

Estimación de entropías a partir de conjuntos limitados de datos

J. A. Bonachela, H. Hinrichsen, and Miguel A. Muñoz
*Departamento de Electromagnetismo y Física de la Materia and
 Instituto de Física Teórica y Computacional Carlos I,
 Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
 Fakultät für Physik und Astronomie, Universität Würzburg*

La estimación de entropías en series limitadas (finitas) de datos no es en absoluto trivial. Estimaciones naïve, basadas en reemplazar probabilidades por frecuencias de aparición lleva a errores no sólo de naturaleza estadística si no también sistemáticos (sesgo). De hecho, dicho sesgo obscurece los resultados para cualquier serie finita y obstaculiza el posible uso de funcionales de entropía (sean de Shannon, Rényi, Tsallis, etc). Distintos “estimadores” han sido propuestos en la literatura para evitar este problema y conseguir estimaciones significativas para conjuntos finitos de datos¹⁻⁵. En particular el método mejorado de Grassberger produce resultados muy satisfactorios cuando el número de estados posibles es grande y el tamaño de la muestra también. Sin embargo, en muchos problemas en biología, bioinformática, análisis de sistemas digitales, etc, el número de estados posibles es pequeño (2 en series binarias, 4 en series de nucleótidos de ADN, etc.).

En esta charla presentaré un nuevo método que min-

imiza simultáneamente los errores sistemáticos y estadísticos para sistemas con pocos estados y series de datos limitadas. El estimador puede ser adaptado para cualquier tipo de entropía. Complementa, por lo tanto, a los mejores estimadores disponibles hasta la fecha⁶.

¹ P. Grassberger, *Phys. Letters A* **128**, 369 (1988).

² M. S. Roulston, *Physica D* **125**, 285 (1999),

³ G. Miller, *Information Theory in Psychology II-B*, ed. H. Quastler, Glencoe, Illinois; Free Press, 95 (1955). See also, G. P. Basharin, *Theory Prob. App.* **4**, 333 (1959).

⁴ B. Harris *Topics on information Theory*, 323, ed. I. Csiszar, Amsterdam, North Holland (1975).

⁵ P. Grassberger, *condmat*, 0307138 (2003).

T. Schürmann, *J. Phys. A* **27**, L295 (2004).

T. Schürmann and P. Grassberger, *Chaos* **6**, 414 (1996).

⁶ J. A. Bonachela, H. Hinrichsen, and M. A. Muñoz, *J. Phys. A*, preprint 2008.