43

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

INTRODUCCION A LA TEORIA DE GRANDES DESVIACIONES

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

ERWIN LUIS HERRERA DOMINGUEZ

DIRECTOR DE TESIS: DRA. MA. EMILIA CABALLERO ACOSTA

LTAD DE CIENCIAS
UNAM

TESIS CON FALLA DE ORIGEN

IDAD NACIONAL
TONOMA DE
MEXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis

INTRODUCCION A LA TEORIA DE GRANDES DESVIACIONES

realizado por **Erwin Luis Herrera Domínguez**

con número de cuenta 9455523-7 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dra. Ma. Emilia Caballero Acosta	
Propietario	Dra. Ana Meda Guardiola	
Propietario	DR. JOSE MARIA GONZALEZ BARRIOS	
Suplente	Dr. Fernando Avila Murillo	
Suplente	M. en C. ALBERTO MOLINA ESCOBAR	

Consejo Departamental de
Matematicas

M. en C. JOSE ANTONIO FLORES DIAZ

.gradecimientos.

·acias a mis padres, Margarita y Raúl por apoyarme siempre y por los princi-)s que me han inculcado.

A mis hermanos, idem, que además han sido mis mejores amigos. A mi orino Dante, le dedico este trabajo.

Gracias a Ma. Emilia Caballero por las observaciones, comentarios , correc-)nes y la paciencia que nos ha tenido para terminar este material.

A la Dra. Ana Meda, los Drs. José María Barrios, Fernando Avila y el M.C. berto Molina por las observaciones y correcciones que hicieron al trabajo.

Gracias al Instituto de Matemáticas por la beca de tésis y el apoyo otorgado.

A Ulises, y porque terminemos pronto lo que resta del tema.

A mis amigos, Chucho, Cruz, Ulises, Ricardo, Moises, Yudico, Cesar, An- es, Carlos, Emilio, Julio, Veronica, Octavio, Gerardo, Salomon, Dulce, Dasha, .mia, Tatiana , Mónica.

Gracias a todos aquellos que han contribuido a mi formación academica y mo persona.

Erwin.

ıdice General

ntroducción.

presente trabajo tiene como objetivo dar una introducción al tema de la teoría grandes desviaciones, el cual es de gran interés actualmente en el campo de teoría de la probabilidad. Mostraremos las ideas principales que motivan el tudio del tema, así como el desarrollo de los resultados teóricos básicos. En ta parte hemos hecho un esfuerzo especial en dar las demostraciones completas detalladas de todos estos resultados, buscando siempre dar la presentación mas cesible de ellos. En general los textos que tratan el tema no hacen esto ya ie comunmente solo dan la idea de la demostración sin hacer los detalles. Esto rmitirá que este trabajo pueda ser usado como introducción para estudiantes los últimos semestres de licenciatura con conocimientos básicos de teoría probabilidad. Además, en este trabajo nos interesa presentar los elementos óricos en los que se fundamenta la teoría y se desarrollan algunas aplicaciones.

El primer resultado en la teoría de *grandes desviaciones* se debe a Harald amér. Cramér era un matemático suizo que daba asesoría a una compañia seguros. Para compañías como estas es fundamental protegerse del riesgo de e ocurran sucesos inesperados o raros y en este contexto podemos hablar de *andes desviaciones*. Es decir, sabemos que el Teorema del Límite Central nos información acerca del comportamiento de alguna distribución de probabili- d alrededor de su media o valor esperado, mientras que la teoría del riesgo en guros se ocupa de eventos *raros*, es decir, que están en la "cola" de la distribu- n de probabilidad. Más adelante mostraremos el teorema debido a Cramér e es fundamental en este trabajo así como la aplicación mencionada en teoría l riesgo, la cual es muy sencilla y además ilustra muy bien el tema.

El desarrollo moderno de la teoría de *desviaciones grandes* se debe en gran dida al trabajo fundamental de S.R.S. Varadhan en el que muestra una clara concisa descripición de los principales resultados. En el libro de [Dem98]

A. Dembo y O. Zeitouni, se pueden encontrar mayores referencias históricas. partir del trabajo de Varadhan el estudio de grandes desviaciones recobra ir portancia y esto puede obedecer a dos razones:

1. Las estimaciones de grandes desviaciones son una herramienta crucial q se utilizan en diversos problemas en estadística, ingeniería, mecánica e tadística y probabilidad aplicada.

2. Las técnicas empleadas en grandes desviaciones nos proporciona métod que funcionan en otras ramas de la probabilidad y han ayudado a resolv problemas en dichas ramas.

Como deciamos anteriormente la teoría de grandes desviaciones es el estud de *eventos raros*, en el sentido de que son sucesos muy poco probables. Es deci es el cálculo asintótico de probabilidades tan pequeñas que se estudian en esca exponencial.

El término *grandes* en grandes desviaciones se debe a lo siguiente: El te rema del Límite Central (TLC) nos dice que : Si $X_1, ..., X_n$ es una sucesión variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) co media μ y varianza $\sigma^2 < \infty$, entonces la media muestral $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ aproximadamente normal con media μ y varianza σ^2/n, y la aproximación válida solamente en una vecindad de radio $\sigma/\sqrt{n}$ alrededor de μ. Esto es, TLC gobierna fluctuaciones aleatorias sólo cerca de la media, desviaciones d orden de $\sigma/\sqrt{n}$. Las fluctuaciones que son del orden de σ son, en relación con l fluctuaciones típicas, mucho más grandes: son grandes desviaciones de la med Ocurren sólo raramente, por lo que la teoría de las grandes desviaciones se d scribe frecuentemente como la *teoría de los eventos raros*, eventos que ocurr lejos de la media, en las colas de la distribución.

En el capítulo 1 estudiamos ciertas propiedades de las "colas" de algunas d tribuciones de probabilidad. En el primer caso, que corresponde a la distribuci Binomial, realizamos un sencillo experimento y nos preguntamos por la pr abilidad de que algún evento se encuentre en las colas de la distribución (*evento raro*). De una forma muy intuitiva y usando elementos gráficos poden encontrar estas probabilidades.

En el segundo caso, el de la distribución normal, hallamos la tasa de conver-
encia (a la que llamaremos función de tasa) de eventos raros, en este caso, de
ue la media muestral diste significativamente del valor esperado. Y en el último
aso, el de la distribución estable, se muestra que no siempre se puede hallar
na tasa de convergencia usando la técnica del caso anterior. Finalmente, enun-
iaremos el concepto de función de tasa y el Principio de Grandes Desviaciones
PGD) en su forma más general.

Para el capítulo 2, estudiaremos el PGD para espacios discretos de dimensión
nita. Utilizaremos métodos de combinatoria para obtener un PGD para la
istribución empírica de procesos que toman valores en un conjunto finito, en
ste caso obtendremos el importante Teorema de Sanov; además emplearemos el
nétodo para obtener grandes desviaciones en el tema de muestreo sin reemplazo,
n procedimiento que surge en muchos problemas estadísticos.

Empleando los métodos combinatorios, también tendremos un PGD para la
nedia empírica de procesos que toman valores en conjuntos finitos, en este caso
btendremos el Teorema de Cramér para conjuntos finitos en $\mathbb{R}$ el cual será una
onsecuencia casi inmediata del Teorema de Sanov.

Como consecuencia de los teoremas antes mencionados, mostramos como se
ueden aplicar estos resultados en el área de teoría del riesgo y teoría de la
nformación.

Es importante señalar que los resultados obtenidos en este capítulo, utilizando
l método de tipos para conjuntos finitos, nos darán una idea aproximada acerca
e los resultados que podríamos esperar para conjuntos más abstractos.

En el capítulo 3 daremos la demostración del Teorema de Cramér en el caso
eal, es decir, obtendremos un PGD para la media empírica de una sucesión de
a.i.i.d. con valores en $\mathbb{R}$ y función de tasa que en este caso será la transformada
e Cramér. Enunciaremos propiedades de la transformada de Cramér que nos
rán útiles en la demostración del teorema. Finalmente, haremos el cálculo
plícito de la función de tasa para algunas distribuciones de probabilidad.

Damos la definición de *tensión exponencial*, así como la noción de PGD débil
el capítulo 4. Obtendremos, como en el capítulo anterior, un PGD para la
edia empírica de una sucesión de v.a.i.i.d. con valores en $\mathbb{R}^d$. Veremos las
opiedades análogas de la transformada de Cramér en $\mathbb{R}^d$ al caso en $\mathbb{R}$.

Por último, en el capítulo 5, tratamos el caso de variables aleatorias que no son independientes. Situaciones de este tipo se encuentran en diversos problemas de probabilidad, por ejemplo, las cadenas de Markov. También obtendremos un PGD para este caso.

Estas son algunas situaciones en las cuales se pueden obtener PGD. Esto no es exhaustivo y desde luego hay muchos otros casos en donde también hay PGD, pero en éste trabajo hemos buscado exponer los casos más accesibles ya que la teoría de grandes desviaciones es una rama activa dentro de la probabilidad y frecuentemente se obtienen nuevos resultados en el área que además pueden llegar a ser sumamente complicados técnicamente y usar herramienta muy sofisticada, todo lo cual queda fuera del objetivo de ésta tesis.

Cabe mencionar que un trabajo posterior, se presentarán distintas aplicaciones de la teoría de Grandes Desviaciones a diversos problemas en estadística, teoría de la información, mecánica estadística y probabilidad aplicada. Este material y lo que se presente porteriormente es resultado del trabajo en equipo que he hecho con Ulises Salas y bajo la dirección de la Dra. Ma. Emilia Caballero y la Dra. Ana Meda.

`# Capítulo 1

El Principio de Grandes Desviaciones.

.1 Colas de la distribución binomial.

'omenzaremos la exposición del tema con un ejemplo que nos ayudará a en-
ender de manera sencilla e intuitiva el concepto subyacente en el principio de
randes desviaciones. Consideremos el experimento de lanzar una moneda n
eces y registramos los resultados. Supongamos que hay dos posibles resultados
ara cada lanzamiento (águila o sol), por lo que tenemos en total 2^n posibles
esultados. ¿Qué podemos decir del número total de águilas? Existen $n+1$
osibles valores, desde 0 hasta n; ahora bien, de los 2^n posibles resultados, C_r^n
epresenta las distintas formas de obtener r águilas en n lanzamientos (C_r^n son
s combinaciones de n tomadas en r). Si la moneda es justa, es decir, que
enemos la misma probabilidad de obtener águila o sol, entonces los resultados
n igualmente probables y por lo tanto la probabilidad de obtener r águilas en
lanzamientos es $\frac{C_r^n}{2^n}$. Por lo que el número promedio de águilas por lanzamien-
tiene $n+1$ posibles valores, $\{0, 1/n, 2/n, ..., 1\}$ y el valor r/n tiene peso $\frac{C_r^n}{2^n}$.
alculemos la probabilidad de que el número promedio de águilas se encuentre
un cierto rango. Sea M_n el número promedio de águilas en n lanzamientos,
ntonces para x y y reales en el intervalo $[0,1]$ tenemos que

$$\mathbb{P}(x < M_n < y) = \sum_{\{r: x < \frac{r}{n} < y\}} \frac{C_r^n}{2^n}.$$

tilizando la expresión anterior, obtenemos los histogramas de la distribución
probabilidad de M_n para distintos valores de n. En la figura 1 se presentan
s histogramas para $n = 16, n = 64$ y $n = 100$.

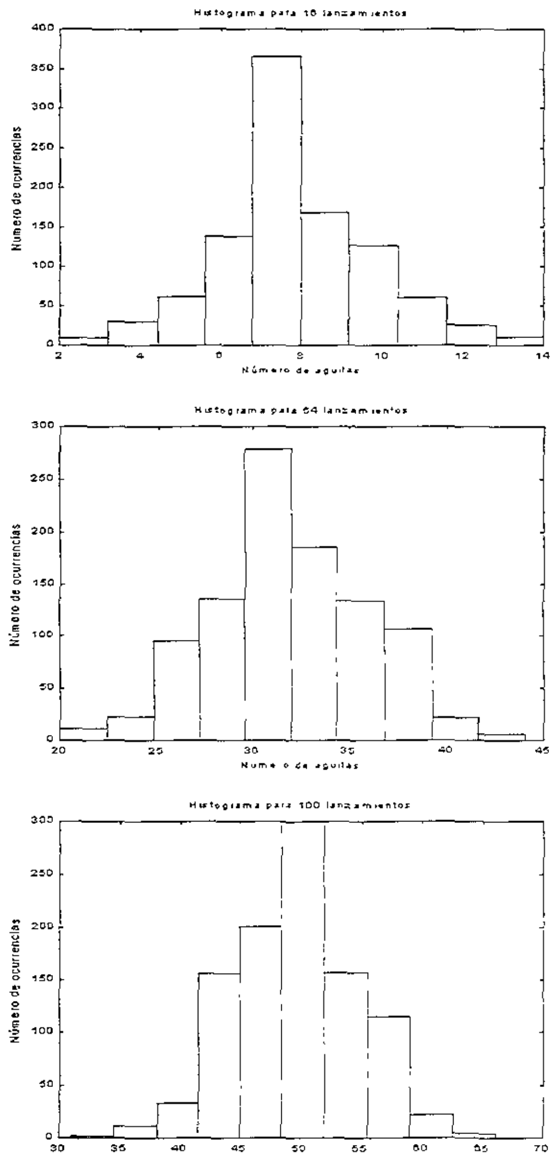

Figura 1. Distribución del número de águilas en n lanzamientos.

Como era de esperarse, observamos en las gráficas que conforme n es m
grande, la distribución se concentra alrededor de la media $1/2$, mientras que

as de la distribución se hacen cada vez más pequeñas. Es decir, los eventos que tan significativamente de la media tienen probabilidad de ocurrencia pequeña os interesará estudiar este fenómeno. Para esto consideremos dos casos:

1. Escogemos cualquier punto x mayor que la media de la distribución (1/2 en éste caso), es decir, el evento *raro* $\{M_n > x\}$, calculamos ahora para un rango de valores de n el logaritmo de la probabilidad de que M_n sea mayor que x. Fijamos el valor $x = 0.7$ y obtenemos las gráficas de $\ln \mathbb{P}(M_n > x)$ contra n que se muestran en la figura 2.

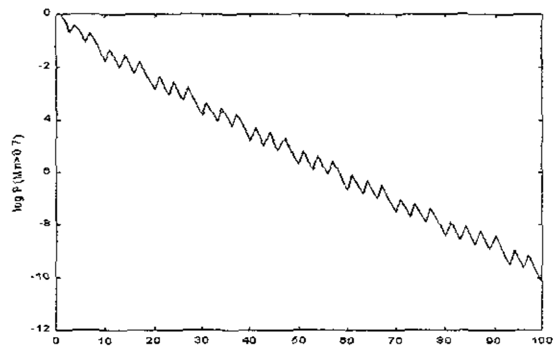

Figura 2. $\log \mathbb{P}(M_n > 0.7)$ contra n.

En las gráficas podemos observar que aunque éstas tienen oscilaciones (las cuales se repiten periódicamente), cuando el valor de n es grande podemos ver que la gráfica de $\log \mathbb{P}(M_n > x)$ contra n tiene un comportamiento aproximadamente lineal. Ahora cambiamos el valor a $x = 0.8$ y después a $x = 0.9$ y obtenemos las gráficas que se muestran en la figura 3; otra vez para valores de n grandes éstas siguen un comportamiento aproximadamente lineal, si denotamos por $-I(x)$ al valor de la pendiente asintótica la cual es negativa, tenemos la siguiente aproximación.

$$\log \mathbb{P}(M_n > x) \approx -nI(x).$$

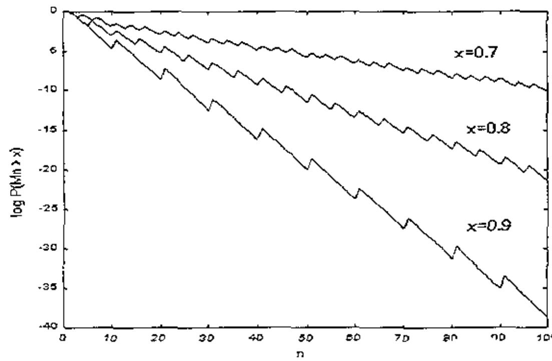

Figura 3. $\log \mathbb{P}(M_n > x)$ para $x = 0.8$ y $x = 0.9$ contra n.

2. Escogemos cualquier punto x menor que la media de la distribución (1, y obtenemos como en el caso anterior las gráficas de $\ln \mathbb{P}(M_n < x)$ con n para distintos valores de x (figura 4), $x = 0.4$, $x = 0.2$, $x = 0.1$ y n va de 1 a 100. Observemos como antes que para n grande, la gráfica tiene comportamiento aproximadamente lineal con pendiente asintótica $-I($ por consiguiente

$$\log \mathbb{P}(M_n < x) \approx -nI(x).$$

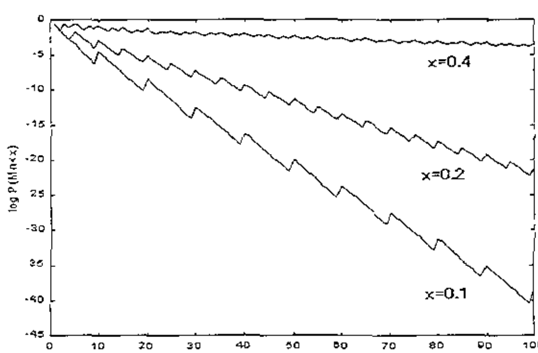

Figura 4. $\log \mathbb{P}(M_n < x)$ para $x = 0.4$ y $x = 0.2$ y $x = 0.1$ contra n.

Finalmente de las gráficas anteriores, para $1/2 < x < 1$ e igualmente p $0 < x < 1/2$ medimos la pendiente asintótica $-I(x)$ en cada caso y graficar s valores de $I(x)$ contra x, es decir, obtenemos la gráfica que corresponde a lo ıe llamaremos la *función de tasa* $I(x)$ que se muestra en la figura 5.

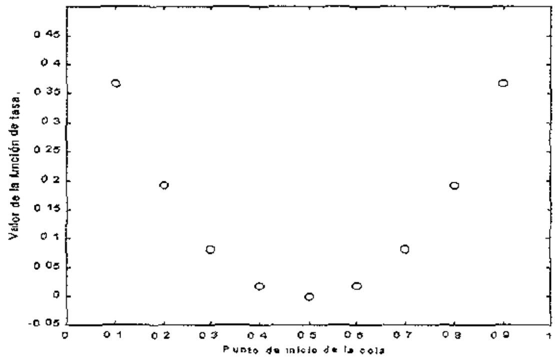

Figura 5. Función de tasa $I(x)$.

Podemos decir entonces que:*La cola de la distribución del número promedio e caras en n lanzamientos decae exponencialmente conforme n se incrementa.* s decir,

$$\begin{aligned} \mathbb{P}(M_n > x) &\approx e^{-nI(x)} \qquad \text{para} \quad x > 1/2 \\ \mathbb{P}(M_n < x) &\approx e^{-nI(x)} \qquad \text{para} \quad x < 1/2 \end{aligned} \tag{1.1.1}$$

n crece.

Como se mencionó en la introducción, la teoría de las grandes desviaciones studia los sucesos poco frecuentes, los sucesos exponenciales pequeños cuando pasa al límite y uno de los objetivos principales es encontrar una manera stemática de calcular la función de tasa. Más adelante mostraremos una forma hacer esto en algunos casos.

Observación. *La Ley Débil de los Grandes Números.*

Una consecuencia del resultado anterior, el cual se demostrará rigurosamente ás adelante, es la Ley Débil de los Grandes Números en el caso del lanzamiento monedas, tenemos que para n grande la distribución de M_n se concentra rededor de la media, es decir, para todo $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|M_n - \frac{1}{2}| < \varepsilon) = 1. \tag{1.1.2}$$

Otra forma de decirlo es que las colas se hacen cada vez más pequeñas, es dec para todo $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|M_n - \frac{1}{2}| > \varepsilon) = 0,$$

para lo cual vemos que

$$\mathbb{P}(|M_n - \frac{1}{2}| > \varepsilon) = \mathbb{P}(M_n > \frac{1}{2} + \varepsilon) + \mathbb{P}(M_n < \frac{1}{2} - \varepsilon).$$

Como las colas de la distribución de M_n para n grande decaen exponencialmen usando 1.1.1 en el miembro derecho de la igualdad anterior tenemos que

$$\begin{aligned} \mathbb{P}(M_n > x) &\approx e^{-nI(x)} \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{para} \quad x > 1/2 \quad \text{y} \\ \mathbb{P}(M_n < x) &\approx e^{-nI(x)} \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{para} \quad x < 1/2 \end{aligned}$$

obteniendo el resultado 1.1.2.

1.2 Colas de la distribución normal.

El siguiente ejemplo, el estudio de las colas gaussianas, es la forma clásica co se aborda la idea subyacente en la teoría de grandes desviaciones. También verá que no siempre funciona este método para el caso de otras distribucion Consideremos $X_1 \ldots X_n$ una sucesión de variables aleatorias independiente idénticamente distribuídas (v.a.i.i.d.), $X_i \sim N(0,1)$. Si

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

entonces

$$\overline{X}_n \sim N(0, 1/n),$$

por lo tanto

$$\mathbb{P}(\overline{X}_n \geq t) \underset{n\to\infty}{\longrightarrow} 0.$$

La idea es entonces buscar la velocidad de la convergencia a cero, y para ha esto hacemos estimaciones de "colas" de gaussianas. Sean $a > 0$ y $Y \sim N(0$

$$\mathbb{P}(Y \geq a) = \int_a^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx.$$

tiene, para $\mathbb{P}(Y \geq a)$, que $(Y/a) \geq 1$, por lo que

$$\begin{aligned}\mathbb{P}(Y \geq a) &\leq \frac{1}{\sqrt{2\pi}}\left[\frac{1}{a}\int_a^{+\infty} x \exp\left(\frac{-x^2}{2}\right) dx\right] \\ &= \frac{1}{\sqrt{2\pi}}\left[\frac{\exp\left(-\frac{a^2}{2}\right)}{a}\right]\end{aligned}$$

ɔara $\mathbb{P}(Y \leq a)$ se tiene que $(Y/a) \leq 1$, por lo que

$$\begin{aligned}\mathbb{P}(Y \leq a) &\geq \frac{1}{\sqrt{2\pi}}\left[\frac{1}{2a}\int_a^{2a} x \exp\left(\frac{-x^2}{2}\right) dx\right] \\ &= \frac{1}{\sqrt{2\pi}}\left[\frac{\exp\left(-\frac{a^2}{2}\right) - \exp\left(-2a^2\right)}{a}\right].\end{aligned}$$

en el caso que nos interesa.

$$\mathbb{P}(Y \geq t) \leq \frac{1}{\sqrt{2\pi}}\left[\frac{1}{t\sqrt{n}} \exp\left(\frac{-nt^2}{2}\right)\right] \quad \text{y}$$

$$\mathbb{P}(Y \leq t) \geq \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{t\sqrt{n}}\left[\exp\left(-\frac{nt^2}{2}\right) - \exp\left(-2nt^2\right)\right],$$

namos logaritmo, dividimos entre n y finalmente tomando límite superior e erior, al cual denotaremos respectivamente como $\overline{\lim}$ y $\underline{\lim}$, tenemos

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\log(\mathbb{P}(\overline{X}_n \geq t)) \leq -\frac{t^2}{2} \quad \text{y} \quad \underline{\lim_{n\to\infty}} \frac{1}{n}\log(\mathbb{P}(\overline{X}_n \leq t)) \geq -\frac{t^2}{2}.$$

Lo que nos da la convergencia hacia $-t^2/2$. Observemos que en este caso la ıción de tasa $I(x)$ como se vió en la sección anterior es

$$I(x) = \frac{x^2}{2}. \qquad 1.2.1$$

Sin embargo, existen casos donde esta forma de resolver el problema no es ıveniente, lo que veremos a continuación.

1.3 Colas de la distribución estable.

Definición 1.3.1. Se dice que una v.a. X tiene distribución estable si par toda $k > 0$, y para toda sucesión $X_1, \ldots, X_k$ de v.a.i.i.d. con la distribución d X, existen constantes $a_k > 0$, b_k tal que

$$\mathcal{L}(X_1 + \cdots + X_k) = \mathcal{L}(a_k X + b_k)$$

donde $\mathcal{L}$ denota la distribución de probabilidad de la variable.

Si X tiene una distribución estable y si además esta distribución es simétric es decir, $F(x) = F(-x)$ para toda $x > 0$, entonces ocurre alguna de las siguiente dos cosas:

(a) X tiene una distribución normal$(\alpha = 2)$.

(b) Existe $\alpha \in (0, 2)$ y tal que la función característica de X es

$$\mathbb{E}[\exp(ixX)] = \phi_\alpha(x) = \exp(-d|x|^\alpha) \qquad d > 0. \qquad 1.3.$$

A α se le llama índice de la distribución estable y d es un número rea y positivo, ver [Bre75] pag. 199 y 204. El hecho de que la distribució estable sea simétrica simplifica significativamente la expresión de la funció característica en el caso general.

También en [Bre75] pag. 202 se prueba la siguiente estimación para la "cola de esta distribución.

Si $M^+(t) := \mathbb{P}(X \geq t)$, $t > 0$, entonces

$$M^+(t) = t^{-\alpha} M^+(1).$$

Es decir,

$$\mathbb{P}(X \geq t) \approx C t^{-\alpha}. \qquad 1.3.$$

: otra parte, si $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ entonces

$$\begin{aligned}
\mathbb{E}[\exp(ix\overline{X}_n)] &= \mathbb{E}\big[\exp\big(\frac{1}{n}ix\sum_{j=1}^{n} X_j\big)\big] \\
&= \mathbb{E}\big[\exp\big(\sum_{j=1}^{n}(\frac{1}{n}ix)X_j\big)\big] = \prod_{j=1}^{n}\mathbb{E}\big[\exp\big(\frac{ix}{n}X_j\big)\big] \\
&= \prod_{j=1}^{n}\phi_\alpha\big(\frac{x}{n}\big) = \big[\exp(-d\big|\frac{x}{n}\big|^\alpha)\big]^n \\
&= \exp\big(-d|x|^\alpha n^{1-\alpha}\big) = \exp\big(-d\big|n^{\frac{1-\alpha}{\alpha}}x\big|^\alpha\big) \\
&= \phi_\alpha\big(n^{\frac{1-\alpha}{\alpha}}x\big) = \mathbb{E}\big[\exp\big(n^{\frac{1-\alpha}{\alpha}}X_1\big)\big].
\end{aligned}$$

r lo que

$$\overline{X}_n \sim n^{\frac{1-\alpha}{\alpha}}X_1. \qquad 1.3.3$$

amos ahora $\mathbb{P}(\overline{X}_n \geq t)$, de 1.3.2 y 1.3.3 tenemos que

$$\mathbb{P}(\overline{X}_n \geq t) \approx Ct^{-\alpha}n^{1-\alpha},$$

nando logaritmo, dividimos entre n, y tomando límite obtenemos

$$\frac{1}{n}\ln\mathbb{P}(\overline{X}_n \geq t) \approx \frac{1}{n}(\ln Ct^{-\alpha} + (1-\alpha)\log n)$$

: lo tanto

$$\frac{1}{n}\ln\mathbb{P}(\overline{X}_n \geq t) \underset{n\to\infty}{\longrightarrow} 0.$$

n esta aproximación no encontramos una función que nos indique la velocidad convergencia hacia cero. Por lo que, obteniendo cotas para la cola de la tribución no siempre se obtiene un resultado satisfactorio, por lo tanto se ablecerá una manera sistematica de hallar la función de tasa.

A continuación daremos la definicion del PGD.

4 El Principio de Grandes Desviaciones.

las secciones anteriores vimos algunas técnicas para encontrar la función tasa, también se vió que no siempre es posible hallar dicha tasa. Ahora se mostrará que bajo ciertas condiciones podremos encontrar una manera sistem tica de hallar esta función de tasa. Para definir el PGD, previamente darem la definición de función de tasa.

Usaremos la siguiente notación: A^o y $\overline{A}$ denota el interior y la cerradura d conjunto A respectivamente y A^c el complemento de A para cualquier conjun A. El ínfimo de una función sobre un conjunto vacio se interpreta como ∞.

Definición 1.4.1. Sea f una función en un espacio métrico X en $[0, +\infty]$, s dice que f es semicontinua inferiormente (s.c.i.) si una de las dos propiedad siguientes se cumple:

1. Si $x_n \to x$ entonces $f(x) \leq \varliminf_{n\to\infty} f(x_n)$.
2. Para todo a, el conjunto de nivel $\{x : f(x) \leq a\}$ es cerrado.

Es claro que una función continua es semicontinua inferiormente.

Definición 1.4.2. Diremos que una función $I : X \to [0, \infty]$ es una *función tasa* si I es semicontinua inferiormente. Además, si para todo $a > 0$ el conjun de nivel $\{x : I(x) \leq a\}$ es compacto, entonces I es una *función de tasa buena*

Definición 1.4.3. PGD. Sea X un espacio métrico y sea $\{\mu_\varepsilon\}, \varepsilon > 0$ un familia de probabilidades sobre un σ-álgebra F sobre X. Diremos que la famil $\{\mu_\varepsilon\}, \varepsilon > 0$ satisface un Principio de Grandes Desviaciones (PGD) con funci de tasa I si para todo $A \in F$ tenemos que

$$\begin{aligned} -\inf\{I(x) : x \in A^o\} &\leq \varliminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \\ &\leq \varlimsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \qquad\qquad 1.4. \\ &\leq -\inf\{I(x) : x \in \overline{A}\}. \end{aligned}$$

Observaciones.

1. El PGD caracteríza el comportamiento en el límite de una familia de m didas de probabilidad $\{\mu\}_\epsilon$ conforme $\epsilon \to 0$ sobre un σ-álgebra F en u espacio métrico X en términos de una función de tasa. Esta caracter zación está dada por las cotas exponenciales asintóticas superior e inferi sobre los valores que μ_ϵ asigna a los subconjuntos medibles de X. En l casos que consideraremos, F será la σ-álgebra de Borel sobre X.

2. Es claro que si $\{\mu\}_\epsilon$ satisface el PGD y $A \in F$ es tal que

$$\inf_{x\in A^o} I(x) = \inf_{x\in \overline{A}} I(x) := I_A, \qquad 1.4.2$$

entonces

$$\lim_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) = -I_A. \qquad 1.4.3$$

Un conjunto A que satisface 1.4.2 es llamado un conjunto de *continuidad.* En general, el PGD implica un limite preciso en 1.4.3 solamente para conjuntos de continuidad A.

3. En cualquier situación que involucre medidas aisladas(no atómicas), es decir, $\mu_\epsilon(\{x\}) = 0$ para todo $x \in X$. Tenemos en la cota inferior de 1.4.1 que

$$-\inf_{x\in A^o} I(x) \leq \underline{\lim}_{\epsilon\to 0} \epsilon \log(0) = -\infty,$$

y si tomamos el ínfimo sobre A en vez de A^o, concluímos que $I(x) = \infty$, lo cual contradice la cota superior en 1.4.1 porque $\mu_\epsilon(X) = 1$ para todo ϵ. Por lo que es necesario que $\inf_{x\in X} I(x) = 0$ y así la cota superior es válida. Y si I es una función de tasa buena, significa que existe al menos un punto x tal que $I(x) = 0$.

Luego, la cota superior es trivial siempre que $\inf_{x\in\overline{A}} I(x) = 0$, de igual manera, la cota inferior también es trivial siempre que $\inf_{x\in A^o} I(x) = \infty$.

Esto nos lleva a una formulación alternativa del PGD, que resulta ser útil cuando este se demuestra. Supongamos que I es una función de tasa y $\Gamma(a) = \{x : I(x) < a\}$ su conjunto de nivel. Entonces, la desigualdad 1.4.1 es equivalente a las siguientes cotas:

(a) *Cota Superior:* Para todo $a < \infty$ y todo conjunto medible A, tal que $\overline{A} \subset \Gamma(a)^c$

$$\overline{\lim}_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \leq -a, \qquad 1.4.4$$

lo anterior se sigue ya que si $x \in \overline{A}$

$$I(x) \geq a$$

y entonces

$$-\inf_{x\in\overline{A}} I(x) \leq -a.$$

(b) *Cota Inferior:* Para cualquier $x \in \{x : I(x) < \infty\}$ y cualquier conjun medible A y $x \in A^o$,

$$\varliminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A) \geq -I(x), \qquad 1.4$$

ya que para $x \in A^o$,

$$I(x) \geq \inf_{x\in A^o} I(x).$$

Por lo tanto

$$-I(x) \leq -\inf_{x\in A^o} I(x).$$

Como deciamos anteriormente, en los casos que consideraremos, F será σ-álgebra de Borel sobre X, y entonces 1.4.1 es equivalente a:

$$\begin{aligned} &\text{para todo } O \text{ abierto } \varliminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(O) \geq -\inf\{I(x) : x \in O\} \qquad \text{y} \\ &\text{para todo } F \text{ cerrado } \varlimsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\inf\{I(x) : x \in F\}. \end{aligned} \qquad 1.4$$

Si ahora consideramos $\{X_n\}_n$ una sucesión de v.a., la media empírica de v.a. es $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ y μ_n es la distribución de la media empírica, diremos q la sucesión de probabilidades μ_n satisface un Principio de Grandes Desviacion en el caso $\varepsilon = 1/n$ con función de tasa I si :

$$\begin{aligned} &\text{para todo } O \text{ abierto } \varliminf_{n\to\infty} \frac{1}{n} \log \mu_n(O) \geq -\inf\{I(x) : x \in O\} \qquad \text{y} \\ &\text{para todo } F \text{ cerrado } \varlimsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf\{I(x) : x \in F\}. \end{aligned}$$

1.4

apítulo 2

GD para Espacios Discretos de 'imensión Finita.

étodos Combinatorios.

este capítulo, utilizaremos métodos combinatorios para obtener el PGD para distribución empírica de procesos que toman valores en un conjunto finito y ra la correspondiente media empírica. El empleo de estos métodos está limilo a espacios de dimensión finita. Estos métodos nos ilustrarán los resultados e se esperarían obtener para conjuntos más abstractos. También, daremos s ejemplos de como se aplica la teoría.

1 El Método de Tipos.

este capítulo, salvo que se indique lo contrario, todas las variables aleatorias e consideremos toman valores en un conjunto finito $\Sigma = \{a_1, \ldots, a_{|\Sigma|}\}$, donde denota la cardinalidad o tamaño de un conjunto A. A Σ se le conoce como ılfabeto subyacente. Denotaremos por $M_1(\Sigma)$ al espacio de todas las medidas probabilidad (leyes) en el alfabeto Σ. $M_1(\Sigma)$ se identifica con un subconjunto $\mathbb{R}^{|\Sigma|}$. El conjunto de todos los vectores $|\Sigma|$-dimensionales con entradas reales -negativas que suman 1. A $M_1(\Sigma)$ se le asocia la topología inducida por los ertos en $\mathbb{R}^{|\Sigma|}$.

Sea $Y_1, Y_2, \ldots, Y_n$ una sucesión de v.a.i.i.d. con distribución de probabilidad $\in M_1(\Sigma)$. Definimos Σ_μ como el soporte de la distribución de probabilidad es decir, $\Sigma_\mu = \{a_i : \mu(a_i) > 0\}$. En general, Σ_μ podría ser un subconjunto propio de Σ, y cuando consideremos una medida μ podemos suponer, sin pérd de generalidad, que $\Sigma_\mu = \Sigma$ al ignorar aquellos elementos que aparecen probabilidad cero.

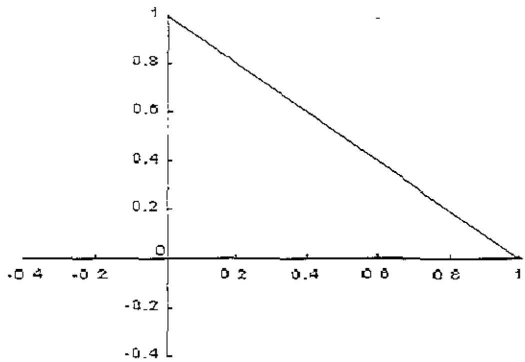

Figura 6. $M_1(\Sigma)$ para $|\Sigma| = 2$.

A continuación daremos la definición de **tipo**, el cual esta asociado a sucesión finita.

Definición 2.1.1. El **tipo** $\mathbf{L}_\mathbf{n}^\mathbf{y}$ de una sucesión finita $\mathbf{y} = (y_1, \ldots, y_n) \in \Sigma$ la distribución de probabilidad empírica (o medida empírica) inducida por sucesión. Explícitamente, $\mathbf{L}_\mathbf{n}^\mathbf{y} = (L_n^\mathbf{y}(a_1), \ldots, L_n^\mathbf{y}(a_{|\Sigma|}))$ es el elemento de M_1 dado por

$$L_n^\mathbf{y}(a_i) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{a_i}(y_j), \quad i = 1, \ldots, |\Sigma|$$

es decir, $L_n^\mathbf{y}(a_i)$ es el promedio de ocurrencias de a_i en la sucesión $y_1, \ldots, y_n$

Definimos $\mathcal{L}_n$ como conjunto de todas las posibles sucesiones de **tipos** longitud n. Entonces $\mathcal{L}_n := \{v : v = L_n^\mathbf{y} \text{ para alguna } \mathbf{y}\} \subset \mathbb{R}^{|\Sigma|}$ y la distribu de probabilidad empírica $L_n^\mathbf{Y}$ asociada con la sucesión $\mathbf{Y} := (Y_1, \ldots, Y_n)$ es elemento aleatorio del conjunto $\mathcal{L}_n$. Los conceptos anteriores son útiles com verá más adelante. Por ahora, implican las siguientes estimaciones del volu y la distancia variacional.

Se define la **distancia variacional** entre dos medidas en $M_1(\Sigma)$ como $d_V(v, v') := \sup_{A \subset \Sigma} |v(A) - v'(A)|$.

Podemos decir que la distancia variacional mide la 'semejanza' entre dos
didas de probabilidad. Para mayores referencias ver [Cor94] pag.52-57.

Tenemos el siguiente lema.

ma 2.1.1. *Con la notación definida arriba,*

ı) $|\mathcal{L}_n| \leq (n+1)^{|\Sigma|}$.

›) Para todo vector de probabilidad $v \in M_1(\Sigma)$

$$d_{\mathrm{V}}(v, \mathcal{L}_n) := \inf_{v' \in \mathcal{L}_n} d_{\mathrm{V}}(v, v') \leq \frac{|\Sigma|}{2n}. \qquad 2.1.1$$

Demostración.

Observemos que cualquier componente del vector $L_n^{\mathbf{y}}$ pertenece al conjunto
$= \{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\}$, cuya cardinalidad es $(n+1)$, y como el vector $L_n^{\mathbf{y}}$ está dado
r a lo más $|\Sigma|$ de tales cantidades, la parte (a) del lema está demostrada. Más
n, como los componentes de dicho vector suman 1, la última entrada está
da automáticamente al conocer las primeras $|\Sigma| - 1$ coordenadas del vector,
r lo que $|\mathcal{L}_n| \leq (n+1)^{|\Sigma|-1}$.

Para demostrar la parte (b), antes probaremos la siguiente expresión equiva-
te a la distancia variacional

$$d'_{\mathrm{V}}(v, v') = \frac{1}{2} \sum_{i=1}^{|\Sigma|} |v(a_i) - v'(a_i)|.$$

prueba de la igualdad anterior se hace en dos partes:

i) $d'_{\mathrm{V}}(v, v') \leq d_{\mathrm{V}}(v, v')$.

Sea $I = \{a_i : v(a_i) > v'(a_i)\}$, entonces,

$$\begin{aligned}
2d'_{\mathrm{V}}(v,v') &= \sum_{i=1}^{|\Sigma|} |v(a_i) - v'(a_i)| \\
&= \sum_{a_i \in I} v(a_i) - v'(a_i) + \sum_{i \in I^C} v'(a_i) - v(a_i) \\
&\leq |\sum_{a_i \in I} v(a_i) - v'(a_i)| + |\sum_{a_i \in I^C} v(a_i) - v'(a_i)| \\
&= |v(I) - v'(I)| + |v(I^c) - v'(I^c)| \\
&\leq 2\sup_{A \subset \Sigma}\{v(A) - v'(A)\} \\
&= 2\, d_{\mathrm{V}}(v,v')
\end{aligned}$$

por lo tanto, $d'_{\mathrm{V}}(v,v') \leq d_{\mathrm{V}}(v,v')$.

ii) $d_{\mathrm{V}}(v,v') \leq d'_{\mathrm{V}}(v,v')$.

Sea $A \subset \Sigma$, entonces,

$$\begin{aligned}
2|v(A) - v'(A)| &= |v(A) - v'(A)| + |v(A) - v'(A)| \\
&= |v(A) - v'(A)| + |v(A^c) - v'(A^c)| \\
&= |\sum_{a_i \in A} v(a_i) - v'(a_i)| + |\sum_{a_i \in A^c} v(a_i) - v'(a_i)| \\
&\leq \sum_{a_i \in A} |v(a_i) - v'(a_i)| + \sum_{a_i \in A^c} |v(a_i) - v'(a_i)| \\
&= \sum_{i=1}^{|\Sigma|} |v(a_i) - v'(a_i)| \\
&= 2d'_{\mathrm{V}}(v,v')
\end{aligned}$$

y por lo tanto,

$$2\sup_{A \subset \Sigma}\{|v(A) - v'(A)|\} \leq 2d'_{\mathrm{V}}(v,v').$$

De esta forma obtenemos que $d_{\mathrm{V}}(v,v') = \dfrac{1}{2}\displaystyle\sum_{i=1}^{\Sigma} |v(a_i) - v'(a_i)|$.

Como se definió anteriormente, $\mathcal{L}_n$ contiene todos los vectores de probabilid compuestos de $|\Sigma|$ componentes del conjunto A, por lo que, para toda $v \in M_1($ existe un $v' \in \mathcal{L}_n$ con $|v(a_i) - v'(a_i) \leq \frac{1}{n}$ para $i = 1, ..., |\Sigma|$; y por la iguald nterior

$$d_V(v, v') = \frac{1}{2}\sum_{i=1}^{|\Sigma|} |v(a_i) - v'(a_i)| \leq \frac{|\Sigma|}{2n}.$$

on lo que se tiene el resultado.

Observemos que para cada $v \in \mathcal{L}_n$ existen distintas sucesiones $y = (y_1, ...y_n)$ ue inducen la medida empírica v. Esto nos lleva a la siguiente definición.

efinición 2.1.2. La clase tipo $\mathbf{T_n(v)}$ de una distribución de probabilidad $\in \mathcal{L}_n$ es el conjunto $T_n(v) = \{\mathrm{y} \in \Sigma^n : L_n^{\mathrm{y}} = v\}$.

Note que una clase tipo consiste de todas las permutaciones de un vector dado n el conjunto. En la siguiente definición se usará la convención $0 \log 0 := 0$ y $\log(0/0) := 0$.

efinición 2.1.3. Entropía y **Entropía Relativa.**

a) La entropía de un vector de probabilidad v es

$$H(v) := -\sum_{i=1}^{|\Sigma|} v(a_i) \log v(a_i).$$

b) La entropía relativa de un vector de probabilidad v con respecto a otro vector de probabilidad μ es

$$H(v \mid \mu) := \sum_{i=1}^{|\Sigma|} v(a_i) \log \frac{v(a_i)}{\mu(a_i)}.$$

Nota : El nombre de entropía de un vector de probabilidad $v, H(v)$, es un rmino conocido en el área de teoría de la información. Esta cantidad es inrpretada como la cantidad promedio de información recibida cuando el valor v es observado. Otra interpretación es , $H(v)$ mide la no-uniformidad de na medida de probabilidad. A la entropía también se le conoce con el nombre entropía de Shannon. Al final de este capítulo veremos un ejemplo en este ntexto.

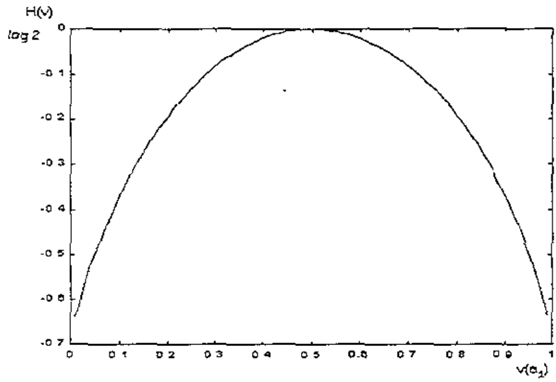

Figura 7. $H(v)$ para $|\Sigma| = 2$.

Demostraremos ahora que la entropía relativa es una función de tasa buena es decir, no negativa, semicontinua inferiormente y para todo $a \in \mathbb{R}$ el conjunt $\{v : H(\cdot|\mu) \geq a\}$ es compacto, (ver definición 1.4.2). Veamos que $H(\cdot|\mu)$ e una función no negativa. La desigualdad de Jensen afirma que: si X es un variable aleatoria y $U : \mathbb{R} \to \mathbb{R}$ una función convexa, entonces, $E(U(X))$ ≥ $U(E(X))$. Si aplicamos esta desigualdad a la función convexa $x \log x$ teneme que $E(x \log x) \geq E(x) \log E(x)$. Sea $v \in M_1(\Sigma)$ y con $x = \frac{v(a_i)}{\mu(a_i)}$ se tiene

$$
\begin{aligned}
H(v|\mu) = \sum_{i=1}^{\Sigma} v(a_i) \log \frac{v(a_i)}{\mu(a_i)} &= \sum_{i=1}^{\Sigma} \frac{v(a_i)}{\mu(a_i)} \log \left(\frac{v(a_i)}{\mu(a_i)} \right) \mu(a_i) \\
&\geq \sum_{i=1}^{\Sigma} \frac{v(a_i)}{\mu(a_i)} \mu(a_i) \log \sum_{i=1}^{\Sigma} \frac{v(a_i)}{\mu(a_i)} \mu(a_i) \\
&= \log 1 = 0
\end{aligned}
$$

por lo que $H(v|\mu)$ es no negativa.

Además, $H(\cdot|\mu)$ es continua en el conjunto compacto $\{v \in M_1(\Sigma) : \Sigma_v \subset \Sigma_\mu$ ya que $x \log x$ es continua en $0 \leq x \leq 1$ y $H(v|\mu) = \infty$ en el complemento d este conjunto, de donde se sigue que $H(\cdot|\mu)$ es una función de tasa buena.

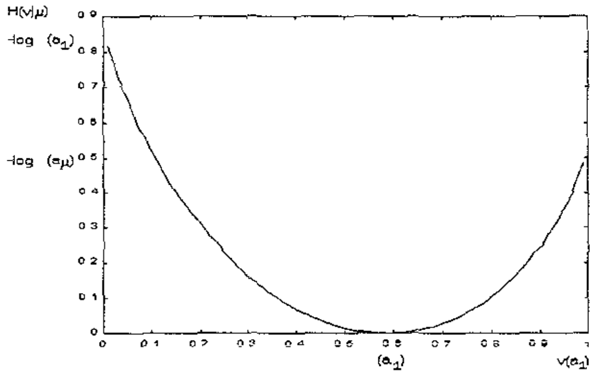

Figura 8. $H(v|\mu)$ para $|\Sigma| = 2$. y $\mu = (0.5842, .4158)$.

Los siguientes tres lemas nos dan estimaciones de las probabilidades de los eventos $\{L_n^{\mathbf{Y}} = v\}$, $v \in \mathcal{L}_n$, los cuales son eventos raros. Primero se demuestra que los resultados que pertenecen a la misma clase tipo son igualmente probables, y luego se estima el crecimiento exponencial de la tasa de cada clase tipo.

Sea $\mathbb{P}_\mu$ la ley de probabilidad $\mu^{\mathbb{Z}_-}$ asociada a una sucesión infinita de v.a.i.i.d. $\{Y_j\}$ distribuidas de acuerdo a $\mu \in M_1(\Sigma)$. En el siguiente lema veremos que las sucesiones que pertenecen a la misma clase tipo tienen la misma probabilidad.

Lema 2.1.2. *Si $y \in T_n(v)$ para $v \in \mathcal{L}_n$, entonces*

$$\mathbb{P}_\mu((Y_1, Y_2, \ldots, Y_n) = \mathbf{y}) = e^{-n[H(v)+H(v|\mu)]}.$$

Demostración.

Por la propiedad de independencia tenemos que

$$\mathbb{P}_\mu((Y_1, Y_2, \ldots, Y_n) = \mathbf{y}) = \prod_{i=1}^{|\Sigma|} \mathbb{P}_\mu(Y_i = y_i)$$

y por la definición de **tipo** (ver 2.1.1) se tiene

$$\begin{aligned}\prod_{i=1}^{|\Sigma|} \mathbb{P}_\mu(Y_i = y_i) &= \prod_{i=1}^{|\Sigma|} \mu(a_i)^{nv(a_i)} \\ &= \exp\left(\log\Big(\prod_{i=1}^{|\Sigma|} \mu(a_i)^{nv(a_i)}\Big)\right)\end{aligned}$$

$$= \exp\left(n\sum_{i=1}^{|\Sigma|} v(a_i)\log\left(\mu(a_i)\right)\right)$$
$$= e^{-n[H(v)+H(v|\mu)]}$$

donde la última igualdad se explica debido a que

$$-[H(v)+H(v\mid\mu)] = -\sum_{i=1}^{|\Sigma|} v(a_i)\left(\log\frac{v(a_i)}{\mu(a_i)} - \log v(a_i)\right)$$
$$= \sum_{i=1}^{|\Sigma|} v(a_i)\log\mu(a_i).$$

En particular, debido a que $H(\mu\mid\mu)=0$, se sigue que para todo $\mu\in\mathcal{L}_n$ y $y\in T_n(\mu)$

$$\mathbb{P}_\mu((Y_1,Y_2,\ldots,Y_n)=y) = e^{-nH(\mu)}. \qquad 2.1$$

Con lo que se concluye la demostración.

Se tiene una estimación de la cardinalidad de la clase tipo.

Lema 2.1.3. *Para todo $v\in\mathcal{L}_n$,*

$$(n+1)^{-|\Sigma|}e^{nH(v)} \le |T_n(v)| \le e^{nH(v)}.$$

Observación.

Como la clase tipo $T_n(v)$ consiste de todas las diferentes permutaciones de elementos del vector dado en el conjunto, de los cuales $nv(a_i)$, $(i=1,\ldots,|\Sigma$ son iguales, tenemos que,

$$|T_n(v)| = \frac{n!}{(nv(a_1))!\ldots(nv(a_{|\Sigma|}))!} \qquad i=1,2,...,|\Sigma|. \qquad 2.$$

Usando la aproximación de Stirling se puede obtener una estimación de $|T_n(v$ la cual se puede ver en [Fel71].

Otra manera de obtener una estimación para $|T_n(v)|$ se muestra aquí.

Demostración.

Bajo $\mathbb{P}_v$, cualquier clase tipo tiene probabilidad a lo más uno y todos eventos son equiprobables. Por lo que, para todo $v\in\mathcal{L}_n$ y por la ecuación 2.

enemos

$$1 \geq \mathbb{P}_v(L_n^{\mathbf{Y}} = v) = \mathbb{P}_v((Y_1, \ldots, Y_n) \in T_n(v)) = e^{-nH(v)}|T_n(v)|$$

›or lo tanto

$$|T_n(v)| \leq e^{nH(v)}.$$

Encontraremos ahora la cota inferior. Sea $v' \in \mathcal{L}_n$ tal que $\Sigma_{v'} \subset \Sigma_v$, y por ›onveniencia de notación restringiremos Σ de manera tal que $\Sigma_v = \Sigma$. Entonces,

$$\begin{aligned}
\frac{\mathbb{P}_v(L_n^Y = v)}{\mathbb{P}_v(L_n^Y = v')} &= \frac{|T_n(v)|\prod_{i=1}^{|\Sigma|} v(a_i)^{nv(a_i)}}{|T_n(v')|\prod_{i=1}^{|\Sigma|} v(a_i)^{nv'(a_i)}} \\
&= \frac{\big(nv'(a_1)\big)!\ldots\big(nv'(a_{|\Sigma|})\big)!}{(nv(a_1))!\ldots(nv(a_{|\Sigma|}))!}\prod_{i=1}^{|\Sigma|} v(a_i)^{\left(nv(a_i)-nv'(a_i)\right)} && \text{por 2.1.3} \\
&= \prod_{i=1}^{|\Sigma|}\frac{(nv'(a_i))!}{(nv(a_i))!} v(a_i)^{\left(nv(a_i)-nv'(a_i)\right)}.
\end{aligned}$$

Esta última expresión es un producto de términos de la forma $\frac{m!}{\ell!}\left(\frac{\ell}{n}\right)^{\ell-m}$. /eamos que $\frac{m!}{\ell!} \geq \ell^{m-\ell}$ para todo $m, \ell \in \mathcal{Z}^+$. Consideremos el caso $m \geq \ell$ (para $n < \ell$ la prueba es similar), entonces existe $k \in \mathcal{Z}^+$ tal que $m = \ell + k$, así

$$\frac{m!}{\ell!} = (m)(m-1)\ldots(m-(k-1)) \geq \ell^k = \ell^{m-\ell}.$$

De la desigualdad anterior con $m = nv'(a_i)$ y $l = nv(a_i)$ obtenemos que

$$\begin{aligned}
\frac{\mathbb{P}_v(L_n^{\mathbf{Y}} = v)}{\mathbb{P}_v(L_n^{\mathbf{Y}} = v')} &= \prod_{i=1}^{|\Sigma|}\frac{(nv'(a_i))!}{(nv(a_i))!} v(a_i)^{nv(a_i)-nv'(a_i)} \\
&\geq \prod_{i=1}^{|\Sigma|}\big(nv(a_i)\big)^{nv'(a_i)-nv(a_i)}\left[v(a_i)^{nv(a_i)-nv'(a_i)}\right] \\
&= \prod_{i=1}^{|\Sigma|} n^{nv'(a_i)-nv(a_i)} \\
&= n^{n\left(\sum_{i=1}^{|\Sigma|} v'(a_i)-\sum_{i=1}^{|\Sigma|} v(a_i)\right)} = 1.
\end{aligned}$$

Notemos que $\mathbb{P}_v(L_n^{\mathbf{Y}} = v) > 0$ solamente cuando $\Sigma_{v'} \subset \Sigma_v$ y $v' \in \mathcal{L}_n$, por tanto lo anterior implica que para todo $v, v' \in \mathcal{L}_n$

$$\mathbb{P}_v(L_n^{\mathbf{Y}} = v) \geq \mathbb{P}_v(L_n^{\mathbf{Y}} = v')$$

con lo que

$$\begin{aligned} 1 &= \sum_{v' \in \mathcal{L}_n} \mathbb{P}_v(L_n^Y = v') \leq |\mathcal{L}_n| \mathbb{P}_v(L_n^{\mathbf{Y}} = v) \\ &= |\mathcal{L}_n| e^{-nH(v)} |T_n(v)| \\ &\leq (n+1)^{|\Sigma|} e^{-nH(v)} |T_n(v)|, \end{aligned}$$

la última desigualdad se sigue del lema 2.1.1 y asi obtenemos la cota inferior

$$|T_n(v)| \geq (n+1)^{-|\Sigma|} e^{nH(v)}.$$

Lema 2.1.4 (Estimación de Grandes Desviaciones). *Para cualquier v* $\mathcal{L}_n$

$$(n+1)^{-|\Sigma|} e^{-nH(v|\mu)} \leq \mathbb{P}_\mu(L_n^{\mathbf{Y}} = v) \leq e^{-nH(v|\mu)}.$$

Demostración. La prueba es sencilla aplicando los lemas antes demostrad Por el lema 2.1.2.

$$\begin{aligned} \mathbb{P}_\mu(L_n^Y = v) &= |T_n(v)| \mathbb{P}_\mu((Y_1, \ldots, Y_n) = \mathbf{y},\, L_n^{\mathbf{y}} = v) \\ &= |T_n(v)| e^{-n[H(v) + H(v|\mu))} \\ &\leq e^{nH(v)} e^{-n[H(v)+H(v|\mu))} = e^{-n[H(v)+H(v|\mu))}. \end{aligned}$$

La ultima desigualdad se tiene del lema 2.1.3 y por la cota inferior del mis lema se concluye la demostración.

Veamos ahora el teorema de Sanov, el cual establece que la familia de d tribuciones de probabilidad $\mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \cdot)$ satisface el PGD con función de ta $H(\cdot|\mu)$. Comparese este resultado con 1.4.7.

Teorema 2.1 (Sanov). *Para todo conjunto* Γ *de vectores de probabilidad* $M_1(\Sigma)$.

$$\begin{aligned} -\inf_{v \in \Gamma^o} H(v \mid \mu) &\leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \Gamma) \\ &\leq \overline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \Gamma) \leq -\inf_{v \in \Gamma} H(v \mid \mu) \end{aligned} \quad 2.$$

donde Γ^o *es el interior de* $\Gamma \subset \mathbb{R}^{|\Sigma|}$.

Demostración.

Por el lema 2.1.1 y la cota superior del lema 2.1.4

$$\begin{aligned}\mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \Gamma) &= \sum_{v\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^{\mathbf{Y}} = v)\\ &\leq \sum_{v\in\Gamma\cap\mathcal{L}_n} e^{-nH(v|\mu)}\\ &\leq |\Gamma\cap\mathcal{L}_n| e^{-n\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v|\mu)}\\ &\leq (n+1)^{|\Sigma|} e^{-n\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v|\mu)}\end{aligned} \tag{2.1.5}$$

: la cota inferior del mismo lema se tiene

$$\begin{aligned}\mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \Gamma) &= \sum_{v\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^{\mathbf{Y}} = v)\\ &= \sum_{v\in\Gamma\cap\mathcal{L}_n} (n+1)^{-|\Sigma|} e^{-nH(v|\mu)}\\ &\geq (n+1)^{-|\Sigma|} e^{-n\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v|\mu)}.\end{aligned} \tag{2.1.6}$$

Dividimos entre n, tomando logaritmo y límite superior en 2.1.5 y 2.1.6, bido a que $\lim_{n\to\infty} \frac{1}{n}\log(n+1)^{|\Sigma|} = 0$ tenemos

$$\begin{aligned}\overline{\lim_{n\to\infty}} \tfrac{1}{n}\log(\mathbb{P}_\mu(L_n^{\mathbf{Y}} \in \Gamma)) &\leq \overline{\lim_{n\to\infty}} \tfrac{1}{n}\log((n+1)^{|\Sigma|} e^{-n\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v|\mu)})\\ &= \lim_{n\to\infty} \tfrac{1}{n}\log(n+1)^{|\Sigma} + \overline{\lim_{n\to\infty}} \left(-\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v\mid\mu)\right)\\ &= \overline{\lim_{n\to\infty}} \left(-\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v\mid\mu)\right)\end{aligned}$$

$$\begin{aligned}\overline{\lim_{n\to\infty}} \log(\mathbb{P}_\mu(L_n^{Y} \in \Gamma)) &\geq \overline{\lim_{n\to\infty}} \log((n+1)^{-|\Sigma|} e^{-n\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v|\mu)})\\ &\geq \overline{\lim_{n\to\infty}} \left(-\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v\mid\mu)\right)\end{aligned}$$

: lo que,

$$\begin{aligned}\overline{\lim_{n\to\infty}} \tfrac{1}{n}\log(\mathbb{P}_\mu(L_n^{Y} \in \Gamma)) &= \overline{\lim_{n\to\infty}} \left(-\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v\mid\mu)\right)\\ &= -\underline{\lim_{n\to\infty}} \left(\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v\mid\mu)\right)\end{aligned} \tag{2.1.7}$$

donde la última igualdad es cierta por propiedades de límite superior e inferi
De igual manera, como en 2.1.7, obtenemos una igualdad para el límite inferi

$$\begin{aligned}\varliminf_{n\to\infty} \tfrac{1}{n}\log(\mathbb{P}_\mu(L_n^Y \in \Gamma)) &= \varliminf_{n\to\infty}\left(-\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v \mid \mu)\right)\\ &= -\varlimsup_{n\to\infty}\left(\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v \mid \mu)\right).\end{aligned} \quad 2.$$

Para obtener la cota superior en la ecuación 2.1.4, observemos que $\Gamma\cap\mathcal{L}_n \subset$ para todo n, se sigue que $\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v \mid \mu) \geq \inf_{v\in\Gamma} H(v \mid \mu)$. Entonces

$$-\varliminf_{n\to\infty}\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v \mid \mu) \leq -\varliminf_{n\to\infty}\inf_{v\in\Gamma} H(v \mid \mu) = -\inf_{v\in\Gamma} H(v \mid \mu)$$

y por la ecuación 2.1.7 se tiene el resultado.

Ahora obtendremos la cota inferior en 2.1.4. Fijemos un punto arbitra $v \in \Gamma^o$ tal que $\Sigma_v \subset \Sigma_\mu$, entonces para un $\delta > 0$ suficientemente pequeño, conjunto $\{v' : d_V(v, v') < \delta\}$ está contenido en Γ, por el inciso (b) del lema 2.1 existe una sucesión $v_n \subset \Gamma$ tal que $v_n \xrightarrow[n\to\infty]{} v$ para toda v en $M_1(\Sigma)$. Además, pérdida de generalidad, supongamos que $\Sigma_{v_n} \subset \Sigma_\mu$, y como $v_n \subset \Gamma\cap\mathcal{L}_n$ pa toda n, entonces

$$\begin{aligned}\inf_{v'\in\Gamma\cap\mathcal{L}_n}\{H(v' \mid \mu)\} &\leq H(v_n \mid \mu)\\ \varlimsup_{n\to\infty}\left\{\inf_{v'\in\Gamma\cap\mathcal{L}_n} H(v' \mid \mu)\right\} &\leq \varlimsup_{n\to\infty} H(v_n \mid \mu) \;= -H(v \mid \mu)\\ -\varlimsup_{n\to\infty}\left\{\inf_{v'\in\Gamma\cap\mathcal{L}_n} H(v' \mid \mu)\right\} &\geq -H(v \mid \mu).\end{aligned}$$

Recordemos que $H(v \mid \mu) = \infty$ cuando el soporte de v no está contenido el soporte de μ, es decir, para algún $i \in \{1, 2, ..., |\Sigma|\}, v(a_i) > 0$ mientras c $\mu(a_i) = 0$. Por lo tanto, por la desigualdad anterior tenemos,

$$\begin{aligned}-\varlimsup_{n\to\infty}\left\{\inf_{v\in\Gamma\cap\mathcal{L}_n} H(v \mid \mu)\right\} &\geq \inf_{v\in\Gamma^o}\{-H(v \mid \mu)\}\\ &= -\inf_{v\in\Gamma^o}\{H(v \mid \mu)\}\end{aligned}$$

y aplicando la igualdad 2.1.8 obtenemos la cota inferior del teorema de Sanc

2 Aplicación en Teoría de la Información.

. esta sección no pretendemos dar una exposición completa de las ideas princiles en teoría de la información. Veremos que, como una aplicacón del teorema Sanov obtenemos un resultado muy importante en esta área. Se verá como la ıción de entropía de Shannon surge a partir de la función de tasa de grandes sviaciones. Consideremos el siguiente problema, sea $\Sigma = \{a_1, ..., a_r\}$ un alfato finito con $|\Sigma| = r$. Pensemos en un lenguaje escrito usando este alfabeto tal forma que asignamos probabilidades a cada una de las letras; estas repreıtan la frecuencia relativa de ocurrencia de las letras en textos escritos con el ıguaje. Escribimos p_k para la frecuencia relativa de la letra a_k. Supongamos e no hay correlación entre las letras y entonces la probabilidad de un texto :á el producto de las probabilidades de las letras. Es decir, la sucesión de letras cualquier texto es una sucesión independiente e identicamente distribuidas de ras aleatorias. Sea Ω_n el conjunto de todos los textos que contienen exactaente n letras; entonces la probabilidad del texto $\omega = (\omega_1, \omega_2, ..., \omega_n)$ es $p_1^{n_1}...p_r^{n_r}$, nde cada una de las ω_i representa una letra y n_k es el número de ocurrenıs de la letra a_k en el texto ω. Escribamos $\mathbb{P}(\omega) = p_1^{n_1}...p_r^{n_r}$ para denotar plícitamente la dependencia de esta probabilidad en el texto.

Claude Shannon hizo un descubrimiento notable: Si algunas letras son más obables, es decir, aparecen mas frecuentemente, entonces existe un subconıto Γ_n de Ω_n que consiste de "textos típicos" con las siguientes propiedades:

- Los textos que no están en Γ_n ocurren muy raramente, de manera que $\mathbb{P}(\Gamma_n) = \sum_{\omega \in \Gamma_n} \mathbb{P}(\omega)$ es casi uno.
- Para n grande, Γ_n es mucho más pequeño que Ω_n

$$\frac{\#\Gamma_n}{\#\Omega_n} \approx e^{-n\delta} \qquad \text{para alguna } \delta > 0.$$

- Todos los textos en Γ_n tienen aproximadamente la misma probabilidad.

Este resultado es conocido como la *Propiedad de Equipartición Asintótica* o *EA*. Como sabemos, para un vector de probabilidad p (donde $p_i = \mathbb{P}(Y = a_i)$, para Y v.a. con $i = 1, \ldots, |\Sigma|)$, la entropía de Shannon está definida por

$$H(p) = \sum_{i=i}^{r} p_i \log p_i.$$

Sabemos que $0 \leq H(p) \leq \log r$; la función de entropía alcanza su valor máxi cuando todas las p_k son iguales. El número de elementos en Ω_n es

$$\#\Omega_n = r^n = e^{n \log r},$$

por lo que la cardinalidad crece exponencialmente en n. El número de elemen en Γ_n también crece exponencialmente en n, y la entropía de Shannon nos dá tasa de crecimiento

$$\#\Gamma_n \approx e^{nH(p)},$$

por lo tanto, por la segunda parte del PEA tenemos que

$$\frac{\#\Gamma_n}{\#\Omega_n} \approx e^{-n(H(p)-\log r)},$$

donde la constante δ es la diferencia en las tasas de crecimiento, es decir, δ $\log r - H(p)$.

Si la distribución de probabilidad de p es muy diferente a la distribución probabilidad uniforme, entonces $H(p) < \log r$ y Γ_n es significativamente n pequeño que Ω_n aún para valores de n cercanos a 1.

Nota: Debido a las aplicaciones de teoría de la información a la computac e ingeniería computacional, un alfabeto binario es frecuentemente considerado tal manera que $r = 2$, $a_1 = 0, a_2 = 1$ y la entropía de Shannon es frecuenteme definida usando logaritmo en base 2 en lugar del logaritmo natural; en este ca la entropía máxima es uno.

Para demostrar la PEA, ubicamos el problema en el contexto del teorema Sanov. Sean Y_i v.a.i.i.d. con distribución de probabilidad p para todo i. variable aleatoria toma la i-ésima letra en el texto, es decir, si $\omega = (\omega_1, \ldots, \omega$ entonces $Y_i(\omega) = \omega_i$. Como antes, para cada $a \in A$, sea $1_a(\cdot)$ la función in cadora definida sobre los subconjuntos de A por

$$1_a(B) = \begin{cases} 1 & \text{si } a \in B \\ 0 & \text{e.o.c.} \end{cases}$$

ι

$$L_n^Y(\omega, B) = \frac{1}{n}\sum_{i=1}^{n} 1_{Y_i(\omega)}(B)$$

nde L_n es llamada la distribución empírica ya que $L_n(\omega, \{a_k\}) = n_k/n$, donde es el número de ocurrencias de la letra a_k en el texto ω. Aplicaremos ahora el ›rema de Sanov a $\{L_n^Y\}$, el cual nos dice que existe una función convexa $I(\cdot)$ ›re el espacio de probabilidad $M(\Sigma)$ tal que para todo $v \in M(\Sigma)$ (v representa medida de distribución empírica de cualquier texto $\omega \in \Omega_n$).

$$\mathbb{P}(L_n^Y = v) \approx e^{-nI(v)}$$

ɪl teorema sabemos que $I(v) = \sum_{i=1}^{r} v(a_i)\log\frac{v(a_i)}{p_i}$, es decir $H(v \mid p)$, a la cual mbién se le conoce por *Divergencia Informacional.*

En el teorema de Sanov, la función de tasa es $H(v \mid p)$; ésta se hace cero si y lo si $v = p$, y tenemos que

$$\lim_{n\to\infty} \mathbb{P}(L_n^Y = p) = 1,$$

se sigue que si elegimos Γ_n como aquellos textos en los cuales las frecuencias ativas de las letras son cercanas a las especificadas por p, entonces $\mathbb{P}(\Gamma_n)$ nvergerá a 1 si n crece. Por lo tanto, Γ_n consiste de los textos más probables y r lo tanto, los textos que no pertenecen a Γ_n ocurren raramente. Para estimar tamaño de Γ_n, aplicaremos el teorema de Sanov por segunda ocasión. Sea β distribución de probabilidad uniforme, la cual asigna probabilidad $1/r$ a cada a de las r letras en Σ; entonces la probabilidad $\mathbb{P}_\beta(\Gamma_n) = \#\Gamma_n/\#\Omega_n$. Ahora, es el conjunto de los textos ω para los cuales $L_n^Y(\omega)$ está cercano a p, por lo ıto podemos aplicar el teorema de Sanov a la distribución de L_n^Y con respecto 3, así,

$$\mathbb{P}_\beta(\Gamma_n) = \mathbb{P}_\beta(L_n^Y = p) \approx e^{-nH(p|\beta)}.$$

›nde

$$\begin{aligned} H(p \mid \beta) &= p_1(\log p_1 - \log \tfrac{1}{r}) + \cdots + p_r(\log p_r - \log \tfrac{1}{r}) \\ &= \log r - H(p) \\ &= \delta \end{aligned}$$

r lo tanto, $\frac{\#\Gamma_n}{\#\Omega_n} \approx e^{-n\delta}$. Con lo que demostramos la segunda parte de la PEA.

2.3 El Teorema de Cramér para Alfabetos Finitos en

Mostraremos , como aplicación del teorema de Sanov, una versión del teorema Cramér para Grandes Desviaciones de la media empírica de v.a.i.i.d. Como la sección anterior , las v.a. toman valores en un conjunto finito. Consideren $\bar{S}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$, donde $X_j = f(Y_j)$, $f : \Sigma \to \mathbb{R}$ y $\{Y_j\} \in \Sigma$ son v.a.i. con distribución de probabilidad μ, sin pérdida de generalidad asumiremos c $\Sigma = \Sigma_\mu$ y que $f(a_1) < f(a_2) < \cdots < f(a_{|\Sigma|})$. El teorema de Cramér establ un PGD asociado a la sucesión de v.a. $\bar{S}_n$ que toman valores en un conju finito. En el siguiente capítulo se demostrará el caso en que las v.a. tom valores reales. Por ahora, en el caso que nos encontramos observemos que v.a. $\bar{S}_n$ toman valores en el intervalo compacto $K := [f(a_1), f(a_{(|\Sigma|)})]$, además se puede escribir de la siguiente manera $\bar{S}_n = \sum_{i=1}^{|\Sigma|} f(a_i)L_n^Y(a_i) := \langle f, L_n^Y\rangle$. F lo tanto, para todo conjunto A y todo entero n,

$$\bar{S}_n \in A \Longleftrightarrow L_n^Y \in \{v : \langle f, v\rangle \in A\} := \Gamma. \qquad 2.$$

Por lo anterior, la siguiente versión del teorema de Cramér es una consecuen directa del Teorema de Sanov.

Teorema 2.2. *Para todo conjunto $A \subset \mathbb{R}$,*

$$\begin{aligned} -\inf_{x\in A^o} I(x) &\leq \varliminf_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}_\mu(\bar{S}_n \in A) \\ &\leq \varlimsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}_\mu(\bar{S}_n \in A) \leq -\inf_{x\in A} I(x) \end{aligned} \qquad 2.$$

donde A^o es el interior de A y definimos $I(x) := \inf_{\{v:\langle f,v\rangle = x\}} H(v \mid \mu)$. función de tasa $I(x)$ es continua para $x \in K$ y satisface además que

$$I(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\}, \qquad 2.$$

donde

$$\Lambda(\lambda) = \log \sum_{i=1}^{|\Sigma|} \mu(a_i)e^{\lambda f(a_i)}. \qquad 2.$$

Observación. Como la función de tasa $I(\cdot)$ es continua en el conjunto co pacto K, se sigue de 2.3.2 que si $A \subset \overline{A^o} \subseteq K$ se cumple que

$$\lim \frac{1}{n}\log \mathbb{P}_\mu(\bar{S}_n \in A) = -\inf_{x\in A} I(x).$$

Comentario. La función de tasa $I(x)$ como se define en 2.3.3 también recibe nombre de transformada de Cramér o transformada de Fenchel-Legendre y la ıción $\Lambda(\lambda)$ en 2.3.4 recibe el nombre de función log-Laplace o función generara logarítmica. En el siguiente capítulo se verán propiedades importantes de ;as funciones.

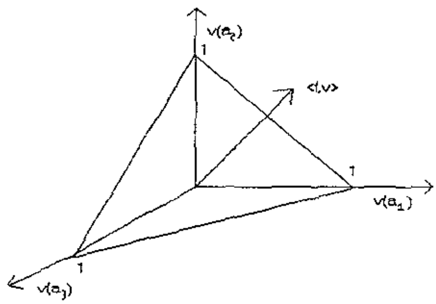

Figura 9. $M_1(\Sigma)$ para $|\Sigma| = 3$ y $< f, v >$.

Demostración.

Cuando el conjunto A es abierto, el conjunto Γ de 2.3.1 también lo es, además, ; cotas del teorema son las mismas que las del teorema de Sanov(2.1) para Γ. ·lo falta mostrar que $I(x)$ satisface la ecuación 2.3.3. Por la desigualdad de nsen tenemos que si X es una v.a., y $U : \mathbb{R} \to \mathbb{R}$ una función cóncava, entonces

$$\mathbb{E}(U(X)) \leq U(\mathbb{E}(X)).$$

Aplicando esta desigualdad a la función concava $\log x$, tenemos que para todo $\in M_1(|\Sigma|)$, y todo $\lambda \in \mathbb{R}$

$$\begin{aligned} \Lambda(\lambda) &= \log \sum_{i=1}^{|\Sigma|} \mu(a_i) e^{\lambda f(a_i)} \\ &= \log \sum_{i=1}^{|\Sigma|} \frac{\mu(a_i) e^{\lambda f(a_i)}}{v(a_i)} v(a_i) \\ &\geq \mathbb{E}_v \left(\log \tfrac{\mu(a_i) e^{\lambda f(a_i)}}{v(a_i)} \right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{|\Sigma|} v(a_i) \log\left(\frac{\mu(a_i)e^{\lambda f(a_i)}}{v(a_i)}\right)\\
&= \sum_{i=1}^{|\Sigma|} v(a_i) \log \mu(a_i) + \sum_{i=1}^{|\Sigma|} v(a_i)\lambda f(a_i) - \sum_{i=1}^{|\Sigma|} v(a_i) \log v(a_i)\\
&= \sum_{i=1}^{|\Sigma|} \lambda f(a_i) v(a_i) - \sum_{i=1}^{|\Sigma|} v(a_i) \log v(a_i) + \sum_{i=1}^{|\Sigma|} v(a_i) \log \mu(a_i)\\
&= \lambda\langle f, v\rangle - H(v \mid \mu)
\end{aligned}
$$

por lo que,

$$\Lambda(\lambda) \geq \lambda\langle f, v\rangle - H(v \mid \mu) \qquad 2.$$

y la igualdad se cumple para $v(a_i) = \mu(a_i)e^{\lambda f(a_i)-\Lambda(\lambda)}$, para verificar esto, sus tuimos el valor de $v(a_i)$ en el miembro derecho de la ecuación anterior, así

$$
\begin{aligned}
\sum_{i=1}^{|\Sigma|} v(a_i) \log \frac{\mu(a_i)e^{\lambda f(a_i)}}{v(a_i)} &= \sum_{i=1}^{|\Sigma|} \mu(a_i)e^{\lambda f(a_i)-\Lambda(\lambda)} \log\left(\frac{\mu(a_i)e^{\lambda f(a_i)}}{\mu(a_i)e^{\lambda f(a_i)-\Lambda(\lambda)}}\right)\\
&= e^{-\Lambda(\lambda)}\Lambda(\lambda) \sum_{i=1}^{|\Sigma|} \mu(a_i)e^{\lambda f(a_i)}\\
&= e^{-\Lambda(\lambda)} \cdot e^{\Lambda(\lambda)} \cdot \Lambda(\lambda)\\
&= \Lambda(\lambda).
\end{aligned}
$$

Como la ecuación 2.3.5 es válida para todo $v \in M_1(\Sigma)$ y todo $\lambda \in \mathbb{R}$, se tien

$$\lambda x - \Lambda(\lambda) \leq \inf_{\{v:\langle f,v\rangle = x\}} H(v \mid \mu) = I(x)$$

por lo tanto,

$$\sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} \leq \inf_{\{v:\langle f,v\rangle = x\}} H(v \mid \mu) = I(x)$$

y la igualdad se da para $x = \langle f, v_\lambda\rangle$.

La función log-Laplace $\Lambda(\lambda)$ es diferenciable (en el siguiente capítulo daren una demostración de este hecho) y su derivada es $\Lambda'(\lambda) = \frac{\mathbb{E}(f(x)e^{\lambda f(x)})}{\mathbb{E}e^{\lambda f(x)}}$. Simpl ıdo la expresión para $\Lambda(\lambda)$, tenemos,

$$\begin{aligned}
\Lambda'(\lambda) &= \frac{\mathbb{E}(f(x)e^{\lambda f(x)})}{\mathbb{E}e^{\lambda f(x)}} \\
&= e^{-\Lambda(\lambda)}\mathbb{E}(f(x)e^{\lambda f(x)}) \\
&= \sum_{i=1}^{|\Sigma|} f(a_i)e^{\lambda f(a_i)}e^{-\Lambda(\lambda)} \\
&= \sum_{i=1}^{|\Sigma|} f(a_i)e^{\lambda f(a_i)-\Lambda(\lambda)} \\
&= \langle f, v_\lambda \rangle.
\end{aligned}$$

ır lo que $\Lambda'(\lambda) = \langle f, v_\lambda \rangle$ y entonces la ecuación 2.3.3 es válida para todas las tales que $x \in \{\Lambda'(\lambda) : \lambda \in \mathbb{R}\}$.

Ya que $\Lambda(\lambda)$ es una función estrictamente convexa, entonces $\Lambda(\lambda)$ es estrictaente creciente, de lo cual, por el supuesto que hicimos al inicio sobre f resulta e $f(a_1) = \inf_\lambda \Lambda'(\lambda)$ y $f(a_|\Sigma|) = \sup_\lambda \Lambda'(\lambda)$. Por lo tanto, 2.3.3 funciona para das las $x \in K^o$. Consideremos ahora el punto inicial de K, $x = f(a_1)$, y sea $(a_1) = 1$ tal que $\langle f, v^* \rangle = x$ por lo que $H(v^* \mid \mu) = -\log \mu(a_1)$, entonces

$$\begin{aligned}
-\log \mu(a_1) &= H(v^* \mid \mu) \\
&\geq I(x) \\
&= \sup_{\lambda \in \mathbb{R}}\{\lambda x - \Lambda(\lambda)\} \\
&= \lim_{\lambda \to \infty}(\lambda x - \Lambda(\lambda) \\
&= -\log \mu(a_1),
\end{aligned}$$

: lo tanto

$$I(x) = \sup_{\lambda \in \mathbb{R}}\{\lambda x - \Lambda(\lambda)\}$$

:umple para $f(a_1)$. Análogamente se prueba para $f(a_{|\Sigma|})$. La continuidad de :) para $x \in K$ es una consecuencia de la continuidad de la entropía relativa $\cdot \mid \mu)$.

2.4 Aplicación en Teoría del Riesgo.

En la sección anterior vimos una versión del teorema de Cramér para la me empírica de una sucesión de v.a.i.i.d que tomaban valores en un alfabeto fini En el siguiente capítulo veremos el teorema de Cramér en el caso que las va ables aleatorias toman valores reales. Supongamos por ahora, para fines de e sección que el teorema en su versión real es válido.

Como mencionamos antes, la teoría de grandes desviaciones se ha aplicad distintos modelos en teoría del riesgo que pueden llegar a ser muy complejos. continuación daremos un ejemplo de un modelo sencillo de como es hecho es

Supongamos que una compañia aseguradora recibe un número fijo de reclan ciones (una reclamación es un pago que debe hacer la aseguradora al beneficia en caso de siniestro) en un periodo de tiempo dado, y supongamos también q recibe un ingreso producto del pago de primas, sea esta una cantidad k diar Como el importe de las reclamaciones es una variable aleatoria, existe el ries de que, al término de un período de tamaño T, la cantidad pagada a los asegu dos exceda los ingresos de la compañia aseguradora en dicho periodo de tiem Este riesgo es inevitable, pero a la compañia le gustaría tener la certeza de q dicho evento ocurra muy raramente. Es decir, estamos interesados en las prob bilidades pequeñas relacionadas con la suma de un número grande de variab aleatorias. Si las cantidades X_T de las reclamaciones son v.a.i.i.d., utilizaren el Teorema de Cramér para calcular la probabilidad de que la cantidad $\sum_{t=1}^{T}$ pagada durante el periodo T exceda los ingresos kT recibidos en dicho perio esto es

$$\mathbb{P}\left(\sum_{t=1}^{T} X_t > kt\right) \approx e^{-TI(k)}.$$

Por lo que, si queremos asegurarnos que el riesgo de ruina sea pequeño, digan de tamaño e^{-r}, para $r > 0$ suficientemente grande, podemos usar la función tasa $I(k)$ para hallar un valor apropiado de k.

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} X_t > k\right) &\approx e^{-r} \\
e^{-TI(k)} &\approx e^{-r} \\
I(k) &\approx \tfrac{r}{T}.
\end{aligned}$$

Dado que $I(x)$ es convexa, es monótonamente creciente para x mayor que la edia de X_t, y por lo tanto la ecuación

$$I(k) = \frac{r}{T}$$

ene solución única.

Por supuesto, para resolver ésta ecuación debemos conocer qué es $I(x)$, es ecir, conocer las estadísticas de las cantidades de las reclamaciones. Veamos os ejemplos:

- En primer lugar , supongamos que el tamaño de cada reclamación se distribuye binomial con media np. Calculamos la función de tasa como se indica en 2.3.3 y tenemos que

$$I(x) = x \log\left(\frac{x}{p}\right) + (n-x)\log\left(\frac{n-x}{1-p}\right) - n\log n.$$

Si fijamos el valor de np, r y T , tenemos que resolver la ecuación

$$k \log\left(\frac{k}{p}\right) + (n-k)\log\left(\frac{n-k}{1-p}\right) - n\log n = \frac{r}{T}.$$

La solución de la ecuación para k en este caso se obtiene numericamente.

- Ahora, supongamos que el tamaño de cada reclamación tiene distribución normal con media m y varianza σ^2. En este caso utilizamos el teorema de Cramér Real y por la ecuación 3.1.1, que veremos más adelante, obtenemos la función de tasa

$$I(x) = \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2.$$

Notese que con $m = 0$ y $\sigma^2 = 1$ se tiene la función de tasa que se obtuvo en 1.2.1

La solución de la ecuación para k es en éste caso: $k = m + \sigma\sqrt{2r/T}$; por lo que el valor de la prima debería ser tal que el ingreso diario sea la media de las reclamaciones más una cantidad adicional para cubrir el riesgo; en éste caso, está dada por $\sigma\sqrt{2r/T}$. Observemos que σ es una medida de las fluctuaciones en la cantidad de las reclamaciones, mientras que $\sqrt{2r/T}$ es fijado por la aseguradora.

2.5 Grandes Desviaciones para Muestreo sin Reemplaz

Usaremos el método de Tipos para un procedimiento de uso común en muc problemas de estadística, el muestreo sin reemplazo. Consideremos una urna c un número finito m de distintos objetos, definimos el vector $\mathrm{y} = (y_1, \ldots, y_m)$, el cual cada entrada representa un elemento de la urna, tomemos una mues sin reemplazo de tamaño n a la que representaremos como una n-ada Y $(y_{i_1}, \ldots, y_{i_n})$, los índices $\{i_1, i_2, \ldots, i_n\}$ se escogen aleatoriamente, de tal for que cada subconjunto de n elementos distintos de $\{1, 2, \ldots, m\}$ es igualme probable.

Supongamos que para todo m, cada elemento del vector $(y_1^{(m)}, y_2^{(m)}, \ldots, y_m^{(}$ pertenece al alfabeto finito $\Sigma = \{a_1, a_2, \ldots, a_{|\Sigma|}\}$, además suponemos que número de elementos en la urna depende del tamaño de la muestra, es de $m = m(n)$, y si $n \to \infty$, los vectores deterministas de frecuencia relativa L_m^{y} $(L_m^{\mathrm{y}}(a_1), \ldots, L_m^{\mathrm{y}}(a_{|\Sigma|}))$ convergen a una medida de probabilidad $\mu \in M_1($ recordemos que

$$L_m^{\mathrm{y}}(a_i) = \frac{1}{m} \sum_{j=1}^{m} 1_{a_i}(y_j^{(m)}), \qquad i = 1, 2, \ldots, |\Sigma|.$$

Supongamos también que Y es un vector aleatorio obtenido del muestreo reemplazo al escoger n de m posibles elementos de la urna. Estableceren el PGD para la media empírica aleatoria L_n^{Y} y veremos el análogo al Teore de Sanov con $m = m(n)$ y $\lim_{n\to\infty} \frac{n}{m(n)} = \beta$, para $0 < \beta < 1$. Para es consideremos la siguiente función de tasa

$$I(v \mid \beta, \mu) := \begin{cases} H(v \mid \mu) + \frac{1-\beta}{\beta} H\left(\frac{\mu - \beta v}{1-\beta} \mid \mu\right) & \text{si} \quad \mu(a_i) \geq \beta v(a_i) \text{ para toda } \\ \infty & \text{en otro caso.} \end{cases}$$

2.

Veamos que cuando $\beta \to 0$, la función $I(\cdot \mid \beta, \mu)$ se simplifica a $H(\cdot \mid \mu)$, para cual desarrollamos $\frac{1-\beta}{\beta} H\left(\frac{\mu-\beta v}{1-\beta} \mid \mu\right)$ y tomando el límite cuando $\beta \to 0$ se tie que

$$\lim_{\beta\to 0} \frac{1-\beta}{\beta} H\left(\frac{\mu - \beta v}{1-\beta} \mid \mu\right) = \lim_{\beta\to 0} \sum_{i=1}^{|\Sigma|} \left(\mu(a_i) - \beta v(a_i)\right) \cdot \frac{\log \frac{\mu(a_i) - \beta v(a_i)}{(1-\beta)\mu(a_i)}}{\beta} = 0$$

ara lo cual se aplicó regla de L'Hopital al segundo factor de la parte derecha. or otra parte, cuando $\beta \to 1$, el dominio de v para el cual $I(v \mid \beta, \mu) < \infty$ se stringe a una única medida de probabilidad $v = \mu$ porque no puede suceder ue $\mu(a_i) > v(a_i)$ para toda i. Intuitivamente vemos que si β se incrementa, el rado de aleatoriedad de la muestra disminuye.

Recordemos del lema 2.1.1 que L_n^Y pertenece al conjunto $\{\mathcal{L}_n\}$ cuyo tamaño rece de manera polinomial en n. Las siguientes estimaciones de las probabililades de grandes desviaciones para $\mathcal{L}_n^Y$ se obtienen a través de los métodos ombinatorios que hemos estado empleando.

ema 2.5.1. *Para todo vector de probabilidad $v \in \mathcal{L}_n$*

(a) Si $I(v \mid \frac{n}{m}, L_m^y) < \infty$, entonces

$$\left|\frac{1}{n}\log \mathbb{P}(L_n^Y = v) + I(v \mid \beta, \mu)\right| \leq 2(|\Sigma| + 1)\left(\frac{\log(m+1)}{n}\right) \qquad 2.5.2$$

(b) Si $I(v \mid \beta, \mu) = \infty$, entonces $\mathbb{P}(L_n^Y = v) = 0$.

Demostración.

En primer lugar demostraremos el inciso (b) que es inmediato.

Notemos que $I(v \mid \frac{n}{m}, L_m^y) = \infty$ si y sólo si se cumple que $nv(a_i) > mL_m^y(a_i)$ ara algún $a_i \in \Sigma$. Esto no es posible en el caso de muestreo sin reemplazo, ya ue para $v \in \mathcal{L}_n$, se tiene que $L_n^Y = v$ y $nL_n^Y(a_i) \leq mL_m^y(a_i)$ para toda $a_i \in \Sigma$.

Ahora demostraremos el inciso (a).

En el caso de muestreo sin reemplazo, la probabilidad del evento $\{L_n^Y = v\}$ ara $v \in \mathcal{L}_n$ es exactamente el número de n-adas $i_1 \neq i_2 \neq \cdots \neq i_n$ resultantes en el tipo v, comparadas con el número total de n-adas, es decir

$$\begin{aligned}
\mathbb{P}(L_n^{\mathrm{Y}} = v) &= \mathbb{P}\big((L_n^{\mathrm{Y}}(a_1), \ldots, L_n^{\mathrm{Y}}(a_{|\Sigma|})) = (v(a_1), \ldots, v(a_{|\Sigma|}))\big) \\
&= \mathbb{P}\big((L_n^{\mathrm{Y}}(a_1) = v(a_1), \ldots, L_n^{\mathrm{Y}}(a_{|\Sigma|}) = v(a_{|\Sigma|}))\big) \\
&= \prod_{i=1}^{|\Sigma|} \mathbb{P}\big(L_n^{\mathrm{Y}}(a_i) = v(a_i)\big) \\
&= \prod_{i=1}^{|\Sigma|} \mathbb{P}\big(\sum_{j=1}^{n} 1_{a_i}(Y_j) = nv(a_i)\big) \qquad 2.\\
&= \frac{\prod_{i=1}^{|\Sigma|} \binom{mL_m^y(a_i)}{nv(a_i)}}{\binom{m}{n}}.
\end{aligned}$$

Por otra parte, sustituyendo $|\Sigma| = 2$, $v(a_1) = k/n$, y $v(a_2) = 1 - k/n$, en ecuación 2.1.3, se tiene que $|T_n(v)| = \binom{n}{k}$ y por el lema 2.1.3 se tiene,

$$\max_{0 \le k \le n} \left| \log \binom{n}{k} - nH\left(\frac{k}{n}\right) \right| \le 2\log(n+1) \qquad 2.$$

donde definimos $H(p) = -p\log p - (1-p)\log(1-p)$. Probaremos ahora afirmación anterior. Sustituyendo $|T_n(v)| = \binom{n}{k}$ en el lema 2.1.3 tenemos,

$$\begin{array}{llll}
 & (n+1)^{-2}e^{nH(v)} & \le \log\binom{n}{k} & \le e^{nH(v)} \\
\iff & -2\log(n+1) & \le \log\binom{n}{k} - nH(v) & \le 0 \\
\iff & -2\log(n+1) & \le \log\binom{n}{k} - nH\left(\frac{k}{n}\right) & \le 2\log(n+1) \\
\iff & & \left|\log\binom{n}{k} - nH\left(\frac{k}{n}\right)\right| & \le 2\log(n+1)
\end{array}$$

y como se cumple para toda $0 \le k \le n$, se tiene el resultado.

Las ecuaciones anteriores nos serán de utilidad para acotar la siguiente expresión.

$$\left| \frac{1}{n}\log \mathbb{P}(L_n^{Y} = v) - \sum_{i=1}^{|\Sigma|} \frac{mL_m^{\mathrm{Y}}(a_i)}{n} H\left(\frac{nv(a_i)}{mL_m^{\mathrm{Y}}(a_i)}\right) + \frac{m}{n}H\left(\frac{n}{m}\right) \right|$$

or 2.5.3 y 2.5.4 tenemos que la ecuación anterior es igual a

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{|\Sigma|}\log\binom{mL_m^y(a_i)}{nv(a_i)} - \sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)}{n}H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right)\right. \\
&\qquad \left.+\frac{m}{n}H\left(\frac{n}{m}\right) - \frac{1}{n}\log\binom{m}{n}\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{|\Sigma|}\left[\log\binom{mL_m^y(a_i)}{nv(a_i)} - mL_m^y(a_i)H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right)\right]\right. \\
&\qquad \left.-\left[\log\binom{m}{n} - mH\left(\frac{n}{m}\right)\right]\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{|\Sigma|}\left|\log\binom{mL_m^y(a_i)}{nv(a_i)} - mL_m^y(a_i)H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right)\right| \\
&\qquad +\left|\log\binom{m}{n} - mH\left(\frac{n}{m}\right)\right| \\
&\leq \frac{1}{n}|\Sigma|\max_{0\leq nv(a_i)\leq mL_m^y y a_i}\left|\log\binom{mL_m^y(a_i)}{nv(a_i)} - mL_m^y(a_i)H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right)\right| \\
&\qquad +\max_{0\leq n\leq m}\left|\log\binom{m}{n} - mH\left(\frac{n}{m}\right)\right| \\
&\leq \frac{2}{n}[|\Sigma|\log(mL_m^y(a_i)+1) + \log(m+1)] \\
&\leq \frac{2}{n}[|\Sigma|\log(m+1) + \log(m+1)] \\
&= \frac{2}{n}(|\Sigma|+1)\log(m+1).
\end{aligned}
$$

Por lo tanto, hemos demostrado que

$$\left|\frac{1}{n}\log \mathbb{P}(L_n^Y = v) - \sum_{i=1}^{|\Sigma|}\frac{mL_m^Y(a_i)}{n}H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right) + \frac{m}{n}H\left(\frac{n}{m}\right)\right|$$

$$\leq 2(N+1)\left(\frac{\log(m+1)}{n}\right) \qquad 2.$$

Sólo falta mostrar que en el lado izquierdo de la desigualdad anterior es cie que $I(v \mid \frac{n}{m}, L_m^y) = -\sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)}{n}H\left(\frac{nv(a_i)}{mL_m^y(a_i)}\right) + \frac{m}{n}H\left(\frac{n}{m}\right)$. Para ver esto usar que $H(p) = -p\log p - (1-p)\log(1-p)$ y desarrollando la parte derecha de igualdad anterior tenemos:

$$-\sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)}{n}\left[\frac{-nv(a_i)}{mL_m^y(a_i)}\log\frac{nv(a_i)}{mL_m^y(a_i)} - \left(1-\frac{nv(a_i)}{mL_m^y(a_i)}\right)\log\left(1-\frac{nv(a_i)}{mL_m^y(a_i}\right.\right.$$
$$+\frac{m}{n}\left[\frac{n}{m}\log\frac{n}{m} - (1-\frac{n}{m})\log\frac{m-n}{m}\right]$$
$$= \sum_{i=1}^{|\Sigma|}v(a_i)\log\frac{nv(a_i)}{mL_m^y(a_i)} + \sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)-nv(a_i)}{n}\log\frac{mL_m^y(a_i)-nv(a_i)}{mL_m^y(a_i)}$$
$$-\log\frac{n}{m} - \frac{m-n}{n}\log\frac{m-n}{m}$$
$$= \sum_{i=1}^{|\Sigma|}v(a_i)(\log nv(a_i) - \log mL_m^y(a_i))$$
$$+\sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)-nv(a_i)}{n}(\log mL_m^y(a_i) - nv(a_i) - \log mL_m^y(a_i))$$
$$-\log\frac{n}{m} - \frac{m-n}{n}\log\frac{m-n}{m}$$
$$= \sum_{i=1}^{|\Sigma|}v(a_i)\log\frac{v(a_i)}{L_m^y(a_i)} + \sum_{i=1}^{|\Sigma|}v(a_i)\log\frac{n}{m} + \sum_{i=1}^{|\Sigma|}\frac{mL_m^y(a_i)-nv(a_i)}{n}\big[\log(mL_m^y($$
$$-nv(a_i)) - \log m - \log L_m^y(a_i) - \log(m-n) + \log(m-n)\big]$$
$$-\log\frac{n}{m} - \frac{m-n}{n}\log\frac{m-n}{m}$$

$$
\begin{aligned}
&= H(v \mid L^{y}_{m}) + \sum_{i=1}^{|\Sigma|} \frac{mL^{y}_{m}(a_i) - nv(a_i)}{n} \log \frac{mL^{y}_{m}(a_i) - nv(a_i)}{(m-n)L^{y}_{m}(a_i)} \\
&\quad + \sum_{i=1}^{|\Sigma|} \frac{mL^{y}_{m}(a_i) - nv(a_i)}{n} \log \frac{m-n}{m} - \frac{m-n}{n} \log \frac{m-n}{m} \\
&= H(v \mid L^{y}_{m}) + \frac{m-n}{n} \sum_{i=1}^{|\Sigma|} \frac{mL^{y}_{m}(a_i) - nv(a_i)}{m-n} \log \frac{mL^{y}_{m}(a_i) - nv(a_i)}{(m-n)L^{y}_{m}(a_i)} \\
&\quad + \left(\frac{\sum_{i=1}^{|\Sigma|} mL^{y}_{m}(a_i) - nv(a_i)}{n} - \frac{m-n}{n} \right) \log \frac{m-n}{m} \\
&= H(v \mid L^{y}_{m}) + \tfrac{m-n}{n} H\left(\tfrac{\mu - \frac{n}{m} v}{1 - \frac{n}{m}} \mid L^{y}_{m} \right) = I(v \mid \tfrac{n}{m}, L^{y}_{m}).
\end{aligned}
$$

esta manera, queda demostrado el lema 2.5.1. Como en la demostración del rema de Sanov, las siguientes estimaciones son las análogas de 2.1.7 y 2.1.8.

ma 2.5.2. *Con $m = m(n)$ y para todo conjunto de vectores de probabilidad* $\in M_1(\sum)$.

$$
\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}(L^{Y}_{n} \in \Gamma) = - \underline{\lim_{n\to\infty}} \left\{ \inf_{v \in \Gamma \cap \mathcal{L}_n} I\left(v \mid \frac{n}{m}, L^{y}_{m} \right) \right\} \tag{2.5.6}
$$

$$
\underline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}(L^{Y}_{n} \in \Gamma) = - \overline{\lim_{n\to\infty}} \left\{ \inf_{v \in \Gamma \cap \mathcal{L}_n} I\left(v \mid \frac{n}{m}, L^{y}_{m} \right) \right\} \tag{2.5.7}
$$

Demostración.

Sea $v \in \mathcal{L}_n$, por el lema 2.3.1 tenemos que,

$$
\begin{aligned}
\mathbb{P}(L^{Y}_{n} \in \Gamma) &= \sum_{v \in \Gamma \cap \mathcal{L}_n} \mathbb{P}(L^{Y}_{n} = v) \\
&\leq \sum_{v \in \Gamma \cap \mathcal{L}_n} \exp[2n(N+1)\frac{\log(m+1)}{n} - nI(v \mid \frac{n}{m}, L^{y}_{m})] \\
&\leq |\Gamma \cap \mathcal{L}_n| \exp[n(2(|\Sigma|+1)\frac{\log(m+1)}{n} - \inf_{v \in \Gamma \cap \mathcal{L}_n} I(v \mid \frac{n}{m}, L^{y}_{m})]
\end{aligned}
$$

aando logaritmos de ambos lados y dividiendo entre n tenemos que

$$
\text{og}\, \mathbb{P}(L^{Y}_{n} \in \Gamma) \leq \frac{1}{n} \log(n+1)^{|\Sigma|} + 2(|\Sigma|+1)\frac{\log(m+1)}{n} - \inf_{v \in \Gamma \cap \mathcal{L}_n} I(v \mid \frac{n}{m}, L^{y}_{m})
$$

Por otra parte :

$$\begin{aligned}
\mathbb{P}(L_n^{\mathrm{Y}} \in \Gamma) &= \sum_{v \in \Gamma \cap \mathcal{L}_n} \mathbb{P}(L_n^{\mathrm{Y}} = v) \\
&\geq \sum_{v \in \Gamma \cap \mathcal{L}_n} \exp[-n(2(|\Sigma|+1)\frac{\log(m+1)}{n} - I(v \mid \frac{n}{m}, L_m^{\mathrm{y}})] \\
&\geq \exp[-n(2(|\Sigma|+1)\frac{\log(m+1)}{n} - \inf_{v \in \Gamma \cap \mathcal{L}_n} I(v \mid \frac{n}{m}, L_m^{\mathrm{y}})]
\end{aligned}$$

tomando logaritmos de ambos lados y dividiendo entre n tenemos que

$$\frac{1}{n} \log \mathbb{P}(\mathrm{L}_n^{\mathrm{Y}} \in \Gamma) \geq \frac{1}{n} \log(n+1)^{-|\Sigma|} - 2(|\Sigma|+1) \log \frac{m+1}{n} - \inf_{v \in \Gamma \cap \mathcal{L}_n} I(v \mid \frac{n}{m}, L$$

y finalmente, tomamos limite superior en ambas desigualdades, y usando q $\overline{\lim}_{n\to\infty}(-A) = -\underline{\lim}_{n\to\infty}(A)$ para cualquier conjunto A, tenemos demostrado 2.5 Análogamente se prueba la igualdad en el caso del límite inferior.

El siguiente lema nos será de utilidad para la demostración del resultado m importante en ésta sección.

Lema 2.5.3. *Sea $\beta_n \in (0,1), \mu_n, v_n \in M_1(\Sigma)$ tales que $\beta_n \to \beta \in (0,1)$ $\mu_n \to \mu$ conforme $n \to \infty$.*

(a) Si $v_n \to v$ e $I(v_n \mid \beta_n, \mu_n) < \infty$ para toda n suficientemente grande, e tonces $\lim_{n\to\infty} I(v_n \mid \beta_n, \mu_n) = I(v \mid \beta, \mu)$.

(b) Si $I(v \mid \beta, \mu) < \infty$, entonces existe una sucesión $v_n \in \mathcal{L}_n$ tal que $v_n \to v$ $\lim_{n\to\infty} I(v_n \mid \beta_n, \mu_n) = I(v \mid \beta, \mu)$.

Demostración.

(a) Se puede verificar de la misma manera que se hizo en el lema anterior q

$$I(v_n \mid \beta_n, \mu_n) = \frac{1}{\beta_n} H(\mu_n) - H(v_n) - \frac{1-\beta_n}{\beta_n} H\left(\frac{\mu_n - \beta_n v_n}{1-\beta_n}\right).$$

Como la función $H(\cdot \mid \beta, \mu)$ es continua, tomando el límite cuando $n \to$ y como $\beta \in (0,1)$, se concluye la demostración.

) Considere primero $v \in M_1(\Sigma)$ para el cual

$$\min_{a_i \in \Sigma}\{\mu(a_i) - \beta v(a_i)\} > 0. \tag{2.5.8}$$

Ahora, por el inciso (b) del lema 2.1.2. existe $v \in \mathcal{L}_n$ tal que $v_n \to v$. Entonces, la desigualdad estricta anterior implica que para toda n grande

$$\min_{a_i \in \Sigma}\{\mu_n(a_i) - \beta_n v_n(a_i)\} \geq 0.$$

Por lo tanto, $I(v_n \mid \beta_n, \mu_n) \leq \infty$, para toda n suficientemente grande, y por el inciso anterior se obtiene que

$$\lim_{n\to\infty} I(v_n \mid \beta_n, \mu_n) = I(v \mid \beta, \mu) < \infty.$$

Supongamos ahora que $\Sigma_\mu = \Sigma$, pero posiblemente no se satisface 2.5.8. Como $\beta < 1$ e $I(v \mid \beta, \mu) \leq \infty$, existe algún $a_i \in \Sigma$ tal que $\mu(a_i) - \beta v(a_i) = 0$, entonces existe $v_k \to v$ tal que $\min_{a_i \in \Sigma}\{\mu(a_i) - \beta v_k(a_i)\} > 0$ para toda k. Ahora bien, para cada k, usando el argumento anterior, existe una sucesión $\{v_{n,k}\}_{n\geq 1}$ tal que $v_{n,k} \in \mathcal{L}_n, v_{n,k} \to v_k$ y se cumple que

$$\lim_{n\to\infty} I(v_{n,k} \mid \beta_n, \mu_n) = I(v_k \mid \beta, \mu).$$

Por el argumento de la diagonalización de Cantor, existe una sucesión v_n tal que

$$\lim_{n\to\infty} v_n = \lim_{k\to\infty} v_k,$$

por lo tanto,

$$\lim_{n\to\infty} I(v_n \mid \beta, \mu) = \lim_{k\to\infty} I(v_k \mid \beta, \mu)$$

Finalmente, si $\Sigma_\mu \neq \Sigma$, con $\Sigma_v \subset \Sigma_\mu$. Como $I(v \mid \beta, \mu) < \infty$, se repite el procedimiento con Σ_μ en lugar de Σ, por lo tanto, existe $v_n \in \mathcal{L}_n$ tal que $\Sigma_{v_n} \subset \Sigma_\mu, v_n \to v$ e $I(v_n \mid \beta_n, \mu_n)$ para n suficientemente grande y aplicando (a) se tiene el resultado.

El siguiente teorema es el análogo al Teorema de Sanov.

Teorema 2.3. *Supóngase que* L_m^y *converge a* μ *y* $\frac{n}{m} \to \beta \in (0,1)$ *confor* $n \to \infty$. *Entonces las medidas empíricas aleatorias* L_n^Y *satisfacen el PGD,* *función de tasa buena* $I(v \mid \beta, \mu)$. *Explícitamente, para todo conjunto* Γ *vectores de probabilidad en* $M_1(\Sigma) \subset \mathbb{R}^{|\Sigma|}$.

$$\begin{aligned} -\inf_{v\in\Gamma^o} &\leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}(L_n^Y \in \Gamma) \\ &\leq \overline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}(L_n^Y \in \Gamma) \\ &\leq -\inf_{v\in\overline{\Gamma}} I(v \mid \beta, \mu) \end{aligned} \qquad 2.$$

hacemos notar que la cota superior es más débil que en el Teorema de Sanov, *el sentido de que el ínfimo sobre la función de tasa es tomado sobre la cerrad* *de* Γ.

Demostración.

Primero veamos que la función $I(\cdot \mid \beta, \mu)$ es una función de tasa buena. P el inciso (a) del lema anterior tenemos que la función $I(\cdot \mid \beta, \mu)$ es contin conjuntamente en $\beta \in (0,1)$, μ y v por lo que es semicontinua inferiormen Además es una función no-negativa ya que $H(\cdot \mid \mu)$ lo es y como el conjunto todos los vectores Σ-dimensionales con entradas reales no negativas que sum 1 es un conjunto compacto, entonces $I(\cdot \mid \beta, \mu)$ es una función de tasa buena

Demostremos ahora la cota superior. Primero, se deduce de 2.5.6 que pa alguna subsucesión infinita n_k, existe una sucesión $\{v_k\} \subset \Gamma$ tal que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}(L_n^Y \in \Gamma) = -\lim_{k\to\infty} I(v_k \mid \frac{n_k}{m_k}, L_{m_k}^y) := -I^* \qquad 2.5$$

donde posiblemente $I^* = -\infty$. La sucesión $\{v_k\}$ tiene un punto limite v^* en conjunto compacto $\overline{\Gamma}$. Pasando a la subsucesión convergente, la semicontinuid inferior de I conjuntamente con β, μ y v implica que

$$I^* \geq I(v^* \mid \beta, \mu) \geq \inf_{v\in\overline{\Gamma}} I(v \mid \beta, \mu).$$

La cota superior se sigue por 2.5.10. Finalmente, para probar la cota inferi considere v arbitrario en Γ^o tal que $I(v \mid \beta, \mu) < \infty$. Entonces, por la parte del lema 2.5.3 existe $v_n \in \mathcal{L}_n$ tal que $v_n \to v$ y

$$\lim_{n\to\infty} I(v_n \mid \frac{n}{m}, L_m^y) = I(v \mid \beta, \mu)$$

lemás, para n suficientemente grande, $v_n \in \Gamma \cap \mathcal{L}_n$ y se sigue que

$$-\overline{\lim_{n\to\infty}}\{\inf_{v'\in\Gamma\cap\mathcal{L}_n} I(v' \mid \frac{n}{m}, L_m^y)\} \geq -I(v \mid \beta, \mu)$$

ombinando la desigualdad anterior con 2.5.7, concluímos que para cada v.

$$\underline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}(L_n^Y \in \Gamma) \geq -I(v \mid \beta, \mu)$$

la cota se sigue por 1.4.5.

;apítulo 3

;l Teorema de Cramér Real.

ı la sección 2.3 como una aplicación del metodo de tipos, se presentó el teorema ; Cramér para grandes desviaciones asociadas a la distribución empírica de a.i.i.d. que toman valores en un conjunto finito. En este capítulo, el teorema 2 se extenderá al caso de v.a.i.i.d que toman valores en $\mathbb{R}$.

.1 Transformada de Cramér.

l teorema de Cramér afirma que la sucesión de medidas de probabilidad μ_n .tisface el PGD con la función de tasa convexa $\Lambda^*(\cdot)$, la cuál recibe el nom-e de Transformada de Cramér.(También es conocida como transformada de ənchel-Legendre). Antes de enunciar el teorema , demostraremos propiedades ıportantes de la transformada de Cramér.

Considerese $(\Omega, \mathbb{F}, \mathbb{P})$ un espacio de probabilidad. Sea $X : \Omega \to \mathbb{R}$ una variable eatoria, y μ la probabilidad sobre $\mathbb{R}$, inducida por X de la siguiente manera: $A) = \mathbb{P}(X^{-1}(A) \in \mathbb{F}) = \mathbb{P}(X \in A)$ con $A \in \mathbb{R}$.

efinición 3.1.1. La función log-Laplace de μ es la función definida por

$$\begin{aligned} \Lambda :\ & \mathbb{R} \to (-\infty, +\infty] \\ & \lambda \mapsto \Lambda(\lambda) = \log[\mathbb{E}(e^{\lambda X})]. \end{aligned}$$

Notemos que la función Λ nunca toma el valor $-\infty$, de ser así, tenemos que valor esperado $\mathbb{E}(e^{\lambda X}) = 0$, lo cual contradice el hecho de que X toma valores ales.

Tenemos las siguientes proposiciones para la función log-Laplace.

Proposición 3.1.1. *Sea $\mathcal{D}(\Lambda) = \{\lambda : \Lambda(\lambda) < \infty\}$. Entonces tenemos:*

1. *Λ es convexa, s.c.i. y $\Lambda(0) = 0$.*

2. *Si λ está en el interior de $\mathcal{D}(\Lambda)$, entonces Λ es derivable con respecto y además,*
$$\Lambda'(\lambda) = \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}.$$

3. *Si $\mu(0, +\infty) > 0$ (resp. si $\mu(-\infty, 0) > 0$), entonces $\lim_{\lambda\to+\infty} \Lambda(\lambda) =$ - (resp. $\lim_{\lambda\to-\infty} \Lambda(\lambda) = +\infty$).*

4. *Si $\mu(-\infty, 0) = 0$ (resp. si $\mu(0, +\infty) = 0$), entonces Λ es creciente (resp. creciente) y $\lim_{\lambda\to-\infty} \Lambda(\lambda) = \log \mu(\{0\})$ (resp. $\lim_{\lambda\to+\infty} \Lambda(\lambda) = \log \mu(\{($*

Demostración.

1. Para mostrar la convexidad vamos a utilizar la desigualdad de Hölder $p = \frac{1}{t}$ y $q = \frac{1}{1-t}$. Si $\lambda_1, \lambda_2 \in \mathcal{D}(\lambda)$ y si $t \in [0, 1]$.
$$\begin{aligned}
\Lambda(t\lambda_1 + (1-t)\lambda_2) &= \log[\mathbb{E}(e^{(t\lambda_1+(1-t)\lambda_2)X})] \\
&= \log[\mathbb{E}(e^{(t\lambda_1)X}\, e^{((1-t)\lambda_2)X})] \\
&= \log[\mathbb{E}((e^{\lambda_1 X})^t \cdot (e^{\lambda_2 X})^{1-t})] \\
&= \log[\mathbb{E}((e^{\lambda_1 X})^{\frac{1}{1/t}} \cdot (e^{\lambda_2 X})^{\frac{1}{1/(1-t)}})] \\
&\leq \log([\mathbb{E}(e^{\lambda_1 X})]^t \cdot [\mathbb{E}(e^{\lambda_2 X})^{1-t}]) \qquad \text{(por Hölder)} \\
&= \log[(\mathbb{E}(e^{\lambda_1 X}))^t] + \log[(\mathbb{E}(e^{\lambda_2 X}))^{1-t}] \\
&= t\log[\mathbb{E}(e^{\lambda_1 X})] + (1-t)\log[\mathbb{E}(e^{\lambda_2 X})].
\end{aligned}$$
por lo tanto, $$\Lambda(t\lambda_1 + (1-t)\lambda_2) \leq t\Lambda(\lambda_1) + (1-t)\Lambda(\lambda_2).$$
Para mostrar que Λ es s.c.i, tomamos $\lambda_n \to \lambda$.
Por el lema de Fatou sabemos que si (f_n) es una sucesión en $\mu^+(x,$ conjunto de funciones medibles positivas, y λ una medida sobre $\mathbb{R}$ entor
$$\varliminf_{n\to\infty} \int f_n d\lambda \geq \int \varliminf_{n\to\infty} f_n\, d\lambda,$$

y usando este lema tenemos que

$$\begin{aligned}
\varliminf_{n\to\infty} \mathbb{E}[e^{\lambda_n X}] &= \varliminf_{n\to\infty} \int_{\mathbb{R}} e^{\lambda_n x}\mu(dx) \\
&\geq \int_{\mathbb{R}} \varliminf_{n\to\infty} e^{\lambda_n x}\mu(dx) \\
&= \int_{\mathbb{R}} e^{\lambda x}\mu(dx) \\
&= \mathbb{E}[e^{\lambda X}],
\end{aligned}$$

después tomamos el logaritmo que es continuo y creciente, y entonces

$$\Lambda(\lambda) \leq \varliminf_{n\to\infty} \Lambda(\lambda_n),$$

por lo que Λ es semicontinua inferiormente.

2. Sea λ_0 en el interior de $\mathcal{D}(\Lambda)$. Vamos a aplicar el teorema de la derivación bajo la integral, para esto, primero verificaremos que $\int_{\mathbb{R}} e^{\lambda X}\mu(dx) < +\infty$, es decir, integrable. Con ω constante, la función

$$\lambda \mapsto e^{\lambda X(\omega)}$$

es derivable con respecto a λ de derivada $X(\omega)e^{\lambda X(\omega)}$. Como tomamos λ en el interior de $\mathcal{D}(\Lambda)$, existe un $\alpha > 0$ tal que $(\lambda_0 - 2\alpha, \lambda_0 + 2\alpha) \subset \mathcal{D}(\Lambda)$. Por otra parte, existe una constante C_α que sólo depende de α tal que,

$$\text{para todo}\, x \in \mathbb{R}, \qquad |x| \leq C_\alpha e^{\alpha|x|}$$

para demostrar lo anterior, hacemos $f(x) = C_\alpha e^{\alpha x} - x$ y tenemos que ver que $f(x)$ es mayor o igual a cero. Más aún, queremos la desigualdad con valor absoluto, entonces vamos a partir el dominio de las x en $(-\infty, 0]$ y $[0, +\infty)$. Sea $x \in [0, +\infty)$, tomamos la derivada de $f(x)$, y vemos que $f'(x) = \alpha C_\alpha e^{\alpha x} - 1$, para que ésta también sea mayor que cero, se tiene que cumplir que

$$\alpha C_\alpha \geq 1$$

ya que $e^{\alpha x}$ siempre es mayor que 1, y las funciones son estrictamente crecientes, de aquí resulta que debemos escoger $C_\alpha \geq 1/\alpha$, lo cual siempre puede hacerse. Con lo cual se prueba la desigualdad. Por lo que,

$$\begin{aligned} Xe^{\lambda X} &\leq C_\alpha e^{\alpha|x|+\lambda X} \\ &\leq C_\alpha\{e^{(\lambda-\alpha)X} + e^{(\lambda+\alpha)X}\}, \end{aligned}$$

es entonces claro que si tomamos $\lambda \in (\lambda_0 - \alpha, \lambda_0 + \alpha)$, $\lambda + \alpha$ y $\lambda - \alpha$ estar en el interior del dominio de Λ y por esto, tenemos que el miembro derec de 3.1.1(2) es integrable, lo que nos permite aplicar el teorema. Deducim entonces la derivabilidad, y así obtenemos la ecuación requerida.

$$\begin{aligned} \Lambda(\lambda) &= \log[\mathbb{E}(e^{\lambda X})] = \log\left[\int_{\mathbb{R}} e^{\lambda X}\mu(dx)\right] \\ \text{y} \qquad \Lambda'(\lambda) &= \frac{1}{\int_R e^{\lambda X}\mu(dx)} \cdot \frac{\partial}{\partial x}\left(\int_{\mathbb{R}} e^{\lambda X}\mu(dx)\right) \\ &= \frac{1}{\mathbb{E}(e^{\lambda X})}\int_{\mathbb{R}} \frac{\partial}{\partial x}\left(e^{\lambda X}\right)\mu(dx) \\ &= \frac{1}{\mathbb{E}(e^{\lambda X})}\int_{\mathbb{R}} Xe^{\lambda X}\mu(dx) \\ &= \frac{\mathbb{E}(Xe^{\lambda X})}{\mathbb{E}(e^{\lambda X})}. \end{aligned}$$

3. Supongamos que $\mu(0,\infty) > 0$, entonces existe $a > 0$ tal que $\mathbb{P}(X \geq a) >$ Entonces

$$\begin{aligned} \mathbb{E}(e^{\lambda X}) &\geq \mathbb{E}(e^{\lambda X}\cdot 1_{\{X\geq a\}}) \\ &\geq \mathbb{E}(e^{\lambda a}\cdot 1_{\{X\geq a\}}) \\ &= e^{\lambda a}\mathbb{E}(1_{\{X\geq a\}}) \\ &= e^{\lambda a}\cdot \mathbb{P}(X \geq a), \end{aligned}$$

así,

$$\begin{aligned} \mathbb{E}(e^{\lambda X}) &\geq e^{\lambda a}\,\mathbb{P}(X\geq a) \\ \log[\mathbb{E}(e^{\lambda X})] &\geq \lambda a + \log[\mathbb{P}(X \geq a)], \end{aligned}$$

por lo que

$$\lim_{\lambda\to\infty}\Lambda(\lambda) \geq \lim_{\lambda\to\infty}(\lambda a + \log[\mathbb{P}(X\geq a)]) = \infty$$

concluyendo que

$$\lim_{\lambda\to\infty}\Lambda(\lambda) = +\infty.$$

4. Supongamos que $\mu(-\infty, 0) = 0$, entonces

$$\Lambda(\lambda) = \log\left\{\int_0^{+\infty} e^{\lambda x}\mu(dx)\right\}$$

como $e^{\lambda X}$ es creciente, y log también, entonces $\Lambda(\lambda)$ es creciente. Por otra parte, si $\lambda < 0$, sobre $\mathbb{R}^+$ se tiene que $e^{\lambda x} \leq 1$, que es integrable. Cuando λ tiende hacia $-\infty$,

$$e^{\lambda x} \longrightarrow \begin{cases} 0 & \text{si} \quad x > 0 \\ 1 & \text{si} \quad x = 0 \end{cases}$$

Ahora, por el teorema de convergencia dominada,

$$\int_0^{+\infty} e^{\lambda x}\mu(dx) \longrightarrow \mu(\{0\})$$

lo que nos da el resultado.

A partir de la función log-Laplace definiremos la función de tasa , es decir, transformada de Cramér de la distribución de probabilidad μ.

efinición 3.1.2. Se llama **transformada de Cramér** de μ a la función Λ^* na función de x) definida sobre $\mathbb{R}$ que toma valores en $[0, +\infty]$ por

$$\Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \in \mathbb{R}\}. \tag{3.1.1}$$

Hacemos notar que $\Lambda^*(x)$ es positiva en la medida en que, cuando $\lambda = 0$, la ınsformada de Cramér es 0.

En todo lo que sigue, X es semi-integrable, es decir, $\mathbb{E}(X^+)$ o $\mathbb{E}(X^-)$ es finito a siguiente notación se da $\bar{x} = \mathbb{E}[X]$.

Recordemos que en la sección 2.3 usamos que $\Lambda(\cdot)$ es diferenciable, en las uientes propiedades demostraremos esa afirmación.

oposición 3.1.2. *Recordando que $\mathcal{D}(\lambda) = \{\lambda : \Lambda(\lambda) < \infty\}$ se tiene que,*

. *$\Lambda^*(x)$ es convexa y s.c.i..*

. *Sı η está en el interior de $\mathcal{D}(\lambda)$ y si $y = \Lambda'(\eta)$, entonces $\Lambda^*(y) = \eta y - \Lambda(\eta)$.*

3. Si $\mathcal{D}(\lambda) = \{0\}$ entonces $\Lambda^*(x) = 0$.

4. $\inf\{\Lambda^*(x) : x \in \mathbb{R}\} = 0$ *y si $\bar{x} \in \mathbb{R}$, entonces $\Lambda^*(\bar{x}) = 0$.*

5. Si existe $\lambda_0 > 0$ tal que $\Lambda(\lambda_0) < \infty$, entonces $\bar{x} < +\infty$ y

$$\textit{para todo} \qquad x \geq \bar{x}, \qquad \Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \geq 0\}.$$

Además, $\Lambda^*(x)$ es creciente sobre $[\bar{x}, +\infty)$ (como caso particular tenemo que $\lim_{y \to +\infty} \Lambda^*(y) = +\infty$).

Demostración.

1. Λ^* se puede ver como $\sup\{ax + b\}$, ya que es una función que depende d x, por lo tanto es convexa, y es s.c.i. como supremo de funciones continuas

2. Sea η en el interior del dominio de Λ, y sea $y = \Lambda'(\eta)$. Escribimos

$$\Lambda^*(y) = \sup\{\lambda y - \Lambda(\lambda) : \lambda \in \mathbb{R}\}$$

y sea

$$g(\lambda) = \lambda y - \Lambda(\lambda),$$

como $\Lambda(\lambda)$ es convexa, entonces $-\Lambda(\lambda)$ es cóncava y λy es afín, entonc $g(\lambda)$ es cóncava. Además, g es derivable y

$$g'(\lambda) = y - \Lambda'(\lambda),$$

por lo que,

$$g'(\eta) = y - \Lambda'(\eta) = \Lambda'(\eta) - \Lambda'(\eta) = 0$$

lo que produce que $g'(\eta) = 0$. Entonces g es una función cóncava cu derivada se anula en η, por lo tanto el supremo se alcanza en este pun de donde concluimos que

$$\Lambda^*(y) = \eta y - \Lambda(\eta).$$

3. Si el dominio de Λ se reduce a 0, entonces $\Lambda(\lambda) \equiv +\infty$, lo que produce $\subseteq$ $\lambda x - \Lambda(\lambda)$ valga $-\infty$ salvo en 0, donde vale 0.

4. Si $\bar{x} \in \mathbb{R}$, se sabe que $\Lambda^*(\bar{x}) \geq 0$. Por lo que nos falta mostrar que $\Lambda^*(\bar{x}) \leq 0$. Para esto, emplearemos la desigualdad de Jensen en el caso de tener funciones cóncavas y es la siguiente: si X es una v.a., y $H : \mathbb{R} \to \mathbb{R}$ una función cóncava, entonces

$$\mathbb{E}(H(X)) = \mathbb{E}(-U(X)) \leq -U(\mathbb{E}(X)) = H(\mathbb{E}(X)).$$

Como el logaritmo es una función cóncava, entonces,

$$\lambda\bar{x} = \lambda\mathbb{E}(X) = \mathbb{E}(\lambda X) = \mathbb{E}(\log(e^{\lambda X})) \leq \log(\mathbb{E}(e^{\lambda X})) = \Lambda(\lambda)$$

de aquí que

$$\begin{array}{rcl} & \lambda\bar{x} & \leq \Lambda(\lambda) \\ \Leftrightarrow & \lambda\bar{x} - \Lambda(\lambda) & \leq 0 \\ \Leftrightarrow & \sup\{\lambda\bar{x} - \Lambda(\lambda)\} & \leq 0 \end{array}$$

por lo que

$$\Lambda^*(\bar{x}) = \sup\{\lambda\bar{x} - \Lambda(\lambda) : \lambda \in \mathbb{R}\} \leq 0.$$

Así, tenemos que $\Lambda^*(\bar{x}) = 0$.

5. Sea $\lambda_0 \geq 0$ tal que $\Lambda(\lambda_0) < \infty$. Como $e^{\lambda_0 X^+} \geq 1 + \lambda_0 X^+$, tenemos que

$$\begin{array}{rcl} & \lambda_0 X^+ & \leq e^{\lambda_0 X^+} - 1 \\ \Leftrightarrow & \mathbb{E}(\lambda_0 X^+) & \leq \mathbb{E}(e^{\lambda_0 X^+} - 1) \\ \Leftrightarrow & \mathbb{E}(X^+) & \leq \dfrac{\mathbb{E}(e^{\lambda_0 X^+} - 1)}{\lambda_0} \end{array}$$

lo que significa que $\bar{x} \leq +\infty$.

Supongamos ahora que $\bar{x} \in \mathbb{R}$. Sea $\lambda < 0$, para $x \geq \bar{x}$ tenemos que $\lambda x \leq \lambda\bar{x}$ debido a que $\lambda < 0$, entonces $\lambda x - \Lambda(\lambda) \leq \lambda\bar{x} - \Lambda(\lambda)$. Tomando supremos de ambos lados obtenemos

$$\sup\{\lambda x - \Lambda(\lambda) : \lambda < 0\} \leq \sup\{\lambda\bar{x} - \Lambda(\lambda) : \lambda < 0\} = \Lambda^*(\bar{x}) = 0$$

lo que prueba que la contribución de las $\lambda < 0$ no cuenta. Finalmente, verificar el crecimiento de $\Lambda^*(x)$ es fácil. Para $x > y > \bar{x}$ tenemos que $\lambda y - \Lambda(\lambda) < \lambda x - \Lambda(\lambda)$ $\lambda \geq 0$, entonces $\Lambda^*(y) \leq \Lambda^*(x)$.

Ahora, daremos el cálculo explícito de Λ y de Λ^* para $\mu = \mathrm{N}(0,1)$, más adelante se mostraran otros ejemplos con otras distribuciones.

$$\begin{aligned}\mathbb{E}[e^{\lambda X}] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left(\lambda x - \frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}}\left[\int_{\mathbb{R}} \exp\left(-\frac{1}{2}(x-\lambda)^2\right) dx\right] e^{\lambda^2/2} \\ &= e^{\lambda^2/2}.\end{aligned}$$

Entonces $\Lambda(\lambda) = \lambda^2/2$, ya que $\Lambda(\lambda) = \log(\mathbb{E}[e^{\lambda X}])$. Como la parte negativa no cuenta

$$\begin{aligned}\Lambda^*(x) &= \sup\{\lambda x - \lambda^2/2 : \lambda \in \mathbb{R}\} \\ &= \sup\{-\tfrac{1}{2}(\lambda - x)^2 + x^2/2 : \lambda \in \mathbb{R}\} \\ &= x^2/2\end{aligned}$$

Podemos ver que teníamos, en el ejemplo de las Colas de la distribución normal, un PGD con función de tasa $\Lambda^*(x)$.

3.2 El Teorema de Cramér.

Teorema 3.1. *Sea X_n una sucesión de v.a. reales independientes y con la misma distribución de probabilidad μ. Sea $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ y μ_n la distribución de probabilidad empírica de $\overline{X}_n$. Entonces la sucesión μ_n satisface un PGD con función de tasa, la función Λ^*, transformada de Cramér de μ, es decir,*

$$\textit{para todo}\, O \text{ abierto } \varliminf_{n\to\infty} \frac{1}{n}\log \mu_n(O) \geq -\inf\{\Lambda^*(x) : x \in O\}$$

$$\textit{para todo}\, F \text{ cerrado } \varlimsup_{n\to\infty} \frac{1}{n}\log \mu_n(F) \leq -\inf\{\Lambda^*(x) : x \in F\}$$

Por otra parte, se tiene la siguiente estimación que resulta ser más fina

$$\mu_n(F) \leq 2\exp(-n \inf\{\Lambda^*(x) : x \in F\}). \qquad 3.1$$

Demostración.

Para simplificar supongamos que μ cumple las hipótesis siguientes:
μ es de soporte compacto y $\mu(-\infty, 0) > 0$, $\mu(0, +\infty) > 0$.

1. **Minoración.**

La minoración consiste en demostrar que para todo O abierto

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log\, \mu_n(O) \geq -\inf\{\Lambda^*(x) : x \in O\}.$$

La idea principal en la demostración será hacer un cambio de probabilidad. Sea entonces O un abierto no vacío. Para probar el resultado basta con demostrar que para toda $x \in O$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(O) \geq -\Lambda^*(x), \qquad 3.2.2$$

esto es,

$$\begin{aligned} &\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mu_n(O) &&\geq -\Lambda^*(x) \\ \iff &\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mu_n(O) &&\geq \sup(-\Lambda^*(x) : x \in O) \\ \iff &\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mu_n(O) &&\geq -\inf(\Lambda^*(x) : x \in O). \end{aligned}$$

Demostrar 3.2.2 equivale a demostrar que para toda $x \in O$ y $\delta > 0$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(x-\delta, x+\delta) \geq -\Lambda^*(x), \qquad 3.2.3$$

pues entonces para toda $x \in O$ existe δ tal que

$$(x-\delta, x+\delta) \subset O,$$

por lo que,

$$\mu_n(x-\delta, x+\delta) \leq \mu_n(O)$$

y

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(O) \geq \underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(x-\delta, x+\delta) \geq -\Lambda^*(x),$$

de donde se concluye 3.2.2.

Para demostrar 3.2.3, vamos a suponer que hemos demostrado que para toda $\delta > 0$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq -\Lambda^*(0). \qquad 3.2.4$$

Partiendo de esto, definimos una nueva v.a. Y_n que sigue una distribució
de probabilidad μ_n de la siguiente manera

$$Y_n = (X_n - x).$$

Y por 3.2.4 tenemos para toda $\delta > 0$ que

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Y_n \in (-\delta, \delta)) \geq -\Lambda_Y^*(0).$$

Como

$$\begin{aligned}
\Lambda_Y(\lambda) &= \log(\mathbb{E}[e^{\lambda Y}]) \\
&= \log\left(\mathbb{E}\left[e^{\lambda(X_n - x)}\right]\right) \\
&= \log\left(e^{-\lambda x} \cdot \mathbb{E}[e^{\lambda(X_n)}]\right) \\
&= -\lambda x + \Lambda(\lambda)
\end{aligned}$$

y

$$\begin{aligned}
\Lambda_Y^*(z) &= \sup\{\lambda z - \Lambda_Y(\lambda)\} \\
&= \sup\{\lambda z + \lambda x - \Lambda(\lambda)\} \\
&= \sup\{\lambda(z + x) - \Lambda(\lambda)\} \\
&= \Lambda^*(z + x)
\end{aligned}$$

entonces para toda $\delta > 0$

$$\begin{aligned}
&\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}((X_n - x) \in (-\delta, \delta) \geq -\Lambda_Y^*(0) = \Lambda^*(x) \\
\iff\ &\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}(X_n \in (x - \delta, x + \delta) \geq \Lambda^*(x) \\
\iff\ &\underline{\lim}_{n\to\infty} \tfrac{1}{n} \log \mu_n(x - \delta, x + \delta) \geq \Lambda^*(x)
\end{aligned}$$

que es 3.2.3.

Demostraremos ahora 3.2.4. Por hipótesis μ es de soporte compacto, que implica que $\mathcal{D}(\Lambda) = \mathbb{R}$. Daremos una expresión explícita de $\Lambda^*(0)$, acuerdo con la proposición 3.1.1 (3), basta encontrar η tal que $\Lambda'(\eta) =$ Pero como μ toma valores en $\mathbb{R}^+$ y $\mathbb{R}^-$, se tiene, según la proposición 3. (3), que $\Lambda(\lambda)$ tiende hacia $+\infty$ cuando $|\lambda| \to +\infty$. Ahora, como Λ convexa, se deduce la existencia de un mínimo, es decir, un punto η tal q $\Lambda'(\eta) = 0$, por lo que

$$\Lambda^*(0) = 0\eta - \Lambda(\eta) = -\Lambda(\eta).$$

Hacemos ahora el cambio de probabilidad, definimos v como,

$$\frac{dv}{d\mu}(x) = \frac{\exp(\eta x)}{\int_{\mathbb{R}} \exp(\eta x)\mu dx} = \exp(\eta x - \Lambda(\eta)),$$

observamos que si $\varepsilon \in (0, \delta)$, entonces $\mu_n(-\delta, \delta) \geq \mu_n(-\varepsilon, \varepsilon)$ y empleando el Teorema de Cambio de Variable

$$\begin{aligned}
\mu_n(-\varepsilon, \varepsilon) &= \int_\Omega 1_{\{-\varepsilon,\varepsilon\}}(X)\mu(dx) \\
&= \int_\Omega \left(1_{\{-\varepsilon,\varepsilon\}} \circ \overline{X}_n\right)(\omega) d\mathbb{P}(\omega) \\
&= \int_\Omega 1_{\{\overline{X}_n \in (-\varepsilon,\varepsilon)\}}(\omega) d\mathbb{P}(\omega) \\
&= \int_\Omega 1_{\{\sum_{i=1}^n X_i \in (-n\varepsilon, n\varepsilon)\}}(\omega) d\mathbb{P}(\omega) \\
&= \int_{\mathbb{R}^n} 1_{\{\sum_{i=1}^n x_i \in (-n\varepsilon, n\varepsilon)\}}(x_i \dots x_n) d(\mu_1 \dots \mu_n)(x_i \dots x_n) \\
&= \int_{\mathbb{R}^n} 1_{\{\sum_{i=1}^n x_i \in (-n\varepsilon, n\varepsilon)\}}(x_i \dots x_n) \mu(dx_1) \dots \mu(dx_n) \\
&= \int_{\{|\frac{1}{n}\sum_{k=1}^n x_k| < \varepsilon\}} \mu(dx_1)\mu(dx_2) \dots \mu(dx_n),
\end{aligned}$$

donde la penúltima desigualdad se sigue por independencia. Sobre el conjunto $\{|\frac{1}{n}\sum_{k=1}^n x_k| < \varepsilon\}$, se tiene que

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{k=1}^n x_k \leq \left|\tfrac{1}{n}\sum_{k=1}^n x_k\right| < \varepsilon \\
\iff & \eta \textstyle\sum_{k=1}^n x_k \leq n\varepsilon|\eta| \\
\iff & \eta \textstyle\sum_{k=1}^n x_k - n\varepsilon|\eta| \leq 0 \\
\iff & \exp\left(\eta \textstyle\sum_{k=1}^n x_k - n\varepsilon|\eta|\right) \leq 1,
\end{aligned}$$

por lo tanto,

$$\mu_n(-\varepsilon, \varepsilon) \geq \int_{\{|\frac{1}{n}\sum\limits_{k=1}^n x_k| < \varepsilon\}} \exp\left(\eta \sum_{k=1}^n x_k - n\varepsilon|\eta|\right) \mu(dx_1)\mu(dx_2) \dots \mu(dx_n)$$

y desarrollando el integrando de la desigualdad anterior se tiene

$$\exp\left(\eta\sum_{k=1}^{n}x_k - n\varepsilon|\eta|\right) = \exp\left(\sum_{k=1}^{n}(\eta x_k - \varepsilon|\eta|)\right) = \prod_{k=1}^{n}\big(\exp(\eta x_k - \varepsilon|\eta|)$$

por lo que

$$\begin{aligned}
\mu_n(-\varepsilon,\varepsilon) &\geq \int_{A_1\times\cdots\times A_n}\prod_{k=1}^{n}\big(\exp(\eta x_k-\varepsilon|\eta|)\big)\mu(dx_1)\mu(dx_2)\ldots\mu(dx_n)\\
&= \int_{A_1}\exp(\eta x_1-\varepsilon|\eta|)\mu(dx_1)\ldots\int_{A_n}\exp(\eta x_n-\varepsilon|\eta|)\mu(dx_n)\\
&= \int_{A_1}\exp(\eta x_1-\Lambda(\eta)+\Lambda(\eta)-\varepsilon|\eta|)\mu(dx_1)\ldots\\
&\qquad\qquad \int_{A_n}\exp(\eta x_n-\Lambda(\eta)+\Lambda(\eta)-\varepsilon|\eta|)\mu(dx_n)\\
&= \exp\big(\Lambda(\eta)-\varepsilon|\eta|\big)\int_{A_1}\exp(\eta x_1-\Lambda(\eta))\mu(dx_1)\ldots\\
&\qquad\qquad \exp\big(\Lambda(\eta)-\varepsilon|\eta|\big)\int_{A_n}\exp(\eta x_n-\Lambda(\eta))\mu(dx_n)\\
&= \exp\big(n\Lambda(\eta)-n\varepsilon|\eta|\big)\\
&\quad\left(\int_{A_1\times\cdots\times A_n}\prod_{k=1}^{n}\exp\big(\eta x_k-\Lambda(\eta)\big)\mu(dx_1)\mu(dx_2)\ldots\mu(dx_n)\right)\\
&= \exp\big(n\Lambda(\eta)-n\varepsilon|\eta|\big)\left(\int_{A_1\times\cdots\times A_n}v(dx_1)\ldots v(dx_n)\right) \qquad 3.2.
\end{aligned}$$

Así, obtenemos la desigualdad 3.2.5. Por la proposición 3.1.1(2), saben que existe η tal que

$$0 = \Lambda'(\eta) = \frac{\mathbb{E}[Xe^{\eta X}]}{\mathbb{E}[e^{\eta X}]},$$

esto nos lleva inmediatamente a decir que $\mathbb{E}[Xe^{\eta X}] = 0$ de donde tenen que

$$\begin{aligned}
\mathbb{E}[Xe^{\eta X}] &= \int_{\mathbb{R}} xe^{\eta x}\mu(dx) \\
&= \int_{\mathbb{R}} xe^{\eta x}\frac{v(dx)}{\exp(\eta x - \Lambda(\eta))} \\
&= \int_{\mathbb{R}} \frac{xe^{\eta x}v(dx)}{e^{\eta x}\cdot\exp(-\Lambda(\eta)} \\
&= \exp(\Lambda(\eta))\int_{R} xv(dx) = 0
\end{aligned}$$

por lo tanto

$$\int_{\mathbb{R}} xv(dx) = 0. \tag{3.2.6}$$

Observamos ahora lo que se obtiene asintóticamente

$$\begin{aligned}
& \mu_n(-\delta,\delta) && \geq \mu_n(-\varepsilon,\varepsilon) \\
\Longleftrightarrow\quad & \log\mu_n(-\delta,\delta) && \geq \log\mu_n(-\varepsilon,\varepsilon) \\
\Longleftrightarrow\quad & \tfrac{1}{n}\log\mu_n(-\delta,\delta) && \geq \tfrac{1}{n}\log\mu_n(-\varepsilon,\varepsilon) \\
\Longleftrightarrow\quad & \varliminf_{n\to\infty}\tfrac{1}{n}\log\mu_n(-\delta,\delta) && \geq \varliminf_{n\to\infty}\tfrac{1}{n}\log\mu_n(-\varepsilon,\varepsilon)
\end{aligned}$$

$$\begin{aligned}
&\geq \varliminf_{n\to\infty}\tfrac{1}{n}\log\left[\left(\int_{\{|\frac{1}{n}\sum_{k=1}^{n}x_k|<\varepsilon\}} v(dx_1)v(dx_2)\ldots v(dx_n)\right)\exp(n\Lambda(\eta)-n\varepsilon|\eta|)\right] \\
&= \varliminf_{n\to\infty}\tfrac{1}{n}\log\left[v_n(-\epsilon,\epsilon)\cdot\exp(n\Lambda(\eta)-n\varepsilon|\eta|)\right] \\
&= \varliminf_{n\to\infty}\left[\tfrac{1}{n}\log < v_n(\varepsilon,\varepsilon)+\Lambda(\eta)-\varepsilon|\eta|\right]
\end{aligned}$$

Por otra parte, la ley de los grandes números nos dice que: *Dadas* $X_1, X_2, \ldots$ *v.a. independientes (no necesariamente con la misma disribución) con media y varianza finita. Supongamos que las varianzas están acotadas uniformemente por* $M < \infty$. *Sea* $S_n = (X_1 + \cdots + X_n)/n$, *entonces* $[S_n - \mathbb{E}(S_n)]/n$ *converge en probabilidad a 0, esto es,*

$$\mathbb{P}\left[\left|\frac{S_n-\mathbb{E}(S_n)}{n}\right| \geq \varepsilon\right] \underset{n\to\infty}{\longrightarrow} 0$$

o, lo que es lo mismo

$$\mathbb{P}\left[\left|\frac{S_n-\mathbb{E}(S_n)}{n}\right| < \varepsilon\right] \underset{n\to\infty}{\longrightarrow} 1$$

Ahora, por la ecuación 3.2.6 tenemos que $\mathbb{E}_v[X] = 0$, y por la ley débil de los grandes números,

$$\mathbb{P}\left[\left|\frac{\overline{X}_n - 0}{n}\right| < \varepsilon\right] \underset{n\to\infty}{\longrightarrow} 1 \Longleftrightarrow \mathbb{P}\left[\left|\frac{\overline{X}_n}{n}\right| < \varepsilon\right] \underset{n\to\infty}{\longrightarrow} 1$$

lo que es igual a

$$v_n(-\varepsilon, \varepsilon) \underset{n\to\infty}{\longrightarrow} 1$$

por lo que

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq \Lambda(\eta) - \varepsilon|\eta|,$$

si $\varepsilon \underset{n\to\infty}{\longrightarrow} 0$ y viendo que el factor $1/n$ no tiene ninguna utilidad, obtenemos

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq \Lambda(\eta) = -\Lambda^*(0)$$

como se quería demostrar.

2. **Mayoración.**

   La prueba que se dará no podrá generalizarse en $\mathbb{R}^d$ ya que utilizaremos la propiedad de $\mathbb{R}$ de poseer orden.

   Sea F un cerrado en $\mathbb{R}$. La mayoración es trivial si tenemos que

$$\inf\{\Lambda^*(x) : x \in F\} = 0,$$

   pues entonces equivale a demostrar que $\mu_n(F) \leq 1$, lo que es claro. Primero notemos que

$$\inf\{\Lambda^*(x) : x \in F\} = 0 \Longleftrightarrow 0 = -\inf\{\Lambda^*(x) : x \in F\}$$

   entonces

$$\begin{array}{lll} & \mu_n(F) & \leq 1 \\ \Longleftrightarrow & \log \mu_n(F) & \leq \log 1 = 0 \\ \Longleftrightarrow & \frac{1}{n} \log \mu_n(F) & \leq 0 \\ \Longleftrightarrow & \overline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(F) & \leq 0 = -\inf\{\Lambda^*(x) : x \in F\} \\ \Longleftrightarrow & \overline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(F) & \leq -\inf\{\Lambda^*(x) : x \in F\} \end{array}$$

Supongamos entonces que $\inf\{\Lambda^*(x) : x \in F\} > 0$. En la proposición 3.1.2 (3) vimos que si $\mathcal{D}(\Lambda) = 0$, entonces $\Lambda^*(x) = 0$, por lo que $\inf\{\Lambda^*\} = 0$, ahora, como estamos negando esta afirmación, tenemos que $\mathcal{D}(\Lambda) \neq \{0\}$.

Supongamos primero que 0 está en el interior de $\mathcal{D}(\Lambda)$. Esto significa que X_1 tiene momentos exponenciales a derecha y a izquierda de 0, ya que podemos dar una vecindad alrededor del 0 completamente contenida en $\mathcal{D}(\Lambda)$ en la que existan $(\lambda_i)_{i\in\mathbb{R}}$ tales que $\Lambda(\lambda_i) < \infty$, por lo que podemos aplicar la proposición 3.1.2 (5) para deducir que $\overline{x} \in \mathbb{R}$ (recordamos que $\bar{x} = \mathbb{E}(X_1)$) y por la afirmación 3.1.2 (4) también que $\Lambda^*(\bar{x}) = 0$. Utilizaremos ahora la desigualdad de Chebychev que nos dice : Si $\alpha > 0$, $t \in \mathbb{R}$, Y v. a. r. y f una función continua entonces

$$\mathbb{P}(f(Y) \geq t) \leq \frac{1}{t^\alpha}\mathbb{E}(f(Y)^\alpha)$$

y si f es la función identidad,

$$\begin{aligned}\mathbb{P}(Y \geq t) &= \mathbb{P}(e^Y \geq e^t)\\ &= \mathbb{P}(e^Y \cdot e^{-t} \geq 1)\\ &\leq \mathbb{E}\left[e^{\alpha(Y-t)}\right].\end{aligned} \qquad 3.2.7$$

Sea entonces $\lambda > 0$ tal que $\Lambda(\lambda) < \infty$, tenemos, según la desigualdad 3.2.7 con $\alpha = n\lambda > 0$ y $Y = \overline{X}$ que

$$\begin{aligned}\mu_n(t,+\infty) &= \mathbb{P}(\overline{X}_n \geq t)\\ &\leq \mathbb{E}\left[e^{n\lambda(\overline{X}_n - t)}\right]\\ &= \mathbb{E}\left[e^{n\lambda\left(\frac{\sum X_i}{n}\right)}e^{(-n\lambda t)}\right]\\ &= e^{-n\lambda t}\,\mathbb{E}[e^{\lambda\sum X_1}]\\ &\underset{\text{indep}}{=} e^{-n\lambda t}\left(\mathbb{E}[e^{\lambda X_1}]\right)^n\\ &= e^{-n\lambda t}\left(\exp(n\log\mathbb{E}[e^{\lambda X_1}])\right)\\ &= e^{-n\lambda t}\cdot\exp[n\,\Lambda(\lambda)]\\ &= \exp\left[n\left(\Lambda(\lambda) - \lambda t\right)\right]\end{aligned}$$

por lo tanto

$$\mu_n(t,+\infty) \leq \exp\left[-n\left(\lambda t - \Lambda(\lambda)\right)\right] \qquad 3.2.8$$

lo cual también se cumple para $\lambda = 0$.

Según la proposición 3.1.2 (5), si tomamos $t \geq \bar{x}$, entonces

$$\Lambda^*(t) = \sup\{\lambda t - \Lambda(\lambda) : \lambda \geq 0\},$$

de donde, tomando el supremo en la desigualdad 3.2.8, encontramos:

$$\begin{aligned} \mu_n(t, +\infty) &= \sup\{\mu_n(t, +\infty)\} \\ &\leq \sup\{\exp\left[-n\left(\lambda t - \Lambda(\lambda)\right)\right]\} \\ &= \exp\left[-n \sup\{\lambda t - \Lambda(\lambda)\}\right] \end{aligned}$$

por lo que

$$\mu_n[t, +\infty] \leq \exp[-n\Lambda^*(t)] \qquad t \geq \bar{x}. \qquad 3.2.$$

de manera análoga mostraremos que

$$\mu_n[-\infty, t] \leq \exp[-n\Lambda^*(t)] \qquad t \leq \bar{x} \qquad 3.2.1$$

para esto, aplicaremos la desigualdad de Chebychev con $\alpha > 0$, $t \in \mathbb{R}$ y una v.a. $\mathbb{P}(Y \leq t)$, equivalente a que:

$$\mathbb{P}(e^Y \leq e^t) = \mathbb{P}(e^{(t-Y)} \geq 1) \leq \mathbb{E}(e^{\alpha(Y-t)}).$$

Por la desigualdad anterior con $\alpha = n\lambda$ y $Y = \overline{X}$ y haciendo unos calcul similares para obtener 3.2.8 se tiene que

$$\mu_n(-\infty, t) \leq \exp\left[-n\left(\lambda t - \Lambda(\lambda)\right)\right],$$

donde, tomando el supremo en la desigualdad anterior se tiene el resulta

Utilicemos ahora que $\Lambda^*(\bar{x}) = 0$. Como hicimos el supuesto de que ínfimo de las $\Lambda^*(x)$ es estrictamente positivo, deducimos necesariamer que $\bar{x} \notin F$, por lo que $\bar{x}$ está en el complemento de F, el cual es un abie en $\mathbb{R}$, entonces la componente conexa de F^c que contiene a $\bar{x}$ es un interv (a, b). De donde,

$$\begin{aligned} F &\subset (-\infty, a] \cup [b, +\infty) & a < \bar{x} < b \\ \mu_n(F) &\leq \mu_n(-\infty, a] + \mu_n[b, +\infty) & a < \bar{x} < b \end{aligned}$$

de ahí que, utilizando 3.2.9 y 3.2.10 se tenga

$$\mu_n(F) \leq \exp[-n\Lambda^*(a)] + \exp[-n\Lambda^*(b)].$$

Si, por ejemplo $\Lambda^*(a) \geq \inf\{\Lambda^*(x) : x \in F\}$, como $a < b$ entonces:

$$\begin{aligned} \lambda a &< \lambda b \\ \lambda a - \Lambda(\lambda) &< \lambda b - \Lambda(\lambda) \\ \sup\{\lambda a - \Lambda(\lambda)\} &< \sup\{\lambda b - \Lambda(\lambda)\} \\ \Lambda^*(a) &< \Lambda^*(b) \end{aligned}$$

por lo que se deduce que $\Lambda^*(b) \geq \inf\{\Lambda^*(x) : x \in F\}$, de lo que obtenemos

$$\begin{aligned} \mu_n(F) &\leq \exp[-n\Lambda^*(a)] + \exp[-n\Lambda^*(b)] \\ &\leq \exp[-n\inf\{\Lambda^*(x) : x \in F\}] + \exp[-n\inf\{\Lambda^*(x) : x \in F\}] \\ &= 2\exp[-n\inf\{\Lambda^*(x) : x \in F\}] \end{aligned}$$

que es exactamente 3.2.1.

Supongamos ahora que 0 no está en el interior de $\mathcal{D}(\Lambda)$, por ejemplo, tenemos para toda $\lambda < 0$, $\Lambda(\lambda) = +\infty$. Entonces el $\lim_{x\to\infty}\Lambda^*(x) = +\infty$, Λ^* creciente sobre $[\bar{x}, +\infty)$. Por la proposición 3.1.2 (5) sabemos que $\bar{x} \in \mathbb{R}^+$, y que $\Lambda^*(\bar{x}) = 0$, si usamos la desigualdad de Chebychev exponencial aplicada a $t > 0$, y la proposición 3.1.2 (5), obtenemos nuevamente que

$$\mu_n(t, \infty) \leq \exp[-n\Lambda^*(t)] \qquad t \geq \bar{x},$$

como por hipótesis, el ínfimo de las $\Lambda^*(x)$ es estrictamente positivo, también en este caso $\bar{x} \notin F$, por lo que $\bar{x}$ está contenido en el abierto $(0, b)$, entonces,

$$F \subset [b, \infty) \qquad 0 < \bar{x} \leq b$$

de donde

$$\mu_n(F) \leq \exp[-n\inf\{\Lambda^*(x) : x \in F\}]$$

dando el resultado más fino que 3.2.1.

El problema que podemos plantearnos es saber cuando hay convergencia y no nplemente las mayoraciones y minoraciones. El corolario siguiente responde rcialmente al problema.

ɔrolario 3.2.1. *Bajo las hipótesis del teorema 3.1, tenemos*

$$\lim_{n\to\infty}\frac{1}{n}\log\mu_n[t, +\infty) = -\inf\{\Lambda^*(x) : x \geq t\}$$

Demostración.

Refinaremos la demostración de la minoración. Utilizando que, para toda $x \geq t$ y toda $\delta > 0$

$$\begin{array}{rrl} & [x, x+\delta) & \subset [t, +\infty) \\ \Longrightarrow & \mu_n[x, x+\delta) & \leq \mu_n[t, +\infty) \\ \Longrightarrow & \underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n[t, +\infty) & \geq \underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(x, x+\delta) \end{array}$$

análogo a la prueba de la minoración, sólo hay que demostrar

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(x, x+\delta) \geq -\Lambda^*(x) \qquad 3.2.1$$

esto implica que

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n([t, +\infty)) \leq -\inf\{\Lambda^*(x) : x \geq t\}$$

para demostrar 3.2.11 se rehace el cambio de variable $Y = X_n - x$ hecho en l prueba de la minoración, y con la suposición que

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n[0, \delta) \geq -\Lambda^*(0)$$

se tiene el resultado. Para mostrar la desigualdad anterior se utiliza el camb de probabilidad que se hizo en la minoración, entonces, para un η tal que $\Lambda^*(0) = -\Lambda(\eta)$ y para $\varepsilon < \delta$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n[0, \delta) \geq \underline{\lim}_{n\to\infty} \left[\Lambda(\eta) - \varepsilon|\eta| + \frac{1}{n} \log v_n[0, \varepsilon)\right]$$

y

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n[0, \delta) \geq \left[\Lambda(\eta) - \varepsilon|\eta| + \underline{\lim}_{n\to\infty} \frac{1}{n} \log v_n[0, \varepsilon)\right].$$

En este caso, para hallar la convergencia de $v_n[0, \varepsilon)$, no se puede utilizar ley débil de los grandes números, pues el intervalo a estimar no es simétri utilizamos entonces el teorema del límite central, para decir que $v_n[0, \varepsilon)$ tien hacia un límite finito estrictamente positivo, lo que permite controlar este últi término.

3.3 Cálculo de $\Lambda(\lambda)$ y $\Lambda^*(\lambda)$ para algunas variables aleatorias

Daremos ahora el cálculo explícito de la función log-Laplace y la transformada de Cramér para distintas distribuciones de probabilidad. También obtendremos la gráfica de la función de tasa correspondiente.

Ejemplo 3.3.1. Considerar X v.a. bernoulli con parámetro (p).

Como sabemos, la función generadora de momentos de esta variable aleatoria es

$$M(\lambda) = 1 + p(e^\lambda - 1)$$

por lo que

$$\log M(\lambda) = \log(1 + p(e^\lambda - 1)).$$

Vamos ahora a calcular el punto λ^* donde se maximiza la función

$$g(\lambda) = \lambda x - \log(1 + p(e^\lambda - 1)).$$

Entonces, se obtiene su derivada y se iguala a cero, así tenemos que

$$g'(\lambda) = x - \frac{pe^\lambda}{1 + p(e^\lambda - 1)} = 0$$

de donde

$$\begin{array}{rcl} x + pxe^\lambda - px & = pe^\lambda \\ \Rightarrow \quad x(1-p) & = pe^\lambda(1-x) \\ \Rightarrow \quad \frac{x(1-p)}{p(1-x)} & = e^\lambda \end{array}$$

se tiene que $\lambda^* = \log\left(\dfrac{x(1-p)}{p(1-x)}\right)$ y si calculamos $g(\lambda^*)$ obtenemos

$$\begin{aligned} \lambda^* x - \log M(\lambda^*) &= x\log\left(\tfrac{x(1-p)}{p(1-x)}\right) - \log\left(1 + p\left(\tfrac{x(1-p)}{p(1-x)} - 1\right)\right) \\ &= x\log\left(\tfrac{x(1-p)}{p(1-x)}\right) - \log\left(\tfrac{1-p}{1-x}\right) \\ &= x\log\left(\tfrac{x}{p}\right) + x\log\left(\tfrac{1-p}{1-x}\right) - \log\left(\tfrac{1-p}{1-x}\right) \\ &= x\log\left(\tfrac{x}{p}\right) + (1-x)\log\left(\tfrac{1-x}{1-p}\right) \end{aligned}$$

por lo que $\Lambda^*(x) = x \log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right)$ para $x \in [0,1]$ y notemos que $\mathcal{D}_\Lambda = \mathbb{R}$, pero $\Lambda^*(x)$ es discontinua.

Consideremos el caso particular de $p = \frac{1}{2}$, entonces tenemos que

$$\Lambda^*(x) = x \log x + (1-x)\log(1-x) + \log 2 \qquad \text{para} \qquad x \in [0,1]$$

y la gráfica de la función de tasa es

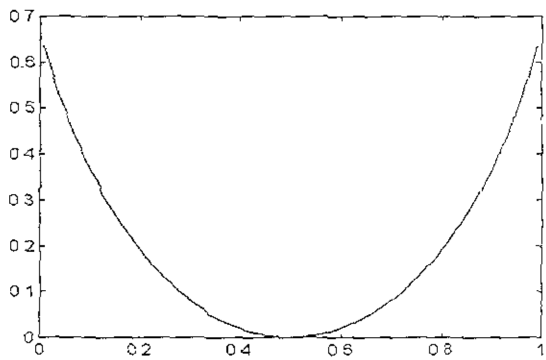

Figura 10. Función de tasa para una v.a. Bernoulli con parámetro $p = 1/2$.

Observación: La función de tasa que aquí obtuvimos corresponde al ejemplo de las colas de la distribución binomial que se presentó al inicio.

Ejemplo 3.3.2. Si X es una v.a. Exponencial con parámetro (θ).

Como sabemos, la función generadora de momentos para una v.a. con esta distribución es $M(\lambda) = \frac{\theta}{\theta - \lambda}$, por lo que $\log M(\lambda) = \log\left(\frac{\theta}{\theta-\lambda}\right)$ y por lo tanto

$$g(\lambda) = \lambda x - \log\left(\frac{\theta}{\theta - \lambda}\right) = \lambda x - \log\theta + \log(\theta - \lambda).$$

Se obtiene su derivada y se iguala a cero

$$g'(\lambda) = x - \frac{1}{\theta - \lambda} = 0$$

de donde

$$\begin{aligned} x &= \frac{1}{\theta - \lambda} \\ \Rightarrow \quad \theta - \lambda &= \frac{1}{x} \\ \Rightarrow \quad \theta - \frac{1}{x} &= \lambda^* \end{aligned}$$

con lo que $\lambda^* = \frac{\theta x - 1}{x}$ y si calculamos $g(\lambda^*)$ obtenemos

$$\begin{aligned} \lambda^* x - \log\theta + \log(\theta - \lambda^*) &= x\left(\frac{\theta x - 1}{x}\right) - \log\theta + \log\left(\theta - \left(\frac{\theta x - 1}{x}\right)\right) \\ &= \theta x - 1 - \log\theta + \log\left(\frac{1}{x}\right) \\ &= \theta x - 1 - \log\theta - \log x \end{aligned}$$

por lo que

$$\Lambda^*(x) = \theta x - 1 - \log(\theta x) \qquad \text{para} \qquad x > 0.$$

En el caso de que el parametro $\theta = 1$ tenemos que la función de tasa es

$$\Lambda^*(x) = x - 1 - \log x \qquad \text{para} \qquad x > 0.$$

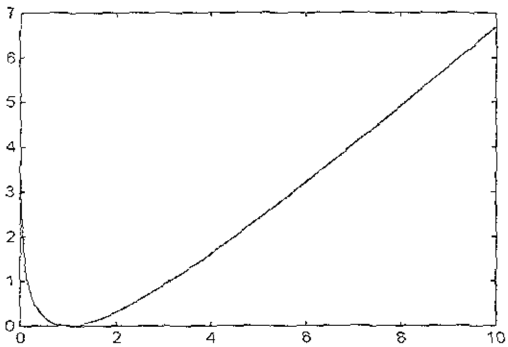

Figura 11. Función de tasa para una v.a. Exponencial con parámetro $\theta = 1$.

jemplo 3.3.3. Para X v.a. Poisson con parámetro (θ).

Como sabemos,la función generadora de momentos en este caso es

$$M(\lambda) = e^{\theta(e^{\lambda}-1)}$$

por lo que

$$\log M(\lambda) = \theta(e^{\lambda} - 1),$$

de esta forma,

$$g(\lambda) = \lambda x - \theta(e^{\lambda} - 1).$$

Se obtiene su derivada y se iguala a cero

$$g'(\lambda) = x - \theta e^{\lambda} = 0$$

de donde

$$\begin{aligned} x &= \theta e^{\lambda} \\ \Rightarrow \quad \frac{x}{\theta} &= e^{\lambda} \\ \Rightarrow \quad \lambda^* &= \log\left(\frac{x}{\theta}\right) \end{aligned}$$

así, tenemos que $\lambda^* = \log(x/\theta)$ y se calcula $g(\lambda^*)$ con lo que se obtiene

$$\begin{aligned} \lambda^* x - \theta(e^{\lambda} - 1) &= x \log\left(\tfrac{x}{\theta}\right) - \theta\left(\tfrac{x}{\theta} - 1\right) \\ &= x \log\left(\tfrac{x}{\theta}\right) - x + \theta \end{aligned}$$

por lo que $\Lambda^*(x) = \theta - x + x\log\left(\dfrac{x}{\theta}\right)$ para $x \geq 0$.

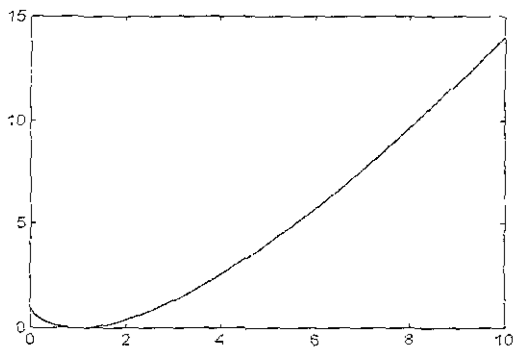

Figura 12. Función de tasa para una v.a. Poisson.

jemplo 3.3.4. Si X es una v.a. Binomial con parámetros (n,p).

En este caso, con la convención $q=(1-p)$, la función generadora de momentos es

$$M(\lambda)=(e^{\lambda}p+q)^n$$

or lo que $\log M(\lambda)=n\log(e^{\lambda}p+q)$,asi, tenemos que

$$g(\lambda)=\lambda x-n\log(e^{\lambda}p+q).$$

e obtiene su derivada y se iguala a cero, así tenemos que

$$g'(\lambda)=x-\frac{npe^{\lambda}}{pe^{\lambda}+q}$$

e donde

$$\begin{aligned} & xe^{\lambda}p+x-px &&= npe^{\lambda}\\ \Rightarrow\quad & x(1-p) &&= pe^{\lambda}(n-x)\\ \Rightarrow\quad & \frac{x(1-p)}{p(n-x)} &&= e^{\lambda}\end{aligned}$$

n lo que $\lambda^*=\log\left(\dfrac{x(1-p)}{p(n-x)}\right)$ y se calcula $g(\lambda^*)$:

$$\begin{aligned} g(\lambda^*) &= x\log\left(\tfrac{x(1-p)}{p(n-x)}\right)-n\log\left(\tfrac{x(1-p)}{p(n-x)}p+(1-p)\right)\\ &= x\log\left(\tfrac{x}{p}\right)+x\log\left(\tfrac{1-p}{n-x}\right)-n\log\left(\tfrac{x(1-p)}{n-x}+(1-p)\right)\\ &= x\log\left(\tfrac{x}{p}\right)+x\log\left(\tfrac{1-p}{n-x}\right)-n\log n-n\log\left(\tfrac{1-p}{n-x}\right)\\ &= x\log\left(\tfrac{x}{p}\right)+(x-n)\log\left(\tfrac{1-p}{n-x}\right)-n\log n\\ &= x\log\left(\tfrac{x}{p}\right)+(n-x)\log\left(\tfrac{n-x}{1-p}\right)-n\log n\end{aligned}$$

lo que $\Lambda^*(x)=x\log\left(\dfrac{x}{p}\right)+(n-x)\log\left(\dfrac{n-x}{1-p}\right)-n\log n \qquad x=0,\ldots,n.$

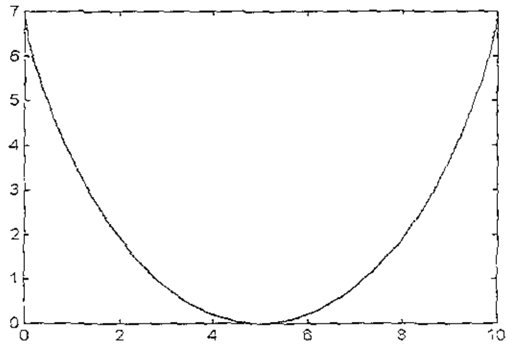

Figura 13. Función de tasa para una v.a. Binomial

Ejemplo 3.3.5. Ahora, si X es una v.a. Geométrica con parámetro (p).

La función generadora de momentos es

$$M(\lambda) = \frac{p}{1 - qe^{\lambda}}$$

por lo que $\log M(\lambda) = \log p - \log(1 - qe^{\lambda})$ y por lo tanto

$$g(\lambda) = \lambda x - \log p + \log(1 - qe^{\lambda}).$$

Obteniendo su derivada e igualándola a cero tenemos que

$$g'(\lambda) = x - \frac{qe^{\lambda}}{1 - qe^{\lambda}} = 0$$

de donde

$$\begin{array}{rcl} & x - qxe^{\lambda} & = qe^{\lambda} \\ \iff & x & = qe^{\lambda}(1 + x) \\ \iff & \dfrac{x}{q(1 + x)} & = e^{\lambda} \end{array}$$

así, se tiene que $\lambda^* = \log\left(\dfrac{x}{q(1+x)}\right)$ y si calculamos $g(\lambda^*)$ tenemos que

$$
\begin{aligned}
g(\lambda^*) &= x\log\left(\tfrac{x}{q(1+x)}\right) + \log\left(1 - q\left(\tfrac{x}{q(1+x)}\right)\right) - \log p \\
&= x\log\left(\tfrac{x}{q(1+x)}\right) + \log\left(1 - \tfrac{x}{1+x}\right) - \log p \\
&= x\log\left(\tfrac{x}{q}\right) + x\log\left(\tfrac{1}{1+x}\right) + \log\left(\tfrac{1}{1+x}\right) - \log p \\
&= x\log\left(\tfrac{x}{q}\right) - x\log(1+x) - \log(1+x) - \log p \\
&= x\log\left(\tfrac{x}{q}\right) - (1+x)\log(1+x) - \log p
\end{aligned}
$$

r lo que $\Lambda^*(x) = x\log\left(\dfrac{x}{q}\right) - (1+x)\log(1+x) - \log p \qquad x = 1, 2, \ldots$

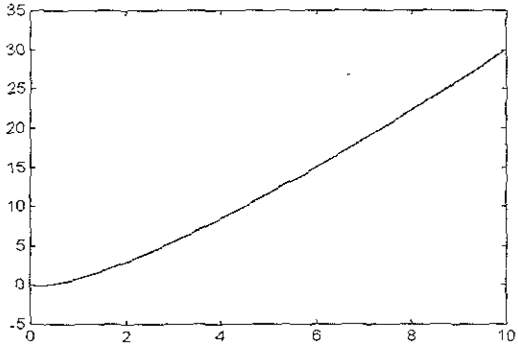

Figura 14. Función de tasa para una v.a. Geometrica

apítulo 4

l Teorema de Cramér $\mathbb{R}^d$ Dimensional.

.1 Tensión exponencial.

omenzamos por un lema que será fundamental en todo el capítulo.

ema 4.1.1. *Sea $N \in N^+$ y $(a_i^\varepsilon)_{1\leq i\leq N}$ una familia de reales estrictamente po-
tivos. Tenemos:*

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log\left(\sum_{i=1}^{N} a_i^\varepsilon\right) = \sup_{1\leq i\leq N}\left\{\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log a_i^\varepsilon\right\}$$

Demostración.

La minoración es evidente, pues una suma de números positivos es siempre ás grande que cada término de la suma, por lo que queda solamente demostrar mayoración. Empezamos por escribir que

$$\log\left(\sum_{i=1}^{N} a_i^\varepsilon\right) \leq \log\left(N \sup_i a_i^\varepsilon\right) = \log N + \log\left(\sup_i a_i^\varepsilon\right)$$

que nos dá

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log\left(\sum_{i=1}^{N} a_i^\varepsilon\right) \leq \overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log\left(\sup_i a_i^\varepsilon\right),$$

sta con probar que,

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log\left(\sup_i a_i^\varepsilon\right) = \sup_i\left\{\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log a_i^\varepsilon\right\} \tag{4.1.1}$$

Probaremos la desigualdad anterior para N=2.

Sean entonces a_1^ε y a_2^ε estrictamente positivos. La minoración es inmedia debido a que

$$\begin{aligned}
\sup_{1\le i\le 2}\{a_i^\varepsilon\} &\ge a_i^\varepsilon && \text{para toda} \quad i=1,2\\
\overline{\lim_{\varepsilon\to\infty}}\,\varepsilon\log\sup_{1\le i\le 2}\{a_i^\varepsilon\} &\ge \overline{\lim_{n\to\infty}}\,\varepsilon\log a_i^\varepsilon && \text{para toda} \quad i=1,2\\
\overline{\lim_{\varepsilon\to\infty}}\,\varepsilon\log\sup_{1\le i\le 2}\{a_i^\varepsilon\} &\ge \sup\{\overline{\lim_{n\to\infty}}\,\varepsilon\log a_i^\varepsilon\}
\end{aligned}$$

Para la mayoración suponemos, por ejemplo, que

$$\overline{\lim_{\varepsilon\to 0}}\,\varepsilon\log(a_1^\varepsilon) \le \overline{\lim_{\varepsilon\to 0}}\,\varepsilon\log(a_2^\varepsilon) = S$$

y por la definición de límite superior, para a_1 y a_2 tenemos que para toda $\delta >$ existe $\varepsilon_0 > 0$ tal que

$$0<\varepsilon<\varepsilon_0 \Longrightarrow \varepsilon\log a^\varepsilon \le S+\delta,$$

así mismo, para toda $\delta > 0$ existe $\varepsilon_1 > 0$ tal que

$$0<\varepsilon<\varepsilon_1 \Longrightarrow \varepsilon\log b^\varepsilon \le S+\delta$$

y como el máximo es uno de estos dos, tenemos claramente el resultado final

Vamos ahora a dar una definición que será fundamental en lo que resta.

Definición 4.1.1. Sea F un σ-álgebra sobre (X,Γ), espacio topológico, q contiene al σ-álgebra de los borelianos de X. A la familia $\{\mu_\varepsilon\}, \varepsilon > 0$ de prob bilidades sobre (X,F) se le llama **exponencialmente tensa** si para todo $M >$ existe un compacto K_M tal que

$$\overline{\lim_{\varepsilon\to 0}}\,\varepsilon\log\mu_\varepsilon(K_M^c) \le -M.$$

Recordemos, por la definición 1.4.2, que una función $I : X \to [0,+\infty]$ s.c.i. una función de tasa buena, si para todo $a > 0$ el conjunto de nivel $\{x : I(x) \le$ es compacto.

Definiremos ahora la noción de PGD débil.

inición 4.1.2. Sea F un σ-álgebra sobre X que contiene al σ-álgebra de los elianos de X. La familia $\{\mu_\varepsilon\}, \varepsilon > 0$ de probabilidades sobre (X, F) satisface PGD débil con función de tasa I si I es una función de X en $[0, +\infty]$ s.c.i. cumple:

$$\begin{array}{ll} \text{Para todo } O \text{ abierto} & \underline{\lim}_{\varepsilon\to 0} \varepsilon \log \mu^\varepsilon(O) \geq -\inf\{I(x) : x \in O\}. \\ \text{Para todo } K \text{ compacto} & \overline{\lim}_{\varepsilon\to 0} \varepsilon \log \mu^\varepsilon(K) \leq -\inf\{I(x) : x \in K\}. \end{array}$$

Es importante hacer notar que la familia de probabilidades puede satisfacer PGD débil sin satisfacer un PGD fuerte. Basta considerar $\mu_\varepsilon = \delta_{1/\varepsilon}$, que sface un PGD débil con $I \equiv +\infty$, y no satisface un PGD fuerte, pues $\mu_\varepsilon(\mathbb{R}) =$ Disponemos de la proposición siguiente.

oposición 4.1.1. *Sea F un σ-álgebra de X que contenga al σ-álgebra de los elianos de X, y $\{\mu_\varepsilon\}, \varepsilon > 0$ una familia exponencialmente tensa de probabildes sobre (X, F), entonces:*

. *Si tenemos que para todo K compacto $\overline{\lim}_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) \geq -\inf\{I(x) : x \in K\}$. Entonces esta mayoración sigue siendo cierta sobre los cerrados.*

. *Si tenemos que para todo O abierto $\underline{\lim}_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(O) \geq -\inf\{J(x) : x \in O\}$. Entonces J es una función de tasa buena.*

Demostración.

. Sea F cerrado. El primer caso evidente es cuando $\inf\{I(x) : x \in F\} = 0$. Si no estamos en este caso, pongamos $M = \inf\{I(x) : x \in F\}$. Como la familia de medidas es exponencialmente tensa, existe entonces un compacto K_M tal que

$$\overline{\lim}_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_M^c) \leq -M.$$

Hacemos entonces la separación siguiente

$$\begin{aligned} \mu_\varepsilon(F) &= \mu_\varepsilon(F \cap \Omega) \\ &= \mu_\varepsilon(f \cap (K_M \cup K_M^c)) \\ &= \mu_\varepsilon(F \cap K_M) + \mu_\varepsilon(F \cap K_M^c) \\ &\leq \mu_\varepsilon(F \cap K_M) + \mu_\varepsilon(K_M^c) \end{aligned}$$

De acuerdo con la ecuación 4.3.1 tenemos que

$$\begin{aligned}\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(F) &\leq \overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log(\mu_\varepsilon(F\cap K_M) + \mu_\varepsilon(K_M^c)) \\ &= \sup\left\{\overline{\lim_{\varepsilon\to 0}}(\varepsilon \log \mu_\varepsilon(F\cap K_M) + \varepsilon\log\mu_\varepsilon(K_M^c))\right\}\end{aligned}$$

por lo que

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(F) \leq \sup\{\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(F\cap K_M) + \overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(K_M^c)\}$$

y como $F\cap K_M$ es compacto, podemos aplicarle la mayoración, queda

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(F) \leq \sup\{-\inf\{I(x) : x\in F\cap K_M\}; -M\}.$$

Como $-\inf\{I(x) : x\in F\cap K_M\} \leq -\inf\{I(x) : x\in F\} = -M$ tenemo resultado.

2. Aquí sólo tenemos que verificar que $I(x)$ es una función de tasa bue Sea $a > 0$, queremos demostrar que el conjunto de nivel $\{x : I(x) \leq$ es compacto. Comencemos por decir que, ya que la familia de medida exponencialmente tensa, existe un compacto K_a tal que

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(K_a^c) \leq -a.$$

Como, por otra parte, K_a^c es un abierto, le aplicamos la minoración

$$-a \geq \underline{\lim_{\varepsilon\to 0}}\, \varepsilon \log \mu_\varepsilon(K_a^c) \geq -\inf\{I(x) : x\in K_a^c\},$$

reuniendo los dos resultados obtenemos

$$\inf\{I(x) : x\in K_a^c\} > a.$$

Lo que produce que $\{x : I(x)\leq a\}\subset K_a$, entonces, tenemos un conju cerrado contenido en un compacto, lo que nos dice que es un compacto que termina la prueba.

4.2 Transformada de Cramér en $\mathbb{R}^d$.

Definición 4.2.1. Sea X una v.a. con valores en $\mathbb{R}^d$ de ley μ. La func log-Laplace de μ es la función de $\mathbb{R}^d$ en $(-\infty,\infty]$, definida como

$$\Lambda(\lambda) = \log\mathbb{E}[e^{\langle\lambda,X\rangle}]$$

y $\mathcal{D}(\lambda) = \{\lambda\in\mathbb{R}^d : \Lambda(\lambda) < \infty\}.$

En lo que resta del capítulo, supondremos que 0 está en el interior de $\mathcal{D}(\lambda)$, ue, al pasar por las marginales implica en particular que $\mathbb{E}(X)$ está definida no vector de $\mathbb{R}^d$. Tenemos la siguiente proposición, análoga a la proposición 1.

oposición 4.2.1. *Λ es convexa, diferenciable en el interior de $\mathcal{D}(\lambda)$ y tenos que*

$$grad_K\Lambda(\lambda) = \frac{\mathbb{E}[X^K \cdot e^{\langle\lambda,X\rangle}]}{\mathbb{E}[e^{\langle\lambda,X\rangle}]}$$

particular, $grad_K\Lambda(0) = \mathbb{E}(X)$.

Demostración.

$$\begin{aligned}
\Lambda(t\lambda_1 + (1-t)\lambda_2) &= \log[\mathbb{E}(e^{\langle(t\lambda_1+(1-t)\lambda_2),X\rangle})] \\
&= \log[\mathbb{E}(e^{t\langle\lambda_1,X\rangle}\, e^{(1-t)\langle\lambda_2,X\rangle})] \\
&= \log[\mathbb{E}((e^{\langle\lambda_1,X\rangle})^t \cdot (e^{\langle\lambda_2,X\rangle})^{1-t})] \\
&= \log[\mathbb{E}((e^{\langle\lambda_1,X\rangle})^{\frac{1}{1/t}} \cdot (e^{\langle\lambda_2,X\rangle})^{\frac{1}{1/(1-t)}})] \\
&\leq \log([\mathbb{E}(e^{\langle\lambda_1,X\rangle})]^t \cdot [\mathbb{E}(e^{\langle\lambda_2,X\rangle})^{1-t}]) \qquad \text{(por Hölder)} \\
&= \log[(\mathbb{E}(e^{\langle\lambda_1,X\rangle}))^t] + \log[(\mathbb{E}(e^{\langle\lambda_2,X\rangle}))^{1-t}] \\
&= t\log[\mathbb{E}(e^{\langle\lambda_1,X\rangle})] + (1-t)\log[\mathbb{E}(e^{\langle\lambda_2,X\rangle})].
\end{aligned}$$

finición 4.2.2. La transformada de Cramér de la ley μ es la función

$$\Lambda^*(x) = \sup\{\langle\lambda, x\rangle - \Lambda(\lambda) : \lambda \in \mathbb{R}^d\}.$$

Tenemos las siguientes propiedades, análogas a las de la proposición 3.1.2.

oposición 4.2.2. *Propiedades de la transformada de Cramér en $\mathbb{R}^d$:*

. *$\Lambda^*(x)$ es convexa, positiva, s.c.i. y aún más, es una función de tasa buena.*

. *Si $y = grad\Lambda(\eta)$, entonces,*

$$\Lambda^*(y) = \langle\eta, y\rangle - \Lambda(\eta).$$

Demostración.

1. Para verificar la convexidad y la s.c.i. se usa el mismo argumento para el caso univariado. Sea un conjunto de nivel $\{x : \Lambda^*(x) \leq a\}$. Co sabemos que se trata de un cerrado, basta mostrar que está acotado y afirmamos que $\Lambda^*(x)$ es una función de tasa buena. Hemos supuesto qu está en el interior de $\mathcal{D}(\Lambda)$, entonces $\mathcal{D}(\Lambda)$ contiene en particular una es $S(0,r)$ para un $r > 0$ contenida totalmente en su interior. Sea entor $x \in \mathbb{R}^d$ perteneciente al conjunto de nivel $\{\Lambda^* \leq a\}$, entonces, como Λ^* es un supremo, es mayor o igual que $\lambda x - \Lambda(\lambda)$, en particular si hacer $\lambda = rx/||x||$ tenemos

$$\begin{aligned} a \geq \Lambda^*(x) &\geq < r\tfrac{x}{||x||}, x > -\Lambda\left(r\tfrac{x}{||x||}\right) \\ &= \tfrac{r}{||x||} < x, x > -\Lambda\left(r\tfrac{x}{||x||}\right) \\ &= \tfrac{r}{||x||}||x||^2 - \Lambda\left(r\tfrac{x}{||x||}\right) \\ &= r||x|| - \Lambda\left(r\tfrac{x}{||x||}\right) \\ &\geq r||x|| - \sup\{\Lambda(\lambda) : \lambda \in S(0;r)\} \end{aligned}$$

(pues $\Lambda^*(x)$ es un supremo y $r\frac{x}{||x||} \in (0:r))$. De donde

$$||x|| \leq \frac{1}{r}(a + \sup\{\Lambda(\lambda) : \lambda \in S(0;r)\}).$$

Ahora, como Λ es diferenciable en el interior de su dominio, en particu es continuo sobre $S(0,r)$, entonces el supremo del miembro de la dere es finito, lo que dá el resultado.

2. Sea $y = grad(\eta)$ y sea $\lambda \in \mathbb{R}^d$. El principio va a ser demostrar que tener

$$\langle \lambda, y\rangle - \Lambda(\lambda) \geq \langle \eta, y\rangle - \Lambda(\eta)$$

para esto, dèfinimos una función g de $[0,1]$ en $\mathbb{R}$ de la manera siguiente $\alpha \in \mathbb{R}^d$,

$$g(\alpha) = \langle \eta, y\rangle + \alpha\langle \lambda - \eta, y\rangle - \Lambda(\eta + \alpha(\lambda - \eta)),$$

g es la suma de una función afín y de una función cóncava, por lo que cóncava. Por otra parte $g(0) = \langle \eta, y\rangle - \Lambda(\eta)$ y $g(1) = \langle \lambda, y\rangle - \Lambda(\lambda)$. El es entonces demostrar que $g(1) \leq g(0)$. Como g es cóncava, tenemos:

$$g(1) - g(0) \leq g'(0) \cdot (1 - 0)$$

y

$$g'(0) = \langle \lambda - \eta, y \rangle - \langle \lambda - \eta, grad \Lambda(\eta) \rangle = 0$$

por lo tanto $\big(\langle \lambda, y \rangle - \Lambda(\lambda)\big) - \big(\langle \eta, y \rangle - \Lambda(\eta)\big) \leq 0$ lo que nos dá el resultado.

3 El Teorema de Cramér Multidimensional.

orema 4.1. *Sea $(X_n)_n$ una sucesión de v.a. con valores en $\mathbb{R}^d$ independi-tes y equidistribuídos de la ley μ tal que 0 esté en el interior de $\mathcal{D}(\Lambda)$. En-nces la sucesión μ_n, la distribución de probabilidad de las medias empíricas $_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ sigue un PGD con función de tasa buena, la función Λ^*, ansformada de Cramér en $\mathbb{R}^d$.*

Demostración.

La prueba se hará en tres etapas: tensión exponencial, ,mayoración sobre los mpactos y minoración sobre los abiertos.

1. **Tensión Exponencial.**

Como primer paso, demostraremos que la familia $\{\mu_n\}$ de probabilidades que corresponden a la distribución de la media empírica $\overline{X_n}$, es una familia exponencialmente tensa y luego la idea será utilizar el teorema de Cramér en $\mathbb{R}$. Sea $M > 0$, vamos a mostrar que

$$\lim_{n \to +\infty} \left[\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mu_n \left([-M, M]^d \right)^c \right] = -\infty. \qquad 4.3.1$$

Pongamos $K^M = [-M, M]^d$. Razonemos sobre las leyes marginales

$$\mu_n(K_M^c) \leq \sum_{i=1}^{d} \{\mu_n^i[M, +\infty) + \mu_n^i(-\infty, -M]\}.$$

Donde μ_n^j es la j-ésima marginal de μ_n, es decir, la ley de $\overline{X_n^i} = \frac{1}{n}\sum_{k=1}^{n} X_k^i$. Según el teorema 3.1, μ_n^i satisface un PGD de función de tasa Λ_i^*, la transformada de Cramér de X_1^i. Como 0 está en el interior del dominio de μ, es

claro que 0 está a su vez en el interior del dominio de Λ_i para todo i.
ahí que, por una parte

$$\lim_{|y|\to+\infty} \Lambda^\star(y) = +\infty$$

y, por otra parte, según 3.2.1

$$\mu_n^i[M, +\infty) \leq 2\exp[-n \inf\{\Lambda^\star(x) : x \geq M\}]$$

y una desigualdad análoga del otro lado, lo que dá finalmente el resulta
cuando hacemos tender M hacia el infinito.

2. **Mayoración sobre los compactos.**

Comenzaremos haciendo mayoraciones sobre las bolas. Sea $\delta > 0$, pon
mos

$$I^\delta(x) = \inf\left\{\Lambda^\star(x) - \delta, \frac{1}{\delta}\right\}$$

Notemos que si $\Lambda^\star(X) < \infty$, entonces, para δ suficientemente peque
$I^\delta(x) = \Lambda^\star(x) - \delta$. Si, por el contrario, $\Lambda^\star(x)$ es infinito, el ínfimo es 1
Además, en ambos casos, $I^\delta(x)$ crece hacia $\Lambda^\star(x)$ cuando δ decrece ha
0. Sea entonces K un compacto de $\mathbb{R}^d$ y $X \in K$. Tenemos $I^\delta(x) < \Lambda^\star($
entonces, como éste último es un supremo, tenemos que existe un vec
$\lambda_x \in \mathbb{R}^d$ tal que

$$\langle\lambda_x, x\rangle - \Lambda(\lambda_x) \geq I^\delta(x). \qquad 4.$$

Por otra parte, existe $\rho_x > 0$ tal que $\rho_x \cdot \|\lambda_x\| \leq \delta$. Hacemos $B_x = B(x, \rho$
Entonces, si $\overline{X_n} \in B_x$

$$\langle\lambda_x, \overline{X_n}\rangle \leq \inf\{\langle\lambda_x, y\rangle : y \in B_x\}$$

y, por la desigualdad de Schwartz

$$|\langle\lambda_x, y - x\rangle| \leq \|\lambda_x\| \cdot \|y - x\| \leq \|\lambda_x\| \cdot \rho_x$$

cambiando y por $\overline{X}_n$ tenemos

$$\langle\lambda_x, \overline{X_n}\rangle \geq \langle\lambda_x, x\rangle - \rho_x \cdot \|\lambda_x\| \geq \langle\lambda_x, x\rangle - \delta$$

según 3.2.7

$$\begin{aligned} \mu_n(B_x) &\leq \mathbb{E}[\exp\{n(\langle\lambda_x, \overline{X_n}\rangle - \langle\lambda_x, x\rangle + \delta)\}] \\ &= \exp(-n\langle\lambda_x, x\rangle + n\delta) \cdot (\mathbb{E}[e^{\langle\lambda_x, X_1\rangle}])^n \\ &= \exp[n(\Lambda(\lambda_x) - \langle\lambda_x, x\rangle + \delta)]. \end{aligned}$$

De ahí, por 4.3.2

$$\frac{1}{n}\log\mu_n(B_x) \leq \Lambda(\lambda_x) - \langle\lambda_x, x\rangle + \delta \leq \delta - I^\delta(x).$$

Ahora, como K es compacto, existen puntos x_k, en un número finito N tales que $K \subset \bigcup_{k=1}^{N} B_{x_k}$ lo que nos dá

$$\frac{1}{n}\log\mu_n(K) \leq \frac{1}{n}\log\left(\sum_{k=1}^{N}\mu_n(B_{x_k})\right).$$

Según el lema 4.1.1

$$\begin{aligned} \overline{\lim_{n\to\infty}} \tfrac{1}{n}\log\mu_n(K) &\leq \sup_{1\leq k\leq N}\left\{\overline{\lim_{n\to\infty}} \tfrac{1}{n}\log\mu_n(B_{x_k})\right\} \\ &\leq \sup_{1\leq k\leq N}\{-I^\delta(x_k) + \delta\} \\ &= -\inf_{1\leq k\leq N}\{-I^\delta(x_k) + \delta\} \\ &\leq -\inf_{x\in K}\{I^\delta(x) + \delta\}. \end{aligned}$$

Para concluír basta con hacer tender δ hacia 0.

3. Minoración sobre los abiertos.

Supongamos que $\mathcal{D}(\Lambda) = \mathbb{R}^d$. Análogo a $\mathbb{R}$, mostraremos el siguiente resultado. Para todo $x \in \mathbb{R}^d$ y para todo $\delta > 0$

$$\underline{\lim_{n\to\infty}} \frac{1}{n}\log\mu_n(B(x,\delta)) \geq -\Lambda^\star(x). \qquad 4.3.3$$

Tenemos dos casos: existe η tal que $x = grad\,\Lambda(\eta)$. En este caso la demostración de 4.3.3 es similar a la del caso real. Si no estamos en el caso anterior, el método será perturbar las X_k con Gaussianas $N(0, \varepsilon^2 \cdot Id)$.

Supongamos entonces que para todo η, X no es igual a $grad\Lambda(\eta)$, y hag
$Y_n = X_n + \varepsilon Z_n$ donde las Z_n son v.a.i.i.d. con distribución $N(0, Id)$ si
entonces la log–Laplace de las Y_n,

$$\begin{aligned}\Lambda_Y(\lambda) &= \log \mathbb{E}[\exp(\langle\lambda, X_1 + \varepsilon Z_1\rangle)] \\ &= \log \mathbb{E}[\exp(\langle\lambda.X_1\rangle)\exp(\langle\varepsilon\lambda, Z_1\rangle)] \\ &= \Lambda(\lambda) + \Lambda_{Z_1}(\varepsilon\lambda).\end{aligned}$$

Además Z_1 es un vector Normal centrado de componentes independie
por lo que su log-Laplace es obtenida como una suma de log-Laplaces d
componentes, y la log-Laplace de una $N(0,1)$ fué calculada anteriorm
Entonces

$$\Lambda_{Z_1}(\varepsilon\lambda) = \sum_{j=1}^{d} \Lambda_{Z_1}^{j}(\varepsilon\lambda_j) = \sum_{j=1}^{d} \frac{\varepsilon^2}{2}\lambda_j^2$$

de ahí que

$$\Lambda_Y(\lambda) = \Lambda(\lambda) + \frac{\varepsilon}{2}\|\lambda\|^2 \geq \Lambda(\lambda),$$

entonces

$$\Lambda_Y^{\star}(x) = \sup\{\langle\lambda, x\rangle - \Lambda_Y(\lambda); \lambda \in \mathbb{R}^d\} \leq \Lambda^{\star}(x).$$

Mostraremos que para Y_n existe η tal que $X = grad\Lambda(\eta)$. Sea $g($
$\langle\lambda, x\rangle - \Lambda_Y(\lambda)$, ésta función es cóncava (ya que $\Lambda_Y(\lambda)$ es convexa
mostramos que $g(\lambda) \to -\infty$ cuando $\|\lambda\| \to \infty$, esto prueba que g pasa
un máximo, lo que garantiza la existencia de un η tal que grad $g(\eta)$
Sea $x = grad\Lambda_Y(\eta)$, de aquí que

$$g(\lambda) = \langle\lambda, x\rangle - \Lambda(\lambda) - \frac{\varepsilon}{2}\|\lambda\|^2 \leq \Lambda^{\star}(x) - \frac{\varepsilon}{2}\|\lambda\|^2$$

entonces

$$\lim_{\|\lambda\|\to\infty} g(\lambda) = C - \frac{\varepsilon}{2}\lim_{\|\lambda\|\to\infty}\|\lambda\|$$

que tiende a $-\infty$.

Podemos ahora aplicar lo que se demostró en 4.3.3, es decir, para tod
δ.

$$\underline{\lim}_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\overline{Y}_n \in B(x,\delta)) \geq -\Lambda_Y^{\star}(x) \geq -\Lambda^{\star}(x)$$

y la última desigualdad se sigue por 4.3.4.

Vamos ahora a mayorar la probabilidad que aparece aquí arriba.

$$\begin{aligned}\{\|\overline{Y_n}-x\|<\delta\} &\subset \{\|\overline{Y_n}-x\|<\delta\}\cap\{\|\varepsilon\overline{Z_n}\|\geq\delta/2\}\\ &\subset \{\|\overline{Y_n}-x\|<\delta\}\cap\{\|\varepsilon\overline{Z_n}\|\geq\delta/2\}\cap\Omega\\ &\subset \{\|\overline{Y_n}-x\|<\delta\}\cap\{\|\varepsilon\overline{Z_n}\|\geq\delta/2\}\cap\{\|\varepsilon\overline{Z}_n\|<\delta/2\}\\ &\qquad\cup\{\varepsilon\|\overline{Z}_n\|\geq\delta/2\}\\ &\subset \{\|\overline{Y_n}-x\|<\delta\}\cap\{\|\varepsilon\overline{Z_n}\|<\delta/2\}\cup\{\|\varepsilon\overline{Z}_n\geq\delta/2\}\end{aligned}$$

y

$$\begin{aligned}\{\|\overline{Y}_n-x\|<\delta\}\cap\{\|\varepsilon\overline{Z}_n\|<\delta/2\} &\subset \{\|\overline{X}_n-x\|+\|\varepsilon\overline{Z}_n\|\leq 2\delta\}\\ &\subset \{\|\overline{X}_n-x\|\leq 3\delta/2\}\end{aligned}$$

y $\overline{Z_n}$ se distribuye con la misma ley que $\frac{Z_1}{\sqrt{n}}$, así,

$$\begin{aligned}\varlimsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\|\varepsilon\overline{Z_n}\|\geq\delta/2) &= \varlimsup_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}\left(\varepsilon^2\tfrac{\|Z_1\|^2}{\sqrt{n}}\geq\left(\tfrac{\delta}{2}\right)^2\right)\\ &\leq \varlimsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\exists j\leq d,\ \left|\tfrac{1}{\sqrt{n}}Z_1^j\right|^2\geq\frac{\delta^2}{4\varepsilon^2 d}\right)\\ &\leq \varlimsup_{n\to\infty}\frac{1}{n}\log\left[\sum_{j=1}^{d}\mathbb{P}\left(\left|\tfrac{1}{\sqrt{n}}Z_1^j\right|^2\geq\frac{\delta^2}{4\varepsilon^2 d}\right)\right]\\ &= \varlimsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left|\tfrac{1}{\sqrt{n}}N(0,1)\right|\geq\frac{\delta}{2\varepsilon\sqrt{d}}\right)=-\frac{\delta^2}{8\varepsilon^2 d}\end{aligned}$$

lo que implica en particular que para toda $M>0$ existe ε_0 tal que para todo $\varepsilon<\varepsilon_0$,

$$\varlimsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\|\varepsilon\overline{Z_n}\|\geq\delta/2)\leq -M \qquad 4.3.6$$

por lo que

$$\mathbb{P}(\|\varepsilon\overline{Z}_n\|\geq\delta/2)\leq\exp(-nM)$$

entonces

$$\mathbb{P}(\|\overline{X_n}-x\|<3\delta/2)\geq\mathbb{P}(\|\overline{Y_n}-x\|<\delta)-\mathbb{P}(\|\varepsilon\overline{Z_n}\|\geq\delta).$$

Sea α constante y M tal que $M\leq\Lambda^*(x)+\alpha$. Como $\Lambda^*(x)$ es un límite inferior, y por 4.3.5 se tiene que existe n_0 tal que para toda $n\geq n_0$

$$\frac{1}{n}\log\mathbb{P}(\|\overline{Y_n}-x\|<\delta)\geq-\Lambda^*(x)-\alpha$$

por lo que

$$\mathbb{P}(\|\overline{Y_n} - x\| < \delta) \geq \exp[-n(\Lambda^\star(x) + \alpha)]$$

y sustituyendo en 4.3.6

$$\mathbb{P}(\|\overline{X_n} - x\| < 3\delta/2) \geq \exp[-n(\Lambda^\star(x) + \alpha)] - \exp(-nM)$$

como $M \geq \Lambda^\star(x) + \alpha$ entonces $-\exp[-nM] \geq -\exp[-n(\Lambda^\star(x) + \alpha)]$, lo que

$$\mathbb{P}(\|\overline{X}_n - x\| < \frac{3\delta}{2}) \geq 2\exp[-n(\Lambda^\star(x) + \alpha)].$$

Como resultado de este teorema se tiene el siguiente corolario.

Corolario 4.3.1. *Si A es un abierto convexo y si $\mathcal{D}(\Lambda) = \mathbb{R}^d$ entonces*

$$\lim_{n\to+\infty} \frac{1}{n} \log \mu_n(A) = -\inf\{\Lambda^\star(x) : x \in A\}.$$

Capítulo 5

PGD en los casos dependientes.

En los capítulos anteriores, nos interesamos exclusivamente en el caso de v.a. independientes. Sin embargo, numerosas situaciones relativamente naturales en teoría de probabilidades hacen intervenir v.a. que no son independientes (basta considerar las cadenas de Markov). Entonces, si queremos estudiar dichas situaciones bajo el ángulo de las **grandes desviaciones**, es necesario encontrar nuevas herramientas para extender las nociones que acabamos de ver. Esa es la finalidad de este capítulo.

5.1 Descripción del teorema.

Sea $(Z_n)_n$ una sucesión de v.a., $Z_n : \Omega \to \mathbb{R}^d$, y sea Λ_n la sucesión de funciones log-Laplace asociadas. En lo siguiente, Z_n jugará el mismo papel que $\overline{X}_n$ en los capítulos anteriores, conviene entonces observar qué propiedades, implícitas o no, hemos utilizado hasta ahora. Sea Ξ la log-Laplace de X_1 y Ξ_n la log-Laplace de $\overline{X}_n$, entonces

$$\begin{aligned}
\Xi_n(\lambda) &= \log \mathbb{E}\left[\exp\left(< \lambda, \frac{1}{n}\sum_{j=1}^{n} X_j >\right)\right] \\
&= \log \mathbb{E}\left[\exp\left(\sum_{j=1}^{n} < \frac{\lambda}{n}, X_j >\right)\right] \\
&= n\Xi\left(\frac{\lambda}{n}\right)
\end{aligned}$$

por lo que

$$\frac{1}{n}\Xi_n(n\lambda) = \Xi(\lambda)$$

y es este cálculo el que motiva la hipótesis siguiente

$$(H.1) \qquad \text{para toda}\, \lambda \in \mathbb{R}^d,\ \frac{1}{n}\Lambda_n(n\lambda) \longrightarrow \Lambda(\lambda) \in [-\infty, +\infty]$$

donde Λ ya no designa necesariamente la log-Laplace de una medida.

La segunda hipótesis también resalta a la vista de lo que vimos en los capítul anteriores:

$$(H.2) \qquad 0 \in \operatorname{Int}(\mathcal{D}(\Lambda))$$

donde $\mathcal{D}(\Lambda)$ es el conjunto de vectores $\lambda \in \mathbb{R}^d$ tales que $\Lambda(\lambda) \in \mathbb{R}$ por lo q definimos

$$\Lambda^*(x) := \sup\{\langle \lambda, x\rangle - \Lambda(\lambda); \lambda \in \mathbb{R}^d\}.$$

Hacemos notar que, como $\Lambda_n(0) = 0$, la hipótesis (H.1) nos lleva a $\Lambda(0) = 0$, que implica en particular, que $\Lambda^*(x) \geq 0$.

La definición siguiente no aparecerá inmediatamente de forma evidente, s embargo, juega un papel muy importante para lo que sigue.

Definición 5.1.1. Un punto x de $\mathbb{R}^d$ es llamado **punto expuesto** de Λ^*, c η hiperplano exponente si

$$\text{para toda}\, y \neq x,\ \Lambda^*(y) > \Lambda^*(x) + < \eta, y - x > .$$

Llamaremos $\mathcal{E}$ al conjunto de puntos expuestos de Λ^* cuyo uno de los hipe planos exponentes está en $\operatorname{Int}(\mathcal{D}(\Lambda))$.

Teorema 5.1 (Gartner-Ellis.). *Si las hipótesis (H.1) y (H.2) se satisface entonces, si llamamos μ_n a la ley de Z_n:*

$$\begin{array}{lll} \textit{Para todo}\, F\ \textit{cerrado}, & \overline{\lim}_{n\to\infty} \frac{1}{n}\log \mu_n(F) \leq -\inf\{\Lambda^*(x); x \in F\} & y \\ \textit{para todo}\, O\ \textit{abierto}, & \underline{\lim}_{n\to\infty} \frac{1}{n}\log \mu_n(O) \geq -\inf\{\Lambda^*(x); x \in O \cap \mathcal{E}\} & \end{array}$$

Además, si $\mathcal{D}(\Lambda) = \mathbb{R}^d$ y si Λ es diferenciable sobre $\mathbb{R}^d$, entonces μ_n satisfa un PGD con función de tasa buena Λ^*.

Observación. Podemos deshechar la hipótesis de la última parte del teore si suponemos simplemente que Λ es **esencialmente regular**, es decir, *i)* Λ es diferenciable en el interior de su dominio, y 0 está en el interior.

ii) Si $\lambda_n \to \lambda$, con $\lambda_n \in (\mathcal{D}(\Lambda))^o$, entonces $|\mathrm{grad}\Lambda(\lambda_n)| \to +\infty$.

ıntes de demostrar el teorema 5.1, comenzaremos por dar algunas propiedades ıtiles de Λ y $\Lambda^\star$.

ı.2 Propiedades de Λ y $\Lambda^\star$.

ʾroposición 5.2.1. *Propiedades:*

(1) Λ *y* $\Lambda^\star$ *son funciónes convexas, y si* $0 \in (\mathcal{D}(\Lambda))^o$*, entonces* $\Lambda(\lambda) \neq -\infty$.

(2) Si $0 \in (\mathcal{D}(\Lambda))^o$*, entonces* $\Lambda^\star$ *es una función de tasa buena.*

(3) Si $y = \mathrm{grad}\Lambda(\eta)$*, entonces* $\Lambda^\star(y) =< \eta, y > -\Lambda(\eta)$*. Por otra parte,* $y \in \mathcal{E}$ *y* η *es uno de sus hiperplanos exponentes.*

Demostración.

(1) Λ es convexa, como límite de funciones convexas, por la hipótesis (H.1), $\Lambda^\star(x)$ es convexa, por ser el supremo de funciones afines. Para la segunda parte negaremos la proposición, entonces suponemos que existe $\lambda \in \mathbb{R}^d$ tal que $\Lambda(\lambda) = -\infty$, y sea $\alpha \in (0, 1)$, como Λ es convexa, tenemos

$$\Lambda(\alpha\lambda) = \Lambda(\alpha\lambda + (1-\alpha)0) \leq \alpha\Lambda(\lambda) + (1-\alpha)\Lambda(0) = \alpha\Lambda(\lambda) = -\infty,$$

entonces el segmento $[0, \lambda]$ no está en $\mathcal{D}(\Lambda)$, lo que prueba que $0 \notin (\mathcal{D}(\Lambda))^o$ (pues entonces podríamos encontrar una bola centrada en 0 y contenida totalmente en $(\mathcal{D}(\Lambda))^o$, bola que intersectaría forzosamente este segmento).

(2) $\Lambda^\star(x)$ es s.c.i. por ser el supremo de funciones continuas (por ser afines). entonces, los conjuntos de nivel $\{\Lambda^\star(x) \leq a\}$ son cerrados, sólo basta demostrar que están acotados, lo que ya hicimos en la proposición 3.1.1 (1) (Notando que Λ es continua en el interior de su dominio, por ser convexa).

(3) Aplicando el método de la proposición 3.1.1 (2), obtenemos fácilmente que $\Lambda^\star(y) = \langle y, \eta \rangle - \Lambda(\eta)$. Lo que es menos evidente es la exposición de y. Sea entonces $x \in \mathbb{R}^d$ tal que $\Lambda^\star(x) \leq \Lambda^\star(y) + \langle \eta, x - y \rangle$, vamos a mostrar que

$x = y$, lo que dará el resultado pasando a la contrapropuesta. Para hacerl vamos a mostrar que $\langle y - x, z\rangle = 0$ para todo $z \in \mathbb{R}^d$. Sea entonces $z \in \mathbb{R}$

$$\begin{aligned}\langle z+\eta, x - \Lambda(z+\eta)\rangle &\leq \Lambda^*(x) \leq \underbrace{\Lambda^*(y)}_{=<y,\eta>-\Lambda(\eta)} + \langle \eta, x - y\rangle \\ &\leq \langle \eta, x\rangle - \Lambda(\eta)\end{aligned}$$

de ahí

$$\langle z, x\rangle - \Lambda(z+\eta) \leq -\Lambda(\eta)$$

es decir

$$\langle z, x\rangle \leq \Lambda(z+\eta) - \Lambda(\eta)$$

de donde para todo $z \in |mathbbR^d$

$$\langle z, x\rangle \leq \frac{\Lambda(\varepsilon z+\eta) - \Lambda(\eta)}{\varepsilon}$$

haciendo tender $\varepsilon \to 0$, encontramos $\forall z, \quad \langle z, x\rangle \leq \langle \underbrace{\mathrm{grad}\Lambda(\eta)}_{=y}, z\rangle$, y car biando z por $-z$ el resultado está probado.

5.3 Teorema de Gartner-Ellis.

Demostración.

1. **Tensión exponencial.**

   Sea $B = \{e_1, \dots, e_d\}$ la base canónica de $\mathbb{R}^d$. Como 0 está en el interior $\mathcal{D}(\Lambda)$, existe, para todo j un α_j tal que $-\alpha_j e_j$ y $\alpha_j e_j$ estén en el interior $\mathcal{D}(\Lambda)$. Sea μ_n^j la j-ésima marginal de μ_n. Utilizando 3.2.7 y el método cálculo de la mayoración en el teorema 3.2, tenemos

$$\begin{aligned}\mu_n^j([t,+\infty)) &\leq \mathbb{E}(\exp[n\alpha_j(\langle e_j, \sum \tfrac{\Xi_i}{n}\rangle - t)]) \\ &= \mathbb{E}(\exp[\langle n\alpha_j e_j, \tfrac{1}{n}\sum \Xi_i\rangle - n\alpha_j t]) \\ &= \exp(-n\alpha_j t)\,\mathbb{E}\left(\exp\langle n\alpha_j e_j, \tfrac{1}{n}\sum \Xi_i\rangle\right) \\ &= \exp(-n\alpha_j t - \Lambda_n(n\alpha_j e_j)) \\ &= \exp(-n(\alpha_j t - \tfrac{\Lambda_n(n\alpha_j e_j)}{n})) \\ &\longrightarrow \exp[-n(\alpha_j t - \Lambda_n(\alpha_j e_j))]\end{aligned}$$

y una desigualdad análoga del otro lado. Terminamos entonces como en la primera parte del Teorema 4.1.

2. **Mayoración sobre los compactos.**

Aplicaremos el mismo razonamiento que en el teorema 4.1. Sea $\delta > 0$, y hagamos:

$$I^\delta(x) = \inf\left\{\Lambda^*(x) - \delta; \frac{1}{\delta}\right\} < \Lambda^*(x).$$

Si fijamos x, existe $\lambda_x \in \mathbb{R}^d$ tal que

$$\langle\lambda_x, x\rangle - \Lambda(\lambda_x) \geq I^\delta(x).$$

Sea $\rho_x > 0$ tal que $\rho_x\|\lambda_x\| < \delta$. Si hacemos $B_x = B(x; \rho_x)$, y procedemos como en la parte B del teorema 3.1, tenemos que

$$\begin{aligned}
\mu_n(B_x) &\leq \mathbb{E}(exp(n(\langle\lambda_x, \Xi_n\rangle - \langle\lambda_x, x\rangle + \delta))) \\
&= \mathbb{E}(\exp(\langle n\lambda_x, \Xi_n\rangle - \langle n\lambda_x, x\rangle + n\delta)) \\
&= \exp(n\delta - n\langle\lambda_x, x\rangle)\mathbb{E}(\exp(\langle n\lambda_x, \Xi_n\rangle) \\
&= \exp(n\delta - n\langle\lambda_x, x\rangle) \exp(\Lambda_n(n\lambda_x)) \\
&\leq \exp[\Lambda_n(\lambda_x) - n\langle\lambda_x, x\rangle + n\delta]
\end{aligned}$$

entonces

$$\frac{1}{n}\log\mu_n(B_x) \leq \frac{1}{n}\Lambda_n(n\lambda_x) - \langle\lambda_x, x\rangle + \delta$$

y como $\frac{1}{n}\Lambda_n(n\lambda_x) \to \Lambda(\lambda_x)$, tenemos, para n suficientemente grande que

$$\frac{1}{n}\log\mu_n(B_x) \leq \Lambda(\lambda_x) - \langle\lambda_x, x\rangle + 2\delta \leq -I^\delta(x) + 2\delta$$

con lo que el final de la prueba es el mismo que la segunda parte del teorema 4.1.

3. **Minoración sobre los abiertos.**

Sea $y \in \mathcal{E}$ de hiperplano expuesto $\eta \in \mathcal{D}(\Lambda(\lambda))$, vamos a mostrar que

$$\text{para toda}\,\delta > 0, \quad \frac{1}{n}\log\mu_n(B(y,\delta)) \geq -\Lambda^*(y). \qquad 5.1.1$$

Supongamos 5.1.1 demostrado, entonces, si O es un abierto de $\mathbb{R}^d$, te
dremos que para toda $y \in \mathbb{R} \cap \mathcal{E}$ existe $\delta > 0$ tal que

$$B(y,\delta) \subset O$$

y entonces

$$\varliminf_{n\to\infty} \frac{1}{n}\mu_n(O) \geq \varliminf_{n\to\infty} \frac{1}{n}\mu_n(B(y,\delta)) \geq -\Lambda^*(y)$$

así

$$\varliminf_{n\to\infty} \frac{1}{n}\mu_n(O) \geq -\inf\{\Lambda^*(y) : y \in O \cap \mathcal{E}\}$$

que es exactamente lo que queremos.

Demostraremos entonces 5.1.1 utilizando de nuevo un cambio de medi
Definimos así v_n por

$$\frac{dv_n}{d\mu_n}(x) = \exp[n\langle \eta, x\rangle - \Lambda_n(n\eta)]$$

Notamos el hecho que $\frac{1}{n}\Lambda_n(n\eta) \to \Lambda(\eta) \in \mathbb{R}$ implica que, al menos pa
$n \geq n_0$ suficientemente grande, $\frac{1}{n}\Lambda_n(n\eta)$ es finito. En el resto del párra
supondremos que $n \geq n_0$. Como anteriormente, tenemos, para todo $\varepsilon <$

$$\begin{aligned}
\mu_n(B(y,\delta)) &\geq \int_{B(y,\varepsilon)} \exp[-n\langle \eta, z\rangle - \Lambda_n(n\eta)]v_n(dz) \\
&= \int_{B(y,\varepsilon)} \exp[-n\langle \eta, z-y\rangle - n\langle \eta, y\rangle + \Lambda_n(n\eta)]v_n(dz).
\end{aligned}$$

Pero, según la desigualdad de Schwartz tenemos

$$|\langle \eta, z-y\rangle| \leq \|\eta\| \cdot \|z-y\| \leq \varepsilon \cdot \|\eta\|$$

entonces

$$\begin{aligned}
\varliminf_{n\to\infty} \tfrac{1}{n}\log\mu_n(B(y,\varepsilon)) &\geq \overbrace{\varliminf_{n\to\infty}\left(\frac{1}{n}\Lambda_n(n\eta)\right)}^{=\Lambda(\eta)} - \varepsilon\cdot\|\eta\| - \langle \eta, y\rangle \\
&\qquad + \varliminf_{n\to\infty} \tfrac{1}{n}\log v_n(B(y,\varepsilon)) \\
&\geq -\Lambda^*(y) - \varepsilon\cdot\|\eta\| + \varliminf_{n\to\infty} \tfrac{1}{n}\log v_n(B(y,\varepsilon)).
\end{aligned}$$

En esta etapa no es posible aplicar los métodos de los teoremas 3.2 y 4.1, ya que en este caso, las variables no son independientes, por lo que tendremos que demostrar el resultado auxiliar siguiente.

ema 5.3.1.

$$v_n(B(y,\varepsilon)) \to 1$$

Demostración.

Demostraremos que $v_n(B(y;\varepsilon)^c) \to 0$, para lo cual definimos la log-Laplace sociada a las v_n, sea

$$\Xi_n(\lambda) = \log\left(\int_{\mathbb{R}^d} e^{\langle\lambda,x\rangle} v_n(dx)\right),$$

ntonces

$$\begin{aligned}
\Xi_n(n\lambda) &= \log\left(\int_{\mathbb{R}^d} e^{\langle n\lambda,x\rangle}\cdot e^{n\langle\eta,x\rangle}\cdot e^{-\Lambda_n(n\eta)}\mu_n(dx)\right)\\
&= \log\left(e^{-\Lambda_n(n\eta)}\int_{\mathbb{R}^d} e^{\langle n(\lambda+\eta),x\rangle}\mu_n(dx)\right)\\
&= \Lambda_n(n(\lambda+\eta)) - \Lambda_n(n\eta)
\end{aligned}$$

le ahí

$$\begin{aligned}
\tfrac{1}{n}\Xi_n(n\lambda) &\longrightarrow \Lambda(\lambda+\eta) - \Lambda(\eta)\\
&= \log\mathbb{E}(e^{\langle\lambda+\eta,x\rangle}) - \log\mathbb{E}(e^{\langle\eta,x\rangle})\\
&= \log\mathbb{E}(e^{\langle\lambda+\eta,x\rangle-\langle\eta,x\rangle})\\
&= \log\mathbb{E}(e^{\langle,x\rangle})\\
&= \Xi(\lambda).
\end{aligned}$$

or lo que Ξ cumple exactamente las hipótesis del teorema 5.1. De este hecho, egún las primeras partes que ya han sido probadas y de la proposición 4.1, lisponemos de mayoraciones sobre los cerrados, es decir, sobre $B(y;\varepsilon)^c$, por lo ue

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log v_n(B(y,\varepsilon)^c) \le -\inf\{\Xi^\star(x) : x\in B(y,\varepsilon)^c\}. \qquad 5.1.2$$

asta con demostrar que

$$\inf\{\Xi^\star(x) : x\in B(y,\varepsilon)^c\} = \alpha > 0. \qquad 5.1.3$$

n efecto, si tenemos 5.1.3, entonces existirá $n_1 \ge n_0$ tal que para todo $n \ge n_1$

$$\frac{1}{n}\log v_n(B(y,\varepsilon)^c) \le -\frac{\alpha}{2}$$

entonces

$$v_n(B(y,\varepsilon)^c) \leq \exp\left[-\frac{n\alpha}{2}\right]$$

lo que dará el resultado.

Para mostrar 5.1.3 probamos anteriormente que la cota inferior se alcanz es decir, existe $x_\infty \notin B(y,\varepsilon)$ tal que

$$\inf\{\Xi^\star(x) : x \in B(y,\varepsilon)^c\} = \Xi^\star(x_\infty).$$

Hacemos notar que podemos suponer que $\Lambda^\star(y) < +\infty$ pues, si no fuera así, desigualdad 5.1.1 es trivial. Por otra parte

$$\begin{aligned} \Xi^\star(x) &= \sup\{\langle\lambda, x\rangle - \Lambda(\eta+\lambda) + \Lambda(\eta) : \lambda \in \mathbb{R}\} \\ &= \sup\{\langle\lambda+\eta, x\rangle - \Lambda(\eta+\lambda) - \langle\eta, x\rangle + \Lambda(\eta) : \lambda \in \mathbb{R}\} \\ &= \Lambda^\star(x) + \Lambda(\eta) - \langle\eta, x\rangle \end{aligned}$$

lo que prueba que $\Xi^\star(y) < \infty$ y que $\Xi^\star$ es una función de tasa buena. De es hecho, si $\alpha = +\infty$, no hay nada mas que decir, ya que tenemos lo que queremo si ahora $\alpha < +\infty$, entonces el conjunto de nivel $\{\Xi^\star(x) \leq \alpha + 1\}$ es compact Como α es una cota inferior (límite inferior), si tomamos una sucesión x_n c puntos $B(y,\varepsilon)^c$ tal que $\Xi^\star(x_n)$ converge hacia α, entonces podemos suponer qu la sucesión x_n toma valores dentro del conjunto de nivel $\{\Xi^\star(x) \leq \alpha + 1\}$ que implica que podemos extraer de ahí una sub-sucesión convergente, de límit $x_\infty \in B(y,\varepsilon)^c$ (porque $B(y,\varepsilon)^c$ es cerrado), y finalmente la s.c.i. de $\Xi^\star$ dá

$$\alpha \leq \Xi^\star(x_\infty \leq \varliminf_{n\to\infty} \Xi^\star(x_n) \leq \alpha$$

lo que muestra que α es alcanzada en x_∞. Probemos ahora que $\alpha > 0$. Com $x_\infty \notin B(y,\varepsilon)$ tenemos, en particular $x_\infty \neq y$. Pero y es un punto expuesto c $\Lambda^\star$, de hiperplano exponente η, entonces

$$\Lambda^\star(x_\infty) > \Lambda^\star(y) + \langle\eta, x_\infty - y\rangle$$

de ahí

$$\Lambda^\star(y) < \Lambda^\star(x_\infty) + \langle\eta, y - x_\infty\rangle$$

y

$$\Lambda^\star(y) \geq \langle\eta, y\rangle - \Lambda(\eta)$$

donde

$$\langle \eta, y\rangle - \Lambda(\eta) < \Lambda^\star(x_\infty) + \langle \eta, y - x_\infty\rangle$$

que, simplificando y teniendo en cuenta que $\Xi^\star(x_\infty) = \Lambda^\star(x_\infty) + \Lambda(\eta) - \langle \eta, x_\infty\rangle$, el resultado buscado, lo que explica el lema y la minoración.

;ibliografía

sh72] R. Ash. *Real Analysis and Probability.* Academic Press.Inc., 1972.

il79] Patrick Billingsley. *Probability and Measure.* New York :J.Wiley, 1979.

re75] L. Breimer. *Probability.* Addison, 1975.

or94] Refugio Trujillo Cortes. *Series Aleatorias.Tesis de Licenciatura.UNAM*, 1994.

em98] O. Dembo, A. & Zeitouni. *Large Deviations Techniques and Applications*, volume 1. Springer Verlag, second edition, 1998.

el71] W. Feller. *An Introduction to Probability Theory and Applications.* Wiley & Sons, 2nd. edition, 1971.

ac98] Annie Jacquet. Grandes deviations et aplicationes, 1998.

ew96] John T. & Russell Raymond Lewis. *An Introduction to Large Deviations for Teletraffic Engineers.* Dublin Institute for Advanced Studies (DIAS), 10 Burlington Road, Dublin 4, Ireland, October 1996. lewis@stp.dias.ie, russell@stp.dias.ie.

ic71] Durret Richard. *Probability: Theory and Examples.* Pacific Grove,California:Wadsworth, 1971.

os88] Sheldon Ross. *A first course in probability.* Mc.millan, 1988.