UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

FACULTAD DE CIENCIAS

"GRANDES DESVIACIONES Y ENTROPÍA"

QUE PARA OBTENER EL GRADO ACADÉMICO DE

MAESTRO EN CIENCIAS

P R E S E N T A

SERGIO IVÁN LÓPEZ ORTEGA

DIRECTORA DE LA TESINA: DRA: ANA MEDA GUARDIOLA

MÉXICO, D.F. ENERO, 2009

Índice

Grandes desviaciones y entropía

Sergio Iván López Ortega,

30 de enero de 2009.

Resumen

En este documento damos una breve introducción a las Grandes Desviaciones, al concepto de Entropía (desde el punto de vista de teoría ergódica), y finalmente un esbozo de algunas relaciones entre ambos conceptos.

La Teoría de Grandes desviaciones y el concepto de Entropía tienen como punto en común que son utilizadas para dar una medición de la cantidad de azar en un sistema. La Entropía nos proporciona una estimación puntual y la teoría de Grandes Desviaciones una función tasa. La relación entre ellas es el objetivo de esta tesina.

1. Teoría de Grandes Desviaciones. Introducción.

En adelante $\{X_i\}_{i\in\mathbb{N}}$ denotará una sucesión de variables aleatorias reales, independientes e idénticamente distribuidas (v.a.i.i.d.). Existen en probabilidad dos teoremas fundamentales que estudian el comportamiento asintótico de la sucesión de sumas parciales $\{S_n = \sum_{i=1}^{n} X_i\}_{n\in\mathbb{N}}$. Cuando X_1 tiene esperanza μ y varianza σ^2 finitas, se tiene:

1. **La Ley Fuerte de los Grandes Números**

$$\frac{S_n}{n} \longrightarrow \mu \quad \mathbb{P}-\ c.s. \quad \textbf{[LFG]}$$

2. **El Teorema de Límite Central**

$$\frac{\sqrt{n}}{\sigma}\Big(\frac{S_n}{n} - \mu\Big) \xrightarrow{w} Z \quad \textbf{[TLC]}$$

donde Z es una variable con distribución normal estándar. (Existen versiones de estos teoremas con supuestos más débiles, pero en esta tesina trabajamos con variables aleatorias con generadora de momentos finita, lo que garantiza las condiciones anteriores).

En términos simples, podemos pensar que la Ley Fuerte de los Grandes Números nos asegura la convergencia de la variable $\frac{S_n}{n}$ a μ, y el Teorema de Límite Central la velocidad de convergencia u orden en que esto ocurre (notando que la convergencia es en el sentido débil). A los eventos en los cuales ocurren desviaciones de la variable S_n respecto a su media $n\,\mu$ de este orden se les llama desviaciones normales, y su probabilidad se puede cuantificar por medio del [TLC]. Un ejemplo para $a > 0$ y $\mu < \infty$ es el siguiente:

$$\Big\{S_n \geq n\,\mu + a\sqrt{n}\Big\}.$$

Por el contrario, a los eventos en los que ocurren desviaciones respecto a la media sin el factor $\sqrt{n}$ se les llama *Grandes Desviaciones.*

Un ejemplo es el *evento raro*

$$\Big\{\frac{S_n}{n} \geq \mu + a\Big\},$$

donde $\mu < \infty$ y $a > 0$, cuya probabilidad tiende a 0 cuando n tiende a infinito (LFGN). El objetivo es cuantificar la tasa a la que ocurre esta disminución. Veremos que en este caso, bajo cierta condición en la cola de la función de distribución de X_1, el decaimiento es exponencial en n, es decir:

$$\mathbb{P}\Big(\frac{S_n}{n} \geq \mu + a\Big) \approx \exp\{-n\, I(a)\}$$

donde I es una función tasa explícita dependiente de la distribución de la variable X_1. Escrito de otra forma

$$\lim_{n\to\infty} \frac{1}{n} \log\, \mathbb{P}(S_n \geq (\mu + a)n) = -I(a) < 0$$

para $a > 0$ y una función tasa I. (En la siguiente sección definiremos lo que es una función tasa formalmente, por ahora nos conformamos con la noción intuitiva mostrada arriba).

El estudio de Grandes Desviaciones no se reduce a analizar desviaciones de la variable aleatoria $\frac{S_n}{n}$; de la misma forma se consideran otros funcionales de $X_1, ..., X_n$ tales como la medida empírica (medida aleatoria) $L_n = \frac{1}{n}(\delta_{X_1} + ... + \delta_{X_n})$ donde δ_x es la medida puntual en el punto x. El grado de generalidad al que nos referiremos se debe a S.R.S. Varadhan desarrollado en [Var1] para medidas de probabilidad sobre un espacio polaco.

Antes de abordar el contexto general de Grandes Desviaciones hagamos algunos cálculos explícitos para mostrar el tipo de resultados que se esperan obtener y generalizar en el estudio de las grandes desviaciones.

La siguiente definición nos permite hacer una comparación para la convergencia de dos sucesiones en el sentido de la convergencia de las tasas de una gran desviación.

Definición 1. *Sean* $\{a_n\}_{n\in\mathbb{N}}, \{b_n\}_{n\in\mathbb{N}}$ *sucesiones de números positivos. Decimos que* $\{a_n\}_{n\in\mathbb{N}}$ *y* $\{b_n\}_{n\in\mathbb{N}}$ *son logarítmicamente equivalentes, denotado por* $\{a_n\}_{n\in\mathbb{N}} \approx \{b_n\}_{n\in\mathbb{N}}$*, si*

$$\lim_{n\to\infty} \frac{1}{n}(\log\, a_n - \log\, b_n) = 0.$$

Tenemos algunas observaciones inmediatas.

Observación 1. *Si* $\{a_n\}_{n\in\mathbb{N}} \approx \{b_n\}_{n\in\mathbb{N}}$*, entonces*

(a). $\limsup_{n\to\infty} \frac{1}{n} \log\, a_n = \limsup_{n\to\infty} \frac{1}{n} \log b_n$,

(b). $\liminf_{n\to\infty} \frac{1}{n} \log\, a_n = \liminf_{n\to\infty} \frac{1}{n} \log b_n$.

Demostración:

(a). Sean $\{c_n\}_{n\in\mathbb{N}}$, $\{d_n\}_{n\in\mathbb{N}}$ sucesiones de números reales tales que

$$\lim c_n - d_n = 0.$$

Entonces

$$0 = \limsup_{n\to\infty}(c_n - d_n) \geq \limsup_{n\to\infty} c_n + \liminf_{n\to\infty}(-d_n) = \limsup_{n\to\infty} c_n - \limsup_{n\to\infty} d_n,$$

lo que implica que $\limsup_{n\to\infty} c_n \geq \limsup_{n\to\infty} d_n$. Tenemos también que $\lim_{n\to\infty} d_n - c_n = 0$, así que por simetría se cumple la desigualdad inversa. Poniendo $c_n = \frac{1}{n}\log a_n$, $d_n = \frac{1}{n}\log b_n$ se sigue el resultado.
(b). La demostración es análoga al punto anterior.

$\square$

Observación 2. *Sean $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ dos sucesiones de números reales positivos. Entonces*

(a). $\{a_n + b_n\}_{n\in\mathbb{N}} \approx \{a_n \vee b_n\}_{n\in\mathbb{N}}$*, donde $s \vee t$ denota al máximo entre s y t.*

(b). $\limsup_{n\to\infty} \frac{1}{n}\log(a_n + b_n) = \limsup_{n\to\infty} \frac{1}{n}\log(a_n \vee b_n)$

(c). $\liminf_{n\to\infty} \frac{1}{n}\log(a_n + b_n) = \liminf_{n\to\infty} \frac{1}{n}\log(a_n \vee b_n)$

Demostración:

(a). Sean $\{a_n\}_{n\in\mathbb{N}}, \{b_n\}_{n\in\mathbb{N}}$ sucesiones de números positivos. Tenemos que para n fijo $a_n + b_n \leq 2(a_n \vee b_n)$, pues a_n y b_n son positivos. También tenemos que $\frac{a_n+b_n}{a_n\vee b_n} \geq 1$. Por lo tanto

$$1 \leq \lim_{n\to\infty} \inf \left(\frac{a_n + b_n}{a_n \vee b_n}\right)^{\frac{1}{n}} \leq \lim_{n\to\infty} \sup \left(\frac{a_n + b_n}{a_n \vee b_n}\right)^{\frac{1}{n}} \leq \lim_{n\to\infty} 2^{\frac{1}{n}} = 1.$$

Entonces

$$\lim_{n\to\infty} \frac{1}{n}(\log(a_n + b_n) - \log(a_n \vee b_n)) = 0.$$

(b) y (c). Se siguen de (a.) y la Observación (1).

$\square$

Esta observación nos dice que la magnitud de una desviación dada por la suma de dos exponentes está dada por la magnitud del exponente mayor; es decir, el crecimiento exponencial dado por $e^{(a_n+b_n)\,n}$ es exactamente el mismo que el crecimiento dado por $e^{(a_n\vee b_n)\,n}$. Claramente se puede extender esta propiedad para un número finito de exponentes.

El siguiente teorema nos muestra cómo calcular la función tasa asociada a la suma de variables aleatorias independientes reales. La función tasa es la transformada de Fenchel-Legendre del logaritmo de la función generadora de momentos de las variables aleatorias. La demostración es laboriosa pero sencilla debido a la independencia de las variables aleatorias.

Teorema 1 (Cramér). *Sea $\{X_i\}_{i\in\mathbb{N}}$ una sucesión de v.a.i.i.d. reales tales que $\phi(t) = \mathbb{E}(e^{tX_1}) < \infty$ para t en una vecindad que contiene al 0. Si $S_n = \sum_{i=1}^n X_i$ entonces para $a > \mathbb{E}(X_1)$ tenemos*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq a\, n) = -I(a)$$

donde

$$I(z) = \sup_{t\in\mathbb{R}} [zt - \log\, \phi(t)].$$

Demostración:

Supongamos válido el resultado para $\{Y_i\}_{i\in\mathbb{N}}$ v.a.'s con $\mathbb{E}(Y_1) < 0$ y $a = 0$, $T_n = \sum_{i=1}^n Y_i$. Sean $\{X_i\}_{i\in\mathbb{N}}$ v.a.'s con $a > \mathbb{E}(Y_1)$, $S_n = \sum_{i=1}^n X_i$. Ponemos $Y_i = X_i - a$, $T_n = \sum_{i=1}^n Y_i$; el supuesto nos dice que

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n \geq 0) = -\sup_{t\in\mathbb{R}} [-\log \phi_Y(t)].$$

Pero

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n \geq 0) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\sum_{i=1}^n (X_i - a) \geq 0\Big) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na)$$

y

$$-\sup_{t\in\mathbb{R}} [-\log \phi_Y(t)] = -\sup_{t\in\mathbb{R}} [-\log(e^{-at}\, \mathbb{E}(e^{tX}))] = -\sup_{t\in\mathbb{R}} [at - \log \phi_X(t)],$$

por lo que el resultado se cumple para la sucesión $\{X_i\}_{i\in\mathbb{N}}$. Por lo tanto podemos suponer S.P.G. que $a = 0$ y $\mathbb{E}(X_1) < 0$.

También podemos suponer que X_1 es no degenerada. Si X es degenerada tenemos que $\mathbb{P}(X_1 = c) = 1$ para algún número real c. Entonces

$$\begin{aligned} \phi(t) &= \mathbb{E}(e^{X_1 t}) = e^{ct} \\ \Rightarrow I(z) &= \sup_{t\in\mathbb{R}} [zt - ct] \end{aligned}$$

de donde se concluye que $I(c) = 0$ y $I(z) = \infty$ para $z \neq c$. Por otro lado, $S_n = nc$, $\mathbb{P} - c.s.$, y como $a > c = \mathbb{E}(X)$, tenemos $\mathbb{P}(S_n \geq na) = 0$. Entonces

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -\infty,$$

cumpliendo el teorema.

Definimos $\rho = \text{ínf}_{t\in\mathbb{R}} \phi(t)$. Notamos que

$$I(0) = \sup_{t\in\mathbb{R}} [-\log\, \phi(t)] = -\text{ínf}_{t\in\mathbb{R}} [\log\, \phi(t)] = -\log[\text{ínf}_{t\in\mathbb{R}} \phi(t)] = -\log\, \rho.$$

Así que debemos probar que

$$\lim_{n\to\infty} \frac{1}{n} \log\, \mathbb{P}(S_n \geq 0) = \log\, \rho.$$

Dependiendo de dónde esté la masa de $\mathbb{P}$, tenemos tres casos:

1. Cuando $\mathbb{P}(X_1 < 0) = 1$. Sean $t < s$ números reales. Como $X_1 < 0$ $\mathbb{P} - c.s.$, tenemos que

$$\begin{aligned} X_1\, t &> X_1\, s \qquad \mathbb{P} - c.s. \\ \Rightarrow e^{X_1\, t} &> e^{X_1\, s} \qquad \mathbb{P} - c.s. \\ \Rightarrow \phi(t) = \mathbb{E}(e^{X_1\, t}) &> \mathbb{E}(e^{X_1\, s}) = \phi(s). \end{aligned}$$

   Se concluye que ϕ es estrictamente decreciente. Por lo que $\lim_{t\to\infty} \phi(t) = \rho$. Calculamos explícitamente:

$$\lim_{t\to\infty} \phi(t) = \lim_{t\to\infty} \mathbb{E}(e^{tX_1}) = \mathbb{E}(e^{\lim_{t\to\infty} tX_1}) = \int_{\{X_1<0\}} e^{\lim_{t\to\infty} tX_1} d\,\mathbb{P} = 0.$$

   Por lo tanto $\rho = 0$ y como $\mathbb{P}(S_n \geq 0) = 0$, se tiene lo deseado.

2. Cuando $\mathbb{P}(X_1 \leq 0) = 1$ con $\mathbb{P}(X_1 = 0) > 0$. También aquí ϕ es decreciente por lo que $\lim_{t\to\infty} \phi(t) = \rho$. Y tenemos

$$\begin{aligned} \lim_{t\to\infty} \phi(t) &= \lim_{t\to\infty} \mathbb{E}(e^{tX_1}) \\ &= \mathbb{E}(e^{\lim_{t\to\infty} tX_1}) \\ &= \int_{\{X_1<0\}} e^{\lim_{t\to\infty} tX_1} d\,\mathbb{P} + \int_{\{X_1=0\}} e^{\lim_{t\to\infty} tX_1} d\,\mathbb{P} \\ &= \mathbb{P}(X_1 = 0), \end{aligned}$$

   por lo que $\rho = \mathbb{P}(X_1 = 0) > 0$. En este caso $\mathbb{P}(S_n \geq 0) = \mathbb{P}(S_n = 0) = \mathbb{P}(X_1 = ... = X_n = 0) = \rho^n$. Entonces

$$\lim_{n\to\infty} \frac{1}{n} \log\, \mathbb{P}(S_n \geq 0) = \lim_{n\to\infty} \frac{1}{n}\, \log\, \rho^n = \log\, \rho,$$

   y se cumple el teorema.

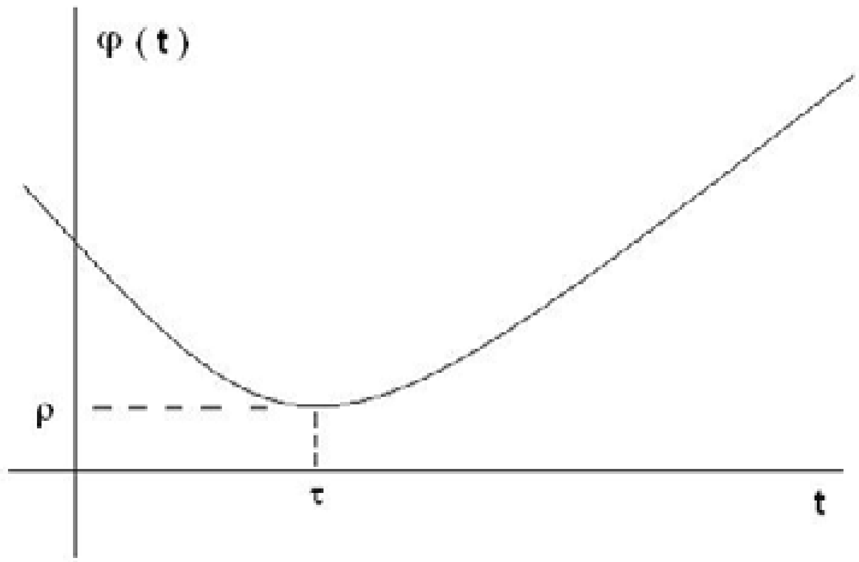
3. Cuando $\mathbb{P}(X_1 < 0), \mathbb{P}(X_1 > 0) > 0$. En este caso $\lim_{t\to\infty} \phi(t) = \lim_{t\to-\infty} \phi(t) = \infty$.

   Sea F la función de distribución de X_1. Debido a que la generadora de momentos de X_1 existe y es finita en una vecindad alrededor de $t = 0$ podemos usar los teoremas de convergencia de la teoría de integración de Lebesgue y tenemos las siguientes igualdades:

$$\begin{aligned} \phi'(t) &= \int_{\mathbb{R}} x\, e^{tx} d\, F(x) \\ \phi''(t) &= \int_{\mathbb{R}} x^2\, e^{tx} d\, F(x). \end{aligned}$$

   La segunda derivada es igual a una integral con integrando estrictamente positivo (excepto en 0) respecto a la función creciente F, por lo que $\phi''(t) > 0$. Por lo tanto, ϕ es una función convexa y $\phi'(0) = \mathbb{E}(X_1) < 0$.

   Debido a que ϕ es estrictamente convexa, entonces es continua en cualquier intervalo abierto, por lo que existe un único número real τ positivo que satisface $\phi(\tau) = \rho$ y $\phi'(\tau) = 0$.



Debido a la desigualdad de Chebyshev [Dur], tenemos la siguiente cota superior

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}(e^{\tau S_n} \geq 1) \leq \mathbb{E}(e^{\tau S_n}) = [\phi(\tau)]^n = \rho^n$$

y entonces

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \leq \log \rho.$$

La cota inferior resulta más difícil de obtener. Utilizaremos una técnica que se puede generalizar en el contexto abstracto de grandes desviaciones. Sea $\{\hat{X}_i\}_{i\in\mathbb{N}}$ una sucesión de variables aleatorias independientes con función de distribución

$$\hat{F}(x) = \frac{1}{\rho}\int_{(-\infty,x]} e^{\tau y} dF(y).$$

A la función $\hat{F}$ se le conoce como la transformada de Cramér de F. Notemos que $\rho = \phi(\tau) = \int_{\mathbb{R}} e^{\tau y} dF(y)$.

Afirmación 1. *Ocurre que* $\mathbb{E}(\hat{X}_1) = \hat{\mu} = 0$ *y* $\mathbb{V}ar(\hat{X}_1) = \hat{\sigma}^2 \in (0,\infty)$.

Sea $\hat{\phi}(t) = \mathbb{E}(e^{t\hat{X}_1})$. Entonces

$$\hat{\phi}(t) = \int_{\mathbb{R}} e^{tx} d\hat{F}(x) = \frac{1}{\rho}\int_{\mathbb{R}} e^{tx}\, e^{\tau x} dF(x) = \frac{1}{\rho}\,\phi(t+\tau) < \infty.$$

Lo cual implica que $\hat{\phi}$ es suave en una vecindad alrededor del 0 y

$$\begin{aligned}
\mathbb{E}(\hat{X}_1) &= \hat{\phi}'(0) = \frac{1}{\rho}\phi'(\tau) = \hat{\mu} = 0 \\
\mathbb{V}ar(\hat{X}_1) &= \hat{\phi}''(0) = \frac{1}{\rho}\phi''(\tau) \in (0,\infty).
\end{aligned}$$

Afirmación 2. *Sea* $\hat{S}_n = \sum_{i=1}^n \hat{X}_i$. *Entonces* $\mathbb{P}(S_n \geq 0) = \rho^n \, \mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}})$.

Para demostrar esta afirmación, basta invertir:

$$\begin{aligned}
\mathbb{P}(S_n \geq 0) &= \int_{\{x_1+\cdots+x_n \geq 0\}} dF(x_1) \cdots dF(x_n) \\
&= \int_{\{x_1+\cdots+x_n \geq 0\}} [\rho \, e^{-\tau \, x_1} d\hat{F}(x_1)] \cdots [\rho \, e^{-\tau \, x_n} d\hat{F}(x_n)] \\
&= \rho^n \int_{\mathbb{R}^n} e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}} d\hat{F}(x_1) \cdots d\hat{F}(x_n) \\
&= \rho^n \, \mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}).
\end{aligned}$$

Afirmación 3. $\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}) \geq 0.$

Por la Afirmación (1) podemos aplicar el [TLC] a la sucesión $\{\hat{S}_n\}$. Tomemos $C > 0$ tal que

$$\frac{1}{\sqrt{2\,\pi}} \int_0^C e^{\frac{-x^2}{2}} dx > \frac{1}{4}$$

y estimemos

$$\begin{aligned}
\mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}) &= \int_{\mathbb{R}} e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}} d\,\mathbb{P} \\
&\geq \int_{\hat{S}_n \in 0, C]} e^{-\tau \, \hat{S}_n} d\,\mathbb{P} \\
&= \int_{\hat{S}_n \in [0, C \, \hat{\sigma} \, \sqrt{n}\,]} e^{\frac{-\tau \, \hat{S}_n}{\hat{\sigma} \, \sqrt{n}}} d\,\mathbb{P} \\
&\geq e^{-\tau C \hat{\sigma} \sqrt{n}} \, \mathbb{P}\Big(\frac{\hat{S}_n}{\hat{\sigma} \sqrt{n}} \in [0, C) \Big)
\end{aligned}$$

donde $\mathbb{P}\Big(\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0, C)\Big) \geq \frac{1}{4}$ para n suficientemente grande. Entonces

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}) \geq \liminf_{n \to \infty} \frac{1}{n} \log \left(\frac{e^{-\tau C \hat{\sigma} \sqrt{n}}}{4} \right) = 0,$$

demostrando la afirmación.

Así concluimos el caso 3.; pues debido a la Afirmación (2) tenemos

$$\begin{aligned}
\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) &= \liminf_{n \to \infty} \frac{1}{n} \log \left(\rho^n \mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}) \right) \\
&= \log \rho + \liminf_{n \to \infty} \frac{1}{n} \log \left(\mathbb{E}(e^{-\tau \, \hat{S}_n} 1_{\{\hat{S}_n \geq 0\}}) \right) \\
&\geq \log \rho
\end{aligned}$$

donde la última desigualdad se debe a la Afirmación (3).

$\square$

Mediante el teorema anterior podemos calcular la función tasa para algunas funciones de distribución particulares. La calcularemos explícitamente para la función de distribución más sencilla: una función de distribución Bernoulli $(\frac{1}{2})$.

En este caso, $\phi_{X_1}(t) = \frac{e^t+1}{2}$ y para $a > \frac{1}{2}$ tenemos

$$\begin{aligned} I(a) &= \sup_{t\in\mathbb{R}} \Big[at - \log\, \phi(t)\Big] \\ &= \sup_{t\in\mathbb{R}} \Big[at - \log\, \Big(\frac{e^t+1}{2}\Big)\Big]. \end{aligned}$$

Y como la función dentro del paréntesis alcanza un máximo en $a(a-1)$, tenemos que

$$I(a) = a\, \log\, a + (1-a)\, \log\,(1-a) + \log\, 2$$

Lo notable en este caso es que $-I(a)$ será precisamente la entropía de una variable Bernoulli de parámetro a. Este tipo de relaciones no es casual y lo estudiaremos en la última sección de este trabajo.

1. Teoría de Grandes Desviaciones. Contexto general

La relación encontrada en el cálculo de la sección anterior aparece también en el contexto general. Planteemos primero este contexto.

Definición 1. *Sea (X, d) un espacio polaco. Decimos que $f : X \longrightarrow [-\infty, \infty]$ es semicontinua inferiormente si ocurre alguna de las siguientes condiciones:*

a. $\liminf_{n\to\infty} f(x_n) \geq f(x)$ *para toda sucesión* $x_n \to x$.

b. $\lim_{\varepsilon\to 0} \inf_{y\in B_\varepsilon(x)} f(y) = f(x)$.

c. Los conjuntos de nivel, que son conjuntos de la forma

$$f^{-1}([-\infty, c])$$

son cerrados, para toda $c \in \mathbb{R}$.

Observación 1. *Una función semicontinua inferiormente alcanza mínimo en compactos no vacíos.*

Demostración:

Sean f una función semicontinua inferiormente y $K \subseteq X$ un conjunto compacto no vacío. Sea $c = \inf_{x\in K} f(x)$. Debido a la definición de ínfimo, existe una sucesión $\{x_n\}_{n=1}^{\infty} \subseteq K$ tal que $f(x_n) \to c$.

CASO 1: $c > -\infty$. Entonces, para cada $n \in \mathbb{N}$, tenemos que el conjunto

$$A_n = \left\{x \in X : f(x) \leq c + \frac{1}{n}\right\} \cap K$$

es no vacío, por la existencia de la sucesión $\{x_n\}_{n=1}^{\infty} \subseteq K$ tal que $f(x_n) \to c$. Ahora, para cada $n \in \mathbb{N}$, $f^{-1}[-\infty, c + \frac{1}{n}]$ es cerrado, ya que f es semicontinua inferiormente, y como K es compacto se sigue que A_n es compacto. Entonces, por el teorema de Bolzano-Weierstrass, existe un punto x en la intersección de los compactos anidados $\{A_n\}_{n\in\mathbb{N}}$, el mínimo de f en K, por construcción.

CASO 2: $c = -\infty$. Este caso es análogo al anterior, basta considerar los conjuntos $A_n = \{x \in X : f(x) \leq n\} \cap K$.

□

Definición 2. *Sea (X, d) un espacio polaco. Decimos que una función $I : X \longrightarrow [0, \infty]$ es una función tasa si cumple que*

(a). I no es idénticamente ∞.

(b). I es semicontinua inferiormente.

(c). I tiene conjuntos de nivel compactos.

Claramente (b) es consecuencia de (c) pero es tradicional incluirlo en la definición. El punto (c) garantiza que la sucesión de probabilidades $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ sea *tensa*, es decir, para todo $M < \infty$ existe un compacto $K_M \subseteq X$ tal que

$$\lim_{n\to\infty}\sup \frac{1}{n}\log \mathbb{P}_n(X \setminus K_M) \leq -M;$$

se puede consultar en [Dem]. Como la condición (c) se puede relajar, en algunos libros la definición anterior se refiere a una *función tasa buena*.

Definición 3. *Una sucesión de medidas de probabilidad $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ en X satisfacen un principio de grandes desviaciones **[PGD]** (con orden n) y tasa I si*

a. I es una función tasa en el sentido definido anteriormente.

b. $\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n(C) \leq -\underline{I}(C)$ para todo $C \subseteq X$ cerrado.

c. $\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n(O) \geq -\underline{I}(O)$ para todo $O \subseteq X$ abierto.

donde $\underline{I}(S) = \inf_{x\in S} I(x)$ para $S \subseteq X$.

Observación 2. *Sea $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ una sucesión de medidas de probabilidad en X que satisfacen un PGD. Entonces la función tasa asociada es única.*

Demostración:

Sean I y J funciones tasa asociadas a $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$. Sea $x \in X$. Definamos $B_N = B_{\frac{1}{N}}(x)$ para $N \in \mathbb{N}$. Entonces

$$\begin{array}{rcll}
-I(x) & \leq & -\underline{I}(B_{N+1}) & \text{pues } x \in B_{N+1} \\
& \leq & \displaystyle\lim_{n\to\infty}\inf \frac{1}{n}\log \mathbb{P}_n(B_{N+1}) & \text{se cumple el (PGD) para } I \\
& \leq & \displaystyle\lim_{n\to\infty}\inf \frac{1}{n}\log \mathbb{P}_n(\bar{B}_{N+1}) & B_{N+1} \subseteq \bar{B}_{N+1} \\
& \leq & \displaystyle\lim_{n\to\infty}\sup \frac{1}{n}\log \mathbb{P}_n(\bar{B}_{N+1}) & \\
& \leq & -\underline{J}(\bar{B}_{N+1}) & \text{se cumple el (PGD) para } J \\
& \leq & -\underline{J}(B_N) & \bar{B}_{N+1} \subseteq B_N.
\end{array}$$

Por lo tanto $-I(x) \leq -\underline{J}(B_N)$ para cualquier natural N. Tomando el límite cuando N crece tenemos que

$$-I(x) \leq \lim_{N\to\infty} -\underline{J}(B_N) = -\lim_{N\to\infty}\inf_{y\in B_N(x)} J(y) = -J(x),$$

donde la última igualdad es consecuencia de que J es una función semicontinua inferiormente. Por lo tanto $J(x) \leq I(x)$ para toda $x \in X$, y por la simetría de la prueba, se cumple la unicidad de la función tasa.

□

El siguiente teorema es importante en la teoría de grandes desviaciones ya que es una de las posibles generalizaciones del teorema de Cramér.

Teorema 1 (Lema de Varadhan)**.** *Sea* $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ *una sucesión de probabilidades que satisface un (PGD) en* X *con orden* n *y función tasa* I*. Sea* $F : X \to \mathbb{R}$ *una función continua acotada superiormente. Entonces*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{n\,F(x)} \mathbb{P}_n(dx) = \sup_{x\in X}[F(x) - I(x)].$$

Demostración:

Para cada $S \subseteq X$ boreliano definimos

$$J_n(S) = \int_S e^{n\,F(x)}\,\mathbb{P}_n(dx), \qquad n \geq 1$$

y ponemos $a = \sup_{x\in X} F(x)$, $b = \sup_{x\in X}[F(x) - I(x)]$. Notamos que $-\infty < b \leq a < \infty$ debido a que $I \geq 0$, I no es idénticamente ∞ y F es acotada superiormente. La demostración se hace por medio de estimaciones de los límites superior e inferior de la sucesión $\{\frac{1}{n} \log J_n(X)\}_{n\in\mathbb{N}}$.

Estimación para el límite superior de la sucesión $\{\frac{1}{n} \log J_n(X)\}_{n\in\mathbb{N}}$.

Dividimos el espacio de acuerdo a los valores de F. Definimos $C = F^{-1}[b, a]$ y para cada N natural a los conjuntos

$$C_j^N = F^{-1}[c_{j-1}^N, c_j^N], \qquad j = 1, ..., N$$

donde $c_j^N = b + \frac{j}{N}(a - b)$. Claramente $\{C_j^N\}_{j=1}^N$ constituye una partición de C. Como F es continua cada C_j^N es cerrado. Entonces, debido al (PGD)

$$\lim_{n\to\infty} \sup \frac{1}{n} \log \mathbb{P}_n(C_j^N) \leq -\underline{I}(C_j^N) \quad \forall\, j = 1, \cdots n \qquad (\star).$$

Por otro lado,

$$\lim_{n\to\infty} \sup \frac{1}{n} \log J_n(C) = \lim_{n\to\infty} \sup \frac{1}{n} \log \sum_{j=1}^{N} \int_{C_j^N} e^{n\,F(x)}\,\mathbb{P}_n(dx).$$

Dado que $F \leq c_j^N$ en C_j^N tenemos

$$\lim_{n\to\infty} \sup \frac{1}{n} \log J_n(C) \;\; \leq \;\; \lim_{n\to\infty} \sup \frac{1}{n} \log \sum_{j=1}^{N} e^{n\,c_j^N}\,\mathbb{P}_n(C_j^N).$$

Debido a la Observación (2), la desviación crece conforme el mayor exponente, es decir

$$\begin{aligned}
\lim_{n\to\infty}\sup \frac{1}{n}\log\, J_n(C) &\leq \lim_{n\to\infty}\sup \frac{1}{n}\log\, \max_{1\leq j\leq N}[e^{n\, c_j^N}\, \mathbb{P}_n(C_j^N)] \\
&= \lim_{n\to\infty}\sup\, \max_{1\leq j\leq N}[c_j^N + \frac{1}{n}\log \mathbb{P}_n(C_j^N)] \\
&\leq \max_{1\leq j\leq N}[c_j^N - \underline{I}(C_j^N)]
\end{aligned}$$

donde la última desigualdad se debe a $(\star)$. Por construcción, $c_j^N \leq \inf_{x\in C_j^N} F(x) + \frac{1}{N}(a-b)$, así que sustituyendo en la desigualdad anterior nos da

$$\begin{aligned}
\lim_{n\to\infty}\sup \frac{1}{n}\log\, J_n(C) &\leq \max_{1\leq j\leq N}\left\{\inf_{x\in C_j^N} F(x) - \inf_{x\in C_j^N} I(x)\right\} + \frac{1}{N}(a-b) \\
&\leq \max_{1\leq j\leq N}\sup_{x\in C_j^N}[F(x)-I(x)] + \frac{1}{N}(a-b) \\
&= \sup_{x\in C}[F(x)-I(x)] + \frac{1}{N}(a-b) \\
&\leq b + \frac{1}{N}(a-b).
\end{aligned}$$

Haciendo tender N a infinito, se tiene la estimación

$$\lim_{n\to\infty}\sup \frac{1}{n}\log\, J_n(C) \leq b.$$

Ahora, como $F(x)\leq b$ para cualquier $x\in X\setminus C$, tenemos la estimación trivial

$$J_n(X\setminus C) = \int_{X\setminus C} e^{n\,F(x)}\mathbb{P}_n(dx) \leq \int_{X\setminus C} e^{nb}\,\mathbb{P}_n(dx) \leq e^{nb}.$$

Usando de nuevo la Observación (2),

$$\begin{aligned}
\lim_{n\to\infty}\sup \frac{1}{n}\log\,[J_n(X)] &= \lim_{n\to\infty}\sup \frac{1}{n}\log\,[J_n(C) + J_n(X\setminus C)] \\
&= \lim_{n\to\infty}\sup \frac{1}{n}\log\, \max\{J_n(C), J_n(X\setminus C)\} \\
&\leq \lim_{n\to\infty}\sup \max\left\{\frac{1}{n}\log\, J_n(C),\, b\right\} \\
&\leq b
\end{aligned}$$

obtenemos la cota superior.

Estimación para el límite inferior de la sucesión $\{\frac{1}{n}\log\, J_n(X)\}_{n\in\mathbb{N}}$

Tomemos $x\in X$ y $\varepsilon>0$ arbitrarios. Definimos al conjunto

$$O_{x,\varepsilon} = \{y\in X : F(y) > F(x)-\varepsilon\},$$

que resulta ser una vecindad abierta debido a la continuidad de F. Debido al (PGD), tenemos que

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(O_{x,\varepsilon}) \geq -\underline{I}(O_{x,\varepsilon}).$$

Como $\underline{I}(O_{x,\varepsilon}) \leq I(x)$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log J_n(O_{x,\varepsilon}) &= \liminf_{n\to\infty} \frac{1}{n} \log \int_{O_{x,\varepsilon}} e^{n\,F(y)} \mathbb{P}_n(dy) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log \left[e^{n\,(F(x)-\varepsilon)} \mathbb{P}_n(O_{x,\varepsilon})\right] \\
&\geq \liminf_{n\to\infty} F(x) - \varepsilon + \frac{1}{n} \log \mathbb{P}_n(O_{x,\varepsilon}) \\
&\geq F(x) - \varepsilon - \underline{I}(O_{x,\varepsilon}) \\
&\geq F(x) - \varepsilon - I(x).
\end{aligned}$$

Ahora usamos el hecho de que $J_n(X) \geq J_n(O_{x,\varepsilon})$, entonces

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(X) \geq F(x) - \varepsilon - I(x).$$

Haciendo ε tender a 0 y tomando después el supremo sobre todo $x \in X$ tenemos

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(X) \geq b.$$

□

Debido a la estructura exponencial subyacente a un (PGD), el Lema de Varadhan es particularmente útil para resolver problemas relacionados con integrales de funcionales exponenciales. Hay diversas aplicaciones en conducción de calor, mecánica estadística, optimización de pruebas de hipótesis en estadística, dinámica hamiltoniana, procesos de difusión, campos aleatorios entre muchas áreas. Dos libros con un buen panorama de aplicaciones son [den], [Ell]. Un compendio de conferencias dadas por Varadhan [Var2] tiene un enriquecedor panorama de la teoría general. Cabe mencionar que la teoría de grandes desviaciones se puede desarrollar también en espacios topológicos arbitrarios [Dem], aunque el contexto de espacios polacos parece ser una buena elección entre generalidad y simplicidad de las pruebas.

1. La entropía como medida de incertidumbre y complejidad. Marco conceptual de Boltzmann

Supongamos que tenemos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ en donde ocurre un evento E. Supongamos que queremos cuantificar por medio de una función $S : \mathcal{F} \to \mathbb{R}$ cuánta incertidumbre está asociada a ese evento. Es razonable pensar que la incertidumbre no depende del evento en particular, sino simplemente de su probabilidad de ocurrencia; es decir, que la función S tan sólo dependa de $\mathbb{P}(E)$ (y por lo tanto basta definir $S : [0,1] \to \mathbb{R}$).

A esta función también le pedimos las siguientes características:

(a). $S(1) = 0$, es decir, cualquier evento seguro carece de incertidumbre.

(b). $S(p)$ es una función decreciente respecto a la probabilidad del evento; cuánto más probable es un evento menos incertidumbre se le asigna.

(c). $S(p)$ es una función continua. Si hay cambios pequeños en probabilidad, esperamos cambios pequeños en la incertidumbre asociada.

(d). Supongamos que tenemos dos eventos independientes E y F. La cantidad $S(E \cap F) - S(E)$ representa la incertidumbre de saber que F ocurre sabiendo que E ha ocurrido. Debido a la independencia, esta cantidad debe ser simplemente $S(F)$, y por lo tanto

$$S(E \cap F) = S(E) + S(F).$$

Debido a la misma independencia, si $\mathbb{P}(E) = p$, $\mathbb{P}(F) = q$, entonces $\mathbb{P}(E \cap F) = pq$. Esta intuición es una motivación para la propiedad

$$S(pq) = S(p) + S(q) \qquad \forall\, p, q \in (0,1].$$

Afirmación 1. *Si S satisface las propiedades (a)-(d) entonces existe $k > 0$ tal que $S(p) = k \log(p)$.*

Demostración:

Usando (d) es fácil ver que $S(p^m) = m\,S(p)$, $m \geq 1$, $p \in (0,1]$. Ahora, para cualquier natural n tenemos

$$S(p) = S((p^{\frac{1}{n}})^n) = n\,S(p^{\frac{1}{n}})$$

por lo que $S(p^{\frac{1}{n}}) = \frac{1}{n}S(p)$. Combinando las dos propiedades anteriores tenemos

$$S(p^{\frac{m}{n}}) = m\,S(p^{\frac{1}{n}}) = \frac{m}{n}\,S(p), \qquad \forall\, m, n \geq 1$$

por lo que para cualquier racional positivo ocurre $S(p^x) = x\,S(p)$. Por la continuidad de S y la densidad de los números racionales se sigue que $S(p^x) = x\,S(p)$ para todo $p \in (0,1]$, $x > 0$.

Poniendo $y = p^x$ tenemos que

$$\frac{S(y)}{\log\, y} = \frac{S(p)}{\log p} = k \qquad \forall p, y \in (0,1].$$

Por lo tanto $S(p) = k\, \log p$, donde $k = -S(e^{-1}) > -S(1) = 0$ por las propiedades (a) y (b).

$\square$

Lo anterior nos da la interpretación probabilista de la famosa fórmula de Boltzmann enunciada en el marco de la entropía termodinámica

$$S = k\, \log \mathbb{P}(W).$$

Es decir, la entropía de un estado macroscópico W es el logaritmo de la probabilidad de su ocurrencia multiplicada por un factor k (la constante de Boltzmann). A continuación desarrollamos este marco conceptual.

Consideremos un conjunto de n partículas con niveles de energía representados por un espacio finito E. Pensemos que una configuración de las partículas y sus niveles de energía se puede representar como un elemento ω del espacio E^n. A esta configuración le llamamos microestado.

Supondremos que los microestados no son observables, pero que podemos observar como se distribuyen los niveles de energía, es decir, la información accesible son elementos μ en el espacio

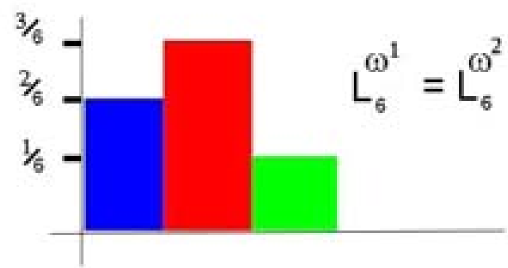
$$\mathcal{M}(E) = \Big\{ \mu : \mu(x) = \frac{m_x}{n},\, m_x \in \mathbb{N},\, \sum_{x \in E} \mu(x) = 1, \mu(x) \geq 0 \Big\}.$$

En particular, $\mathcal{M}(E)$ es un subespacio de las medidas sobre E, y cada μ puede ser pensada de dos maneras: como un histograma para n realizaciones, o como un macroestado de un sistema de n partículas. Al conjunto $\mathcal{M}(E)$ se le llama el espacio de histogramas de longitud n de E.

Definimos $N_n(\mu)$ como el número de microestados ω que llevan al macroestado μ. Para un microestado $\omega \in E^n$ definimos la medida aleatoria

$$L_n^\omega = \frac{1}{n} \sum_{i=1}^{n} \delta_{\omega_i}.$$

L_n^ω describe como se distribuyen las partículas sobre los niveles de energía y se le conoce como distribución empírica o histograma asociado al microestado ω.



Entonces podemos expresar a $N_n(\mu)$ en términos de L_n^ω:

$$N_n(\mu) \stackrel{\circ}{=} \#\{\omega \in E^n : L_n^\omega = \mu\}$$

que además, contando, resulta ser

$$N_n(\mu) = \frac{n!}{\prod_{x\in E} m_x!} = \frac{n!}{\prod_{x\in E}(n\,\mu(x))!}.$$

La idea de Boltzmann era que la entropía (o cantidad de incertidumbre) del estado ω cuando sólo conocemos μ se puede cuantificar por el logaritmo de microestados ω que llevan a μ. Entonces la entropía de un histograma de n partículas μ resulta ser

$$H(\mu) = \log\, N_n(\mu).$$

Ahora, si en lugar de querer calcular la entropía para una medida de probabilidad en el espacio $\mathcal{M}(E)$, lo queremos hacer para una medida de probabilidad μ_0 arbitraria sobre E, podemos aproximar por una sucesión $\{\mu_n\}_{n\in\mathbb{N}}$ de macroestados con n partículas tales que $\mu_n \to \mu_0$ (donde usamos la distancia de variación total, $d(\mu,\nu) = \frac{1}{2}\sum_{e\in E}|\mu\{e\} - \nu\{e\}|$) y definir la incertidumbre de μ_0 como el límite de la incertidumbre promedio de μ_n por partícula, es decir

$$H(\mu_0) = \lim_{n\to\infty}\frac{1}{n}H(\mu_n) = \lim_{n\to\infty}\frac{1}{n}\log\, N_n(\mu_n).$$

Verifiquemos la existencia de este límite y calculémoslo.

Teorema 1. *Sea $\{\mu_n\}_{n\in\mathbb{N}}$ una sucesión de medidas de probabilidad sobre un espacio finito E tales que $n\,\mu_n(x) \in \mathbb{N}$ para todo $x \in E$, $n \in \mathbb{N}$ y $\mu_n \to \mu_0$. Entonces el límite*

$$\lim_{n\to\infty}\frac{1}{n}\log\, N_n(\mu_n)$$

existe y es igual a

$$H(\mu) = -\sum_{x\in E} \mu_0(x)\ \log\ \mu_0(x).$$

Demostración:

La demostración es directa utilizando la aproximación para $n!$ dada por la fórmula de Stirling:

$$n! \sim \sqrt{2\pi\, n}\, n^n\, e^{-n}.$$

Entonces

$$\begin{aligned} N_n(\mu_n) &= \frac{n!}{\prod_{x\in E}(n\,\mu_n(x))!} \\ &\sim \frac{\sqrt{2\pi\, n}\, n^n\, e^{-n}}{\prod_{x\in E}\sqrt{2\pi\, n\mu_n(x)}\,(n\,\mu_n(x))^{n\,\mu_n(x)} e^{-n\,\mu_n(x)}}. \end{aligned}$$

Y podemos calcular

$$\begin{aligned} \frac{1}{n}\log\, N_n(\mu_n) &= \frac{1}{n}\Big[\log\sqrt{2\pi n} + n\,\log\left(\frac{n}{e}\right) - \sum_{x\in E}\log\,\left(\sqrt{2\pi n\mu_n(x)}\left(\frac{n\mu_n(x)}{e}\right)^{n\,\mu_n(x)}\right] \\ &= \frac{1}{n}\Big[\log\sqrt{2\pi n} + n\,\log\left(\frac{n}{e}\right) - \sum_{x\in E}\log\,\sqrt{2\pi n\mu_n(x)} \\ &- \log\,\left(\frac{n}{e}\right)\sum_{x\in E} n\mu_n(x) - n\sum_{x\in E}\mu_n(x)\log\mu_n(x)\Big] \\ &= \frac{1}{n}\Big[\log\sqrt{2\pi n} - \sum_{x\in E}\log\,\sqrt{2\pi n\mu_n(x)} - n\sum_{x\in E}\mu_n(x)\log\mu_n(x)\Big] \end{aligned}$$

Debido a la continuidad de la función $\log(x)$ y a que $\mu_n(x) \to \mu_0(x)$, tomando el límite en la expresión anterior concluimos que

$$\lim_{n\to\infty}\frac{1}{n}\log\, N_n(\mu_n) = -\sum_{x\in E}\mu_0(x)\log\mu_0(x)$$

□

1. Entropía de una partición

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $\xi = \{A_1, \cdots, A_k\} \subseteq \mathcal{F}$ una partición finita de Ω. Consideremos que la partición ξ representa las salidas de un experimento realizado en el espacio muestral Ω. Definamos a la entropía de ξ, $H(\xi)$, como la cantidad de incertidumbre eliminada al realizar el experimento con salidas $\{A_1, \cdots, A_k\}$. Notemos que debido a la naturaleza aleatoria del experimento, la cantidad de incertidumbre depende tan sólo de las probabilidades $\{\mathbb{P}(A_1), \cdots, \mathbb{P}(A_k)\}$. Extendiendo nuestro resultado de la sección anterior, podemos definirla como

$$H(\xi) = H(\mathbb{P}(A_1), \cdots, \mathbb{P}(A_k)) = -\sum_{i=1}^{k} \mathbb{P}(A_i) \log \mathbb{P}(A_i).$$

Otra manera de obtener esta fórmula es pensar que la entropía de la partición es el valor esperado de la entropía de cada elemento de la partición (calculado como el logaritmo de la probabilidad, al inicio de la sección anterior).

Ahora, supongamos que tenemos dos particiones $\xi = \{A_1, \cdots, A_k\}$ y $\nu = \{B_1, \cdots, B_l\}$ asociadas a los resultados de dos experimentos en el mismo espacio de estados $(\Omega, \mathcal{F}, \mathbb{P})$. Buscamos una manera de medir la incertidumbre del resultado de ξ dado el resultado de ν. Sabemos que la probabilidad de que ocurra A_i dado que ocurrió B_j está dada por $\mathbb{P}(A_i|B_j)$ y entonces, la incertidumbre del experimento ξ dado que ocurrió B_j es

$$H(\mathbb{P}(A_1|B_j), \cdots, \mathbb{P}(A_k|B_j)) = -\sum_{i=1}^{k} \mathbb{P}(A_i|B_j) \log \mathbb{P}(A_i|B_j). \qquad (\star)$$

Teniendo definida la entropía de ξ para cada conjunto $B_j \in \nu$, definimos a la entropía de ξ dada ν como:

$$H(\xi|\nu) = -\sum_{j=1}^{l} \Big[\sum_{i=1}^{k} \mathbb{P}(A_i|B_j) \log \mathbb{P}(A_i|B_j) \Big] \mathbb{P}(B_j).$$

Notemos que esta definición es precisamente la esperanza de la función $(\star)$.

Tenemos distintas expresiones para la entropía condicional:

$$\begin{aligned}
\mathbb{E}(H(\mathbb{P}(A_1|\cdot), \cdots, \mathbb{P}(A_k|\cdot)) &= \sum_{j=1}^{l} \mathbb{P}(B_j) H(\mathbb{P}(A_1|B_j), \cdots, \mathbb{P}(A_k|B_j)) \\
&= \sum_{j=1}^{l} \mathbb{P}(B_j) \Big(-\sum_{i=1}^{k} \mathbb{P}(A_i|B_j) \log \mathbb{P}(A_i|B_j) \Big) \\
&= -\sum_{i,j} \mathbb{P}(B_j) \mathbb{P}(A_i|B_j) \log \mathbb{P}(A_i|B_j) \\
&= -\sum_{i,j} \mathbb{P}(A_i \cap B_j) \log \frac{\mathbb{P}(A_i \cap B_j)}{\mathbb{P}(B_j)}.
\end{aligned}$$

Ahora podemos ver que nuestra elección para la fórmula de la entropía no es arbitraria. Definamos al conjunto

$$\triangle_k = \Big\{(p_1, \cdots, p_k) \in \mathbb{R}^k : p_i \geq 0, \sum_{i=1}^{k} p_i = 1\Big\},$$

y la función $H : \bigcup_{k=1}^{\infty} \triangle_k \to \mathbb{R}$ dada por

$$H(p_1, \cdots, p_k) = -\sum_{i=1}^{k} p_i \log p_i,$$

con la convención $p_i \log p - i = 0$ si $p_i = 0$. Khinchine [Khi] probó que cualquier función $F : \bigcup_{k=1}^{\infty} \triangle_k \to \mathbb{R}$ que cumple las propiedades siguientes es de la forma $F = \lambda H$ con $\lambda > 0$.

a. $F(p_1, \cdots, p_k) \geq 0$ con igualdad sólo cuando $p_i = 1$ para algún i.

b. F restringida a $\triangle_k$ es continua.

c. F restringida a $\triangle_k$ es simétrica (sobre los argumentos $p_1, \cdots, p_k$).

d. F restringida a $\triangle_k$ alcanza su valor máximo en $(\frac{1}{k}, \cdots, \frac{1}{k})$.

e. F cumple que para cualquier par $(p_1, \cdots, p_k) \in \triangle_k$, $(q_1, \cdots, q_l) \in \triangle_l$,

$$F(p_1\, q_1, \cdots, p_1\, q_l, \cdots, p_k\, q_1, \cdots, p_k\, q_l) = F(p_1, \cdots, p_k) + F(q_1, \cdots, q_l)$$

f. $F(p_1, \cdots, p_k, 0) = F(p_1, \cdots, p_k)$.

Estas propiedades son deseables para nuestra función entropía. (a) dice que el único experimento que no proporciona ninguna información es el experimento con un resultado certero. (d) nos dice que, entre los experimentos con k resultados, el que tiene más incertidumbre es aquél con salidas equiprobables.

Sean dos particiones $\xi = \{A_1, \cdots, A_k\}$ y $\nu = \{B_1, \cdots, B_l\}$ asociadas a dos experimentos. Definimos a la partición junta de las particiones ξ y ν como sigue

$$\xi \vee \nu = \{A_i \cap B_j : 1 \leq i \leq k, 1 \leq j \leq m\},$$

que representa los resultados combinados de ambos experimentos. Llamemos $p_i = \mathbb{P}(A_i)$ y $q_j = \mathbb{P}(B_j)$. Entonces (e) nos dice que la información ganada por realizar dos experimentos ξ y ν es la información obtenida por realizar ξ más la información obtenida por realizar ν dado ξ

$$F(\xi \vee \nu) = F(\xi) + F(\nu|\xi),$$

en el caso particular en que $\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\, \mathbb{P}(B_j)$, para toda i, j; es decir, cuando los experimentos son independientes.

Enunciamos algunas propiedades de la entropía y de la entropía condicional. La demostración del siguiente lema es elemental y se omite; la prueba se puede consultar en [Wal].

Lema 1. *Sean ξ,ν y η particiones finitas de Ω. Entonces ocurren las siguientes propiedades:*

1. $H(\nu) \geq 0$.
2. $H(\nu|\xi) \geq 0$.
3. $H(\nu \vee \xi|\eta) = H(\nu|\eta) + H(\xi|\nu \vee \eta)$.
4. $H(\nu) \geq H(\nu|\xi)$.
5. $H(\nu \vee \xi) \leq H(\nu) + H(\xi)$.

Nuestra definición de entropía es lo suficientemente general para abordar distintos problemas, pero nos enfocaremos en algunos aspectos relacionados a probabilidad. A continuación definimos la entropía para variables aleatorias discretas.

Definición 1. *Sea X una v.a. discreta que toma valores en $T = \{t_1, t_2, \cdots\}$. Definimos la entropía de X, $H(X)$ como $H(\xi)$ donde ξ es la partición dada por $\xi = \{X^{-1}(t_i) : t_i \in T\}$. Entonces*

$$H(X) = H(\xi) = -\sum_{t_i \in T} \mathbb{P}(X = t_i) \log \mathbb{P}(X = t_i).$$

Tenemos la extensión para variables aleatorias con densidad, conocida en la literatura como entropía diferencial [Geo].

Definición 2. *Sea X una v.a. con densidad f cuyo soporte es T. Definimos la entropía de X como*

$$H(X) = -\int_T f(x) \log f(x) dx$$

cuando la integral existe, y $H(X) = \infty$ en caso contrario.

Ahora probamos un pequeño resultado con un doble propósito: hacer cálculos de entropías específicas y enunciar una definición equivalente de la entropía condicional en un caso particular.

Lema 2. *Sean $(p_1, \cdots, p_k), (q_1, \cdots, q_k) \in \triangle_k$. Entonces*

$$-\sum_{i=1}^{k} p_i \log p_i \leq -\sum_{i=1}^{k} p_i \log q_i$$

con igualdad sólo cuando $p_i = q_i$ para toda i.

Demostración:

Notemos que $log\, x \leq x - 1$ con igualdad sólo cuando $x = 1$. Entonces

$$\sum_{i=1}^{k} p_i \log \left(\frac{q_i}{p_i}\right) \leq \sum_{i=1}^{k} p_i \left(\frac{q_i}{p_i} - 1\right) = \sum_{i=1}^{k} (q_i - p_i) = 0$$

$$\Leftrightarrow -\sum_{i=1}^{k} p_i \log(p_i) \leq -\sum_{i=1}^{k} q_i \log(q_i)$$

y la igualdad se da sólo cuando $\frac{q_i}{p_i} = 1$ para cada $i = 1, \cdots, k$.

$\square$

Notamos que el lema anterior se puede extender cuando el conjunto de valores $\{p_i\}$ es numerable, o a su forma continua

$$-\int_T f(x)\,\log\, f(x)\,dx \leq -\int_T f(x)\,\log\, g(x)\,dx$$

haciendo exactamente lo mismo.

Del lema anterior se sigue que podemos definir, equivalentemente, la entropía condicional para $(p_1, \cdots, p_k), (q_1, \cdots, q_k) \in \triangle_k$ como sigue

$$H(p_1, \cdots, p_k | q_1, \cdots, q_k) = -\sum_{i=1}^{k} p_i\,\log\,\frac{p_i}{q_i}$$

con igualdad sólo cuando $(p_1, \cdots, p_k) = (q_1, \cdots, q_k)$.

Calculamos la entropía de v.a. que maximizan la entropía, dadas distintas condiciones.

Teorema 1. *Sean U, G, E, Z v.a. que distribuyen Uniforme(n), Geométrica$(\frac{1}{\lambda})$, Exp$(\frac{1}{\lambda})$, Normal$(0, 1)$ respectivamente. Entonces*

(a). Si X es una v.a. con soporte en $\{1, \cdots, n\}$ entonces $H(X) \leq H(U)$.

(b). Si X es una v.a. con soporte en $\{1, 2, \cdots\}$ con media $\lambda > 1$ entonces $H(X) \leq H(G)$.

(c). Si X es una v.a. con densidad en $[0, \infty)$ y media λ entonces $H(X) \leq H(E)$.

(d). Si X es una v.a. con densidad en $\mathbb{R}$ tal que $\mathbb{E}(X^2) = 1$ entonces $H(X) \leq H(Z)$.

Demostración:

Todos los puntos son análogos (una aplicación directa del lema anterior), pero los calculamos para tener la fórmula especifica de la entropía en cada caso.
(a). Sea X una v.a. con soporte en $\{1, \cdots, n\}$ y probabilidades $p_i = \mathbb{P}(X = i)$, $i = 1, \cdots, n$.

$$H(X) = -\sum_{i=1}^{n} p_i\,\log\,p_i \leq -\sum_{i=1}^{n} p_i\,\log\,\frac{1}{n} = \log\,n\; = -\sum_{i=1}^{n} \frac{1}{n}\,\log\,\frac{1}{n} = H(U).$$

(b). Sea X una v.a. con soporte en $\{1, 2, \cdots\}$ con media $\lambda > 1$ y probabilidades $p_i = \mathbb{P}(X = i)$, $i = 1, 2, \cdots$.

$$\begin{aligned}
H(X) = -\sum_{i\in\mathbb{N}} p_i\,\log\,p_i &\leq -\sum_{i\in\mathbb{N}} p_i\,\log\,\left[\Big(\frac{1}{\lambda}\Big)\Big(1 - \frac{1}{\lambda}\Big)^{i-1}\right] \\
&= -\sum_{i\in\mathbb{N}} p_i\Big(i(-\,\log\,(\lambda) + \log\,(\lambda - 1)) - \log(\lambda - 1)\Big) \\
&= -\Big[(-\,\log\,(\lambda)) + \log\,(\lambda - 1)\mathbb{E}(X) - \log(\lambda - 1)\Big] \\
&= \log\Big(\frac{\lambda^{\lambda}}{(\lambda - 1)^{\lambda - 1}}\Big).
\end{aligned}$$

Y por otro lado

$$\begin{aligned} H(G) &= -\sum_{i\in\mathbb{N}} \Big(\frac{1}{\lambda}\Big)\Big(1-\frac{1}{\lambda}\Big)^{i-1} \log\Big[\Big(\frac{1}{\lambda}\Big)\Big(1-\frac{1}{\lambda}\Big)^{i-1}\Big] \\ &= -\sum_{i\in\mathbb{N}} \Big(\frac{1}{\lambda}\Big)\Big(1-\frac{1}{\lambda}\Big)^{i-1} \Big[-\log\lambda + i\,\log\Big(\frac{\lambda-1}{\lambda}\Big) + \log\frac{\lambda}{\lambda-1}\Big] \\ &= -\mathbb{E}(G)\,\log\Big(\frac{\lambda-1}{\lambda}\Big) + \log(\lambda-1) \\ &= \log\Big(\frac{\lambda^\lambda}{(\lambda-1)^{\lambda-1}}\Big). \end{aligned}$$

(c). Sea X una v.a. con densidad $f(x)$ en $[0,\infty)$ y media λ.

$$\begin{aligned} H(X) = -\int_{[0,\infty)} f(x)\,\log f(x)\,dx &\leq -\int_{[0,\infty)} f(x)\,\log\Big(\frac{1}{\lambda}\,e^{-\frac{x}{\lambda}}\Big)\,dx \\ &= -\int_{[0,\infty)} f(x)\Big(\log\frac{1}{\lambda} - \frac{x}{\lambda}\Big)\,dx \\ &= -\log\frac{1}{\lambda} + \frac{\mathbb{E}(X)}{\lambda} \\ &= 1+\log\lambda \end{aligned}$$

donde

$$H(E) = -\int_{[0,\infty)} \frac{e^{\frac{-x}{\lambda}}}{\lambda}\log\frac{e^{\frac{-x}{\lambda}}}{\lambda}dx = -\int_{[0,\infty)} \frac{e^{\frac{-x}{\lambda}}}{\lambda}\Big(\log\Big(\frac{1}{\lambda}\Big)-\frac{x}{\lambda}\Big)dx = -\log\Big(\frac{1}{\lambda}\Big)+\frac{\mathbb{E}(X)}{\lambda} = 1+\log\lambda$$

(d). Sea X v.a. con densidad en $\mathbb{R}$ tal que $\mathbb{E}(X^2)=1$

$$\begin{aligned} H(X) = -\int_{\mathbb{R}} f(x)\,\log f(x)\,dx &\leq -\int_{\mathbb{R}} f(x)\,\log\Big(\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\Big)\,dx \\ &= -\int_{\mathbb{R}} f(x)\Big(\log\Big(\frac{1}{\sqrt{2\pi}}\Big) - \frac{x^2}{2}\Big) \\ &= -\log\Big(\frac{1}{\sqrt{2\pi}}\Big) + \frac{1}{2}\mathbb{E}(X^2) \\ &= \frac{1}{2} + \log(\sqrt{2\pi}). \end{aligned}$$

Y por otro lado

$$\begin{aligned} H(Z) &= -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\,e^{\frac{-x^2}{2}}\,\log\Big(\frac{1}{\sqrt{2\pi}}\,e^{\frac{-x^2}{2}}\Big)\,dx \\ &= -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\,e^{\frac{-x^2}{2}}\Big(-\log(\sqrt{2\pi}) - \frac{-x^2}{2}\Big)\,dx \\ &= \log(\sqrt{2\pi}) + \frac{\mathbb{E}(X^2)}{2} = \frac{1}{2} + \log(\sqrt{2\pi}). \end{aligned}$$

$\square$

Como hemos visto el concepto de entropía nos da una medición de qué tan aleatoria es una variable aleatoria X. A diferencia de la variancia que es una medida de dispersión de los valores que toma la variable, la entropía no depende de la magnitud de los valores que toma X, tan sólo de la probabilidad de que los tomen.

Para cerrar esta sección hacemos notar que a pesar de la dificultad para definir la entropía de una v.a. arbitraria (o de su ley asociada μ) siempre se puede definir la entropía condicional entre dos leyes de probabilidad.

Definición 3. *Sean μ y ν medidas de probabilidad sobre un espacio medible $(\Omega, \mathcal{F})$. Definimos a la entropía (relativa) de ν dado μ*

$$H(\nu|\mu) = \left\{ \begin{array}{cc} -\int h \log h \, d\mu & \textit{cuando } \nu << \mu, \ \textit{donde } h = \frac{d\nu}{d\mu} \\ \infty & \textit{en otro caso.} \end{array} \right\}$$

1. Relaciones entre grandes desviaciones y la entropía

Supongamos que tiramos una moneda justa (equiprobable) n veces y contamos el número de soles k. El número de maneras de obtener k soles está dado por $\binom{n}{k}$. Si usamos la aproximación de Stirling tenemos que

$$\begin{aligned}\binom{n}{k} &\approx \frac{\sqrt{2\,\pi n}e^{-n}n^n}{\sqrt{2\,\pi k}e^{-k}k^k\sqrt{2\,\pi(n-k)}e^{-(n-k)}\,(n-k)^{n-k}} \\ &= \frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}}\exp\left[n\Big[-\frac{k}{n}\log\frac{k}{n}-\frac{n-k}{n}\log\frac{n-k}{n}\Big]\right]\end{aligned}$$

Si fijamos la proporción de soles $\frac{k}{n}$ y águilas $\frac{n-k}{n}$ como p y q respectivamente, tenemos que

$$\binom{n}{k} \approx \frac{1}{\sqrt{2\,\pi n\,pq}}\exp\{nH(p,q)\}.$$

En este caso la entropía es el factor del crecimiento exponencial del número de realizaciones posibles que corresponden a un número fijo k de soles. La probabilidad de obtener k soles al lanzar n volados es aproximada por

$$\mathbb{P}_n(k) \approx \frac{1}{\sqrt{2\,\pi n\,p\,q}}\exp\{n[H(p,q)-\log\,2]\}.$$

Ahora supongamos que tenemos una moneda arbitraria, con probabilidad α de sacar un sol y $1-\alpha$ de sacar un águila. En este caso nuestra aproximación es

$$\mathbb{P}_n(k) \approx \frac{1}{\sqrt{2\,\pi n\,p\,q}}\exp\{n[H(p,q)+p\,\log\,\alpha+(1-p)\,\log\,(1-\alpha)]\}$$

es decir, la entropía $H(p,q)$ como factor del volúmen, y un término extra, dependiente de p y α, al que se conoce como el *término de energía*. Podemos reordenar el exponente para expresar la aproximación de manera conceptualmente diferente

$$\mathbb{P}_n(k) \approx \frac{1}{\sqrt{2\,\pi n\,p\,q}}\exp\{n\,H(p,q\,|\,\alpha,1-\alpha)\},$$

es decir, en este caso nuestra función tasa de decaimiento de grandes desviaciones, para α fija, está dada por

$$I_\alpha(p) = -H(p,1-p\,|\,\alpha,1-\alpha).$$

Concluimos que, en el caso de v.a. independientes con distribución Bernoulli(α), nuestra función de decaimiento para el evento $\{S_n = k\}$, está dada por $-H(\frac{k}{n},1-\frac{k}{n}\,|\,\alpha,1-\alpha)$. La función de entropía condicional, definida puntualmente, coincide con nuestra función tasa.

A continuación un teorema en donde sucede lo mismo para otro ejemplo de grandes desviaciones.

Teorema 1 (Sanov)**.** *Sea $S = \{1, \cdots, r\}$ un espacio métrico finito y ρ una medida de probabilidad sobre $(S, \mathcal{B}(S))$. Supongamos que tenemos una sucesión $\{X_i\}_{i\in\mathbb{N}}$ de variables aleatorias independientes con ley ρ donde $\rho_i \doteq \rho\{i\} > 0$ para todo $i \in S$. Definamos a la medida empírica*

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

y la distancia entre medidas sobre S como la distancia de variación total, es decir

$$d(\nu, \mu) = \frac{1}{2}\sum_{s\in S} |\mu_s - \nu_s|.$$

Entonces, para todo $a > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) = -\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu),$$

donde $B_a(\rho)$ es la bola de radio a, centrada en ρ, en el espacio de medidas definidas en S, $\mathcal{M}(S)$ y $I_\mu(\nu) = -H(\nu|\mu)$ representa la función tasa de decaimiento exponencial.

Notemos que el enunciado anterior es prácticamente afirmar que las medidas $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ asociadas a las medidas empíricas $\{L_n\}_{n\in\mathbb{N}}$ satisfacen un (PGD) con función tasa I_ρ; la diferencia radica en que en un (PGD) debemos probar las desigualdades requeridas para conjuntos abiertos y cerrados arbitrarios, en lugar de bolas abiertas, como en esta versión. La demostración es una prueba similar a la del ejemplo anterior; se utiliza la aproximación de Stirling para estimar las probabilidades, la prueba puede consultarse en [den]. La generalización de este teorema a espacios arbitrarios se encuentra en [Oli] o en [Dem]; la prueba requiere varios resultados técnicos.

Lo que el teorema de Sanov expresa, es que entre todas las realizaciones posibles de los histogramas o leyes empíricas, las más probables son aquellas que tienen un histograma más cercano a μ, donde la cercanía está dada por la entropía relativa.

Finalmente, presentamos una observación de Varadhan [Var3], en la que se establece una relación existente entre cualquier colección de medidas que satisfagan un (PGD) en un espacio polaco y una entropía relativa.

Teorema 2. *Sea $\mathbb{P}$ una medida de probabilidad sobre S un espacio polaco. Definamos*

$$l = \acute{\imath}nf\{-H(\mathbb{P}|\mathbb{Q}) : \mathbb{Q}(A) = 1\}$$

Entonces

a. $\mathbb{P}(A) \leq e^{-l}$.

b. Supongamos que $\mathbb{Q} << \mathbb{P}$. *Entonces*

$$\mathbb{P}(A) \geq \mathbb{Q}(A)\, \exp\Big\{ -l - \frac{1}{\mathbb{Q}(A)} \mathbb{E}_{\mathbb{Q}} \Big| \log \frac{d\mathbb{Q}}{d\mathbb{P}} - l \Big| \Big\}$$

Demostración:
Si $\mathbb{P}$ es una medida de probabilidad y queremos estimar $\mathbb{P}(A)$, entonces lo podemos hacer si estimamos $l \doteq \text{ínf}\,\{-H(\mathbb{P}|\mathbb{Q}) : \mathbb{Q}(A) = 1\}$. Lo mostramos a continuación.

Sea g una función medible cualquiera. Apliquemos la desigualdad de Jensen a la función $g - \log \frac{d\mathbb{Q}}{d\mathbb{P}}$, bajo la función exponencial y la medida $\mathbb{Q}$:

$$\exp\Big\{ \mathbb{E}_{\mathbb{Q}} \Big(g - \log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}} \Big) \Big\} \leq \mathbb{E}_{\mathbb{Q}} \Big(\exp\Big\{ g - \log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}} \Big\} \Big).$$

Entonces

$$\begin{aligned}
\exp\Big\{ \int g - \log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}}\, d\,\mathbb{Q} \Big\} &\leq \int \exp\Big\{ g - \log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}} \Big\}\, d\,\mathbb{Q} \\
\Rightarrow \quad \exp\Big\{ \int g\, d\,\mathbb{Q} - \int \frac{d\,\mathbb{Q}}{d\,\mathbb{P}} \log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}}\, d\,\mathbb{P} \Big\} &\leq \int \frac{\exp(g)}{\exp\{\log \frac{d\,\mathbb{Q}}{d\,\mathbb{P}}\}}\, d\,\mathbb{Q} \\
\Rightarrow \quad \exp\{\, \mathbb{E}_{\mathbb{Q}}(g) + H(\mathbb{P}|\mathbb{Q})\,\} &\leq \mathbb{E}_{\mathbb{P}}(\exp(g)).
\end{aligned}$$

Por lo tanto

$$\mathbb{E}_{\mathbb{Q}}(g) \leq -H(\mathbb{P}|\mathbb{Q}) + \log\, \mathbb{E}_{\mathbb{P}}(\exp(g)).$$

Sea $c > 0$. Tomando $g = c\,1_A$ en la desigualdad anterior tenemos que

$$\mathbb{Q}(A) \leq \frac{1}{c}[-H(\mathbb{P}|\mathbb{Q}) + \log\,[e^c\,\mathbb{P}(A) + (1 - \mathbb{P}(A))].$$

Tomando el ínfimo sobre las medidas $\mathbb{Q}$ con $\mathbb{Q}(A) = 1$, obtenemos:

$$\begin{aligned}
1 &\leq \frac{1}{c}[l + \log\,[e^c\,\mathbb{P}(A) + (1 - \mathbb{P}(A))] \\
\Leftrightarrow l &\geq c - \log\,[1 - \mathbb{P}(A) + e^c\,\mathbb{P}(A)]
\end{aligned}$$

Haciendo c tender a infinito en el lado derecho de la ecuación, tenemos:

$$\begin{aligned}
\lim_{c\to\infty} c - \log[1 - \mathbb{P}(A) + e^c\,\mathbb{P}(A)] &= \log\Big(\lim_{c\to\infty} \exp\{c - \log[1 - \mathbb{P}(A) + e^c\,\mathbb{P}(A)]\} \Big) \\
&= \log\Big(\lim_{c\to\infty} \frac{e^c}{1 - \mathbb{P}(A) + e^c\,\mathbb{P}(A)} \Big) \\
&= -\log \mathbb{P}(A).
\end{aligned}$$

Por lo que tenemos $l \geq -\log\,\mathbb{P}(A)$, es decir $\mathbb{P}(A) \leq e^{-l}$.

Ahora supongamos $\mathbb{Q} << \mathbb{P}$. Llamemos $\mathbb{Q}(A) = q$, entonces

$$\begin{aligned}
\mathbb{P}(A) = \int_A \frac{d\mathbb{P}}{d\mathbb{Q}}\, d\mathbb{Q} &= q\,\frac{1}{q}\int_A \exp\Big\{ -\log\frac{d\mathbb{Q}}{d\mathbb{P}}\Big\}\, d\mathbb{Q} \\
&\geq q\,\exp\Big\{\frac{1}{q}\int_A \Big(-\log\frac{d\mathbb{Q}}{d\mathbb{P}}\Big)\, d\mathbb{Q}\Big\} \\
&\geq q\,\exp\Big\{ -l - \frac{1}{q}\,\mathbb{E}_{\mathbb{Q}}\Big|\log\frac{d\mathbb{Q}}{d\mathbb{P}} - l\Big|\Big\}.
\end{aligned}$$

$\square$

La primera parte del teorema nos muestra cómo l proporciona la desigualdad precisa (a) para el cálculo de la probabilidad de A. La segunda parte es una estimación más débil dada por la desigualdad (b); necesitamos tener una cercanía "en entropía"(es decir que $\mathbb{E}_{\mathbb{Q}}\Big|\log\frac{d\mathbb{Q}}{d\mathbb{P}} - l\Big|$ pequeño) y $Q(A) = 1$, para que nuestra estimación $\mathbb{P}(A) \approx e^{-l}$ sea buena. Este teorema nos muestra que los cálculos en grandes desviaciones consisten en calcular entropías, donde la medida Q es desconocida y se trata de aproximar. Existen muchas elecciones para la medida Q; la elección óptima se convierte un problema de control óptimo donde la entropía es el funcional de costo [Var3].

1. Bibliografía

[den] Den Hollander F. Fields Institute Monographs, 14. American Mathematical Society, Providence, RI, 2000. x+143 pp. ISBN: 0-8218-1989-5.

[Dem] Dembo, Amir; Zeitouni, Ofer Large deviations techniques and applications. Second edition. Applications of Mathematics (New York), 38. Springer-Verlag, New York, 1998. xvi+396 pp.

[Dur] Durrett, Richard(1-CRNL) Probability: theory and examples. Second edition. Duxbury Press, Belmont, CA, 1996. xiii+503 pp.

[Ell] Ellis, Richard S. Entropy, large deviations, and statistical mechanics. Reprint of the 1985 original. Classics in Mathematics. Springer-Verlag, Berlin, 2006. xiv+364 pp.

[Khi] Khinchin, A. I. Mathematical foundations of information theory. Translated by R. A. Silverman and M. D. Friedman. Dover Publications, Inc., New York, N. Y., 1957. ii+120 pp.

[Geo] Georgii, Hans-Otto(D-MNCH-MI) Probabilistic aspects of entropy. (English summary) Entropy, 37–54, Princeton Ser. Appl. Math., Princeton Univ. Press, Princeton, NJ, 2003..

[Oli] Olivieri, Enzo(I-ROME2); Vares, Maria Eulália(BR-CBPF) Large deviations and metastability. Encyclopedia of Mathematics and its Applications, 100. Cambridge University Press, Cambridge, 2005. xvi+512 pp.

[Var1] Varadhan, S. R. S. Large deviations. Ann. Probab. 36 (2008), no. 2, 397–419.

[Var2] Varadhan, S. R. S. Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics, 46. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984. v+75 pp.

[Var3] Varadhan, S. R. S. Large deviations and entropy. Entropy, 199–214, Princeton Ser. Appl. Math., Princeton Univ. Press, Princeton, NJ, 2003.

[Wal] Walters, Peter. An introduction to ergodic theory. Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982. ix+250 pp.