

ESTIMAÇÃO DE FUNÇÕES

DE DENSIDADE

HELENO BOLFARINE

DISSERTAÇÃO APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA

UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF. DR. PEDRO ALBERTO MORETTIN

- SÃO PAULO, ABRIL DE 1976 -

Quero dedicar este trabalho à minha esposa pelo incentivo que sempre me deu e a meus pais pela formação que me propiciaram.

AGRADECIMENTOS

Ao terminarmos este trabalho, não podemos esquecer toda a ajuda que muitos, direta ou indiretamente, nos prestaram. Em especial, gostaríamos de mencionar

- o Prof. Dr. Pedro Alberto Morettin que sugeriu o tema desta dissertação e nos orientou de forma efetiva durante todo o trabalho;
- o apoio e estímulo dos membros do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo, e o cuidado dispensado por nosso colega João Baptista Esteves de Oliveira, na datilografia dos manuscritos.

INDICE

| | |
|---|----|
| CAP. 0 - INTRODUÇÃO. | 1 |
| CAP. 1 - COLOCAÇÃO DO PROBLEMA E ALGUMAS CONSIDERAÇÕES GERAIS. . . | 3 |
| CAP. 2 - ESTIMAÇÃO DE UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE POR FUNÇÕES PESO. | 14 |
| 1. A Classe dos Estimadores | 14 |
| 2. Não viciosa assintótica e consistência de $\hat{f}_n(x)$. . . | 18 |
| 3. Normalidade assintótica de $\hat{f}_n(x)$ | 21 |
| 4. Algumas extensões para o caso Multivariado | 22 |
| 5. Aplicações à Teoria da Confiabilidade e a Estimação da função de risco. | 27 |
| 6. Algumas Comparações com a Estimação da Função de Densi- dade expectral | 33 |
| CAP. 3 - ESTIMAÇÃO DE UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE POR SISTEMAS ORTOGONAIS | 39 |
| 1. A Classe dos Estimadores | 39 |
| 2. Não Viciosa, Normalidade assintóticas e Consisten- cia de $\hat{f}_n(x)$ | 42 |
| 3. O uso de sistemas ortogonais trigonométricos | 48 |
| CAP. 4 - OUTROS MÉTODOS DE ESTIMAÇÃO | 59 |
| APÊNDICE 1 | 67 |
| APÊNDICE 2 | 74 |
| REFERÊNCIAS. | 81 |

CAPITULO 0

INTRODUÇÃO

O problema de estimação de uma função de densidade de probabilidade tem recebido pouca atenção na literatura, visto que o primeiro trabalho usando técnicas modernas sobre o assunto surgiu com Roseblatt, em 1956. No seu trabalho, primeiramente ele discute algumas propriedades gerais da estimação de uma função de densidade e, num segundo momento, sugere um estimador e estuda suas propriedades principais. Um segundo trabalho importante da teoria, usando técnicas modernas, surgiu com Parzem, em 1962. Tanto Parzem como Roseblatt utilizaram estimadores baseados em funções peso. Além do que foi feito por Roseblatt, Parzem procurou estabelecer condições sobre o estimador por ele considerado, para que fosse assintoticamente normal.

Outros estimadores interessantes surgiram. Sugerido pelas noções de aproximação (sob certas condições) de uma função por uma série (de funções ou de polinômios) ortogonal e usando as idéias básicas dos artigos precedentes, surgiu um novo estimador de uma função de densidade, cujos aspectos e propriedades foram desenvolvidos principalmente por Schwartz (1967), Watson (1969) e Roseblatt (1971). Outros aspectos da teoria da estimação de uma função de densidade foram abordados por Watson & Leadbetter (1963) e, por James Pickands III (1969), utilizando técnicas baseadas no decrescimento algébrico e exponencial de funções caracterís

ticas. Um outro método, ainda seria a estimação de uma função de densidade por estimadores de máxima verossimilhança, que foram introduzidos, principalmente por Robertson (1967), Wegan (1969) e Wegan (1970).

O objetivo do nosso trabalho é discutir detalhadamente duas das classes de estimadores citadas acima (por nós julgadas as mais importantes), procurando apresentar os resultados em forma didática e dar a eles uma maior formalização. Não examinaremos neste trabalho, o problema da escolha do estimador a usar. Vamos dividir o trabalho em quatro capítulos, que organizamos da seguinte maneira: na primeira seção há a proposição teórica do problema, juntamente com algumas propriedades gerais dos estimadores; no capítulo dois, vamos discutir a classe de estimadores, baseado no que chamaremos de "funções peso"; no capítulo três vamos discutir outra classe de estimadores, baseada em funções ortogonais e, finalmente no quarto capítulo faremos uma discussão breve e bastante sucinta a respeito dos outros métodos citados acima. Vamos ainda apresentar um apêndice com as definições, teoremas e resultados utilizados para o desenvolvimento deste trabalho, bem como a bibliografia adequada para o estudo da estimação de uma função de densidade.

CAPÍTULO I

COLOCAÇÃO DO PROBLEMA E

CONSIDERAÇÕES GERAIS

Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da distribuição de uma variável aleatória real X , com função de densidade f , que é desconhecida. Nosso objetivo é estimar f . Podemos ter então duas situações. Por um lado, se o pesquisador conhece suficientemente a forma da distribuição da população a qual está estudando bastaria que ele estimasse os parâmetros dessa distribuição por procedimentos paramétricos e ele teria a densidade estimada. Por outro lado, se a forma da distribuição é bastante complicada, o experimentador para obter melhores resultados, deverá usar os estimadores ótimos (classe de estimadores que discutiremos); mas que são bem mais complicados do que os métodos de aproximação por histogramas, para a função de densidade, e pela função de distribuição amostral, para a função de distribuição. Esses métodos são mais primitivos porém mais simples, e é claro trazem resultados menos precisos.

Quem primeiro se preocupou com o problema de estimação de uma função de densidade, foi K. Pearson [9] e [10]. Seu objetivo era obter uma expressão matemática que representasse satisfatoriamente as observações. O sistema de distribuições de Pearson é representado pelo sistema de equações diferenciais

$$(1.1) \quad \frac{df}{dx} = \frac{(x - a)f}{b_0 + b_1x + b_2x^2}$$

que é a maneira de se expressar o limite da série hipergeométrica. As constantes a, b_0, b_1, b_2 são de tal forma que todos os momentos dessas distribuições podem ser encontrados a partir delas e, reciprocamente, podemos expressar as quatro constantes a_0, b_0, b_1, b_2 , em termos dos momentos em torno da média μ_2, μ_3, μ_4 , definidos por

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu_1^r)^r dF, \quad r=0,1,2,\dots$$

onde $\mu_r^r, r=1,2,\dots$ é o r -ésimo momentos amostral. Ver Kendall e Stuart [3]. O sistema de equações diferenciais (1.1) envolve uma grande família de distribuições. Podemos citar entre outras, a distribuição Beta, denominada do tipo I, a distribuição Gama, denominada do tipo III e a distribuição do tipo IV, que é usualmente escrita como

$$(1.2) \quad f(x) = k \left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\psi \arctg(x/a)}$$

Esta distribuição é ilimitada em ambas as direções e unimodal. Não é muito fácil manusear a expressão (1.2) na prática. A constante k aparece na maioria das distribuições de Pearson e, para seu cálculo, é necessário tabelas especiais de integração. O que é notável contudo, é que o sistema Pearsoniano contém quase todas as distribuições conhecidas em sua era e, atingiu grande popularidade. O problema de decidir qual tipo usar face aos dados de uma população com distribuição desconhecida também foi tratado por Pearson. Ele considera a constante

$$(1.3) \quad k = \frac{\beta_1 (\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}$$

onde

$$\beta_{2n+1} = \frac{\mu_3 \mu_{2n+3}}{\mu_2^{n+3}}, \quad \beta_{2n} = \frac{\mu_{2n+2}}{\mu_2^{n+1}},$$

como um critério para distinguir qual o tipo da distribuição adequada de modo que se $k > 1$, a distribuição a usar seria do tipo I; se k é infinito, seria a do tipo III e, se k está entre 0 e 1, usaríamos a distribuição IV. Todas as distribuições Pearsonianas são determinadas pelos momentos μ_1^i , μ_2^i , μ_3^i , μ_4^i , exceto algumas distribuições degeneradas que são determinadas por menos do que 4 momentos. Basicamente, o método de ajustamento de Pearson consiste em:

- 1) determinação dos valores numéricos dos quatro primeiros momentos da distribuição observada.
- 2) Cálculo dos valores numéricos de β_1 , β_2 e da constante k em (1.3) e então determinar a qual tipo a distribuição pertence.
- 3) igualar os momentos observados com os momentos da distribuição apropriada, expressada em termos de seus parâmetros.
- 4) resolver as equações resultantes para aqueles parâmetros, determinando-se então a distribuição.

Para ilustrar o processo, reproduziremos um exemplo de Kendall e Stuart [3].

EXEMPLO 1.1 - A tabela abaixo mostra a distribuição de frequências de 9.440 feijões, de acordo com o comprimento e largura.

| * ** | 17 | 16,5 | 16 | 15,5 | 15 | 14,5 | 14 | 13,5 | 13 | 12,5 | 12 | 11,5 | 11 | 10,5 | 10 | 9,5 | TOTAL |
|---------|----|------|-----|------|------|------|------|------|-----|------|-----|------|----|------|----|-----|-------|
| 9,125 | - | 2 | - | - | 3 | - | - | - | - | - | - | - | - | - | - | - | 5 |
| 8,875 | 4 | 8 | 17 | 19 | - | - | - | - | - | - | - | - | - | - | - | - | 48 |
| 8,625 | 2 | 23 | 101 | 156 | 93 | 23 | 2 | - | - | - | - | - | - | - | - | - | 400 |
| 8,375 | - | 18 | 105 | 494 | 574 | 227 | 56 | 9 | - | - | - | - | - | - | - | - | 1483 |
| 8,125 | - | 4 | 44 | 375 | 956 | 913 | 362 | 73 | 12 | 3 | - | - | - | - | - | - | 2742 |
| 7,875 | - | - | 7 | 61 | 385 | 871 | 794 | 330 | 89 | 19 | 3 | - | - | - | - | - | 2579 |
| 7,625 | - | - | 1 | 4 | 65 | 236 | 469 | 361 | 175 | 55 | 27 | 4 | - | - | - | - | 1397 |
| 7,375 | - | - | - | - | 6 | 23 | 91 | 137 | 124 | 76 | 37 | 22 | 11 | - | 1 | - | 530 |
| 7,125 | - | - | - | - | - | 1 | 13 | 18 | 28 | 35 | 25 | 32 | 11 | 6 | 1 | - | 170 |
| 6,875 | - | - | - | - | - | - | - | 1 | 9 | 8 | 21 | 12 | 13 | 7 | 1 | - | 72 |
| 6,625 | - | - | - | - | - | - | - | - | - | - | 2 | - | 1 | 4 | 3 | - | 18 |
| 6,375 | - | - | - | - | - | - | - | - | - | 1 | - | - | - | 1 | 1 | 1 | 4 |
| TOTAL | 6 | 55 | 275 | 1128 | 2082 | 2294 | 1787 | 929 | 437 | 199 | 115 | 70 | 36 | 18 | 7 | 1 | 9440 |

* Comprimento em mm (valores centrais)

** Largura em mm (valores centrais).

De acordo com (1) vamos encontrar em primeiro lugar os momentos, usando a correção de Sheppard (ver Kendall e Stuart [3]).

$$\begin{aligned}\mu_1' \text{ (origem até 14,5)} &= -0,190783898 \\ \mu_2 &= 3,238424951 \\ \mu_3 &= -5,306566352 \\ \mu_4 &= 50,999624044\end{aligned}$$

de onde vem que

$$\begin{aligned}\beta_1 &= 0,829135838 \quad \sqrt{\beta_1} = -0,910569 \\ \beta_2 &= 4,862944362\end{aligned}$$

Quanto ao tipo da distribuição temos

$$k = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} = \frac{51,262}{84,04}$$

Como k está entre 0 e 1, temos que a distribuição adequada para os dados é a do tipo IV. Escrevendo

$$\text{tang } \theta = x/a \quad \text{e} \quad 2m-2 = r$$

na equação (1.2) nós encontramos

$$\mu_n' = k \int_{-\pi/2}^{\pi/2} a^{n+1} \cos^{r-n}\theta \sin^n\theta e^{-v\theta} d\theta$$

Integrando por partes (com $\cos^{r-n}\theta \sin \theta$ como uma parte) temos

$$\mu_n' = \frac{a}{r-n+1} \{ (n-1) a \mu_{n-2}' - v \mu_{n-1}' \}$$

Então, os momentos em torno da média são dados por

$$\begin{aligned}\mu_1' &= -\frac{av}{r} \\ \mu_2' &= \frac{a^2}{r^2(r-1)}(r^2+v^2) \\ \mu_3' &= -\frac{4a^3v(r^2+v^2)}{r^3(r-1)(r-2)}\end{aligned}$$

$$\mu_4 = \frac{3a^4(r^2+v^2)\{(r+6)(r^2+v^2)-8v^2\}}{r^4(r-1)(r-2)(r-3)}$$

de onde segue-se que

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{2\beta_2 - 3\beta_1 - 6}$$

$$v = \frac{r(r-2)\sqrt{\beta_1}}{\sqrt{\{16(r-1) - \beta_1(r-2)^2\}}}$$

$$a = \sqrt{\left[\frac{\mu_2}{16} \{16(r-1) - \beta_1(r-2)^2\} \right]}$$

Substituindo β_2 , β_1 e μ_2 por seus valores numéricos, nós encontramos:

$$r = 14,69772 \quad m = 8,34886$$

$$v = 18,38043 \quad a = 4,15949$$

Calculando a constante k através de tabelas de integração, obtemos:

$$f(x) = 0,395121 \left(1 + \frac{x^2}{17,30134} \right)^{-8,3488} e^{-18,38043 \operatorname{arctg} \frac{x}{4,15949}} \quad \parallel$$

Esse método tem sido criticado quando os dados observados são encarados como amostra de uma população e dese^{ja}-se encontrar uma expressão matemática para representar a população. Em tais casos, os momentos calculados a partir das observações são somente estimadores dos momentos populacionais e eles em geral, não levam a estimadores eficientes dos parâmetros populacionais. Outros métodos de ajustamento de curvas serão apresentados no apêndice.

Nosso objetivo neste trabalho não é procurar expres

sões matemáticas para representar uma população, mas sim, procurar definir estimadores da função de densidade que tenham boas propriedades e que possam ser melhorados de acordo com a teoria da estimação. Como veremos, existem situações em que o próprio estimador (que é função das observações) é uma função de densidade. Veremos também (o que é um tanto desalentador) que não existe estimador não viciado de uma função de densidade. Grande parte da teoria de estimação não pode então ser utilizada neste caso; aquela que trata de estimadores não viciados.

Sejam então X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas, com função de densidade f contínua. Seja $\hat{f}(y, x_1, \dots, x_n)$ um estimador de $f(y)$ baseado nas n observações e satisfazendo as condições:

- 1) $\hat{f}(y, x_1, \dots, x_n) \geq 0, \forall y.$
- 2) $\hat{f}(y, x_1, \dots, x_n)$ é conjuntamente Borel mensurável em $(y, x_1, \dots, x_n).$

Com estas suposições sobre a densidade f e sobre o estimador $\hat{f}(y, x_1, \dots, x_n)$, Roseblatt [14] apresenta um dos resultados mais importantes da teoria, que está contido no seguinte

TEOREMA 1.1 - Não existe estimador não viciado de $f(y)$ baseado na n -pla ordenada $(X^{(1)}, \dots, X^{(n)})$.

DEMONSTRAÇÃO - Suponhamos por absurdo que

$$3) \quad E\hat{f}(y, X_1, \dots, X_n) = f(y), \forall f \text{ e } \forall y.$$

Então, a condição (3) acima implica para cada y que

$$E\hat{f}(y, X_1, X_2, \dots, X_n) < \infty.$$

Vamos assumir que $\hat{f}(y, x_1, \dots, x_n)$ seja uma função simétrica de x_1, \dots, x_n ; desde que a n -pla ordenada $(X^{(1)}, \dots, X^{(n)})$ é uma estatística suficiente para o problema. Como:

$$\begin{aligned} E \int_a^b \hat{f}(y, X_1, \dots, X_n) dy &= \int_a^b E \hat{f}(y, X_1, \dots, X_n) dy = \\ &= \int_a^b f(y) dy = F(b) - F(a) \end{aligned}$$

temos que

$$\int_a^b \hat{f}(y, X_1, \dots, X_n) dy$$

é um estimador simétrico e não viciado de $F(b) - F(a)$.

Por outro lado, como a n -pla ordenada $(X^{(1)}, \dots, X^{(n)})$ é também, uma estatística completa temos que $F_n(b) - F_n(a)$ é o único estimador de $F(b) - F(a)$ não viciado e simétrico baseado na n -pla ordenada $(X^{(1)}, X^{(2)}, \dots, X^{(n)})$ onde $F_n(y)$ é a função de distribuição amostral definida em (2.1.1). Isto decorre do fato de que $F_n(b) - F_n(a)$ é não viciado para $F(b) - F(a)$ (também mostrado no capítulo 2), simétrico nas observações e é baseado na estatística completa e suficiente, que é a n -pla ordenada $(X^{(1)}, \dots, X^{(n)})$. Temos portanto que

$$F_n(b) - F_n(a) = \int_a^b \hat{f}(y, x_1, x_2, \dots, x_n) dy$$

para todo a e b , $a < b$ e para quase toda (X_1, \dots, X_n) . Concluimos então, que $F_n(y)$ é absolutamente contínua em y para quase toda (X_1, \dots, X_n) , o que é absurdo. \square

OBSERVAÇÃO - Para a prova do teorema utilizamos o fato de que a n -pla ordenada $(X^{(1)}, \dots, X^{(n)})$ é uma estatística suficiente e completa, de onde se conclui que a função de distribuição amostral F_n é o estimador UMVU para F . A prova deste fato pode ser encontrado em [5].

Aqui estimador UMVU significa: estimador não viciado e de variância uniformemente mínima. É o estimador não viciado cuja variância coincide com o limite inferior da desigualdade de Cramér-Rao.

Um outro trabalho interessante, que aborda exclusivamente o problema de existência do estimador não viciado de f , foi publicado em 1970 por A.H. Seheult e C.P. Quesenberry [12]. Eles formularam o problema num contexto bastante geral, não fazendo nenhuma suposição sobre f e, obtiveram os resultados em termos de estimadores não viciados de medidas de probabilidade (ou funções de distribuição), os quais sempre existem. Para discutirmos os teoremas de existência, seja (X, A, μ) um espaço mensurável euclidiano onde μ é uma medida σ -finita e seja P uma família de medidas de probabilidade P em A , dominada por μ . Seja agora X uma variável aleatória com distribuição P e densidade p (p é a derivada de Radon-Nikodym de P com respeito a μ); $X^{(n)} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória de tamanho n da distribuição de X e $x^{(n)} = (x_1, \dots, x_n)$ uma observação de $X^{(n)}$. Vamos denotar por $X^{(n)}$ o espaço amostral das observações $x^{(n)}$; por $A^{(n)}$ a σ -álgebra produto dos subconjuntos de $X^{(n)}$ e por $Q^{(n)}$ a medida produto em $A^{(n)}$ correspondendo a qualquer medida Q em A . Seja T uma função mensurável (Estatística) de $(X^{(n)}, A^{(n)})$ no espaço euclidiano (T, B) e denotamos por Q_*^T uma versão da medida de probabilidade condicional (dado T) induzida em $A^{(n)}$ por Q em A . Então, a classe de todos os cilindros da forma $X^{(n-1)} \times A$ para $A \in A$ é uma sub- σ -álgebra de $A^{(n)}$ e nós definimos uma medida de probabilidade Q^T em A , por $Q^T = Q_*^T(X^{(n-1)} \times A)$, para todo $A \in A$. Exclusivamente nesta secção vamos denotar por \hat{P} o estimador de P e, por \hat{p} o estimador de p .

DEFINIÇÃO 1.1 - Um estimador $\hat{P} = \hat{P}(\cdot, T)$ é o estimador UMVU de P , se $\hat{P}(A, t)$ é o estimador UMVU de $P(A)$ para todo $A \in A$, onde T é uma estatística suficiente e completa para P .

O lema seguinte garante a existência do estimador não viciado de P , baseado na estatística completa e suficiente T , que é o estimador UMVU.

LEMA 1.1 - Sempre existe um estimador não viciado de P dado por $\hat{P} = P^T$, onde P^T é qualquer determinação da medida de probabilidade condicional dado T , que comentamos acima. Além disso, se T é uma estatística completa e suficiente para P , então P^T é o estimador UMVU de P .

A demonstração segue-se como consequência imediata das propriedades da esperança condicional e do fato de que se P^T é o estimador não viciado para P e baseado na estatística completa e suficiente, então pelo Teorema de Lehmann-Scheffé, temos que P^T é o estimador UMVU de P . Se colocarmos $X = R^k$ e $A = \{(a_1, \dots, a_k); a_1 \leq x_1, \dots, a_k \leq x_k\}$ temos o caso particular da estimação de uma função de distribuição. Como consequência do Lema temos também que todos os estimadores \hat{P} que consideraremos, serão medidas de probabilidade em A , para cada $t \in T$.

DEFINIÇÃO 1.2 - Dizemos que $\hat{p} = \hat{p}(\cdot; T)$ é o estimador UMVU de p , se T é suficiente e completa para P , $\hat{p}(x, t)$ é $A \times B$ -mensurável e $E\hat{p}(x, T) = p(x)$ q.c.

Os dois teoremas seguintes darão as condições para a existência do estimador não viciado para p .

TEOREMA 1.2 - Um estimador não viciado \hat{p} de p existe se e somente se existe um estimador $\hat{P} = P^T$ de P tal que:

- a) para cada $t \in T$ o estimador é absolutamente contínuo com respeito a μ ,
- b) a derivada de Radon Nikodym é $A \times B$ -mensurável.

PROVA - Suponhamos primeiramente que $E\hat{p} = p$, ou seja \hat{p} é um estimador não viciado de p e consideremos:

$$\tilde{P}(A) = \int_A \hat{p} d\mu.$$

Assim,

$$E\tilde{P}(A) = E \int_A \hat{p} d\mu = \int_A (E\hat{p}) d\mu = \int_A p d\mu = P(A),$$

aplicando-se o teorema de Fubini. Então, \tilde{P} é não viciado e μ -contínuo, por construção. Suponhamos agora que $E\tilde{P} = P$ e $\tilde{P} \ll \mu$, denotando $\hat{p} = \frac{d\tilde{P}}{d\mu}$. Então,

$$\int_A p d\mu = E \int_A \hat{p} d\mu = \int_A (E\hat{p}) d\mu$$

Assim, pelo teorema de Radon Nikodym, temos que

$$E\hat{p} = p \quad \text{q.c.}\mu. \quad \square$$

TEOREMA 1.3 - Se T é uma estatística suficiente e completa para P , então um estimador UMVU \hat{p} de p existe, se e somente se, P^T satisfaz as condições a) e b) do teorema 1.2. Além disso, quando tal \hat{p} existe, ele é dado por

$$\frac{dP^T}{d\mu}.$$

PROVA: Se \hat{p} é UMVU, ele precisa ser, de acordo com a definição, uma função de T e, então

$$\int_A \hat{p} d\mu$$

é um estimador UMVU de P .

Segue-se pelo Lema 1.1 que $P^T \ll \mu$. Por outro lado, se $P^T \ll \mu$ então

$$\frac{dP^T}{d\mu} = \hat{p}$$

é um estimador não viciado de p , pelo teorema 1.2. Como ele

é uma função de T , temos que é o UMVU. \square

Note que o teorema 1.2 dá as condições para que exista o estimador não viciado de p , enquanto, que o teorema 1.3, complementando o teorema 1.2, dá as condições para que este estimador seja UMVU.

EXEMPLO 1.1 - Aplicação dos resultados dos Teorema 1.2 e 1.3 à família F de todas funções de distribuição absolutamente contínuas k -variadas.

Sabemos que $T = (X^{(1)}, \dots, X^{(n)})$, vetor das estatísticas de ordem é uma estatística suficiente e completa para F e que a função de distribuição amostral k -variada que é baseada em T , é o estimador UMVU para a função de distribuição desconhecida F em F .

Mas, este estimador não é absolutamente contínuo. Então, pela parte a) do teorema 1.2, temos que não existe estimador não viciado de uma densidade, ou seja qualquer estimador não viciado de uma função de densidade não é uma densidade, o que não é muito bom. \square

Note-se que o teorema 1.1 segue como um caso particular do exemplo 1.1. Partindo, então, do fato que não existem estimadores não viciados para o caso das funções de distribuição absolutamente contínuas, vamos procurar encontrar o estimador ótimo, que minimize o erro quadrático, médio e também possua boas propriedades assintóticas. Como uma medida da conveniência das classes particulares de estimadores tomamos a integral do erro quadrático médio, que tem sido sugeridas por um grande número de autores. Nos capítulos seguintes, discutiremos com detalhes duas classes de estimadores e faremos uma discussão breve de algumas outras classes. Discutiremos aplicações, à teoria da confiabilidade e à estimação de funções de risco, que consideraremos no Capítulo 2.

CAPÍTULO 2

ESTIMAÇÃO DE UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE POR FUNÇÕES PESO

No capítulo anterior, apresentamos os resultados mais gerais da teoria da estimação de função de densidade e chegamos a importante conclusão de que não existe estimador não viciado. Neste capítulo, vamos discutir com detalhes a estimação de uma função de densidade por estimadores baseados em funções peso, que alguns autores denominaram estimadores do tipo algébrico. Primeiro apresentamos a definição e as propriedades principais da função de distribuição amostral. Depois, vamos definir a classe de estimadores, apresentando um exemplo particular e, em seguida, procuraremos condições convenientes, sobre a classe considerada, para que apresente resultados ótimos no sentido de boas propriedades assintóticas e erro quadrático médio mínimo. Finalmente, vamos fazer aplicações do método para a teoria da confiabilidade e para a estimação da função de risco. Faremos também algumas extensões para o caso multivariado e algumas comparações com a estimação do espectro. Os resultados que discutiremos encontram-se basicamente em Roseblatt [15], Parzen [8] e Epanechnikov [2].

1. A CLASSE DE ESTIMADORES

Seja X_1, \dots, X_n uma amostra aleatória de tamanho n da variável aleatória X , cuja função de distribuição $F(x)$ é

absolutamente contínua, com função de densidade $f(x)$.

No capítulo anterior, mostramos que o estimador UMVU para $F(x)$ existe e é dado pela função de distribuição amostral, definida por:

$$(2.1.1) \quad F_n(x) = \frac{\text{n}^\circ \text{ dos } X_i \leq x \text{ na amostra}}{n},$$

para um dado ponto x .

Com relação a função de distribuição amostral definida por (2.1.1), pode-se notar que:

- 1) Se $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ denotam as observações ordenadas e $x_{(0)} = 0, x_{(n+1)} = \infty$, então $F_n(x)$ salta $1/n$ para cada ponto $x_{(i)}$ e tem valor constante $1/n$ para todo ponto x tal que $x_{(i)} \leq x < x_{(i+1)}$, $i=1, 2, \dots, n$.
- 2) Uma definição alternativa de $F_n(x)$ é a seguinte:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x-X_i),$$

onde

$$I(x) = \begin{cases} 1 & \text{para } x \geq 0 \\ 0 & \text{para } x < 0 \end{cases}$$

Então, a variável aleatória $I(x-X)$ toma valores 0 e 1 com probabilidades $1-F(x)$ e $F(x)$, respectivamente. Portanto, a variável

$$nF_n(x) = \sum_{i=1}^n I(x-X_i)$$

é a soma de n variáveis aleatórias independentes e identicamente distribuídas com distribuição de Bernoulli. Então,

$$P[F_n(x)=i|n] = \binom{n}{i} [F(x)]^i [1-F(x)]^{n-i}, \quad i=0,1,\dots,n,$$

de onde segue-se que

$$(2.1.2) \quad E[F_n(x)] = F(x)$$

$$(2.1.3) \quad \text{Var}[F_n(x)] = \frac{[F(x)][1-F(x)]}{n}$$

Então, (2.1.2) e (2.1.3) acima, implicam que $F_n(x)$ é um estimador consistente e não viciado para $F(x)$.

3) Teorema de Glivenko-Cantelli. A convergência de $F_n(x)$ para $F(x)$ é uniforme em x , ou seja

$$P\left[\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{} 0\right] = 1$$

Para a demonstração do teorema acima o leitor poderá consultar Loève [6].

4)

$$\text{Cov}[F_n(x), F_n(y)] = \frac{1}{n} F(x)[1-F(y)]$$

para $y \geq x$.

DEFINIÇÃO 2.1.1 - Uma função integrável $\omega(\cdot)$ é chamada função peso, se ela satisfaz às condições:

$$(2.1.4) \quad \text{a) } 0 \leq \omega(y) < c < \infty$$

$$(2.1.5) \quad \text{b) } \omega(y) = \omega(-y)$$

$$(2.1.6) \quad \text{c) } \int_{-\infty}^{\infty} \omega(y) dy = 1$$

$$(2.1.7) \quad \text{d) } \int_{-\infty}^{\infty} \omega^2(y) y dy = 1$$

Note que as condições especificadas acima sobre $\omega(\cdot)$, implicam que

$$\int \omega^2(y) dy < \infty,$$

ou seja $\omega(\cdot)$ é de classe L^2 . Desde que todas as integrais neste trabalho são integrais definidas sobre a reta real, todas as integrais sem limites inferiores e superiores tem limites de integração implicitamente definidos de $-\infty$ a $+\infty$. Vamos, agora, definir nossa classe de estimadores.

DEFINIÇÃO 2.1.2 - Uma classe de estimadores $\hat{f}_n(x)$ de $f(x)$ é uma função dos valores amostrais de X , dada por

$$(2.1.8) \quad \hat{f}_n(x) = \frac{1}{nb(n)} \sum_{j=1}^n \omega\left(\frac{x-X_j}{b(n)}\right)$$

onde $b(n)$ é uma função de n , que tende a zero para $n \rightarrow \infty$.

Com as suposições feitas sobre $\omega(\cdot)$, temos que $\hat{f}_n(x)$ definido acima é uma função de densidade, pois

$$\hat{f}_n(x) \geq 0, \quad \forall x$$

e

$$\int \hat{f}_n(y) dy = \int \omega(y) dy = 1.$$

EXEMPLO 2.1.1 - Temos que

$$(2.1.9) \quad \hat{f}_n(x) = \frac{F_n(x+b(n)) - F_n(x-b(n))}{2b(n)}$$

é um particular estimador para $f(x)$, da forma (2.1.8).

Para mostrar isso, temos:

$$\begin{aligned}\hat{f}_n(x) &= \frac{F_n(x+b(n)) - F_n(x-b(n))}{2b(n)} = \frac{1}{2b(n)} \int_{x-b(n)}^{x+b(n)} dF_n(y) = \\ &= \frac{1}{b(n)} \int_{-\infty}^{\infty} \omega\left(\frac{x-y}{b(n)}\right) dF_n(y) = \frac{1}{nb(n)} \sum_{j=1}^n \omega\left(\frac{x-X_j}{b(n)}\right)\end{aligned}$$

onde

$$(2.1.10) \quad \omega(y) = \begin{cases} 1/2 & \text{se } |y| \leq 1 \\ 0 & \text{se } |y| > 1 \end{cases} \quad \square$$

O estimador (2.1.9) que é uma média ponderada da função de distribuição amostral, foi sugerido pela definição de derivada e foi introduzido por Roseblatt [14] em 1956. Isso sugeriu que se pode ter uma grande quantidade de estimadores de $f(x)$, pois ao invés de $\omega(y)$, definida em (2.1.10), poderíamos ter escolhidos outras funções $\omega(\cdot)$ que satisfaçam as condições da definição 2.1.1. Daí a justificativa da nossa definição 2.1.2.

Nas próximas duas secções procuraremos discutir algumas propriedades do estimador definido em (2.1.8). Na secção 2 vamos abordar a média, a variância e o erro quadrático médio de $\hat{f}_n(x)$; na secção 3 nos preocuparemos com a normalidade do estimador.

2. NÃO VICIOSIDADE ASSINTÓTICA E CONSISTÊNCIA DE $\hat{f}_n(x)$

LEMA 2.2.1 - a) O estimador $\hat{f}_n(x)$ é assintoticamente não vi-

b)

$$\text{var } \hat{f}_n(x) = \sigma^2(\hat{f}_n(x)) \cong \frac{f(x)}{nb(n)} \int \omega^2(u) du, \text{ para } n \rightarrow \infty.$$

PROVA: a) pela definição de esperança, temos:

$$\begin{aligned} E\hat{f}_n(x) &= \int \hat{f}_n(y) f(y) dy = \frac{1}{b(n)} \int \omega\left(\frac{x-y}{b(n)}\right) f(y) dy \\ (2.2.1) \quad &= \int \omega(u) f(x-b(n)u) du. \end{aligned}$$

Fazendo $n \rightarrow \infty$, o resultado segue-se.

$$\begin{aligned} b) \quad \text{Var}(\hat{f}_n(x)) &= \text{Var}\left\{\frac{1}{nb(n)} \sum_{j=1}^n \omega\left(\frac{x-X_j}{b(n)}\right)\right\} = \frac{1}{n} \text{var}\left\{\frac{1}{b(n)} \omega\left(\frac{x-X}{b(n)}\right)\right\} = \\ (2.2.2) \quad &= \frac{1}{n} \left[E\left\{\frac{1}{b(n)} \omega\left(\frac{x-X}{b(n)}\right)\right\}^2 - \left\{E \frac{1}{b(n)} \omega\left(\frac{x-X}{b(n)}\right)\right\}^2 \right] = \\ &= \frac{1}{n} \left[\frac{1}{b(n)} \int \omega^2(u) f(x-b(n)u) du - \left[\int \omega(u) f(x-b(n)u) du \right]^2 \right] \\ &= \frac{1}{nb(n)} \int \omega^2(u) f(x-b(n)u) du - \frac{1}{n} \left[\int \omega(u) f(x-b(n)u) du \right]^2 \end{aligned}$$

Mas, para $n \rightarrow \infty$ temos

$$\int \omega^2(u) f(x-b(n)u) du \rightarrow f(x) \int \omega^2(u) du$$

e

$$\int \omega(u) f(x-b(n)u) du \rightarrow f(x),$$

e então o resultado segue-se. \square

DEFINIÇÃO 2.2.1 - O erro quadrático médio do estimador $\hat{f}_n(x)$ é definido por

$$(2.2.3) \quad E[\hat{f}_n(x) - f(x)]^2 = \sigma^2(\hat{f}_n(x)) + [E\hat{f}_n(x) - f(x)]^2$$

É uma medida da dispersão do estimador $\hat{f}_n(x)$ em torno de $f(x)$. Quando

$$\lim_{n \rightarrow \infty} E[\hat{f}_n(x) - f(x)]^2 = 0$$

dizemos que $\hat{f}_n(x)$ é consistente em média quadrática.

O teorema seguinte dará uma condição para que o estimador definido em (2.1.7) seja consistente em média quadrática.

TEOREMA 2.2.1 - Se $nb(n) \rightarrow \infty$, quando $n \rightarrow \infty$, temos que o estimador $\hat{f}_n(x)$, definido em (2.1.8), é consistente em média quadrática.

PROVA: De acordo com (2.2.1) e (2.2.2), temos que

$$E[\hat{f}_n(x) - f(x)]^2 = \frac{1}{nb(n)} \int \omega^2(u) f(x-b(n)u) du - \frac{1}{n} \left[\int \omega(u) f(x-b(n)u) du \right]^2 + \left[\int \omega(u) \{f(x-b(n)u) - f(x)\} du \right]^2$$

Usando o Lema 1, o resultado segue-se. \square

Então, as condições que nós obtivemos sobre as funções peso $\omega(\cdot)$, para que os estimadores definidos em (2.1.8) sejam assintoticamente não viciados e consistentes em média quadrática, além das já estabelecidas na Definição 2.1.1, são:

$$(2.2.4) \quad \lim_{n \rightarrow \infty} b(n) = 0,$$

$$(2.2.5) \quad \lim_{n \rightarrow \infty} nb(n) = \infty.$$

Na secção 4 deste capítulo apresentaremos alguns exemplos de funções peso e discutiremos também o problema da

função peso ótima.

3. NORMALIDADE ASSINTÓTICA DE $\hat{f}_n(x)$

Temos que o estimador $\hat{f}_n(x)$ dado por (2.1.8) pode ser escrito como a média

$$\hat{f}_n(x) = \frac{1}{n} \sum_{j=1}^n V_j$$

onde $V_j = \frac{1}{b(n)} W\left(\frac{x-X_j}{b(n)}\right)$, com $b(n) \rightarrow 0$, para $n \rightarrow \infty$, e $\omega(\cdot)$ satisfazendo as condições da definição 2.1.1. De acordo com o que foi discutido no Lema 2.2.1 sobre a média e a variância de $\hat{f}_n(x)$, usando o teorema do limite central (quando $nb(n) \rightarrow \infty$, para $n \rightarrow \infty$, $b(n) \rightarrow 0$), nós obtemos uma aproximação normal para a distribuição de probabilidade de $\hat{f}_n(x)$, ou seja,

$$\lim_{n \rightarrow \infty} P \left[\frac{\hat{f}_n(x) - E\hat{f}_n(x)}{\sigma[\hat{f}_n(x)]} \leq c \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \Phi(c)$$

Uma idéia da proximidade, ou seja, do erro na aproximação, é dado pelo limite de Berry-Esseen (Loeve [6] pág. 288): para uma conveniente constante C, temos:

$$\sup_{-\infty < c < \infty} \left| P \left[\frac{\hat{f}_n(x) - E[\hat{f}_n(x)]}{\sigma[\hat{f}_n(x)]} \leq c \right] - \Phi(c) \right| \leq \frac{CE|V^3|^3}{n^{1/2}\sigma^3[V]}$$

$$= \frac{1}{[nb(n)f(x)]^{1/2}} \frac{\int |\omega(y)|^3 dy}{\left[\int \omega^2(y) dy \right]}$$

com $V = \frac{1}{b(n)} \omega\left(\frac{x-X}{b(n)}\right)$.

4. ALGUMAS EXTENSÕES PARA O CASO MULTIVARIADO

Agora, procuraremos estender para o caso multivariado, alguns dos resultados obtidos nas secções anteriores, onde os artigos básicos foram Roseblatt [14] e [15] e Parzen [8]. Existem poucos trabalhos neste sentido. Vamos discutir, basicamente o trabalho de Epanechnikov [2], que apresenta alguns dos resultados no caso multivariado. Vamos, primeira mente, definir nossa classe de estimadores.

Seja então,

$$X_i = (X_1^{(i)}, X_2^{(i)}, \dots, X_k^{(i)}) \quad i=1, 2, \dots, n$$

uma amostra aleatória de tamanho n , da distribuição de uma variável aleatória k -dimensional X com função de densidade de probabilidade k -variada $f(x_1, \dots, x_k)$.

DEFINIÇÃO 2.4.1 - O estimador $\hat{f}_n(x_1, \dots, x_k)$ de $f(x_1, \dots, x_k)$ baseado na amostra de tamanho n da variável aleatória k -dimensional X será dado por

$$(2.4.1) \quad \hat{f}_n(x_1, \dots, x_k) = \sum_{i=1}^n \prod_{j=1}^k \frac{1}{nb_j(n)} \omega_j \left(\frac{x_j - x_j^{(i)}}{b_j(n)} \right),$$

onde as constantes $b_j(n) \rightarrow 0$ para $n \rightarrow \infty$ e cada uma das funções peso satisfazem as condições estabelecidas na definição 2.4.1.

OBSERVAÇÃO: Como as variáveis aleatórias k -dimensionais X_1, \dots, X_n são independentes, pois formam uma amostra aleatória de tamanho n da distribuição da variável aleatória X , temos que

$$(2.4.2) \quad \omega \left(\frac{x_1 - x_1^{(i)}}{b_1(n)}, \dots, \frac{x_k - x_k^{(i)}}{b_k(n)} \right) = \prod_{j=1}^k \omega_j \left(\frac{x_j - x_j^{(i)}}{b_j(n)} \right) \quad i=1, \dots, n$$

Daí a justificativa de definirmos nossa classe de estimadores da forma (2.4.1). Ainda se $\omega_1 = \omega_2 = \dots = \omega_n = \omega$ a expressão (2.4.2) acima ficará:

$$\left[\omega \left(\frac{x - x^{(i)}}{b(n)} \right) \right]^n$$

No caso bivariado, tomando $b_1(n) = b_2(n) = b(n)$, podemos definir nossa classe de estimadores como

$$\hat{f}_n(x_1, x_2) = \frac{1}{nb^2(n)} \sum_{i=1}^n \omega \left(\frac{x_1 - x_1^{(i)}}{b(n)}; \frac{x_2 - x_2^{(i)}}{b(n)} \right)$$

que é a forma apresentada por Roseblatt [15], quando se procura estimar uma densidade bivariada.*

Vamos, agora, estudar as propriedades assintóticas principais do estimador definido em (2.4.1). Temos, então:

$$(2.4.3) \quad E\hat{f}_n(x_1, \dots, x_k) = \int \dots \int \left[\prod_{j=1}^k \omega \left(\frac{x_j - y_j}{b_j(n)} \right) \right] f(y_1, \dots, y_k) dy_1, \dots, dy_k$$

Como $\omega(\cdot)$ é uma função par, usando a transformação

$$u_j = \frac{y_j - x_j}{b_j(n)}, \quad j=1, \dots, k$$

temos que a expressão (2.4.3) toma a forma

$$(2.4.4) \quad E\hat{f}_n(x_1, \dots, x_k) = \int \dots \int \left[\prod_{j=1}^k \omega(u_j) \right] f(x_1 + u_1 b_1(n), \dots, x_k + u_k b_k(n)) dy_1 \dots dy_k$$

Observemos que a expressão (2.2.1) é obtida a partir de (2.4.4) no caso em que $k=1$. Expandindo:

$$f(x_1 + b_1(n)u_1, \dots, x_k + b_k(n)u_k)$$

em séries de Taylor com respeito a todos os x_i , em torno do ponto (x_1, \dots, x_k) , integrando o lado direito de (2.4.4) e fazendo $n \rightarrow \infty$, temos que o vício do estimador definido em (2.4.1) será dado por

$$E[\hat{f}_n(x_1, \dots, x_k) - f(x_1, \dots, x_k)] = \frac{1}{2} \sum_{j=1}^k \frac{\partial^2 f(x_1, \dots, x_k)}{\partial x_j^2} b_j^2(n)$$

Para o erro quadrático médio, temos:

$$\begin{aligned}
 & E[\hat{f}_n(x_1, \dots, x_k) - f(x_1, \dots, x_k)]^2 = \\
 & = E[\hat{f}_n^2(x_1, \dots, x_k)] - 2f(x_1, \dots, x_k)E\hat{f}_n(x_1, \dots, x_k) + f^2(x_1, \dots, x_k) = \\
 & = \frac{1}{n^2} \sum_{j=1}^k \frac{1}{b_j^2(n)} E \left[\sum_{i=1}^n \omega_j \left(\frac{x_j - x_j^{(i)}}{b_j(n)} \right) \right]^2 - \\
 & - 2f(x_1, \dots, x_k) \frac{1}{n} \sum_{j=1}^k \frac{1}{b_j(n)} E \left[\sum_{i=1}^n \omega_j \left(\frac{x_j - x_j^{(i)}}{b_j(n)} \right) \right] + f^2(x_1, \dots, x_k) = \\
 & \frac{1}{n} \sum_{j=1}^k \frac{1}{b_j(n)} \int \dots \int \left[\sum_{j=1}^k \omega_j(u_j) \right]^2 f(x_1 + b_1(n)u_1, \dots, x_k + b_k(n)u_k) du_1 \dots du_k + \\
 & + \frac{n-1}{n} \left[\int \dots \int \sum_{j=1}^k \omega_j(u_j) f(x_1 + b_1(n)u_1, \dots, x_k + b_k(n)u_k) du_1 \dots du_k \right]^2 - \\
 & - 2f(x_1, \dots, x_k) \int \dots \int \sum_{j=1}^k \omega_j(u_j) f(x_1 + b_1(n)u_1, \dots, x_k + b_k(n)u_k) du_1 \dots du_k + \\
 & + f^2(x_1, \dots, x_k).
 \end{aligned}
 \tag{2.4.5}$$

Expandindo novamente $f(x_1 + b_1(n)u_1, \dots, x_k + b_k(n)u_k)$ em série de Taylor, com respeito a x_i , em torno do ponto (x_1, \dots, x_k) , no lado direito de (2.4.5) e fazendo $n \rightarrow \infty$, temos:

$$(2.4.6) \quad E[\hat{f}_n(x_1, \dots, x_k) - f(x_1, \dots, x_k)]^2 \sim f(x_1, \dots, x_k) \prod_{j=1}^k \left[\frac{1}{nb_j(n)} \int \omega_j^2(y) dy \right] + \\ + \frac{1}{4} \left[\sum_{j=1}^k \frac{\partial^2 f(x_1, \dots, x_k)}{\partial x_j^2} b_j^2(n) \right]^2$$

Provamos então

TEOREMA 2.4.1 - Suponha que

$$b_j(n) \rightarrow 0 \quad \text{e} \quad n \prod_{j=1}^k b_j(n) \rightarrow \infty$$

quando $n \rightarrow \infty$. Então o estimador definido em (2.4.1) é um estimador consistente em média quadrática da verdadeira função de densidade f , para cada ponto (x_1, \dots, x_k) .

Para o estimador (2.4.1) podemos definir a integral do erro quadrático médio (I.E.Q.M.), que chamaremos de J , por

$$(2.4.7) \quad J = \int \dots \int E[\hat{f}_n(x_1, \dots, x_k) - f(x_1, \dots, x_k)]^2 dx_1 \dots dx_k$$

Usando o resultado (2.4.6) na expressão acima,

$$J \sim \frac{1}{n} \prod_{j=1}^k \frac{1}{b_j(n)} \int \omega_j^2(y) dy + \frac{1}{4} \int \dots \int \left[\sum_{j=1}^k \frac{\partial^2 f(x_1, \dots, x_k)}{\partial x_j^2} b_j^2(n) \right]^2 \cdot dx_1 \dots dx_k,$$

que é uma expressão assintótica ($n \rightarrow \infty$) para I.E.Q.M. Colocando $\omega_j = \omega$ e $b_j(n) = b(n)$ para $j=1, 2, \dots, k$, podemos dispor o resultado acima como

$$(2.4.8) \quad J \sim \frac{L^k}{nb^k(n)} + \frac{1}{4} b^4(n)M,$$

onde

$$(2.4.9) \quad L = \int \omega^2(y) dy,$$

$$M = \int \dots \int \left[\sum_{j=1}^k \frac{\partial^2 f(x_1, \dots, x_k)}{\partial x_j^2} \right]^2 dx_1 \dots dx_k$$

De acordo com a disposição (2.4.8); para encontrar a função peso $\omega_0(\cdot)$ que minimiza I.E.Q.M. é suficiente minimizar a expressão (2.4.9) para $b(n)$ e n fixados e $\omega(\cdot)$ satisfazendo as condições da definição 2.1.1. Epanechnikov [2] e Rosenblatt [15] encontram então que o mínimo dessa expressão; que é a função peso ótima, é dada por

$$(2.4.10) \quad \omega_0(y) = \begin{cases} \frac{3}{4} 5^{-1/2} (1-y/5) & \text{se } |y| \leq \sqrt{5} \\ 0 & \text{caso contrário} \end{cases}$$

Podemos ver que a função peso ótima $\omega_0(\cdot)$ é independente da verdadeira função de densidade, do tamanho da amostra e da dimensão do espaço, o que é bastante importante.

Abaixo temos a tabela 1, reproduzida de Epanechnikov [2], que apresenta exemplos de funções peso com o correspondente valor de L e a razão $r = L/L_0$, onde

$$L_0 = \int \omega_0^2(y) dy.$$

TABELA 1

| $\omega_0(y)$ | L | r |
|--|---------------|-------|
| $\omega_0(y)$ (2.4.10) | $3/5\sqrt{5}$ | 1 |
| $1/\sqrt{6} - y /6$ para $ y \leq \sqrt{6}$ 0 para $ y > \sqrt{6}$ | $\sqrt{6/9}$ | 1,015 |
| $(2\pi)^{-1/2} e^{-y^2/2}$ | $1/2\sqrt{5}$ | 1,051 |
| $1/2\sqrt{3}$ para $ y \leq \sqrt{3}$ 0 para $ y > \sqrt{3}$ | $1/2\sqrt{3}$ | 1,077 |
| $\frac{1}{2} e^{-\sqrt{2} y }$ | $1/4\sqrt{2}$ | 1,320 |

5. APLICAÇÕES À TEORIA DA CONFIABILIDADE E À ESTIMAÇÃO DA FUNÇÃO DE RISCO

Nesta secção vamos procurar definir estimadores para a confiabilidade e para a função de risco, baseados no estimador definido por (2.1.8). Procuraremos condições para que sejam consistentes e assintoticamente normais. Uma outra abordagem sobre o assunto, poderá ser encontrada em Murthy [7]. Seja, então, X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n, da distribuição da variável aleatória X, com função de distribuição F(x) que é absolutamente contínua e com função de densidade f(x). Em particular, quando a variável aleatória X é o tempo de falha de um item em um equipamento, temos que

$$(2.5.1) \quad F(x) = P[X \leq x]$$

é a probabilidade de que o item falhe até o instante x, en-

quanto que

$$R(x) = 1-F(x)$$

é a probabilidade de que o ítem falhe após o instante x . $R(x)$ é chamada *função de confiabilidade*, enquanto que f é a função

$$(2.5.2) \quad Z(x) = \frac{f(x)}{1-F(x)}$$

será referida como *função de risco*.

Como um estimador para $R(t)$, podemos ter

$$(2.5.3) \quad R_n(t) = \frac{\text{n}^\circ \text{ dos } X_i > t \text{ na amostra}}{n}$$

que é não viciado e consistente, pois, como já mostramos anteriormente para $F_n(t)$, temos que $R_n(t)$ é uma variável aleatória binomialmente distribuída com

$$E R_n(t) = R(t)$$

$$\text{Var } R_n(t) = \frac{R(t)[1-R(t)]}{n}$$

Usando para $\omega(\cdot)$ as suposições da definição 2.1.1 e escolhendo a sequência $b(n)$ como na definição 2.1.2, vamos definir nosso estimador para $R(t)$, como

$$(2.5.4) \quad R_n^*(t) = \int_t^\infty \hat{f}_n(x) dx = \frac{1}{nb(n)} \sum_{j=1}^n \int_t^\infty \omega \left[\frac{x-X_j}{b(n)} \right] dx$$

Definido o estimador, vamos procurar mostrar que ele é assintoticamente não viciado, consistente e assintóti-

camente normal. Então,

$$\begin{aligned} E R_n^*(t) &= E \left[\frac{1}{nb(n)} \sum_{j=1}^n \int_t^{\infty} \omega \left(\frac{x-X_j}{b(n)} \right) dx \right] = \\ &= \frac{1}{b(n)} \int_{-\infty}^{\infty} \int_t^{\infty} \omega \left(\frac{x-y}{b(n)} \right) f(y) dx dy = \\ &= \int_{-\infty}^{\infty} \int_t^{\infty} \omega(u) f(x-b(n) \cdot u) du dx \end{aligned}$$

Assim,

$$(2.5.5) \quad \lim_{n \rightarrow \infty} E R_n^*(t) = \int_{-\infty}^{\infty} \omega(u) du \int_t^{\infty} f(x) dx = R(t)$$

que prova a não viciosidade assintótica do estimador por nós definido em (2.5.4). Por outro lado, temos

$$\begin{aligned} (2.5.6) \quad \text{Var } R_n^*(t) &= \text{Var} \left[\frac{1}{nb(n)} \sum_{j=1}^n \int_t^{\infty} \omega \left(\frac{x-X_j}{b(n)} \right) dx \right] = \\ &= \frac{1}{n} E \left[\frac{1}{nb(n)} \sum_{j=1}^n \int_t^{\infty} \omega \left(\frac{x-X_j}{b(n)} \right) dx \right]^2 - \frac{1}{n} \left[E \left[\frac{1}{nb(n)} \sum_{j=1}^n \int_t^{\infty} \omega \left(\frac{x-X_j}{b(n)} \right) dx \right] \right]^2 \end{aligned}$$

onde

$$\begin{aligned} E \left[\frac{1}{nb(n)} \sum_{j=1}^n \int_t^{\infty} \omega \left(\frac{x-X_j}{b(n)} \right) dx \right]^2 &= \int_{-\infty}^{\infty} \left[\frac{1}{b(n)} \int_t^{\infty} \omega \left(\frac{x-y}{b(n)} \right) dx \right]^2 f(y) dy = \\ (2.5.7) \quad &= 1 - \frac{2}{b^2(n)} \int_{-\infty}^{\infty} \left[\int_t^{\infty} \omega \left(\frac{x-y}{b(n)} \right) dx \right] \omega \left(\frac{t-y}{b(n)} \right) F(y) dy, \end{aligned}$$

após uma integração por partes. Agora, fazendo as substituições

$$\frac{x-y}{b(n)} = z \quad \text{e} \quad \frac{t-y}{b(n)} = u$$

(2.5.7) pode escrito como

$$E \left[\frac{1}{b(n)} \int_t^\infty \omega \left(\frac{x-X_j}{b(n)} \right) dx \right]^2 = 1-2 \int_{-\infty}^\infty \left[\int_u^\infty \omega(z) dz \right] \omega(u) F(t-b(n)u) du$$

Assim, tomando limite de ambos os lados na expressão acima, temos

$$(2.5.8) \quad \lim_{n \rightarrow \infty} E \left[\frac{1}{b(n)} \int_t^\infty \omega \left(\frac{x-X_j}{b(n)} \right) dx \right]^2 = 1 - F(t)$$

desde que

$$\int_{-\infty}^\infty \left[\int_u^\infty \omega(x) dx \right] \omega(u) du = 1/2 \quad (\text{Murthy, [7]}) .$$

Comparando os resultados (2.5.5), (2.5.6), (2.5.8), temos

$$\lim_{n \rightarrow \infty} n \text{Var } R_n^*(t) = R(t)[1-R(t)]$$

que prova a consistência do estimador (2.5.4). Note-se que $R_n(t)$ e $R_n^*(t)$ tem a mesma variância assintótica.

Chamando as variáveis

$$\int_t^\infty \frac{1}{b(n)} \omega \left(\frac{x-X_j}{b(n)} \right) dx = V_j$$

temos que o estimador (2.5.4) pode ser escrito como a média:

$$f_n(x) = \frac{1}{n} \sum_{j=1}^n V_j$$

de n variáveis aleatórias independentes e todas com a mesma distribuição. De acordo com (2.5.5), (2.5.8) e com os resultados da seção 2.3, usando o teorema do limite central, obtemos a aproximação normal desejada.

Quanto à função de risco dada por

$$Z(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{R(t)}$$

vamos propor inicialmente o estimador

$$(2.5.9) \quad Z_n(t) = \frac{\hat{f}_n(t)}{R_n(t)}$$

onde $\hat{f}_n(t)$ e $R_n(t)$ são dados respectivamente por (2.1.8) e (2.5.3). Já mostramos que $\hat{f}_n(t)$ é um estimador consistente de $f(t)$ e, que $R_n(t)$ é um estimador consistente de $R(t)$. Então, um teorema de convergência bastante conhecido que pode ser encontrado em Cramer [1], página 254 conclui a consistência de $Z_n(t)$ ou seja, $Z_n(t)$ é um estimador consistente de $Z(t)$.

Por outro lado, na seção 2.3, mostramos que $\hat{f}_n(x)$ é assintoticamente normal, ou seja

$$(2.5.10) \quad \lim_{n \rightarrow \infty} P \left[\left[nb(n) \right]^{\frac{1}{2}} \cdot \frac{\hat{f}_n(t) - f(t)}{\left[f(t) \int_{-\infty}^{\infty} \omega^2(x) dx \right]^{1/2}} < x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

Usando (2.5.10) e o fato de que $R_n(t)$ é um estimador consistente de $R(t)$, temos de acordo com o mesmo teorema

de convergência citado acima que

$$(2.5.11) \quad \lim_{n \rightarrow \infty} P \left[\left[nb(n) \right]^{\frac{1}{2}} \cdot \frac{\frac{\hat{f}_n(t)}{R_n(t)} - \frac{f(t)}{R(t)}}{\left(\frac{f(t)}{R^2(t)} \int_{-\infty}^{\infty} \omega^2(x) dx \right)^{1/2}} < x \right] = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Considerando agora,

$$y_n = \left[nb(n) \right]^{1/2} \left(R_n(t) - R(t) \right)$$

temos

$$E y_n = 0$$

$$\text{Var } y_n = b(n)R(t) \left(1 - R(t) \right),$$

de onde conclui-se que y_n é um estimador consistente. Este resultado e (2.5.11) nos permitem concluir com auxílio do teorema de convergência especificado acima que

$$\lim_{n \rightarrow \infty} P \left[\left[nb(n) \right]^{\frac{1}{2}} \cdot \frac{Z_n(t) - Z(t)}{\left(\frac{Z(t)}{R(t)} \int_{-\infty}^{\infty} \omega^2(x) dx \right)^{1/2}} < x \right] = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

o que prova que $Z_n(t)$ é assintoticamente normal. Os resultados de consistência e normalidade assintótica que discutimos, encontram-se em Murthy [7].

Um outro estimador, que podemos considerar para a função de risco é dado por

$$Z_n^*(t) = \frac{\hat{f}_n(t)}{R_n^*(t)}.$$

A prova de que este estimador é consistente e assintoticamente normal, é exatamente igual a que fizemos para provar que o estimador (2.5.9) é consistente e assintoticamente normal.

6. ALGUMAS COMPARAÇÕES COM A ESTIMAÇÃO DA FUNÇÃO DE DENSIDADE ESPECTRAL

Nesta secção procuraremos algumas relações entre a estimação de função de densidade e a estimação de de uma função de densidade espectral. Faremos observações sobre alguns resultados da teoria da estimação espectral que são comparáveis ou relacionam-se com aqueles mencionados na estimação de uma função de densidade de probabilidade.

Seja $X(t)$, $t = \dots, -1, 0, 1, \dots$ um processo estacionário vetorial de dimensão r , com parâmetro discreto. Se o momento de segunda ordem existe, então

$$(2.6.1) \quad \underline{X}(t) = \int_{-\pi}^{\pi} e^{it\lambda} d\underline{z}(\lambda)$$

onde $\underline{z}(\lambda)$ é um processo vetorial de dimensão r (com componentes complexos) de incrementos ortogonais tais que

$$(2.6.2) \quad E d\underline{z}(\lambda) d\underline{z}(u)' = \delta(\lambda+u) dF(\lambda), \quad -\pi < \lambda, u < \pi$$

Em (2.6.2)

$$\delta(\lambda) = \begin{cases} 1 & \text{se } \lambda = 0 \\ 0 & \text{caso contrário,} \end{cases}$$

F é uma função matricial Hermitiana $r \times r$ não decrescente ($F(\lambda) - F(\mu)$ é positiva semidefinida), e para toda matriz A , A' denota sua transposta. Como os componentes de $\underline{X}(t)$ são reais,

$$\overline{d\underline{Z}(-\lambda)} = d\underline{Z}(\lambda)$$

e $d\underline{F}(\lambda) = d\underline{F}(-\lambda)'$. Sem perda de generalidade podemos supor que

$$E\underline{X}(t) = \underline{0}$$

A matriz de covariância dada por

$$\underline{F}(t) = E\underline{X}(u)\underline{X}(t+u)'$$

esta relacionada com a função de distribuição espectral de 2.^a ordem \underline{F} por

$$\underline{r}(t) = \int_{-\pi}^{\pi} e^{i\lambda t} d\underline{F}(\lambda)$$

Considerando a existência de todos os momentos, temos que estes são dados por

$$\begin{aligned} m_{a_1, \dots, a_k}(t_1, \dots, t_k) &= EX_{a_1}(t) \dots X_{a_k}(t) = \\ &= m_{a_1, \dots, a_k}(t_1+t, \dots, t_k+t) \end{aligned}$$

onde $X_a(t)$ é o a -ésimo componente de $\underline{X}(t)$. As representações de Fourier para os momentos são dadas por

$$(2.6.3) \quad m_{a_1, \dots, a_k}(t_1, \dots, t_k) = \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} e^{i\sum t_j \omega_j} dG_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k)$$

onde

$$dG_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k) = E \left[\prod_{j=1}^k dZ_{a_j}(\omega_j) \right]$$

e G é de variação limitada. A estacionaridade do processo $\underline{X}(t)$ implica que

$$d\underline{G} = \underline{0} \text{ a menos que } \sum_{j=1}^k \omega_j = 0 \pmod{2\pi}$$

A representação (2.6.3) pode também ser colocada na forma

$$(2.6.4) \quad c_{a_1, \dots, a_k}(t_1, \dots, t_k) = \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} e^{i \sum_{j=1}^k t_j \omega_j} dF_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k)$$

onde

$$dF_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k) = c\{dZ_{a_j}(\omega_j); j=1, \dots, k\}$$

e $c(\cdot)$ é a função cumulante correspondente. É conveniente considerar que os cumulantes

$$c_{a_1, \dots, a_k}(t_1, \dots, t_k)$$

estão em L_1 como função de seus $k-1$ (ou k) argumentos, pois isto implica que

$$dF_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k) \eta \left(\sum_{j=1}^k \omega_j \right) = f_{a_1, \dots, a_k}(\omega_1, \dots, \omega_k) \eta \left(\sum_{j=1}^k \omega_j \right) d\omega_1, \dots, d\omega_k$$

($\eta(\omega) = 0$ se $\omega \neq 0 \pmod{2\pi}$) e é igual a 1, caso contrário) onde a função de densidade espectral f acima é contínua e é uma função de $k-1$ variáveis ω , já que

$$\sum_{j=1}^k \omega_j = 0 \pmod{2\pi}.$$

Para um processo estacionário real $X(t)$ de dimensão 1, podemos considerar para a função de densidade espectral, o estimador

$$(2.6.5) \quad f^{(N)}(\lambda) = \frac{1}{2\pi} \sum_{j=-N}^N \omega_j^{(N)} R_j^{(N)} e^{-ij\lambda}$$

com

$$\omega_j^{(N)} = \int_{-\pi}^{\pi} e^{ij\lambda} W_N(\lambda) d\lambda$$

e

$$R_j^{(N)} = \frac{1}{N} \sum_{\substack{t=1 \\ t-u=j}}^N \sum_{u=1}^N X(t) X(u).$$

A expressão (2.6.5) também pode ser escrita como

$$f^{(N)}(\lambda) = \int_{-\pi}^{\pi} W_N(\lambda-\alpha) I^{(N)}(\alpha) d\alpha$$

com

$$I^{(N)}(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N X(t) e^{-it\lambda} \right|^2,$$

o periodograma que sob certas condições de regularidade é assintoticamente não viciado, sendo no entanto bastante instavel, pois

$$\text{Var } I(\lambda) = f^2(\lambda) + o(N^{-1}).$$

É conveniente supormos que

$$(2.6.6) \quad W_N(u) = A_N B_N^{-1} W(B_N^{-1} u) \text{ se } |u| \leq \pi$$

com $B_N \rightarrow 0$ quando $n \rightarrow \infty$ e A_N como um fator de normalização de modo que

$$\int_{-\pi}^{\pi} W_N(u) du = 1$$

Se W tem segundo momento finito, é não negativa e simétrica em torno do zero e $NB_N \rightarrow \infty$, com W_N estendida periodicamente, de período 2π , podemos mostrar que

$$(2.6.7) \quad B_N^{-2} \left(E f^{(N)}(\lambda) - f(\lambda) \right) = \int u^2 w(u) du \frac{f''(\lambda)}{2} + o(B_N^{-3})$$

e

$$(2.6.8) \quad \text{Cov} \left\{ f^{(N)}(\lambda), f^{(N)}(\mu) \right\} = 2\pi N^{-1} \left\{ \int_{-\pi}^{\pi} W_N(\lambda - \alpha) W_N(\mu + \alpha) f^2(\alpha) d\alpha + \int_{-\pi}^{\pi} W_N(\lambda - \alpha) W_N(\mu - \alpha) f^2(\alpha) d\alpha + o(N^{-1}) \right\}$$

De (2.6.6) e (2.6.8), segue-se que:

$$(2.6.9) \quad \lim_{N \rightarrow \infty} NB_N \text{Cov}(f^{(N)}(\lambda), f^{(N)}(\mu)) =$$

$$= \begin{cases} 0 & \text{se } \lambda \neq \mu \\ 2\pi f^2(\lambda) \int_{-\pi}^{\pi} W(u) du & \text{se } \lambda = \mu \neq 0, \pi \\ 4\pi f^2(\lambda) \int_{-\pi}^{\pi} W^2(u) du & \text{se } \lambda = \mu = 0, \pi \end{cases}$$

quando $0 \leq \lambda, \mu \leq \pi$.

Note, com relação ao estimador $\hat{f}_n(x)$, que a utilização de (2.2.1) e o desenvolvimento de Taylor para

$$\hat{f}_n(x - b(n)u)$$

no ponto x , nos dá

$$(2.6.10) \quad b_n^{-2} \left(E\hat{f}_n(x) - f(x) \right) = \int u^2 \omega(u) du \frac{f''(x)}{2} + o(b^2(n))$$

que de certa forma é bastante parecido com o resultado (2.6.7). Os resultados assintóticos obtidos na estimação espectral diferem um pouco daqueles obtidos na estimação de uma função de densidade. O domínio básico onde (2.6.9) vale é $0 < \lambda < \pi$, com resultado ligeiramente diferentes para $\lambda = 0$ e π . Por outro lado, o resultado definido em (b) do Lema 2.2.1 para a estimação da função de densidade de probabilidade vale sobre todo o domínio da variável X , sendo que a constante $f(x)$ em (b) do Lema 2.2.1 é substituído em (2.6.9) por $2\pi f^2(\lambda)$. Quanto ao erro quadrático médio, os resultados obtidos por Epanechnikov [2] que discutimos na seção 4 continuam valendo para $W(\cdot)$, ou seja, a função peso ótima $W_0(\cdot)$ é ainda da forma (2.4.10).

CAPÍTULO 3

ESTIMAÇÃO DE UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE POR SISTEMAS ORTOGONAIS

Na secção anterior nós consideramos o problema de estimar uma função de densidade por funções peso. Vimos então que para se obter um estimador com boas propriedades é necessário que as funções peso $\omega(\cdot)$ satisfaçam propriedades convenientes. Neste capítulo nós consideraremos o problema de estimação de uma função de densidade por séries ortogonais. Discutiremos basicamente os artigos de Roseblatt [15], Schwartz [17], Watson [19] e Kronmal & Tarter [4]. A idéia de se estimar uma função de densidade através de Sistemas Ortogonais surgiu da expansão de uma função de densidade em séries de polinômios ortogonais, onde os mais usados são os polinômios de Tchebychef-Hermite (ver apêndice 2). Essas aproximações são usadas para se obter uma expressão matemática que represente os dados de maneira satisfatória. Alguns desses resultados serão discutidos no apêndice e encontram-se basicamente em Kendall & Stuart [3].

1. A CLASSE DE ESTIMADORES

Seja $\{\phi_j(x)\}$ um sistema ortogonal completo com relação a uma dada função peso $\omega(\cdot)$, que é considerada ser não negativa. Vamos exigir que o sistema seja ortonormal completo, porque esta suposição é mais conveniente para certos

propriedades dos estimadores que discutiremos. No apêndice procuraremos fazer uma breve e objetiva discussão sobre funções ortogonais. Suponhamos que a densidade f seja de quadrado integrável (de classe L^2), de modo que a série ortogonal

$$(3.1.1) \quad \sum_j c_j \phi_j(x)$$

com

$$(3.1.2) \quad c_j = \int f(x) \overline{\phi_j(x)} \omega(x) dx$$

converge para f em média quadrática em L^2 . Do ponto de vista estatístico, a questão importante não é se uma série infinita pode representar uma função de densidade, mas se um número finito de termos podem conduzir a uma aproximação satisfatória. O que pode ocorrer claramente é que a soma de um número finito de termos pode ser negativa, particularmente próximo das caudas. O nosso problema se resume então em estimar as constantes c_i , $i=1,2,\dots$, que são denominadas constantes de Fourier. Seja então, X_1, \dots, X_n uma amostra aleatória de tamanho n de uma população com função de distribuição absolutamente contínua $F(x)$, função de densidade $f(x)$ e seja $F_n(x)$ a função de distribuição amostral baseada nessas n observações. Então, em (3.1.2) se passarmos para a integral de Stieljes, usando a função de distribuição $F_n(x)$, teremos que um estimador plausível de c_j será dado por

$$(3.1.3) \quad \hat{c}_j(n) = \int \overline{\phi_j(x)} \omega(x) dF_n(x) = \frac{1}{n} \sum_{k=1}^n \overline{\phi_j(X_k)} \omega(X_k)$$

Vamos ver, agora, algumas propriedades estatísticas desse estimador. Assim, como

$$(3.1.4) \quad E\hat{c}_j(n) = \int \frac{1}{n} \sum_{k=1}^n \overline{\phi_j(X_k)} \omega(x) f(x) dx = \int \overline{\phi_j(x)} f(x) \omega(x) dx = c_j,$$

temos que \hat{c}_j é um estimador não viciado de c_j . Quanto à covariância deste estimador temos

$$\begin{aligned}
 \text{cov}(\hat{c}_j(n), \hat{c}_k(n)) &= \text{cov}\left(\frac{1}{n} \sum_{t=1}^n \phi_j(X_t) \omega(X_t); \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) \omega(X_i)\right) = \\
 &= \frac{1}{n} \text{cov}(\phi_j(X) \omega(X), \phi_k(X) \omega(X)) = \\
 &= \frac{1}{n} \left\{ E \omega^2(X) \phi_j(X) \overline{\phi_k(X)} - E \phi_j(X) \omega(X) \overline{E \phi_k(X) \omega(X)} \right\} = \\
 (3.1.5) \quad &= \frac{1}{n} \left\{ \int \phi_j(x) \overline{\phi_k(x)} \omega^2(x) f(x) dx - c_j \bar{c}_k \right\}
 \end{aligned}$$

Então, de (3.1.5) segue-se que

$$(3.1.6) \quad \text{var } \hat{c}_j(n) = \frac{1}{n} \left\{ \int |\phi_j(x)|^2 \omega^2(x) f(x) dx - |c_j|^2 \right\}$$

Da expressão (3.1.6) podemos concluir que \hat{c}_j é um estimador consistente de c_j , no sentido de a variância do estimador vai para zero quando $n \rightarrow \infty$.

Baseado no que discutimos até agora nesta secção e com a suposição adicional de que as funções $\phi_j(x)$ são contínuas no domínio de definição, temos a

DEFINIÇÃO 3.1.1 - Uma classe de estimadores $\hat{f}_n(x)$ de $f(x)$ é dada por

$$(3.1.7) \quad \hat{f}_n(x) = \sum_{j=1}^{\infty} \hat{c}_j(n) \lambda_j(n) \phi_j(x)$$

onde a sequência $\{\lambda_j(n)\}$ é escolhida de modo a melhorar as propriedades de $f(x)$.

Procuraremos condições sobre as constantes $\lambda_j(n)$,

de modo que o estimador definido em (3.1.7) seja consistente, assintoticamente não viciado e minimize a integral do erro quadrático médio definida neste caso por

$$E \int (\hat{f}_n(x) - f(x))^2 dx$$

2 - NÃO VICIOSIDADE, NORMALIDADE ASSINTÓTICAS E CONSISTÊNCIA DE $\hat{f}_n(x)$

Mostramos em (3.1.4) que \hat{c}_j é um estimador não viciado de c_j . Isso implica que o estimador definido em (3.1.7) é um estimador viciado da densidade f e o vício do estimador é dada por

$$(3.2.1) \quad E(\hat{f}_n(x) - f(x)) = \sum_{j=0}^{\infty} c_j (\lambda_j(n) - 1) \phi_j(x)$$

Usando (3.1.5) temos que a covariância do estimador (3.1.7) é dado por

$$\begin{aligned} \text{Cov}(\hat{f}_n(x), \hat{f}_n(y)) &= \sum_{j,k} \phi_j(x) \overline{\phi_k(x)} \lambda_j(n) \lambda_k(n) \text{cov}(\hat{c}_j(n), \hat{c}_k(n)) = \\ &= \frac{1}{n} \sum_{j,k} \phi_j(x) \overline{\phi_k(x)} \lambda_j(n) \lambda_k(n) \cdot \left\{ \int \phi_j(u) \overline{\phi_k(u)} \omega^2(u) f(u) du - c_j \overline{c_k} \right\} \end{aligned}$$

de onde segue-se que

$$\text{var } \hat{f}_n(x) = \frac{1}{n} \sum_{j=0}^{\infty} \lambda_j^2(n) |\phi_j(x)|^2 \left\{ \int |\phi_j(x)|^2 \omega^2(x) f(x) dx - |c_j|^2 \right\}$$

Quanto à I.E.Q.M., temos

$$\begin{aligned} E \int (f(x) - \hat{f}_n(x))^2 dx &= E \int \left(\sum_{j=0}^{\infty} c_j \phi_j(x) - \sum_j \hat{c}_j(n) \lambda_j(n) \phi_j(x) \right)^2 dx = \\ &= E \int \left(\sum_{j=0}^{\infty} (c_j - \hat{c}_j(n) \lambda_j(n)) \right)^2 |\phi_j(n)|^2 dx = \\ &= \sum_{j=0}^{\infty} E (c_j - \hat{c}_j(n) \lambda_j(n))^2 \end{aligned}$$

Como

$$E(c_j - \hat{c}_j(n) \lambda_j(n)) = c_j (1 - \lambda_j(n))$$

e

$$\begin{aligned} E(c_j - \hat{c}_j(n) \lambda_j(n))^2 &= c_j^2 (1 - \lambda_j(n))^2 = \\ &= \frac{1}{n} \lambda_j^2(n) \text{var } \phi_j(X) \end{aligned}$$

temos que

$$(3.2.2) \quad E \int (\hat{f}_n(x) - f(x))^2 = \sum_{j=0}^{\infty} \left\{ c_j^2 (1 - \lambda_j(n))^2 + \frac{1}{n} \lambda_j^2(n) \text{var } \phi_j(X) \right\}.$$

Observamos, então, de acordo com (3.2.1) e (3.2.2), que se para cada j fixado $\{\lambda_j(n)\} \rightarrow 1$ quando $n \rightarrow \infty$, o estimador (3.1.7) é assintoticamente não viciado e consistente no sentido da I.E.Q.M. Por outro lado, derivando o lado direito de (3.2.2) com relação a $\lambda_j(n)$, temos que o mínimo da expressão (3.2.2) é atingido para

$$(3.2.3) \quad \lambda_j(n) = \frac{\frac{c_j^2}{E\phi_j^2(X)}}{\frac{1}{n} \left[1 + \frac{(n-1)c_j^2}{E\phi_j^2(X)} \right]}$$

Assim, substituindo (3.2.3) em (3.2.2) temos um mínimo para a integral do erro quadrático médio, ou seja,

$$(3.2.4) \quad \min E \int (\hat{f}_n(x) - f(x))^2 dx = \sum_{j=0}^{\infty} \left[\frac{c_j^2 (E\phi_j^2(X) - c_j^2)}{E\phi_j^2(X) + (n-1)c_j^2} \right]$$

$$= \sum_{j=0}^{\infty} \frac{c_j \text{ var } \phi_j(X)}{\text{var } \phi_j(X) + c_j^2}$$

Como de (3.2.3), temos que para j fixado, $\lambda_j(n) \rightarrow 1$, para $n \rightarrow \infty$, isto nos sugere que

$$(3.2.5) \quad \lambda_j(n) = \begin{cases} 1 & \text{para } j = 1, 2, \dots, M(n) \\ 0 & \text{para } j = M(n)+1, \dots \end{cases}$$

de onde segue-se que (3.2.2) pode ser escrito como

$$(3.2.6) \quad E \int (\hat{f}_n(x) - f(x))^2 dx = \sum_{j=0}^{M(n)} n^{-1} \text{var } \phi_j(x) + \sum_{j=M(n)+1}^{\infty} c_j^2$$

Comparando (3.2.4) com (3.2.6) a escolha de $\{\lambda_j(n)\}$ será ótima quando $M(n)$ for escolhido de modo que os $\lambda_j(n)$ ($j \leq M(n)$) sejam grandes comparados com $n^{-1} \text{var } \phi_j(n)$ e os $\lambda_j(n)$ ($j > M(n)$) sejam desprezíveis. A escolha ótima dos $M(n)$ consta basicamente de Watson [19] e Schwartz [17].

EXEMPLO 3.2.1 - Seja

$$H_j(x) = (-1)^j e^{x^2} \left(\frac{d^j e^{-x^2}}{dx^j} \right)$$

o j -ésimo polinômio de Hermite. Sansone [16], pg.307-308, temos que o sistema

$$\phi_j(x) = \frac{e^{-x^2/2} H_j(x)}{2^j j! \pi^{1/2}} \quad j=0,1,2,\dots$$

é um sistema ortonormal completo em $L_1 \cap L_2$. Neste caso, temos que o estimador de c_j é dado por

$$\hat{c}_j = \frac{1}{n} \sum_{k=1}^n \frac{e^{-x_k^2/2} H_j(x_k)}{2^j j! \pi^{1/2}}$$

onde a sequência de constantes $\{\lambda_j(n)\}$ pode ser escolhida de acordo com (3.2.5) como

$$(3.2.7) \quad \lambda_j(n) = \begin{cases} 1 & \text{para } j \leq n \\ 0 & \text{para } j > n \end{cases}$$

Aqui $\omega(x) \equiv 1$. □

Como um segundo exemplo, vamos considerar um sistema particular de funções ortogonais, que discutiremos adiante com detalhes.

EXEMPLO 3.2.2 - O sistema trigonométrico

$$\phi_j(x) = \frac{e^{ijx}}{\sqrt{2\pi}} \quad -\pi < x < \pi$$

é também um sistema completo e ortogonal com relação as fun

ções peso $\omega(x) = 1$. Neste caso temos

$$\hat{c}_j = \frac{1}{n} \sum_{k=1}^n \frac{e^{-ijX_k}}{\sqrt{2\pi}}$$

onde a sequência de constantes $\{\lambda_j(n)\}$ pode ser escolhida da forma (3.2.7). \square

Para que possamos concluir a normalidade assintótica do estimador (3.1.7) é conveniente escrevê-lo na forma

$$(3.2.8) \quad \hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \omega(X_i) k_n(x, X_i),$$

onde

$$k_n(x, y) = \sum_{j=1}^{\infty} \lambda_j(n) \lambda_j(x) \overline{\phi_j(y)}$$

Notemos que esta nova forma, sugerida por Watson [19] é apenas uma nova maneira de escrever o estimador (3.1.7), pois

$$\begin{aligned} n^{-1} \sum_{i=1}^n \omega(X_i) k_n(x, X_i) &= n^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} \lambda_j(n) \phi_j(x) \overline{\phi_j(X_i)} \omega(X_i) \\ &= \sum_{j=1}^{\infty} \lambda_j(n) \phi_j(x) \frac{1}{n} \sum_{k=1}^n \overline{\phi_j(X_k)} \omega(X_i) \\ &= \sum_{j=1}^{\infty} \hat{c}_j(n) \lambda_j(n) \phi_j(x) = \hat{f}_n(x). \end{aligned}$$

Temos que a esperança do estimador (3.2.8) é calculada de maneira análoga à do estimador (3.1.7); é claro que (3.2.8) é assintoticamente não viciado. Quanto a variância, temos:

$$\begin{aligned} \text{var } \hat{f}_n(x) &= \text{var} \left\{ \frac{1}{n} \sum_{k=1}^n k_n(x, X_k) \omega(X_k) \right\} = \frac{1}{n^2} \text{var} \sum_{k=1}^n k_n(x, X_k) \omega(X_k) = \\ &= \frac{1}{n} \left\{ E |k_n(x, u)|^2 \omega^2(u) - |E k_n(x, u) \omega(u)|^2 \right\} = \\ (3.2.9) \quad &= \frac{1}{n} \left\{ \int |k_n(x, u)|^2 f(u) \omega(u) du - \left| \int k_n(x, u) \omega(u) f(u) du \right|^2 \right\} \end{aligned}$$

Para o caso em que $\omega(x) \equiv 1$, $k_n(x, u) = k_n(x-u)$,

$$\int k_n(x-u) du = 1$$

e, por conveniência, $k_n(u)$ é uma função peso não negativa que se comporta de maneira que

$$\int_{|u| \geq \epsilon} k_n(u) du \rightarrow 0$$

$\forall \epsilon > 0$, temos que o estimador definido em (3.2.8) continua sendo assintoticamente não viciado e a variância em (3.2.9) se reduz a

$$\text{var}[\hat{f}_n(x)] \cong \frac{f(x)}{n} \int k_n^2(u) du.$$

Analogamente temos

$$E |k_n(x, u) - E k_n(x, X)|^4 \cong f(x) \int k_n^4(u) du$$

e então pela forma do teorema do limite central de Liapunov a condição suficiente sobre a função peso $k(u)$ para que o estimador (3.2.7) seja assintoticamente normal, para $n \rightarrow \infty$, é que

$$\frac{\int k_n^4(u) du}{n \left(\int k_n^2(u) du \right)^2} \rightarrow 0$$

Este resultado encontra-se em Roseblatt [15]. Uma pequena observação que temos que fazer é sobre o problema da estimação de uma função de densidade multivariada por sistemas ortogonais. Se a densidade pode ser representada na forma

$$f(x_1, \dots, x_k) = \sum_{j_1, \dots, j_k} c_{j_1 \dots j_k} \phi_{j_1}(x_1) \dots \phi_{j_k}(x_k),$$

temos que c_{j_1, \dots, j_k} é estimado de maneira análoga a (3.1.3) e em (3.2.7) usamos a função peso multivariada como definida na secção 4 do capítulo 2. De maneira análoga podemos encontrar expressões para a esperança e variância, mas pouco se sabe sobre a normalidade assintótica.

3 - O USO DE SISTEMAS ORTOGONAIS TRIGONÔMÉTRICOS

Nas secções anteriores deste capítulo, ao discutirmos o problema da estimação de uma função de densidade por sistemas ortogonais, procuramos considerá-lo de modo amplo e geral. Nesta secção vamos considerar o mesmo problema, mas vamos conduzi-lo de modo a chegarmos a estimadores mais particulares. Para isso usaremos como sistema ortogonal uma classe bastante particular que chamaremos de sistema ortogonal trigonométrico. Esses resultados encontram-se basicamente em Kronmal e Tarter [4].

Seja então $\{\phi_k(x)\}$ um sistema ortogonal trigonométrico. Esse sistema tem a propriedade de que

$$(3.3.1) \quad \left\{ \frac{\partial \phi_k(x)}{\partial x} = c_k(x) \right\}$$

forma ainda um sistema ortogonal trigonométrico. Vamos então procurar um estimador para $F(x)$, baseado no sistema ortogonal trigonométrico $\{\phi_k(x)\}$ e, por derivação, obter um estimador para $f(x)$, que seja da forma

$$(3.3.2) \quad \hat{f}_m(x) = \sum_{k=0}^m \hat{a}_k \phi_k(x)$$

onde o número de termos m é escolhido de modo a melhorar as propriedades do estimador. Notemos que as formas (3.3.2) e (3.1.7) são as mesmas. Como

$$F(x) = \sum_{k=0}^{\infty} A_k \phi_k(x)$$

com

$$A_k = \int \phi_k(x) F(x) \omega(x) dx$$

onde $\{\phi_k(x)\}$ é um sistema ortogonal trigonométrico, podemos definir nosso estimador para $F(x)$ como

$$(3.3.3) \quad \hat{F}_n(x) = \sum_{k=0}^m \hat{A}_k \phi_k(x)$$

com \hat{a}_k e \hat{A}_k escolhidos de modo que

$$(3.3.4) \quad \begin{cases} E(\hat{A}_k) = A_k \\ E(\hat{a}_k) = a_k \end{cases}$$

De (3.3.4) segue-se imediatamente

$$E(F(x) - \hat{F}_m(x)) = \sum_{k=m+1}^{\infty} A_k \phi_k(x).$$

que é o vício do estimador (3.3.3).

O teorema seguinte de uma expressão para a integral do erro quadrático médio do estimador (3.3.3), expondo as vantagens da condição (3.3.4). A integral do erro quadrático médio é importante no sentido de que facilita as avaliações simultâneas dos valores ótimos de m e n para $\hat{F}_m(x)$ e $\hat{f}_m(x)$ onde n é o tamanho da amostra considerada e m é o número de termos em \hat{F}_m e em \hat{f}_m . Para facilitar vamos chamar de $J(\hat{F}_m(x))$ e $J(\hat{f}_m(x))$ as I.E.Q.M. dos estimadores $\hat{F}_m(x)$ e $\hat{f}_m(x)$.

TEOREMA 3.3.1 - Suponhamos que a condição (3.3.4) esteja verificada. Então

$$(3.3.5) \quad J(\hat{F}_m(x)) = \int F^2(x) \omega(x) dx - \sum_{k=0}^m (A_k^2 - \text{var } \hat{A}_k)$$

$$(3.3.6) \quad J(\hat{f}_m(x)) = \int f^2(x) \omega(x) dx - \sum_{k=0}^m (a_k^2 - \text{var } \hat{a}_k)$$

onde $\hat{F}_m(x)$ e $\hat{f}_m(x)$ são dados respectivamente por (3.3.3) e (3.3.2).

(A prova do teorema será feita apenas para (3.3.5))

DEMONSTRAÇÃO - Temos

$$\begin{aligned} \int \{F(x) - \hat{F}_m(x)\}^2 \omega(x) dx &= \int \left\{ F(x) - \sum_{k=0}^m \hat{A}_k \phi_k(x) \right\}^2 \omega(x) dx = \\ &= \int F^2(x) \omega(x) dx - 2 \sum_{k=0}^m \hat{A}_k \int F(x) \phi_k(x) \omega(x) dx + \int \left[\sum_{k=0}^m \hat{A}_k \phi_k(x) \right]^2 \omega(x) dx = \\ &= \int F^2(x) \omega(x) dx - 2 \sum_{k=0}^m \hat{A}_k A_k + \int \left[\sum_{k=0}^m \hat{A}_k^2 \phi_k^2(x) + \sum_{k \neq \ell} \sum_{\ell=0}^m A_\ell A_k \phi_k(x) \phi_\ell(x) \right] \omega(x) dx = \\ &= \int F^2(x) \omega(x) dx - 2 \sum_{k=0}^m \hat{A}_k \int F(x) \phi_k(x) \omega(x) dx + \sum_{k=0}^m A_k^2 \int \phi_k^2(x) \omega(x) dx + \end{aligned}$$

$$\begin{aligned} & + \sum_{k \neq \ell}^m \sum_{\ell}^m A_k A_\ell \int \phi_k(x) \phi_\ell(x) \omega(x) dx = \\ & = \int F^2(x) \omega(x) dx + \sum_{k=0}^m (\hat{A}_k - A_k)^2 - \sum_{k=0}^m A_k^2 \end{aligned}$$

Tomando esperança de ambos os lados temos

$$J(\hat{F}_m) = \int F^2(x) \omega(x) dx - \sum_{k=0}^m A_k^2 + \sum_{k=0}^m E(\hat{A}_k - A_k)^2$$

e então

$$J(\hat{F}_m(x)) = \int F^2(x) \omega(x) dx - \sum_{k=0}^m A_k^2 + \sum_{k=0}^m \text{var } \hat{A}_k$$

de onde segue (3.3.5). □

COROLÁRIO - Se as funções $\{\phi_k(x)\}$ formam um sistema ortogonal completo e (3.3.5) é válida, então

$$(3.3.7) \quad J(\hat{F}_m(x)) = \sum_{k=0}^m \text{var}(\hat{A}_k) + \sum_{k=m+1}^{\infty} A_k^2$$

A demonstração de (3.3.7) segue-se imediatamente do fato de que

$$\int F^2(x) \omega(x) dx = \sum_{k=0}^{\infty} A_k^2$$

ou seja, neste caso é válida a igualdade de Bessel.

Vamos, agora, encontrar um estimador para A_k que seja não viciado. Seja então X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da densidade f , que queremos estimar. A função de distribuição amostral relativa as n observações consideradas é dada por

$$(3.3.8) \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n \left[I_{(X_i, b)}(x) + \frac{1}{2} I_{[X_i, X_i]}(x) \right]$$

sendo que a densidade f tem região de suporte $[a, b]$, onde a e b são números reais. É então natural considerarmos

$$(3.3.9) \quad \hat{A}_k = \int \phi_k(x) F_n(x) \omega(x) dx = \frac{1}{n} \sum_{i=1}^n \int \phi_k(x) \varepsilon(x - X_i) \omega(x) dx$$

onde

$$\varepsilon(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0. \end{cases}$$

Pode-se mostrar que o estimador (3.3.9) é um estimador não viciado de A_k . Este resultado consta no seguinte

TEOREMA 3.3.2 - Seja A_k o k -ésimo coeficiente de Fourier da função de distribuição $F(x)$, cuja região de suporte é o intervalo $[a, b]$, onde a e b são reais e, sobre os quais as funções ortogonais $\{\phi_k\}$ estão definidas. Seja \hat{A}_k definido em (3.3.9). Então

$$E(\hat{A}_k) = A_k$$

DEMONSTRAÇÃO - Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da distribuição F . Como \hat{A}_k é simétrico em X_1, X_2, \dots, X_n , temos que

$$\begin{aligned} E(\hat{A}_k) &= E \left[\frac{1}{n} \sum_{i=1}^n \int \phi_k(x) \varepsilon(x - X_i) \omega(x) dx \right] \\ &= E \left[\int \phi_k(x) \varepsilon(x - X_i) \omega(x) dx \right] \end{aligned}$$

e essa expressão só é não nula se $X_i \leq x \leq b$.

Assim,

$$\begin{aligned} E(\hat{A}_k) &= \left[\int_y^b \phi_k(x) \omega(x) dx \right] = E \left[- \int_b^y \phi_k(x) \omega(x) dx \right] = \\ &= - \int_a^b \left[\int_b^y \phi_k(x) \omega(x) dx \right] dF(y). \end{aligned}$$

Usando integração por partes temos

$$E(\hat{A}_k) = - \left[F(y) \int_b^y \phi_k(x) \omega(x) dx \right]_a^b + \int_a^b F(y) \phi_k(y) \omega(y) dy = A_k \quad \square$$

Os sistemas ortogonais trigonométricos completos $\{\cos k\pi x\}, \{\text{sen } k\pi x\}, e^{ik\pi x}$ são os sistemas trigonométricos mais comumente usados e se $\{\phi_k(x)\}$ for escolhido como um deles, temos que a condição (3.3.1) estará verificado. Para o caso de estimação restrita a intervalos finitos $[a,b]$, temos que a função de distribuição F pode ser definida de maneira conveniente a esquerda do ponto a . Vamos então procurar escrever os sistemas trigonométricos apresentados acima, de maneira a atender a funções de distribuição restrita a intervalos finitos do tipo $[a,b]$. Então, cada uma das sequências ortogonais

$$(3.3.10) \quad \left\{ \cos k\pi \left(\frac{x-a}{b-a} \right) \right\}$$

$$(3.3.11) \quad \left\{ \text{sen } k\pi \left(\frac{x-a}{b-a} \right) \right\} \quad \text{ou}$$

$$(3.3.12) \quad \left\{ e^{ik\pi \left(\frac{x-a}{b-a} \right)} \right\}$$

são completas com respeito a uma função definida para ser:

i) par no intervalo $[2a-b,b]$

ii) ímpar no intervalo $[2a-b, b]$

ou

iii) zero no intervalo $[2a-b, a]$

A função peso $\omega(x)$ é por conveniência definida para toda reta real. Todavia, quando a função de distribuição F é restrita a intervalos finitos $[a, b]$ a função peso $\omega(x)$ será definida como

$$(3.3.13) \quad \omega(x) = \frac{2I_{[a,b]}(x)}{b-a}$$

e quando consideramos densidades com regiões de suporte infinita, tal como a distribuição normal, temos que as técnicas obtidas para densidades com região de suporte finito dão métodos de estimação suficientemente precisos para essas famílias mais gerais de distribuições. Se tomarmos

$$(3.3.14) \quad \left\{ \phi_k(x) = \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] \right\}$$

temos

$$A_0 = 0 \text{ e, para } k \geq 1$$

$$(3.3.15) \quad A_k = \frac{1}{n} \sum_{i=1}^n \int_a^b \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] e^{(x-X_i)} \omega(x) dx$$

e a expressão (3.3.15) só é não nula se $x \geq X_i$. Assim,

$$\begin{aligned} A_k &= \sum_{i=1}^n \frac{2I_{[a,b]}(X_i)}{n(b-a)} \int_a^b \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] dx = \\ &= \sum_{i=1}^n \frac{2I_{[a,b]}(X_i)}{n(b-a)} \frac{(b-a)}{k\pi} \left(-\cos \left[k\pi \left(\frac{x-a}{b-a} \right) \right] \right)_{X_i}^b = \end{aligned}$$

$$\begin{aligned}
 &= \frac{2}{k\pi} \left(\frac{1}{n} \sum_{i=1}^n \cos \left[k\pi \left(\frac{X_i - a}{b-a} \right) \right] I_{[a,b]}(X_i) - (-1)^k \right) = \\
 &= \frac{(b-a) \bar{c}_k - 2(-1)^k}{k\pi}
 \end{aligned}$$

onde $\cos k\pi = (-1)^k$ e,

$$(3.3.16) \quad \bar{c}_k = \frac{2}{(b-a)n} \sum_{i=1}^n \cos \left[k\pi \left(\frac{X_i - a}{b-a} \right) \right] I_{[a,b]}(X_i)$$

onde X_i é qualquer uma das n variáveis X_1, \dots, X_n que representam uma amostra aleatória de tamanho n da distribuição F . Concluimos então que o estimador de F baseado no sistema ortogonal (3.3.13) é, dado por

$$(3.3.17) \quad \hat{F}_m(x) = \sum_{k=1}^m \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] \left(\frac{(b-a) \bar{c}_k - 2(-1)^k}{k\pi} \right)$$

Usando o fato de que \bar{c}_k é desprezível para $k > m$, para um certo m , na maioria das distribuições que podemos considerar, e que

$$- \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] = \frac{x-a}{b-a}$$

podemos reescrever o estimador (3.1.17) como

$$(3.3.18) \quad \hat{F}_m(x) = \frac{x-a}{b-a} + \sum_{k=1}^m \frac{(b-a) \bar{c}_k}{k\pi} \text{sen} \left(k\pi \left(\frac{x-a}{b-a} \right) \right)$$

Uma observação que temos a fazer é que um resultado interessante obtido por Kronmal e Tarter foi

$$\lim_{m \rightarrow \infty} \hat{F}_m(x) = F_n(x)$$

onde F_n é dada por (3.3.8), de onde segue-se que

$$E\left[\lim_{m \rightarrow \infty} \hat{F}_m(x)\right] = F(x)$$

e daí, que $\lim_{m \rightarrow \infty} \hat{F}_m(x)$ é um estimador não viciado de $F(x)$.

Derivando (3.3.18) em relação a x temos

$$(3.3.19) \quad \hat{f}_m(x) = \frac{\bar{c}_0}{2} + \sum_{k=1}^m \bar{c}_k \cos k\pi \left(\frac{x-a}{b-a}\right)$$

onde

$$\frac{\bar{c}_0}{2} = \frac{1}{b-a}$$

Temos que $\hat{f}_m(x)$ é um estimador de forma (3.3.2), onde

$$\left\{ \phi_k(x) = \cos \left[k\pi \left(\frac{x-a}{b-a} \right) \right] \right\}$$

com $\hat{a}_0 = \bar{c}_0/2$, $\hat{a}_k = \bar{c}_k$, $k=1, \dots, m$ e satisfaz às condições do teorema (3.3.1), pois

$$E\hat{a}_k = a_k.$$

De maneira análoga podemos mostrar que o estimador para $f(x)$ baseado no sistema (3.3.10) é dado por.

$$(3.3.20) \quad \hat{f}_m^1(x) = \sum_{k=1}^m \bar{s}_k \operatorname{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right]$$

onde

$$\bar{s}_k = \frac{2}{n(b-a)} \sum_{i=1}^n \operatorname{sen} \left[k\pi \left(\frac{X_i - a}{b-a} \right) \right] I_{[a,b]}(X_i)$$

O novo estimador (3.3.20) é baseado no sistema ortogonal trigonométrico

$$\left\{ \phi_k(x) = \text{sen} \left[k\pi \left(\frac{x-a}{b-a} \right) \right] \right\}$$

com $\hat{a}_k = \bar{s}_k$. Para este estimador, também as condições do teorema (3.3.1) estão verificadas, ou seja, podemos utilizar este teorema para avaliar $J(\hat{f}_m(x))$ e $J(\hat{f}_m^1(x))$. Uma objeção que podemos fazer contra estes estimadores é que eles próprios não são densidades, ao contrário dos estimadores definidos no capítulo 2, os quais sempre definem uma densidade. Kronmal e Tarter [4] procuraram condições sobre m e n , de modo que $J(\hat{f}_m(x))$ converge para zero. Este resultado consta do seguinte

TEOREMA 3.3.3 - Se a série de Fourier de cosenos da densidade f converge uniformemente e se $m = o(\sqrt{n})$, então

$$\lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty}} E(\hat{f}_m(x) - f(x))^2 = 0$$

Uma observação importante a ser feita é que a importância do teorema 3.3.3 não reside no fato de que sob certas condições a I.E.Q.M. vai para zero no limite e sim devido ao fato de que a condição $m = o(\sqrt{n})$ indica que o número de termos do estimador pode ser mantido constante relativo ao tamanho da amostra. Outro fato interessante a ser ressaltado é que nós podemos expressar os estimadores trigonométricos com cosenos como estimadores algébricos da forma como definimos no capítulo 2, usando

$$\omega(y) = \frac{1}{2\pi} \left[\frac{\text{sen}(y/2)}{y/2} \right]^2$$

Kronmal e Tarter discutiram ainda a escolha do número de termos em $\hat{f}_m(x)$. Eles concluíram que o número ótimo de termos, m , depende de $(b-a)/B$ e, de natureza da distribuição a ser estimada. Aqui a , b e B são respectivamente o limite inferior e o limite superior do domínio da densidade e um parâmetro de escala da densidade de interesse.

CAPÍTULO 4

OUTROS MÉTODOS DE ESTIMAÇÃO

Para concluir, faremos uma breve discussão sobre outros métodos de estimação, procurando discutir algumas de suas propriedades. Os estimadores que consideramos no capítulo 2 são considerados estimadores do tipo algébrico, que são casos particulares do que os autores chamam de "Kernel Estimates". Diversos autores se preocuparam com esse tipo mais geral de estimadores. Um trabalho interessante que vale a pena ressaltar é o trabalho de Watson e Leadbetter [18].

Um passo importante além de considerar a consistência de um estimador "ótimo" é determinar se um estimador prático (não baseado na densidade desconhecida) pode ser assintoticamente equivalente a um estimador "ótimo". Esse passo foi considerado por Watson e Leadbetter [18]. Os estimadores por eles considerados, são de um tipo mais geral que aqueles que consideramos no capítulo 2, podendo ser expressados na forma abaixo, com k_n satisfazendo condições convenientes,

$$(4.1) \quad \hat{f}_n(x) = \frac{1}{n} \sum_{j=1}^n k_n(x, X_j)$$

onde X_1, \dots, X_n representa uma amostra aleatória de tamanho n da distribuição da densidade f , desconhecida. Notemos que os estimadores propostos no capítulo 2 para a densidade f são tais que

$$(4.2) \quad k_n(x,y) = \frac{1}{b(n)} \omega\left(\frac{x-y}{b(n)}\right)$$

onde $\omega(\cdot)$ é definida de acordo com a definição 2.1.1. Watson e Leadbetter consideraram a I.E.Q.M. como uma medida da bondade de seu estimador e encontraram o $k_n(x,y)$ que minimiza a I.E.Q.M., denotado por δ_n^* . Em geral δ_n^* é uma expressão complicada, a qual depende em forma explícita da densidade f . Watson e Leadbetter mostraram que o comportamento do estimador depende do comportamento da função característica ϕ_f . Colocando $k_n(x,y) = \delta_n(x-y)$ na definição (4.1) acima e usando a fórmula de Parseval, temos que mínimo da I.E.Q.M. é obtido tomando-se δ_n dado por

$$(4.3) \quad \phi_{\delta_n}(t) = \frac{|\phi_f(t)|^2}{(1/n) + [(n-1)/n] |\phi_f(t)|^2}$$

Temos então que o mínimo da I.E.Q.M. é dado por

$$(4.4) \quad J^*(\hat{f}_n(x)) = \frac{\delta_n^*(0)}{n} - \frac{1}{2\pi} \int \frac{|\phi_f(t)|^4}{1+(n-1)|\phi_f(t)|^2} dt$$

ou

$$(4.5) \quad J^*(\hat{f}_n(x)) = \frac{\delta_n^*(0)}{n} - o(1/n)$$

desde que o segundo termo em (4.4) seja dominado por

$$\frac{1}{2\pi(n-1)} \int |\phi_f(t)|^2 dt,$$

onde δ_n^* é o δ_n ótimo, que é encontrado invertendo-se a expressão (4.3). Como podemos ver o mínimo da I.E.Q.M. J^* cor

respondendo a δ_n^* depende de ϕ_f .

EXEMPLO 4.1 -

1) Distribuição de Cauchy,

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad \phi_f(t) = c^{-(t)}.$$

$$\delta_n^*(x) = \frac{n e^{-\pi x/2}}{(n-1)(1-e^{-\pi x})} \operatorname{sen} \frac{x \lg(n-1)}{2} + \frac{2n}{\pi} \sum_{r=1}^{\infty} \left(\frac{-1}{n-1}\right)^{r+1} \frac{r}{x^2+r^2}.$$

De (4.5) segue-se que $(n/\lg n)J^* \rightarrow (1/2\pi)$.

2) Distribuição dada por

$$f(x) = \frac{1-\cos x}{\pi x^2}, \quad \text{com } \phi_f(t) = \begin{cases} 1-|t| & \text{se } |t| \leq 1 \\ e & \\ 0 & \text{se } |t| > 1 \end{cases}$$

$$\delta_n^*(x) = \frac{n}{(n-1)\pi} \left[\frac{\operatorname{sen} x}{x} - \int_0^1 \frac{\cos xt}{1+(n-1)(1-t)^2} dt \right]$$

Assim, de (4.5) temos $nJ^* \rightarrow 1/\pi$. □

DEFINIÇÃO 4.1 - Diz-se que a função característica $\phi_f(t)$ de cresce algebricamente com grau $p > 0$ se

$$\lim_{|t| \rightarrow \infty} |t|^p |\phi_f(t)| = c > 0.$$

Um estimador da forma (4.1) é dito ter forma algébrica se $\delta_n(x) = b^{-1}(n)k(b^{-1}(n)x)$. Se sabemos que $\phi_f(t)$ tem decrescimento algébrico com grau p , um estimador da forma

algébrica $\hat{f}_n(x)$ pode ser encontrado cuja I.E.Q.M. $J(\hat{f}_n(x))$, satisfaz

$$(4.6) \quad \lim_{n \rightarrow \infty} \frac{J(\hat{f}_n(x))}{J^*(f_n(x))} = 1,$$

onde $J^*(\hat{f}_n(x))$ é a I.E.Q.M. correspondente a $\delta_n^*(x)$. Isto significa que se a eficiência assintótica do estimador é definida pelo lado esquerdo da expressão (4.6), então o estimador tem eficiência assintótica 1. O estimador é o do tipo algébrico com

$$\phi_{\delta_n}(t) = (1+|t|^{2p})^{-1}, \quad e \quad b(n) = cn^{-1/2p}$$

DEFINIÇÃO 4.2 - Diz-se que a função característica ϕ decresce exponencialmente com coeficiente $\rho > 0$ se

$$|\phi_f(t)| \leq Ae^{-\rho|t|}$$

para alguma constante a e todo t e se

$$\lim_{v \rightarrow \infty} \int_0^1 \left[1 + e^{2\rho v} |\phi_f(vt)|^2 \right]^{-2} dt = 0$$

Um estimador (4.1) tem a forma exponencial se a função característica de δ_n é dada por

$$(4.7) \quad \phi_{\delta_n}(t) = h(A_n e^{-a|t|})$$

onde $A_n \rightarrow 0$ para $n \rightarrow \infty$, $a > 0$ e h é limitada e de quadrado integrável. Aqui também $\delta_n(x-y) = k_n(x,y)$. Como no caso algébrico, um estimador da forma exponencial pode ser encontrado para o qual (4.6) é verdadeira. O estimador do tipo ex-

ponencial, com $|1-h(t)| \leq B|t|$ para $|t| \leq 1$ (B é uma constante), $A_n = cn^{-b}$ ($b \leq 1/2$) e $a = 2\rho b$ é um estimador com a propriedade assintótica ótima (4.6).

Estimadores de uma função de densidade que aparecem em diferentes contextos são os estimadores de máxima verossimilhança encontrados principalmente em Robertson [13] e Wegan [20],[21]. Em geral, estimadores de máxima verossimilhança para uma função de densidade não existem mas se algumas restrições apropriadas são colocadas na classe das densidades, dos quais o estimador pode ser selecionado, então um estimador de máxima verossimilhança sobre essa classe pode existir. O tipo de restrição considerada aqui é que os estimadores são mensuráveis com respeito a uma " σ -lattice". Seja $(\Omega, \mathcal{A}, \mu)$ um espaço mensurável, no qual a densidade está definida e é totalmente finita. Uma " σ -lattice" $L(L \subset \mathcal{A})$ de subconjuntos de Ω é por definição uma classe fechada sobre a formação de uniões enumeráveis e intersecções, contendo Ω e o conjunto vazio. A função f é L -mensurável se o conjunto $\{f > a\}$ está em L , para todo número real a . Robertson [13], usou o critério de máxima verossimilhança para escolher um estimador, ou seja procurou encontrar uma densidade \hat{f} L -mensurável, tal que o produto dos valores de \hat{f} pelos pontos observados é pelo menos tão grande quanto o produto dos valores de qualquer outra densidade L -mensurável por aqueles pontos. Um caso interessante considerado por Robertson [13] é o caso de uma densidade unimodal, com moda conhecida. Uma densidade unimodal com moda conhecida pode ser caracterizada como mensurável com respeito a uma σ -lattice L , de intervalos contendo a moda (isso é equivalente a: f é não decrescente para $x < M$ e f é não crescente para $x > M$, onde M é a moda). O estimador de Robertson tem a forma

$$(4.8) \quad \hat{f}_n(x) = \sum_{j=1}^k \frac{n_j}{n} \frac{I_{A_j}(x)}{\lambda(A_j)}$$

onde A_j , $j=1, \dots, k$ é um intervalo em L e x_1, \dots, x_n é o particular conjunto de observações. A função I_{A_j} é o indicador de A_j e n_j é o número de observações em A_j , $j=1, 2, \dots, k$. Aqui $\lambda(A_j)$ é a medida de Lebesgue do intervalo A_j , $j=1, \dots, k$.

O exemplo que consideramos a seguir, encontra-se em Wegan [21]. Observaremos que uma característica peculiar do estimador de Robertson é um "pico" próximo da moda.

EXEMPLO 4.1 - A figura abaixo é baseada em uma amostra de tamanho noventa, obtida pelo método de Monte Carlo, de uma densidade triangular. Este é o estimador de máxima verossimilhança proposto por Robertson [13], com moda conhecida igual a $1/2$. Aqui, a densidade é dada por

$$f(x) = \begin{cases} 4x & 0 \leq x < 1/2 \\ -4x+4 & 1/2 \leq x \leq 1, \end{cases}$$

e a reta foi dividida em 19 intervalos (em (4.8), $k=19$).

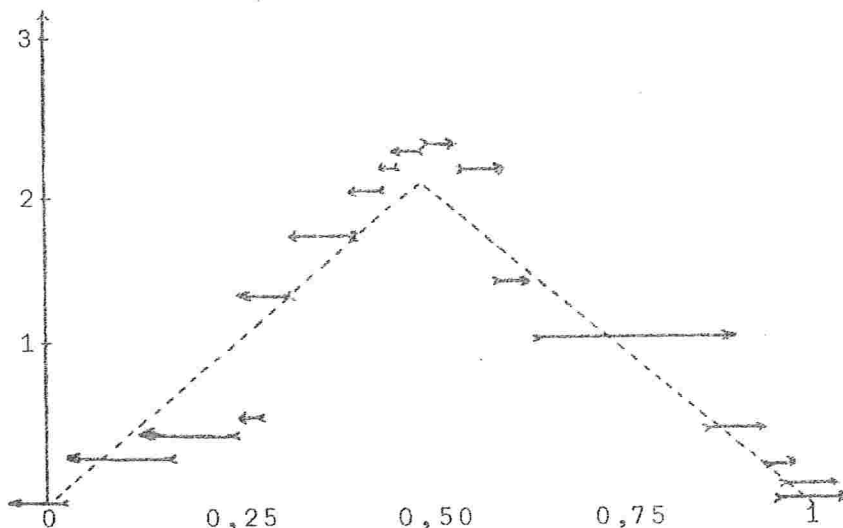


Fig.1 - Densidade triangular (linha interrompida) junto com estimativa de máxima verossimilhança (linha sólida) quando a moda é conhecida e igual a 0,5.

TABELA 1

| INTERVALO I | VALOR DO ESTIMADOR EM I |
|--------------------|-------------------------|
| $[-\infty, 0,011)$ | 0,0 |
| $[0,011, 0,135)$ | 0,360 |
| $[0,135, 0,248)$ | 0,490 |
| $[0,248, 0,264)$ | 0,674 |
| $[0,264, 0,345)$ | 1,37 |
| $[0,375, 0,414)$ | 1,59 |
| $[0,414, 0,432)$ | 2,04 |
| $[0,432, 0,448)$ | 2,17 |
| $[0,448, 0,493)$ | 2,34 |
| $[0,493, 0,500)$ | 4,02 |
| $[0,500, 0,505]$ | 4,54 |
| $(0,505, 0,510]$ | 2,26 |
| $(0,510, 0,553]$ | 1,74 |
| $(0,553, 0,583]$ | 1,48 |
| $(0,583, 0,834]$ | 0,973 |
| $(0,834, 0,890]$ | 0,605 |
| $(0,890, 0,915]$ | 0,427 |
| $(0,915, 0,977]$ | 0,357 |
| $(0,977, \infty]$ | 0,0 |

Robertson mostrou que se f é contínua em x , então

$$\lim_n \hat{f}_n(x) = f(x)$$

com probabilidade um e, sob a suposição da continuidade de f , mostrou também a consistência quase uniforme com probabilidade um. Uma falha particular desse estimador, que pressupõe o conhecimento da moda, está no mau comportamento do estimador próximo da moda, como podemos observar na figura acima. Wegan [21] procurou melhorar o comportamento do estimador próximo da moda. Ele obteve então, estimadores de máxima verossimilhança quando a moda é desconhecida. Sendo $[L,R]$ o intervalo mo

dal, ele provou que ao se determinar o estimador de máxima verossimilhança com moda central desconhecida, temos que as sumir que um dos pontos finais do intervalo $[L,R]$ pertence ao conjunto $\{y_1, y_2, \dots, y_n\}$, que é o conjunto das observações ordenadas. Temos então $2n$ intervalos da forma $[L,R]$ onde L ou R pertencem a $\{y_1, y_2, \dots, y_n\}$, dos quais somente um é selecionado, aquele mais provável. Para maiores detalhes sobre estes estimadores, o leitor poderá consultar os artigos citados, pois nosso objetivo aqui foi, numa discussão geral apresentar as condições bastante especiais em que funcionam

Como um método bastante recente que não foi abordado neste trabalho, podemos citar a estimação de funções de densidade por "spline functions" (ver por exemplo, Schoenberg, I.J. (1968) - Monosplines and quadrature formulae. - Theory and Application of Spline Functions (T.N.E.Greville, ed.) Academic, New York). Pretendemos num próximo trabalho discutir particularmente este método de estimação.

APÊNDICE 1

Vamos tratar, agora, alguns resultados básicos da teoria das funções ortogonais que utilizamos frequentemente no desenvolvimento do capítulo 3. A maioria das demonstrações dos resultados que apresentaremos poderão ser encontradas em Sansone [16].

DEFINIÇÃO A.1.1 - Uma função ϕ de quadrado integrável é dita normal, se

$$\int_{-\infty}^{\infty} \phi^2(t) dt = 1.$$

DEFINIÇÃO A.1.2 - Um sistema $\{\phi_k(x)\}$ de funções é dito ortogonal com respeito a função peso $\omega(x)$, se

$$\int_{-\infty}^{\infty} \phi_k(x) \phi_j(x) \omega(x) dx = \delta_{kj}(x)$$

onde

$$\delta_{kj} = \begin{cases} 1 & \text{se } k = j \\ 0 & \text{se } k \neq j \end{cases}$$

No caso em que são válidas simultaneamente as condições das duas definições para o sistema $\{\phi_k(x)\}$, dizemos

que $\{\phi_k(x)\}$ é um sistema ortonormal.

Outro problema importante a ser considerado, consiste em determinar uma combinação linear

$$c_1 f_1 + \dots + c_n f_n$$

de n funções f_1, f_2, \dots, f_n todas de quadrado integrável e ortonormais que estão a uma distância mínima de uma outra função F , também de quadrado integrável. Nosso problema consiste em determinar as constantes c_1, c_2, \dots, c_n de modo que

$$(A.1.1) \quad \int_G (F - c_1 f_1 - c_2 f_2 - \dots - c_n f_n)^2 dt$$

assume seu valor mínimo. Temos então

$$\int_G (F - \sum c_i f_i)^2 dt = \int_G F^2 dt + \int_G \left(\sum_{i=1}^n c_i f_i \right)^2 dt - 2 \sum_{i=1}^n c_i \int_G F f_i dt$$

Como f_1, f_2, \dots, f_n é um sistema ortonormal, temos

$$\int_G (F - \sum c_i f_i)^2 dt = \int_G F^2 dt + \sum_{i=1}^n c_i^2 - 2 \sum_{i=1}^n c_i \int_G F f_i dt.$$

Igualando o zero as derivadas parciais, com respeito a c_1, c_2, \dots, c_n encontramos

$$(A.1.2) \quad c_i = \int_G F f_i dt \quad (i=1, 2, \dots, n).$$

Reciprocamente, substituindo (A.1.2) na integral (A.1.1), temos:

$$\begin{aligned} \int_G (F - \sum c_i f_i)^2 dt &= \int_G F^2 dt + \sum_{i=1}^n c_i^2 - 2 \sum_{i=1}^n c_i \int_G F f_i dt = \\ &= \int_G F^2 dt + \sum_{i=1}^n \left(c_i - \int_G F f_i dt \right)^2 - \sum_{i=1}^n \left(\int_G F f_i dt \right)^2 \end{aligned}$$

O único termo que depende de c_i no segundo membro é

$$\sum \left(c_i - \int_G F f_i dt \right)^2$$

que assume o valor zero quando os coeficientes c_i são dados por (A.1.2). Provamos então que

TEOREMA A.1.1 - Se F é de quadrado integrável em G , e f_1, \dots, f_n são ortogonais em G , a combinação linear $\sum c_i f_i$, na qual os c_i , $i=1, \dots, n$ são os coeficientes de Fourier de F , é a combinação linear de f_1, f_2, \dots, f_n que é a melhor aproximação em média para F de todas as possíveis combinações lineares de f_1, \dots, f_n .

As constantes c_i , $i=1, \dots, n$ dadas por (A.1.2) são chamadas os coeficientes de Fourier de F com respeito a f_1, \dots, f_n .

TEOREMA A.1.2 - Se $\{f_n\}$ é uma sequência de funções ortogonais em G , F é de quadrado integrável em G , $\{c_n\}$ é a sequência de Fourier de F com respeito a sequência $\{f_n\}$ então a série

$$\sum_{n=1}^{\infty} c_n^2$$

é convergente e, mais ainda, as constantes c_i , $i=1,2,\dots$ satisfazem a desigualdade de Bessel

$$\sum_{n=1}^{\infty} c_n^2 \leq \int_G F^2 dt$$

TEOREMA A.1.3 (Fischer-Riesz) - Se G é um conjunto de medida finita, $\{\phi_n(x)\}$ é uma sequência de funções ortogonais em G , $\{a_n\}$ é uma sequência de constantes para as quais a série $\sum a_n^2$ é convergente e finalmente, se nós escrevemos

$$f_n = a_1 \phi_1 + \dots + a_n \phi_n,$$

então a sequência $\{f_n\}$ converge em média em G para uma função f , a qual tem por coeficientes de Fourier os termos da sequência $\{a_n\}$, ou seja

$$a_i = \int_G f \phi_i dt, \quad i=1,2,\dots$$

DEFINIÇÃO A.1.3 - Seja G um conjunto mensurável de pontos na reta, $p > 0$ e $|f(t)|^p$ integrável em G . Então, $f(t)$ é chamada integrável L^p em G , e nos escrevemos $f(t) \in L^p$ em G .

DEFINIÇÃO A.1.4 - Seja $\{f_n\}$ uma sequência de funções integráveis L^p em um conjunto mensurável G e $f(t)$ integrável em G . Então $\{f_n\}$ é dita convergir em média de ordem p em G para $f(t)$, se

$$\lim_{n \rightarrow \infty} \int_G |f(t) - f_n(t)|^p dt = 0$$

Se $p = 2$ na definição acima, dizemos simplesmente que $\{f_n(t)\}$ converge em média para $f(t)$.

DEFINIÇÃO A.1.5 - Um sistema de funções ortogonais em G é chamado completo em G , quando a única função de quadrado integrável a qual é ortogonal em G a todas as funções do sistema é a função nula q.c.

TEOREMA A.1.4 - Um sistema completo em um conjunto G de medida finita consiste de uma quantidade enumerável de funções.

DEFINIÇÃO A.1.6 - A série $a_1\phi_1 + a_2\phi_2 + \dots + a_n\phi_n$ é chamada convergente em média para f em G se

$$\lim_{n \rightarrow \infty} \int_G \left(f - \sum_{i=1}^n a_i \phi_i \right)^2 dt = 0$$

TEOREMA A.1.5 - Se $\{f_n\}$ é um sistema completo de funções ortogonais em um conjunto G de medida finita e se f é qualquer função de quadrado integrável em G , então a sequência $\{f_n\}$, onde

$$f_n = a_1\phi_1 + a_2\phi_2 + \dots + a_n\phi_n$$

com

$$a_i = \int f\phi_i dt, \quad i=1,2,\dots$$

converge em média para f , isto é

$$\lim_{n \rightarrow \infty} \int_G |f - f_n|^2 dt = 0$$

e conseqüentemente

$$\sum_{i=1}^{\infty} a_i^2 = \int_G f^2 dt$$

Notemos que convergência em média da série

$$\sum_{n=1}^{\infty} a_n \phi_n$$

para a função f não implica em convergência pontual da série, pois se t é um ponto em G , não segue necessariamente que

$$f(t) = a_1 \phi_1(t) + a_2 \phi_2(t) + \dots + a_n \phi_n(t) + \dots$$

TEOREMA A.1.6 - Uma condição necessária e suficiente para que a sequência ortonormal $\{\phi_n\}$ em um conjunto G de medida finita forme um sistema completo é que, para toda função f de quadrado integrável em G ,

$$\int_G f^2 dt = \sum_{n=1}^{\infty} \left(\int_G \phi_n f dt \right)^2$$

Os resultados que apresentaremos a seguir dizem respeito a secção 3 do capítulo 3, onde discutimos o uso de sistemas ortogonais trigonométricos.

DEFINIÇÃO A.1.7 - Seja f uma função integrável de período 2π . Então as constantes dadas pelas expressões

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx$$

$k=1, 2, \dots$, são chamados constantes de Fourier de $f(x)$ e a sé

rie

$$\frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

é chamada série de Fourier de $f(x)$.

TEOREMA A.1.7 - Seja $f(x)$ integrável em $(-\pi, \pi)$. Se $f(x)$ satisfaz a condição $f(x) = f(-x)$, ou seja $f(x)$ é uma função par, então sua série de Fourier se reduz a série dos cosenos. Se $f(x)$ é tal que $f(x) = -f(-x)$, ou seja $f(x)$ é uma função ímpar, então sua série de Fourier se reduz a série dos senos.

Notemos que os estimadores (3.3.19) e (3.3.20) são bastante parecidos com a expansão de Fourier dos cosenos e dos senos respectivamente, de uma função com $a_k = \bar{c}_k$ em (3.3.19) e $a_k = \bar{s}_k$ em (3.3.20).

TEOREMA A.1.8 - Os sistemas de funções

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos kx}{\sqrt{\pi}}, \frac{\sin kx}{\sqrt{\pi}}, k=1, 2, \dots$$

são completos em $[0, 2\pi]$, isto é se $\theta(x)$ é de quadrado integrável em $[0, 2\pi]$, de modo que

$$\int_0^{2\pi} \theta(x) dx = 0; \int_0^{2\pi} \theta(x) \cos kx dx = 0; \int_0^{2\pi} \theta(x) \sin kx dx = 0, \\ k=1, 2, \dots$$

então $\theta(x) = 0$ em quase toda parte em $[0, 2\pi]$.

Como já frisamos anteriormente, a demonstração de todos os resultados que apresentamos neste apêndice estão em Sansone [16].

APÊNDICE 2

No capítulo 1, para motivarmos o problema da estimação de uma função de densidade, apresentamos o método introduzido por Pearson, através do qual se procura uma expressão matemática para representar uma população. Neste apêndice discutiremos outra representação, que consiste basicamente em expressar a função de densidade por uma expressão em série. Nos referiremos a esta expansão como séries de Grand-Charlier do tipo A, cujos resultados constam de Kendall e Stuart [3]. Escrevendo a normal padronizada como

$$\alpha(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

denotando

$$D = \frac{d}{dx}$$

e considerando as derivadas sucessivas com respeito a x , obtemos

$$D\alpha(x) = -x\alpha(x)$$

$$D^2\alpha(x) = (x^2-1)\alpha(x)$$

$$D^3\alpha(x) = (3x-x^2)\alpha(x)$$

e assim por diante. O resultado é, em geral, um polinômio em x multiplicado por $\alpha(x)$. Podemos definir os polinômios de

Tchebychev-Hermite $H_r(x)$ pela identidade

$$(-D)^r \alpha(x) = H_r(x) \alpha(x)$$

Evidentemente H_r é de grau r em x e o coeficiente de x^r é a unidade. Por convenção $H_0 = 1$. Temos

$$\alpha(x-t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2 + tx - \frac{1}{2}t^2\right) = \alpha(x) \exp\left(tx - \frac{1}{2}t^2\right)$$

e então pelo toerema de Taylor

$$\alpha(x-t) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} t^j D^j \alpha(x) = \alpha(x) \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(x)$$

Consequentemente $H_r(x)$ é o coeficiente de $\frac{t^r}{r!}$ em $\exp\left(tx - \frac{1}{2}t^2\right)$, de onde segue-se

$$(A.2.1) \quad H_r(x) = x^r - \frac{r^{[2]}}{2 \cdot 1!} x^{r-2} + \frac{r^{[4]}}{2^2 \cdot 2!} x^{r-4} - \frac{r^{[6]}}{2^3 \cdot 3!} x^{r-6} + \dots$$

Aqui,

$$r^{[k]} = r(r-1)\dots(r-k+1)$$

Os primeiros 10 polinômios são:

$$\begin{aligned} H_0 &= 1 \\ H_1 &= x \\ H_2 &= x^2 - 1 \\ H_3 &= x^3 - 3x \\ H_4 &= x^4 - 6x^2 + 3 \\ H_5 &= x^5 - 10x^3 + 15x \\ H_6 &= x^6 - 15x^4 + 45x^2 - 15 \\ H_7 &= x^7 - 21x^5 + 105x^3 - 105x \end{aligned}$$

$$\begin{aligned}H_8 &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105 \\H_9 &= x^9 - 36x^7 + 378x^5 - 1.260x^3 + 945x \\H_{10} &= x^{10} - 45x^8 + 630x^6 - 3.150x^4 + 4.725x^2 - 945\end{aligned}$$

Esses polinômios têm uma propriedade ortogonal importante, ou seja

$$(A.2.2) \quad \int_{-\infty}^{\infty} H_m(x)H_n(x)\alpha(x)dx = 0, \quad m \neq n \\ = n!, \quad m=n.$$

Pode-se mostrar também que o sistema

$$\frac{e^{-x^2/2}H_r(x)}{2^r r! \sqrt{\pi}}$$

é ortonormal em $(-\infty, \infty)$. A prova desse resultado pode ser encontrada em Sansone [16], página 308.

Suponhamos agora que uma função de densidade pode ser expressada formalmente em uma série de derivadas de $\alpha(x)$. Mais tarde, discutiremos em que condições tal expansão é válida. Temos então

$$f(x) = \sum_{j=0}^{\infty} c_j H_j(x) \alpha(x).$$

Multiplicando ambos os lados da expressão acima por $H_r(x)$ e integrando de $-\infty$ a ∞ , temos em virtude de (A.2.2) que

$$(A.2.3) \quad c_r = \frac{1}{r!} \int_{-\infty}^{\infty} f(x)H_r(x)dx.$$

Substituindo em (A.2.3) os valores explícitos de $H_r(x)$ dados por (A.2.1), temos

$$c_r = \frac{1}{r!} \left\{ \mu_r - \frac{r^{[2]}}{2.1!} \mu_{r-2} + \frac{r^{[4]}}{2.2!} \mu_{r-4} - \dots \right\}$$

de onde segue-se que

$$(A.2.4) \quad \begin{cases} c_0 = 1 \\ c_1 = 0 \\ c_2 = \frac{1}{2}(\mu_2 - 1) \\ c_3 = \frac{1}{6}\mu_3 \\ c_4 = \frac{1}{24}(\mu_4 - 6\mu_2 + 3) \\ c_5 = \frac{1}{120}(\mu_5 - 10\mu_3) \\ c_6 = \frac{1}{720}(\mu_6 - 15\mu_4 + 45\mu_2 - 15) \end{cases}$$

e assim por diante. Temos então a expansão formal

$$(A.2.5) \quad f(x) = \alpha(x) \left\{ 1 + \frac{1}{2}(\mu_2 - 1)H_2 + \frac{1}{6}\mu_3H_3 + \frac{1}{24}(\mu_4 - 6\mu_2 + 3)H_4 + \dots \right\}$$

Se a densidade \tilde{e} é expressa em medida padronizada (mediante o uso da transformação $(x-\mu)/\sigma$), temos que

$$(A.2.6) \quad f(x) = \alpha(x) \left\{ 1 + \frac{1}{6}\mu_3H_3 + \frac{1}{24}(\mu_4 - 3)H_4 + \dots \right\}$$

As séries (A.2.5) e (A.2.6) são chamadas *séries de Gran-Charlier do tipo A*.

Em medida padronizada as relações (A.2.4) tornam-se em temos de cumulantes:

$$(A.2.7) \quad \begin{cases} c_0 = 1 \\ c_1 = c_2 = 0 \\ c_3 = k_3/6 \\ c_4 = k_4/24 \\ c_5 = k_5/120 \\ c_6 = \frac{1}{720}(k_6 + 10k_3^2) \end{cases}$$

onde

$$\begin{aligned} k_3 &= \mu_3 \\ k_4 &= \mu_4 - 3\mu_2^2 \\ k_5 &= \mu_5 - 10\mu_3\mu_2 \\ k_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \end{aligned}$$

Notemos que ao transformarmos uma distribuição com cumulantes k_r para a distribuição padronizada correspondente, os cumulantes da distribuição padronizada k_r^* são dados por

$$k_r^* = k^r / \sigma^r$$

onde σ é o desvio padrão da distribuição.

Em situações práticas, ao representarmos uma função de densidade por uma série do tipo A, vamos levar em conta somente alguns termos. Usualmente vamos representar a densidade desconhecida pela série

$$(A.2.8) \quad \alpha(x) \left(k + \frac{k_3}{6}H_3 + \frac{k_4}{24}H_4 + \frac{k_5}{120}H_5 + \frac{k_6 + 10k_3^2}{720}H_6 \right)$$

desde que os termos remanescentes sejam desprezíveis.

EXEMPLO A.2.1 - Utilizando novamente o problema dos feijões introduzido no exemplo (1.1) vamos encontrar uma expressão para $f(x)$ mediante o uso de (A.2.8). No exemplo 1.1 já tínhamos encontrado os primeiros quatro momentos. Em medida padronizada, para os primeiros seis momentos, encontramos:

$$\mu_3 = -0,910569$$

$$\mu_4 = 4,862944$$

$$\mu_5 = -12,574125$$

$$\mu_6 = 53,221083$$

Então a série é dada por

$$9.440\alpha(x) \{1 - 0,151762H_3 + 0,0776227H_4 - 0,0289036H_5 + 0,0142735H_6\}$$

A tabela seguinte dá as frequências tomando-se os primeiros 3, primeiros 4 e primeiros 5 termos na série acima respectivamente.

TABELA

| 3 TERMOS | 4 TERMOS | 5 TERMOS |
|-------------|-------------|-------------|
| 16,3 | -15,2 | 2,0 |
| 12,8 | 13,7 | -35,3 |
| 25,6 | 116,6 | 22,3 |
| 241,7 | 370,4 | 438,1 |
| 1.012,7 | 926,2 | 1.214,0 |
| 2.155,4 | 1.833,0 | 1.866,9 |
| 2.593,0 | 2.506,4 | 2.112,8 |
| 1.788,4 | 2.082,6 | 1.916,7 |
| 713,4 | 921,6 | 1.183,4 |
| 280,7 | 199,0 | 371,2 |
| 258,7 | 132,1 | 66,9 |
| 206,2 | 178,1 | 101,2 |
| 98,7 | 117,0 | 107,1 |
| 29,6 | 43,5 | 54,0 |
| 5,9 | 10,0 | 15,4 |
| { 0,9 | { 1,7 | { 3,3 |
| - | - | - |
| 9.440 | 9.440 | 9.440 |

Olhando a tabela vemos que a série com 3 termos comporta-se de modo mais razoável do que as séries com 4 e 5 termos, visto que estas últimas apresentam valores negativos. \square

Os dois resultados abaixo estabelecem as condições em que existe uma expansão em série (convergente) do tipo A da densidade $f(x)$.

- 1) Kendall e Stuart [3], pg.161. Se $f(x)$ é uma função a qual tem derivada contínua tal que

$$\int_{-\infty}^{\infty} \left(\frac{df(x)}{dx} \right) e^{x^2/2} dx$$

converge, e se $f(x)$ tende a zero para $|x|$ tendendo ao infinito, então $f(x)$ pode ser desenvolvida na série

$$(A.2.9) \quad f(x) = \sum_{j=0}^{\infty} \frac{c_j}{j!} D^j \alpha(x)$$

onde c_j é dado por

$$c_j = \int_{-\infty}^{\infty} f(x) H_j(x) dx$$

Esta série é absolutamente e uniformemente convergente para $-\infty \leq x \leq \infty$.

- 2) Kendall e Stuart [3], pg.161. Se $f(x)$ é de variação limitada em todo intervalo finito e se

$$\int_{-\infty}^{\infty} |f(x)| e^{x^2/4} dx$$

existe, então a expansão de $f(x)$ na série (A.2.9) converge em toda parte para a soma $\frac{1}{2}\{f(x+0) + f(x-0)\}$. A convergência é uniforme em todo intervalo de continuidade finito.

REFERÊNCIAS

- [1] - CRAMÉR, H. (1946) - Mathematical Methods of Statistics, Princeton Univ. Press.
- [2] - EPANECHNIKOV, V.A. (1969) - "Non-parametric estimation of a multivariate probability density", Theor. Prob. Appl., 14, pp.153-158.
- [3] - KENDALL, M.G. and STUART, A. (1958) - The Advanced Theory of Statistics 1, Charles Griffin and Co.Ltd., London.
- [4] - KRONMAL, R. and TARTER, M. (1968) - "The estimation of probability densities and cumulatives by Fourier Series methods", J. Amer. Statist. Assoc., 63, pp. 925-952.
- [5] - LEHMANN, E. (1950) - Notes on the theory of estimation, University of California, Berkeley.
- [6] - LOEVE, M. (1962) - Probability Theory, D.Van Nostrand Co.Inc., N. York.
- [7] - MURTHY, V.K. (1965) - "Estimation of probability density", Ann. Math. Statist., 36, pp.1.027-1.031.
- [8] - PARZEN, E. (1962) - "On the estimation of a probability density function and the mode", Ann. Math. Statist., 33, pp.1065-1076.
- [9] - PEARSON, K. (1902) - "On the systematic fitting of curves to observations and measurements, I", Biometrika 1, pp.265-303.
- [10] - PEARSON, K. (1902) - "On the systematic fitting of curves to observations and measurements, II", Biometrika 2, pp.1-23.
- [11] - PICKANDS, J. (1969) - "Efficient estimation of a probability density function", Ann. Math. Statist., 40, pp. 854-864.

- [12] - QUESENBERRY, C.P. and SCHEULT, A.H. (1971) - "On unbiased estimation of density functions", Ann. Math. Statist., 42, pp.1434-1438.
- [13] - ROBERTSON, T. (1967) "On estimating a density which is measurable with respect to a σ -lattice", Ann. Math. Statist., 42, pp.1815-1842.
- [14] - ROSENBLATT, M. (1956) - "Remarks on some nonparametric estimation of a density function", Ann. Math. Statist., 27, pp.832-837.
- [15] - ROSENBLATT, M. (1971) - "Curve estimates", Ann. Math. Statist., 42, pp. 1.815-1.842.
- [16] - SANSONE, G. (1959) - "Ortogonal Functions", Interscience Publishers Ltd., London.
- [17] - SCHWARTZ, S.C. (1967) - "Estimation of a probability density by an orthogonal series", Ann. Math. Statist., 38, pp.1.261-1.265.
- [18] - WATSON, G.S. and LEADBETTER, M.R. (1963) - "On estimating a probability density", Ann. Math. Statist., 34, pp.480-491.
- [19] - WATSON, G.S. (1969) - "Density estimation by orthogonal series", Ann. Math. Statist., 40, pp.1.496-1.498.
- [20] - WEGMAN, E.J. (1969) - "A note on estimating a unimodal density", Ann. Math. Statist., 40, pp.1.661-1.667.
- [21] - WEGMAN, E.J. (1970) - "Maximum likelihood estimation of a unimodal density function", Ann. Math. Statist., 41, pp.457-471.

- □ -

IMPRESSO NO SETOR GRÁFICO
DO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO

- o -