

conoce como estimador *naive*. La expresión anterior puede ser escrita de la siguiente forma

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x-x_i}{h}\right) \text{ siendo } w(x) = \begin{cases} \frac{1}{2} & |x| < 1 \\ 0 & \text{c.c} \end{cases}$$

Podemos pensar que este estimador es un intento de construir un histograma donde cada punto es el centro de su respectiva clase, logrando así que no dependa tanto de la elección del origen. La elección de la amplitud de cada clase depende del parámetro h , que controla el valor mediante el cual los datos son suavizados. La generalización del estimador *naive* es el estimador *kernel*. Sustituyendo la función w por la función kernel K , el estimador *kernel* se define como

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \left(\frac{x-x_i}{h}\right)$$

donde h es el parámetro de alisamiento. La función K determina la forma de estas funciones mientras que el parámetro h determina su amplitud. Pero así como la elección de la función K (rectangular, triangular, Gaussiana o Epanechnikov) no influye de manera sustancial³, la adecuada elección del parámetro h es muy importante. Si elegimos un valor muy pequeño de h , introducimos ruido inherente al sistema (figura 1d) pero si lo elegimos demasiado grande, entonces el estima-

dor *kernel* será demasiado suavizado y características importantes, como la multimodalidad de los datos, podrían quedar ocultas (figura 1f). Es por ello que la elección del parámetro h siempre involucra un balance comparativo entre estas dos consideraciones^{3,4} (figura 1e).

Como conclusión, el histograma suavizado evita algunos de los problemas más frecuentes que se encuentran con los histogramas tradicionales, representando con mayor claridad la naturaleza continua de los datos. Su construcción con el software actual resulta tan sencilla como la del histograma tradicional, en especial con el paquete estadístico STATA^{5,6}. Sin embargo su forma final todavía se ve afectada, aunque en menor medida respecto al histograma tradicional, por dos decisiones relativamente arbitrarias, la elección de la función *kernel* y del parámetro de alisamiento.

Aurelio Tobías Garcés¹
Manuel González Scotto²

¹ *Unitat de Recerca Respiratòria i Ambiental*
Institut Municipal d'Investigació Mèdica
Universitat Autònoma de Barcelona. Barcelona
² *Unidade de Engenharia e Tecnologia Naval*
Instituto Superior Técnico. Lisboa, Portugal

Bibliografía

1. Sánchez-Cantalejo E, Ocaña Riola R. Actualizaciones en regresión: suavizando las relaciones. *Gac Sanit* 1997;11:24-32.
2. Härdle W. *Smoothing Techniques, with Implementation in S*. New York: Springer-Verlag; 1991.
3. Jacoby WG. *Statistical Graphics for Univariate and Bivariate Data*. CA: Sage; 1997.

4. Salgado-Ugarte IH, Shimizu M, Taniuchi T. snp6.2: Practical rules for bandwidth selection in univariate density estimation. *Stat Technical Bulletin* 1995;25:5-19.
5. Salgado-Ugarte IH, Shimizu M, Taniuchi T. snp6: Exploring the shape of univariate kernel density estimators. *Stat Technical Bulletin* 1993;16:8-19.
6. Salgado-Ugarte IH, Shimizu M, Taniuchi T. snp6.1: WARPing and kernel density estimation for univariate data. *Stat Technical Bulletin* 1995;26:23-31.

ACTIVIDADES CIENTÍFICAS

Diploma en Gestión Hospitalaria: la Práctica Directiva

Barcelona, 21 de octubre de 1997 al 16 de junio de 1998

Información

ESADE/Centro de Desarrollo Directivo (Laura Nualart)
Av. de Esplugues, 92-96 - 08034 Barcelona
Tel. (93) 280 40 08. Fax (93) 204 81 05/495 20 67
E-mail: cdd@esade.es
<http://www.esade.es/escsup/cdddo2e.htm>

1er Congreso Mundial de Salud y Medio Ambiente Urbano

Madrid, del 6 al 10 de julio de 1998

Información

TILESA O.P.C.
C/. Londres, 17 - 28080 Madrid
Tel. (91) 361 26 00 - Fax. (91) 355 92 08
E-Mail: tilesa@wpa.es