Carlos Tenreiro

Uma introdução à estimação não-paramétrica da densidade

Coimbra, 2010

Prefácio

Decorria o ano de 1956 e nos *Annals of Mathematical Statistics* surgia um artigo de Murray Rosenblatt onde era proposta uma família de estimadores da densidade de probabilidade subjacente às observações realizadas que ficariam conhecidos na literatura como estimadores do núcleo (*kernel estimates*). Apesar de outros autores, como Fix e Hodges (1951) ou Akaike (1954), terem previamente considerado estimadores do mesmo tipo no contexto da análise discriminante não-paramétrica, é sem dúvida o trabalho seminal de Rosenblatt que dá origem a toda uma vasta literatura sobre estimação não-paramétrica de funções, sejam elas densidades de probabilidade, derivadas da densidade, funções de distribuição, funções de regressão e suas derivadas, entre outras.

De entre os múltiplos estimadores não-paramétricos que os utilizadores têm presentemente à sua disposição (ver Prakasa Rao, 1983; Devroye e Györfi, 1985; Silverman, 1986; Bosq e Lecoutre, 1987; Härdle, 1990, 1991; Thompson e Tapia, 1990; Scott, 1992; Wand e Jones, 1995; Simonoff, 1996; Bowman e Azzalini, 1997; Fan e Gijbels, 1997; Härdle et al., 2004; Wasserman, 2006), o estimador do núcleo é, sem dúvida, o mais popular. Algumas das razões que podem ser avançadas para justificar tal facto, são seguramente a simplicidade da sua definição, a sua versatilidade e as suas boas propriedades teóricas e práticas. Se centrarmos a nossa atenção no caso da estimação não-paramétrica da densidade de probabilidade, tópico que estudamos neste mini-curso, a popularidade do estimador do núcleo só é superada pela do estimador clássico do histograma que usamos em cur-

sos introdutórios de Estatística e que é em muitos *softwares* estatísticos, o único estimador da densidade disponibilizado ao utilizador.

Por estes motivos, mas também por razões didácticas que se tornarão claras no decorrer deste mini-curso, decidimos organizar o presente texto em torno destes dois estimadores da densidade. Depois de no Capítulo 1 nos dedicarmos a questões de índole geral sobre a estimação não-paramétrica da densidade de probabilidade, nos Capítulos 2 e 3 estudamos os estimadores do histograma e do núcleo, respectivamente. Apesar de centrarmos a nossa atenção no estudo do caso unidimensional, não deixaremos de abordar a generalização dos dois estimadores ao contexto multivariado.

Tomando como ponto de partida os assuntos que expomos neste texto, teremos ainda oportunidade de abordar durante o curso, mesmo que de forma breve, outros tópicos aqui não incluídos como são os casos da estimação da densidade sob condições de dependência, dos testes de ajustamento baseados no estimador do núcleo da densidade ou da estimação pelo método do núcleo de outros parâmetros funcionais de interesse.

À Comissão Organizadora do XVIII Congresso Anual da Sociedade Portuguesa de Estatística agradeço o convite que simpaticamente me formulou para leccionar este mini-curso. Devo um agradecimento especial ao meu colega Paulo Eduardo Oliveira pelo apoio e estímulo sempre demonstrados.

Carlos Tenreiro

Coimbra, Julho de 2010

Índice