \sigma(\mathcal{P})) \\ \omega &\mapsto P(\omega, \cdot) . \end{aligned}$$

Neste caso, a distribuição de P , digamos ηP^{-1} é a medida de probabilidade a priori sobre $(\mathcal{P}, \sigma(\mathcal{P}))$.

Com o que foi exposto, o problema bayesiano não-paramétrico requer uma classe de medidas de probabilidade aleatórias, ou uma classe de processos estocásticos, ou uma classe de medidas de probabilidade a priori.

Neste contexto, é conveniente a consideração de uma classe de processos estocásticos tal que:

- (I) Seja suficientemente grande a ponto de diferentes tipos de opinião possuir representação ou representação aproximada.
- (II) Seja fechada em relação à operação bayesiana, isto é, a priori escolhida nessa classe produz uma posteriori na mesma classe. Classes de distribuições satisfazendo essa propriedade também são chamadas de conjugadas.

As idéias fundamentais sobre Inferência Bayesiana Não-Paramétrica são discutidas por Dalal (1980) e uma revisão dos avanços nas duas últimas décadas é feita por Ferguson et al. (1992).

2.3 O processo de Dirichlet

A classe de processos de Dirichlet, introduzida por Ferguson (1973) num artigo fundamental sobre Estatística Bayesiana Não-Paramétrica, é particularmente conveniente, pois esta família satisfaz as propriedades (I) e (II) da seção anterior. Brevemente, o processo de Dirichlet pode ser descrito como segue.

Seja $(\mathcal{X}, \mathcal{A})$ um espaço mensurável, e α uma medida finita não-nula sobre $(\mathcal{X}, \mathcal{A})$. O processo de Dirichlet é uma medida de probabilidade aleatória P sobre $(\mathcal{X}, \mathcal{A})$ tal que, para toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$; a distribuição do vetor aleatório $(P(A_1), \dots, P(A_k))$ é Dirichlet singular com parâmetro $(\alpha(A_1), \dots, \alpha(A_k))$.

Ferguson (1973) provou, através do Teorema de Existência de Kolmogorov, que existe uma medida de probabilidade sobre $([0, 1]^{\mathcal{A}}, \sigma([0, 1]^{\mathcal{A}}))$ sustentando as distribuições de Dirichlet finito-dimensionais anteriores. Aqui $[0, 1]^{\mathcal{A}}$ é o espaço produto, onde cada fator é o intervalo unitário fechado $[0, 1]$, sendo tantos os fatores quantos elementos tenha a σ -álgebra $\mathcal{A}$. Na maior parte das aplicações $\mathcal{A}$ é infinita não-enumerável. Equivalentemente,

$[0, 1]^{\mathcal{A}}$ pode ser visto como a classe de funções de conjuntos definidas sobre $\mathcal{A}$ assumindo valores em $[0, 1]$. Também, $\sigma([0, 1]^{\mathcal{A}})$ é a σ -álgebra produto para $[0, 1]^{\mathcal{A}}$, gerada por cilindros mensuráveis tendo uma base finita, i.é, conjuntos cilindro cuja base é determinada por restrições mensuráveis sobre um número finito de coordenadas de $[0, 1]^{\mathcal{A}}$. Então $\sigma([0, 1]^{\mathcal{A}})$ é precisamente o conjunto de cilindros mensuráveis tendo uma base enumerável. Considerando $[0, 1]^{\mathcal{A}}$ como uma classe de funções de conjuntos, cada conjunto em $\sigma([0, 1]^{\mathcal{A}})$ pode ser definido por restrições sobre uma coleção enumerável $\{x(A_1), x(A_2), \dots\}$, onde $\{A_1, A_2, \dots\}$ é um subconjunto enumerável de $\mathcal{A}$ e x é um elemento de $[0, 1]^{\mathcal{A}}$.

Observemos que, quando $\mathcal{A}$ é não-enumerável, os conjuntos unitários em $[0, 1]^{\mathcal{A}}$ não são mensuráveis em $\sigma([0, 1]^{\mathcal{A}})$. Também, a classe $\mathcal{P}$ de todas as medidas de probabilidade sobre $(\mathcal{X}, \mathcal{A})$ não pertence a $\sigma([0, 1]^{\mathcal{A}})$, pois $\mathcal{P}$ não é determinada por um número enumerável de restrições quando $\mathcal{A}$ é não-enumerável. Assim, sentenças como “o processo de Dirichlet da probabilidade 1 à classe $\mathcal{P}$ ” carecem de significado. Uma discussão mais detalhada sobre esses assuntos pode ser encontrada em Tiwari (1981).

??
Exemplos?

Por outro lado, a idéia básica na definição do processo de Dirichlet segundo Sethuraman, é a construção de uma transformação mensurável definida sobre algum espaço de probabilidade, assumindo valores em $(\mathcal{P}, \sigma(\mathcal{P}))$ através da qual obtém-se uma medida de probabilidade aleatória P sobre $(\mathcal{X}, \mathcal{A})$ que é um processo de Dirichlet.

Sendo que nosso trabalho trata fundamentalmente da classe de processos de Dirichlet, tanto de suas representações quanto de suas aplicações, esta seção tem como objetivo descrever com algum detalhe as construções de Ferguson, e de Sethuraman, de tal família de processos.

2.3.1 Construção de Ferguson

Consideramos o espaço mensurável $(\mathcal{X}, \mathcal{A})$. Construimos uma medida de probabilidade

aleatória P sobre $(\mathcal{X}, \mathcal{A})$, definindo a distribuição do vetor aleatório $(P(A_1), \dots, P(A_k))$, para todo inteiro positivo k e para toda partição mensurável $\{A_i : i = 1, \dots, k\}$ de $\mathcal{X}$. Logo definimos a distribuição do vetor aleatório $(P(B_1), \dots, P(B_m))$, onde $B_1, B_2, \dots, B_m$ são conjuntos arbitrários de $\mathcal{A}$.

Dados $B_1, \dots, B_m \in \mathcal{A}$, consideramos os $k = 2^m$ conjuntos de $\mathcal{A}$

$$A_{\xi_1, \dots, \xi_m} = \bigcap_{j=1}^m B_j^{\xi_j}$$

$$\xi_j = 0 \quad \text{ou} \quad 1 \tag{2.3.1}$$

$$B_j^1 = B_j, \quad B_j^0 = B_j^c$$

então $\{A_{\xi_1, \dots, \xi_m} : \xi_j = 0 \text{ ou } 1, j = 1, 2, \dots, m\}$ é uma partição mensurável de $\mathcal{X}$.

Se temos a distribuição do 2^m -vetor

$$(P(A_{\xi_1, \dots, \xi_m}) : \xi_j = 0 \quad \text{ou} \quad 1, j = 1, \dots, m) \tag{2.3.2}$$

então a distribuição do vetor $(P(B_1), \dots, P(B_m))$ é obtida de (2.3.2) definindo para $i = 1, 2, \dots, m$

$$P(B_i) = \sum_{\substack{(\xi_1, \dots, \xi_m) \\ \xi_i = 1}} P(A_{\xi_1, \dots, \xi_m}) \tag{2.3.3}$$

Se $\{B_1, \dots, B_m\}$ for uma partição de $\mathcal{X}$, isto não leva a contradições, sob o suposto que

$$P(\phi) = 0 \quad \text{q.c.} \tag{2.3.4}$$

Vamos supor também que a primeira condição, equação (36.4) do Billingsley (1986), do Teorema de Kolmogorov é satisfeita em nosso caso. } ?

Dado um sistema de distribuições de $(P(A_1), \dots, P(A_k))$ para todo inteiro $k > 0$ e toda partição mensurável $\{A_1, \dots, A_k\}$, de $\mathcal{X}$, consideramos o seguinte critério de consistência:

Condição C:

Se $\{A'_1, \dots, A'_{k'}\}$ e $\{A_1, \dots, A_k\}$ são duas partições de $\mathcal{X}$ e se $\{A'_1, \dots, A'_{k'}\}$ é uma refinação de $\{A_1, \dots, A_k\}$ com $A_1 = \bigcup_1^{r_1} A'_i, \dots, A_k = \bigcup_{r_{k-1}+1}^{k'} A'_i$, então a distribuição do vetor $(\sum_1^{r_1} P(A'_i), \dots, \sum_{r_{k-1}+1}^{k'} P(A'_i))$ determinada da distribuição conjunta de $(P(A'_1), \dots, P(A'_{k'}))$ é idêntica à distribuição conjunta de $(P(A_1), \dots, P(A_k))$.

Observação 2.3.1

A função P , como descrita na seção anterior, é chamada de medida de probabilidade aleatória sobre $(\mathcal{X}, \mathcal{A})$, quando a condição C é satisfeita e se $P(\mathcal{X}) = 1$ q.c.

O lema seguinte mostra que a condição C é suficiente para a validação das condições do Teorema de Existência de Kolmogorov (ver por exemplo Billingsley, 1986), para as distribuições de $(P(B_1), \dots, P(B_m))$ como definidas anteriormente.

Lema 2.3.1

Se um sistema de distribuições de $(P(A_1), \dots, P(A_k))$ para todo inteiro $k > 0$ e toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, é definido satisfazendo a condição C, e se para $B_1, B_2, \dots, B_m \in \mathcal{A}$, arbitrários, a distribuição de $(P(B_1), \dots, P(B_m))$ é definida como em (2.3.1), (2.3.2) e (2.3.3) então existe uma medida de probabilidade, digamos $\mathbb{P}$, sobre $([0, 1]^{\mathcal{A}}, \sigma[0, 1]^{\mathcal{A}})$, onde $[0, 1]^{\mathcal{A}} = \{x : \mathcal{A} \rightarrow [0, 1]\}$ e $\sigma[0, 1]^{\mathcal{A}}$ é a σ -álgebra gerada pelos conjuntos cilindros de $[0, 1]^{\mathcal{A}}$, talque o processo variável-coordenada $\{Z_A : A \in \mathcal{A}\}$, $Z_A : [0, 1]^{\mathcal{A}} \rightarrow [0, 1]$
 $P \mapsto P(A)$, tem o sistema de distribuições original como o seu sistema de distribuições finito-dimensionais.

Prova

Ver Ferguson (1973).

□

Definição 2.3.1

Seja α uma medida finita, não-nula, não-negativa e finitamente aditiva sobre $(\mathcal{X}, \mathcal{A})$. Dizemos que a medida de probabilidade aleatória P sobre $(\mathcal{X}, \mathcal{A})$ é um processo de Dirichlet sobre $(\mathcal{X}, \mathcal{A})$ com parâmetro α , se para todo inteiro $k > 0$ e para toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, a distribuição do vetor aleatório $(P(A_1), \dots, P(A_k))$ é Dirichlet singular com parâmetro $(\alpha(A_1), \dots, \alpha(A_k))$. **Notação:** $P \in \mathcal{D}(\alpha)$.

Observação 2.3.2

A condição C para o processo de Dirichlet corresponde à propriedade da distribuição de Dirichlet descrita na Proposição 6 do Apêndice A.

Segue do Lema 2.3.1 e do Teorema de Kolmogorov que isto, de fato, define um processo estocástico. Notemos também que $P(\mathcal{X}) = 1$ q.c.

Ferguson (1973) mostrou também que as condições (I) e (II) da seção 2.2 são satisfeitas. No contexto da Estatística Bayesiana Não-Paramétrica a propriedade (II) pode ser apresentada como:

Teorema 2.3.1

Seja $P \in \mathcal{D}(\alpha)$ e seja $X_1, \dots, X_n$ uma amostra aleatória de tamanho n de P , i.é, dado P , $X_1, \dots, X_n$ são independentes e identicamente distribuídas segundo P , onde $P \in \mathcal{D}(\alpha)$.

Então a distribuição condicional de P dada $X_1, \dots, X_n$ é $\mathcal{D}(\alpha + \sum_{i=1}^n \delta_{X_i})$, onde

$$\delta_x(A) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}.$$

Prova

Ver Ferguson (1973). □

2.3.2 Construção de Sethuraman

Consideramos $(\mathcal{X}, \mathcal{A})$ e α como antes.

Seja $\mathbf{W} = \{(U_1, \theta_1), (U_2, \theta_2), \dots\}$ a sequência de vetores aleatórios bidimensionais tal que:

- (i) As variáveis aleatórias U_i 's são independentes e identicamente distribuídas assumindo valores em $\mathcal{X}$, com distribuição comum dada por $\bar{\alpha}(\cdot) = \alpha(\cdot)/\alpha(\mathcal{X})$.
- (ii) As variáveis aleatórias θ_j 's são independentes e identicamente distribuídas assumindo valores em $[0, 1]$ com distribuição comum Beta $(1, \alpha(\mathcal{X}))$.
- (iii) As variáveis aleatórias U_i 's e θ_j 's são mutuamente independentes.

Seja $(\Omega, \sigma(\Omega), \eta_\alpha)$ o espaço produto, onde $\Omega = (\mathcal{X} \times [0, 1])^\infty$, i.é, o espaço das realizações de $\mathbf{W}$, $\sigma(\Omega) = (\mathcal{A} \times \mathcal{B}[0, 1])^\infty$, onde $\mathcal{B}[0, 1]$ é a usual restrição da σ -álgebra de Borel ao intervalo $[0, 1]$, e η_α é a lei de $\mathbf{W}$.

Definimos a função:

$$\begin{aligned} P : (\Omega, \sigma(\Omega)) &\rightarrow (\mathcal{P}, \sigma(\mathcal{P})) \\ \omega &\mapsto P(\omega, \cdot) \end{aligned} \quad (2.3.5)$$

tal que para cada $\omega \in \Omega$ e para todo $A \in \mathcal{A}$,

$$P(\omega, A) = \sum_{j=1}^{\infty} p_j(\omega) \delta_{U_j(\omega)}(A) \quad (2.3.6)$$

onde

$$p_1 = \theta_1, \quad p_2 = (1 - \theta_1)\theta_2, \quad p_3 = (1 - \theta_1)(1 - \theta_2)\theta_3, \dots \quad (2.3.7)$$

então a medida induzida por P , digamos $D^\alpha = \eta_\alpha P^{-1}$ e a medida de Dirichlet com parâmetro α .

Esse resultado devido a Sethuraman, que aparece provado com algum detalhe em Tiwari (1981), requer várias etapas na sua demonstração.

Teorema 2.3.2

Para qualquer $A \in \mathcal{A}$, a variável aleatória $P(A)$ tem distribuição Beta $(\alpha(A), \alpha(A^c))$.

□

Este último teorema é provado considerando o seguinte:

$$P(A) = p_1 \delta_{U_1}(A) + (1 - p_1)P'(A) \quad (2.3.8)$$

onde

$$P'(A) = \sum_{j=2}^{\infty} p'_j \delta_{U_j}(A) \quad (2.3.9)$$

$$\text{e } p'_j = \frac{p_j}{1 - p_1} \quad j = 2, 3, \dots \quad (2.3.10)$$

Notemos que as variáveis aleatórias p_1 , $\delta_{U_1}(A)$ e $P'(A)$ são mutuamente independentes.

Dado que

$$p'_2 = \frac{p_2}{1 - p_1} = \theta_2, \quad p'_3 = (1 - \theta_2)\theta_3, \dots \quad (2.3.11)$$

então $P(A) \sim P'(A)$. A relação (2.3.8) pode ser colocada como

$$P(A) \sim p_1 \delta_{U_1}(A) + (1 - p_1)P(A) \quad (2.3.12)$$

Logo é mostrado que $P(A) \sim \text{Beta}(\alpha(A), \alpha(A^c))$ é uma solução da relação (2.3.12), ver Proposição 3.1.1 de Tiwari (1981).

O fato que $P(A) \sim \text{Beta}(\alpha(A), \alpha(A^c))$ é a única solução da relação, é garantido pelo seguinte lema, cuja demonstração pode ser encontrada em Tiwari (1981, Proposição 3.1.2).

Lema 2.3.2

Sejam U, V variáveis aleatórias tais que $0 < U < 1$ q.c.

Seja X uma variável aleatória tal que $X \amalg (U, V)$.

Se a relação $X \sim UX + V$ possui duas soluções X e Y , então $X \sim Y$.

□

Observação 2.3.3

Quando $A = \mathcal{X}$, temos que $P(\mathcal{X}) \sim \text{Beta}(\alpha(\mathcal{X}), 0)$, i.é, $P(\mathcal{X}) = 1$ q.c. $[\eta_\alpha]$. Isto mostra que P é uma medida de probabilidade aleatória sobre $(\mathcal{X}, \mathcal{A})$.

Considerando agora uma partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, conclui-se que $(P(A_1), \dots, P(A_k)) \sim \mathcal{D}_s(\alpha(A_1), \dots, \alpha(A_k))$ é uma solução da relação $(P(A_1), \dots, P(A_k)) \sim (p_1 \delta_{U_1}(A_1) + (1 - p_1)P(A_1), \dots, p_1 \delta_{U_1}(A_k) + (1 - p_1)P(A_k))$. Um lema de tipo vetorial análogo ao Lema 2.3.2 garante que esta solução é única.

Finalmente, temos:

Teorema 2.3.3

A medida de probabilidade aleatória P sobre $(\mathcal{X}, \mathcal{A})$ definida em (2.3.6) é um processo de Dirichlet de parâmetro α . □

Observação 2.3.4

Da definição do processo de Dirichlet P em (2.3.5) e (2.3.6), temos que quase toda realização de P é uma medida de probabilidade discreta sobre $(\mathcal{X}, \mathcal{A})$, i.é, se $\mathcal{P}_0$ é a família de medidas de probabilidade discretas sobre $(\mathcal{X}, \mathcal{A})$ então $D^\alpha(\mathcal{P}_0) = 1$. Este último fato poderia levar a dúvidas sobre o uso do processo de Dirichlet como distribuição a priori em problemas de inferência que envolvem funções de distribuição contínuas. Mas, no caso de $(\mathcal{X}, \mathcal{A})$ ser um espaço de Borel - um espaço métrico separável e completo $\mathcal{X}$, com $\mathcal{A}$ sendo a σ -álgebra gerada pelos abertos de $\mathcal{X}$ - caso muito usual nas aplicações, Basu e Tiwari (1982, Teorema 8.4) mostraram que sob, a condição de que a α -medida de todo aberto de $\mathcal{X}$ seja positiva, para quase toda realização do processo de Dirichlet P , tem-se que se $\{x_1, x_2, \dots\}$ é o conjunto de pontos onde a realização do processo coloca massa, então $\{x_1, x_2, \dots\}$ é denso em $\mathcal{X}$.

2.3.3 Notas bibliográficas

Uma outra definição do processo de Dirichlet, baseada em processos Gamma com incrementos independentes, foi dada por Ferguson (1973). Outros trabalhos na teoria de processos de Dirichlet são os de Blackwell (1973), Blackwell e Mac-Queen (1973), e Berk e Savage (1975).

Problemas de convergência na teoria de processos de Dirichlet são estudados por Korwar e Hollander (1973), Sethuraman e Tiwari (1982), Lo (1983). Alguns resultados sobre distribuições de funcionais do processo foram obtidos por Hannum et al. (1981). Paralelamente à teoria de processos de Dirichlet, algumas extensões foram desenvolvidas: misturas de processos de Dirichlet são introduzidas por Antoniak (1974), e também é tema de interesse de Dalal (1978) e Kuo (1986). O conceito de processos de Dirichlet invariantes é considerado nos trabalhos de Dalal (1979), e Hannum e Hollander (1983).

2.4 Aplicações do processo de Dirichlet a alguns problemas não-paramétricos

Nesta seção, e a título de ilustração, descrevemos algumas aplicações do processo de Dirichlet no problema de estimação bayesiana não-paramétrica.

Exemplo 2.4.1

Consideramos o problema de estimar uma função de sobrevivência S , definida sobre o conjunto dos números reais positivos.

Sejam $\mathcal{X} = (0, \infty)$, $\mathcal{A} = \mathcal{B}(0, \infty)$ σ -álgebra dos borelianos de $(0, \infty)$. O espaço de ações do estatístico é o conjunto de todas as funções de sobrevivência sobre $(0, \infty)$.

Seja a função de perda quadrática

$$L(S, \hat{S}) = \int [S(t) - \hat{S}(t)]^2 dW(t) \quad (2.4.1)$$

onde W é uma função de peso e $S(t) = P(t, \infty)$.

Se $P \in \mathcal{D}(\alpha)$ então $S(t) \sim \text{Beta}(\alpha(t, \infty); \alpha(0, t])$.

O risco bayesiano é

$$\mathbb{E}L(S, \hat{S}) = \int \mathbb{E}[S(t) - \hat{S}(t)]^2 dW(t) \quad (2.4.2)$$

expressão que é minimizada escolhendo $\hat{S}(t)$ para cada t a minimizar $\mathbb{E}[S(t) - \hat{S}(t)]^2$, logo $\hat{S}(t) = \mathbb{E}S(t) = \frac{\alpha(t, \infty)}{\alpha(0, \infty)}$ para cada t , i.é, a regra de Bayes para o problema não-amstral é o valor esperado a priori:

$$S_0(t) = \frac{\alpha(t, \infty)}{\alpha(0, \infty)} \quad (2.4.3)$$

Para uma amostra $X_1, X_2, \dots, X_n$ de S , a regra de Bayes é, usando o Teorema 2.3.1:

$$\hat{S}_n(t) = \frac{(\alpha + \sum_1^n \delta_{X_i})(t, \infty)}{(\alpha + \sum_1^n \delta_{X_i})(0, \infty)} = \frac{\alpha(t, \infty) + \sum_1^n I(X_i > t)}{\alpha(0, \infty) + n},$$

sendo $I(A)$ a função indicadora do conjunto A . Chamando

$$r_n = \frac{\alpha(0, \infty)}{\alpha(0, \infty) + n} \quad (2.4.4)$$

e

$$S_n(t) = \frac{1}{n} \sum_1^n I(X_i > t) \quad (2.4.5)$$

a função de sobrevivência empírica da amostra $X_1, \dots, X_n$, então

$$\hat{S}_n(t) = r_n \cdot S_0(t) + (1 - r_n) \cdot S_n(t). \quad (2.4.6)$$

Vemos então, de (2.4.6), que o estimador bayesiano de $S(t)$, sob perda quadrática, é uma combinação convexa do valor esperado a priori e a função de sobrevivência empírica da amostra.

Notemos que $\alpha(0, \infty)$ grande [pequeno] quer dizer maior peso [menor peso] da priori e menor peso [maior peso] da amostra.

Também temos pelo Teorema de Glivenko-Cantelli que dado $r_n \searrow 0$ quando $n \rightarrow \infty$, então $\hat{S}_n(t)$ converge a $S(t)$, uniformemente (q.c.), i.é $\sup_{t \in (0, \infty)} |\hat{S}_n(t) - S(t)| \rightarrow 0$, quando

$n \rightarrow \infty$, com probabilidade 1. Logo $\hat{S}_n$ é um estimador fortemente consistente da função de sobrevivência S .

Exemplo 2.4.2

No Exemplo 2.4.1, consideramos a seguinte função de perda 0-1

$$L((S(t), \tilde{S}(t))) = \begin{cases} 0 & \text{se } |S(t) - \tilde{S}(t)| \leq c \\ 1 & \text{se } |S(t) - \tilde{S}(t)| > c \end{cases} \text{ para cada } t \quad (2.4.7)$$

onde $c > 0$.

O risco bayesiano é:

$$\mathbb{E}L(S(t), \tilde{S}(t)) = Pr\{|S(t) - \tilde{S}(t)| > c\} \text{ para cada } t \quad (2.4.8)$$

logo a regra de Bayes para o problema não-amstral é a moda da priori, i.é o estimador de máxima probabilidade a priori

$$M_0(S(t)) = \frac{\alpha(t, \infty) - 1}{\alpha(0, \infty) - 2} \quad (2.4.9)$$

e a regra de Bayes para uma amostra $X_1, \dots, X_n$ de S é, novamente usando o Teorema 2.3.1,

$$\tilde{S}_n(t) = r_n \cdot M_0(S(t)) + (1 - r_n) \cdot S_n(t) \quad (2.4.10)$$

sendo $r_n = \frac{\alpha(0, \infty) - 2}{\alpha(0, \infty) + n - 2}$.

As mesmas propriedades obtidas para $\hat{S}_n$ valem para o estimador $\tilde{S}_n$. $\tilde{S}_n$ é o chamado estimador de máxima probabilidade a posteriori, também conhecido como estimador de máxima verossimilhança generalizado.

Exemplo 2.4.3

Consideramos o problema de estimar a média de uma função de distribuição desconhecida F .

Seja $\mathcal{X} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$, borelianos de $\mathbb{R}$, e consideremos a função de perda quadrática $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$ onde $\mu = \int x dP(x)$, $F(t) = P(-\infty, t]$.

Suponhamos que $P \in \mathcal{D}(\alpha)$ e que $\int |x| d\alpha(x) < \infty$.

O risco de Bayes é $\mathbb{E}L(\mu, \hat{\mu}) = \mathbb{E}(\mu - \hat{\mu})^2$, que é minimizado quando $\hat{\mu} = \mathbb{E}\mu = \mathbb{E} \int x dP(x)$, expressão que devemos calcular.

Mostramos primeiro, dois resultados mais gerais.

Proposição 2.4.1

Seja P como definida em (2.3.6) e $\varphi : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ uma função $\mathcal{A}$ - $\mathcal{B}$ -mensurável.

Então

- (a) $\int_{\mathcal{X}} \varphi dP = \sum_{j=1}^{\infty} p_j \varphi(U_j)$,
- (b) $\mathbb{E} \int_{\mathcal{X}} \varphi dP = \int_{\mathcal{X}} \varphi d(\mathbb{E}P)$.

Prova de (a):

Seja $\varphi = 1_A \in \mathcal{A}$, então

$$\int_{\mathcal{X}} \varphi dP = P(A) = \sum_{j=1}^{\infty} p_j \delta_{U_j}(A) = \sum_{j=1}^{\infty} p_j 1_{\{U_j \in A\}} .$$

Usando a técnica usual da teoria da medida, segue o resultado para funções simples não-negativas, logo para $\varphi \geq 0$ $\mathcal{A}$ - $\mathcal{B}$ -mensurável, e finalmente consideramos $\varphi = \varphi^+ - \varphi^-$, onde $\varphi^+ = \sup(\varphi, 0)$ e $\varphi^- = \sup(-\varphi, 0)$.

Prova de (b):

Seja $\varphi = 1_A \in \mathcal{A}$, logo

$$\mathbb{E} \int_{\mathcal{X}} \varphi dP = \mathbb{E} \int_{\mathcal{X}} 1_A dP = \mathbb{E}P(A) = \int_{\mathcal{X}} 1_A d(\mathbb{E}P)$$

e aplica-se a mesma técnica anterior. □

Proposição 2.4.2

Com o mesmo enunciado da proposição anterior.

Se $\int_{\mathcal{X}} |\varphi| d\alpha < \infty$ então $\int_{\mathcal{X}} |\varphi| dP < \infty$ q.c., e

$$\mathbb{E} \int_{\mathcal{X}} \varphi dP = \frac{1}{\alpha(\mathcal{X})} \int_{\mathcal{X}} \varphi d\alpha \quad (2.4.11)$$

Prova:

Da Proposição 2.4.1, (a), temos que $\int_{\mathcal{X}} |\varphi| dP = \sum_1^{\infty} |\varphi(U_j)| p_j$, pelo Teorema da Convergência Monótona e pelo fato das variáveis aleatórias U_j 's e p_j 's serem mutuamente independentes, temos que $\mathbb{E} \int_{\mathcal{X}} |\varphi| dP = \sum_{j=1}^{\infty} \mathbb{E} |\varphi(U_j)| \mathbb{E} p_j = \int_{\mathcal{X}} \varphi d\bar{\alpha} \sum_{j=1}^{\infty} \mathbb{E} p_j$ pois as variáveis aleatórias U_j 's são independentes e identicamente distribuídas com distribuição $\bar{\alpha}(\cdot) = \alpha(\cdot)/\alpha(\mathcal{X})$, finalmente $\mathbb{E} \int_{\mathcal{X}} |\varphi| dP = \frac{1}{\alpha(\mathcal{X})} \int_{\mathcal{X}} |\varphi| d\alpha$, pois $\sum_1^{\infty} p_j = 1$ q.c. Logo $\int_{\mathcal{X}} |\varphi| d\alpha < \infty \Rightarrow \mathbb{E} \int_{\mathcal{X}} |\varphi| dP < \infty \Rightarrow \int_{\mathcal{X}} |\varphi| dP < \infty$ q.c.

É fácil ver que a série $\sum_1^{\infty} \varphi(U_j) p_j$ é absolutamente convergente q.c. e de fato limitada por $\sum_1^{\infty} |\varphi(U_j)| p_j \in L^1$.

Aplicando o Teorema da Convergência Dominada, obtemos $\mathbb{E} (\int_{\mathcal{X}} \varphi dP) = \sum_{j=1}^{\infty} \mathbb{E} \varphi(U_j) \mathbb{E} p_j = \frac{1}{\alpha(\mathcal{X})} \int_{\mathcal{X}} \varphi d\alpha$. $\square$

Voltando ao nosso exemplo e aplicando a Proposição 2.4.2, temos que $\varphi(x) = x$, logo a regra de Bayes para o problema não-amstral é

$$\mu_0 = \frac{1}{\alpha(\mathbb{R})} \int_{\mathbb{R}} x d\alpha(x)$$

e para uma amostra $X_1, \dots, X_n$ de F , a regra de Bayes é:

$$\hat{\mu}_n = \frac{1}{\alpha(\mathbb{R}) + n} \int_{\mathbb{R}} x d \left(\alpha + \sum_1^n \delta_{X_i}(x) \right) = r_n \mu_0 + (1 - r_n) \bar{X}_n \quad (2.4.12)$$

sendo $r_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R}) + n}$ e $\bar{X}_n = \frac{1}{n} \sum_1^n X_i$ é a média da amostra. Novamente obtemos uma combinação convexa, esta vez da média da priori e da média da amostra. Também $\hat{\mu}_n \rightarrow \mu$

q.c. quando $n \rightarrow \infty$, i.é o estimador é fortemente consistente na classe das distribuições com primeiro momento finito.

Observação 2.4.2

É possível estabelecer uma proposição para momentos de segunda ordem, análoga à Proposição 2.4.2, i.é sob o suposto $\int_{\mathcal{X}} |\varphi_1| d\alpha < \infty$, $\int_{\mathcal{X}} |\varphi_2| d\alpha < \infty$ e $\int_{\mathcal{X}} |\varphi_1 \cdot \varphi_2| d\alpha < \infty$ então $\mathbb{E} \int_{\mathcal{X}} \varphi_1 dP \cdot \int_{\mathcal{X}} \varphi_2 dP = \frac{\sigma_{12}}{\alpha(\mathcal{X})+1} + \mu_1 \cdot \mu_2$, sendo $\mu_i = \frac{1}{\alpha(\mathcal{X})} \int_{\mathcal{X}} \varphi_i dP$, $i = 1, 2$ e $\sigma_{12} = \frac{1}{\alpha(\mathcal{X})} \int_{\mathcal{X}} \varphi_1 \varphi_2 dP$.

Este resultado é usado no cálculo do estimador bayesiano da variância de uma função de distribuição desconhecida F , neste caso temos que

$$\hat{\sigma}_n^2 = \frac{\alpha(\mathbb{R}) + n}{\alpha(\mathbb{R}) + n + 1} \text{Var } \hat{F}_n \quad (2.4.13)$$

onde

$$\hat{F}_n(t) = r'_n \cdot F_0(t) + (1 - r'_n) \cdot F_n(t) \quad (2.4.14)$$

é o estimador de F sob perda quadrática, sendo

$$r'_n = \frac{\alpha(-\infty, t]}{\alpha(\mathbb{R}) + n}, \quad (2.4.15)$$

$$F_0(t) = \frac{\alpha(-\infty, t]}{\alpha(\mathbb{R})} \quad (2.4.16)$$

é a média da distribuição a priori e $F_n(t)$ é a função de distribuição empírica.

Exemplo 2.4.4

Sejam F e G duas funções de distribuição sobre $\mathbb{R}$ desconhecidas e mutuamente independentes, sejam $X_1, \dots, X_m$ uma amostra de F e $Y_1, \dots, Y_n$ uma amostra de G , mutuamente independentes. O problema é estimar $\Delta = \Pr(X_1 \leq Y_1) = \int F dG$, sob perda quadrática $L(\Delta, \hat{\Delta}) = (\Delta - \hat{\Delta})^2$.

Consideramos como distribuição a priori para (F, G) dois processos de Dirichlet independentes, i.é $F(t) = P_1(-\infty, t]$ e $G(t) = P_2(-\infty, t]$, onde $P_1 \in \mathcal{D}(\alpha_1)$ e $P_2 \in \mathcal{D}(\alpha_2)$, o

qual escrevemos $P_1 \amalg P_2 / (\alpha_1, \alpha_2)$.

Analogamente como nas Proposições 2.4.2 e 2.4.3, é possível mostrar que a regra de Bayes para problema não-amostrai é

$$\Delta_0 = \mathbb{E} \int F dG = \int F_0 dG_0 \quad (2.4.17)$$

onde F_0 e G_0 são como em (2.4.16).

Logo, a regra de Bayes para as duas amostras $X_1, \dots, X_m; Y_1, \dots, Y_n$ é:

$$\hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n) = \int \hat{F}_m d\hat{G}_n \quad (2.4.18)$$

sendo $\hat{F}_m$ e $\hat{G}_n$ como em (2.4.14), i.é: $\hat{F}_m(t) = r_m^1 F_0(t) + (1 - r_m^1) F_m(t)$ e $\hat{G}_n(t) = r_n^2 G_0(t) + (1 - r_n^2) G_n(t)$, $r_m^1 = \frac{\alpha_1(-\infty, t]}{\alpha_1(\mathbb{R}) + m}$ e $r_n^2 = \frac{\alpha_2(-\infty, t]}{\alpha_2(\mathbb{R}) + n}$, $F_0(t) = \frac{\alpha_1(-\infty, t]}{\alpha_1(\mathbb{R})}$ e $G_0(t) = \frac{\alpha_2(-\infty, t]}{\alpha_2(\mathbb{R})}$, e $F_m(t)$ e $G_n(t)$ são as respectivas funções de distribuições empírica.

Um cálculo simples mostra que:

$$\begin{aligned} \hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n) &= r_m^1 r_n^2 \Delta_0 + r_m^1 (1 - r_n^2) \frac{1}{n} \sum_{j=1}^n F_0(Y_j) \\ &\quad + (1 - r_m^1) r_n^2 \int (1 - G_0) dF_m + (1 - r_m^1) (1 - r_n^2) \frac{U_{mn}}{mn} \end{aligned}$$

sendo U_{mn} = número de pares (X_i, Y_j) tais que $X_i \leq Y_j$, que é a estatística de Mann-Whitney.

Notamos que se $\alpha_1(\mathbb{R})$ e $\alpha_2(\mathbb{R}) \rightarrow 0$ então

$$\hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n) \rightarrow \frac{1}{mn} U_{mn}$$

o estimador não-paramétrico usual.

2.4.1 Notas bibliográficas

Aplicações do processo de Dirichlet à Estatística Bayesiana Não-Paramétrica, incluindo a abordagem bayesiana empírica, têm sido muito frequentes nas últimas décadas.

O problema de estimação de uma função de distribuição (e de seus funcionais) tem sido de interesse de Korwar e Hollander (1976), Ferguson (1982), e Ghosh et al. (1989). Outros trabalhos relacionados são os de Tiwari e Zalkikar (1985), Tiwari e Jammalamadaka (1985), Zalkikar et al. (1986), e Hjort (1990). Problemas relativos à dependência no caso bivariado são estudados por Dalal e Phadia (1983). Inferências em modelos de locação baseadas sobre informação parcial são consideradas por Doksum e Lo (1990). Outras aplicações podem ser vistas em Ferguson (1983), Tiwari e Triphati (1990), e Lahiri e Tiwari (1990).

CAPÍTULO 3

Um processo de Dirichlet multivariado

3.1 Introdução

Neste capítulo introduzimos uma classe de processos de Dirichlet multivariado sobre um espaço mensurável $(\mathcal{X}, \mathcal{A})$. O objetivo fundamental da construção desta classe de processos é o de resolver o problema de risco competitivo, o qual foi apresentado na introdução deste trabalho, sob uma abordagem bayesiana não-paramétrica.

Para uma maior compreensão e simplicidade nos cálculos, consideramos primeiro a construção de um processo de Dirichlet bivariado. Brevemente, esse processo pode ser descrito como segue. Seja $(\mathcal{X}, \mathcal{A})$ um espaço mensurável, e α_1, α_2 medidas finitas e não-nulas sobre $(\mathcal{X}, \mathcal{A})$. Um processo de Dirichlet bivariado $\mathbf{P}$ sobre $(\mathcal{X}, \mathcal{A})$ é uma família de vetores aleatórios bidimensionais definidos sobre algum espaço de probabilidade, $\{\mathbf{P}(A) : A \in \mathcal{A}\}$, tal que para toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, o vetor $2k$ -dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k))$ tem distribuição de Dirichlet singular com parâmetro $(\alpha_1(A_1), \alpha_2(A_1), \dots, \alpha_1(A_k), \alpha_2(A_k))$.

Duas provas da existência do processo de Dirichlet são dadas. Uma é não-constructiva, usa as condições de consistência do Teorema de Kolmogorov e é dada sem maiores detalhes, pois é completamente análoga à demonstração desenvolvida por Ferguson (1973). A outra definição é dada em função de dois processos de Dirichlet usuais e de uma certa variável aleatória com distribuição Beta. Esta última representação está inspirada no problema de estimação de uma função de distribuição (ou de sobrevivência), quando as observações são aleatoriamente censuradas. Mostramos também que o processo de Dirichlet bivariado satisfaz uma relação análoga à estabelecida por Sethuraman, relação que foi exposta na

seção 2.3.2 deste trabalho.

Dedicamos a última seção deste capítulo a estender os resultados do caso bivariado ao caso multivariado. Assim, estabelecemos o processo de Dirichlet multivariado.

3.2 Um processo de Dirichlet bivariado no contexto de Ferguson

No problema de estimação de uma função de distribuição (ou de sobrevivência), quando as observações são aleatoriamente censuradas, vemos sob uma abordagem não-paramétrica que, da função de verossimilhança o parâmetro é um vetor bidimensional de funções de subsobrevivências, tal que a soma dos seus componentes corresponde à função de sobrevivência da variável efetivamente observada.

Do ponto de vista bayesiano, podemos pensar em uma distribuição a priori para o modelo anterior, considerando um processo estocástico definido sobre um certo espaço de probabilidade $(\Omega, \mathcal{F}, \eta)$, sendo o conjunto de índices a σ -álgebra $\mathcal{A}$, e assumindo valores em $\mathcal{S}_2 = \{(x_1, x_2) : x_1, x_2 \geq 0, x_1 + x_2 \leq 1\}$, o simplex de $\mathbb{R}^2$. Assim, temos uma família de vetores aleatórios bidimensionais $\{\mathbf{P}(A) : A \in \mathcal{A}\}$, de tal maneira que para cada $\omega \in \Omega$, a função $\mathbf{P}(\omega, \cdot) : \mathcal{A} \rightarrow \mathcal{S}_2$, é uma função de conjuntos sobre $(\mathcal{X}, \mathcal{A})$ assumindo valores vetoriais no simplex $\mathcal{S}_2$ e para cada $A \in \mathcal{A}$, $\mathbf{P}(\cdot, A) : \Omega \rightarrow \mathcal{S}_2$ é um vetor aleatório bidimensional.

Fazendo uma analogia com o exposto na seção 2.2 deste trabalho, podemos entender esta classe de processos como uma função de conjuntos aleatória definida sobre o espaço mensurável $(\mathcal{X}, \mathcal{A})$.

Da mesma forma como foi construído o processo de Dirichlet por Ferguson (1973) (ver seção 2.3.1), construímos uma função de conjuntos aleatória sobre $(\mathcal{X}, \mathcal{A})$, definindo a distribuição do vetor aleatório $2k$ -dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k))$, para todo inteiro $k > 0$ e toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$. Logo é possível definir a distribuição

do vetor aleatório $(P(B_1), \dots, P(B_m))$, onde $B_1, \dots, B_m$ são conjuntos arbitrários de $\mathcal{A}$. Este último é feito usando as equações 2.3.1 - 2.3.3.

Uma condição de consistência similar à condição da seção 2.3.1, pode ser estabelecida para nosso caso.

Denotamos por $\mathcal{S}_2^{\mathcal{A}}$ o espaço produto, onde cada fator é $\mathcal{S}_2 = \{(x_1, x_2) : x_1, x_2 \geq 0; x_1 + x_2 \leq 1\}$ o conjunto simplex de $\mathbb{R}^2$. Equivalentemente, $\mathcal{S}_2^{\mathcal{A}}$ pode ser visto como a classe de funções de conjuntos definidas sobre $\mathcal{A}$ e assumindo valores em $\mathcal{S}_2$. Também, $\sigma(\mathcal{S}_2^{\mathcal{A}})$ denota a σ -álgebra produto para $\mathcal{S}_2^{\mathcal{A}}$, i.é a σ -álgebra gerada pelos cilindros mensuráveis tendo uma base finita.

Assim, acompanhando o enunciado do Lema 2.3.1 com as respectivas modificações, concluímos o seguinte:

Lema 3.2.1

Existe uma medida de probabilidade sobre $(\mathcal{S}_2^{\mathcal{A}}, \sigma(\mathcal{S}_2^{\mathcal{A}}))$ tal que o processo variável-coordenada $\{Z_A : A \in \mathcal{A}\}$,

$$\begin{aligned} Z_A : \mathcal{S}_2^{\mathcal{A}} &\rightarrow \mathcal{S}_2 \\ P &\mapsto P(A) , \end{aligned}$$

tem o sistema de distribuições original como o seu sistema de distribuições finito-dimensionais.

Prova: Análoga à do Lema 2.3.1. □

A definição rigorosa do nosso processo de Dirichlet bivariado é como segue:

Definição 3.2.1

Sejam α_1, α_2 medidas finitas, não-nulas, não negativas e finitamente aditivas sobre $(\mathcal{X}, \mathcal{A})$. Dizemos que a função de conjuntos aleatória P definida sobre $(\mathcal{X}, \mathcal{A})$ e assu-

mando valores no simplex $\mathcal{S}_2$ de $\mathbb{R}^2$, é um processo de Dirichlet bivariado sobre $(\mathcal{X}, \mathcal{A})$ com parâmetro (α_1, α_2) , se para todo inteiro $k > 0$ e para toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, a distribuição do vetor $2k$ -dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k))$ é Dirichlet singular com parâmetro $(\alpha_1(A_1), \alpha_2(A_2), \dots, \alpha_1(A_k), \alpha_2(A_k))$, o qual escrevemos $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k)) \sim \mathcal{D}_s(\alpha_1(A_1), \alpha_2(A_1), \dots, \alpha_1(A_k), \alpha_2(A_k))$. **Notação:** $\mathbf{P} \in \mathcal{DB}(\alpha_1, \alpha_2)$.

Observações 3.2.1

- (a) O Lema 3.2.1 diz que existe uma medida de probabilidade sobre $(\mathcal{S}_2^{\mathcal{A}}, \sigma(\mathcal{S}_2^{\mathcal{A}}))$ sustentando as distribuições finito-dimensionais do processo de Dirichlet bivariado.
- (b) Uma versão não-singular da definição do processo de Dirichlet bivariado, obtém-se considerando a distribuição do vetor $2(k-1)$ -dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_{k-1}))$, a qual será Dirichlet com parâmetro $(\alpha_1(A_1), \alpha_2(A_1), \dots, \alpha_1(A_{k-1}), \alpha_2(A_{k-1}); \alpha_1(A_k) + \alpha_2(A_k))$, i.é, $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_{k-1})) \sim \mathcal{D}(\alpha_1(A_1), \alpha_2(A_1), \dots, \alpha_1(A_{k-1}), \alpha_2(A_{k-1}); \alpha_1(A_k) + \alpha_2(A_k))$.
- (c) Decorre-se, também, que no caso de considerar $\mathcal{X} = A \cup A^c$, então o vetor bidimensional $\mathbf{P}(A)$ possui distribuição de Dirichlet com parâmetro $(\alpha_1(A), \alpha_2(A); \alpha_1(A^c) + \alpha_2(A^c))$, i.é, $\mathbf{P}(A) \sim \mathcal{D}(\alpha_1(A), \alpha_2(A); \alpha_1(A^c) + \alpha_2(A^c))$.
- (d) De acordo com a nossa conveniência, principalmente nos cálculos, as formas singular e não-singular da distribuição de Dirichlet, são usadas indistintamente.

3.3 Um processo de Dirichlet bivariado a partir de dois processos de Dirichlet e de uma variável aleatória Beta

Nesta seção, consideramos o problema de construir um processo de Dirichlet bivariado baseando-nos em dois processos de Dirichlet independentes, e de uma certa variável

aleatória com distribuição Beta, também independente dos processos anteriores.

Esta representação do processo de Dirichlet bivariado está inspirada no problema de estimação de uma função de distribuição (ou de sobrevivência), quando as observações são aleatoriamente censuradas e será portanto de utilidade na resolução desse problema. Em primeira instância, nosso parâmetro de interesse é um vetor bidimensional de funções de subsobrevivências, tal que a soma dos seus componentes é a função de sobrevivência da variável efetivamente observada.

Do ponto de vista bayesiano, a idéia básica é construir uma função mensurável

$$\mathbf{P} : (\Omega, \mathcal{F}, \eta) \rightarrow (\mathcal{P}^*, \sigma(\mathcal{P}^*))$$

onde $(\Omega, \mathcal{F}, \eta)$ é um certo espaço de probabilidade, e $\mathcal{P}^* = \{(P_1^*, P_2^*) : P_1^*, P_2^* \text{ são submedidas de probabilidade não-nulas sobre } (\mathcal{X}, \mathcal{A}), P_1^* + P_2^* \text{ é uma medida de probabilidade sobre } (\mathcal{X}, \mathcal{A})\}$, sendo $\sigma(\mathcal{P}^*)$ a menor σ -álgebra tal que a função $\mathbf{P}^* \in \mathcal{P}^* \mapsto \mathbf{P}^*(A) \in \mathcal{S}_2$ é $\sigma(\mathcal{P}^*)$ -mensurável para todo $A \in \mathcal{A}$; tal que a medida induzida por $\mathbf{P}$, digamos $\eta^{\mathbf{P}^{-1}}$, é a nossa medida de Dirichlet bivariada.

Equivalentemente, podemos pensar em um certo processo estocástico tendo como conjunto de índices a σ -álgebra $\mathcal{A}$, i.é, uma família de vetores aleatórios $\{\mathbf{P}(A) : A \in \mathcal{A}\}$ tal que para cada $\omega \in \Omega$, $\mathbf{P}(\omega, \cdot)$ é um elemento de $\mathcal{P}^*$, i.é, $\mathbf{P}(\omega, \cdot) = (P_1^*(\omega, \cdot), P_2^*(\omega, \cdot))$, sendo $P_1^*(\omega, \cdot)$ e $P_2^*(\omega, \cdot)$ submedidas de probabilidade não-nulas sobre $(\mathcal{X}, \mathcal{A})$ com $P_1^*(\omega, \cdot) + P_2^*(\omega, \cdot)$ medida de probabilidade sobre $(\mathcal{X}, \mathcal{A})$, e para cada $A \in \mathcal{A}$, $\mathbf{P}(\cdot, A)$ é um vetor aleatório bidimensional a valores em $\mathcal{S}_2$.

Sejam $(\mathcal{X}, \mathcal{A})$ nosso espaço mensurável, e α_1, α_2 as medidas sobre $(\mathcal{X}, \mathcal{A})$ como descritas na Definição 3.2.1.

De acordo com a construção de Sethuraman (ver seção 2.3.2 deste trabalho), consideramos o seguinte:

Sejam $\mathbf{W}^i = \{(U_1^i, \theta_1^i), (U_2^i, \theta_2^i), \dots\}$, $i = 1, 2$, duas sequências de vetores aleatórios bidimensionais tais que:

- (i) As variáveis aleatórias $(U_1^i, U_2^i, \dots)$ são independentes e identicamente distribuídas, assumindo valores em $\mathcal{X}$, com $U_1^i \sim \alpha_i(\cdot)/\alpha_i(\mathcal{X})$, $i = 1, 2$.
- (ii) As variáveis aleatórias $(\theta_1^i, \theta_2^i, \dots)$ são independentes e identicamente distribuídas, assumem valores em $[0, 1]$, com $\theta_1^i \sim \text{Beta}(1, \alpha_i(\mathcal{X}))$, $i = 1, 2$.
- (iii) As variáveis aleatórias $\{U_j^1\}$, $\{U_j^2\}$, $\{\theta_j^1\}$ e $\{\theta_j^2\}$ são mutuamente independentes.

Consideramos então o espaço de probabilidade produto $(\Omega, \sigma(\Omega), \eta_{\alpha_1, \alpha_2})$, onde $\Omega = (\mathcal{X} \times [0, 1])^\infty \times (\mathcal{X} \times [0, 1])^\infty$, i.é, o espaço das realizações de $\mathbf{W} = (\mathbf{W}^1, \mathbf{W}^2)$, $\sigma(\Omega)$ é a σ -álgebra produto $(\mathcal{A} \times \mathcal{B}([0, 1]))^\infty \times (\mathcal{A} \times \mathcal{B}([0, 1]))^\infty$, e $\eta_{\alpha_1, \alpha_2}$ é a lei de $\mathbf{W}$.

Seja ρ uma variável aleatória definida sobre $(\Omega, \sigma(\Omega), \eta_{\alpha_1, \alpha_2})$ independente de $\mathbf{W}^i$, $i = 1, 2$, tal que $\rho \sim \text{Beta}(\alpha_1(\mathcal{X}), \alpha_2(\mathcal{X}))$.

Sejam $p_1^i = \theta_1^i$, $p_j^i = \theta_j^i \prod_{\ell=1}^{j-1} (1 - \theta_\ell^i)$, $j \geq 2$; $i = 1, 2$.

Para cada $\omega \in \Omega$ e para todo $A \in \mathcal{A}$, definimos

$$P_i(\omega, A) = \sum_{j=1}^{\infty} p_j^i(\omega) \delta_{U_j^i(\omega)}(A), \quad i = 1, 2,$$

e logo de acordo com a construção de Sethuraman (cf. Seção 2.3.2), P_1 e P_2 definem dois processos de Dirichlet sobre $(\mathcal{X}, \mathcal{A})$, sendo P_1 e P_2 independentes.

Notamos que os elementos aleatórios ρ , P_1 e P_2 , assim definidos, são mutuamente independentes, o que escrevemos $\rho \amalg P_1 \amalg P_2 / (\alpha_1, \alpha_2)$.

Assim, a representação do processo de Dirichlet bivariado é dada no seguinte teorema.

Teorema 3.3.1

Sejam $(\mathcal{X}, \mathcal{A})$ espaço mensurável, α_1 e α_2 medidas finitas, não-nulas, não-negativas e finitamente aditivas sobre $(\mathcal{X}, \mathcal{A})$.

Sejam ρ, P_1 e P_2 elementos aleatórios definidos sobre o espaço de probabilidade $(\Omega, \sigma(\Omega), \eta_{\alpha_1, \alpha_2})$, tais que $\rho \amalg P_1 \amalg P_2 / (\alpha_1, \alpha_2)$, sendo ρ uma variável aleatória com distri-

buição Beta de parâmetros $\alpha_1(\mathcal{X})$ e $\alpha_2(\mathcal{X})$, $\rho \sim \text{Beta}(\alpha_1(\mathcal{X}), \alpha_2(\mathcal{X}))$, e $P_i \in \mathcal{D}(\alpha_i)$, $i = 1, 2$.

Seja $\mathbf{P}$ a função mensurável de $(\Omega, \sigma(\Omega), \eta_{\alpha_1, \alpha_2})$ em $(\mathcal{P}^*, \sigma(\mathcal{P}^*))$, ou equivalentemente o processo estocástico bidimensional $\{\mathbf{P}(A) : A \in \mathcal{A}\}$ definido por

$$\mathbf{P}(\omega, A) = \left(\rho(\omega)P_1(\omega, A), (1 - \rho)(\omega)P_2(\omega, A) \right), \quad (3.3.1)$$

onde $\omega \in \Omega$, $A \in \mathcal{A}$.

Então $\mathbf{P} \in \mathcal{DB}(\alpha_1, \alpha_2)$.

Prova

A prova do teorema é feita em várias etapas.

Notamos primeiro que $\mathbf{P}$ é bem definida, i.é, sendo ρ uma variável Beta e P_1 e P_2 processos de Dirichlet então claramente para cada $\omega \in \Omega$, os componentes de $\mathbf{P}(\omega, \cdot)$ são submedidas de probabilidade sobre $(\mathcal{X}, \mathcal{A})$, tal que a sua soma é uma medida de probabilidade sobre $(\mathcal{X}, \mathcal{A})$, portanto, $\mathbf{P}(\omega, \cdot)$ é um elemento de $\mathcal{P}^*$.

Notamos que a variável aleatória $\mathbf{P}(\cdot, A)$, para cada $A \in \mathcal{A}$, assume valores em $\mathcal{S}_2$ o conjunto simplex de $\mathbb{R}^2$.

Agora consideramos partições mensuráveis do espaço $\mathcal{X}$ em ordem crescente de refinação, e no que segue denotamos $\mathbf{P}(A)$ em vez de $\mathbf{P}(\cdot, A)$, o mesmo para P_1 e P_2 .

(a) $\mathcal{X} = A \cup A^c$, $A \in \mathcal{A}$.

Neste caso devemos provar que

$$\mathbf{P}(A) = (\rho P_1(A), (1 - \rho)P_2(A)) \sim \mathcal{D}(\alpha_1(A), \alpha_2(A); \alpha_1(A^c) + \alpha_2(A^c)) \quad (3.3.2)$$

Mostramos que

$$(\rho P_1(A), \rho(1 - P_1(A)), (1 - \rho)P_2(A)) \sim \mathcal{D}(\alpha_1(A), \alpha_1(A^c), \alpha_2(A); \alpha_2(A^c)) \quad (3.3.3)$$

ou na sua forma singular

$$(\rho P_1(A), \rho(1 - P_1(A)), (1 - \rho)P_2(A), (1 - \rho)(1 - P_2(A))) \sim \mathcal{D}_s(\alpha_1(A), \alpha_1(A^c), \alpha_2(A), \alpha_2(A^c)) \quad (3.3.4)$$

Tendo-se que $1 - \rho \sim \text{Beta}(\alpha_1(\mathcal{X}), \alpha_2(\mathcal{X}))$, $P_1(A) \sim \text{Beta}(\alpha_1(A), \alpha_1(A^c))$, e $1 - \rho \perp\!\!\!\perp P_1(A) / (\alpha_1, \alpha_2)$, segue da Proposição 4 do Apêndice A,

$$(1 - \rho, \rho P_1(A)) \sim \mathcal{D}(\alpha_2(\mathcal{X}), \alpha_1(A); \alpha_1(A^c)) \quad (3.3.5)$$

ou equivalentemente, $(\rho P_1(A), 1 - \rho) \sim \mathcal{D}(\alpha_1(A), \alpha_2(A) + \alpha_2(A^c); \alpha_1(A^c))$.

Usando a Proposição 3 do Apêndice A, temos que as variáveis aleatórias

$$\begin{aligned} \phi_1 &= \rho P_1(A) \sim \text{Beta}(\alpha_1(A), \alpha_1(A^c) + \alpha_2(A) + \alpha_2(A^c)) \\ \phi_2 &= \frac{1 - \rho}{1 - \rho P_1(A)} \sim \text{Beta}(\alpha_2(A) + \alpha_2(A^c), \alpha_1(A^c)) \end{aligned}$$

e são mutuamente independentes.

Consideramos então as variáveis aleatórias

$$\begin{aligned} \psi_1 &= \phi_1 = \rho P_1(A) \sim \text{Beta}(\alpha_1(A), \alpha_1(A^c) + \alpha_2(A) + \alpha_2(A^c)) \\ \psi_2 &= 1 - \phi_2 = \frac{\rho(1 - P_1(A))}{1 - \rho P_1(A)} \sim \text{Beta}(\alpha_1(A^c), \alpha_2(A) + \alpha_2(A^c)) \\ \psi_3 &= P_2(A) \sim \text{Beta}(\alpha_2(A), \alpha_2(A^c)) \end{aligned} \quad (3.3.6)$$

que também são mutuamente independentes.

Novamente, fazendo uso da Proposição 4 do Apêndice A, definimos as variáveis aleatórias $Y_1 = \psi_1 = \rho P_1(A)$, $Y_2 = (1 - \psi_1)\psi_2 = \rho(1 - P_1(A))$, e $Y_3 = (1 - \psi_1)(1 - \psi_2)\psi_3 = (1 - \rho)P_2(A)$, e obtemos que $(Y_1, Y_2, Y_3) \sim \mathcal{D}(\alpha_1(A), \alpha_1(A^c), \alpha_2(A); \alpha_2(A^c))$ que corresponde a (3.3.3).

(b) $\mathcal{X} = A_1 \cup A_2 \cup A_3$, onde $\{A_1, A_2, A_3\}$ é uma partição mensurável de $\mathcal{X}$.

Neste caso devemos provar que

$$\begin{aligned} (\mathbf{P}(A_1), \mathbf{P}(A_2)) &= (\rho P_1(A_1), (1 - \rho)P_2(A_1), \rho P_1(A_2), (1 - \rho)P_2(A_2)) \\ &\sim \mathcal{D}(\alpha_1(A_1), \alpha_2(A_1), \alpha_1(A_2), \alpha_2(A_2); \alpha_1(A_3) + \alpha_2(A_3)) \end{aligned} \quad (3.3.7)$$

Mostramos que

$$\begin{aligned} & (\rho P_1(A_1), \rho P_1(A_2), \rho P_1(A_3), (1 - \rho)P_2(A_1), (1 - \rho)P_2(A_2)) \\ & \sim \mathcal{D}(\alpha_1(A_1), \alpha_1(A_2), \alpha_1(A_3), \alpha_2(A_1); \alpha_2(A_2) + \alpha_2(A_3)) \end{aligned} \quad (3.3.8)$$

ou na sua forma singular,

$$\begin{aligned} & (\rho P_1(A_1), \rho P_1(A_2), \rho P_1(A_3), (1 - \rho)P_2(A_1), (1 - \rho)P_2(A_2), (1 - \rho)P_2(A_3)) \\ & \sim \mathcal{D}_s(\alpha_1(A_1), \alpha_1(A_2), \alpha_1(A_3), \alpha_2(A_1), \alpha_2(A_2), \alpha_2(A_3)) \end{aligned} \quad (3.3.9)$$

Tendo-se que $1 - \rho \sim \text{Beta}(\alpha_2(\mathcal{X}), \alpha_1(\mathcal{X}))$, $(P_1(A_1), P_1(A_2)) \sim \mathcal{D}(\alpha_1(A_1), \alpha_1(A_2); \alpha_1(A_3))$ e $1 - \rho \perp\!\!\!\perp (P_1(A_1), P_1(A_2)) / (\alpha_1, \alpha_2)$, segue da Proposição 5 do Apêndice A que,

$$(1 - \rho, \rho(P_1(A_1), P_1(A_2))) \sim \mathcal{D}(\alpha_2(\mathcal{X}), \alpha_1(A_1), \alpha_1(A_2); \alpha_1(A_3)) \quad (3.3.10)$$

ou equivalentemente,

$$(\rho P_1(A_1), \rho P_1(A_2), 1 - \rho) \sim \mathcal{D}(\alpha_1(A_1), \alpha_1(A_2), \alpha_2(\mathcal{X}); \alpha_1(A_3))$$

Usando a Proposição 3 do Apêndice A obtemos que as variáveis aleatórias

$$\begin{aligned} \phi_1 &= \rho P_1(A_1) \sim \text{Beta}(\alpha_1(A_1), \alpha_1(A_2) + \alpha_1(A_3) + \alpha_2(\mathcal{X})) \\ \phi_2 &= \frac{\rho P_1(A_2)}{1 - \rho P_1(A_1)} \sim \text{Beta}(\alpha_1(A_2), \alpha_1(A_3) + \alpha_2(\mathcal{X})) \\ \phi_3 &= \frac{1 - \rho}{1 - \rho P_1(A_1) - \rho P_1(A_2)} \sim \text{Beta}(\alpha_2(\mathcal{X}), \alpha_1(A_3)) \end{aligned}$$

e são mutuamente independentes.

Consideremos então as variáveis aleatórias

$$\begin{aligned} \psi_1 &= \rho P_1(A_1) \sim \text{Beta}(\alpha_1(A_1), \alpha_1(A_2) + \alpha_1(A_3) + \alpha_2(A_1) + \alpha_2(A_2) + \alpha_2(A_3)) \\ \psi_2 &= \frac{\rho P_1(A_2)}{1 - \rho P_1(A_1)} \sim \text{Beta}(\alpha_1(A_2), \alpha_1(A_3) + \alpha_2(A_1) + \alpha_2(A_2) + \alpha_2(A_3)) \\ \psi_3 &= 1 - \phi_3 = \frac{\rho(1 - P_1(A_1) - P_1(A_2))}{1 - \rho P_1(A_1) - \rho P_1(A_2)} \sim \text{Beta}(\alpha_1(A_3), \alpha_2(A_1) + \alpha_2(A_2) + \alpha_2(A_3)) \\ \psi_4 &= P_2(A_1) \sim \text{Beta}(\alpha_2(A_1), \alpha_2(A_2) + \alpha_2(A_3)) \\ \psi_5 &= \frac{P_2(A_2)}{1 - P_2(A_1)} \sim \text{Beta}(\alpha_2(A_2), \alpha_2(A_3)) \end{aligned} \quad (3.3.11)$$

que são mutuamente independentes.

Mais uma vez, usando a Proposição 4 do Apêndice A, definimos as variáveis aleatórias $Y_1 = \psi_1 = \rho P_1(A_1)$, $Y_2 = (1 - \psi_1)\psi_2 = \rho P_1(A_2)$, $Y_3 = (1 - \psi_1)(1 - \psi_2)\psi_3 = \rho P_1(A_3)$, $Y_4 = (1 - \psi_1)(1 - \psi_2)(1 - \psi_3)\psi_4 = (1 - \rho)P_2(A_1)$, $Y_5 = (1 - \psi_1)(1 - \psi_2)(1 - \psi_3)(1 - \psi_4) = (1 - \rho)P_2(A_2)$, e obtemos que

$$(Y_1, Y_2, Y_3, Y_4, Y_5) \sim \mathcal{D}(\alpha_1(A_1), \alpha_1(A_2), \alpha_1(A_3), \alpha_2(A_1), \alpha_2(A_2); \alpha_2(A_3))$$

de onde segue o resultado.

(c) Consideramos agora o caso geral, $\mathcal{X} = \bigcup_{i=1}^k A_i$, $\{A_1, \dots, A_k\}$ partição mensurável de $\mathcal{X}$.

Usando o mesmo raciocínio anterior, notamos que $1 - \rho \sim \text{Beta}(\alpha_2(\mathcal{X}), \alpha_1(\mathcal{X}))$, $(P_1(A_1), \dots, P_1(A_{k-1})) \sim \mathcal{D}(\alpha_1(A_1), \dots, \alpha_1(A_{k-1}); \alpha_1(A_k))$ e que $1 - \rho \amalg (P_1(A_1), \dots, P_1(A_{k-1})) / (\alpha_1, \alpha_2)$.

É possível mostrar então que

$$\begin{aligned} & (\rho P_1(A_1), \dots, \rho P_1(A_k), (1 - \rho)P_2(A_1), \dots, (1 - \rho)P_2(A_k)) \\ & \sim \mathcal{D}_s(\alpha_1(A_1), \dots, \alpha_1(A_k), \alpha_2(A_1), \dots, \alpha_2(A_k)) \end{aligned} \quad (3.3.12)$$

isto, considerando as seguintes variáveis aleatórias ψ_j 's, mutuamente independentes, $(2k - 1)$ em total

$$\begin{aligned} \psi_1 &= \rho P_1(A_1) \sim \text{Beta} \left(\alpha_1(A_1), \sum_{i=2}^k \alpha_1(A_i) + \sum_{i=1}^k \alpha_2(A_i) \right) \\ \psi_2 &= \frac{\rho P_1(A_2)}{1 - \rho P_1(A_1)} \sim \text{Beta} \left(\alpha_2(A_2), \sum_{i=3}^k \alpha_1(A_i) + \sum_{i=1}^k \alpha_2(A_i) \right) \\ &\dots \\ \psi_k &= \frac{\rho P_1(A_k)}{1 - \rho \sum_{i=1}^{k-1} P_1(A_i)} \sim \text{Beta} \left(\alpha_1(A_k), \sum_{i=1}^k \alpha_2(A_i) \right) \\ \psi_{k+1} &= P_2(A_1) \sim \text{Beta} \left(\alpha_2(A_1), \sum_{i=2}^k \alpha_2(A_i) \right) \\ \psi_{k+2} &= \frac{P_2(A_2)}{1 - \sum_{i=1}^{k-2} P_2(A_i)} \sim \text{Beta} \left(\alpha_2(A_2), \sum_{i=3}^k \alpha_2(A_i) \right) \end{aligned} \quad (3.3.13)$$

...

$$\psi_{2k-1} = \frac{P_2(A_{k-1})}{1 - \sum_{i=1}^{k-2} P_2(A_i)} \sim \text{Beta}(\alpha_2(A_{k-1}), \alpha_2(A_k))$$

Definindo $Y_1 = \psi_1$, e $Y_j = \prod_{i=1}^{j-1} (1 - \psi_i) \psi_j$ $j = 2, \dots, 2k - 1$ obtemos que

$$(Y_1, Y_2, \dots, Y_{2k-1}) \sim \mathcal{D}(\alpha_1(A_1), \dots, \alpha_1(A_k), \alpha_2(A_1), \dots, \alpha_2(A_{k-1}); \alpha_2(A_k)),$$

o que implica

$$(\mathbf{P}(A_1), \mathbf{P}(A_2), \dots, \mathbf{P}(A_k)) \sim \mathcal{D}_s(\alpha_1(A_1), \alpha_2(A_1), \dots, \alpha_1(A_k), \alpha_2(A_k))$$

o que completa a prova do teorema. $\square$

Baseando-nos na construção de Sethuraman do processo de Dirichlet (ver seção 2.3.2 deste trabalho), mostramos na proposição seguinte que é possível estabelecer uma relação do tipo “ $\sim$ ” para o processo de Dirichlet bivariado como construído no Teorema 3.3.1.

Proposição 3.3.1

Seja $\mathbf{P} \in \mathcal{DB}(\alpha_1, \alpha_2)$ como no Teorema 3.3.1.

Então, se $A \in \mathcal{A}$, tem-se que $\mathbf{P}(A)$ é a única solução da relação

$$\mathbf{P}(A) \sim (\rho, 1 - \rho) \bullet (p_1^1 \delta_{U_1^1}(A), p_1^2 \delta_{U_1^2}(A)) + (1 - p_1^1, 1 - p_1^2) \bullet \mathbf{P}(A) \quad (3.3.14)$$

onde as variáveis aleatórias U_1^i , p_1^i , $i = 1, 2$, são como em (i)-(iii) desta seção e $(x_1, x_2) \bullet (y_1, y_2) = (x_1 y_1, x_2 y_2)$ é o produto direto entre vetores.

Prova. Consideramos $\mathbf{P}(A) = (\rho P_1(A), (1 - \rho) P_2(A))$, onde $P_i \in \mathcal{D}(\alpha_i)$, $i = 1, 2$, $\rho \sim \text{Beta}(\alpha_1(\mathcal{X}), \alpha_2(\mathcal{X}))$, tais que $\rho \amalg P_1 \amalg P_2 / (\alpha_1, \alpha_2)$. Da construção de Sethuraman do processo de Dirichlet, temos que $P_i(A)$ é a única solução da relação $P_i(A) \sim p_1^i \delta_{U_1^i}(A) + (1 - p_1^i) P_i(A)$, $i = 1, 2$. Então usando o Lema 2.3.2 e os argumentos usados na demonstração do Teorema 3.3.1, obtemos que $\rho P_1(A)$ é a única solução à relação $\rho P_1(A) \sim \rho p_1^1 \delta_{U_1^1}(A) +$

$\rho(1 - p_1^1)P_1(A)$, e analogamente $(1 - \rho)P_2(A)$ é a única solução à relação $(1 - \rho)P_2(A) \sim (1 - \rho)p_1^2\delta_{U_1^2}(A) + (1 - \rho)(1 - p_1^2)P_2(A)$.

Do Teorema 3.3.1 concluímos (3.3.14) e logo a demonstração. $\square$

Observação 3.3.1

A generalização da Proposição 3.3.1, no caso de considerar uma partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, é imediata. Assim, o vetor $2k$ -dimensional

$$(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k)) = (\rho P_1(A_1), (1 - \rho)P_2(A_1), \dots, \rho P_1(A_k), (1 - \rho)P_2(A_k))$$

é a única solução da relação

$$\begin{aligned} (\mathbf{P}(A_1), \dots, \mathbf{P}(A_k)) \sim & (\rho, 1 - \rho, \dots, \rho, 1 - \rho) \bullet (p_1\delta_{U_1}(A_1), q_1\delta_{V_1}(A_1), \dots, \\ & p_1\delta_{U_1}(A_k), q_1\delta_{V_1}(A_k)) + (1 - p_1, 1 - q_1, \dots, 1 - p_1, 1 - q_1) \bullet (\mathbf{P}(A_1), \dots, \mathbf{P}(A_k)) \end{aligned} \quad (3.3.15)$$

3.4 A extensão ao processo de Dirichlet multivariado

Esta seção está dedicada fundamentalmente a estender os resultados do caso bivariado ao caso multivariado. Assim, nosso objetivo é o de estabelecer uma classe de processos de Dirichlet multivariado. Sendo que a maior parte dos desenvolvimentos são análogos aos feitos nas seções anteriores, os resultados serão estabelecidos sem provas. Principalmente nos dedicamos a definir uma notação adequada para este caso.

Primeiramente podemos considerar um processo de Dirichlet multivariado, da maneira como foi construído o processo bivariado na seção 3.2. Assim, seja $\{\mathbf{P}(A) : A \in \mathcal{A}\}$ uma família de vetores aleatórios r -dimensionais sobre um certo espaço de probabilidade $(\Omega, \mathcal{F}, \eta)$, tal que para cada $\omega \in \Omega$, $\mathbf{P}(\omega, \cdot)$ é uma função de conjuntos definida sobre a σ -álgebra $\mathcal{A}$ e assumindo valores no simplex r -dimensional

$$\mathcal{S}_r = \{(x_1, \dots, x_r) : x_i \geq 0 \quad \forall i, \quad \sum_{i=1}^r x_i \leq 1\} \quad \text{de } \mathbb{R}^r,$$

e tal que para cada $A \in \mathcal{A}$, $\mathbf{P}(\cdot, A)$ é um vetor aleatório r -dimensional assumindo valores no simplex $\mathcal{S}_r$.

Sejam $\alpha_1, \alpha_2, \dots, \alpha_r$ r medidas finitas, não-nulas, não-negativas e finitamente aditivas sobre $(\mathcal{X}, \mathcal{A})$, e consideramos a função de conjunto e a valores vetoriais, $\vec{\alpha} : \mathcal{A} \rightarrow [0, \infty)^r$, dada por $\vec{\alpha}(A) = (\alpha_1(A), \alpha_2(A), \dots, \alpha_r(A))$, onde $[0, \infty)^r = \underbrace{[0, \infty) \times [0, \infty) \times \dots \times [0, \infty)}_{r \text{ vezes}}$.

A definição do processo de Dirichlet r -variado é como segue.

Definição 3.4.1

Consideramos as r medidas $\alpha_1, \dots, \alpha_r$ sobre $(\mathcal{X}, \mathcal{A})$.

A função de conjuntos aleatória $\mathbf{P}$, definida sobre $(\mathcal{X}, \mathcal{A})$ e assumindo valores no simplex $\mathcal{S}_r$ de $\mathbb{R}^r$, é um processo de Dirichlet r -variado sobre $(\mathcal{X}, \mathcal{A})$ com parâmetro $(\alpha_1, \dots, \alpha_r)$, se para todo inteiro $k > 0$ e para toda partição mensurável $\{A_1, \dots, A_k\}$ de $\mathcal{X}$, a distribuição do vetor rk -dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k))$ é Dirichlet singular com parâmetro $(\vec{\alpha}(A_1), \dots, \vec{\alpha}(A_k))$, i.é, $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_k)) \sim \mathcal{D}_s(\vec{\alpha}(A_1), \dots, \vec{\alpha}(A_k))$.

Notação: $\mathbf{P} \in \mathcal{DM}_r(\vec{\alpha})$. □

Observações 3.4.1

- (a) Uma condição de consistência análoga à Condição C da construção de Ferguson (ver seção 2.3.1), junto ao Teorema de Kolmogorov, garante que existe uma medida de probabilidade sobre $(\mathcal{S}_r^{\mathcal{A}}, \sigma(\mathcal{S}_r^{\mathcal{A}}))$ sustentando as distribuições finito-dimensionais do processo de Dirichlet r -variado.
- (b) É possível obter uma versão não-singular da definição do processo de Dirichlet r -variado, considerando a distribuição do vetor $r(k-1)$ dimensional $(\mathbf{P}(A_1), \dots, \mathbf{P}(A_{k-1}))$, a qual será de Dirichlet com parâmetro $(\vec{\alpha}(A_1), \dots, \vec{\alpha}(A_{k-1}); \sum_{i=1}^r \alpha_i(A_k))$.
- (c) Decorre, também, que no caso de considerar $\mathcal{X} = A \cup A^c$, o vetor r -dimensional

$\mathbf{P}(A)$ possui distribuição de Dirichlet com parâmetro $(\vec{\alpha}(A); \sum_{i=1}^r \alpha_i(A^c))$.

- (d) O processo de Dirichlet multivariado pode ser visto como uma extensão do processo de Dirichlet usual.

Consideramos $r = 1$, então a função de conjuntos $\vec{\alpha}$ transforma-se em uma medida escalar, digamos α , sobre o espaço $(\mathcal{X}, \mathcal{A})$. Também, a função de conjuntos $\mathbf{P}$ passa a ser simplesmente uma medida de probabilidade aleatória sobre $(\mathcal{X}, \mathcal{A})$, digamos P (notemos que $\mathcal{S}_1$ é o intervalo fechado unitário $[0, 1]$). Assim, da Definição 3.4.1, obtemos a definição do processo de Dirichlet usual, i.é, para todo inteiro $k > 0$ e para toda partição mensurável $\{A_1, A_2, \dots, A_k\}$ de $\mathcal{X}$, $(P(A_1), \dots, P(A_k)) \sim \mathcal{D}_s(\alpha(A_1), \dots, \alpha(A_k))$.

Sendo o principal objetivo deste trabalho a resolução do problema de risco competitivo, sob uma ótica bayesiana, na última parte desta seção nos dedicamos a formalizar o processo de Dirichlet r -variado em termos de r processos de Dirichlet usuais independentes, e mutuamente independentes com um certo vetor aleatório com distribuição Dirichlet. Os raciocínios e a notação são análogos aos da Seção 3.3.

Consideremos $(\mathcal{X}, \mathcal{A})$ espaço mensurável, e a função $\vec{\alpha} : \mathcal{A} \rightarrow [0, \infty)^r$, dada por $\vec{\alpha}(A) = (\alpha_1(A), \alpha_2(A), \dots, \alpha_r(A))$, como no início desta seção.

Trata-se, então, de construir uma função mensurável $\mathbf{P}$, de um certo espaço de probabilidade $(\Omega, \mathcal{F}, \eta)$ no espaço $(\mathcal{P}^*, \sigma(\mathcal{P}^*))$, onde $\mathcal{P}^* = \{(P_1^*, P_2^*, \dots, P_r^*) : P_i^* \text{ é uma submedida de probabilidade não-nula sobre } (\mathcal{X}, \mathcal{A}) \quad \forall i = 1, 2, \dots, r; \text{ e } \sum_{i=1}^r P_i^* \text{ é uma medida de probabilidade sobre } (\mathcal{X}, \mathcal{A})\}$, sendo $\sigma(\mathcal{P}^*)$ a menor σ -álgebra tal que a função $\mathbf{P}^* \in \mathcal{P}^* \mapsto \mathbf{P}^*(A) \in \mathcal{S}_r$ é $\sigma(\mathcal{P}^*)$ -mensurável para todo $A \in \mathcal{A}$. De tal maneira que a medida induzida por $\mathbf{P}$, $\eta\mathbf{P}^{-1}$ é a medida de Dirichlet r -variada.

Equivalentemente, podemos pensar em um processo estocástico $\{\mathbf{P}(A) : A \in \mathcal{A}\}$ tal que para cada $\omega \in \Omega$, $\mathbf{P}(\omega, \cdot) \in \mathcal{P}^*$; e para cada $A \in \mathcal{A}$, $\mathbf{P}(\cdot, A)$ é um vetor aleatório

r -dimensional assumindo valores no simplex $\mathcal{S}_r$ de $\mathbb{R}^r$.

Consideramos, então o seguinte conjunto de sequências de vetores aleatórios,

$$\mathbf{W}_i = \{(U_1^i, \theta_1^i), (U_2^i, \theta_2^i), \dots\} \quad i = 1, 2, \dots, r,$$

tal que:

- (i) As variáveis aleatórias $(U_1^i, U_2^i, \dots)$ são independentes e identicamente distribuídas, assumindo valores em $\mathcal{X}$, com $U_1^i \sim \alpha_i(\cdot)/\alpha_i(\mathcal{X})$, $i = 1, 2, \dots, r$.
- (ii) As variáveis aleatórias $(\theta_1^i, \theta_2^i, \dots)$ são independentes e identicamente distribuídas, assumindo valores em $[0, 1]$, com $\theta_1^i \sim \text{Beta}(1, \alpha_i(\mathcal{X}))$, $i = 1, 2, \dots, r$.
- (iii) As variáveis aleatórias $\{U_j^1\}, \dots, \{U_j^r\}, \{\theta_j^1\}, \dots, \{\theta_j^r\}$, são mutuamente independentes.

Consideramos então o espaço de probabilidade produto $(\Omega, \sigma(\Omega), \eta_{\vec{\alpha}})$, onde $\Omega = \times_{i=1}^r \Omega_i$, $\sigma(\Omega) = \times_{i=1}^r \sigma(\Omega_i)$, sendo $\Omega_i = (\mathcal{X} \times [0, 1])^\infty$, $\sigma(\Omega_i) = (\mathcal{A} \times \mathcal{B}([0, 1]))^\infty$, $i = 1, 2, \dots, r$, e $\eta_{\vec{\alpha}} = \eta_{\alpha_1} \times \dots \times \eta_{\alpha_r}$ é a lei do vetor $(\mathbf{W}^1, \dots, \mathbf{W}^r)$.

Seja $\boldsymbol{\rho}$, vetor aleatório r -dimensional, definido sobre $(\Omega, \sigma(\Omega), \eta_{\vec{\alpha}})$, independente de $\mathbf{W}^i$, $i = 1, 2, \dots, r$, tal que $\boldsymbol{\rho} \sim \mathcal{D}_s(\alpha_1(\mathcal{X}), \dots, \alpha_r(\mathcal{X}))$.

Sejam $p_1^i = \theta_1^i$, $p_j^i = \theta_j^i \prod_{\ell=1}^{j-1} (1 - \theta_\ell^i)$, $j \geq 2$, $i = 1, 2, \dots, r$.

Para cada $\omega \in \Omega$ e para todo $A \in \mathcal{A}$, definimos

$$P_i(\omega, A) = \sum_{j=1}^{\infty} p_j^i(\omega) \delta_{U_j^i(\omega)}(A), \quad i = 1, 2, \dots, r,$$

portanto $P_1, P_2, \dots, P_r$ definem r processos de Dirichlet mutuamente independentes.

Notamos também que os elementos aleatórios $\boldsymbol{\rho}, P_1, P_2, \dots, P_r$ são mutuamente independentes, i.é, $\boldsymbol{\rho} \amalg P_1 \amalg P_2 \amalg \dots \amalg P_r / \vec{\alpha}$.

Nestes termos, a caracterização do processo de Dirichlet r -variado é dada no seguinte teorema.

Teorema 3.4.1

Sejam $(\mathcal{X}, \mathcal{A})$, e $\vec{\alpha}$ como antes.

Sejam $\rho = (\rho_1, \dots, \rho_r)$; $P_i, i = 1, \dots, r$ elementos aleatórios definidos sobre o espaço de probabilidade $(\Omega, \sigma(\Omega), \eta_{\vec{\alpha}})$, tais que $\rho \amalg P_1 \amalg \dots \amalg P_r / \vec{\alpha}$, com $\rho \sim \mathcal{D}_s(\alpha_1(\mathcal{X}), \dots, \alpha_r(\mathcal{X}))$ e $P_i \in \mathcal{D}(\alpha_i), i = 1, \dots, r$.

Seja $\mathbf{P}$ a função mensurável de $(\Omega, \sigma(\Omega), \eta_{\vec{\alpha}})$ em $(\mathcal{P}^*, \sigma(\mathcal{P}^*))$ ou equivalentemente o processo estocástico r -dimensional $\{\mathbf{P}(A) : A \in \mathcal{A}\}$ definido por

$$\mathbf{P}(\omega, A) = (\rho_1(\omega)P_1(\omega, A), \dots, \rho_r(\omega)P_r(\omega, A)) , \quad (3.4.1)$$

onde $\omega \in \Omega$, e $A \in \mathcal{A}$.

Então $\mathbf{P} \in \mathcal{DM}_r(\vec{\alpha})$.

Prova

A demonstração é feita acompanhando a prova do Teorema 3.3.1. $\square$

Uma proposição análoga à Proposição 3.3.1 pode ser estabelecida neste caso.

Proposição 3.4.1

Seja $\mathbf{P} \in \mathcal{DM}_r(\vec{\alpha})$ como no Teorema 3.4.1.

Então, se $A \in \mathcal{A}$, tem-se que $\mathbf{P}(A)$ é a única solução da relação

$$\mathbf{P}(A) \sim \mathbf{p}_1 \bullet \delta_{U_1}(A) + (1 - \mathbf{p}_1) \bullet \mathbf{P}(A) , \quad (3.4.2)$$

onde $\mathbf{p}_1 \delta_{U_1}(A) = (p_1^1 \delta_{U_1^1}(A), \dots, p_1^r \delta_{U_1^r}(A))$, $1 - \mathbf{p}_1 = (1 - p_1^1, \dots, 1 - p_1^r)$. $\square$

Observação 3.4.2

O processo de Dirichlet multivariado é utilizado, no próximo capítulo, como uma classe de distribuições a priori na resolução do problema de riscos competitivos, sob uma abordagem bayesiana não-paramétrica.

CAPÍTULO 4

Estimação bayesiana não-paramétrica em modelos de riscos competitivos

4.1 Introdução e Preliminares.

Este capítulo é dedicado à solução bayesiana não-paramétrica de situações com observações censuradas aleatoriamente. Uma introdução ao problema foi dada com algum detalhe no capítulo 1 deste trabalho.

Neste tipo de problemas existem algumas contribuições onde o processo de Dirichlet é frequentemente utilizado. Os primeiros avanços foram obtidos por Susarla e Van Ryzin (1976,1978a), e Phadia (1980). Esses trabalhos têm por objetivo a estimação de uma função de sobrevivência baseada sobre dados censurados, supondo como medida de probabilidade a priori um processo de Dirichlet. Um estudo assintótico dos estimadores obtido pode ser visto em Susarla e Van Ryzin (1978b), e um estudo de simulação considerando diferentes estimadores é desenvolvido por Rai et al. (1980). Outros trabalhos relacionados são os de Jammalamadaka e Tiwari (1985), e Zenwirth (1985). Ver também as contribuições de Blum e Susarla (1977). Por outra parte, Papadopoulos e Tiwari (1989) fazem um estudo do tipo bootstrap bayesiano usando o processo de Dirichlet para a estimação de funções de confiabilidade e taxas de falhas, esse estudo é feito mediante técnicas de simulação. Mais recentemente, Morales et al. (1990) consideram a estimação de uma curva de sobrevivência em modelos com taxas de falhas proporcionais com dados duplamente censurados, e Tiwari e Zalkikar (1990) estudam modelos com dados censurados parcialmente observáveis. Por outra parte, problemas com dados truncados aleatoriamente são considerados por Tiwari e Zalkikar (1991).

Na sequência definimos o problema em questão, da maneira que foi apresentado no capítulo introdutório deste trabalho. Nosso objetivo é utilizar o processo de Dirichlet bivariado como uma classe de distribuições a priori na procura de soluções bayesianas não-paramétricas, como também propor novas soluções não-paramétricas para o problema com dados censurados, e portanto esse o tema de interesse na segunda seção. Este ponto de vista não nos parece estar estudado na literatura. A natural extensão ao problema de estimação na teoria de riscos competitivos é considerada na última seção deste capítulo, utilizando como distribuição a priori o processo de Dirichlet multivariado.

Consideramos, então, (X_i, Y_i) , $i = 1, 2, \dots, n$, n pares de variáveis aleatórias independentes e identicamente distribuídas tais que $X_i \sim X$ e $Y_i \sim Y$. Denotamos por $S_1(t) = \Pr\{X > t\}$ e $S_2(t) = \Pr\{Y > t\}$ as funções de sobrevivência das variáveis aleatórias X e Y , respectivamente. Assumimos que as variáveis X_i 's e Y_i 's são independentes para todo $i = 1, 2, \dots, n$.

No problema com dados censurados, as variáveis aleatórias, (X_i, Y_i) , $i = 1, 2, \dots, n$, não sempre são totalmente observadas, e o estatístico só tem disponível os dados:

$$\begin{aligned} Z_i &= \min(X_i, Y_i) \\ \delta_i &= \begin{cases} 1 & \text{se } Z_i = X_i \\ 2 & \text{se } Z_i = Y_i \end{cases}, \quad i = 1, 2, \dots, n \end{aligned} \quad (4.1.1)$$

Isto é, são observadas para cada par (X_i, Y_i) , $i = 1, 2, \dots, n$, o mínimo entre X_i e Y_i , e uma função indicadora que diz em que variável esse mínimo foi atingido. A função de sobrevivência de $Z_i = \min(X_i, Y_i)$ é denotada por $S(t) = \Pr(Z_i > t)$.

Embora, o problema com dados censurados é simétrico em X_i e Y_i , é comum referir-se a X_i como a variável de interesse e referir-se a Y_i como a variável que censura. Assim, dizemos que a observação Z_i é não-censurada se $\delta_i = 1$ e censurada se $\delta_i = 2$.

Dado que as variáveis aleatórias em questão representam em geral tempos de sobrevivência, vamos supor que elas assumem valores no intervalo $(0, \infty)$. Problemas de interesse são: estimar as funções de sobrevivência $S_1(\cdot)$ e $S_2(\cdot)$, o funcional $\Pr(X < Y)$, e

a função de sobrevivência $S(\cdot)$.

Nós não vamos impor a restrição, a qual é usual, que as funções de sobrevivência $S_1(\cdot)$ e $S_2(\cdot)$ são contínuas. A hipótese a ser considerada sobre as funções $S_1(\cdot)$ e $S_2(\cdot)$ é que estas funções não possuem pontos de descontinuidade em comum. Esta última suposição é necessária para que o problema seja bem definido, em particular evita ambiguidade na definição (4.1.1) de δ_i .

Definimos as funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$ por:

$$S_1^*(t) = \Pr(Z > t, \delta = 1) \quad , \quad S_2^*(t) = \Pr(Z > t, \delta = 2) \quad (4.1.2)$$

As seguintes propriedades das funções de subsobrevivência são bem conhecidas e são dadas sem prova.

Propriedade 1:

Dado que as variáveis aleatórias são independentes, a função de subsobrevivência $S_1^*(\cdot)$ pode ser expressa em função das funções de sobrevivência $S_1(\cdot)$ e $S_2(\cdot)$ como:

$$S_1^*(t) = \int_t^\infty S_2(s)[-dS_1(s)] \quad (4.1.3)$$

Notamos que a integral de Riemann-Stieltjes é bem definida, pois integrador e integrando não possuem pontos de salto em comum.

Propriedade 2:

- (a) $S_1^*(0) = \Pr(Z > 0, \delta = 1) = \Pr(\delta = 1) = \Pr(X < Y)$
- (b) $S_2^*(0) = \Pr(Z > 0, \delta = 2) = \Pr(\delta = 2) = \Pr(Y < X)$
- (c) $S_1^*(0) + S_2^*(0) = 1$
- (d) $S_1^*(t) + S_2^*(t) = \Pr(\min(X, Y) > t) = S_1(t)S_2(t) = \Pr(Z > t) = S(t)$

Propriedade 3:

(a) Os pontos de descontinuidade de $S_j^*(\cdot)$ são os pontos de descontinuidade de $S_j(\cdot)$, $j = 1, 2$.

(b) Da hipótese que $S_1(\cdot)$ e $S_2(\cdot)$ não possuem pontos de descontinuidade em comum, decorre-se que $S_1^*(\cdot)$ e $S_2^*(\cdot)$ também não possuirão.

Propriedade 4:

Seja $t_G = \sup(t > 0 : G(t) > 0)$. Então $\min(t_{S_1}, t_{S_2}) = \max(t_{S_1^*}, t_{S_2^*}) \equiv t^*$.

Pelo menos uma das duas subsobrevivência $S_1^*(t)$ e $S_2^*(t)$ é positiva para $t < t^*$, e ambas são zero para $t \geq t^*$.

A proposição seguinte é simples, mas importante. Dado que a distribuição conjunta das observações (Z_i, δ_i) no problema com dados censurados é $\Pr(Z_i > t, \delta_i = j) \equiv S_j^*(t)$, $j = 1, 2$, temos que:

Proposição 4.1

As duas funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$ determinam a distribuição das observações (Z_i, δ_i) no problema com dados censurados. $\square$

Um interesse natural é obter uma expressão para a função de sobrevivência $S_1(\cdot)$ em termos das funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$.

É conhecido (cf. Breslow e Crowley, 1974, eq. 7.5) que para funções de sobrevivência contínuas $S_1(\cdot)$ e $S_2(\cdot)$, a função de sobrevivência $S_1(\cdot)$ pode ser expressa em termos das funções de subsobrevivência como segue:

$$S_1(t) = \exp \left[\int_0^t \frac{dS_1^*(s)}{S_1^*(s) + S_2^*(s)} \right], \quad (4.1.4)$$

mas, esta expressão não é válida quando as funções de sobrevivência possuem pontos de descontinuidade. Notamos que a integral de Riemann-Stieltjes em (4.1.4) não está definida nesse caso, pois integrador e integrando têm saltos em comum.

Uma generalização da expressão (4.1.4) foi dada por Peterton (1977), quando a única restrição sobre $S_1(\cdot)$ e $S_2(\cdot)$ é que estas funções não tenham pontos de descontinuidade em comum.

Teorema 4.1.1 (Fórmula de Peterson).

(a) As funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$ determinam unicamente a função de sobrevivência $S_1(t)$ para $t \leq t^*$ de acordo com a seguinte expressão:

$$S_1(t) = \varphi(S_1^*(\cdot), S_2^*(\cdot), t) \quad (4.1.5)$$

onde,

$$\varphi(F(\cdot), G(\cdot), t) = \exp \left[c \int_0^t \frac{dF(s)}{F(s) + G(s)} \right] \prod_{\substack{s \leq t \\ s \text{ ponto de} \\ \text{salto de } F(\cdot)}} \frac{F(s^+) + G(s^+)}{F(s^-) + G(s^-)} \quad (4.1.6)$$

sendo $c \int_0^t$: integração sobre a união dos intervalos abertos de pontos menores que t para os quais F é contínua.

(b) Se $S_1(t^*) = 0$, então $S_1(t) = 0$ para $t > t^*$. □

Observações 4.1

(a) O primeiro fator de (4.1.5) considera a parte contínua de $S_1^*(\cdot)$, e coincide com a expressão (4.1.4). O segundo fator de (4.1.5) considera a parte discreta de $S_1^*(\cdot)$.

(b) A Fórmula de Peterson nos será de utilidade na definição de um estimador não-paramétrico da função de sobrevivência $S_1(\cdot)$.

(c) Propriedades do funcional φ são discutidas no capítulo seguinte, em particular no estudo assintótico das soluções obtidas.

4.2 Uma abordagem bayesiana não-paramétrica para o problema com observações aleatoriamente censuradas

Nesta seção, consideramos o problema de estimação de uma curva de sobrevivência quando as observações são aleatoriamente censuradas, sob uma ótica bayesiana não-paramétrica. Na nossa abordagem, utilizamos a classe de processos de Dirichlet bivariado, desenvolvida no capítulo anterior, como uma classe de distribuições a priori para nosso problema. A utilização do processo de Dirichlet bivariado, está inspirada fundamentalmente na estrutura da função de verosimilhança, da qual decorre-se que em uma primeira instância o parâmetro de interesse é o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$. Portanto, aparece de maneira natural o fato de introduzir uma classe de distribuições a priori sobre o espaço de funções definido por esses vetores. Agora, dada a simplicidade analítica, particularmente na obtenção da distribuição a posteriori, e também tendo em mente as propriedades (I) e (II) do processo de Dirichlet usual (ver seção 2.2 deste trabalho), escolhemos a classe de processos de Dirichlet bivariado como nossa classe de distribuições a priori.

Notamos que nosso esquema é diferente do adotado por Susarla e Van Ryzin (1976). Nesse trabalho, a distribuição a priori é incorporada na função de sobrevivência $S_1(\cdot)$ de X e, eles não consideram distribuição a priori alguma sobre a distribuição da variável censora Y . No nosso caso, estamos introduzindo uma classe de distribuições a priori sobre o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, portanto estamos tomando em consideração ambas variáveis X e Y , como também a função indicadora δ , que indica em qual dessas variáveis o mínimo é atingido. Dessa maneira pensamos estar incorporando toda a informação resultante das observações nos cálculos dos estimadores bayesianos, uma vez que a distribuição a posteriori seja determinada.

Como foi formulado em (4.1.1), as observações são $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$; onde $Z_i = \min(X_i, Y_i)$, $\delta_i = \begin{cases} 1 & \text{se } Z_i = X_i \\ 2 & \text{se } Z_i = Y_i \end{cases}$, $i = 1, 2, \dots, n$. Ou equivalentemente, para cada t ,

observamos as variáveis aleatórias:

$$nS_{1n}^* = \sum_{i=1}^n I(Z_i > t, \delta_i = 1) \quad ; \quad nS_{2n}^* = \sum_{i=1}^n I(Z_i > t, \delta_i = 2) \quad (4.2.1)$$

onde $I(A)$ é a função indicadora do conjunto A .

As funções $S_{1n}^*(\cdot)$ e $S_{2n}^*(\cdot)$ são chamadas funções de subsobrevivência empírica. Notamos que se $S_n(\cdot)$ é a função de distribuição empírica correspondente às observações $Z_1, \dots, Z_n$, então para cada t ,

$$nS_n(t) = nS_{1n}^*(t) + nS_{2n}^*(t) \quad (4.2.2)$$

Neste contexto, para cada t , a função de verossimilhança do problema corresponde à verossimilhança de uma modelo Multinomial $(n; p_1(t), p_2(t), p_3(t))$, sendo $p_1(t) = S_{1n}^*(t)$, $p_2(t) = S_{2n}^*(t)$ e $p_3(t) = 1 - S(t) = 1 - S_{1n}^*(t) - S_{2n}^*(t)$, i.é.,

$$\text{Lik} = \Pr(nS_{1n}^*(t) = j_1(t), nS_{2n}^*(t) = j_2(t)) \propto S_{1n}^*(t)^{j_1(t)} S_{2n}^*(t)^{j_2(t)} (1 - S(t))^{n-j_1(t)-j_2(t)} \quad (4.2.3)$$

Logo, em primeira instância, o parâmetro de interesse é o vetor de subsobrevivências $(S_{1n}^*(\cdot), S_{2n}^*(\cdot))$.

Consideramos $\mathcal{X} = (0, \infty)$ e $\mathcal{A} = \mathcal{B}(0, \infty)$, dado que estamos supondo que as variáveis aleatórias envolvidas no problema assumem valores em $(0, \infty)$.

O espaço de ações do estatístico é o conjunto de todos os vetores bidimensionais de subsobrevivências definidos sobre $(0, \infty)$, tais que a soma dos seus componentes define uma função de sobrevivência.

Na primeira parte da análise bayesiana, utilizamos uma função de perda quadrática nos cálculos de soluções bayesianas de algumas quantidades de interesse, tais como o vetor de subsobrevivências $(S_{1n}^*(\cdot), S_{2n}^*(\cdot))$, a função de sobrevivência $S(\cdot)$ e o funcional $\Pr(X < Y)$. Logo, usando a fórmula de Peterson, propomos um estimador não-paramétrico para a

função de sobrevivência $S_1(\cdot)$. Na segunda parte, uma análise similar é feita, considerando uma função de perda $0 - 1$.

A distribuição a priori para o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$ é construída a partir da caracterização do processo de Dirichlet bivariado dada no Teorema 3.3.1.

Lembramos que no nosso caso $\mathcal{X} = (0, \infty)$ e $\mathcal{A} = \mathcal{B}(0, \infty)$, e consideramos α_1 e α_2 medidas finitas, não-nulas, não-negativas e finitamente aditivas sobre $(\mathcal{X}, \mathcal{A})$.

Definamos,

$$\rho = \Pr(X < Y) = \Pr(\delta = 1) \quad (4.2.4)$$

e, as funções de sobrevivência condicionais,

$$\begin{aligned} T_1^*(t) &= \Pr(Z > t/\delta = 1) \\ T_2^*(t) &= \Pr(Z > t/\delta = 2) . \end{aligned} \quad (4.2.5)$$

Logo é fácil ver que o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, pode ser expresso como:

$$S_1^*(\cdot), S_2^*(\cdot) = (\rho T_1^*(\cdot), (1 - \rho)T_2^*(\cdot)) \quad (4.2.6)$$

e obtemos o seguinte:

Lema 4.2.1

Se $\rho \sim \text{Beta}(\alpha_1(0, \infty), \alpha_2(0, \infty))$, $T_1^* \in \mathcal{D}(\alpha_1)$ e $T_2^* \in \mathcal{D}(\alpha_2)$; tais que $\rho \amalg T_1^* \amalg T_2^* / (\alpha_1, \alpha_2)$. Então $(S_1^*, S_2^*) \in \mathcal{DB}(\alpha_1, \alpha_2)$.

Prova: É consequência direta do Teorema 3.3.1. □

Observação 4.2.1

Decorre-se do lema anterior que, para cada t , se $\rho \sim \text{Beta}(\alpha_1(0, \infty), \alpha_2(0, \infty))$, $T_1^*(t) \sim \text{Beta}(\alpha_1(t, \infty), \alpha_2(0, t])$ e $T_2^*(t) \sim \text{Beta}(\alpha_2(t, \infty), \alpha_2(0, t])$, e dado o vetor de medidas (α_1, α_2) , estas variáveis aleatórias são condicionalmente independentes, então

$$(S_1^*(t), S_2^*(t)) \sim \mathcal{D}(\alpha_1(t, \infty), \alpha_2(t, \infty); \alpha_1(0, t] + \alpha_2(0, t]) \quad (4.2.7)$$

ou na sua forma singular,

$$(S_1^*(t), S_2^*(t), 1 - S(t)) \sim \mathcal{D}_s(\alpha_1(t, \infty), \alpha_2(t, \infty), \alpha_1(0, t] + \alpha_2(0, t]) \quad (4.2.8)$$

Agora, de (4.2.3) e (4.2.7), e utilizando a Proposição 1 do Apêndice A, obtemos a distribuição a posteriori do vetor $(S_1^*(t), S_2^*(t))$ dada as observações.

Para cada t ,

$$\begin{aligned} (S_1^*(t), S_2^*(t))/\text{Dados} &\sim \mathcal{D}(\alpha_1(t, \infty) + j_1(t), \alpha_2(t, \infty) + j_2(t); \\ &\alpha_1(0, t] + \alpha_2(0, t] + n - j_1(t) - j_2(t)) \end{aligned} \quad (4.2.9)$$

ou na forma singular, para cada t ,

$$\begin{aligned} (S_1^*(t), S_2^*(t), 1 - S(t))/\text{Dados} \\ \sim \mathcal{D}_s(\alpha_1(t, \infty) + j_1(t), \alpha_2(t, \infty) + j_2(t), \alpha_1(0, t] + \alpha_2(0, t] + n - j_1(t) - j_2(t)) \end{aligned} \quad (4.2.10)$$

4.2.1 Análise bayesiana sob uma função de perda quadrática

Consideramos, para o problema de estimação do vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, a seguinte função de perda quadrática:

$$L(\mathcal{S}^*, \hat{\mathcal{S}}^*) = \int \| \mathcal{S}^*(t) - \hat{\mathcal{S}}^*(t) \|^2 dW(t) \quad (4.2.11)$$

onde, $\mathcal{S}^*(t) = (S_1^*(t), S_2^*(t))$, e $\hat{\mathcal{S}}^*(t) = (\hat{S}_1^*(t), \hat{S}_2^*(t))$ denota o estimador de $\mathcal{S}^*$. Também, $\| \cdot \|$ denota a norma euclidiana em $\mathbb{R}^2$, e W é uma função de peso.

O risco bayesiano é,

$$\mathbb{E}L(\mathcal{S}^*, \hat{\mathcal{S}}^*) = \int \mathbb{E} \| \mathcal{S}^*(t) - \hat{\mathcal{S}}^*(t) \|^2 dW(t) ,$$

minimizar o risco bayesiano equivale a minimizar $\mathbb{E} \| \mathcal{S}^*(t) - \hat{\mathcal{S}}^*(t) \|^2$, para cada t .

Logo, temos o seguinte resultado:

Proposição 4.2.1

Considerando o processo de Dirichlet bivariado de parâmetro (α_1, α_2) como distribuição a priori para o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, i.é, $(S_1^*, S_2^*) \in \mathcal{DB}(\alpha_1, \alpha_2)$, e a função de perda quadrática dada por (4.2.11), temos que o estimador bayesiano do vetor de subsobrevivências (S_1^*, S_2^*) é dado pela expressão:

$$\begin{aligned} \hat{\mathcal{S}}^*(t) = (\hat{S}_1^*(t), \hat{S}_2^*(t)) &= p_n \left(\frac{\alpha_1(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)}, \frac{\alpha_2(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} \right) \\ &+ (1 - p_n)(S_{1n}^*(t), S_{2n}^*(t)), \quad t \in (0, \infty) \end{aligned} \quad (4.2.12)$$

onde,

$$p_n = \frac{\alpha_1(0, \infty) + \alpha_2(0, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \quad (4.2.13)$$

e S_{1n}^* e S_{2n}^* são as funções de subsobrevivência empírica definidas em (4.2.1).

Prova.

É bem conhecido que o estimador bayesiano, $\hat{\mathcal{S}}^*(\cdot)$, é aquele que minimiza o risco a posteriori e a solução é dada pelo vetor média da distribuição a posteriori. Portanto, para cada t ,

$$\begin{aligned} \hat{\mathcal{S}}^*(t) &= (\hat{S}_1^*(t), \hat{S}_2^*(t)) = \\ &= \left(\frac{\alpha_1(t, \infty) + j_1(t)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n}, \frac{\alpha_2(t, \infty) + j_2(t)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \right) \\ &= \left(\frac{\alpha_1(t, \infty) + \sum_{i=1}^n I(Z_i > t, \delta_i = 1)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n}, \frac{\alpha_2(t, \infty) + \sum_{i=1}^n I(Z_i > t, \delta_i = 2)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \right) \end{aligned}$$

e desta última expressão segue (4.2.12) □

Observações 4.2.2

(a) Notamos que o estimador bayesiano do vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$ em (4.2.12), é uma combinação convexa entre o vetor dada priori (ver 4.2.7) e o vetor cujos componentes são as funções de subsobrevivência empírica.

(b) Do Teorema de Glivenko-Cantelli (cf. Apêndice B), e dado que $p_n \downarrow 0$, quando $n \rightarrow \infty$, temos que $\hat{\mathcal{S}}^* = (\hat{S}_1^*, \hat{S}_2^*)$ converge ao verdadeiro vetor de subsobrevivências (S_1^*, S_2^*) , uniformemente com probabilidade 1. Isto é,

$$\|\hat{\mathcal{S}}^* - \mathcal{S}^*\|_\infty := \max\left\{\sup_{t \in (0, \infty)} |\hat{S}_1^*(t) - S_1^*(t)|, \sup_{t \in (0, \infty)} |\hat{S}_2^*(t) - S_2^*(t)|\right\}$$

converge a zero, com probabilidade 1. Trata-se então de um estimador fortemente consistente.

(c) Notamos que os estimadores das funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$; dados pela expressão (4.2.12) são:

Para cada t ,

$$\hat{S}_1^*(t) = p_n \frac{\alpha_1(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) S_{1n}^*(t) \quad (4.2.14)$$

$$\hat{S}_2^*(t) = p_n \frac{\alpha_2(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) S_{2n}^*(t) \quad (4.2.15)$$

Estes estimadores são os estimadores bayesianos de $S_1^*(\cdot)$ e $S_2^*(\cdot)$, sob respectivas funções de perda quadrática, quando consideramos as distribuições a priori:

Para cada t ,

$$\begin{aligned} S_1^*(t) &\sim \text{Beta}(\alpha_1(t, \infty), \alpha_1(0, t] + \alpha_2(0, \infty)) \\ S_2^*(t) &\sim \text{Beta}(\alpha_2(t, \infty), \alpha_2(0, t] + \alpha_1(0, \infty)) \end{aligned} \quad (4.2.16)$$

Este fato é totalmente coerente com a nossa escolha do processo de Dirichlet bivariado como distribuição a priori para o vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, ver também expressões (4.2.7) e (4.2.9).

As mesmas propriedades descritas em (a) e (b) para $\hat{S}^*$, também valem para $\hat{S}_1^*$ e $\hat{S}_2^*$ por separado. Em particular, ambos estimadores são fortemente consistentes para S_1^* e S_2^* , respectivamente.

Como foi visto no capítulo introdutório, o problema com observações aleatoriamente censuradas corresponde a um sistema com dois componentes em série. Dessa maneira, $S(t) = \Pr(Z > t)$ representa o tempo de sobrevivência do sistema e $\rho = \Pr(X < Y)$ é chamada a confiabilidade do sistema. Vemos que estas duas quantidades de interesse podem ser estimadas a partir da estrutura definida anteriormente e dos resultados já obtidos.

Proposição 4.2.2

Considerando a distribuição a priori como em (4.2.8), o estimador bayesiano da função de sobrevivência da variável Z , $S(t) = \Pr(Z > t)$, sob função de perda quadrática é dado pela expressão:

$$\begin{aligned}\hat{S}(t) &= \hat{S}_1^*(t) + \hat{S}_2^*(t) = \frac{\alpha_1(t, \infty) + \alpha_2(t, \infty) + \sum_{i=1}^n I(Z_i > t)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \\ &= p_n \frac{\alpha_1(t, \infty) + \alpha_2(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) S_n(t) \quad t \in (0, \infty)\end{aligned}\tag{4.2.17}$$

onde p_n é com em (4.2.13) e S_n é a função de distribuição empírica correspondente às observações $Z_1, \dots, Z_n$, dada em (4.2.2).

Prova. Direta de (4.2.14) e (4.2.15). □

Proposição 4.2.3

Sejam $\rho = \Pr(X < Y) = \Pr(\delta = 1)$, e a função de sobrevivência condicional $T_1^*(\cdot)$ como no Lema 4.2.1.

Suponhamos que a função $f_1(\cdot) = \alpha_1(\cdot, \infty)$ é contínua pela direita no ponto zero.

Então o estimador bayesiano de ρ sob função de perda quadrática é dado por

$$\hat{\rho} = \lim_{t \downarrow 0} \hat{S}_1^*(t) = p_n \frac{\alpha_1(0, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) \frac{1}{n} \sum_{i=1}^n I(\delta_i = 1) \quad (4.2.18)$$

Prova.

Temos que,

$$\begin{aligned} \lim_{t \downarrow 0} \hat{S}_1^*(t) &= p_n \frac{\alpha_1(0, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) S_{1n}^*(0) = \\ &= p_n \frac{\alpha_1(0, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} + (1 - p_n) \frac{1}{n} \sum_{i=1}^n I(\delta_i = 1) , \end{aligned}$$

notando que a distribuição a priori para ρ é, $\rho \sim \text{Beta}(\alpha_1(0, \infty), \alpha_2(0, \infty))$ e que a verossimilhança, nesse caso, é definida pelo modelo binomial $\sum_{i=1}^n I(\delta_i = 1) \sim \text{Bin}(n, \rho)$, obtemos a conclusão. $\square$

Observações 4.2.3.

- (a) O estimador bayesiano de ρ , $\hat{\rho}$, é uma combinação convexa entre a média da priori $\alpha_1(0, \infty)/(\alpha_1(0, \infty) + \alpha_2(0, \infty))$ e o estimador clássico de ρ , $\frac{1}{n} \sum_{i=1}^n I(\delta_i = 1) = \frac{1}{n} \{\text{número de pares } (X_i, Y_i), \text{ tais que } X_i < Y_i\}$.
- (b) Dado que $p_n \downarrow 0$, pela lei dos grandes números $\hat{\rho} \xrightarrow{P} \rho$, quando $n \rightarrow \infty$.
- (c) O Teorema do limite central garante que $n^{\frac{1}{2}}(\hat{\rho} - \rho) \xrightarrow{D} N(0, \rho(1 - \rho))$, quando $n \rightarrow \infty$.
- (d) Sob a hipótese de continuidade pela direita da função $f_2(\cdot) = \alpha_2(\cdot, \infty)$, o estimador bayesiano do funcional $\Pr(Y < X)$ é dado por $\lim_{t \downarrow 0} \hat{S}_2^*(t)$.

Naturalmente a estimação das funções de sobrevivência $S_1(\cdot)$ e $S_2(\cdot)$, é o relevante no problema com observações aleatoriamente censuradas. Assim, usando a Proposição 4.2.2 e a fórmula de Peterson dada no Teorema 4.1.1, propomos $\hat{S}_1$ como um estimador não-paramétrico da função de sobrevivência S_1 de X , onde

$$\hat{S}_1(t) = \varphi(\hat{S}_1^*(\cdot), \hat{S}_2^*(\cdot); t) \quad t \leq t^* , \quad (4.2.19)$$

onde t^* foi definido na Propriedade 4 da seção 4.1.

O resultado mais importante desta seção é dado no teorema seguinte.

Teorema 4.2.1

Suponhamos que as funções $f_1(\cdot) = \alpha_1(\cdot, \infty)$ e $f_2(\cdot) = \alpha_2(\cdot, \infty)$ são contínuas em $(0, t)$ e S_1 e S_2 não possuem pontos de descontinuidade em comum. Então,

$$\hat{S}_1(t) = \hat{S}(t)I(t)\Pi(t), \quad t \in (0, Z_{(m)}^*) \quad (4.2.20)$$

onde $\hat{S}(t)$ é o estimador bayesiano da função de sobrevivência $S(t) = \Pr(Z > t)$, dado por (4.2.17),

$$I(t) = \exp \left\{ \frac{-1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \int_0^t \frac{d\alpha_2(s, \infty)}{\hat{S}(s)} \right\}$$

$$\Pi(t) = \prod_{i: Z_{(i)}^* \leq t} \frac{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i - d_i}{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i},$$

onde $Z_{(1)}^* < Z_{(2)}^* < \dots < Z_{(m)}^*$ são os m valores ($m \leq n$) diferentes das estatísticas de ordem correspondentes às observações $Z_1, \dots, Z_n$, $n_j = \sum_{i=1}^n I(Z_i \geq Z_{(j)}^*)$ e $d_j = \sum_{i=1}^n I(Z_i = Z_{(j)}^*, \delta_i = 1)$, $j = 1, 2, \dots, m$.

Prova.

De (4.2.19) e o fato que $\hat{S}(\cdot) = \hat{S}_1^*(\cdot) + \hat{S}_2^*(\cdot)$, temos que

$$\hat{S}_1(t) = \exp \left\{ \int_0^t \frac{d\hat{S}_1^*(s)}{\hat{S}(s)} \right\} \prod_{\substack{s \leq t \\ s \text{ ponto de} \\ \text{salto de } \hat{S}_1^*(\cdot)}} \frac{\hat{S}_1^*(s^+) + \hat{S}_2^*(s^+)}{\hat{S}_1^*(s^-) + \hat{S}_2^*(s^-)}, \quad (4.2.21)$$

para $t \leq t^*$. Notamos que $d\hat{S}_1^*(s) = d\hat{S}(s) - \frac{d\alpha_2(s, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n}$, logo o primeiro termo em (4.2.21) é

$$\exp \left\{ \int_0^t \frac{d\hat{S}_1^*(s)}{\hat{S}(s)} \right\} = \hat{S}(t) \exp \left\{ -\frac{1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n} \int_0^t \frac{d\alpha_2(s, \infty)}{\hat{S}(s)} \right\}$$

$$= \hat{S}(t)I(t).$$

Notamos que, para cada t , $f_2(\cdot) = \alpha_2(\cdot, \infty)$ é monótona (decrecente), e contínua em $(0, t)$, e $\frac{1}{\hat{S}(\cdot)}$ é monótona (crescente) em $(0, t)$. Portanto, é possível decompor de maneira única $f_2(\cdot)$ numa diferença de funções monótonas e contínuas, cf. Teorema 6.27, Corolário 1 de Rudin (1964), e $\frac{1}{\hat{S}(\cdot)}$ numa diferença de funções monótonas. Assim, o Teorema 6.9 de Rudin (1964) garante que a integral $\int_0^t \frac{d\alpha_2(s, \infty)}{\hat{S}(s)}$ é bem definida.

Por outro lado, o segundo fator de (4.2.21) é

$$\begin{aligned} \prod_{\substack{s \leq t \\ s \text{ ponto de} \\ \text{salto de } \hat{S}_1^*(\cdot)}} \frac{\hat{S}_1^*(s^+) + \hat{S}_2^*(s^+)}{\hat{S}_1^*(s^-) + \hat{S}_2^*(s^-)} &= \prod_{s \leq t} \frac{\alpha_1(s, \infty) + \alpha_2(s, \infty) + \sum_{i=1}^n I(Z_i > s^+)}{\alpha_1(s, \infty) + \alpha_2(s, \infty) + \sum_{i=1}^n I(Z_i > s^-)} \\ &= \prod_{i: Z_{(i)}^* \leq t} \frac{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i - d_i}{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i} \\ &= \Pi(t) \quad \square \end{aligned}$$

Observações 4.2.4

(a) Definindo $\alpha_1(t, \infty) + \alpha_2(t, \infty) = \alpha(t, \infty)$, para cada t , em (4.2.20), o produto $B(t)\Pi(t)$ é a equação (3.2) de Tsai (1986). [Notamos que n_{i+1} no denominador da fórmula (3.2) de Tsai (1986) deveria ser n_i .] Além disso, $B(t)\Pi(t)$ é similar ao estimador bayesiano de S_1 obtido por Susarla e Van Ryzin (1976), embora todos estes estimadores tenham sido derivados usando diferentes distribuições a priori.

(b) Sendo a integral o segundo fator, $I(t)$, em (4.2.20), nosso estimador de S_1 é mais suave que os estimadores de Tsai (1986) e Susarla e Van Ryzin (1976).

(c) Se $\alpha_1(0, \infty)$ e $\alpha_2(0, \infty)$ convergem a zero, então o estimador $\hat{S}_1$ é uma versão modificada do estimador PL de Kaplan e Meier (1958).

4.2.1.1 Cálculo numérico do estimador $\hat{S}_1$.

Na expressão (4.2.20) do estimador $\hat{S}_1(t)$, a integral $\phi(t) := \int_0^t \frac{d\alpha_2(s, \infty)}{\hat{S}(s)}$ deve ser apro-

ximada numericamente.

Suponhamos $d\alpha_2(s, \infty)/\hat{S}(s) = g_2(s)ds$, onde $g_2(s) = f'_2(s)/\hat{S}(s)$, i.é., $\alpha_2 \ll \lambda$ sendo λ a medida de Lebesgue em $\mathbb{R}$, então devemos resolver a equação diferencial

$$\begin{cases} \phi'(t) = g_2(t), & t \in [0, Z_{(m)}^*) \\ \phi(0) = 0 \end{cases}.$$

O método Runge-Kutta de ordem 4 [cf. Davis e Rabinowitz, 1984, capítulo 2] pode ser aplicado em cada um dos intervalos, $[Z_{(0)}^*, Z_{(1)}^*)$, $[Z_{(1)}^*, Z_{(2)}^*)$, ..., $[Z_{(j-1)}^*, Z_{(j)}^*)$, ..., $[Z_{(m-1)}^*, Z_{(m)}^*)$, onde $Z_{(0)}^* := 0$.

Primeiramente resolvemos aproximadamente a equação diferencial $\phi'(t) = g_2(t)$ com $\phi(0) = 0$ no intervalo $[0, Z_{(1)}^*)$, e logo as equações diferenciais $\phi'(t) = g_2(t)$ com $\phi(Z_{(j-1)}^*) = \mathcal{J}_{j-1}^*$ nos intervalos $[Z_{(j-1)}^*, Z_{(j)}^*)$, onde $\mathcal{J}_{j-1}^*$ é obtido resolvendo a equação diferencial no intervalo anterior $[Z_{(j-2)}^*, Z_{(j-1)}^*)$, $j = 2, \dots, m$.

Em geral, seja $[a, b]$ o intervalo em questão, e $\{t_0, \dots, t_k\}$ uma partição de $(a, b]$ tal que $t_\ell - t_{\ell-1} = \frac{b-a}{k} \equiv h$, $\ell = 1, \dots, k$. Logo uma aproximação numérica da função $\phi(t)$ no intervalo $[a, b]$ é obtida usando as fórmulas:

$$\mathcal{J}_\ell = \mathcal{J}_{\ell-1} + \frac{h}{6} \left[g_2(t_{\ell-1}) + 4g_2(t_{\ell-1} + \frac{h}{2}) + g_2(t_{\ell-1} + h) \right] \quad \ell = 1, \dots, k \quad (4.2.22)$$

$$\mathcal{J}_0 = \phi(a), \quad t_0 = a.$$

onde $\mathcal{J}_\ell$ é o valor calculado da solução no ponto t_ℓ .

Assim, a expressão $\hat{S}_1(t) = B(t)I(t)\Pi(t)$ em (4.2.20) pode ser aproximada numericamente no intervalo $[a, b]$ pelos pares $\{(t_\ell, B(t_\ell)\tilde{\mathcal{J}}_\ell\Pi(t_\ell); \ell = 0, 1, \dots, k\}$, onde $\tilde{\mathcal{J}}_j = \exp\{-\mathcal{J}_j/(\alpha_1(0, \infty) + \alpha_2(0, \infty) + n)\}$, $\ell = 0, 1, \dots, k$.

4.2.1.2 Um exemplo

A título de ilustração no cálculo do estimador não-paramétrico $\hat{S}_1$, foram geradas duas amostras independentes $X_1, \dots, X_n; Y_1, \dots, Y_n$, tais que $X_i \sim \exp(1)$, $Y_i \sim \text{Weibull}(1, 2)$, $i = 1, 2, \dots, n$.

GRAFICO 4.1: Estimador versus Funcao de

Sobrevivencia Exponencial(1)

— Estimador

- - - Exponencial

GRAFICO 4.2: Estimador versus Funcao de

Sobrevivencia Weibull(1,2)

— Estimador

--- Weibull

A distribuição a priori é escolhida considerando $\alpha_i(t, \infty) = \frac{M-t}{M}$, $i = 1, 2$, onde M é tal que $\Pr(X > M) < 0,001$, o valor de M considerado é 8.0. Notamos que $\alpha_i(0, \infty) = 1$, $i = 1, 2$. Assim, obtemos como distribuição a priori para o funcional $\rho = \Pr(X < Y) \sim \text{Beta}(1, 1) \equiv U(0, 1)$, i.é, estamos considerando uma priori não-informativa.

Os resultados obtidos são apresentados nos Gráficos 4.1 e 4.2. No primeiro gráfico aparece o estimador não-paramétrico $\hat{S}_1(t)$ e a função de sobrevivência exponencial $S_1(t) = \Pr(X > t) = e^{-t}$. Também o problema simétrico é resolvido e no segundo gráfico apresentamos o estimador $\hat{S}_2(t)$ da função de sobrevivência da distribuição Weibull(1,2), $S_2(t) = e^{-t^2}$.

Notamos que observa-se uma maior proximidade das curvas no primeiro caso, i.é, na estimação da função de sobrevivência da distribuição exponencial de média 1.

4.2.2 Análise bayesiana sob uma função de perda 0-1

Consideramos para o problema de estimação do vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$, a seguinte função de perda 0-1.

Para cada t , seja

$$L(\mathcal{S}^*(t), \tilde{\mathcal{S}}^*(t)) = \begin{cases} 0 & \text{se } \|\mathcal{S}^*(t) - \tilde{\mathcal{S}}^*(t)\| \leq c \\ 1 & \text{se } \|\mathcal{S}^*(t) - \tilde{\mathcal{S}}^*(t)\| > c \end{cases} \quad (4.2.23)$$

onde $\mathcal{S}^*(t) = (S_1^*(t), S_2^*(t))$, e $\tilde{\mathcal{S}}^*(t)$ denota o estimador do vetor $\mathcal{S}^*(t)$. Também, $\|\cdot\|$ denota a norma euclidiana em $\mathbb{R}^2$, e $c > 0$.

Para cada t , o risco bayesiano é

$$\mathbb{E}L(\mathcal{S}^*(t), \tilde{\mathcal{S}}^*(t)) = \Pr(\|\mathcal{S}^*(t) - \tilde{\mathcal{S}}^*(t)\| > c) \quad (4.2.24)$$

Para cada t , minimizar o risco bayesiano é equivalente a maximizar $\Pr(\|\mathcal{S}^*(t) - \tilde{\mathcal{S}}^*(t)\| \leq c)$.

Seja $\pi(\cdot/\text{Dados})$ a função de densidade a posteriori do vetor $(S_1^*(t), S_2^*(t))$, i.é., a função de densidade do vetor $(S_1^*(t), S_2^*(t))/\text{Dados}$, cuja distribuição é dada em (4.2.9).

Então o estimador bayesiano do vetor $(S_1^*(\cdot), S_2^*(\cdot))$, sob a função de perda 0-1 dada por 4.2.19, é obtido resolvendo o problema:

Para cada t ,

$$\max_a \int_{\|\theta - a\| \leq c} \pi(\theta/\text{Dados}) d\theta \quad (4.2.25)$$

onde $\theta = (S_1^*(t), S_2^*(t))$ é o vetor de parâmetros.

O anterior é equivalente a encontrar, para cada t , o vetor $\tilde{\theta}$ tal que

$$\pi(\tilde{\theta}/\text{Dados}) = \sup_{\theta} \pi(\theta/\text{Dados}) \quad (4.2.26)$$

i.é., $\tilde{\theta} := (\tilde{S}_1^*(t), \tilde{S}_2^*(t))$ é o vetor modal da distribuição a posteriori.

Definição 4.2.1

O estimador de $\theta, \tilde{\theta}$, definido em (4.2.26) é dito estimador de máxima verossimilhança generalizado ou também estimador de máxima probabilidade a posteriori.

Temos os seguintes resultados:

Proposição 4.2.4

O estimador de máxima probabilidade a posteriori da função de sobrevivência $S(t) = \Pr(Z > t)$ é dado pela expressão:

$$\hat{S}(t) = q'_n \frac{\alpha_1(t, \infty) + \alpha_2(t, \infty) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) - 2} + (1 - q'_n) S_n(t) \quad (4.2.27)$$

$$\text{onde } q'_n = \frac{\alpha_1(0, \infty) + \alpha_2(0, \infty) - 2}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 2}$$

Prova.

Notamos que de (4.2.9) a distribuição a posteriori de $S(t)$ é $S(t)/\text{Dados} \sim \text{Beta}(\alpha_1(t, \infty) + \alpha_2(t, \infty) + j_1(t) + j_2(t), \alpha_1(0, t] + \alpha_2(0, t] + n - j_1(t) - j_2(t))$ e a moda desta distribuição

(cf. Apêndice A) é:

$$\tilde{S}(t) = \frac{\alpha_1(t, \infty) + \alpha_2(t, \infty) + j_1(t) + j_2(t) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 2},$$

de onde segue (4.2.27). $\square$

Proposição 4.2.5

O estimador de máxima probabilidade a posteriori do vetor de subsobrevivências (S_1^*, S_2^*) é dado pela expressão:

$$\begin{aligned} \tilde{S}^*(t) &= (\tilde{S}_1^*(t), \tilde{S}_2^*(t)) = \\ &= q_n \left(\frac{\alpha_1(t, \infty) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) - 3}, \frac{\alpha_2(t, \infty) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) - 3} \right) \\ &\quad + (1 - q_n)(S_{1n}^*(t), S_{2n}^*(t)), \quad t \in (0, \infty) \end{aligned} \tag{4.2.28}$$

$$\text{onde } q_n = \frac{\alpha_1(0, \infty) + \alpha_2(0, \infty) - 3}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 3}.$$

Prova.

Do Apêndice A, a moda da distribuição de Dirichlet dada pela expressão (4.2.9) é:

$$\tilde{S}^*(t) = (\tilde{S}_1^*(t), \tilde{S}_2^*(t)) = \left(\frac{\alpha_1(t, \infty) + j_1(t) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 3}, \frac{\alpha_2(t, \infty) + j_2(t) - 1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 3} \right)$$

de onde segue (4.2.27) $\square$

Observações 4.2.5

(a) Notamos que o estimador de máxima probabilidade a posteriori do vetor de subsobrevivências $(S_1^*(\cdot), S_2^*(\cdot))$ [da função de sobrevivência $S(\cdot)$] é uma combinação convexa entre o vetor modal [moda] da priori e o vetor de subsobrevivências empíricas [a função de sobrevivência empírica].

(b) Do Teorema de Glivenko-Cantelli e dado que q_n e q'_n convergem a zero quando $n \rightarrow \infty$, os estimadores $\tilde{S}^* = (\tilde{S}_1^*, \tilde{S}_2^*)$ e $\tilde{S}$ são fortemente consistentes para $S^* = (S_1^*, S_2^*)$ e S , respectivamente.

Com base na Proposição 4.2.5 e na formula de Peterson, propomos $\tilde{S}_1(t)$ como estimador da função de sobrevivência $S_1(t) = \Pr(X > t)$, onde

$$\tilde{S}_1(t) = \varphi(\tilde{S}_1^*(\cdot), \tilde{S}_2^*(\cdot), t) \quad t \leq t^* \quad (4.2.29)$$

Teorema 4.2.2

Suponhamos que as funções $f_1(\cdot) = \alpha_1(\cdot, \infty)$ e $f_2(\cdot) = \alpha_2(\cdot, \infty)$ são contínuas em $(0, t)$ e S_1 e S_2 não possuem pontos de descontinuidade em comum. Então,

$$\tilde{S}_1(t) = B(t)I(t)\Pi(t), \quad t \in (0, Z_{(m)}^*] \quad (4.2.30)$$

onde

$$\begin{aligned} B(t) &= \frac{\alpha_1(t, \infty) + \alpha_2(t, \infty) + \sum I(Z_i > t) - 2}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 3}, \\ I(t) &= \exp \left\{ \frac{-1}{\alpha_1(0, \infty) + \alpha_2(0, \infty) + n - 3} \int_0^t \frac{d\alpha_2(s, \infty)}{\tilde{S}_1^*(s) + \tilde{S}_2^*(s)} \right\} \\ &= \exp \left\{ - \int_0^t \frac{d\alpha_2(s)}{\alpha_1(s, \infty) + \alpha_2(s, \infty) + nS_n(s) - 2} \right\} \\ \Pi(t) &= \prod_{i: Z_{(i)}^* \leq t} \frac{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i - d_i - 2}{\alpha_1(Z_{(i)}^*, \infty) + \alpha_2(Z_{(i)}^*, \infty) + n_i - 2}, \end{aligned}$$

onde $Z_{(1)}^* < Z_{(2)}^* < \dots < Z_{(m)}^*$, n_j e d_j foram definidos no enunciado do Teorema 4.2.1.

Prova.

Análoga à prova do Teorema 4.2.1. □

Observação 4.2.6

Um método similar ao descrito na seção 4.2.1.1 pode ser aplicado neste caso, para a obtenção de uma aproximação numérica da expressão $\tilde{S}_1(t) = B(t)I(t)\Pi(t)$ em (4.2.30).

4.3 Estimação bayesiana não-paramétrica na teoria de riscos competitivos

Nesta seção os resultados obtidos na seção anterior são estendidos ao problema de estimacão bayesiana não-paramétrica na teoria de riscos competitivos.

O problema pode ser formulado da seguinte maneira:

Consideramos $(X_{1i}, X_{2i}, \dots, X_{ri})$, $i = 1, 2, \dots, n$, n tuplas de variáveis aleatórias independentes e identicamente distribuídas, tais que $X_{1i} \sim X_1$, $X_{2i} \sim X_2$, $\dots$, $X_{ri} \sim X_r$, $i = 1, 2, \dots, n$. Denotamos por $S_l(t) = \Pr(X_l > t)$, $l = 1, 2, \dots, r$, as respectivas funções de sobrevivência das variáveis $X_1, X_2, \dots, X_r$. Assumimos que as variáveis aleatórias $X_{1i}, X_{2i}, \dots, X_{ri}$ são mutuamente independentes, para todo $i = 1, 2, \dots, n$, e tomando valores no intervalo $(0, \infty)$.

O estatístico só tem disponível os dados:

$$\begin{aligned} Z_i &= \min(X_{1i}, X_{2i}, \dots, X_{ri}) \\ \delta_i &= \begin{cases} 1 & \text{se } Z_i = X_{1i} \\ 2 & \text{se } Z_i = X_{2i} \\ \dots & \\ r & \text{se } Z_i = X_{ri} \end{cases}, \quad i = 1, 2, \dots, n \end{aligned} \quad (4.3.1)$$

o que brevemente escrevemos $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$. A função de sobrevivência da variável Z_i é denotada por $S(t) = \Pr(Z_i > t)$, $i = 1, 2, \dots, n$. Igualmente ao problema com observações censuradas, vamos supor que as funções de sobrevivência $S_l(\cdot)$, $l = 1, 2, \dots, r$, não possuem pontos de descontinuidade em comum.

O modelo descrito corresponde, na teoria da confiabilidade, a um sistema com r componentes em série, e pode ser de interesse a estimacão do tempo de sobrevivência de cada componente do sistema (ou de um grupo de componentes), como também a estimacão do tempo de sobrevivência do sistema. Na terminologia anterior, podemos estar interessados na estimacão das funções de sobrevivência $S_l(\cdot)$, $l = 1, \dots, r$ (ou em algumas delas), funcionais do tipo $\Pr(\delta = l)$, $l = 1, 2, \dots, r$, como também na função de sobrevivência $S(\cdot)$.

As chamadas funções de subsobrevivência são:

$$S_l^*(t) = \Pr(Z > t, \delta = l) \quad , \quad l = 1, 2, \dots, r \quad (4.3.2)$$

Também se Λ é subconjunto de índices não-vazio de $\{1, 2, \dots, r\}$, escrevemos:

$$S_\Lambda^*(t) = \Pr(Z > t, \delta \in \Lambda) \quad , \quad (4.3.3)$$

que é dita a função de subsobrevivência do conjunto de riscos Λ .

Existem propriedades análogas às já expostas na seção 4.1, algumas delas são:

- (i) $S_\ell^*(0) = \Pr(Z > 0, \delta = \ell) = \Pr(\delta = \ell)$, $\ell = 1, 2, \dots, r$, $S_1^*(0) + \dots + S_r(0) = 1$ e $\rho_\Lambda := S_\Lambda^*(0) = \Pr(\delta \in \Lambda)$, que é a confiabilidade do grupo de componentes Λ .
- (ii) $S_1^*(t) + \dots + S_r^*(t) = \Pr(Z > t) = S(t)$.
- (iii) Os pontos de descontinuidade de $S_l^*(\cdot)$ são os pontos de descontinuidade de $S_l(\cdot)$, $l = 1, 2, \dots, r$. E portanto, as funções $S_l^*(\cdot)$ ($l = 1, 2, \dots, r$) não possuem pontos de descontinuidade em comum.
- (iv) Neste caso, $t^* = \min(t_{S_1}, \dots, t_{S_r}) = \max(t_{S_1^*}, \dots, t_{S_r^*})$, onde lembramos que $t_G := \sup\{t > 0 : G(t) > 0\}$.
- (v) A Fórmula de Peterson (cf. Teorema 4.1.1) pode ser generalizada.

Seja Λ um subconjunto de índices não-vazio de $\{1, 2, \dots, r\}$, e denotamos por

$$S_\Lambda(t) = \Pr(\min_{l \in \Lambda} (X_l) > t) \quad (4.3.4)$$

então para cada $t \leq t^*$, temos que

$$S_\Lambda(t) = \varphi(S_\Lambda^*(\cdot), S_{\Lambda^c}^*(\cdot); t) \quad , \quad (4.3.5)$$

onde φ foi definido no Teorema 4.1.1, e S_Λ^* em (4.3.3).

Como foi formulado em (4.3.1), as observações são $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$, onde $Z_i = \min(X_{li} : l = 1, 2, \dots, r)$, e $\delta_i = l$, se $Z_i = X_{li}$; $l = 1, 2, \dots, r$. Equivalentemente, para

cada t , observamos as variáveis aleatórias

$$nS_{ln}^*(t) = \sum_{i=1}^n I(Z_i > t, \delta_i = l) \quad , \quad l = 1, 2, \dots, r \quad (4.3.6)$$

onde as funções $S_{ln}^*(\cdot)$, $l = 1, 2, \dots, r$, são chamadas funções de subsobrevivência empírica.

Notamos que

$$nS_n(t) = n \sum_{l=1}^r S_{ln}^*(t) \quad (4.3.7)$$

Logo, a função de verossimilhança pode ser expressa em termos de um modelo Multinomial $(n; p_1(t), \dots, p_r(t), p_{r+1}(t))$, sendo $p_l(t) = S_l^*(t)$, $l = 1, \dots, r$ e $p_{r+1}(t) = 1 - S(t)$, i.é:

$$\begin{aligned} \text{Lik} &= \Pr(nS_{1n}^*(t) = j_1(t), \dots, nS_{rn}^*(t) = j_r(t)) \\ &\propto S_1^*(t)^{j_1(t)} S_2^*(t)^{j_2(t)} \dots S_r^*(t)^{j_r(t)} (1 - S(t))^{n - \sum_{l=1}^r j_l(t)} \end{aligned} \quad (4.3.8)$$

A distribuição a priori para o vetor de subsobrevivências $(S_1^*(\cdot), \dots, S_r^*(\cdot))$ é construída a partir da caracterização do processo de Dirichlet r -variado dada no Teorema 3.4.2.

Consideramos $\mathcal{X} = (0, \infty)$ e $\mathcal{A} = \mathcal{B}(0, \infty)$, $\vec{\alpha} = (\alpha_1, \dots, \alpha_r)$ vetor de r medidas finitas, não-nulas, não-negativas e finitamente aditivas sobre $(\mathcal{X}, \mathcal{A})$.

Como antes, definimos

$$\rho_l = \Pr(\delta = l) \quad l = 1, 2, \dots, r \quad (4.3.9)$$

$$T_l^*(t) = \Pr(Z > t/\delta = l) \quad l = 1, 2, \dots, r \quad (4.3.10)$$

e temos que

$$(S_1^*(\cdot), \dots, S_r^*(\cdot)) = (\rho_1 T_1^*(\cdot), \dots, \rho_r T_r^*(\cdot)) \quad (4.3.11)$$

O seguinte lema é consequência do Teorema 3.4.2, lembramos antes que $\vec{\alpha}(A) := (\alpha_1(A), \dots, \alpha_r(A))$.

Lema 4.3.1

Se $(\rho_1, \dots, \rho_r) \sim \mathcal{D}_s(\vec{\alpha}(0, \infty))$, $T_l^* \in \mathcal{D}(\alpha_l)$, $l = 1, 2, \dots, r$, tais que $(\rho_1, \dots, \rho_r) \amalg T_1^* \dots \amalg T_r^* / \vec{\alpha}$. Então $(S_1^*, \dots, S_r^*) \in \mathcal{DM}_r(\vec{\alpha})$. $\square$

Portanto, para cada t , temos como distribuição a priori

$$(S_1^*(t), \dots, S_r^*(t)) \sim \mathcal{D}(\vec{\alpha}(t, \infty); \sum_{l=1}^r \alpha_l(0, t]) \quad (4.3.12)$$

ou na sua forma singular,

$$(S_1^*(t), \dots, S_r^*(t), 1 - S(t)) \sim \mathcal{D}_s(\vec{\alpha}(t, \infty), \sum_{l=1}^r \alpha_l(0, t]) \quad (4.3.13)$$

Agora, utilizando a Proposição 1 do Apêndice, obtemos a distribuição a posteriori, para cada t ,

$$(S_1^*(t), \dots, S_r^*(t))/\text{Dados} \sim \mathcal{D}(\vec{\beta}(t, \infty); \sum_{l=1}^r \alpha_l(0, t] + n - \sum_{l=1}^r j_l(t)) \quad (4.3.14)$$

ou também

$$(S_1^*(t), \dots, S_r^*(t), 1 - S(t))/\text{Dados} \sim \mathcal{D}_s(\vec{\beta}(t, \infty), \sum_{l=1}^r \alpha_l(0, t] + n - \sum_{l=1}^r j_l(t)) \quad (4.3.15)$$

onde $\vec{\beta}(t, \infty) = (\alpha_1(t, \infty) + j_1(t), \dots, \alpha_r(t, \infty) + j_r(t))$.

4.3.1 Análise sob uma função de perda quadrática

Consideramos a seguinte função de perda quadrática:

$$L(\mathcal{S}^*, \hat{\mathcal{S}}^*) = \int \| \mathcal{S}^*(t) - \hat{\mathcal{S}}^*(t) \|^2 dW(t) \quad (4.3.16)$$

onde $\mathcal{S}^*(t) = (S_1^*(t), \dots, S_r^*(t))$, $\hat{\mathcal{S}}^*(t) = (\hat{S}_1^*(t), \dots, \hat{S}_r^*(t))$ é o estimador de $\mathcal{S}^*(t)$, $\| \cdot \|$ é a norma euclidiana em $\mathbb{R}^r$, e W é uma função de peso.

Os resultados obtidos nesta situação são análogos aos obtidos na seção 4.2.1 deste trabalho e são resumidos na seguinte proposição.

Denotamos por Λ um subconjunto de índices não-vazio de $\{1, \dots, r\}$.

Proposição 4.3.1

(a) Considerando a função de perda quadrática (4.3.16) e o processo de Dirichlet r -variado $DM_r(\vec{\alpha})$ como distribuição a priori temos que:

- (i) O estimador bayesiano do vetor de subsobrevivências $(S_\ell^*(\cdot))_{\ell \in \Lambda}$ é dado por $(\hat{S}_\ell^*(\cdot))_{\ell \in \Lambda}$, onde

$$\hat{S}_\ell^*(t) = p'_n \frac{\alpha_\ell(t, \infty)}{\sum_{\ell=1}^r \alpha_\ell(0, \infty)} + (1 - p'_n) S_{\ell_n}^*(t) \quad (4.3.17)$$

onde

$$p'_n = \frac{\sum_{\ell=1}^r \alpha_\ell(0, \infty)}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) + n} \quad (4.3.18)$$

e $S_{\ell_n}^*(t)$, $\ell \in \Lambda$, são as funções de subsobrevivência empírica definidas em (4.3.6).

- (ii) O estimador bayesiano da função de subsobrevivência do conjunto de riscos Λ , $S_\Lambda^*(t) = \Pr(Z > t, \delta \in \Lambda)$ é dado por

$$\hat{S}_\Lambda^*(t) = \sum_{\ell \in \Lambda} \hat{S}_\ell^*(t) = p'_n \frac{\sum_{\ell \in \Lambda} \alpha_\ell(t, \infty)}{\sum_{\ell=1}^r \alpha_\ell(0, \infty)} + (1 - p'_n) S_{\Lambda_n}^*(t) \quad (4.3.19)$$

onde $S_{\Lambda_n}^*$ é a função de subsobrevivência empírica do conjunto de riscos Λ , i.é.

$$S_{\Lambda_n}^*(t) = \sum_{\ell \in \Lambda} S_{\ell_n}^* = \frac{1}{n} \sum_{i=1}^n \sum_{\ell \in \Lambda} I(Z_i > t, \delta_i = \ell) = \frac{1}{n} \sum_{i=1}^n I(Z_i > t, \delta_i \in \Lambda). \quad (4.3.20)$$

Em particular, se $\Lambda = \{1, 2, \dots, r\}$ obtemos o estimador bayesiano da função de sobrevivência do sistema $S(t) = \Pr(Z > t)$,

$$\hat{S}(t) = p'_n \frac{\sum_{\ell=1}^r \alpha_\ell(t, \infty)}{\sum_{\ell=1}^r \alpha_\ell(0, \infty)} + (1 - p'_n) S_n(t) \quad (4.3.21)$$

onde $S_n(t)$ é a função de distribuição empírica correspondente a $Z_1, \dots, Z_n$.

- (iii) Se as funções $f_\ell(\cdot) = \alpha_\ell(\cdot, \infty)$, $\ell \in \Lambda$, são contínuas pela direita no ponto zero, o estimador bayesiano de $\rho_\Lambda = \Pr(\delta \in \Lambda)$ é dado por

$$\hat{\rho}_\Lambda = \lim_{t \downarrow 0} \hat{S}_\Lambda^*(t) = p'_n \frac{\sum_{\ell \in \Lambda} \alpha_\ell(0, \infty)}{\sum_{\ell=1}^r \alpha_\ell(0, \infty)} + (1 - p'_n) \frac{1}{n} \sum_{i=1}^n I(\delta_i \in \Lambda) \quad (4.3.22)$$

- (b) Se as funções $f_\ell(\cdot) = \alpha_\ell(\cdot, \infty)$, $\ell = 1, \dots, r$ são contínuas em $(0, t)$, e se as funções de subsobrevivência S_Λ^* e $S_{\Lambda^c}^*$ não possuem pontos de descontinuidade em comum, então

o estimador não-paramétrico, definido pela fórmula de Peterson generalizada (4.3.5), da função de sobrevivência do conjunto riscos Λ , $S_\Lambda(\cdot)$:

$$\hat{S}_\Lambda(t) = \varphi(\hat{S}_\Lambda^*(\cdot), \hat{S}_{\Lambda^c}^*(\cdot); t), \quad t \in (0, t^*]$$

onde $t^* = \min\{t_{S_\ell} : \ell = 1, \dots, r\}$ é dado por

$$\hat{S}_\Lambda(t) = \hat{S}(t) I_\Lambda(t) \Pi_r(t), \quad t \in (0, Z_{(m)}^*] \quad (4.3.23)$$

sendo $\hat{S}(t)$ o estimador bayesiano de $S(t)$ dado por (4.3.21),

$$\begin{aligned} I_\Lambda(t) &= \exp \left\{ \frac{-1}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) + n} \sum_{\ell \in \Lambda^c} \int_0^t \frac{d\alpha_\ell(s, \infty)}{\hat{S}(s)} \right\} \\ \Pi_r(t) &= \prod_{i: Z_{(i)}^* \leq t} \frac{\sum_{\ell=1}^r \alpha_\ell(Z_{(i)}^*, \infty) + n_i - d_i}{\sum_{\ell=1}^r \alpha_\ell(Z_{(i)}^*, \infty) + n_i} \end{aligned}$$

Prova.

É uma direta generalização das Proposições 4.2.1 - 4.2.3 e do Teorema 4.2.1. $\square$

Observações 4.3.1

(a) É claro que os estimadores $(\hat{S}_\ell^*)_{\ell \in \Lambda}$, $\hat{S}_\Lambda^*$ e $\hat{\rho}_\Lambda$ são combinações convexas entre as médias das respectivas distribuições a priori, e os respectivos elementos empíricos. Notamos também que esses estimadores são fortemente consistentes para $(S_\ell^*)_{\ell \in \Lambda}$, S_Λ^* e ρ_Λ , respectivamente.

(b) Os mesmos argumentos dados na prova do Teorema 4.2.1 garantem a existência das integrais $\int_0^t \frac{d\alpha_\ell(s, \infty)}{\hat{S}(s)}$, $\ell \in \Lambda$ na expressão (4.3.23).

(c) Notamos que $\hat{S}_\Lambda(t) = \hat{S}(t) I_\Lambda(t) \Pi_r(t)$, $t \in (0, Z_{(m)}^*]$ em (4.3.23) define a família dos $2^r - 1$ estimadores não-paramétricos baseados na fórmula de Peterson, associados às $2^r - 1$ possíveis funções de sobrevivência no modelo de riscos competitivos.

Em particular se Λ é unitário, digamos $\Lambda = \{\ell\}$ a expressão (4.3.23) define o estimador não-paramétrico da função de sobrevivência $S_\ell(t) = \Pr(X_\ell > t)$.

(d) Se o conjunto de riscos é $\{1, 2\}$ e $\Lambda = \{1\}$, então (4.3.23) é simplesmente o estimador da função de sobrevivência $S_1(t) = \Pr(X_1 > t)$, dado pela expressão (4.2.20) no Teorema 4.2.1.

(e) O algoritmo descrito na seção 4.2.1.1 pode ser aplicado no cálculo numérico do estimador $\hat{S}_\Lambda(t)$.

4.3.2 Análise sob uma Função de Perda 0 - 1

Consideramos uma função de perda 0 - 1 análoga à descrita em (4.2.19), i.é, para cada t

$$L(S^*(t), \tilde{S}^*(t)) = \begin{cases} 0 & \text{se } \|S^*(t) - \tilde{S}^*(t)\| \leq c \\ 1 & \text{se } \|S^*(t) - \tilde{S}^*(t)\| > c \end{cases} \quad (4.3.24)$$

onde, neste caso, $S^*(t) = (S_1^*(t), \dots, S_r^*(t))$ é o vetor de subsobrevivências, e $\tilde{S}^*(t)$ é o estimador de $S^*(t)$. Também, $\|\cdot\|$ denota a norma euclidiana em $\mathbb{R}^r$, e $c > 0$.

Nesta seção, estamos particularmente interessados na estimação da função de sobrevivência do grupo $\Lambda, S_\Lambda(\cdot)$, definida em (4.3.4). Os cálculos são baseados fundamentalmente nas distribuições a priori, e logo a posteriori, do vetor de subsobrevivências $(S_1^*(\cdot), \dots, S_r^*(\cdot))$, dadas pelas expressões (4.3.12) e (4.3.14).

Proposição 4.3.2

(a) O estimador de máxima probabilidade a posteriori da função de sobrevivência do sistema $S(t) = \Pr(Z > t)$ é dado por:

$$\tilde{S}(t) = q'_n \frac{\sum_{\ell=1}^r \alpha_\ell(t, \infty) - 1}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) - 2} + (1 - q'_n) S_n(t) \quad (4.3.24)$$

$$\text{onde } q'_n = \frac{\sum_{\ell=1}^r \alpha_\ell(0, \infty) - 2}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) + n - 2}.$$

(b) O estimador de máxima probabilidade a posteriori do vetor de subsobrevivências

$S^*(\cdot) = (S_\Lambda^*(\cdot), S_{\Lambda^c}^*(\cdot))$, onde $S_\Lambda^*(t) = \Pr(Z > t, \delta \in \Lambda)$, é dado pela expressão:

$$\begin{aligned} \tilde{S}^*(t) &= (\tilde{S}_\Lambda^*(t), \tilde{S}_{\Lambda^c}^*(t)) \\ &= q_n \left(\frac{\sum_{\ell \in \Lambda} \alpha_\ell(t, \infty) - 1}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) - 3}, \frac{\sum_{\ell \in \Lambda^c} \alpha_\ell(t, \infty) - 1}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) - 3} \right) + (1 - q_n)(S_{\Lambda_n}^*(t), S_{\Lambda_n^c}^*(t)) \end{aligned} \quad (4.3.25)$$

onde $q_n = \frac{\sum_{\ell=1}^r \alpha_\ell(0, \infty) - 3}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) + n - 3}$ e $S_{\Lambda_n}^*(\cdot)$ é a função de subsobrevivência empírica do conjunto de risco Λ definida em (4.3.20)

(c) Se as funções $f_\ell(\cdot) = \alpha_\ell(\cdot, \infty)$, $\ell = 1, \dots, r$ são contínuas em $(0, t)$ e se as funções de subsobrevivência S_Λ^* e $S_{\Lambda^c}^*$ não possuem pontos de descontinuidade em comum, então o estimador não-paramétrico, definido pela fórmula de Peterson generalizada (4.3.5), da função de sobrevivência $S_\Lambda(t) = \Pr(\min_{\ell \in \Lambda} X_\ell > t)$:

$$\tilde{S}_\Lambda(t) = \varphi(\tilde{S}_\Lambda^*(\cdot), \tilde{S}_{\Lambda^c}^*(\cdot); t), \quad t \in (0, t^*]$$

é dado por

$$\tilde{S}_\Lambda(t) = B(t)I_\Lambda(t)\Pi_r(t), \quad t \in (0, Z_{(m)}^*] \quad (4.3.26)$$

onde

$$\begin{aligned} B(t) &= \frac{\sum_{\ell=1}^r \alpha_\ell(t, \infty) + \sum_{i=1}^n I(Z_i > t) - 2}{\sum_{\ell=1}^r \alpha_\ell(0, \infty) + n - 3}, \\ I_\Lambda(t) &= \exp \left\{ - \sum_{\ell \in \Lambda^c} \int_0^t \frac{d\alpha_\ell(s, \infty)}{\sum_{\ell=1}^r \alpha_\ell(s, \infty) + nS_n(s) - 2} \right\}, \\ \Pi_r(t) &= \prod_{i: Z_{(i)}^* \leq t} \frac{\sum_{\ell=1}^r \alpha_\ell(Z_{(i)}^*, \infty) + n_i - d_i - 2}{\sum_{\ell=1}^r \alpha_\ell(Z_{(i)}^*, \infty) + n_i - 2}, \end{aligned}$$

onde $Z_{(1)}^* < \dots < Z_{(m)}^*$, n_j d_j foram definidas no enunciado do Teorema 4.2.1

Prova.

São diretas generalizações das Proposições 4.2.4 e 4.2.5 e do Teorema 4.2.2, respectivamente. $\square$

Observações 4.3.2

- (a) Os estimadores $\tilde{S}_\ell$ e $(\tilde{S}_\Lambda^*, \tilde{S}_{\Lambda^c}^*)$, são combinações convexas entre as modas das respectivas distribuições a priori e os respectivos elementos empíricos. Notamos também que esses estimadores são fortemente consistentes para S e $(S_\Lambda^*, S_{\Lambda^c}^*)$, respectivamente.
- (b) Valem aqui também as observações 4.3.1 (b) - (e).

CAPÍTULO 5

Propriedades assintóticas das soluções obtidas

5.1 Introdução

Este capítulo é dedicado ao estudo assintótico das soluções obtidas no capítulo anterior. Notamos que resultados assintóticos para estimadores de algumas quantidades de interesse, tanto na análise sob função de perda quadrática, quanto na análise sob função de perda 0 – 1, foram obtidos no capítulo anterior. Esses resultados aparecem descritos nas observações 4.2.2(b), 4.3.3(b)-(c), 4.2.5(b), 4.3.1(a) e 4.3.2(a).

Assim, o objetivo principal deste capítulo é estudar a convergência do estimador da função de sobrevivência $S_1(t) = \Pr(X > t)$ no problema com dados aleatoriamente censurados, o qual foi nosso tema de interesse na Seção 4.2. Especificamente, estudamos a consistência, a convergência fraca, e estabelecemos uma lei do logaritmo iterado para o estimador não-paramétrico $\hat{S}_1$, obtido com base em uma análise bayesiana sob função de perda quadrática e na fórmula de Peterson (ver Teorema 4.2.1).

A consistência forte de $\hat{S}_1$ é provada utilizando argumentos de continuidade do funcional φ da fórmula de Peterson definido na expressão (4.1.6). Por outro lado, para a convergência fraca e na obtenção de uma lei do logaritmo iterado para $\hat{S}_1$, utilizamos a representação funcional de estatísticas (cf. von Mises, 1947), e resultados conhecidos referentes à convergência fraca e lei do logaritmo iterado para processos empíricos. Os conceitos e resultados utilizados aparecem descritos no Apêndice B.

A extensão dos resultados para o estimador da função de sobrevivência corresponde a um conjunto de riscos $\hat{S}_\Lambda$, como também de outros estimadores, digamos $\tilde{S}_1, \tilde{S}_\Lambda$, é uma consequência direta dos desenvolvimentos feitos no estudo das propriedades assintóticas

de $\hat{S}_1$.

No que segue utilizamos o fato que

$$\hat{S}_1(t) = \varphi(\hat{S}_1^*(\cdot), \hat{S}_2^*(\cdot); t), \quad t \leq t^* \quad (5.1.1)$$

onde $\hat{S}_1^*(\cdot)$ e $\hat{S}_2^*(\cdot)$ são os estimadores bayesianos sob função de perda quadrática das funções de subsobrevivência $S_1^*(\cdot)$ e $S_2^*(\cdot)$, dados pelas expressões (4.2.14) e (4.2.15), e $t^* = \min(t_{S_1}, t_{S_2}) = \max(t_{S_1^*}, t_{S_2^*})$, onde $t_G := \sup(t > 0 : G(t) > 0)$.

5.2 Propriedades assintóticas do estimador $\hat{S}_1$

Nesta seção estudamos as propriedades assintóticas do estimador não-paramétrico da função de sobrevivência $S_1(t) = \Pr(X > t)$ no problema com dados aleatoriamente censurados.

5.2.1 Consistência forte de $\hat{S}_1$

Primeiramente consideramos o seguinte lema, o qual requeremos para provar a consistência forte do estimador não-paramétrico $\hat{S}_1$. Esse lema estabelece a continuidade do funcional φ da fórmula de Peterson.

Lema 5.2.1

Seja $D[0, T]$ o espaço das funções *cadlag* sobre $[0, T]$ munido com a norma do supremo $\|F\|_T = \sup_{0 \leq t \leq T} |F(t)|$. Denotamos $D[0, T] \times D[0, T]$ o espaço produto, onde $\|(F, G)\|_T = \max(\|F\|_T, \|G\|_T)$. Então o funcional φ , com domínio $D[0, T] \times D[0, T]$ e assumindo valores em $D[0, T]$, definido por

$$\varphi(F, G)(t) = \exp \left\{ c \int_0^t \frac{dF(s)}{F(s) + G(s)} \right\} \prod_{\substack{s \leq t \\ \text{s: ponto de} \\ \text{salto de } F(\cdot)}} \frac{F(s^+) + G(s^+)}{F(s^-) + G(s^-)}$$

é um funcional contínuo.

Prova.

A demonstração é feita considerando uma sequência (F_n, G_n) em $D[0, T] \times D[0, T]$, tal que $\|(F_n, G_n) - (F, G)\|_T \rightarrow 0$ quando $n \rightarrow \infty$, para algum $(F, G) \in D[0, T] \times D[0, T]$.

Seja $H_n(\cdot) = \varphi(F_n, G_n)(\cdot)$, então é possível provar que $\|H_n - H\|_T \rightarrow 0$, quando $n \rightarrow \infty$, onde $H(\cdot) = \varphi(F, G)(\cdot)$. $\square$

Proposição 5.2.1

O estimador $\hat{S}_1$ é fortemente consistente, i.é, $\|\hat{S}_1 - S_1\|_{t^*} = \sup_{0 \leq t \leq t^*} |\hat{S}_1(t) - S(t)| \rightarrow 0$, quando $n \rightarrow \infty$, com probabilidade 1.

Prova.

É uma consequência direta do lema anterior e da consistência forte do estimador bayesiano do vetor de subsobrevivências $(\hat{S}_1^*, \hat{S}_2^*)$ (cf. Observação 4.2.2(b)). $\square$

5.2.2 Convergência fraca e lei do logaritmo iterado para $\hat{S}_1$

Para estabelecer a convergência fraca do estimador $\hat{S}_1$, primeiro mostramos as convergências fracas do estimador vetorial $(\hat{S}_1^*, \hat{S}_2^*)$ e do estimador $\hat{S}$ da função de sobrevivência $S(t) = \Pr(Z > t)$

Proposição 5.2.2

(a) Definamos $(U_n^*, V_n^*) \in D[0, t^*] \times D[0, t^*]$ por $U_n^* = n^{1/2}(\hat{S}_1^* - S_1^*)$ e $V_n^* = n^{1/2}(\hat{S}_2^* - S_2^*)$. Então o processo (U_n^*, V_n^*) converge fracamente a um processo gaussiano bivariado (U^*, V^*) com média $(0, 0)$ e núcleo de covariância, para $s \leq t < t^*$, dado por

$$\begin{aligned} \text{cov}(U^*(s), U^*(t)) &= (1 - S_1^*(s))S_1^*(t) \\ \text{cov}(U^*(s), V^*(t)) &= -S_1^*(s)S_2^*(t) \\ \text{cov}(V^*(s), U^*(t)) &= -S_2^*(s)S_1^*(t) \\ \text{cov}(V^*(s), V^*(t)) &= (1 - S_2^*(s))S_2^*(t) \end{aligned} \tag{5.2.1}$$

(b) Definamos $W_n \in D[0, t^*]$ por $W_n = n^{1/2}(\hat{S} - S)$, onde $\hat{S}$ é o estimador bayesiano da função de sobrevivência $S(t) = \Pr(Z > t)$. Então o processo W_n converge fracamente a um processo gaussiano W com média 0 e núcleo de covariância dado, para $s \leq t \leq t^*$, por

$$\text{cov}(W(s), W(t)) = (1 - S(s))S(t) \quad (5.2.2)$$

Prova.

$$(a) \quad (U_n^*(t), V_n^*(t)) = n^{1/2}p_n \left[\left(\frac{\alpha_1(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)}, \frac{\alpha_2(t, \infty)}{\alpha_1(0, \infty) + \alpha_2(0, \infty)} \right) - (S_1^*(t), S_2^*(t)) \right] \\ + n^{1/2}(1 - p_n)[(S_{1n}^*(t), S_{2n}^*(t)) - (S_1^*(t), S_2^*(t))]$$

Dado que $p_n \downarrow 0$ e $n^{1/2}p_n = o(n^{-1/2})$ quando $n \uparrow \infty$, é suficiente estudar a convergência fraca do processo empírico $(E_n^1, E_n^2) = n^{1/2}[(S_{1n}^*, S_{2n}^*) - (S_1^*, S_2^*)]$.

Mas, da estrutura do processo empírico,

$$(E_n^1(t), E_n^2(t)) = n^{1/2} \sum_{i=1}^n [I(Z_i > t, \delta_i = 1) - S_1^*(t), I(Z_i > t, \delta_i = 2) - S_2^*(t)]$$

segue que as distribuições finito-dimensionais de (E_n^1, E_n^2) convergem às distribuições finito-dimensionais do processo (U^*, V^*) , com núcleo de covariância dado, para $s \leq t$, por

$$\begin{aligned} & \text{cov}(U^*(s), U^*(t)) \\ &= \mathbb{E}[I(Z > s, \delta = 1)I(Z > t, \delta = 1)] - \mathbb{E}[I(Z > s, \delta = 1)]\mathbb{E}[I(Z > t, \delta = 1)] \\ &= \mathbb{E}[I(Z > t, \delta = 1)] - \mathbb{E}[I(Z > s, \delta = 1)]\mathbb{E}[I(Z > t, \delta = 1)] \\ &= S_1^*(t)(1 - S_1^*(s)) , \\ & \text{cov}(U^*(s), V^*(t)) \\ &= \mathbb{E}[I(Z > s, \delta = 1)I(Z > t, \delta = 2)] - \mathbb{E}[I(Z > s, \delta = 1)]\mathbb{E}[I(Z > t, \delta = 2)] \\ &= -S_1^*(s)S_2^*(t) , \end{aligned}$$

analogamente para $\text{cov}(V^*(s), U^*(t))$ e $\text{cov}(V^*(s), V^*(t))$ e logo obtemos (5.2.1).

Por outro lado, é bem conhecido (cf. Billingley, 1968, Teorema 16.4) que a sequência de distribuições induzida por E_n^1 é *tight* sobre o espaço $D[0, t^*]$, e o mesmo acontece para E_n^2 . Consequentemente, (E_n^1, E_n^2) induz uma sequência de distribuições *tight* sobre o espaço produto $D[0, t^*] \times D[0, t^*]$. Assim, concluímos a demonstração.

(b) Similar à prova de (a). □

A demonstração da convergência fraca do estimador $\hat{S}_1$ é feita considerando a proposição anterior e uma abordagem baseada no cálculo diferencial em espaços vetoriais normados, a qual considera a representação funcional de estatísticas, como também o teorema da função implícita. Detalhes sobre esses tópicos e suas aplicações na inferência não-paramétrica podem ser encontrados no Apêndice B.

O seguinte lema, que aparece provado em Tsai (1986), estabelece que a função de sobrevivência $S_1(t) = \Pr(X > t)$ é a única solução de uma certa equação implícita. Embora esse lema não seja utilizado diretamente nos desenvolvimentos posteriores, é de interesse conhecer o funcional que define tal equação implícita.

Lema 5.2.2

Se as funções de sobrevivência $S_1(t) = \Pr(X > t)$ e $S_2 = \Pr(Y > t)$ não possuem pontos de descontinuidade em comum, então a equação

$$H(S_1^*, S_2^*, S_1)(t) = 0 \tag{5.2.3}$$

tem como única solução $S_1(t)$ para cada $t \leq t^*$, onde H é definido sobre $D[0, t^*] \times D[0, t^*] \times D[0, t^*]$ em $D[0, t^*]$ pela expressão:

$$H(S_1^*, S_2^*, S_1)(t) = -S_1(t) + S_1^*(t) + S_2(t) - S_1(t) \int_0^t \frac{1}{S_1(u^-)} dS_2^*(u) . \tag{5.2.4}$$

□

Observações 5.2.1

(a) A solução explícita da equação (5.2.4) é dada pela fórmula de Peterson, $S_1(t) = \varphi(S_1^*(\cdot), S_2^*(\cdot); t)$, $t \leq t^*$.

(b) O estimador não-paramétrico $\hat{S}_1$ é solução da equação

$$H(\hat{S}_1^*, \hat{S}_2^*, \hat{S}_1)(t) = 0, \quad t \leq t^* \quad (5.2.5)$$

e, dado que os estimadores $\hat{S}_1^*$ e $\hat{S}_2^*$ são estimadores consistentes de S_1^* e S_2^* , respectivamente, segue que $\hat{S}_1$ é um estimador que possui a *propriedade de auto-consistência generalizada* definida por Tsai (1986).

As propriedades do funcional H são dadas no seguinte lema, cuja demonstração é análoga à do Lema 4.3 em Tsai e Crowley (1985).

Lema 5.2.3

O funcional H é continuamente diferenciável no ponto (S_1^*, S_2^*, S_1) , suas derivadas parciais calculadas nesse ponto e avaliadas em certos pontos de interesse são dadas pelas expressões

$$\begin{aligned} H'_1(S_1^*, S_2^*, S_1; \hat{S}_1^* - S_1^*)(t) &= (\hat{S}_1^* - S_1^*)(t) \\ H'_2(S_1^*, S_2^*, S_1; \hat{S}_2^* - S_2^*)(t) &= (\hat{S}_2^* - S_2^*)(t) - S_1(t) \int_0^t \frac{d(\hat{S}_2^* - S_2^*)(u)}{S_1(u^-)} \\ H'_3(S_1^*, S_2^*, S_1; \hat{S}_1 - S_1)(t) &= (\hat{S}_1 - S_1)(t) - (\hat{S}_1 - S_1)(t) \int_0^t \frac{d(S_2^*(u))}{S_1(u^-)} \\ &\quad + S_1(t) \int_0^t \frac{(\hat{S}_1 - S_1)(u)}{[S_1(u^-)]^2} dS_2^*(u), \end{aligned} \quad (5.2.6)$$

onde H'_i denota o diferencial parcial de H em relação à i -ésima variável, $i = 1, 2, 3$.

Além disso, H'_3 possui inversa e

$$\begin{aligned} & -[H'_3(S_1^*, S_2^*, S_1; H'_1(S_1^*, S_2^*, \hat{S}_1^* - S_1^*) + H'_2(S_1^*, S_2^*, \hat{S}_2^* - S_2^*))^{-1}(t) \\ &= -S_1(t) \left[\int_0^t \frac{(\hat{S}_1 - S_1)(u)}{[S(u^-)]^2} dS_1^*(u) + \frac{(\hat{S}_1^* - S_1^*)(t)}{S(t)} - \int_0^t \frac{(\hat{S}_1^* - S_1^*)(u)}{[S(u^-)]^2} dS(u) \right] \end{aligned} \quad (5.2.7)$$

O resultado mais importante desta parte é estabelecido no seguinte teorema.

Teorema 5.2.1

- (a) Se as funções de sobrevivência S_1 e S_2 não possuem pontos de descontinuidade em comum, então o processo $n^{1/2}(\hat{S}_1 - S_1)$ converge fracamente ao processo gaussiano

$$U(t) = -S_1(t) \left[\int_0^t \frac{W(u)}{[S(u^-)]^2} dS_1^*(u) + \frac{U^*(t)}{S(t)} - \int_0^t \frac{U^*(u)}{[S(u^-)]^2} dS(u) \right], \quad (5.2.8)$$

com núcleo de covariância dado por

$$\text{cov}(U(s), U(t)) = S_1(s)S_1(t) \int_0^s \frac{dS_1^*(u)}{[S(u^-)]^2}, \quad s \leq t \leq t^*, \quad (5.2.9)$$

onde U^* e W são os processos gaussianos definidos em (5.2.1) e (5.2.2).

- (b) Se as funções de sobrevivência S_1 e S_2 são contínuas, então

$$\|\hat{S}_1 - S_1\|_{t^*} = \sup_{0 \leq t \leq t^*} |\hat{S}_1(t) - S_1(t)| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ q.c.}, \quad (5.2.10)$$

quando $n \rightarrow \infty$.

Prova.

$$\begin{aligned} (a) \quad n^{1/2}(\hat{S}_1 - S_1) &= n^{1/2}[\varphi(\hat{S}_1^*, S_2^*) - \varphi(S_1^*, S_2^*)] \\ &= \varphi'(S_1^*, S_2^*; n^{1/2}[(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)]) + o(n^{1/2}\|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*}) \end{aligned} \quad (5.2.11)$$

onde φ' denota diferencial de Frechet.

Do Teorema 1 do Apêndice B temos que $n^{1/2}\|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*} = O_P(1)$, isto implica que o segundo termo de (5.2.11) é $o_P(1)$.

Para estudar a convergência do primeiro termo de (5.2.11), usamos o Lema 5.2.3 (notamos que esse lema implica que H satisfaz as condições do teorema da função implícita) e os argumentos descritos na última seção do Apêndice B.

Assim, usando o teorema da função implícita e (5.2.7),

$$\begin{aligned} n^{1/2}(\hat{S}_1 - S_1)(t) &= -[H'_3(S_1^*, S_2^*, S_1; H'_1(S_1^*, S_2^*; n^{1/2}(\hat{S}_1^* - S_1^*))) \\ &\quad + H'_2(S_1^*, S_2^*; n^{1/2}(\hat{S}_2^* - S_2^*))]^{-1}(t) + o_P(1) \\ &= -S_1(t)[A_n(t) + B_n(t)] + o_P(1), \end{aligned}$$

onde

$$\begin{aligned} A_n(t) &= \int_0^t n^{1/2} \frac{(\hat{S} - S)(u)}{[S(u^-)]^2} dS_1^*(u) \\ B_n(t) &= n^{1/2} \frac{(\hat{S}_1^* - S_1^*)(t)}{S(t)} - \int_0^t n^{1/2} \frac{(\hat{S}_1^* - S_1^*)(u)}{[S(u^-)]^2} dS(u) \end{aligned} \quad (5.2.12)$$

Usando argumentos da seção 4 do Apêndice B e o fato que $n^{1/2}(\hat{S}_1^* - S_1^*) \xrightarrow{\omega} U^*$ e $n^{1/2}(\hat{S} - S) \xrightarrow{\omega} W$ (ver Proposição 5.5.2), obtemos que $n^{1/2}(\hat{S}_1 - S_1)$ converge fracamente ao processo gaussiano U , onde o núcleo de covariância do processo U segue dos resultados de Breslow e Crowley (1974). Notamos que $A_n(t)$ e $B_n(t)$ em (5.2.12) são análogos aos termos A_N e B_N em Breslow e Crowley (1974).

$$\begin{aligned} (b) \quad \|\hat{S}_1 - S_1\|_{t^*} &= \|\varphi(\hat{S}_1^*, \hat{S}_2^*) - \varphi(S_1^*, S_2^*)\|_{t^*} \\ &\leq \|\varphi'(S_1^*, S_2^*; (\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*))\|_{t^*} + o(\|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*}) \\ &\leq \|\varphi'\| \|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*} + o(\|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*}), \end{aligned}$$

onde $\|\varphi'\|$ é a norma do funcional linear φ' . Notamos também que dado que $\varphi'(F, G; \cdot)$ é linear e contínuo, então ele é limitado.

Por outro lado, da hipótese de continuidade de S_1 e S_2 , decorre a continuidade de S_1^* e S_2^* , e logo vale a lei do logaritmo iterado para $\hat{S}_1^*$ e $\hat{S}_2^*$ (cf. Apêndice B), i.é,

$$\|\hat{S}_1^* - S_1\|_{t^*} = \|\hat{S}_2^* - S_2\|_{t^*} = O\left(\sqrt{\frac{\log \log n}{n}}\right),$$

e portanto

$$\|(\hat{S}_1^*, \hat{S}_2^*) - (S_1^*, S_2^*)\|_{t^*} = \max\{\|\hat{S}_1^* - S_1\|_{t^*}, \|\hat{S}_2^* - S_2\|_{t^*}\} = O\left(\sqrt{\frac{\log \log n}{n}}\right).$$

Assim, obtemos (5.2.10). □

APÊNDICE A

A distribuição de Dirichlet

Este apêndice é dedicado ao estudo da distribuição de Dirichlet, suas caracterizações e propriedades.

Definição

Seja $\alpha_i \geq 0$ $i = 1, 2, \dots, k+1$, com $\sum_{i=1}^{k+1} \alpha_i > 0$. Dizemos que o vetor $(Y_1, \dots, Y_k)$ tem distribuição de Dirichlet com parâmetro $(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$, se a função de densidade conjunta de $(Y_1, \dots, Y_k)$ é

$$f(y_1, \dots, y_k; \alpha_1, \dots, \alpha_k; \alpha_{k+1}) = \frac{\Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} \prod_{i=1}^k y_i^{\alpha_i-1} \left(1 - \sum_{i=1}^k y_i\right)^{\alpha_{k+1}-1}$$

definida sobre el simplex k -dimensional

$$\mathcal{S}_k = \{(y_1, \dots, y_k) : y_i \geq 0 \ i = 1, 2, \dots, k, \sum_{i=1}^k y_i < 1\}$$

Notação: $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$

- Se $\alpha_i = 0$ para algum i , então a correspondente $Y_i = 0$ q.c.
- Se $\alpha_1 = \alpha_2 = \dots = \alpha_{k+1} = 1$ então $(Y_1, \dots, Y_k) \sim U(\mathcal{S}_k)$.
- Se $k = 1$, $Y_1 \sim \text{Beta}(\alpha_1, \alpha_2)$.

Se $\alpha_1 = 0$ então $Y_1 = 0$ q.c.

Se $\alpha_2 = 0$ então $Y_1 = 1$ q.c.

Se $\alpha_1 = \alpha_2 = 1$ então $Y_1 \sim U(0, 1)$

Momentos e moda da distribuição de Dirichlet

Seja $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$.

$$(a) \quad m_{r_1, \dots, r_k} = \mathbb{E} \left[\prod_{i=1}^k Y_i^{r_i} \right] = \frac{\Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} \frac{\prod_{i=1}^k \Gamma(\alpha_i + r_i)}{\Gamma(\sum_{i=1}^{k+1} (\alpha_i + r_i))} \text{ onde } r_{k+1} = 0$$

$$(b) \quad \text{De (a): } \mathbb{E}Y_i = m_{0, \dots, 1, \dots, 0} = \frac{\alpha_i}{\sum_{j=1}^{k+1} \alpha_j}, i = 1, 2, \dots, k \text{ e } \mathbb{E}Y_i^2 = \frac{\alpha_i(\alpha_i + 1)}{\alpha^2 + \alpha} \quad i = 1, \dots, k, \\ \text{onde } \alpha = \sum_{j=1}^{k+1} \alpha_j.$$

Notamos que esses são os momentos de primeira e segunda ordem da distribuição Beta, i.é., considerando o fato que $Y_1 \sim \text{Beta}(\alpha_1, \alpha_2)$ então $\mathbb{E}Y_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}$ e $\mathbb{E}Y_1^2 = \frac{\alpha_1(\alpha_1 + 1)}{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_2 + 1)}$

(c) A moda da distribuição de Dirichlet, i.é., o vetor y^* tal que

$$f(y_1^*, \dots, y_k^*/\alpha_1, \dots, \alpha_k; \alpha_{k+1}) = \sup_{y \in \mathcal{S}_k} f(y_1, \dots, y_k/\alpha_1, \dots, \alpha_k; \alpha_{k+1}),$$

$$\text{e dado por } y_j^* = \frac{\alpha_j - 1}{\sum_{i=1}^{k+1} \alpha_i - (k+1)}; \quad j = 1, 2, \dots, k.$$

No caso da distribuição Beta(α_1, α_2), temos que $y^* = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$.

Observação

Se definirmos $Y_{k+1} = 1 - \sum_{i=1}^k Y_i$, então a distribuição do vetor $(Y_1, \dots, Y_{k+1})$ é singular em relação à medida de Lebesgue λ_{k+1} sobre $\mathbb{R}^{k+1}$. O suporte desta distribuição singular é o conjunto k -dimensional

$$E_{k+1} = \{(y_1, \dots, y_{k+1}) : y_i \geq 0 \quad i = 1, 2, \dots, k+1, \sum_{i=1}^{k+1} y_i = 1\}$$

A função densidade conjunta, neste caso, pode ser representada como $C \prod_{i=1}^{k+1} y_i^{\alpha_i - 1}$ em relação à medida de Lebesgue k -dimensional sobre E_{k+1} . $\square$

A proposição seguinte mostra à família de distribuições de Dirichlet como a natural família conjugada para os parâmetros de uma distribuição multinomial.

Proposição 1

Consideramos o modelo estatístico de uma variável aleatória X que assume $(k+1)$ -valores com vetor de probabilidades $(p_1, \dots, p_{k+1})$. Seja $\mathcal{D}_S(\alpha_1, \dots, \alpha_{k+1})$ o modelo de probabilidade a priori para os parâmetros $(p_1, \dots, p_{k+1})$. Então, com as observações independentes de X , dando origem às frequências amostrais $n_1, \dots, n_{k+1}$ para os $(k+1)$ -valores de X , temos que a distribuição a posteriori de $(p_1, \dots, p_{k+1})$ é $\mathcal{D}_S(\alpha_1 + n_1, \dots, \alpha_{k+1} + n_{k+1})$. $\square$

Proposição 2

Seja $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Se $i_1, \dots, i_{k+1}$ é uma permutação de $1, 2, \dots, k+1$ então $(Y_{i_1}, \dots, Y_{i_k}) \sim \mathcal{D}(\alpha_{i_1}, \dots, \alpha_{i_k}; \alpha_{i_{k+1}})$. $\square$

A caracterização da distribuição de Dirichlet em termos de variáveis aleatórias Beta mutuamente independentes é dada pelas proposições seguintes.

Proposição 3

Seja $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Definamos $U_1 = Y_1, U_i = \frac{Y_i}{1 - Y_1 - \dots - Y_{i-1}}$ para $i = 2, \dots, k$. Então $U_i \sim \text{Beta}(\alpha_i, \alpha_{i+1} + \dots + \alpha_{k+1}), i = 1, 2, \dots, k$, e as variáveis aleatórias $U_1, \dots, U_k$ são mutuamente independentes. $\square$

Vale também a recíproca da última proposição.

Proposição 4

Sejam $U_i \sim \text{Beta}(\alpha_i, \alpha_{i+1} + \dots + \alpha_{k+1}), i = 1, 2, \dots, k$ independentes e definamos

$$Y_1 = U_1, \quad Y_2 = (1 - U_1)U_2, \dots, \quad Y_k = (1 - U_1)(1 - U_2) \cdots (1 - U_{k-1})U_k,$$

então $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. $\square$

Combinando as proposições 3 e 4 temos:

Proposição 5

Sejam $U \sim \text{Beta}(\alpha_1, \alpha_2 + \dots + \alpha_{k+1})$ e $\mathbf{Z} = (Z_2, \dots, Z_k) \sim \mathcal{D}(\alpha_2, \dots, \alpha_k; \alpha_{k+1})$ mutuamente independentes, e definamos:

$$Y_1 = U, \quad Y_2 = (1 - U)Z_2, \dots, \quad Y_k = (1 - U)Z_k$$

então $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. $\square$

Corolário 1

Seja $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$

- (a) Da proposição 3, temos que $Y_1 \sim \text{Beta}(\alpha_1, \alpha - \alpha_1)$, onde $\alpha = \sum_{i=1}^{k+1} \alpha_i$.
- (b) Isto último junto com a proposição 2, nos dá que $(Y_1, \dots, Y_r) \sim \mathcal{D}(\alpha_1, \dots, \alpha_r; \alpha - \sum_{i=1}^r \alpha_i)$, onde $r < k$.
- (c) Uma extensão mais geral é dada pela proposição 2 e a proposição 4.

Para qualquer subconjunto $\{i_1, \dots, i_r\}$ de $\{1, 2, \dots, k+1\}$ tem-se que $(Y_{i_1}, \dots, Y_{i_r}) \sim \mathcal{D}(\alpha_{i_1}, \dots, \alpha_{i_r}; \alpha - \sum_{i=1}^r \alpha_{i_j})$. $\square$

O resultado seguinte é consequência das proposições 3 e 4.

Proposição 6

Seja $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Então, para qualquer inteiro r , tal que $2 \leq r \leq k$ temos que o vetor

$$\mathbf{Z} = \left(\frac{Y_r}{1 - Y_1 - \dots - Y_{r-1}}, \dots, \frac{Y_k}{1 - Y_1 - \dots - Y_{r-1}} \right)$$

é independente do vetor $(Y_1, \dots, Y_{r-1})$.

Também $\mathbf{Z} \sim \mathcal{D}_S(\alpha_r, \alpha_{r+1}, \dots, \alpha_k)$. $\square$

Proposição 7

Sejam $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$, e $r_1, \dots, r_{l+1}$ inteiros tais que $0 < r_1 < r_2 < \dots < r_{l+1} = k + 1$. Então

$$\left(\sum_1^{r_1} Y_i, \sum_{r_1+1}^{r_2} Y_i, \dots, \sum_{r_{l-1}+1}^{r_l} Y_i \right) \sim \mathcal{D} \left(\sum_1^{r_1} \alpha_i, \sum_{r_1+1}^{r_2} \alpha_i, \dots, \sum_{r_{l-1}+1}^{r_l} \alpha_i; \sum_{r_l+1}^k \alpha_i \right) \quad \square$$

A distribuição de Dirichlet pode também ser definida em termos de variáveis aleatórias Gamma mutuamente independentes.

Proposição 8

Sejam $Z_1, \dots, Z_{k+1}$ variáveis aleatórias mutuamente independentes com distribuição $\text{Gamma}(\alpha_i, \beta)$ $i = 1, 2, \dots, k + 1$, onde $\alpha_i, \beta > 0, i = 1, \dots, k + 1$.

Sejam $Z = \sum_1^{k+1} Z_i$ e $Y_i = Z_i/Z$ $i = 1, 2, \dots, k + 1$.

Então $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$ e $(Y_1, \dots, Y_k)$ é independente de Z . $\square$

Vale também a recíproca da última proposição.

Proposição 9

Se $Z \sim \text{Gamma}(\alpha, \beta)$, $\alpha = \sum_1^{k+1} \alpha_i > 0$ e Z é independente de $(Y_1, \dots, Y_k)$, onde $(Y_1, \dots, Y_k) \sim \mathcal{D}(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$, então as variáveis aleatórias $Z_i = ZY_i$, $i = 1, 2, \dots, k$ e $Z_{k+1} = Z(1 - \sum_1^k Y_i)$ são mutuamente independentes e $Z_i \sim \text{Gamma}(\alpha_i, \beta)$, $i = 1, \dots, k + 1$. $\square$

Da proposição 8, obtemos o seguinte.

Corolário 2

Sejam $Z_1 \sim \text{Gamma}(\alpha_1, \beta)$; $Z_2 \sim \text{Gamma}(\alpha_2, \beta)$ independentes. Então $Z_1/Z_1 + Z_2$ e $Z_1 + Z_2$ são independentes. Também $Z_1/Z_1 + Z_2 \sim \text{Beta}(\alpha_1, \alpha_2)$ e $Z_1 + Z_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$. $\square$

Para maiores detalhes sobre a distribuição de Dirichlet, em particular demonstrações destes resultados, ver Wilks (1962), Basu e Tiwari (1982) e Pereira e Basu(1982).

APÊNDICE B

Os processos empíricos na estatística não-paramétrica

Consideramos (ξ_i) sequência de variáveis aleatórias independentes e identicamente distribuídas, com função de distribuição F . Para cada amostra de tamanho n , $\{\xi_1, \dots, \xi_n\}$, a correspondente *função de distribuição empírica* F_n é construída colocando em cada observação ξ_i uma massa $1/n$. Assim F_n pode ser representada como

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n I(\xi_i \leq t), \quad t \in \mathbb{R}.$$

Para cada amostra fixada $\{\xi_1, \dots, \xi_n\}$, $F_n(\cdot)$ é uma função de distribuição considerada como uma função de t . Por outro lado, para cada valor de t fixado, $F_n(t)$ é uma variável aleatória, considerada como uma função da amostra. Considerando ambas características em conjunto, $F_n(\cdot)$ é uma *função de distribuição aleatória* e pode ser tratada então como um particular *processo estocástico*, i.é, um *elemento aleatório* de um apropriado espaço de funções.

1. $F_n(t)$ como um estimador pontual de $F(t)$

O aspecto mais simples de F_n é que, para cada t fixado, $F_n(t)$ serve como um estimador de $F(t)$. Por exemplo, notamos que $F_n(t)$ é *não-viesado*, i.é, $\mathbb{E}F_n(t) = F(t)$. Também $\text{Var } F_n(t) = \frac{F(t)(1-F(t))}{n}$ que converge a zero, quando $n \rightarrow \infty$.

Logo $F_n(t)$ é *consistente em média quadrática* para estimação de $F(t)$. Além disso, aplicando a lei forte dos grandes números, $F_n(t)$ é *fortemente consistente*, i.é, $F_n(t) \rightarrow F(t)$ com probabilidade 1. Tem-se que esta última convergência é *uniforme em t* [ver seção 2].

Notamos que a *distribuição exata* de $nF_n(t)$ é simplesmente Binomial $(n, F(t))$ e, imediatamente do Teorema do Limite Central, a *distribuição assintótica* é dada por:

Para cada t fixado, $-\infty < t < \infty$,

$$n^{\frac{1}{2}}(F_n(t) - F(t)) \xrightarrow{\mathcal{D}} N(0, F(t)(1 - F(t))) .$$

Esta convergência pode ser colocada também em termos da teoria de processos empíricos.

2. A distância de Kolmogorov-Smirnov

Além de considerar, para cada t , $F_n(t)$ um estimador pontual de $F(t)$, é também de interesse caracterizar *globalmente* a estimação de F por F_n . Para isto, uma medida de cercania de F_n a F muito útil é a chamada distância de Kolmogorov-Smirnov

$$\|F_n - F\|_{\infty} = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| .$$

Esta medida é utilizada na construção de bandas de confiança para $F(t)$, $t \in \mathbb{R}$, como também na construção dos testes de aderência baseados na função de distribuição empírica.

No que segue estabelecemos desigualdades de probabilidade, convergência quase certa e lei do logaritmo iterado, para a distância de Kolmogorov-Smirnov.

No caso unidimensional, uma desigualdade de probabilidade do tipo exponencial é a seguinte:

Teorema 1 (Dvoretzky, Kiefer e Wolfowitz).

Seja F uma função de distribuição definida sobre $\mathbb{R}$. Existe uma constante C (finita que não depende de F) tal que

$$\Pr (\|F_n - F\|_{\infty} > d) \leq C e^{-2nd^2} , \quad \text{onde } d > 0 ,$$

para todo $n = 1, 2, \dots$ □

É claro que a desigualdade anterior pode ser colocada na forma

$$\Pr(n^{\frac{1}{2}}\|F_n - F\|_{\infty} > d) \leq Ce^{-2d^2}, \quad \text{onde } d > 0,$$

para todo $n = 1, 2, \dots$. Obtemos assim que a sequência $(n^{\frac{1}{2}}\|F_n - F\|_{\infty})_n$ é limitada em probabilidade, o qual escrevemos

$$n^{\frac{1}{2}}\|F_n - F\|_{\infty} = O_P(1).$$

A convergência quase-certa de $\|F_n - F\|_{\infty}$ é estabelecida no seguinte teorema:

Teorema 2 (Glivenko-Cantelli).

$$\|F_n - F\|_{\infty} = \sup_t |F_n(t) - F(t)| \rightarrow 0 \quad \text{com probabilidade } 1. \quad \square$$

As flutuações extremas na convergência de $\|F_n - F\|_{\infty}$ a 0 são caracterizadas pela seguinte lei do logaritmo iterado.

Teorema 3.

Com probabilidade 1,

$$\lim_{n \uparrow \infty} \frac{n^{\frac{1}{2}}\|F_n - F\|_{\infty}}{(2 \log \log n)^{\frac{1}{2}}} = c(F)$$

onde $c(F) = \sup_t \{F(t)(1 - F(t))\}^{\frac{1}{2}}$.

Se F é contínua então $c(F) = 1/2$ e obtemos

$$\begin{aligned} n^{\frac{1}{2}}\|F_n - F\|_{\infty} &= O((\log \log n)^{\frac{1}{2}}) \\ \text{ou} \quad \|F_n - F\|_{\infty} &= O\left(\left(\frac{\log \log n}{n}\right)^{\frac{1}{2}}\right) \end{aligned}$$

□

3. Convergência fraca de processos empíricos

É de nosso interesse o estudo de distribuições limites de funcionais definidos sobre sequências de processos estocásticos, em particular funcionais definidos sobre processos empíricos. Processos $\{X_n(t) : t \in T\}$, onde T é um subconjunto da reta real estendida $[-\infty, \infty]$, são identificados com elementos aleatórios de algum espaço de funções definidas sobre T , um espaço suficientemente grande de maneira que contenha as trajetórias de todo X_n .

Muitos dos bem conhecidos processos estocásticos possuem trajetórias que são *cadlag*, i.é, funções contínuas pela direita (*continue à droite*) e tendo limites pela esquerda (*limite à gauche*) em cada ponto de T . Quando $T = [0, 1]$, esses processos podem ser analisados como elementos aleatórios de $D[0, 1]$, o conjunto de todas funções *cadlag* definidas sobre $[0, 1]$ a valores em $\mathbb{R}$. Notamos que continuidade pela direita em 1 e existência do limite pela esquerda em 0, carecem de sentido neste caso. A análise para outros espaços de funções *cadlag* sobre compactos, por exemplo $D[-\infty, \infty]$, o natural espaço para o estudo das funções de distribuição empírica sobre a reta real, e extensões a conjuntos de índices não-compactos, são em geral variações sobre métodos desenvolvidos para $D[0, 1]$.

Na sequência descrevemos o Teorema do Limite Central empírico, tanto no caso uniforme, quanto no caso geral. Esse resultado é conhecido como Teorema de Donsker.

Como toda função *cadlag* sobre $[0, 1]$ é limitada, o espaço $D[0, 1]$ é munido da métrica uniforme:

$$\|x - y\|_{\infty} = \sup\{|x(t) - y(t)| : 0 \leq t \leq 1\} .$$

Teorema 4.

A partir de uma sequência de variáveis aleatórias (η_i) independentes e identicamente distribuídas com distribuição comum $U(0, 1)$, construímos o processo empírico

$$U(t) = n^{-\frac{1}{2}} \sum_{i=1}^n [I(\eta_i \leq t) - t] = n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n I(\eta_i \leq t) - t \right), \quad 0 \leq t \leq 1.$$

Então, o processo empírico (U_n) converge em distribuição, como elementos aleatórios de $D[0, 1]$, à ponte browniana U , i.é, a um processo estocástico U definido sobre $[0, 1]$ com trajetórias contínuas e

(i) $U(0) = U(1) = 0$

(ii) as distribuições finito-dimensionais de U são normais multivariada, com média 0 e núcleo de covariância dado por $\mathbb{E}U(s)U(t) = s(1-t)$, para $s \leq t$. $\square$

Um resultado análogo ao Teorema 4 é possível obter para o processo empírico E_n , construído a partir de uma amostra proveniente de uma distribuição F arbitrária.

Denotamos $D[-\infty, \infty]$ o espaço das funções cadlag sobre $[-\infty, \infty]$. Continuidade pela direita de E_n em $-\infty$ é obtida, colocando $E_n(-\infty) = 0$; também colocamos o limite pela esquerda de E_n em $+\infty$ igual a zero, o valor natural para $E_n(+\infty)$.

Teorema 5.

Consideramos (ξ_i) sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição comum F (arbitrária), e o respectivo processo empírico

$$\begin{aligned} E_n(t) &= n^{-\frac{1}{2}} \sum_{i=1}^n [I(\xi_i \leq t) - F(t)] \\ &= n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n I(\xi_i \leq t) - F(t) \right) = n^{\frac{1}{2}} (F_n(t) - F(t)). \end{aligned}$$

Então, o processo empírico (E_n) converge em distribuição, como elementos aleatórios de $D[-\infty, \infty]$, ao processo gaussiano E , com média 0 e núcleo de covariância dado por $\mathbb{E}E(s)E(t) = F(s \wedge t) - F(s)F(t)$ $s, t \in \mathbb{R}$, ou $\mathbb{E}E(s)E(t) = F(s)(1 - F(t))$, para $s \leq t$.

□

4. Convergência fraca de funcionais definidos sobre sequências de processos empíricos

Nesta seção descrevemos os argumentos, baseados no cálculo diferencial, que são utilizados para estudar propriedades assintóticas de funcionais definidos sobre um certo processo empírico.

Para uma maior simplicidade na exposição, consideramos a estimação de $\varphi(F)$ através de $\varphi(F_n)$, onde F é uma função de distribuição sobre $(0, \infty)$ e F_n é a função de distribuição empírica baseada em uma amostra $\xi_1, \dots, \xi_n$ de F . Os resultados podem ser facilmente estendidos a casos mais gerais, utilizando argumentos similares.

Denotamos por

$\mathcal{F}$: espaço das funções de distribuição sobre $(0, \infty)$,

$\mathcal{F}_D$: espaço vetorial gerado pelas diferenças $F_2 - F_1$, onde $F_1, F_2 \in \mathcal{F}$, munido com a norma $\|h\|_\infty = \sup_{0 \leq t < \infty} |h(t)|$,

B_1 : espaço vetorial normado e

φ : um funcional de $\mathcal{F}$ em B_1 .

Definição 1

φ é dito ser diferenciável no ponto $F_1 \in \mathcal{F}$, se existe uma transformação linear $\varphi'(F_1, \cdot)$ de $\mathcal{F}_D$ em B_1 , tal que

$$\varphi(F_2) - \varphi(F_1) = \varphi'(F_1; F_2 - F_1) + o(\|F_2 - F_1\|_\infty),$$

quando $\|F_2 - F_1\|_\infty \rightarrow 0$.

Logo, sob a hipótese de diferenciabilidade de φ e usando o Teorema 1 deste Apêndice,

$$\begin{aligned} n^{1/2}[\varphi(F_n) - \varphi(F)] &= \varphi'(F; n^{1/2}(F_n - F)) + o(n^{1/2}\|F_n - F\|_\infty) \\ &= \varphi'(F; n^{1/2}(F_n - F)) + o_P(1) . \end{aligned}$$

Agora, pelo Teorema de Donsker e uma hipótese de continuidade sobre $\varphi'(F, \cdot)$ temos que

$$n^{1/2}[\varphi(F_n) - \varphi(F)] \xrightarrow{\omega} \varphi'(F; W) ,$$

onde W é um processo gaussiano com média 0 e núcleo de covariância $\text{cov}(W(s), W(t)) = F(s)(1 - F(t))$, $s \leq t$, e $\xrightarrow{\omega}$ denota convergência fraca.

Supondo F contínua e dado que $\varphi'(F, \cdot)$ é linear e contínua, implica que $\varphi'(F, \cdot)$ é limitada, segue que

$$\begin{aligned} \|\varphi(F_n) - \varphi(F)\|_{B_1} &\leq \|\varphi'(F; F_n - F)\|_{B_1} + o(\|F_n - F\|_\infty) \\ &\leq \|\varphi'\| \|F_n - F\|_\infty + o(\|F_n - F\|_\infty) \\ &= O(\|F_n - F\|_\infty) = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ q.c.,} \end{aligned}$$

quando $n \rightarrow \infty$, onde $\|\varphi'\|$ é a norma da transformação linear contínua $\varphi'(F, \cdot)$ e na última igualdade usamos o Teorema 3 deste Apêndice. Dado que $\varphi'(F, \cdot)$ é contínua se e somente se φ é contínua e diferenciável em F , uma condição conveniente de ser satisfeita por φ é que φ seja continuamente diferenciável em F .

Em algumas situações a função, digamos $G = \varphi(F)$ não pode ser expressa explicitamente em termos da distribuição F , só pode ser expressa implicitamente através de uma equação $H(F, G) = 0$. Em outras situações, embora possamos ter uma equação explícita, pode ser mais conveniente para o estudo assintótico contar com uma equação do tipo implícito. O teorema da função implícita dá condições suficientes para a existência, unicidade e diferenciabilidade de soluções locais $G = T(F)$.

Definição 2

Sejam B_i , $i = 1, 2, 3$, espaços vetoriais normados e seja $H : B_1 \times B_2 \rightarrow B_3$. A função H é dita ser parcialmente diferenciável em relação à primeira variável no ponto (F_1, F_2) , se e somente se $g(F) = H(F, F_2)$ é diferenciável em F_1 . Neste caso, escrevemos $g'(F_1; \cdot) = H'_1(F_1, F_2; \cdot)$. Chamamos H'_1 o diferencial parcial em relação à primeira variável. A definição de H'_2 , o diferencial parcial em relação à segunda variável, é análoga.

Teorema 6 (Teorema da função implícita).

Seja $H : \mathcal{O} \subset B_1 \times B_2 \rightarrow B_3$ continuamente diferenciável, onde $\mathcal{O}$ é um aberto de $B_1 \times B_2$, B_2 e B_3 são espaços de Banach.

Suponhamos que para algum ponto $(F_1, F_2) \in \mathcal{O}$, $H(F_1, F_2) = 0$, e o diferencial parcial $H'_2(F_1, F_2, \cdot)$ possui inversa.

Então existe $b > 0$ e uma função φ continuamente diferenciável da vizinhança aberta $N_b = \{G \in B_1 : \|G - F_1\|_{B_1} < b\}$ em B_2 tal que:

- (a) $(G, \varphi(G)) \in \mathcal{O}$ para toda $G \in N_b$;
- (b) $F_2 = \varphi(F_1)$ e $H(G, \varphi(G)) = 0$ para toda $G \in N_b$ e
- (c) φ é continuamente diferenciável com

$$\varphi'(F_1; \cdot) = -[H'_2(F_1, \varphi(F_1); H'_1(F_1, \varphi(F_1); \cdot))]^{-1} . \quad \square$$

Se $H(F, G) = 0$, $H(F_n, \hat{G}_n) = 0$, onde $\hat{G}_n$ é um estimador consistente de G , então pelo teorema da função implícita existe uma transformação continuamente diferenciável φ tal que $G = \varphi(F)$ e $\hat{G}_n = \varphi(F_n)$ para n suficientemente grande. Logo, usando os argumentos do início desta seção, é possível estabelecer a convergência fraca e uma lei do logaritmo iterado para o estimador $\hat{G}_n$. Esses fatos são estabelecidos no seguinte teorema.

Teorema 7

Seja $\hat{G}_n$ uma solução de $H(F_n, \hat{G}_n) = 0$. Se H satisfaz as condições do teorema da função implícita, então:

$$(a) \quad n^{1/2}(\hat{G}_n - G) \xrightarrow{\omega} \varphi^*(F, G; W) ,$$

onde W é um processo gaussiano com média 0 e núcleo de covariância $\text{cov}(W(s), W(t)) = F(s)(1 - F(t))$, $s \leq t$, e

$$\varphi^*(F, G; W) = -[H'_2(F, G; H'_1(F, G; W))]^{-1} . \quad \square$$

Para maiores detalhes dos tópicos tratados neste apêndice, algumas referências são Billingsley (1968), Serfling (1980), Pollard (1984) e Tsai e Crowley (1985).

REFERÊNCIAS

- ANTONIAK, C.E.(1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152-1174.
- BASU, A.P.(1984). Censored data. In *Handbook of Statistics 4* (P.R. Krishnaiah and P.K. Sen, Eds.), 551-578.
- BASU, D. and TIWARI, R.C.(1982). A Note on Dirichlet process. In *Statistics and Probability: Essays in Honor of C.R. Rao* (G. Kallinpur, P.R. Krishnaiah and J.K. Ghosh, Eds.) 83-103.
- BERK, R.H. and SAVAGE, I.R.(1975) Dirichlet processes produce discrete measures: An elementary proof. In *Contributions to Statistics. Jaroslav Hájek Memorial Volume*, 25-31. Academia. Prague: North-Holland.
- BERMAN, S.M.(1963). Note on extreme values, competing risks and semi-Markov processes. *Ann. Math. Statist.* **34**, 1104-1106.
- BILLINGSLEY, P.(1968). *Convergence of Probability Measures*. New York: Wiley.
- BILLINGSLEY, P.(1986). *Probability and Measure*. Second edition. New York: Wiley.
- BLACKWELL, D.(1973). Discreteness of Ferguson selections. *Ann. Statist.* **1**, 356-358.
- BLACKWELL, D. and MAC QUEEN, J.B.(1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353-355.

BLUM, J. and SUSARLA, V.(1977). On the posterior distribution of a Dirichlet process given randomly right censored observations. *Stoch. Proc. Appl.* **5**, 207-211.

BOLFARINE, H.; RODRIGUES, J. and ACHCAR, J.(1991). *Análise de Sobre-vivência*. 2a. Escola de Modelos de Regressão. IM-UFRJ. Rio de Janeiro, Brasil.

BRESLOW, N. and CROWLEY, J.(1974). A large sample study of the life table and product limit estimators under random censorship. *Ann. Statist.* **2**, 437-453.

DALAL, S.R.(1978). A note on the adequacy of mixtures of Dirichlet processes. *Sankhya* **40**, Series A, 185-191.

(DALAL, S.R.(1979). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stoch. Proc. Appl.* **9**, 99-107.

DALAL, S.R.(1980). Nonparametric Bayes Decision Theory. In *Bayesian Statistics* . (J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F. Smith , Eds.). 521-534, University Press, Valencia, Spain.

DALAL, S.R. and PHADIA, E.G.(1983). Nonparametric Bayes inference for concordance in bivariate distributions. *Comm. Statist.-Theor. Meth.* **12**(8), 947-963.

DAVIS, P.J. and RABINOWITZ, P.(1984). *Methods of Numerical Integration*. Second edition. New York: Academic Press.

DOKSUM, K.(1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Prob.* **2**, 183-201.

DOKSUM, K. and LO, A.Y.(1990). Consistent and robust Bayes procedures for location based on partial information. *Ann. Statist.* **18**, 443-453.

FERGUSON, T.S.(1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

FERGUSON, T.S.(1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615-629.

FERGUSON, T.S.(1982). Sequential estimation with Dirichlet process priors. In *Statistical Decision Theory and related topics III*, vol.1 (S.S. Gupta and J.O. Berger, Eds.). 385-401.

FERGUSON, T.S.(1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances of Statistics: Papers in Honor of H. Chernoff* (M.H. Rizvi, J. Rustagi and D. Siegmund, Eds.). 287-302. New York: Academic Press.

FERGUSON, T.S.; PHADIA, E.G. and TIWARI, R.C.(1992). Bayesian nonparametric inference. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu* (M. Ghosh and P.K. Pathak, Eds.) IMS Lec. Notes Monograph Ser. 17.

GAUDARD, M. and HADWIN, D.(1989). Sigma-algebras on spaces of probability measures. *Scand. J. Statist.* **16**, 169-175.

GHOSH, M.; LAHIRI, P. and TIWARI, R.C.(1989). Nonparametric empirical Bayes estimation of the distribution function and mean. *Commun. Statist.-Theor. Meth.*, **18**(1), 121-146.

HANNUM, R.C. and HOLLANDER, M.(1983). Robustness of Ferguson's Bayes estimator of a distribution function. *Ann. Statist.* **11**, 632-639.

HANNUM, R.C.; HOLLANDER, M. and LANGBERG, N.A.(1981). Distributional results of random functional of a Dirichlet process. *Ann. Prob.* **9**, 665-670.

HJORT, N.L.(1990). Bayes estimators based on Beta processes in models for life history data. *Ann. Statist.* **18**, 1259-1294.

JAMMALAMADAKA, S.R. and TIWARI, R.C.(1985). Estimation of survival function and failure rate. *Statistic* **16**, 535-540.

KAPLAN, E.L. and MEIER, P.(1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

KORWAR, R.M. and HOLLANDER, M.(1973). Contributions of the theory of Dirichlet processes. *Ann. Prob.* **1**, 705-711.

KORWAR, R.M. and HOLLANDER, M.(1976). Empirical Bayes estimation of a distribution function. *Ann. Statist.* **4**, 581-588.

KUO, L.(1986). Computations of mixtures of Dirichlet processes. *SIAM J. Sci. Statist. Comput.* **7**, 60-71.

LAHIRI, P. and TIWARI, R.C.(1990). Empirical Bayes rank order estimation with Dirichlet prior. *Statist. Decis.* **8**, 231-245.

LO, A.Y.(1983). Weak convergence for Dirichlet processes. *Sankhya* **45**. Series A. 105-111.

MARTZ, H.F. and WALLER, R.A.(1982). *Bayesian Reliability Analysis*. New York: Wiley.

MILLER, R.G.(1981). *Survival Analysis*. New York: Wiley.

MORALES, D.; PARDO, L. and QUESADA, V.(1990). Estimation of a survival function with doubly censored data and Dirichlet process prior knowledge on the observable variable. *Commun. Statist.-Simula. Computa.* **19**(1), 349-361.

PAPADOUPOLOS, A.S. and TIWARI, R.C.(1989). Bayesian Bootstrap lower confidence interval. Preprint.

PARTHASARATHY, K.R.(1967). *Probability Measures on Metric Spaces*. New York: Academic Press.

PEREIRA, C.A.B. and BASU, D.(1982) On the Bayesian analysis of categorical data: The problem of nonresponse. *J. Statist. Plann. Inference* **6**, 345-362.

PETERSON, A.V.(1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72**, 854-858.

PHADIA, E.G.(1980). A note on empirical Bayes estimation of a distribution function based on censored data. *Ann. Statist.* **8**, 226-229.

POLLARD, D.(1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

RAI, K.; SUSARLA, V. and VAN RYZIN, J.(1980). Shrinkage estimation in non-parametric survival analysis: A simulation study. *Commun. Statist. -Simula. Computa.* **9**(3), 271-298.

RUDIN, W.(1964). *Principles of Mathematical Analysis*. Second edition. Tokyo: Mac Graw-Hill.

SERFLING, R.J.(1980). *Approximations Theorems of Mathematical Statistics*. New York: Wiley.

SETHURAMAN, J. and TIWARI, R.C.(1982). Convergence of Dirichlet measures and interpretation of their parameter. In *Statistical Decision Theory and related topics III*, vol.2 (S.S. Gupta and J.O. Berger, Eds.), 305-312.

SUSARLA, V. and VAN RYZIN, J.(1976). Nonparametric Bayesian estimation of survival curves from imcomplete observations. *J. Amer. Statist. Assoc.* **71**, 897-902.

SUSARLA, V. and VAN RYZIN, J.(1978a). Empirical Bayesian estimation of a distribution (survival) function from right censored observations. *Ann. Statist.* **6**, 740-754.

SUSARLA V. and VAN RYZIN, J.(1978b). Large sample theory for a Bayesian non-parametric survival curve estimator based on censored samples. *Ann. Statist.* **6**, 755-768. Addendum(1980), *Ann. Statist.* **8**, 693.

TIWARI, R.C.(1981). A Mathematical Study of Dirichlet Process. Ph.D. Dissertation. Dept. of Statistics, Florida State University.

TIWARI, R.C. and JAMMALAMADAKA, S.R.(1985). On the Bayes estimate of a functional of two distributions. *Calcutta Statist. Ass. Bull.* **34**, 15-22.

TIWARI, R.C. and ZALKIKAR, J.N.(1985). Empirical Bayes estimation of functionals of unknown probability measures. *Comm. Statist.-Theor. Meth.* **14**(12), 2963-2995.

TIWARI, R.C. and ZALKIKAR, J.N.(1990). Nonparametric Bayes estimation of the survival function from record of failures and follow-ups. *Commun. Statist. -Theor. Meth.* **19**(11), 4331-4354.

TIWARI, R.C. and ZALKIKAR, J.N.(1991). Nonparametric Bayesian estimation under random left truncation. Preprint.

TIWARI, R.C. and TRIPHATI, R.C.(1990). Nonparametric Bayes estimation of the probability of discovering a new specie. *Comm. Statist.-Theor. Meth.* **19**(3), 877-890.

TSAI, W.Y.(1986). Estimation of survival curves from dependent censorship models via generalized self-consistent property with nonparametric Bayesian estimation application. *Ann. Statist.* **14**, 238-249.

TSAI, W.Y. and CROWLEY, J.(1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self consistency. *Ann. Statist.* **13**, 1317-1334.

VON MISES, R.(1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Statist.* **18**, 309-348.

WILKS, S.S.(1962). *Mathematical Statistics*. New York: Wiley.

ZALKIKAR, J.N.; TIWARI, R.C. and JAMMALAMADAKA, S.R.(1986). Bayes and empirical Bayes estimation of the probability that $Z > X + Y$. *Comm. Statist.-Theor. Meth.* **15**(10), 3079-3101.

ZEHNWIRTH, B.(1985). Nonparametric linear Bayes estimation of survival curves from incomplete observations. *Comm. Statist.-Theor. Meth.* **14**(8), 1769-1778.