

Discriminación de distribuciones de probabilidad usando distancias entrópicas

Miguel A. Ré¹, Domingo P. Prato¹ y Pedro W. Lamberti^{1,2}

¹Facultad de Matemática, Astronomía y Física
 Universidad Nacional de Córdoba
 Ciudad Universitaria - (5010) - Córdoba - Argentina
²CONICET
e-mail: re@famaf.unc.edu.ar

La divergencia de Jensen-Shannon (JSD), una versión simetrizada de la divergencia de Kullback-Leibler, es una medida de distancia en teoría de probabilidad. Esta medida y generalizaciones de la misma han sido utilizadas en diferentes contextos, por ejemplo en el análisis de secuencias simbólicas, en el estudio de textos literarios, y en la discriminación de estados cuánticos. Si bien la extensión de su definición a distribuciones continuas de probabilidad es directa, la aplicación de la JSD ha estado restringida, en general, al estudio de secuencias discretas. En esta comunicación se analiza la aplicación del método para la detección de segmentación de secuencias asociadas a distribuciones continuas de probabilidad. En particular se consideran secuencias generadas con igual distribución de probabilidad pero distinto valor medio y secuencias generadas con distribuciones de probabilidad distintas con igual valor medio.

Jensen-Shannon divergence (JSD), a symmetrical version of Kullback-Leibler divergence, is a measure of distance in probability theory. JSD and generalizations of its definition have been used to treat different problems such as analysis of symbolic sequences, examination of texts in literature or separation of quantum states. Although JSD is well defined for continuous stochastic variables distributions, its use in segmenting sequences have been restricted to discrete random variables sequences. We study in this communication the application of JSD to detect segmentation in continuous range stochastic variables sequences. We consider in particular sequences generated with the same probability density with a different mean value and sequences generated with different probability densities, but with the same mean value.

Pacs N^o 87.23.-n, 05.40.-a

1 Introducción

El problema de discriminación entre distribuciones de probabilidad es de importancia central en la teoría de inferencia estadística. Desde los trabajos pioneros de Fisher⁽¹⁾ en esta área, ha quedado en claro la relevancia de la noción de distancia entre distribuciones de probabilidad para un planteo formal de la inferencia estadística. En estos trabajos se introdujo la cantidad hoy conocida como información de Fisher y que puede interpretarse como la distancia entre una distribución de probabilidades continua, $p(x)$ y su versión desplazada, $p(x + \delta)$. A su vez el estudio de la estabilidad de funcionales termodinámicos⁽²⁾, la cuantificación del concepto de complejidad⁽³⁾ o la distinción entre procesos estocásticos y caóticos⁽⁴⁾ son problemas de la Física Estadística que requieren para su resolución la distinción entre distribuciones de probabilidad. Dada la naturaleza probabilística de la teoría, en Mecánica Cuántica, la distinción entre distribuciones de probabilidad está íntimamente relacionada

con la distinguibilidad de estados cuánticos^(5,6). Para dar una estimación cuantitativa de la diferencia entre dos distribuciones de probabilidad se pueden considerar diversas medidas de distancia entre las mismas⁽⁷⁾. Sin embargo, muchas de las definiciones de distancia presentan inconvenientes, tanto formales como de aplicabilidad^(8,9). En los últimos años se ha estudiado una noción de distancia denominada divergencia de Jensen-Shannon (JSD), una versión simetrizada de la divergencia de Kullback-Leibler, que ha mostrado ser de gran utilidad en diversos contextos⁽¹⁰⁾. En términos de la entropía de Gibbs: $H(P) = -\sum_j p^{(j)} \ln(p^{(j)})$ (o la correspondiente versión para variables de rango continuo $H(P) = -\int dx P(x) \ln(P(x))$), puede expresarse como

$$JS(P_1, P_2) = \frac{H(\pi_1 P_1 + \pi_2 P_2) - \pi_1 H(P_1) - \pi_2 H(P_2)}{\pi_1 + \pi_2} \quad (1)$$

con π_1 y π_2 dos números positivos que cumplen $\pi_1 + \pi_2 = 1$.

En estudios previos se ha usado la JSD

como herramienta de detección de segmentación de secuencias⁽¹¹⁾, *i.e.* del cambio de contenido simbólico en las secuencias. En estos estudios las secuencias consisten de valores de variables aleatorias de rango discreto. Esencialmente el método consiste en dividir una secuencia S de n valores en dos tramos de longitud arbitraria ν_1 y ν_2 respectivamente y estimar las frecuencias de aparición de cada uno de los valores discretos en cada tramo. Los pesos de cada distribución en la definición de la JSD se toman como $\pi_i = \nu_i/n$. Se verifica un valor extremo de la JSD cuando $\nu_1 = n_1$, detectándose de esta forma el punto de segmentación.

En esta comunicación proponemos considerar la JSD como una medida que permita distinguir la segmentación de secuencias asociadas a distribuciones continuas de probabilidad. En particular se consideran secuencias generadas con igual distribución de probabilidad pero distinto valor medio y secuencias generadas con distribuciones de probabilidad distintas con igual valor medio.

2 Segmentación de secuencias

Sean X_1 y X_2 dos variables estocásticas de rango continuo distribuidas según las densidades de probabilidad $\phi_1(x)$ y $\phi_2(x)$ respectivamente: la probabilidad de que $x < X_i \leq x + dx$ es $\phi_i(x)dx$. Consideremos ahora una secuencia S de n valores conformada de la siguiente manera: una subsecuencia S_1 de n_1 valores de la variable X_1 seguida de una subsecuencia S_2 de n_2 valores de X_2 , con $n_1 + n_2 = n$.

Suponiendo que la única información disponible es la secuencia S , resultando desconocidas tanto las densidades f_i como los largos n_i , se pretende determinar el punto de segmentación o equivalentemente los valores de n_1 y n_2 .

La extensión de este método al caso de secuencias de variables aleatorias de rango continuo no es directa. La estimación de las respectivas distribuciones a través de las frecuencias relativas resulta complicado en este caso. Por esta razón hemos elegido seguir un camino alternativo dividiendo el rango de la variable continua en un número finito de intervalos y estimando la probabilidad de que el valor de la variable aleatoria pertenezca a dicho intervalo a través de las frecuencias relativas. La definición de los intervalos se basa en valores característicos de la distribución como la media o la varianza, estimados en cada tramo por el promedio o la varianza de la muestra.

Nuestra propuesta se basa así en la división de la secuencia en dos tramos de longitudes arbitrarias ν_1 y ν_2 ($\nu_1 + \nu_2 = 1$). En cada tramo calculamos la media m_i y definimos una nueva variable aleatoria en cada uno de ellos según la siguiente regla

$$Y_i = \begin{cases} 0 & X_i \leq m_i \\ 1 & X_i > m_i \end{cases} \quad (2)$$

Las nuevas variables son de rango discreto, aún cuando a priori las distribuciones de probabilidad

son desconocidas. En este caso podemos adoptar el método antes mencionado⁽¹¹⁾, calculando las frecuencias relativas en cada tramo y a partir de éstas la JSD.

Repetimos el cálculo modificando los valores de ν_1 y ν_2 en el rango $[2, n - 1]$ y verificamos la detección de la segmentación como un extremo de la JSD. Cabe mencionar que en algunas situaciones la división del rango de valores continuos en dos intervalos resulta insuficiente como se ilustra en uno de los ejemplos presentados a continuación.

3 Ejemplos de ilustración

Consideramos a continuación tres casos ejemplo para ilustrar el funcionamiento del método de detección de cambio en la distribución de probabilidades que caracteriza un proceso. En estos casos se generaron 2000 secuencias de 2000 valores cada una. Cada secuencia se conformó como sigue: los primeros 1500 valores se generaron por el método de Monte Carlo a partir de una densidad de probabilidad ϕ_1 y los 500 restantes a partir de una segunda densidad ϕ_2 . En cada uno de los ejemplos incluidos se indican las densidades de probabilidad usadas.

Cada secuencia generada se divide en dos tramos, indicando por ν_1 la longitud del primer tramo. Dando valores a ν_1 entre 2 y 1998, se calcula en cada tramo el valor medio de la muestra $m_i(\nu_1)$. De esta manera generamos una secuencia de valores discretos según la regla de asignación (2) y calculamos las correspondientes frecuencias en cada tramo

$$f_i(a) = \frac{r_i(a)}{\nu_i} \quad (3)$$

con $f_i(a)$ la frecuencia de repetición del valor a (0 o 1) en el tramo i y $r_i(a)$ el número de veces que se repite el valor a .

A partir de las frecuencias determinadas se calcula la JSD como función de ν_1 y se promedia su valor entre las 2000 secuencias. Finalmente graficamos la JSD en función de ν_1 . Como puede apreciarse en las figuras de cada ejemplo, el valor de segmentación ($\nu_1 = n_1$) aparece como un extremo de la JSD.

3.1 Caso 1

En este ejemplo elegimos la misma forma funcional para ambas densidades de probabilidad: $\phi_i = 1/\mu_i \exp(-t/\mu_i)$, *i.e.* una distribución exponencial para cada subsecuencia. La diferencia entre ambas densidades surge del valor medio elegido: $\mu_1 = 1$ y $\mu_2 = 2$ respectivamente. Siguiendo el procedimiento general antes descrito obtenemos los resultados ilustrados en la figura 1. En este caso el punto de segmentación aparece asociado a un mínimo de la JSD. En la figura 2 se incluye el histograma de valores para una secuencia particular para el total de valores y para cada subsecuencia por separado.

El resultado puede interpretarse tomando en cuenta que la JSD es definida positiva y es cero

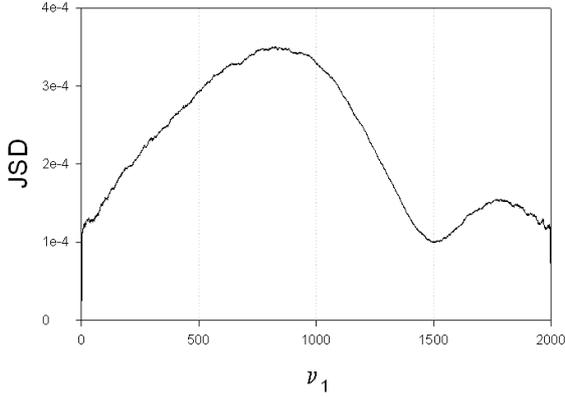


Figura 1: JSD en función de ν_1 , largo del primer tramo. Notamos que el punto de segmentación ($\nu_1 = 1500$) aparece asociada con un mínimo de la JSD.

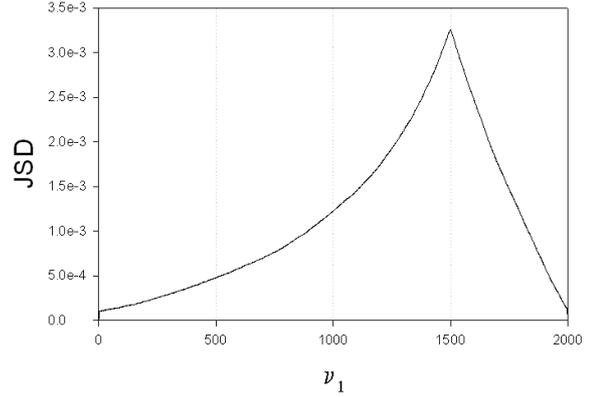


Figura 3: JSD en función de ν_1 , largo del primer tramo. Notamos que el punto de segmentación ($\nu_1 = 1500$) aparece asociada con un máximo de la JSD.

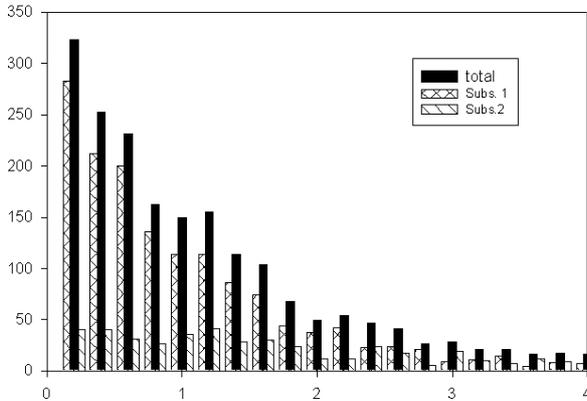


Figura 2: Histograma de una secuencia particular generada para el ejemplo 1.

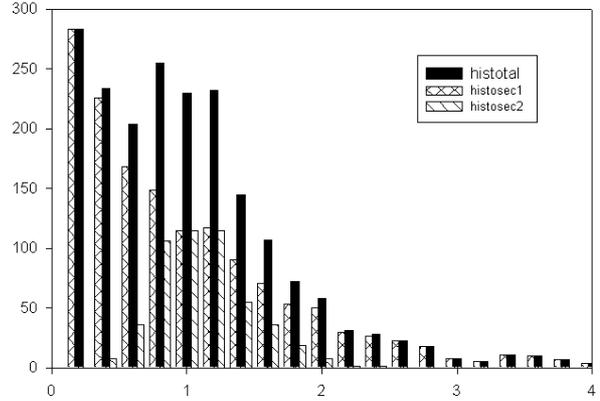


Figura 4: Histograma de una secuencia particular generada para el ejemplo 2.

sólo cuando las distribuciones de probabilidad son idénticas. En el ejemplo en consideración, atento a las densidades de probabilidad elegidas, las distribuciones de probabilidad para la variable Y en cada tramo se hacen iguales para $\nu_1 = n_1$.

3.2 Caso 2

En este ejemplo elegimos las distribuciones

$$\begin{aligned} \phi_1(x) &= \frac{1}{\mu} e^{-x/\mu} \\ \phi_2(x) &= \frac{\eta}{\mu} \left(\frac{\lambda \eta x}{\mu} \right)^{\eta-1} \frac{1}{\Gamma(\eta)} e^{-\eta x/\mu} \end{aligned} \quad (4)$$

i.e. una densidad exponencial para la primer subsecuencia y una densidad tipo Erlang para la segunda. Con la elección de los parámetros los valores medios serán los mismos para ambas distribuciones: $\mu_1 = \mu_2 = \mu$ y por lo tanto, en promedio, la media en cada tramo es independiente del valor de ν_1 . Sin embargo la probabilidad en cada subsecuencia será distinta y estamos en el caso general analizado en⁽¹¹⁾.

Los resultados obtenidos se muestran en la figura 3

y el histograma de valores para una secuencia particular en la figura 4. Puede verse aquí que el punto de segmentación está indicado por un máximo de la JSD.

3.3 Caso 3

Como último ejemplo elegimos una situación en la que no es suficiente dividir el rango en dos intervalos. Las densidades de probabilidad a partir de las que se generan las subsecuencias son

$$\begin{aligned} \phi_1(x) &= \frac{1}{\pi} \frac{1}{1+x^2} \\ \phi_2(x) &= \frac{1}{\sqrt{(2\pi)}} e^{-x^2} \end{aligned} \quad (5)$$

i.e. una distribución de Cauchy para la primer subsecuencia y una densidad normal para la segunda. La elección de parámetros es tal que ambas distribuciones tienen media 0 y además en ambos casos $P(x < \mu)$ es la misma. El método usado en los dos ejemplos anteriores no da una diferencia significativa al calcular la JSD. Sin embargo, si dividimos el rango de valores en cuatro intervalos obtenemos un extremo en la JSD.

De esta forma modificamos el método hasta aquí utilizado calculando en cada tramo el promedio y la

varianza de la muestra. Modificamos así la regla de asignación de valores para las variables discretas Y_i según

$$Y_i^{(j)} = \begin{cases} 0 & \text{si } x_i^{(j)} \leq m_i - s_i \\ 1 & \text{si } m_i - s_i < x_i^{(j)} \leq m_i \\ 2 & \text{si } m_i < x_i^{(j)} \leq m_i + s_i \\ 3 & \text{si } m_i + s_i < x_i^{(j)} \end{cases} \quad (6)$$

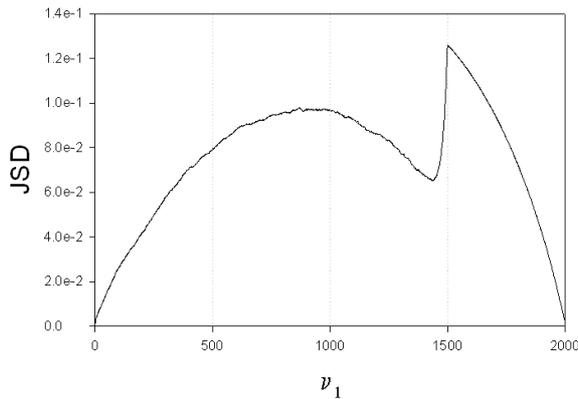


Figura 5: JSD en función de ν_1 , largo del primer tramo. Notamos que el punto de segmentación ($\nu_1 = 1500$) aparece asociada con un máximo de la JSD.

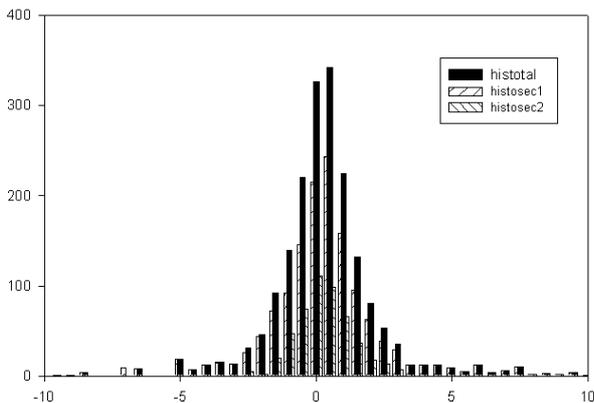


Figura 6: Histograma de una secuencia particular generada para el ejemplo 3.

Las frecuencias de repetición para cada valor se calculan según (3, donde a toma ahora un valor de cuatro posibles.

A partir de las frecuencias determinadas calculamos ahora la JSD, obteniendo los resultados de la figura 5. Nuevamente encontramos que el punto de segmentación aparece relacionado con el máximo de la JSD. En la figura 6 mostramos un histograma de valores para una secuencia particular.

4 Conclusiones

Se corrobora la utilidad de la JSD como herramienta para análisis de segmentación de secuencias, extendiendo su aplicación a secuencias de valores con rango continuo.

Para su aplicación se propone la definición de nuevas variables aleatorias de rango discreto a partir de los valores registrados en la secuencia y de sus propiedades estadísticas: media y desviación estándar en cada tramo. Esta propuesta permite una aplicación práctica del método.

El punto de segmentación aparece claramente indicado en cada caso como un valor extremo de la JSD, demostrando el método una buena sensibilidad.

Agradecimientos: Los autores agradecen el financiamiento de SeCyT-UNC y CONICET, PIP 6036/05, para este proyecto.

Referencias

- [1] R. Fisher, *Proc. Roy. Soc. A*, **121**, 654 (1928).
- [2] B. Lesche, *J. Stat. Phys.* **27**, 419-423 (1982).
- [3] R. Lopez-Ruiz, H. Mancini and X. Calbet, *Phys. Lett. A* **209**, 321-326 (1995).
- [4] O. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino y M. A. Fuentes, *Phys. Rev. Lett.*, **99**, 154102 (2007).
- [5] A. Majtey, P.W. Lamberti, M.T. Martin and A. Plastino, *Eur. Phys. J. D* **32**, 413-418 (2005).
- [6] A. Majtey, P.W. Lamberti and D. Prato, *Phys. Rev. A* **72** 052310 (2005)
- [7] Elena Deza y Michel-Marie Deza "Dictionary of Distances" (First Edition) Elsevier B.V. (2006).
- [8] A.N. Shyriaev, *Probability Theory*, Springer, 1996.
- [9] S. Kullback and R.A. Leibler, *Ann. Math. Stat.* **22**, 79-86 (1951).
- [10] P.W. Lamberti, A. Majtey, M. Madrid y M. Pereyra, "Proceeding of XV Conference on Non-equilibrium Statistical Mechanics and Nonlinear Physics" Eds. O. Descalzi, O. Rosso y H. Larrondo, American Institute of Physics, pag. 32-37 (2007)
- [11] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, H. E. Stanley, *Phys. Rev. E* **65**, 041905 (2002).