

ESTIMACIÓN DE DENSIDAD POR NÚCLEO (KERNEL) PARA DATOS CIRCULARES

Salgado Ugarte, Isaías Hazarmabeth

Laboratorio de Biometría y Biología Pesquera, Facultad de Estudios Superiores "Zaragoza", Batalla 5 de Mayo S/N esq. Fuerte de Loreto, Ejército de Oriente, Iztapalapa 09230, México, D.F. México;
isalgado@unam.mx

Introducción.

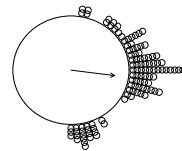
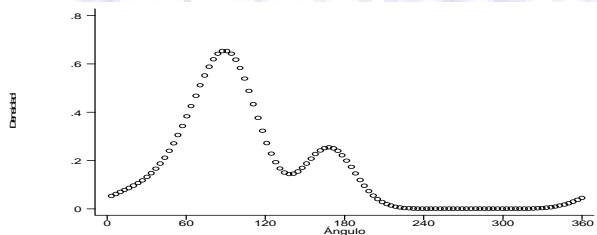
Los datos circulares aparecen con frecuencia en las investigaciones biológicas. Este tipo de información, también conocida como datos direccionales, requiere para su análisis procedimientos diferentes a los utilizados en la Estadística convencional. Como con los datos lineales, es necesario el entendimiento del nivel, dispersión, distribución y la existencia de casos extraordinarios. Para investigar la distribución, procedimientos muy eficientes son los estimadores de densidad por núcleo (kernel). Estos estimadores resuelven los problemas del origen, discontinuidad y ancho fijo de intervalo que poseen los tradicionales histogramas. En esta contribución se presenta un algoritmo que utiliza las funciones ponderales tradicionales y se agrega el núcleo de von Mises, equivalente al núcleo gaussiano en los estimadores de densidad de datos lineales. Asimismo, se incluyen reglas prácticas para la elección del parámetro de suavización (amplitud de banda).

Metodología.

Para la integración del algoritmo se tomaron como base los programas de trabajos anteriores (Salgado-Ugarte, 2002; Cox, 2001; 2004). Los programas corren en el paquete estadístico Stata. El programa para estimar la densidad es una versión aumentada del programa de Cox (2001) y permite el uso de las funciones ponderales uniforme, triangular, Epanechnikov, cuártica, triponderada, Gaussiana, coseno y la "normal circular" de von Mises. Se presenta además un programa que calcula varias reglas prácticas para elección de amplitud de banda. Se incluyen (como referencia) reglas lineales (sobresuavizadas y óptimas) y una regla "plug in" para datos circulares (Taylor, 2008).

Resultados y discusión.

Se aplicaron los algoritmos a datos de estratificación cruzada y de orientación de libélulas, usados por estudios anteriores. El programa calculador de reglas prácticas lineales produjo estimaciones iniciales susceptibles a ajustes posteriores. Las estimaciones de la densidad con el núcleo de von Mises en general produjo estimaciones más suaves. Resulta evidente la necesidad de continuar con la integración de rutinas con métodos complementarios para el entendimiento de la distribución de los datos circulares.



Dirección media 97.6; vector potencia 0.76

Conclusiones.

La estimación de densidad por núcleo es una de las herramientas más potentes para analizar la distribución de los datos. Se presentan algoritmos originales que permiten el extender el uso de estos importantes estimadores para datos circulares. Se destaca su utilidad para visualizar la distribución, característica fundamental de variables cuantitativas, independientemente del área que provengan.

Palabras clave. Datos circulares; estimación de densidad; funciones ponderales; distribución von Mises.

Bibliografía.

- Cox, N.J. 2001. Analysing circular data in Stata. North American Stata Users Group Meeting Proceedings. Marzo, 2001. Boston, EUA.
- Cox, N.J. 2004. Circular statistics in Stata, revisited. United Kingdom Stata Users' Meeting Proceedings. Junio, 2004. Londres, Reino Unido.
- Salgado-Ugarte, I.H. 2002. Suavización no paramétrica para análisis de datos. FES Zaragoza y DGAPA, UNAM, México,
- Taylor, C.C. 2008. Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*. **52**(7): 3493-3500.