

Estimador kernel da função de densidade

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná
Curitiba/PR, 81531-990, Brasil

email: lucambio@ufpr.br

Agosto de 2008

1 Introdução

Desde 1890 diferentes formas de estimar uma função de densidade têm sido propostas. Uma destas é devida a Karl Pearson (1857-1936) entre 1890 e 1900 e é obtida como solução de uma equação diferencial Johnson & Kotz (1988).

A partir de 1956 os métodos de estimação de funções de densidade de probabilidade não paramétricos têm-se consolidado como uma alternativa sofisticada ao tratamento tradicional de estudar conjuntos de dados. Esta alternativa baseia-se na possibilidade de analisar os dados sem assumir um comportamento distribucional específico.

Sobre o problema de estimação de funções de densidade de probabilidade trata a seção 2. Descrevemos também de maneira resumida algumas das propostas para obter estes estimadores e definimos propriedades estatísticas.

Depois, a seção 3, dedica-se ao estudo da proposta mais importante dentre os estimadores da função de densidade, o chamado estimador kernel ou de Rosenblatt-Parzen. As primeiras idéias deste estimador devem-se a Rosenblatt (1956), posteriormente generalizadas por Parzen (1962).

2 Estimadores da função de densidade

A função de densidade é um conceito fundamental em Estatística. Define-se como uma função real mensurável e não-negativa satisfazendo

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

A necessidade de utilizar estimadores da função de densidade aparece em situações tais como:

1. *Análise Exploratória*, onde descrevem-se aspectos como multimodalidade, cauda, assimetria. Isto devido ao gráfico da densidade estimada resumir convenientemente a informação relativa à forma da distribuição da amostra.
2. *Análise Confirmatória*, para a tomada de decisões através de diferentes métodos, como análise discriminante não paramétrica, análise de clusters, testes para a moda, etc.

Histograma do tempo de duração das erupções

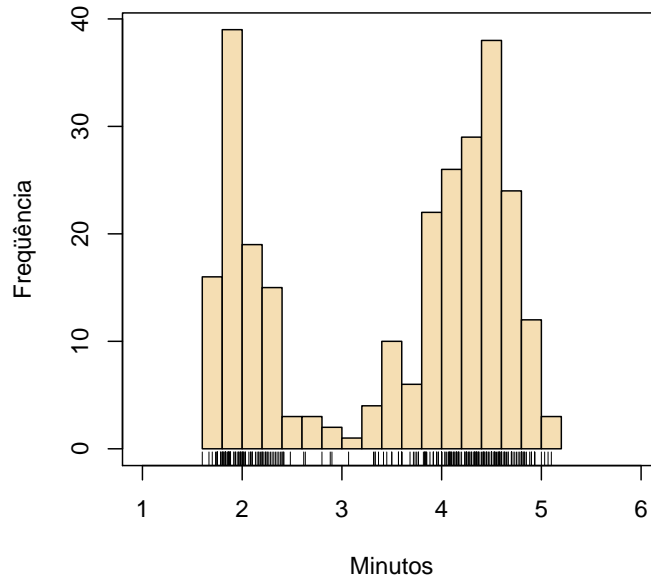


Figura 1: Histograma do tempo de duração das erupções do geysir Old Faithful, no parque nacional Yellowstone, Wyoming, USA. Número de intervalos igual a 20.

Para exemplificar o estimador mais simples de uma função de densidade de probabilidade, chamado de histograma, utilizaremos o tempo de duração das erupções do geysir Old Faithful, no parque nacional Yellowstone, Wyoming, USA. Existem diversas versões destes dados, em Azzalini & Bowman (1990) utiliza-se uma versão mais completa desses dados. Na Figura 1 podemos observar o comportamento destes dados.

Robertson (1967) provou que, dados os intervalos I_1, I_2, \dots, I_k , o histograma \hat{f} é um estimador de máxima verossimilhança dentre os estimadores expressados como funções simples e semi-contínuas superiormente, isto se o fecho de cada intervalo contém duas ou mais observações, veja também Wegman (1975).

Este estimador é calculado segundo a expressão:

$$\hat{f} = \frac{n_i}{n|I_i|},$$

para todo $x \in I_i$ e para cada $i = 1, \dots, k$. Na expressão anterior, $|I_i|$ representa o comprimento do intervalo i e n_i o número de elementos na amostra em I_i . A Figura 1 mostra a forma do histograma segundo a proposta de Robertson (1967).

Pode-se observar que este estimador tem duas limitações importantes: a dependência do comprimento do intervalo e o fato de o histograma não constituir uma função contínua. A primeira destas limitações foi amplamente estudada por Wegman (1975). Ele provou que os pontos extremos de cada intervalo I_i devem ser coincidentes com observações e que, se o número mínimo de observações em cada intervalo m aumenta, conforme aumenta o tamanho da amostra, o estimador \hat{f} é consistente.

A segunda limitação importante do histograma, isto é, o fato de ele não constituir uma função contínua, incentivou diversos estudos na procura de estimadores contínuos como, por exemplo, Parzen (1962) e Grenander (1981) dentre outros.

2.1 Alguns métodos de estimação

Um método amplamente utilizado em estatística para obter estimadores é o método de máxima verossimilhança. Na situação específica da estimação de funções de densidade de probabilidade este método é descrito da seguinte forma.

Suponhamos que X_1, X_2, \dots, X_n seja uma amostra aleatória da variável aleatória X com função de densidade contínua f_0 , desconhecida. Definimos a verossimilhança de uma densidade qualquer f como

$$L(f) = \prod_{i=1}^n f(X_i).$$

O problema de otimização que se tenta solucionar é

$$\text{maximizar}_{f \in \mathfrak{S}} L(f),$$

onde \mathfrak{S} é o conjunto de todas as funções reais mensuráveis tais que

$$\int_{\mathbb{R}} f(t) dt = 1 \quad \text{e} \quad f \geq 0, \text{ para todo } t \in \mathbb{R}.$$

Este problema não tem solução. Por exemplo, a função *delta de Dirac*, definida pelo físico teórico Paul Dirac (Paul Adrien Maurice Dirac, 1902 - 1984). Define-se como $\delta(x)$ satisfazendo ser iguala zero zero em quase toda a reta real exceto em $x = 0$ onde assume infinito de maneira que a integral total seja 1. É a análoga contínua da função discreta conhecida como *delta de Kronecker*.

Significa que que $\delta \in \mathfrak{S}$ e, portanto, $L(f) \rightarrow \infty$. Duas propostas surgiram na literatura: uma forma de conseguir que a função de verossimilhança $L(f)$ tenha máximo é penalizá-la. Para isso define-se um funcional integrável $\Phi : H \rightarrow \mathbb{R}$ e dada uma amostra X_1, \dots, X_n verossimilhança penalizada por Φ de $f \in H$ é definida como

$$L_{\Phi}(f) = \prod_{i=1}^n f(X_i) \exp(-\Phi(f)).$$

Dependen da escolha de H e da função Φ , pode ser gerada uma grande variedade destes estimadores. Outra forma de maximizar $L(f)$ foi proposta por Grenander (1981) (veja Lucambio (1998)). A idéia nesta nova situação é restringir o conjunto H , no qual se faz a maximização, da seguinte forma: para cada $m > 0$, escolhe-se subconjunto $S_m \subset H$ onde o estimador de máxima verossimilhança da densidade f exista. Fazendo estes subconjuntos S_m crescerem com o tamanho da amostra, consegue-se que o estimador achado em cada S_m convirja à função de densidade que originou a amostra. Os estimadores assim achados são chamados de *sieves* ou de Grenander.

2.2 Propriedades assintóticas

Fryer (1976), por exemplo, estuda os estimadores da função de densidade mantendo finito o tamanho da amostra em situações específicas. No entanto, na maioria dos trabalhos as propriedades estatísticas destes estimadores são provadas assintoticamente no tamanho de amostra.

De maneira natural define-se o estimador \hat{f} como não tendencioso ou não viciado se

$$E_f\{\hat{f}(x)\} = f(x),$$

para todo x real. No entanto, esta definição não tem utilidade na estimação de densidades pois não existe estimador não viciado para toda função de densidade (Rosenblatt, 1956). Esta propriedade importante de todo estimador define-se então da seguinte forma.

Definição 1 *A sequência de estimadores $\{\hat{f}_n\}$, da função de densidade f , se diz assintoticamente não tendenciosa se, e somente se,*

$$\lim_{n \rightarrow \infty} E_f\{\hat{f}_n(x)\} = f(x),$$

para todo x real.

Em geral, juntamente com a propriedade de não tendenciosidade estuda-se a sua consistência. Dependendo da forma de convergência escolhida, observam-se diferentes formas de consistência \hat{f} . Se $\hat{f}(x) \rightarrow f(x)$ em probabilidade para todo x real, afirm-se que \hat{f} é fracamente consistente. Se a convergência acontece quase certamente, \hat{f} é fortemente consistente. Outros tipos de convergência dependem do que se entenda por critério de erro.

Por exemplo, se a função f é suposta de quadrado integrável o ajuste obtido por \hat{f} é medido segundo o erro quadrado médio (EQM). Se o $\text{EQM}\{\hat{f}(x)\} \rightarrow 0$ para todo x real quando $n \rightarrow \infty$, diz-se que \hat{f} é um estimador consistente fracamente de f em média quadrática.

Outros critérios importantes medem como a função \hat{f} , como um todo, estima f . Um deles é o erro quadrático médio integral (EQMI).

Definição 2 *Define-se o erro quadrático médio integral (EQMI) como*

$$\text{EQMI}\{\hat{f}\} = \int_{-\infty}^{\infty} \text{EQM}\{\hat{f}(x)\} dx.$$

Um problema com estas formas de medir o ajuste obtido com \hat{f} é que o estudo das caudas das densidade têm pouca importância. Estas e outras objeções ao EQM, EQMI e outras medidas de ajuste encontram-se em Donoho & Johnstone (1989).

Numa tentativa de avaliar globalmente o ajuste do estimador \hat{f} , diminuindo as limitações das medidas de ajuste mencionadas, Devroye & Györfi (1985) e Devroye (1987) propuseram o erro absoluto integral (EAI).

Definição 3 *O estimador \hat{f} diz-se globalmente bem ajustado se o erro absoluto integral, definido como*

$$\text{EAI}\{\hat{f}\} = \int_{-\infty}^{\infty} |\hat{f}(x) - f(x)| dx,$$

assumir valores pequenos.

Esta medida possui como propriedades a invariância a transformações monótonas e ser limitada em $0 \leq \text{EAI} \leq 2$, além de ser útil na definição de consistência do estimador.

Se a convergência a zero do erro absoluto integral é quase certa, \hat{f} diz-se fortemente consistente. Esta forma de medir a consistência está relacionada com a distância de Kullback-Liebler, definida em Kullback & Leibler (1951). Tem-se observado que a obtenção de resultados de consistência segundo a definição do EAI é muito mais complicada do que aqueles esforços necessários para obter resultados semelhantes segundo o EQMI e suas modificações.

3 Estimador kernel (núcleo)

A partir dos trabalhos de Rosenblatt (1956) e Parzen (1962), o estimador kernel ou núcleo tem sido bastante estudado. Veja, por exemplo, os trabalhos de Silverman (1986) e Wand & Jones (1995). Fatores que contribuíram para esta ampla utilização são a simplicidade e as boas propriedades e desde então conhecido como estimadores Rosenblatt-Parzen, também chamado estimador núcleo edenotado por \hat{f}

Seja X_1, \dots, X_n uma amostra aleatória de uma variável aleatória contínua com função de densidade f e função de distribuição F . Um possível estimador do valor $F(x)$ é dado pela função de distribuição empírica, definida como

$$\hat{F}_n(x) = \frac{1}{n} \{\text{número de observações na amostra} \leq x\} = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}.$$

Na definição acima $1\{X_i \leq x\}$ representa a função indicadora. Vejamos algumas propriedades importantes destes estimadores. Note que $n\hat{F}_n(x) \sim B(n, F(x))$. Com isso, $E\{n\hat{F}_n(x)\} = nF(x)$, logo $E\{\hat{F}_n(x)\} = F(x)$, também $\text{Var}\{\hat{F}_n(x)\} = F(x)(1 - F(x))/n$. Portanto $\hat{F}_n(x)$ é um estimador não tendencioso de $F(x)$, pode-se provar também este estimador é consistente e converge em probabilidade a $F(x)$ e podemos concluir que este é um bom estimador de $F(x)$.

A derivada de $\hat{F}_n(x)$ seria um estimador lógico de $f(x)$, no entanto, $\hat{F}_n(x)$ não é diferenciável, pois sempre que x assume um dos alores X_i , a função $\hat{F}_n(x)$ tem salto igual a $1/n$. Rosenblatt (1956) propôs utilizar como estimador de $f(x)$ uma aproximação limite que define um estimador da função de densidade em termos de $\hat{F}_n(x)$, isto é, dado que

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h},$$

Rosenblatt propôs como estimador de $f(x)$ a função

$$\tilde{f}(x) = \lim_{h \rightarrow 0} \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h}, \quad (1)$$

para h um número real positivo pequeno, que será chamado de parâmetro de alisamento e n grande e o chamaremos de estimador de Rosenblatt da função de densidade. Na Figura 2 mostramos a forma deste estimador segundo diversos valores de h , neste gráfico mostra-se a dependência deste estimador do parâmetro de alisamento.

3.1 Estimador de Rosenblatt-Parzen

A partir do trabalho de Rosenblatt (1956), foram direcionados os esforços para a obtenção de estimadores da função de densidade contínuos. Observemos que o estimador em (1) pode ser escrito como

$$\tilde{f}(x) = \frac{1}{2nh} \sum_{i=1}^n 1_{(x-h, x+h]}(X_i). \quad (2)$$

Se definimos

$$K(x) = \frac{1}{2} \sum_{i=1}^n 1_{(-1, 1]}(x),$$

Estimador de Rosenblatt

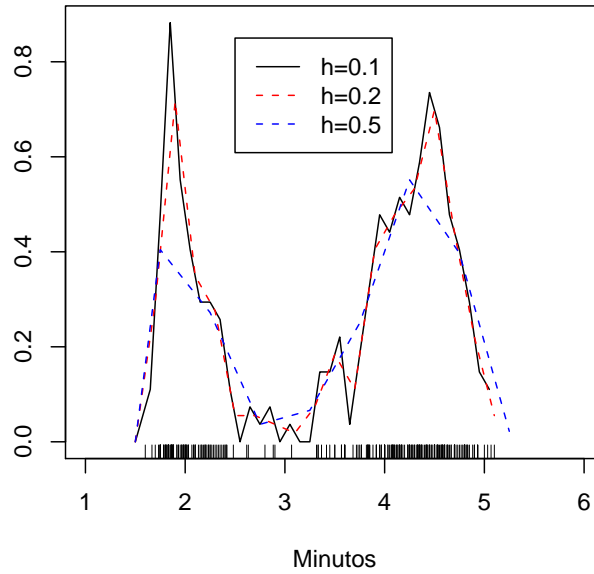


Figura 2: Estimador de Rosenblatt, segundo diversos valores do parâmetro de alisamento h , do tempo de duração das erupções do geysir Old Faithful, no parque nacional Yellowstone, Wyoming, USA.

o estimador $\tilde{f}(x)$ constitui uma integral de Riemann-Stieltjes com relação a $\tilde{F}_n(x)$, isto é,

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - y}{h}\right) d\tilde{F}_n(y).$$

A proposta de Parzen (1962) foi substituir as funções indicadoras na soma em (2) por determinadas funções K , chamadas de funções núcleo. Mudando a função K , obtemos da expressão anterior uma grande variedade de estimadores.

Definição 4 Seja X_1, \dots, X_n uma amostra aleatória da função de densidade f . Diremos que \hat{f} é o estimador kernel ou de Rosenblatt-Parzen de f se

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - y}{h}\right) dy,$$

onde $h > 0$ é chamado de parâmetro de alisamento e K é uma função real positiva satisfazendo $\int_{-\infty}^{\infty} K(x) dx = 1$.

Da definição anterior obtemos que

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = 1,$$

portanto $\hat{f}(x)$ é uma função de densidade e, além disso, se K é contínua e derivável, então $\hat{f}(x)$ também o será.

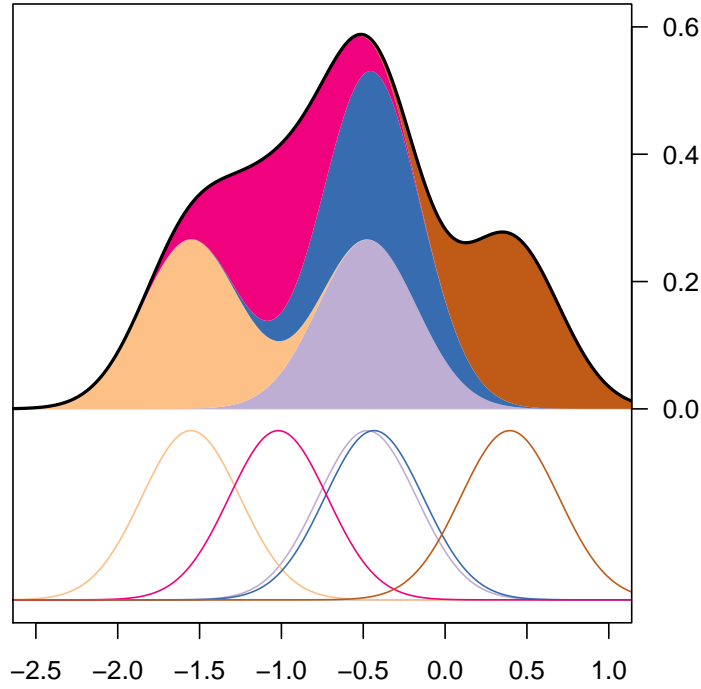


Figura 3: Estimador Rosenblatt-Parzen ou kernel. Influência de cada observação no estimador.

Podemos observar da definição que cada observação tem o mesmo peso relativo e que o resultados final, isto é, a curva estimada depende mais de como estejam agrupadas as observações. A Figura 3 mostra o processo de construção do estimador como soma ponderada do valor da função núcleo em cada valor de x .

Direcionaremos nossa atenção somente nas funções núcleo que sejam contínuas e deriváveis e, dada a restrição na definição do estimador kernel de terem integral um, consideraremos somente funções núcleo que sejam funções de densidade com derivadas contínuas. Além destas, outras condições sobre a função núcleo e sobre o parâmetro de alisamento h permitirão obter propriedades estatísticas do $\hat{f}(x)$.

3.1.1 Propriedades do estimador de Rosenblatt-Parzen

Como é natural no estudo de estimadores, estamos interessados nos momentos, com este objetivo observemos que

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n V_i(x),$$

onde

$$V_i(x) = \frac{1}{h} K\left(\frac{x - X_i}{h}\right),$$

são variáveis aleatórias independentes e igualmente distribuídas.

Escrito desta forma temos que

$$\mathbb{E}\{\hat{f}(x)\} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{V_i(x)\} = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - y}{h}\right) f(y) dy, \quad (3)$$

e

$$\text{Var} \left\{ \widehat{f}(x) \right\} = \int_{-\infty}^{\infty} \frac{1}{h^2} K^2 \left(\frac{x-y}{h} \right) f(y) dy - \left(\int_{-\infty}^{\infty} \frac{1}{h} K \left(\frac{x-y}{h} \right) f(y) dy \right)^2. \quad (4)$$

Dos resultados em (3) e (4) chegamos à conclusão de que os momentos do estimador se reduzem aos momentos da função núcleo utilizada e independem de n . Isto é importante conceitualmente pois mostra que, mesmo tomando uma amostra grande, não será possível reduzir o vício. Torna-se necessário então procurar condições para obter estimadores assintoticamente não tendenciosos.

Seguindo os estudos de Parzen (1962), Silverman (1986) e Wand & Jones (1995), vejamos condições para que este estimador seja assintoticamente não tendencioso. Para isto consideraremos o parâmetro de alisamento h como dependente do tamanho da amostra, isto é, $h = h_n$ e

$$\lim_{n \rightarrow \infty} h_n = 0.$$

Teorema 1 *Suponhamos que a função núcleo K satisfaz que $\lim_{n \rightarrow \infty} yK(y) = 0$ e que $\{h_n\}$ seja uma seqüência positiva convergindo a zero.*

Então, a esperança de $\widehat{f}(x)$ pode ser escrita como

$$E\{\widetilde{f}_n(x)\} = \frac{1}{h_n} \int_{-\infty}^{+\infty} K \left(\frac{y}{h_n} \right) f(x-y) dy,$$

e em cada ponto de continuidade de f ,

$$\lim_{n \rightarrow \infty} E\{\widetilde{f}_n(x)\} = f(x).$$

Deste teorema temos que $\widehat{f}(x)$ é assintoticamente não tendencioso. Outra propriedade geralmente exigida de todo estimador é a consistência. No seguinte teorema se obtém o limite da variação do estimador de Rosenblatt-Parzen.

Teorema 2 *Seja a função kernel limitada e de quadrado integrável. Neste caso o estimador kernel da função de densidade $\widehat{f}(x)$ tem variância satisfazendo*

$$\lim_{n \rightarrow \infty} nh_n \text{Var} \left\{ \widetilde{f}_n(x) \right\} = f(x) \int_{-\infty}^{+\infty} K^2(y) dy,$$

em todo ponto x de continuidade de f , se $h_n \rightarrow 0$ quando $n \rightarrow \infty$.

Estes dois teoremas foram demonstrados por Parzen (1962). Destes resultados concluímos que para o estimador de Rosenblatt-Parzen ser consistente exigimos que $nh_n \rightarrow \infty$ quando $n \rightarrow \infty$. Nesta situação

$$\lim_{n \rightarrow \infty} \text{Var} \left\{ \widetilde{f}_n(x) \right\} = 0.$$

Observemos que para garantir boas propriedades destes estimadores a função núcleo deve ser real, positiva, de integral 1, limitada, de quadrado integrável e tal que $\lim_{n \rightarrow \infty} xK(x) = 0$. Para o parâmetro de alisamento exigimos depender do tamanho da amostra de maneira que $\lim_{n \rightarrow \infty} h_n = 0$ e também ser tal que $\lim_{n \rightarrow \infty} nh_n = +\infty$.

3.1.2 Diferentes funções núcleo

Além das exigências já formalizadas acerca da função núcleo ou kernel esta deve ser admissível, isto é, ela tem que ser simétrica, unimodais e invariantes a transformações. Isto mostra que a escolha de tais funções é possível independentemente da escolha do parâmetro de alisamento.

Várias são as funções que podem servir como núcleo, dentre estas estão: gaussiana, epanechnikov, rectangular, triangular, biweight, coseno e optcosine. Mostramos na Figura 4 as funções mencionadas.

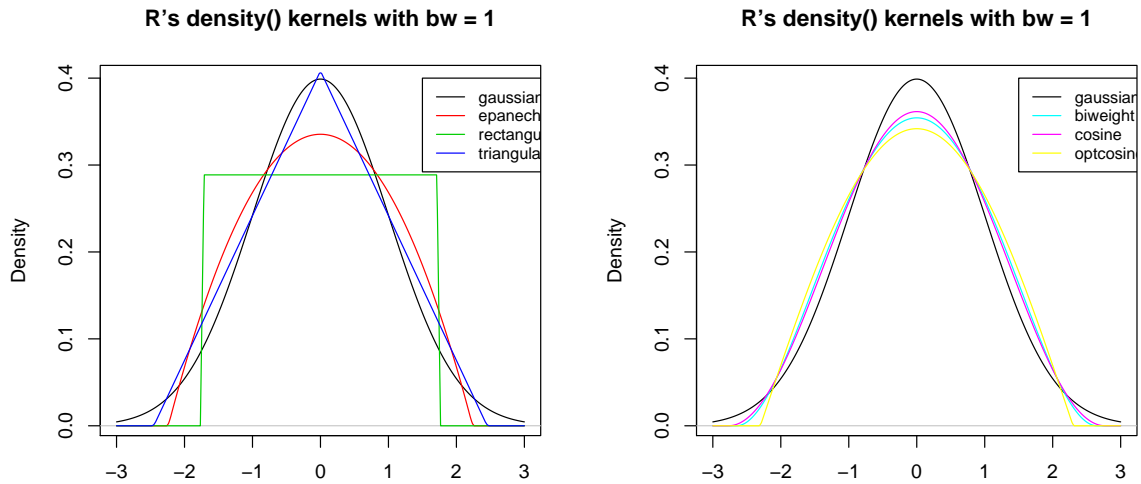


Figura 4: Diferentes funções núcleo admissíveis.

Pode-se observar que a forma funcional destas não é muito diferente, isto implica que na maioria das situações práticas basta utilizar a função núcleo gaussiana.

3.1.3 Estimação do parâmetro de alisamento

Segundo a definição do estimador $\hat{f}(x)$ observamos que depois de escolhida a função núcleo K , para se ter completamente especificada a forma do estimador, temos que achar uma expressão para o parâmetro de alisamento h .

Muito é conhecido acerca da relação entre a razão de convergência de $\tilde{f}_n(x)$ para f e o grau de dependência do parâmetro h como função do tamanho da amostra n . No entanto, para n fixo, o estimador $\tilde{f}_n(x)$ é sensível à escolha de h e não existe uma metodologia geral para a determinação deste parâmetro.

Diversos métodos de estimação de h estão disponíveis: `nrd0`, `nrd`, `sj`, `bcv` e `ucv`. Os dois primeiros são simples e baseiam-se na idéia de que a densidade a ser estimada é muito parecida à gaussiana.

Os métodos `sj`, `bcv` e `ucv` são mais complexos e serão os recomendados. Os métodos `bcv` e `ucv` estimam o valor de h via validação cruzada e `sj` via estimação de derivadas do estimador. A Figura 5 mostra o resultado de utilizar diferentes estimadores para o parâmetro de alisamento fixando a função kernel gaussiana. Percebemos que o resultado obtido quando utilizados os métodos `nrd0` e `nrd` são muito parecidos e diferem daqueles resultados obtidos quando utilizamos os critérios `sj`, `bcv` e `ucv`. Percebemos também que os resultados obtidos quando utilizamos estes os métodos `sj`, `bcv` e `ucv` são parecidos entre si.

Densidade do tempo de duração das erupções

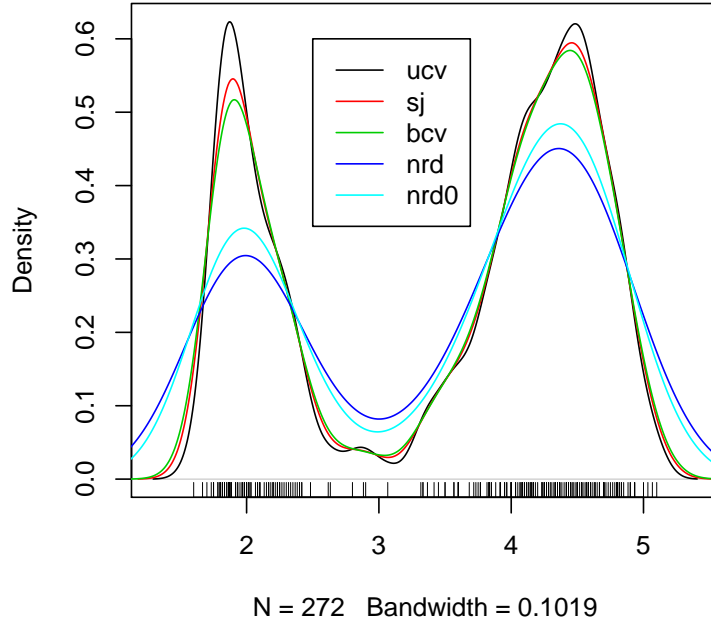


Figura 5: Diferentes estimadores segundo o método de estimação de h .

A idéia geral dos métodos sj , bcv e ucv é a estimação de h via validação cruzada. Este procedimento mede a habilidade do estimador de interpretar ou ajustar os dados observados. O parâmetro de alisamento será escolhido como aquele que maximiza esta medida. A validação cruzada é obtida eliminando-se uma observação e calculando-se este estimador na observação eliminada.

Denotemos por $\hat{f}_{n-1}^i(x)$ o estimador $\hat{f}(x)$ calculado depois de eliminar a i -ésima observação, isto é

$$\hat{f}_{n-1}^i(x) = \frac{1}{(n-1)h} \sum_{j \neq i=1}^n K\left(\frac{x - X_j}{h}\right).$$

Este estimador não depende de X_i e $\hat{f}_{n-1}^i(x)$ pode ser escolhido como medida apropriada da contribuição de X_i ao valor de h : se $\hat{f}_{n-1}^i(x)$ for grande, pode ser dito que $\hat{f}_{n-1}^i(x)$ antecipa a observação X_i e que h é apropriado. Valores pequenos de $\hat{f}_{n-1}^i(x)$ sugerem que a observação X_i pode ser esquecida e interpretá-la como evidência de que o valor de h é impróprio.

Variando i ao longo da amostra, obtemos n de tais medidas de ajuste que então podem ser combinadas na expressão de verossimilhança

$$L(h) = \prod_{i=1}^n \hat{f}_{n-1}^i(x),$$

ou utilizar as $\hat{f}_{n-1}^i(x)$ numa expressão aproximada ao erro quadrático integral, dada por

$$EQI_h \approx \int_{-\infty}^{+\infty} \hat{f}_n(x)^2 dx - 2 \frac{1}{n} \sum_{i=1}^n \hat{f}_{n-1}^i(x_i) + \int_{-\infty}^{+\infty} f(x)^2 dx.$$

Densidade do tempo de duração das erupções

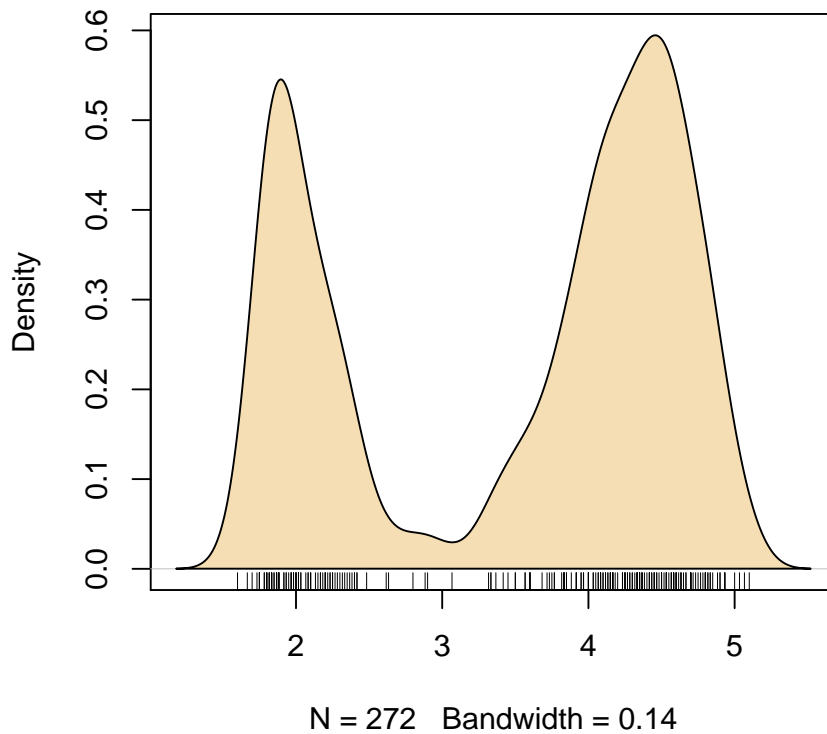


Figura 6: Estimador da densidade do tempo de duração das erupções do geysir Old Faithful, no parque nacional Yellowstone, Wyoming, USA. Estimação do parâmetro de alisamento h via validação cruzada.

Escolhendo \hat{h}_n como aquele que maximiza a verossimilhança ou minimiza o erro quadrático médio e , assim obtemos o estimador de h . Na Figura 6 mostramos o estimador da densidade do tempo de duração das erupções do geysir Old Faithful, no parque nacional Yellowstone, Wyoming, USA. Esta figura foi obtida estimando via validação cruzada o parâmetro de alisamento.

Referências

- Azzalini, A. & Bowman, A.W. (1990). A look at some data on the Old Faithful geysir. *Applied Statistics*, **39**, 357–365.
- Devroye, L. (1987). *A course in density estimation*. Birkhäuser.
- Devroye, L. & Györfi, L. (1985). *Nonparametric density estimation: the L_1 view*. John Wiley & Sons.
- Donoho, D.L. & Johnstone, I.M. (1989). Projection-based approximation and a duality with kernel methods. *Annals of Mathematical Statistics*, **17**, 58–106.
- Fryer, M.J. (1976). Some errors associated with the nonparametric estimation of density function. *Journal of the Institute of Mathematics and its Applications*, **18**, 371–380.

- Grenander, U. (1981). *Abstract Inference*. John Wiley & Sons.
- Johnson, N. & Kotz, S. (1988). *Encyclopedia of Statistical Science*. John Wiley & Sons.
- Kullback, A. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.
- Lucambio, F. (1998). *Estudo comparativo entre os métodos de Rosenblatt-Parzen e Grenander na estimação de densidades*. Master's thesis, Instituto de Matemática, Estatística e Computação Científica - IMECC, Universidade Estadual de Campinas - UNICAMP, Campinas, São Paulo, Brasil.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065–1076.
- Robertson, T. (1967). On estimating a density which is measurable with respect of a density function. *Annals of Mathematical Statistics*, **38**, 482–493.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimate of a density function. *Annals of Mathematical Statistics*, **27**, 832–837.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.
- Wand, M.P. & Jones, M.C. (1995). *Kernel smoothing*. Chapman & Hall.
- Wegman, E.J. (1975). Maximum likelihood estimation of a probability density function. *Sankyā*, **37**, 211–224.