

Distribuciones de máxima entropía, incremento de aleatoriedad y teoremas límite en probabilidad

Evgueni I. Gordienko¹ y Adolfo Minjárez-Sosa²

¹ Departamento de Matemáticas, Universidad Autónoma Metropolitana, Iztapalapa

² Departamento de Matemáticas, Universidad de Sonora

¹ gord@xanum.uam.mx, ² adolfo.minjarez@unison.mx

Resumen

En este artículo se explora el concepto de distribuciones de máxima entropía en el contexto de la teoría de la probabilidad. Se analiza la relación entre incremento en la aleatoriedad, máxima entropía y los teoremas límite en probabilidad, proporcionando un enlace entre la incertidumbre estadística y los principios termodinámicos, en particular con la segunda ley de la termodinámica. En términos generales, veremos que el aumento en la entropía va acompañado con el aumento de la incertidumbre o aleatoriedad en un sistema, lo cual puede ser interpretado como una transición a un estado de equilibrio termodinámico. Finalmente algunos de estos resultados se presentan en el marco de la teoría de la información.

Palabras Clave: Entropía; Incertidumbre; Teoremas Límite en Probabilidad; Termodinámica; Teoría de la Información.

DOI: 10.36788/sah.v8i1.146

Recibido: 30 de enero de 2024

Aceptado: 26 de junio de 2024

1. Introducción

La palabra “entropía” viene de griego “ $\epsilon\nu\tau\rho\omicron\pi\acute{\iota}\alpha$ ” y significa evolución o transformación. Dicho término fue acuñado en 1865 por el físico alemán Rudolf Clausius en el contexto de la termodinámica clásica, una rama de la física dedicada al estudio de las relaciones entre las “propiedades macroscópicas” observables de gases, líquidos, etc., tales como presión, temperatura, energía interna, conversión de calor en trabajo mecánico, entre otras. Es por esta razón que la entropía de Clausius se conoce actualmente como entropía termodinámica la cual, en términos generales, puede entenderse como “la porción de la energía térmica de un sistema que no puede convertirse en trabajo mecánico” .

Al mismo tiempo, a través de numerosos estudios, y gracias a una extensa base experimental, se ha establecido firmemente que en sistemas aislados físicos y químicos la entropía aumenta a medida que los sistemas transitan de “estados de transición” a “estados de equilibrio termodinámico” estos últimos caracterizados por una entropía máxima. Por lo tanto, en

sistemas aislados, la entropía siempre aumenta o se mantiene constante. El siguiente ejemplo ilustra esta idea.

Supongamos que la mitad izquierda de un recipiente aislado contiene gas a temperatura constante, mientras que la mitad derecha está vacía (ver Figura 1.1).

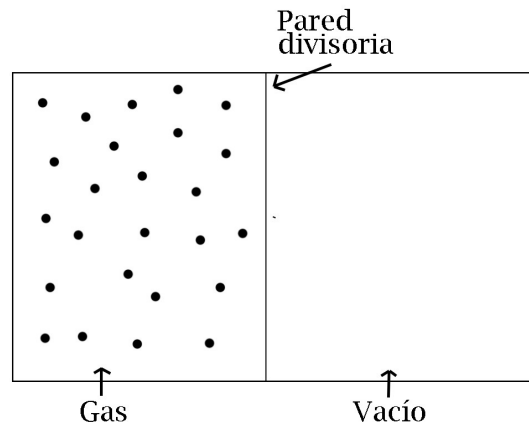


Figura 1:

Si la pared divisoria es removible y, en algún momento, se desliza fuera del recipiente sin romper la hermeticidad del aire, entonces el sistema se encontrará en un estado de desequilibrio, y para restablecer el equilibrio el gas se distribuirá uniformemente en todo el volumen del recipiente en cuestión de milisegundos (ver Figura 1.2).

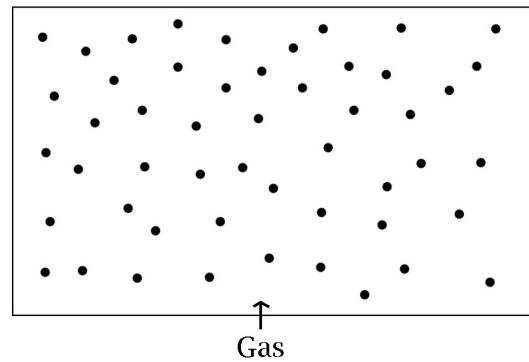


Figura 2:

Posteriormente, veremos que la transición al estado del sistema en la Figura 1.2 está acompañada por un aumento en la entropía, y dicho estado constituye el estado de entropía máxima del gas en el recipiente sin pared divisoria. Es importante destacar que tal transición es irreversible en el tiempo: el gas “nunca” pasará espontáneamente del estado de equilibrio (Figura 1.2) al estado presentado en la Figura 1.3.

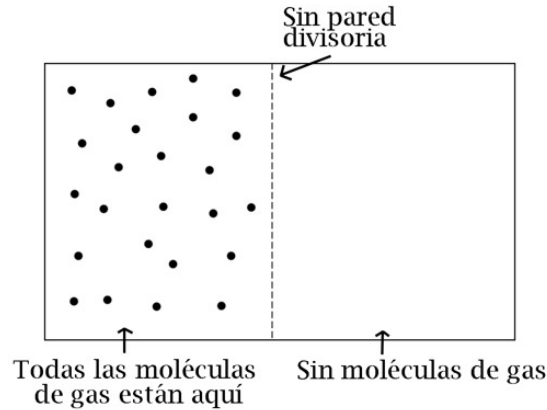


Figura 3:

En resumen, un hecho experimental y parcialmente demostrado matemáticamente es que en sistemas termodinámicos aislados, la entropía no puede disminuir, y dichos sistemas se desplazan de manera irreversible hacia el equilibrio termodinámico, que por lo general se caracteriza por una entropía máxima. Esta esencia constituye la famosa segunda ley de la termodinámica, proporcionando “la flecha del tiempo” es decir, la dirección inevitable del tiempo desde el pasado hacia el futuro.

Años después de la conceptualización de Clausius, el científico austriaco Ludwig Boltzmann introdujo en 1877 la formulación matemática actual de la entropía, y luego J. Willard Gibbs propuso una breve extensión de esta definición. Desde la perspectiva de la teoría de la probabilidad moderna, ambas definiciones utilizan la densidad de probabilidad f_t (en cada tiempo t) de posiciones y velocidades de moléculas de un sistema termodinámico (gas en nuestro caso), definida sobre el espacio fase M , concluyendo que la entropía es definida por el siguiente número:

$$H(f_t) = - \int_M f_t(x) \log [f_t(x)] dx, \quad (1)$$

donde x representa las posiciones y velocidades de todas las partículas. Tomando como base esta definición y utilizando ecuaciones de la mecánica clásica para el movimiento de partículas, Boltzmann se dio a la tarea de justificar matemáticamente la segunda ley de la termodinámica. Es decir, demostró que para un “gas ideal” en un recipiente aislado, la entropía $H(f_t)$ no disminuye en función del tiempo, y bajo algunas restricciones, la distribución límite de posiciones es uniforme en el contenedor y la de velocidades es normal. Inicialmente Boltzmann pensó que esta demostración del aumento de la entropía se fundamentaba exclusivamente en las ecuaciones de la mecánica clásica reversibles en el tiempo. Sin embargo, se hizo evidente posteriormente que, de hecho, estaba empleando una suposición probabilística implícita que conduce a una dirección unidireccional del tiempo. En términos probabilísticos, este aumento en la entropía se relaciona con el crecimiento de la incertidumbre o “aleatoriedad” en el sistema, que puede interpretarse como una transición hacia un “estado de equilibrio”: a mayor entropía, mayor es la incertidumbre asociada a la variable aleatoria.

Es en este punto donde se relaciona la teoría de la probabilidad y la entropía. En efecto, una parte importante de la teoría de la probabilidad moderna la constituyen los llamados teoremas límite que estudian el comportamiento asintótico de las distribuciones de sumas y productos de variables aleatorias cuando el número de términos crece a infinito. Estos teoremas juegan un papel esencial en la teoría de procesos estocásticos, estadística matemática y, por supuesto, en termodinámica estadística, facilitando la comprensión de por qué las distribuciones de probabilidad como la normal, uniforme y de Poisson, son comunes en diversos fenómenos naturales, ámbitos económicos y contextos científicos. Por ejemplo, el teorema límite más conocido, el llamado Teorema Central de Límite, establece que la distribución límite de la suma normalizada de variables aleatorias independientes e idénticamente distribuidas con media y varianza finitas, converge a la distribución normal estándar. El punto importante en este teorema es que la distribución límite, la normal, no depende de una distribución particular de las variables aleatorias. En este sentido la distribución límite es universal.

En el ámbito de la entropía, hacia finales del siglo XX, distintos estudios se centraron en el hecho de que las distribuciones límite, en varios teoremas límite, exhiben la máxima entropía dentro de ciertas clases de distribuciones. Un ejemplo de esto es que la densidad normal estándar posee la entropía máxima entre las densidades de variables aleatorias con media cero y varianza unitaria. Mas aún, en el contexto del teorema central del límite, se ha demostrado que las entropías de las sumas de variables aleatorias aumentan a medida que se incrementa el número de términos, llegando su límite a la entropía de la densidad normal estándar, como es de esperarse. En consecuencia, se observa un fenómeno análogo a la segunda ley de la termodinámica. Este resultado no sorprende, dado que las distribuciones de entropía máxima, como la normal o la uniforme, implican mayores niveles de incertidumbre y deben “acumular” las incertidumbres de todas las variables aleatorias hacia las distribuciones límite.

El objetivo del presente artículo es exponer, desde el punto de vista matemático, los aspectos antes descritos, poniendo énfasis en los aspectos intuitivos y de interpretación.

Por otro lado, aun cuando la conceptualización de la entropía tiene sus orígenes en la termodinámica, en 1948 el matemático e ingeniero electricista americano Claude Shannon involucró este concepto en la *Teoría de la Información*. Esta teoría tiene su fundamento en la cuantificación de la información contenida en mensajes que se transmiten a través de canales donde está presente algún ruido aleatorio. Básicamente, el trabajo de Shannon consiste en relacionar la entropía termodinámica con la cantidad promedio de información en los mensajes. En este sentido, concluiremos el artículo presentando el papel que juega la entropía en la teoría de información con un ejemplo específico.

2. Entropía como una medida de incertidumbre (“aleatoriedad”)

Con el fin de ilustrar el papel que juega la entropía como una medida de la incertidumbre, presentamos algunos ejemplos considerando variables aleatorias discretas y absolutamente continuas. Para este fin, observe que la versión discreta de (1) es la siguiente. Sea X una variable aleatoria con distribución \mathcal{D}_X que toma valores x_1, x_2, \dots . Entonces

$$H(\mathcal{D}_X) = - \sum_k p_k \log(p_k), \quad (2)$$

donde $p_k = P[X = x_k]$, $k = 1, 2, \dots$

Ahora, sea X la variable aleatoria de Bernoulli con parámetro $p \in [0, 1]$. En este caso, X asume solo los valores 0 y 1, y su distribución \mathcal{D}_X es: $P(X = 1) = p$; $P(X = 0) = 1 - p$. Entonces, por (2)

$$H(\mathcal{D}_X) \equiv H(p) = -[p \log(p) + (1 - p) \log(1 - p)]. \quad (3)$$

En (3), $0 \log(0) \stackrel{\text{def}}{=} 0$. La gráfica de la función $H(p)$, $p \in [0, 1]$, está dada en la Figura 4. Observe que el máximo de la entropía es alcanzado en $p = 0.5$, i.e. cuando $P(X = 1) = P(X = 0) = 0.5$ y su valor es

$$H(0.5) \approx 0.69315. \quad (4)$$

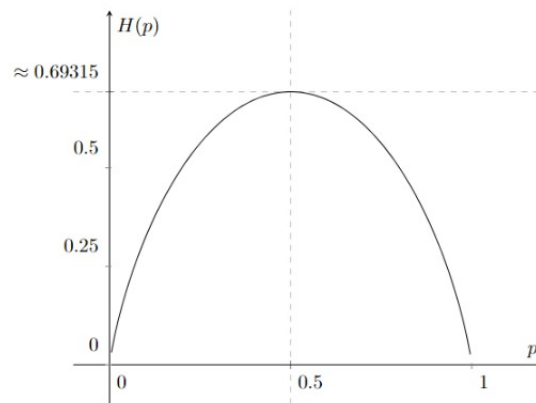


Figura 4:

Este es el caso de *máxima incertidumbre* sobre el posible valor de X . Por ejemplo cuando se lanza una moneda justa, difícilmente se puede predecir el resultado.

Sea ahora $p = 0.99999$, i.e. $P(X = 1) = 0.99999$, $P(X = 0) = 0.00001$. Uno puede “casi seguramente” predecir el valor 1, con lo cual podemos decir que estamos en un caso de *baja incertidumbre*. La entropía en este caso es $H(0.99999) \approx 0.000125129$.

En términos generales, la variable aleatoria con distribución de Bernoulli con $p = 0.99999$ es “menos aleatoria” que la variable aleatoria con distribución de Bernoulli con $p = 0.5$.

Consideremos otro ejemplo. Supongamos que la variable aleatoria X tiene la densidad uniforme f_X en el intervalo $(0,1)$ cuya gráfica se muestra en la Figura 5.

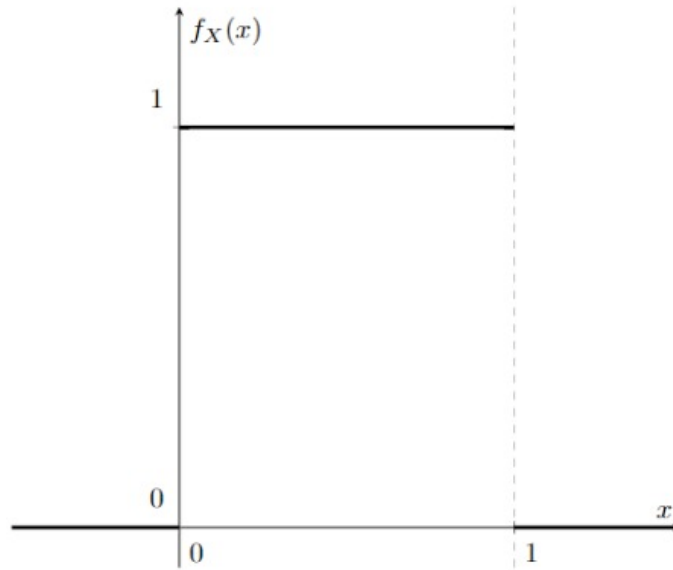


Figura 5:

Entonces, aplicando la definición (1), la entropía de f_X es

$$H(f_X) = - \int_0^1 1 \log(1) dx = 0.$$

La densidad uniforme en la Figura 5 se acopla a nuestro sentimiento intuitivo de *completa incertidumbre* al “escoger un punto *al azar*” del intervalo $(0, 1)$. En efecto, consideremos la variable aleatoria Y con densidad uniforme f_Y en el intervalo $(0.45, 0.55)$ (ver Figura 6). En este caso, estamos bastante seguros de que Y tomará algún valor cercano a 0.5. Por otro lado, la entropía es

$$H(f_Y) = - \int_{0.45}^{0.55} 10 \log(10) \approx -2.30258,$$

la cual es mucho menor que $H(f_X) = 0$. Ahora, sea f_Z la densidad uniforme en el intervalo $[-100, 100]$ (ver Figura 7). Entonces, el grado de incertidumbre sobre los posibles valores de Z es bastante alto, y la entropía $H(f_Z) \approx 5.29832$ es muy grande.

A partir de estos ejemplos, una primera conclusión es que entre mas grande es el valor de la entropía, mayor es la incertidumbre asociada a la variable aleatoria.

Otra situación donde se presenta un incremento de la entropía, es decir mayor incertidumbre, es en la suma de variables aleatorias, lo cual toma importancia en algunos teoremas

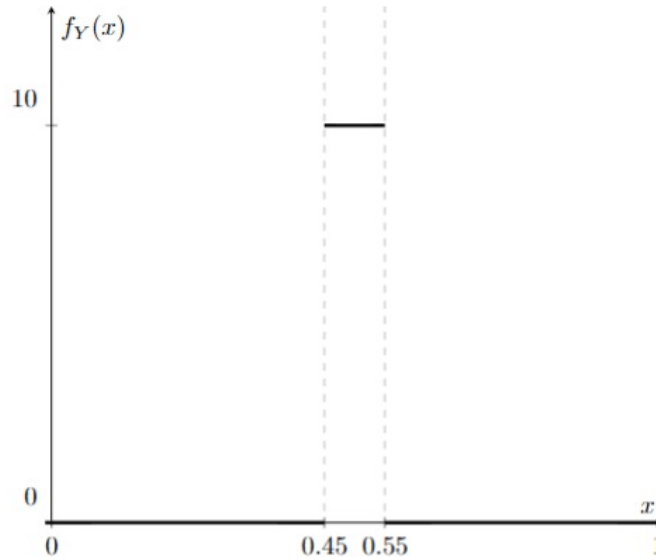


Figura 6:

límite de la Teoría de la Probabilidad. En efecto, como veremos en la Sección 4, similarmente a la segunda ley de la termodinámica, en algunos teoremas límite, la entropía crece conforme aumenta el número de sumandos hasta alcanzar la entropía máxima de ciertas distribuciones.

Vale la pena ilustrar mediante el siguiente ejemplo el por qué la entropía (y la incertidumbre) pueden crecer al aumentar el número de sumandos de variables aleatorias independientes.

Ejemplo 1. Sean X_1, X_2, \dots, X_5 variables aleatorias independientes e idénticamente distribuidas (i.i.d.), y sea X_1 la variable aleatoria de Bernoulli con parámetro $p = 0.95$. Entonces $P(X_1 = 1) = 0.95$, $P(X_1 = 0) = 0.05$, y realmente esperaríamos la aparición del valor 1 (digamos, “en algún experimento”). Aplicando (3),

$$H(\mathcal{D}_{X_1}) = -[0.95 \log(0.95) + 0.05 \log(0.05)] \approx 0.19852, \quad (5)$$

lo cual es significativamente menor que la entropía en (4).

Por otro lado, la suma $S_5 = X_1 + \dots + X_5$ tiene la distribución binomial \mathcal{D}_{S_5} con parámetros $n = 5$ y $p = 0.95$. Usando la fórmula binomial, encontramos que:

$$P(S_5 = 0) \approx 0.0000003125;$$

$$P(S_5 = 1) \approx 0.000029687;$$

$$P(S_5 = 2) \approx 0.0011281;$$

$$P(S_5 = 3) \approx 0.021434;$$

$$P(S_5 = 4) \approx 0.20363;$$

$$P(S_5 = 5) \approx 0.77378.$$

Entonces la incertidumbre ha aumentado. En efecto:

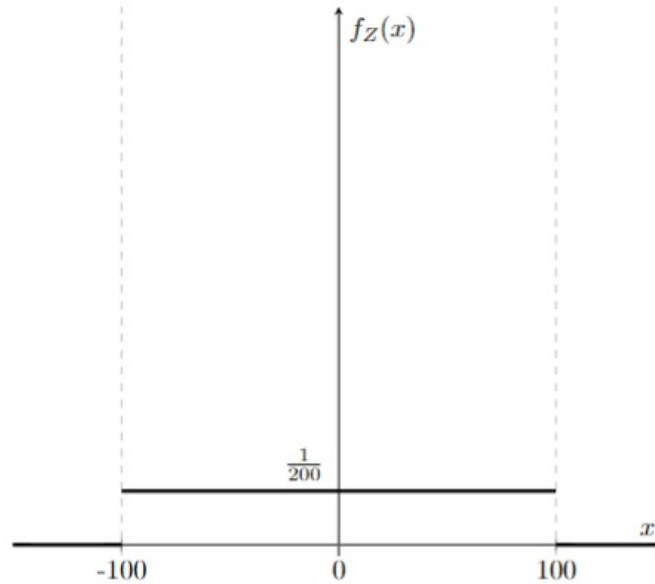


Figura 7:

- (a) ahora la distribución está “dispersa” en el conjunto $\{0, 1, \dots, 5\}$;
- (b) a pesar de que hay solamente dos valores “realmente probables”, $S_5 = 4$ y $S_5 = 5$, no estamos muy seguros de cuál podría aparecer.

Lo anterior se refleja en la entropía ya que en este caso, aplicando (2), fácilmente obtenemos:

$$H(\mathcal{D}_{S_5}) \approx 0.61285,$$

lo cual es mayor que la entropía en (5).

3. Densidades de máxima entropía

Para un entero dado $m \geq 1$, sea \mathbb{R}^m el espacio euclidiano m -dimensional, y sea \mathcal{B}_m la σ -álgebra de Borel de subconjuntos de \mathbb{R}^m . Sea también $f : \mathbb{R}^m \rightarrow [0, \infty)$ una densidad (de algún vector aleatorio absolutamente continuo \bar{X}).

Definición 2. De acuerdo con (1), la **entropía** de f se define como el siguiente número (finito o no)

$$H(f) \stackrel{\text{def}}{=} - \int_{\mathbb{R}^m} f(\bar{x}) \log[f(\bar{x})] d\bar{x}, \quad (6)$$

siempre que la integral en (6) esté bien definida.

Como siempre, en (6) $0 \log(0) \stackrel{\text{def}}{=} 0$. Las demostración de todas las afirmaciones en esta sección pueden encontrarse, por ejemplo, en el trabajo [5].

3.1. Caso I: Densidades en un conjunto acotado

Sea $B \in \mathcal{B}_m$ un subconjunto acotado de \mathbb{R}^m dado, con volumen (medida de Lebesgue) $\ell(B)$.

Definición 3. Se dice que una densidad f_{unif} es **uniforme** en B si

$$f_{unif}(\bar{x}) = \begin{cases} 1/\ell(B), & \text{si } \bar{x} \in B, \\ 0, & \text{de otro modo.} \end{cases}$$

En la Figura 8 se muestra la gráfica de f_{unif} para el caso $m = 2$.

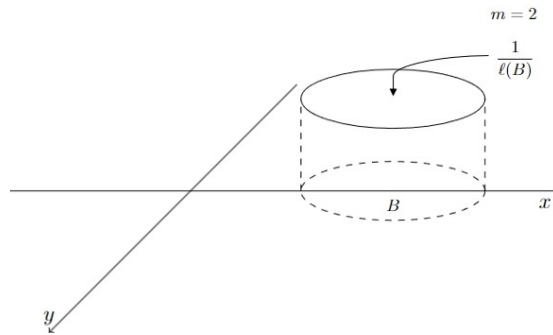


Figura 8:

Por (6), la entropía es

$$H(f_{unif}) = - \int_B 1/\ell(B) \log[1/\ell(B)] d\bar{x} = \log[\ell(B)],$$

i.e.

$$H(f_{unif}) = \log[\ell(B)]. \quad (7)$$

Particularmente, en la recta real, si $B = (a, b)$, entonces

$$H(f_{unif}) = \log(b - a). \quad (8)$$

Consideremos la familia \mathcal{D}_B de todas las densidades $f : B \rightarrow [0, \infty)$ con soporte el conjunto B , i.e. $f(x) = 0$ para cada $x \notin B$.

Proposición 4. Para cada $f \in \mathcal{D}_B$, $H(f) \leq H(f_{unif}) = \log[\ell(B)]$. Además, para $f \in \mathcal{D}_B$, $H(f) = \log[\ell(B)]$ si y solo si $f = f_{unif}$ (en “casi todos los puntos de B ”).

Esta proposición afirma que entre todas las densidades con soporte en un conjunto acotado fijo, la densidad uniforme tiene entropía máxima.

Observación 5. La Proposición 4 y (7) dan argumentos adicionales a favor de la segunda ley de la termodinámica, y explican el cambio en la distribución del gas, por ejemplo, en las Figuras 1.1 y 1.2 en la Sección 1. Mas aún, de (7) vemos que la distribución uniforme en un volumen grande tiene mayor entropía.

3.2. Caso II: Densidades en \mathbb{R}^m con promedio y una matriz de covarianza fijas

Sea $\bar{X} = (X_1, \dots, X_m)$ un vector aleatorio (con valores en \mathbb{R}^m) tal que $E\|\bar{X}\|^2 < \infty$. Entonces, el promedio $\bar{a} = (EX_1, \dots, EX_m)$ y la matriz de covarianza Σ con elementos $Cov(X_i, X_j) \stackrel{\text{def}}{=} E[(X_i - EX_i)(X_j - EX_j)]$, $i, j = 1, 2, \dots, n$ están bien definidos.

Fijemos un vector arbitrario $\bar{a} \in \mathbb{R}^m$ y una matriz de $m \times m$ simétrica y positiva definida Σ , y denotemos por $\mathcal{D}_{\bar{a}, \Sigma}$ a la familia de todas las densidades de vectores aleatorios de cuadrado integrables continuos en \mathbb{R}^m con promedio \bar{a} y matriz de covarianza Σ . Sea también

$$f_N(\bar{x}) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\bar{x} - \bar{a})^T \Sigma^{-1} (\bar{x} - \bar{a}) \right\} \quad (9)$$

la densidad Normal del vector aleatorio Normal \bar{X} con promedio \bar{a} y matriz de covarianza Σ .

En (9), $|\Sigma|$ es el *determinante* de Σ y Σ^{-1} es la matriz inversa. Aplicando (6), no es difícil calcular la entropía de f_N lo cual se establece en el siguiente resultado.

Proposición 6. *La entropía de la densidad f_N está dada por*

$$H(f_N) = \frac{1}{2} [m + \log(2\pi |\Sigma|)]. \quad (10)$$

Proposición 7. *Para cada densidad $f \in \mathcal{D}_{\bar{a}, \Sigma}$, $H(f) \leq H(f_N)$. Además, para $f \in \mathcal{D}_{\bar{a}, \Sigma}$, $H(f) = H(f_N)$ (dado en (10)) si y solo si $f = f_N$ (ver (9)).*

Por lo tanto, de nuevo, en la clase de todas las densidades con promedio y matriz de covarianza fijos, la **densidad Normal** tiene **entropía máxima**.

El caso uno-dimensional es el más importante para nosotros. La fórmula (9) resulta ser:

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (11)$$

Si $X \sim \text{Norm}(a, \sigma)$ es la variable aleatoria Normal con promedio $a = EX$ y varianza $\text{Var}(X) = \sigma^2$, i.e. $f_X = f_N$, entonces podemos decir que dadas las restricciones anteriores, la variable aleatoria X se interpreta como la variable aleatoria “más incierta” (o, informalmente, “la más aleatoria”).

En vista de (10), la entropía de $f_N = f_X$ con $X \sim \text{Norm}(a, \sigma)$ es

$$H(f_N) = \frac{1}{2} + \log(\sqrt{2\pi}\sigma), \quad (12)$$

y para la variable aleatoria Normal estándar: $\eta \sim \text{Norm}(0, 1)$ (i.e. con $a = 0$ y $\sigma = 1$)

$$H(f_\eta) = \frac{1}{2} + \log(\sqrt{2\pi}) \approx 1.41894. \quad (13)$$

Observación 8. (a) La Proposición 7 arroja algo de luz sobre la presencia de la densidad Normal (de una velocidad) ya que se sostiene el principio general de que en la termodinámica los estados de equilibrio tienen entropía máxima.

(b) Quizá, es la propiedad de maximalidad de entropía de las densidades Normales lo que puede explicar (al menos parcialmente) la prevalencia de cantidades normalmente distribuidas en la naturaleza y la ciencia. (Ver también la Sección 4.1 sobre la conexión entre el teorema central del límite y el crecimiento de entropía.)

3.3. Caso III: Densidades en $(0, \infty)$ con un promedio fijo

Sea $\lambda > 0$ un número dado, y sea \mathcal{D}_λ la familia de todas las densidades f_X de variables aleatorias continuas **positivas e integrables** tales que $EX = 1/\lambda$. Uno de los miembros de \mathcal{D}_λ es la densidad de la **variable aleatoria exponencial** $X_{Exp} \sim Exp(\lambda)$ con el parámetro $\lambda > 0$. Esto es:

$$f_{Exp}(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

cuya gráfica se muestra en la Figura 9.

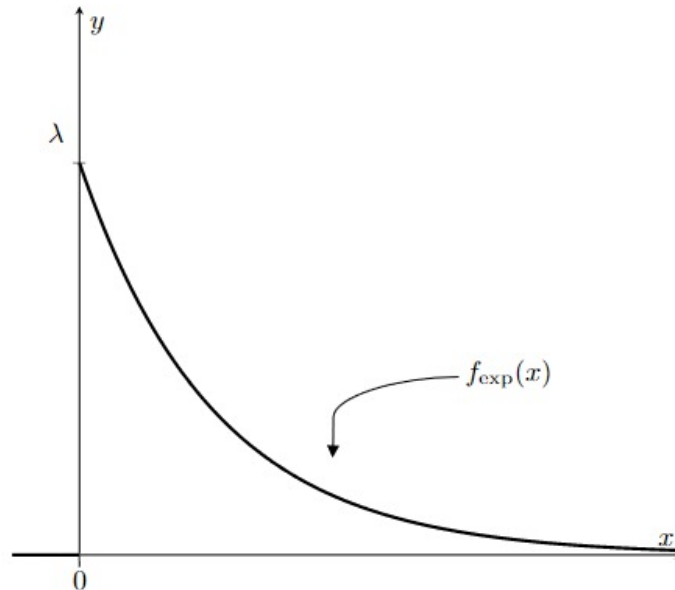


Figura 9:

Proposición 9. Para cada $f \in \mathcal{D}_\lambda$, $H(f) \leq H(f_{Exp})$. Además, para $f \in \mathcal{D}_\lambda$, $H(f) = H(f_{Exp})$ si y solo si $f = f_{Exp}$.

Por lo tanto, la *densidad exponencial* es de *máxima entropía* en la clase \mathcal{D}_λ , la cual, por

(6), está dada como:

$$H(f_{Exp}) = - \int_0^{\infty} \lambda e^{-\lambda x} \log(\lambda e^{-\lambda x}) dx = -\log \lambda + 1, \text{ i.e.}$$

$$H(f_{Exp}) = 1 - \log \lambda. \quad (14)$$

Ejemplo 10. Sea $X \sim Exp(\lambda = 1)$ (ver Figura 10),

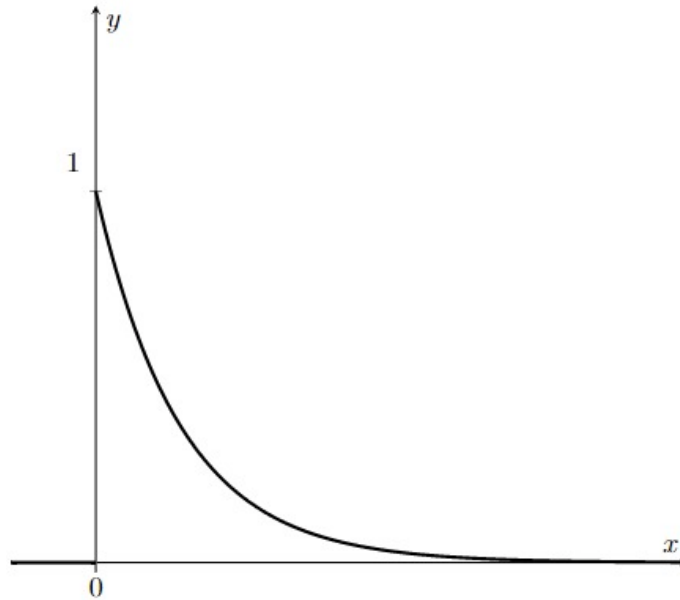


Figura 10:

y supóngase que la variable aleatoria Y tiene densidad $f_Y(x) = 0.5e^{-|x|}$, $x \in \mathbb{R}$. Observe que $f_Y \notin \mathcal{D}_\lambda$ (ver Figura 11).

Por (14), $H(f_X) = 1$ y con cálculos simples, $H(f_Y) \approx 1.6932$. Se puede observar de las Figuras 10 y 11 que comparado con Y , la variable aleatoria X es “menos incierta”.

4. Densidades de máxima entropía y algunos teoremas límite

Los teoremas del límite central y otros teoremas límite asociados con ciertos esquemas de suma de variables (o vectores) aleatorias juegan un papel prominente no solo en la Teoría de Probabilidad moderna, si no también en estadística matemática, la teoría de procesos estocásticos y sus numerosas aplicaciones.

Como sabemos de cálculo, la suma de una serie convergente puede ser cualquier número. Cuando se suman (y normalizan apropiadamente) variables aleatorias independientes, las distribuciones límite comúnmente resultan ser **universales** (i.e. las mismas para diferentes

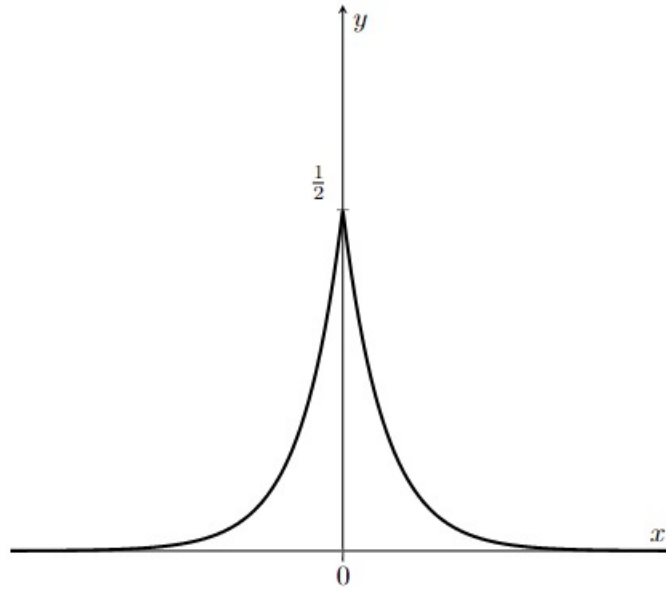


Figura 11:

distribuciones de sumandos). Es incluso más curioso que esas distribuciones (densidades) límite tienen entropía máxima en clases específicas de densidades. Podría decirse que la suma “acumula (en cierto sentido) incertidumbres” y las lleva a la densidad del límite. Esto se compara con la segunda ley de la termodinámica.

4.1. Teorema Central del Límite y crecimiento de entropía

Una de las formas más simples del Teorema Central del Límite (TCL) es la siguiente.

Sean X_1, X_2, \dots variables aleatorias i.i.d. tales que $\sigma^2 = \text{Var}(X_i) \in (0, \infty)$. Denotando $a = EX_1 (= EX_2 = EX_3 = \dots)$, sean $S_n = X_1 + X_2 + \dots + X_n$, $n = 1, 2, \dots$, y

$$Y_n = \frac{S_n - na}{\sigma\sqrt{n}}, \quad n \geq 1, \quad (15)$$

éstas últimas llamadas sumas normalizadas. Sea también $\eta \sim N(0, 1)$ la variable aleatoria Normal estándar. Entonces la “versión canónica” del TCL dice:

Para cada $x \in \mathbb{R}$, cuando $n \rightarrow \infty$, $F_{Y_n}(x) \rightarrow F_\eta(x)$, donde F_{Y_n} y $F_\eta(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ son las funciones de distribución correspondientes.

Es bien sabido que suponiendo que X_1 tiene una densidad, bajo ciertas hipótesis adicionales obtenemos la *convergencia de densidades* (ver e.g. [10]), esto es:

- ya sea

$$f_{Y_n}(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}, \quad \text{cuando } n \rightarrow \infty, \quad (16)$$

- o (bajo otras hipótesis complementarias) tenemos la convergencia en la llamada norma L_1 :

$$\int_{-\infty}^{\infty} |f_{Y_n}(x) - f_{\eta}(x)| dx \rightarrow 0 \text{ cuando } n \rightarrow \infty. \quad (17)$$

En (16) y (17), f_{Y_n} es la densidad de la variable aleatoria Y_n y

$$f_{\eta}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (18)$$

es la densidad de la variable aleatoria Normal estándar.

Para simplificar la notación (y, de hecho, sin pérdida de generalidad), a partir de ahora supondremos que $a = 0$ y $\sigma^2 = 1$. Entonces las sumas normalizadas en (15) tomarán la forma:

$$Y_n = \frac{S_n}{\sqrt{n}}, \quad n = 1, 2, \dots \quad (19)$$

Además, supongamos que la variable aleatoria X_1 tiene una densidad. Como antes, la **entropía** de Y_n es

$$H(f_{Y_n}) \stackrel{\text{def}}{=} - \int_{-\infty}^{\infty} f_{Y_n}(x) \log[f_{Y_n}(x)] dx. \quad (20)$$

Para establecer el siguiente resultado usaremos la siguiente definición (ver, e.g., [9]).

Definición 11. Sean f y g densidades definidas en \mathbb{R} tal que si $f > 0$ entonces $g > 0$. Definimos la entropía relativa de f con respecto a g como

$$H(f|g) = \int_{\mathbb{R}} f(x) \log \left[\frac{f(x)}{g(x)} \right] dx,$$

cuando la integral existe.

Un resultado que relaciona la entropía relativa con las densidades es la llamada desigualdad de Pinsker-Csiszár-Kullback (ver [11]):

$$\int_{\mathbb{R}} |f(x) - g(x)| dx \leq \sqrt{2H(f|g)}. \quad (21)$$

Ahora, observe que la entropía relativa de Y_n con respecto a la densidad Normal f_{η} en (18) es

$$H(f_{Y_n}|f_{\eta}) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_{Y_n}(x) \log \left[\frac{f_{Y_n}(x)}{f_{\eta}(x)} \right] dx. \quad (22)$$

En 1986, A. Barron demostró la siguiente “versión entrópica” del TCL.

Teorema 12. (Ver [2])

(a) Si para algún $n_0 \geq 1$, $H(f_{Y_{n_0}}|f_\eta)$ es **finito**, entonces, cuando $n \rightarrow \infty$

$$H(f_{Y_n}|f_\eta) \rightarrow H(f_\eta|f_\eta) = 0. \quad (23)$$

(b) Si para algún $n_0 \geq 1$, $H(f_{Y_{n_0}})$ es **finito**, entonces, cuando $n \rightarrow \infty$,

$$H(f_{Y_n}) \rightarrow H(f_\eta) = 0.5 + \log(\sqrt{2\pi}). \quad (24)$$

Observacion 13. En (15), $EY_n = \frac{1}{\sqrt{n}}E(X_1 + \dots + X_n) = 0$, $n = 1, 2, \dots$, y $Var(Y_n) = \frac{1}{n}Var(X_1 + \dots + X_n) = 1$, $n = 1, 2, \dots$. Además, $E\eta = 0$ y $Var(\eta) = 1$. Por lo tanto (ver Proposición 7), (24) afirma que cuando $n \rightarrow \infty$, la entropía de f_{Y_n} se acerca a la **entropía máxima** en la clase de densidades de variables aleatorias con media cero y varianza unitaria.

De (23) y la desigualdad (21) se obtiene la L_1 -convergencia de las densidades como se establece en el siguiente resultado.

Corolario 14. Bajo las condiciones de la parte (a) del Teorema 12,

$$\int_{-\infty}^{\infty} |f_{Y_n}(x) - f_\eta(x)| dx \rightarrow 0 \text{ cuando } n \rightarrow \infty. \quad (25)$$

Observacion 15. (Ver [2]) (a) Las condiciones del Teorema 12(a) y el Corolario 14 se satisfacen si para algún n la densidad f_{Y_n} es acotada.

(b) Existe una variable aleatoria X_1 tal que f_{Y_n} es no acotada para cada $n \geq 1$, pero (24) y (25) son verdaderas, a pesar de que la densidad f_η es acotada.

La siguiente pregunta interesante es si la convergencia de entropía en (24) es *monótona*. La respuesta positiva fue dada en el año 2004 por S. Arstein, K. Ball, F. Barthe y A. Naor (ver [1]).

Teorema 16. [1] Para cada $n = 1, 2, \dots$

$$H(Y_{n+1}) \geq H(Y_n) \quad (26)$$

o

$$H\left(\frac{S_{n+1}}{\sqrt{n+1}}\right) \geq H\left(\frac{S_n}{\sqrt{n}}\right).$$

De (26) vemos que en (24) la entropía de las densidades de sumas normalizadas converge monótonamente (i.e. no decrecientemente) a su máximo valor $H(f_\eta)$.

Ejemplo 17. Sea $X_1 \sim Unif(-\sqrt{3}, \sqrt{3})$ (la variable aleatoria uniforme en el intervalo $(-\sqrt{3}, \sqrt{3})$). Entonces $a = EX_1 = 0$; $\sigma^2 = Var(X_1) = 1$, y en (15) $Y_1 = \frac{X_1}{\sqrt{1}} = X_1$, con $H(Y_1) \approx 1.2425$ (ver (8)). En vista de los Teoremas 12 y 16, podemos ver que dado el “valor inicial” 1.2425, la entropía $H\left(\frac{S_n}{\sqrt{n}}\right)$ converge monótonamente a $H(f_\eta) \approx 1.41894$ (ver (13)).

Bajo las condiciones del Teorema 12 la entropía relativa $H(f_{Y_n}|f_\eta)$ tiende a cero, cuando $n \rightarrow \infty$. En un trabajo relativamente reciente [4], S. Bobkov, G. Chistyakov y F. Götze establecieron la tasa de convergencia. En efecto, si $EX_1^4 < \infty$ y la sucesión $\{H(f_{Y_n}|f_\eta)\}$ está acotada, entonces para alguna constante C se cumple

$$H(f_{Y_n}|f_\eta) \leq \frac{C}{n}, \quad n = 1, 2, \dots, \quad (27)$$

y, aplicando la desigualdad (21),

$$\int_{-\infty}^{\infty} |f_{Y_n}(x) - f_\eta(x)| dx \leq \frac{\sqrt{2C}}{\sqrt{n}}, \quad n = 1, 2, \dots \quad (28)$$

Nótese que la cota en (28) es la típica tasa de convergencia en el TCL.

4.2. Suma de variables aleatorias independientes en grupos. Convergencia a las distribuciones uniformes y crecimiento de entropía

En la Sección 3 vimos que una densidad uniforme tiene la entropía máxima entre las densidades con soporte un conjunto acotado. Hay una amplia clase de teoremas del límite que tratan con la suma de variables aleatorias en grupos compactos. Este es el caso cuando las densidades límite son uniformes. Solo consideraremos un esquema particular simple: *adición módulo 1* en el grupo (intervalo) $[0, 1) \subset \mathbb{R}$.

Consideremos al grupo (intervalo) $G = [0, 1)$ equipado con la operación \oplus que representa la adición módulo 1, i.e.

$$x \oplus y = \begin{cases} x + y & \text{if } x + y < 1, \\ x + y - 1 & \text{if } x + y \geq 1. \end{cases} \quad (29)$$

Entonces (G, \oplus) es un ejemplo simple de un *grupo conmutativo compacto*.

En este caso, la distribución, la densidad (si existe) y la independencia de variables aleatorias con valores en G se definen de manera usual, considerando a $G = [0, 1)$ como un subconjunto de Borel de \mathbb{R} . En este sentido, una función (medible) $f_X : [0, 1) \rightarrow [0, \infty)$ es la *densidad* de una variable aleatoria X en G si $P(X \in B) = \int_B f_X(x) dx$, para cada subconjunto de Borel $B \subset [0, 1)$. Particularmente, $f_{unif}(x) = 1$, $x \in [0, 1)$ es llamada la *densidad uniforme*.

Sean X_1, X_2, \dots variables aleatorias con valores en G . Usando la operación (29) se define la variable aleatoria $X_1 \oplus X_2$ en G . Luego, por inducción, la suma $S_n \stackrel{\text{def}}{=} X_1 \oplus X_2 \oplus \dots \oplus X_n$ se define para cada entero $n = 1, 2, \dots$.

No es difícil demostrar que si las variables aleatorias X_1, X_2, \dots son i.i.d. y X_1 tiene una densidad (i.e. X_1 es “absolutamente continua”), entonces para cada $n \geq 1$, la variable aleatoria S_n tiene una densidad, que se denota por f_{S_n} .

El siguiente teorema fue demostrado en 1972 por R. N. Bhattacharya (ver [3]).

Teorema 18. Sean X_1, X_2, \dots variables aleatorias i.i.d. en G con una densidad común f_X . Asíumase que hay una constante $\alpha > 0$ tal que

$$\inf_{x \in [0,1]} f_X(x) \geq \alpha. \quad (30)$$

Entonces:

$$\int_0^1 |f_{S_n}(x) - f_{unif}(x)| dx \equiv \int_0^1 |f_{S_n}(x) - 1| dx \leq 2(1 - \alpha)^n, \quad n = 1, 2, \dots \quad (31)$$

Observe que la condición (30), aunque es muy fuerte, garantiza que la tasa de convergencia en (31) de las densidades f_{S_n} a la uniforme, $f_{unif} \equiv 1$, sea geométrica, la cual es más rápida en comparación con la tasa en (28).

Ejemplo 19. Sea $f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & x \in (0, 1) \\ 1/2, & x = 0. \end{cases}$

Entonces en (30), $\alpha = 0.5$ y, en efecto, $\int_0^1 |f_{S_{30}}(x) - 1| dx \leq 0.00000000187$.

En los Teoremas 12 y 16 notamos que la convergencia de las densidades de las sumas normalizadas $Y_n = \frac{S_n}{\sqrt{n}}$ a la densidad Normal estándar $f_\eta(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (de máxima entropía) está acompañada por un *incremento monótono* de la entropía $H(Y_n)$ hacia su máximo valor $H(f_\eta)$. A partir de lo anterior, surge la siguiente pregunta: ¿Qué pasa con el crecimiento de entropía en el Teorema 18? Se tiene una respuesta a partir de la desigualdad inversa en (21). En efecto, en el artículo [11], bajo ciertas condiciones adicionales, la desigualdad contraria a (21) fue demostrada. Esto es, para alguna constante *positiva* γ :

$$\left\{ \int_{\mathbb{R}} |f(x) - g(x)| dx \right\}^2 \geq \gamma H(f|g). \quad (32)$$

Puede demostrarse que bajo las hipótesis del Teorema 18, las condiciones mencionadas para (32) se satisfacen. Por lo tanto, combinando (31) y (32), y observando que $H(f_{S_n}|f_{unif}) = -H(f_{S_n})$ obtenemos el siguiente resultado.

Proposición 20. Bajo las hipótesis del Teorema 18,

$$0 \geq H(f_{S_n}) \rightarrow H(f_{unif}) = 0, \quad \text{cuando } n \rightarrow \infty. \quad (33)$$

Vale la pena enfatizar nuevamente que en el lado derecho de (33) tenemos la *entropía máxima* de densidades en $[0, 1)$.

4.3. El Teorema de Rényi y la convergencia a la densidad exponencial

Observe que la densidad exponencial entra en una de las clases de densidades de máxima entropía que examinamos en la Sección 3. Nos preguntamos ahora si con esta densidad existen

resultados análogos a los teoremas límite analizados anteriormente. Para responder este punto presentaremos el bien conocido hecho de que una distribución exponencial es el límite débil de las llamadas *sumas geométricas*. Esto es parte del Teorema de Rényi publicado por el matemático Alfréd Rényi en 1956.

Sean X_1, X_2, \dots variables aleatorias no negativas i.i.d. tales que $EX_1^2 < \infty$ y $a \stackrel{\text{def}}{=} EX_1 > 0$. Sea también N una variable aleatoria *geométrica* con parámetro $p \in (0, 1)$. Asíumase que N es *independiente* de X_1, X_2, \dots . Recordemos que la variable aleatoria N es geométrica si $P(N = k) = p(1 - p)^{k-1}$, $k = 1, 2, \dots$.

Consideremos las siguientes *sumas geométricas*:

$$S_N \stackrel{\text{def}}{=} X_1 + X_2 + \dots + X_N. \quad (34)$$

Nótese que el número de términos en (34) es aleatorio y sigue la distribución geométrica.

Teorema 21 (Rényi). *Cuando $p \rightarrow 0$,*

$$pS_N \Rightarrow Y, \quad (35)$$

donde $Y \sim \text{Exp}(\lambda = 1/a)$ es la variable aleatoria con la densidad exponencial con parámetro $\lambda = 1/a$.

El símbolo “ \Rightarrow ” en (35) significa la convergencia *débil*, i.e. la *convergencia en distribución*. En el libro [8], Capítulo 3, V. Kalashnikov demostró un resultado más fuerte.

Teorema 22. *Bajo las hipótesis del Teorema 21 existe una constante $\beta < \infty$ tal que*

$$\sup_{x \in \mathbb{R}} |P(pS_N \leq x) - P(Y \leq x)| \leq \beta p, \quad p \in (0, 1). \quad (36)$$

En particular, el lado izquierdo de (36) tiende a cero cuando $p \rightarrow 0$.

Si adicionalmente suponemos que X_1 tiene una densidad f_{X_1} , entonces la variable aleatoria pS_N también tiene una densidad f_{pS_N} . Desafortunadamente, escapa a nuestro conocimiento bajo qué condiciones se puede afirmar que, cuando $p \rightarrow 0$,

$$\int_0^\infty |f_{pS_N}(x) - \frac{1}{a}e^{-x/a}| dx \rightarrow 0. \quad (37)$$

La convergencia en (36) sugiere que $P(pS_N \in I) \rightarrow P(Y \in I)$ uniformemente sobre **todos los intervalos** $I \subset \mathbb{R}$, mientras que (37) es equivalente a $P(pS_N \in B) \rightarrow P(Y \in B)$ uniformemente sobre todos los **subconjuntos de Borel** de \mathbb{R} . La última convergencia es referida como *convergencia fuerte*.

En el Teorema 12 y el Corolario 14 observamos que la convergencia fuerte está estrechamente relacionada con el *incremento de la entropía* de sumas de variables aleatorias a

la entropía máxima en las clases apropiadas. Ahora, en el contexto del Teorema de Rényi, podemos notar que por la igualdad de Wald

$$E(pS_N) = pEX_1EN = EX_1 = a.$$

Recordemos que (ver Sección 3) en la clase de todas las densidades en $(0, \infty)$ con un promedio fijo a , la densidad exponencial f_{exp} tiene la entropía máxima $H(f_{exp}) = 1 + \log(a)$ (ver (14)), obtenemos que para cada $p \in (0, 1)$, $H(f_{pS_N}) \leq H(f_{exp}) = 1 + \log(a)$. Sin embargo, no tenemos información sobre el siguiente hecho: $H(f_{pS_N}) \rightarrow H(f_{exp})$ cuando $p \rightarrow 0$. Tal vez este es un problema abierto.

5. Crecimiento de entropía y teoría de la información

La entropía en la teoría de la información guarda una conexión con la entropía termodinámica. Como lo hemos comentado anteriormente, en el contexto de la termodinámica, la entropía se basa en el análisis de un sistema de partículas con estados que representan posición y velocidad, y los cuales poseen una distribución de probabilidad específica. En la teoría de la información clásica (ver, e.g., [6]), para una distribución discreta \mathcal{D}_X , la *entropía de Shannon* $H_{Sh} = H$ se define mediante la misma ecuación (2), y eso estima la cantidad promedio de información contenida en los valores de la variable aleatoria X con distribución \mathcal{D}_X . Por ejemplo, como vimos en la Sección 2, si \mathcal{D}_X es la distribución de Bernoulli con $p = 0.5$, entonces al observar la realización del valor $X = 1$, hemos obtenido la información $H_{Sh} \approx 0.69315$. Pero, si $p = 0.99999$ entonces $H_{Sh} \approx 0.000125129$ (un “poco de información” ya que el valor $X = 1$, en este caso, es muy posible.)

Ilustraremos la observación anterior mediante el siguiente ejemplo (ver también [7]).

Sea $p \in (0, 1)$ un número dado y X_1, X_2, \dots variables aleatorias i.i.d. tomando valores en $\{-1, 1\}$, con distribución común

$$P[X_1 = 1] = p \text{ y } P[X_1 = -1] = 1 - p.$$

Además, definimos la variable aleatoria

$$Y_n := X_0 \cdot X_1 \cdot \dots \cdot X_n, \quad (38)$$

donde $X_0 = 1$. Es claro que la variable aleatoria Y_n también toma valores en $\{-1, 1\}$.

En determinado contexto, las variables aleatorias X_0, X_1, \dots y Y_n podrían tener la siguiente interpretación en los términos de la retransmisión múltiple de algún mensaje.

Supongamos que el evento $[X_0 = 1]$ representa que alguna persona ha recibido un mensaje completamente veráz que solo acepta significados de “sí” o “no”. Este mensaje es transmitido a otro sitio, ya sea correctamente, lo cual se representa con el evento $[X_1 = 1]$, con probabilidad p , o cambiando su contenido al lado opuesto representado por el evento $[X_1 = -1]$, con probabilidad $1 - p$. Ahora, este primer receptor retransmite el mensaje bajo el mismo procedimiento, esto es,

- con probabilidad p el mensaje es retransmitido en la forma en la cual ha sido recibido, lo cual se representa por el evento $[X_2 = 1]$;
- con probabilidad $1 - p$ el mensaje es retransmitido cambiando su contenido a lo contrario, representado por el evento $[X_2 = -1]$.

Una vez que el segundo receptor recibe el mensaje, este procedimiento se repite una y otra vez.

Bajo este escenario, observe que $[Y_n = 1]$ representa el evento de que el mensaje ha sido recibido correctamente después de n retransmisiones. Entonces es importante analizar el comportamiento de Y_n para n suficientemente grande y la relación con su entropía. Para este fin, primero calculemos la distribución de Y_n .

Observe que

$$EX_1 = 1 \cdot p + (-1)(1 - p) = 2p - 1.$$

Además, por la independencia de las variables aleatorias X_1, X_2, \dots ,

$$EY_n = EX_1 EX_2 \cdots EX_n = (2p - 1)^n. \quad (39)$$

Por otro lado, sea $p_n = P[Y_n = 1]$ y $1 - p_n = P[Y_n = -1]$. Entonces

$$EY_n = 1 \cdot p_n + (-1)(1 - p_n) = 2p_n - 1. \quad (40)$$

Combinando (39) y (40) tenemos

$$2p_n - 1 = (2p - 1)^n,$$

con lo cual obtenemos la distribución de Y_n dada por

$$P[Y_n = 1] = p_n = \frac{(2p - 1)^n + 1}{2} \quad (41)$$

y

$$P[Y_n = -1] = 1 - p_n = \frac{1 - (2p - 1)^n}{2}, \quad n = 1, 2, \dots \quad (42)$$

Ahora, como $p \in (0, 1)$, se cumple $|2p - 1| < 1$. Entonces, de (41) y (42), cuando $n \rightarrow \infty$,

$$P[Y_n = 1] \rightarrow \frac{1}{2} \quad \text{y} \quad P[Y_n = -1] \rightarrow \frac{1}{2}, \quad (43)$$

con una tasa de convergencia exponencial.

Observe que este comportamiento es el mismo para cada valor de $p \in (0, 1)$. Entonces, el n -ésimo receptor, i.e., el último receptor, para n suficientemente grande, conocerá la

veracidad o la falsedad del mensaje con la misma probabilidad $1/2$, es decir con completa incertidumbre.

Analicemos esto último desde el punto de vista de la entropía considerando el siguiente caso particular con $p = P[X_1 = 1] = 0.9999$. Observe (ver (2)) que la entropía de la variable aleatoria $Y_1 = X_1$ es aproximadamente 0.00082108 , muy cercana a cero. De acuerdo a (43), cuando $n \rightarrow \infty$, la entropía de Y_n crece a su máximo valor $-\log(0.5) \approx 0.69315$, que es la entropía de la variable aleatoria

$$Y_\infty = \begin{cases} 1, & \text{con probabilidad } 1/2; \\ -1 & \text{con probabilidad } 1/2, \end{cases}$$

la cual, como era de esperarse, expresa completa incertidumbre.

A partir de los resultados presentados en el ejemplo previo, resulta relevante compararlos con los resultados de las secciones anteriores. Primero, la situación descrita presenta una clara analogía en el ámbito de la física. En efecto, cuando un sistema de gas aislado pasa a un estado de equilibrio termodinámico, la distribución límite de moléculas, uniforme en posición y normal en velocidad, no tiene memoria de la distribución original. Observe que en los teoremas límite, como el teorema del límite central, también se evidencia cierta pérdida de información. Es decir, la distribución límite normal estándar carece de información sobre las distribuciones específicas de las variables aleatorias involucradas en la suma, a excepción de la media y la varianza. Además, las distribuciones límite alcanzan entropía máxima en determinadas clases, y el proceso de transición a dicho límite conlleva un aumento en la entropía. Todos estos aspectos se reflejan en el ejemplo a partir de que independientemente del valor de p , la distribución límite es la misma, y es de entropía máxima.

Agradecimientos. Este trabajo fue apoyado por el Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT - MÉXICO) bajo el proyecto Ciencia de Frontera CF-2019 87787.

Referencias

- [1] Sh. Artstein, K. Ball, F. Barthe and A. Naor, Solution of Shannon's problem on the monotonicity of entropy, *Journal of the American Mathematical Society*, 17, 975-982 (2004).
- [2] A. R. Barron, Entropy and the central limit theorem, *The Annals of Probability*, 14, 336-342 (1986).
- [3] R. N. Bhattacharya, Speed of convergence of n-th fold convolution of a probability measure on a compact group, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 25, 1-10 (1972).
- [4] S. G. Bobkov, G.P. Chistyakov and F. Götze, Berry-Essen bounds in the entropic central limit theorem, *Probab. Theory Relat Fields*, 435-478 (2014).
- [5] K. Conrad, Probability distributions and maximal entropy, Expository paper (2013)
URL: <https://kconrad.math.uconn.edu/blurbs/analysis/entropypost.pdf>
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, Wiley, 2006.
- [7] E. Gordienko and X. I. Popoca Jimenez, *Introducción a la Teoría de Probabilidad y Métricas Probabilísticas con Aplicaciones en Seguros y Finanzas*, Instituto de Matemáticas UNAM, México, 2018.
- [8] V. V. Kalashnikov, *Geometric Sums: Bounds for Rare Events with Applications*, Kluwer Academic Publishers, 1997.
- [9] M. C. Mackey, *Time's Arrow: The Origins of Thermodynamic Behavior*, Springer-Verlag, New York, 1992.
- [10] V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, Berlin, 1975.
- [11] I. Sason, On reverse Pinsker inequalities, *IEEE. Trans. on information Theory*, 61 (2015).

Como citar este artículo: E. I. Gordienko y J. A. Minjárez-Sosa, "Distribuciones de máxima entropía, incremento de aleatoriedad y teoremas límite en probabilidad", *Sahuarus. Revista Electrónica de Matemáticas*, vol. 8, no. 1, pp. 23-44, 2024. <https://doi.org/10.36788/sah.v8i1.146>.