UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"LEYES CONDICIONALES DE LOS GRANDES NUMEROS Y SU APLICACION A LA MECANICA ESTADISTICA"

T E S I S

QUE PARA OBTENER EL TITULO DE:

M A T E M A T I C O

P R E S E N T A :

JUAN MARTIN BARRIOS VARGAS

DIRECTORA DE TESIS: DRA. ANA MEDA GUARDIOLA

FACULTAD DE CIENCIAS
UNAM

2005

VNIVERSIDAD NACIONAL AVTÓNOMA DE MEXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Juan Martín Barrios Vargas

FECHA: 1/12/05

FIRMA:

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
"Leyes Condicionales de los Grandes Números y su Aplicación a la Mecánica Estadística"

realizado por Juan Martín Barrios Vargas

con número de cuenta 09904560-5 , quien cubrió los créditos de la carrera de: Matemáticas

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dra.	Ana Meda Guardiola	
Propietario	Dra.	María Asunción Begoña Fernández Fernández	
Propietario	Dr.	Ramón Peralta y Fabi	
Suplente	Dr.	Javier Paéz Cárdenas	
Suplente	Act.	Nelson Omar Muriel Torrero	

Consejo Departamental de Matemáticas

M. en C. Alejandro Bravo Mojica

Quiero agradecer a mis padres, que me han apoyado en todas mis decisiones y dado *un jalón de orejas* cuando lo necesito.
Agradezco a la Dra. Ana Meda por su amistad, la cual me hizo mantener la calma a lo largo de este trabajo, y por el tiempo que dedicó en las discusiones surgidas.
A todos los sinodales que se molestaron en leer la tesis y que me hicieron comentarios para mejorarla.
A mi hermano que me ayuda a entender conceptos físicos.
A mi hermana por las sonrisas que tanto me ayudan.
A mis tíos, Rubí y Jorge, que al igual que mis padres me han apoyado en mis estudios.
Y a todos mis amigos que han hecho mi paso por la Facultad, más agradable.

Índice general

Introducción

En la tesis nos concentraremos en dos temas principales: las Grandes Desviaciones y la Mecánica Estadística. También expondremos una relación entre ambas teorías. En particular hablaremos de un concepto básico para ambas, la *entropía*.

Las Grandes Desviaciones se centran en propiedades sobre la convergencia de sistemas estocásticos. A lo largo de la tesis consideraremos uno de los ejemplos más conocidos de la teoría de la probabilidad: La media muestral S_n/n de una sucesión $X_1, X_2, \ldots$ de variables aleatorias independientes e idénticamente distribuidas, donde S_n corresponde a la n–ésima suma parcial de la sucesión. Es bien sabido, por la Ley Débil de los Grandes Números, que esta media muestral converge en probabilidad a su esperanza $\bar{x}$.

Las Grandes Desviaciones muestran que, bajo ciertas condiciones sobre las variables, la convergencia de la probabilidad de los eventos $\{|S_n/n - \bar{x}| > \epsilon\}$ es exponencial (en particular nos permite garantizar una convergencia casi segura de S_n/n, utilizando el Lema de Borel–Cantelli). Uno de los aspectos importantes es que este decaimiento exponencial es calculado con base en funciones de entropía (Funciones de Grandes Desviaciones).

El nombre *entropía* tiene sus orígenes en la Mecánica Estadística. Éste apareció en los trabajos de Rudolf Clausius y Ludwig Boltzmann en la segunda mitad del siglo XIX, quienes estudiaron la relación entre la entropía y las características macroscópicas de sistemas físicos.

Esta tesis está dividida en tres capítulos.

En el primer capítulo nos concentramos en presentar algunas definiciones de la Teoría de Grandes Desviaciones, siguiendo con un teorema que explotaremos en los capítulos posteriores. Este último es la Ley Condicional Débil de los Grandes Números (E. Nummelin, 1989), y se basa en el concepto de punto dominante, el cual tiene relación con la entropía.

En el siguiente capítulo presentamos teoremas de convergencia para las distribuciones empíricas, microcanónicas y canónicas. Los resultados de este capítulo son los que usaremos más adelante para la aplicación a la Mecánica Estadística. También cabe destacar que estos teoremas son consecuencia del punto dominante y de la Ley Condicional Débil de los Grandes Números.

Por último, en el tercer capítulo, utilizaremos todas las herramientas antes desarrolladas para aplicarlas a un ejemplo clásico de la Mecánica Estadística, el modelo del gas ideal. Serán de nuestro interés dos problemas: el primero es un sistema de partículas estacionario que sólo sufre cambios en la energía; el segundo es más complicado y corresponde al estudio de una pequeña subregión de un sistema más grande. En ambos casos calculamos la distribución de probabilidad con la cual se puede obtener información sobre las variables macroscópicas del problema. Aunque en este capítulo nos concentramos en resolver los problemas para el gas ideal, es fácil ver que se pueden extender los resultados a modelos más generales.

Esperamos que al final de la tesis se vea una relación más clara entre las Grandes Desviaciones y la Mecánica Estadística, además de presentar un método para abordar problemas de esta última.

La principal referencia de la tesis es el artículo de E. Nummelin y T. Lehtonen [LN90]; en éste se basan los Capítulos 2 y 3. La parte introductoria a las Grandes Desviaciones se basa en el libro de A. Dembo y O. Zetouni [DZ98]. La Ley Débil de los Grandes Números es demostrada en [Num89], mientras que la existencia del punto dominante y sus propiedades fueron consultadas del artículo de P. Ney [Ney83].

Notación

$\mathbb{P}$	Probabilidad.
$\mathbb{E}$	Esperanza.
c.r.a	*Con respecto a .*
i.i.d	*Independientes e idénticamente distribuidas .*
$\delta_y(dx)$	Distribución de Dirac.
$\mathcal{X}$	Espacio, $\mathcal{X} = \mathbb{R}$ o $\mathcal{X} = \mathbb{R}^d$.
$M_1(\mathcal{X})$	Espacio de todas las medidas de probabilidad sobre $\mathcal{X}$.
$\langle \cdot, \cdot \rangle$	Producto interno en $\mathcal{X}$.
$\lvert x \rvert$	Norma euclideana de $\mathcal{X}$.
$\lVert x \rVert_\infty$	Norma definida por $\lVert x \rVert_\infty = \text{máx}_i \lvert x_i \rvert$, donde x_i es la i-ésima coordenada del vector x.
$B(x, r)$	Bola abierta centrada en x y radio r, $B(x, r) = \{y \,:\, \lvert x - y \rvert < r\}$.
E°	Interior del conjunto E.
$\bar{E}$	Cerradura del conjunto E.
$\lvert E \rvert$	Volumen (área) del conjunto E.
$\mathcal{F}$	σ-álgebra del espacio $\mathcal{X}$.
$\mathcal{B}_{\mathcal{X}}$	σ-álgebra de Borel de $\mathcal{X}$.
$\Psi_I(\alpha)$	Conjuntos de nivel $\{x \,:\, I(x) \leq \alpha\}$ de la función I.
$\mathcal{D}_I$	Dominio efectivo de I, $\mathcal{D}_I = \{x \,:\, I(x) < \infty\}$.
$I^\delta(x)$	δ-Función de Grandes Desviaciones.
$I(E)$	$I(E) = \text{ínf}_{x \in E} I(x)$.
v_E	Punto dominante del conjunto E.
$M(\lambda)$	Función Generadora de Momentos. Transformada de Laplace.
$\Lambda(\lambda)$	Logaritmo de la Función Generadora de Momentos.
$I^*(x)$	Transformada de Legendre–Fenchel de I. Conjugada convexa de I.

Capítulo 1

Preliminares de Grandes Desviaciones

Las Grandes Desviaciones tienen un origen relativamente reciente. La Teoría de Grandes Desviaciones forma parte de los problemas clásicos de la probabilidad, ya que al igual que las Leyes de los Grandes Números y el Teorema del Límite Central, analiza el comportamiento asintótico de los sistemas estocásticos.
En el presente capítulo presentamos las Grandes Desviaciones y teoremas clásicos de esta teoría. En la última parte del capítulo veremos la definición de punto dominante introducido por Peter Ney [Ney83] y su relación con la entropía de conjuntos. A continuación presentaremos la Ley Condicional Débil de los Grandes Números, que se basa en la existencia del punto dominante y sus propiedades.

1.1. Principio de Grandes Desviaciones

Un problema clásico de la probabilidad es investigar qué pasa con los promedios o sumas de Cèsaro de una sucesión de variables aleatorias. Este problema, para el caso de variables aleatorias independiente e idénticamente distribuidas, es resuelto por las Leyes de los Grandes Números las cuales nos aseguran una convergencia a la media de las variables. Una pregunta natural, después de haber resuelto este problema, es qué tan rápida es esta convergencia.

El *principio de grandes desviaciones* (PGD) es un acercamiento en esta dirección. Nos da información del comportamiento asintótico de la probabilidad de los eventos.
Esto se hace por medio de cotas superiores e inferiores sobre la medida de los conjuntos en $\mathcal{F}$, las cuales dependen de una función, llamada la *función de grandes desviaciones* (FGD). En esta sección definiremos el principio de grandes desviaciones y también daremos algunas equivalencias a éste, dependiendo del espacio en que estemos trabajando o cómo sea la sucesión de medidas.

Definición 1.1. Una función de grandes desviaciones I es una función $I : \mathcal{X} \longrightarrow [0, \infty]$ que es semicontinua inferiormente (es decir que los conjuntos de nivel $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ son cerrados). Además diremos que I es una buena función de grandes desviaciones si los conjuntos de nivel $\Psi_I(\alpha)$ son compactos en $\mathcal{X}$.

Basta ver que en sucesiones se tiene que dada $\{x_n\}$ una sucesión que converge a x, entonces $\liminf_{x_n \to x} I(x_n) \geq I(x)$, para verificar que I es semicontinua inferiormente.

Definición 1.2. Decimos que una familia de medidas de probabilidad en $\mathcal{X}$, $\{\rho_\epsilon\}_{\epsilon>0}$, tiene la propiedad de grandes desviaciones para la FGD I, si para todo $E \in \mathcal{F}$ tenemos

$$-I(E^\circ) \leq \liminf_{\epsilon \to 0} \epsilon \log \rho_\epsilon(E) \leq \limsup_{\epsilon \to 0} \epsilon \log \rho_\epsilon(E) \leq -I(\bar{E}),$$

donde $I(F) = \inf_{x \in F} I(x)$.

Notemos que si ρ_ϵ cumple la propiedad de grandes desviaciones con FGD I, y tenemos que

$$I(E^\circ) = I(\bar{E}) = I(E),$$

para algún $E \in \mathcal{F}$, entonces

$$\lim_{\epsilon \to 0} \epsilon \log \rho_\epsilon(E) = -I(E).$$

A los conjuntos con esta propiedad se les llama *conjuntos de continuidad de I*.

Como $\rho_\epsilon(\mathcal{X}) = 1$, basta $I(\mathcal{X}) = 0$ para que tengamos la cota superior de la Definición 1.2. Si I es una buena FGD existe al menos un punto x tal que $I(x) = 0$, esto porque $\bigcap_{n\in\mathbb{N}} \Psi_I(1/n) \neq \emptyset$ por la compacidad de estos conjuntos de nivel y que $\Psi_I(1/m) \subsetneq \Psi_I(1/n)$ si $m \geq n$. También tenemos que las cotas se tienen trivialmente si $I(E^\circ) = \infty$ y $I(\bar{E}) = 0$. Otra caracterización de la PGD la podemos dar apoyándonos en los conjuntos de nivel $\Psi_I(\alpha)$, antes mencionados.

(a) (*Cota superior*) Para todo $\alpha \in \mathbb{R}$ y cualquier conjunto medible E tal que $E \subsetneq \Psi_I(\alpha)^c$ tenemos que

$$\limsup_{\epsilon\to 0} \epsilon \log \rho_\epsilon(E) \leq -\alpha. \tag{1.1}$$

(b) (*Cota inferior*) Para todo $x \in \mathcal{D}_I$ y todo medible E tal que $x \in E^\circ$ se tiene que

$$\liminf_{\epsilon\to 0} \epsilon \log \rho_\epsilon(E) \geq -I(x), \tag{1.2}$$

donde $\mathcal{D}_I = \{x : I(x) < \infty\}$ es el dominio efectivo de I.

Es fácil ver que las condiciones anteriores son equivalentes al hecho de que $\{\rho_\epsilon\}$ tenga la PGD con FGD I. Otra cosa que es importante hacer notar es que la segunda condición resalta la naturaleza local de la cota inferior.

Algunas veces para demostrar la cota superior resulta más fácil trabajar con funciones cuyo rango es acotado, es por ello que definimos una función de rango acotado y equivalente, desde el punto de vista de validez de la cota, a la función I.

Definición 1.3. Para cualquier FGD I y $\delta > 0$, definimos la δ-FGD como

$$I^\delta(x) = \min\{I(x) - \delta, \frac{1}{\delta}\}. \tag{1.3}$$

Vemos que la función I^δ no necesariamente es una FGD. Sin embargo tenemos que

$$\lim_{\delta\to 0} \inf_{x\in E} I^\delta(x) = I(E).$$

Es por esto que la cota superior puede ser reformulada, haciendo uso de esta función, como

$$\limsup_{\epsilon\to 0} \epsilon \log \rho_\epsilon(E) \leq I^\delta(E),$$

para todo E medible y cualquier $\delta > 0$.

Ahora, si tenemos que la σ–álgebra de Borel de $\mathcal{X}$ está contenida en $\mathcal{F}$, podemos enunciar otra caracterización para el PGD:

(a) (*Cota superior*) Para todo $F \subseteq \mathcal{F}$ cerrado

$$\limsup_{\epsilon\to 0} \epsilon \log \rho_\epsilon(F) \leq -\inf_{x\in F} I(x). \tag{1.4}$$

(b) (*Cota inferior*) Para cualquier $G \subseteq \mathcal{F}$ abierto

$$\liminf_{\epsilon\to 0} \epsilon \log \rho_\epsilon(G) \geq -\inf_{x\in G} I(x). \tag{1.5}$$

A lo largo de la tesis trabajaremos con un familia numerable de medidas $\{\rho_n\}$, donde ρ_n será la ley de la media muestral[1] de una sucesión de variables aleatorias. En el caso de familias numerables de medidas el PGD con sucesión $\{a_n\}_{n\in\mathbb{N}}$, lo podemos enunciar como

$$-\inf_{x\in E^\circ} I(x) \leq \liminf_{n\to\infty} a_n \log \rho_n(E) \leq \limsup_{n\to\infty} a_n \log \rho_n(E) \leq -\inf_{x\in \bar{E}} I(x),$$

donde $E \in \mathcal{F}$ y $\{a_n\}$ es una sucesión que converge a cero. En nuestro caso usaremos a $a_n = \frac{1}{n}$. En varias ocasiones demostrar la validez del PGD es complicado, es por esto que es necesario definir condiciones equivalentes que resulten en la práctica mucho más fácil de trabajar.

Definición 1.4. Supongamos que todo compacto de $\mathcal{X}$ está en $\mathcal{F}$. Decimos que una familia de medida de probabilidad $\{\rho_\epsilon\}$ satisface débilmente el principio de grandes desviaciones con función de grandes desviaciones I, si la cota superior

$$\limsup_{\epsilon\to 0} \epsilon \log \rho_\epsilon(K) \leq -\alpha; \qquad K \subseteq \Psi_I(\alpha)^c$$

[1] Si $\{X_n\}$ es una sucesión de variables aleatoria la media muestral, es $\frac{1}{n}\sum_{i=1}^n X_i$.

es válida para todo $\alpha < \infty$ y todos los subconjuntos K compactos de $\Psi_I(\alpha)^c$, y además la cota inferior

$$\liminf_{\epsilon\to 0} \epsilon \log \rho_\epsilon(E) \geq -I(x),$$

vale para cualquier $E \in \mathcal{F}$ y $x \in E^\circ$.

Definición 1.5. Supongamos que todo compacto de X está en $\mathcal{F}$. Una familia de medidas de probabilidad $\{\rho_\epsilon\}$ tiene tensión exponencial si para toda $\alpha < \infty$, existe un compacto K_α tal que

$$\limsup_{\epsilon\to 0} \epsilon \log \rho_\epsilon(K_\alpha^c) < -\alpha. \tag{1.6}$$

Claramente estas dos condiciones son más relajadas que las pedidas por el PGD. Ahora veremos cómo nos ayudarán a probar el PGD. Antes demostraremos un lema técnico que nos servirá más adelante.

Lema 1.6. *Sea N un entero fijo. Entonces para toda $a_\epsilon^i > 0$, tenemos que*

$$\limsup_{\epsilon\to 0} \epsilon \log\left(\sum_{i=1}^{N} a_\epsilon^i\right) = \max_{i=1\ldots N} \limsup_{\epsilon\to 0} \epsilon \log a_\epsilon^i. \tag{1.7}$$

Demostración. Para toda ϵ tenemos

$$0 \leq \epsilon \log\left(\sum_{i=1}^{N} a_\epsilon^i\right) - \max_{i=1,2,\ldots,N} \epsilon \log a_\epsilon^i \leq \epsilon \log N.$$

Ahora, como N es fijo, $\epsilon \log N \to 0$ si $\epsilon \to 0$ y

$$\limsup_{\epsilon\to 0} \max_{i=1\ldots N} \epsilon \log a_\epsilon^i = \max_{i=1\ldots N} \limsup_{\epsilon\to 0} \epsilon \log a_\epsilon^i, \tag{1.8}$$

de donde se tiene el resultado. □

Lema 1.7. *Sea $\{\rho_\epsilon\}$ una familia de medidas de probabilidad con tensión exponencial.*

(a) Si la cota superior (1.1) *se tiene para alguna* α *y para todos los compactos de* $\Psi(\alpha)^c$, *entonces se tiene para todos los medibles E tales que* $\bar{E} \subseteq \Psi(\alpha)^c$.

(b) Si la cota inferior (1.2) *se tiene para todos los medibles, entonces* $I(\cdot)$ *es una buena FGD.*

Demostración. (a) Primero demostraremos (1.1). Tomemos $E \in \mathcal{F}$ y $\alpha < \infty$ tal que $\bar{E} \subseteq \Psi(\alpha)^c$. Sea K_α como en (1.6). Notemos que $\bar{E} \cap K_\alpha \in \mathcal{F}$ y $K_\alpha^c \in \mathcal{F}$, entonces

$$\rho_\epsilon(E) \leq \rho_\epsilon(\bar{E} \cap K_\alpha) + \rho_\epsilon(K_\alpha^c).$$

También tenemos que $\bar{E} \cap K_\alpha \subseteq \Psi_I(\alpha)^c$, y por lo tanto $\inf_{x \in \bar{E} \cap K_\alpha} I(x) \geq \alpha$. Ahora por cómo se tomó K_α, la hipótesis para el compacto $\bar{E} \cap K_\alpha$, y del Lema 1.6 tenemos que

$$\limsup_{\epsilon \to 0} \epsilon \log \rho_\epsilon(E) \leq \max \left\{ \limsup_{\epsilon \to 0} \epsilon \log \rho_\epsilon(\bar{E} \cap K_\alpha), \limsup_{\epsilon \to 0} \epsilon \log \rho_\epsilon(K_\alpha^c) \right\} \leq -\alpha.$$

(b) Notamos que $K_\alpha^c \in \mathcal{F}$ es abierto, y como (1.6) se cumple, tenemos que $\inf_{x \in K_\alpha^c} I(x) \geq \alpha$. Por lo tanto $\Psi_I(\alpha) \subseteq K_\alpha$, de donde se tiene la compacidad de $\Psi_I(\alpha)$, y por lo tanto $I(\cdot)$ es una buena FGD.

□

Observación. I. Si en la proposición anterior tenemos que $\mathcal{B}_X \subseteq \mathcal{F}$, ésta se puede reformular como:

(a) (Cota superior) Si la cota (1.4) se tiene para los compactos, en particular se tiene para cerrados.

(b) (Buena FGD) Si para todo abierto se cumple (1.5), entonces $I(\cdot)$ es una buena FGD.

II. Cada vez que digamos que una familia $\{\rho_\epsilon\}$ cumple el principio débil de grandes desviaciones o tiene tensión exponencial inferiremos que todo compacto de X está en $\mathcal{F}$.

1.2. Teoremas Centrales de la Teoría de Grandes Desviaciones

En la sección anterior vimos algunas definiciones de la Teoría de Grandes Desviaciones. También enunciamos algunas equivalencias de este principio dependiendo del tipo de medidas o de la σ–álgebra $\mathcal{F}$.

Ahora nos abocaremos a presentar tres teoremas que son de los principales en la teoría. El Teorema de Cramér en su versión para medidas en $\mathbb{R}$ y en $\mathbb{R}^d$; presentamos el de la versión en $\mathbb{R}$ ya que nos ilustra la existencia de un punto en el cual veremos que la FGD alcanzará el ínfimo sobre conjuntos suficientemente *decentes* (convexos). Esta idea se puede generalizar a cualquier espacio $\mathbb{R}^d$. El último teorema de la sección corresponde a la Ley Condicional Débil de los Grandes Números, en la cual basaremos mucha de la teoría que desarrollaremos para la Mecánica Estadística. Este último teorema fue demostrado por Nummelin [Num89].

A lo largo de la sección tomaremos $X_1, X_2, \cdots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) en $\mathbb{R}$ o en $\mathbb{R}^d$, con ley ρ y ρ_n la ley de la media muestral de estas variables. Denotaremos por

$$\Lambda(\lambda) = \log \mathbb{E}\left[e^{\langle \lambda, X_1 \rangle}\right], \tag{1.9}$$

al logaritmo de la función generadora de momentos de X_1. También es importante decir que trabajaremos con la σ-álgebra de Borel, siempre que no se diga lo contrario.

1.2.1. Teorema de Cramér para medidas en $\mathbb{R}$

El teorema de Cramér para medidas en $\mathbb{R}$ es un de los primeros para la teoría de Grandes Desviaciones, también se le conoce como Teorema de Chernoff, y es el que mejor nos ilustra varios de los conceptos que usaremos a lo largo de la tesis.

Definición 1.8. Sea $f : \mathbb{R} \to \mathbb{R}$ una función convexa. La transformada de Legendre–Fenchel de f (o conjugada convexa de f) es

$$f^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - f(\lambda)\}. \tag{1.10}$$

Aplicamos esta definición a la función Λ, de modo que

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}.$$

También denotaremos al dominio efectivo de Λ por

$$\mathscr{D}_\Lambda = \{\lambda \in \mathbb{R} \ : \ \Lambda(\lambda) < \infty\}.$$

Como la transformada de Legendre–Fenchel sólo la definimos para funciones convexas, por supuesto verificaremos que Λ es convexa. Presentamos este resultado y algunas propiedades más en el siguiente Lema.

Lema 1.9. *Sea X una variable aleatoria con ley ρ.*

(a) Λ es una función convexa y Λ^* es una función de grandes desviaciones convexa.

(b) Si $\mathscr{D}_\Lambda = \{0\}$, entonces $\Lambda^* \equiv 0$. Si $\Lambda(\lambda) < \infty$ para alguna $\lambda > 0$ entonces $\bar{x} = \mathbf{E}X < \infty$, y para toda $x \geq \bar{x}$

$$\Lambda^*(x) = \sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\}, \tag{1.11}$$

es una función no decreciente. Análogamente, si $\Lambda(\lambda) < \infty$ para alguna $\lambda < 0$ entonces $\bar{x} > -\infty$, y para toda $x \leq \bar{x}$,

$$\Lambda^*(x) = \sup_{\lambda \leq 0} \{\lambda x - \Lambda(x)\}, \tag{1.12}$$

es una función no creciente.

Si $\bar{x}$ es finita, $\Lambda^*(\bar{x}) = 0$ y siempre $\inf_{x \in \mathbb{R}} \Lambda^*(x) = \Lambda^*(\bar{x}) = 0$.

(c) $\Lambda(\cdot)$ *es diferenciable en* $\mathcal{D}_\Lambda^\circ$ *con*

$$\Lambda'(\eta) = \frac{1}{M(\eta)}\mathbb{E}\left[Xe^{\eta X}\right], \tag{1.13}$$

donde $M(\lambda) = \mathbb{E}[e^{\lambda X}]$ *es la función generadora de momentos o transformada de Laplace. Más aún, si* $\Lambda'(\eta) = y$ *se tiene que*

$$\Lambda^*(y) = \eta y - \Lambda(\eta). \tag{1.14}$$

Demostración. (a) La convexidad de la función Λ es consecuencia inmediata de la desigualdad de Hölder, tomando como exponentes conjugados $\frac{1}{t}$ y $\frac{1}{1-t}$, con $t \in (0,1)$, tenemos

$$\mathbb{E}[\exp\{t\lambda_1 X + (1-t)\lambda_2 X\}] \le \mathbb{E}[e^{\lambda_1 X}]^t \mathbb{E}[e^{\lambda_2 X}]^{1-t}.$$

La convexidad de la función Λ^* es consecuencia inmediata de la definición y de la convexidad de Λ.

Sólo nos falta demostrar que Λ^* es una FGD. Para esto notemos que $\Lambda(0) = \log \mathbb{E}[1] = 0$ y por lo tanto para cualquier $x \in \mathbb{R}$ tenemos $\Lambda^*(x) \ge 0x - \Lambda(0) = 0$. Falta ver que Λ^* es semicontinua inferiormente; para esto tomemos $\{x_n\}$ que converja a x, entonces para toda $\lambda \in \mathbb{R}$

$$\liminf_{x_n \to x} \Lambda^*(x_n) \ge \liminf_{x_n \to x} \{\lambda x_n - \Lambda(\lambda)\} = \lambda x - \Lambda(\lambda),$$

por lo tanto

$$\liminf_{x_n \to x} \Lambda^*(x_n) \ge \sup_{\lambda} \{\lambda x - \Lambda(\lambda)\} = \Lambda^*(x),$$

(b) Si $\mathcal{D}_\Lambda = \{0\}$, entonces $\Lambda^*(x) = \Lambda(0) = 0$ para toda $x \in \mathbb{R}$ y por lo tanto $\Lambda^* \equiv 0$. Ahora, si $\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X}] < \infty$ para alguna $\lambda > 0$,

$$\int_0^\infty x\rho(dx) \le \frac{\mathbb{E}[e^{\lambda X}]}{\lambda} < \infty,$$

porque $\lambda x \leq e^{\lambda x}$, y por lo tanto $\bar{x} < \infty$ (tal vez $\bar{x} = -\infty$). Aplicando la desigualdad de Jensen tenemos

$$\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X}] \geq \mathbb{E}[\log e^{\lambda X}] = \lambda \bar{x}.$$

Entonces si $\bar{x} = -\infty$, $\Lambda(\lambda) = \infty$ para $\lambda < 0$, y por lo tanto tenemos

$$\Lambda^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}. \tag{1.15}$$

Ahora, si $\bar{x}$ es finita es claro que $\Lambda^*(\bar{x}) = 0$, y en este caso, para toda $x \geq \bar{x}$ y $\lambda < 0$

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

de donde se sigue (1.15). Observemos que (1.15) implica la monotonía de Λ^* en $(\bar{x}, \infty)$, ya que para toda $\lambda \geq 0$, la función $\lambda x - \Lambda(\lambda)$ es una función monótona con respecto a x. Si tenemos que $\Lambda(\lambda) < \infty$ para alguna $\lambda < 0$, tenemos que todo lo anterior es válido para $-X$, y por lo tanto

$$\Lambda^*(x) = \sup_{\lambda \leq 0}\{\lambda x - \Lambda(\lambda)\},$$

y Λ^* es decreciente en $(-\infty, \bar{x})$.

Sólo resta ver que $\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0$, lo cual ya vimos en el caso de que $\mathcal{D}_\Lambda = \{0\}$, y también cuando $\bar{x}$ es finito (donde tenemos que $\Lambda^*(\bar{x}) = 0$). Consideremos el caso $\bar{x} = -\infty$, es decir $\Lambda(\lambda) < \infty$ para alguna $\lambda < 0$, entonces por la desigualdad de Tchebyshev y por (1.15) tenemos

$$\log \rho([x, \infty)) \leq \inf_{\lambda \geq 0} \log \mathbb{E}[e^{\lambda(X - x)}] = -\sup_{\lambda > 0}\{\lambda x - \Lambda(\lambda)\} = -\Lambda^*(x),$$

entonces

$$\lim_{x \to -\infty} \Lambda^*(x) \leq \lim_{x \to -\infty} -\log \rho([x, \infty)) = 0.$$

Y análogamente hacemos el caso cuando $\bar{x} = \infty$.

(c) La fórmula (1.13) es una consecuencia inmediata de intercambiar el orden de integración y diferenciación, esto observando que $f_\epsilon(x) = \frac{1}{\epsilon}(e^{(\lambda+\epsilon)x} - e^{\lambda x})$ converge puntualmente a $xe^{\lambda x}$ y que podemos aplicar el Teorema de Convergencia Dominada, ya que si $\lambda \in \mathscr{D}_\Lambda$, existe $\delta > 0$ tal que $\lambda + \delta \in \mathscr{D}_\Lambda$ y por lo tanto $|f_\epsilon(x)| \leq \frac{e^{\lambda}(e^{\eta x}-1)}{\eta}$ para $\epsilon \in (-\eta, \eta)$.

Sea $\Lambda'(\eta) = y$ y tomemos $g(\lambda) = \lambda y - \Lambda(\lambda)$, observamos que $g(\cdot)$ es cóncava y que $g'(\eta) = 0$, de donde concluimos que $g(\eta) = \sup_{\lambda\in\mathbb{R}} g(\lambda)$.

□

En el lema hemos visto propiedades importantes de las funciones Λ y Λ^*. Como éstas dos son funciones convexas se les puede aplicar toda la herramienta que existe de Análisis Convexo; como éste no es el fin de la tesis nos quedaremos por ahora con estas propiedades.

Teorema 1.10 (Chernoff–Cramér). *Sea $\{X_n\} \in \mathbb{R}$ una sucesión de variables aleatorias i.i.d., y tomemos ρ_n la distribución de la media muestral. Entonces $\{\rho_n\}$ satisface el principio de grandes desviaciones con FGD Λ^*, es decir, tenemos:*

(a) Para cualquier conjunto cerrado $F \subseteq \mathbb{R}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \rho_n(F) \leq \inf_{x\in F} \Lambda^*(x). \tag{1.16}$$

(b) Para cualquier conjunto abierto $G \subseteq \mathbb{R}$

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n(G) \geq \inf_{x\in G} \Lambda^*(x). \tag{1.17}$$

Obsérvese que Λ^* en general no será una buena FGD, y esto es consecuencia inmediata del Lema 1.9.

Demostración. (a) Tomemos F un cerrado no vacío. Claramente la desigualdad (1.16) se tiene si $I_F = \inf_{x\in F} \Lambda^*(x) = 0$, por lo que suponemos que $I_F > 0$. Supongamos que existe

$\lambda \geq 0$ tal que $\Lambda(\lambda) < \infty$, por la parte (b) del Lema 1.9 se sigue que $\bar{x} < \infty$ posiblemente $\bar{x} = -\infty$. Entonces, dada $x \in \mathbb{R}$ tenemos, aplicando la desigualdad de Tchebyshev,

$$\begin{aligned}\rho_n([x,\infty)) &\leq \mathbb{E}[\mathbf{1}_{S_n/n-x>0}] \leq \frac{\mathbb{E}[e^{n\lambda S_n/n}]}{e^{n\lambda x}} = \mathbb{E}[e^{n\lambda(S_n/n-x)}] = \\ &= e^{-n\lambda x}\mathbb{E}[e^{\lambda S_n}] = e^{-n\lambda x}\prod_{i=1}^{n}\mathbb{E}[e^{\lambda X_i}] = e^{-n[\lambda x-\Lambda(\lambda)]}.\end{aligned}$$

Por lo tanto, como $\bar{x} < \infty$,para todo $x > \bar{x}$

$$\rho_n([x,\infty)) \leq e^{-n\Lambda^*(x)}. \tag{1.18}$$

Análogamente, si existe $\lambda \leq 0$ tal que $\Lambda(\lambda) < \infty$, entonces $\bar{x} > -\infty$ y para todo $x < \bar{x}$

$$\rho_n((-\infty,x]) \leq e^{-n\Lambda^*(x)}. \tag{1.19}$$

Si suponemos que $\bar{x}$ es finita entonces $\Lambda^*(\bar{x}) = 0$ y por lo tanto, como $I_F > 0$, tenemos que $\bar{x} \notin F$. Sea (x_-, x_+) la unión de todos los intervalos $(a,b) \subseteq F^c$ tales que $\bar{x} \in (a,b)$. Tenemos que $x_- < x_+$, y que x_- o x_+ deben ser finitos, porque F no es vacío.

Si x_- es finito, entonces $x_- \in F$, y en consecuencia

$$\Lambda^*(x_-) \geq I_F.$$

Por la misma razón $\Lambda^*(x_+) \geq I_F$ siempre y cuando x_+ sea finito. Ahora por (1.18) y (1.19), tenemos que

$$\rho_n(F) \leq \rho_n((-\infty,x_-]) + \rho_n([x_+,\infty)) \leq 2e^{-nI_F}. \tag{1.20}$$

Ahora bien, si $\bar{x} = -\infty$ entonces, por ser Λ^* no decreciente, se sigue que $\lim_{x\to-\infty}\Lambda^*(x) = 0$, y por lo tanto $x_+ = \inf\{x : x \in F\}$ es finito. Como F es cerrado, $x_+ \in F$, y por lo tanto $\Lambda^*(x_+) \geq I_F$. Por último, notamos que $F \subseteq [x_+,\infty)$ de donde tenemos lo deseado.

Análogamente hacemos el caso donde $\bar{x} = \infty$.

(b) Ahora probaremos que para toda $\delta > 0$ y toda distribución marginal $\rho \in M_1(\mathbb{R})$

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\delta,\delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0). \tag{1.21}$$

En efecto como la transformación $Y = X - x$ cumple que $\Lambda_Y(\lambda) = \Lambda(\lambda) - \lambda x$ y entonces $\Lambda_Y^*(\cdot) = \Lambda^*(\cdot + x)$, se sigue de la desigualdad anterior que para toda x y toda $\delta > 0$

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n((x-\delta, x+\delta)) \geq -\Lambda^*(x) \tag{1.22}$$

Entonces para todo abierto G, para todo $x \in G$ y para toda $\delta > 0$ tal que $(x-\delta, x+\delta) \subseteq G$, se tiene lo deseado.

Para demostrar (1.21), primero supongamos que $\rho((-\infty,0)) > 0$ y $\rho((0,\infty)) > 0$, y que ρ tiene soporte acotado en $\mathbb{R}$. Bajo estos supuestos tenemos que $\Lambda(\lambda)$ tiende a infinito si hacemos $|\lambda|$ tender a infinito y también tenemos $\Lambda(\lambda)$ es finita para cualquier λ. Entonces tenemos que $\Lambda(\cdot)$ es diferenciable y por lo tanto continua. Por la convexidad de Λ existe η finito tal que

$$\Lambda(\eta) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) \quad \text{y} \quad \Lambda'(\eta) = 0.$$

Definimos ahora para esta η fija

$$\tilde{\rho}(dx) = e^{\eta x - \Lambda(\eta)} \rho(dx),$$

entonces $\tilde{\rho}$ es una medida de probabilidad, conocida como *medida de Gibbs con dirección* η.

Sea $\tilde{\rho}_n$ la ley que gobierna a S_n/n, donde las $\tilde{X}_i$ son variables aleatorias i.i.d. con ley $\tilde{\rho}$. Notemos que para toda $\epsilon > 0$,

$$\begin{aligned}
\rho_n((-\epsilon,\epsilon)) &= \int_{|\sum_{i=1}^n x_i|<n\epsilon} \rho(dx_1)\rho(dx_2)\ldots\rho(dx_n) \\
&\geq e^{-n\epsilon|\eta|} \int_{|\sum_{i=1}^n x_i|<n\epsilon} \exp\left(\eta \sum_{i=1}^n x_i\right) \rho(dx_1)\ldots\rho(dx_n) \\
&= e^{-n\epsilon|\eta|} e^{n\Lambda(\eta)} \tilde{\rho}((-\epsilon,\epsilon)).
\end{aligned} \tag{1.23}$$

Por (1.13), y la elección de η, tenemos

$$\mathbb{E}[\tilde{X}_1] = \frac{1}{M(\eta)} \int_{\mathbb{R}} x e^{\eta x} \rho(dx) = \Lambda'(\eta) = 0.$$

entonces, por la Ley Débil de los Grandes Números,

$$\lim_{n\to\infty} \tilde{\rho}_n((-\epsilon, \epsilon)) = 1.$$

Por lo tanto de (1.23) tenemos que para todo $0 < \epsilon < \delta$

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\delta, \delta)) \geq \liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\epsilon, \epsilon)) \geq \Lambda(\eta) - \epsilon|\eta|.$$

Haciendo ϵ tender a cero, tenemos que

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\delta, \delta)) \geq \Lambda(\eta) = \Lambda^*(0).$$

Ahora, si ρ no tiene soporte acotado, y $\rho((-\infty, 0)) > 0$, $\rho((0, \infty)) > 0$, fijemos M tal que $\rho([-M, 0)) > 0$ y $\rho((0, M]) > 0$, y sea

$$\Lambda_M(\lambda) = \log \int_{-M}^{M} e^{\lambda x} \rho(dx).$$

Sea ν la ley de X condicionada a que $\{|X| \leq M\}$, y sea ν_n la ley de S_n/n condicionada a $B = \{|X_i| \leq M : i = 1, 2, \ldots, n\}$. Entonces, para todo n y $\delta > 0$

$$\begin{aligned} \rho_n((-\delta, \delta)) &= \mathbb{P}(S_n/n \in (-\delta, \delta) \mid B)\mathbb{P}(B) + \mathbb{P}(S_n/n \in (-\delta, \delta) \mid B^c)\mathbb{P}(B^c) \\ &\geq \nu_n((-\delta, \delta))\rho([-M, M])^n. \end{aligned}$$

Notemos que la prueba anterior de la desigualdad (1.21) se aplica para ν_n, con la función de momentos logarítmicos de ν, $\Lambda_M(\cdot) - \log \rho([-M, M])$, entonces

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\delta, \delta)) &\geq \log([-M, M] + \liminf_{n\to\infty} \frac{1}{n} \log \nu_n((-\delta, \delta)) \\ &\geq \inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda). \end{aligned}$$

Si definimos $I_M = -\operatorname{ínf}_{\lambda\in\mathbb{R}} \Lambda_M(\lambda)$ e $I^* = \operatorname{lím\,sup}_{M\to\infty} I_M$, tenemos que

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n((-\delta,\delta)) \geq -I^*. \tag{1.24}$$

Podemos ver que $\Lambda_M^* \geq \Lambda_N^*$ si $M \geq N$, es decir las funciones Λ_M son no decrecientes en M, y entonces también $-I_M$ es no decreciente en M, más aún, $-I_M \leq \Lambda_M(0) \leq \Lambda(0) = 0$ por lo cual $-I^* \leq 0$. Por lo tanto los conjuntos de nivel $\{\lambda : \Lambda_M(\lambda) \leq -I^*\}$ son no vacíos, compactos y decrecientes con respecto a M, y entonces existe al menos un punto, que denotaremos como λ_0, en la intersección de estos. Ahora por el teorema de convergencia monótona, $\Lambda(\lambda_0) = \operatorname{lím}_{M\to\infty} \Lambda_M(\lambda_0) \leq -I^*$ y por lo tanto tenemos la cota deseada para medidas de soporte no acotado.

Por último si ρ es una medida tal que $\rho((-\infty,0)) = 0$ o $\rho((0,\infty)) = 0$, tenemos que $\Lambda(\cdot)$ es una función monótona con $\operatorname{ínf}_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \log \rho(\{0\})$. Entonces la cota se sigue de

$$\rho_n((-\delta,\delta)) \geq \rho_n(\{0\}) = \rho(\{0\})^n. \tag{1.25}$$

□

Observación. (a) En la demostración de la primera parte del teorema se usa la desigualdad de Tchebyshev, además de la geometría de $\mathbb{R}$.

(b) En la segunda parte de la demostración tratamos de simplificar las cosas trabajando con medidas de soporte compacto, ya que esto nos daba un punto en el cual podíamos hacer un *twist* (giro) de nuestra medida, y al igual que en la primera parte, usamos la geometría de $\mathbb{R}$.

Estas dos observaciones nos hacen ver que no es trivial la extensión de este teorema a dimensiones mayores.

Otra cosa que podemos notar es que a lo largo de la demostración usamos el Lema 1.9, en el cual teníamos las principales propiedades de la Transformada de Lengendre-Fenchel, de las cuales algunas podremos rescatar para $\mathbb{R}^d$.

Debemos tener en cuenta que todas las cotas pueden ser triviales y no necesariamente útiles. Podemos notar que, si nuestras variables aleatorias tienen transformada de Laplace finita en una vecindad del cero, estas cotas sirven.

1.2.2. Teorema de Cramér en $\mathbb{R}^d$

Nuestro interés es tener el PGD para $\mathbb{R}^d$; es por ello que necesitamos la generalización a estas dimensiones. Como vimos en la sección anterior la extensión del Teorema de Cramér a dimensiones mayores no puede ser demostrado usando las técnicas que usamos para el caso en $\mathbb{R}$.

La primera parte la dedicaremos a enunciar y demostrar el teorema de Cramér en $\mathbb{R}^d$, el cual exige que las variables tengan transformada de Laplace finita ($\mathcal{D}_\Lambda = \mathbb{R}^d$). En la última parte de la sección nos dedicaremos a debilitar esta hipótesis. Al debilitarla no tendremos el PGD sino sólo la cota superior de grandes desviaciones. Esto no es del todo malo ya que con esto basta para demostrar la Ley Condicional Débil de los Grandes Números .

Lema 1.11. *Supongamos que $\mathcal{D}_\Lambda = \mathbb{R}^d$. Entonces*

(a) $\Lambda(\cdot)$ *es una función convexa y diferenciable en todas partes y* $\Lambda^*(\cdot)$ *es una función de grandes desviaciones.*

(b) Si $y = \nabla\Lambda(\eta)$ *se tiene que*

$$\Lambda^*(y) = \langle \eta, y \rangle - \Lambda(\eta). \tag{1.26}$$

Demostración. (a) La prueba de la convexidad y diferenciabilidad de Λ es idéntica a la hecha en el Lema 1.9.

Como $\Lambda(0) = 0$ tenemos que Λ^* es no negativa. Ahora si tomamos $\{x_n\}$ una sucesión que converja a x tenemos que

$$\liminf_{n\to\infty} \Lambda^*(x_n) \geq \liminf_{n\to\infty}\{\langle\lambda, x_n\rangle - \Lambda(\lambda)\} = \langle\lambda, x\rangle - \Lambda(\lambda),$$

para cualquier $\lambda \in \mathbb{R}^d$, de donde

$$\liminf_{n\to\infty} \Lambda^*(x_n) \geq \Lambda^*(x),$$

y por lo tanto Λ^* es semicontinua inferiormente.

Notemos que para toda $x \in \mathbb{R}^d$ y $r > 0$,

$$\Lambda^*(x) \geq r\langle x, x\rangle - \sup_{|\lambda|=r} \Lambda(\lambda).$$

En particular se tiene que los conjuntos de nivel $\Psi_{\Lambda^*}(\alpha)$ son acotados y por la semicontinuidad inferior son compactos.

(b) Sea $y = \nabla\Lambda(\eta)$. Fijemos un punto arbitrario $\lambda \in \mathbb{R}^d$ y sea

$$g(\alpha) = \alpha\langle\lambda - \eta, y\rangle - \Lambda(\eta + \alpha(\lambda - \eta)) + \langle\eta, y\rangle, \qquad \alpha \in [0, 1].$$

Dado que Λ es convexa y $\Lambda(\eta)$ es finita, $g(\cdot)$ es cóncava y $|g(0)| < \infty$. Entonces

$$g(1) - g(0) \leq \liminf_{\alpha\downarrow 0} \frac{g(\alpha) - g(0)}{\alpha} = \langle\lambda - \eta, y - \nabla\Lambda(\eta)\rangle = 0.$$

Por lo tanto para toda λ,

$$g(1) = \langle\lambda, y\rangle - \Lambda(\lambda) \leq g(0) = \langle\eta, y\rangle - \Lambda(\eta) \leq \Lambda^*(y),$$

y tomando supremo sobre $\lambda \in \mathbb{R}^d$ de ambos lados, se tiene lo deseado.

□

Observación. Si en este último Lema sustituimos la hipótesis sobre el dominio efectivo de Λ tenemos que los mismos resultados se valen, sólo que en el interior de $\mathscr{D}_\Lambda$.

Este Lema es útil otra vez ya que nos caracteriza el comportamiento de la función Λ^*, el cual nos ayudará en la demostración del teorema de Cramér. Algo más que es importante destacar es que los conjuntos de nivel $\Psi_{\Lambda^*}(\alpha)$ son compactos, por lo que al demostrar que $\{\rho_n\}$ cumple el PGD con FGD Λ^*, automáticamente tendremos que es una buena FGD.

Teorema 1.12 (Cramér). *Supongamos que $\mathcal{D}_\Lambda = \mathbb{R}^d$. Entonces ρ_n satisface el PGD con FGD Λ^*.*

Demostración. Demostraremos que para toda $\delta > 0$ y para cualquier cerrado $F \subseteq \mathbb{R}^d$

$$\limsup_{n\to\infty} \frac{1}{n} \log \rho_n(F) \leq \delta - \inf_{x\in F} I^\delta(x),$$

que es equivalente a demostrar la cota superior de GD y tiene la ventaja de que I^λ es acotada. Fijemos $\Gamma \subseteq \mathbb{R}^d$. Para todo $q \in \Gamma$ tomemos $\alpha_q \in \mathbb{R}^d$ tal que

$$\langle \alpha_q, q\rangle - \Lambda(\alpha_q) \geq I^\delta(q).$$

Esto es posible dada la definición de I^δ. Ahora para cada q tomemos $r_q > 0$ tal que $r_q|\alpha_q| \leq \delta$. Observemos que para todo n y todo $\lambda \in \mathbb{R}^d$ y cualquier conjunto medible $G \subseteq \mathbb{R}^d$ se tiene, por la desigualdad de Tchebyshev, que

$$\rho_n(G) = \mathbb{E}[\mathbf{1}_{\{S_n/n\in G\}}] \leq \mathbb{E}[\exp(\langle \lambda, S_n/n\rangle - \inf_{x\in G}\langle \lambda, x\rangle)],$$

en particular, para cada $n \in \mathbb{N}$ y $q \in \Gamma$

$$\rho_n(B(q, r_q)) \leq \mathbb{E}[\exp(n\langle \alpha_q, S_n/n\rangle) \exp\left(- \inf_{x\in B(q,r_q)} n\langle \alpha_q, x\rangle\right),$$

también para cada $q \in \Gamma$

$$- \inf_{x\in B(q,r_q)} \langle \alpha_q, x\rangle \leq r_q|\alpha_q| - \langle q, \alpha_q\rangle \leq \delta - \langle q, \lambda_q\rangle.$$

Como Γ es compacto, podemos extraer una subcubierta finita de la cubierta $\bigcup_{q\in\Gamma} B(q, r_q)$ que consiste de $N = N(\delta, \Gamma)$ bolas y sus centros $q_1, q_2, \ldots, q_n$ en Γ. Como ρ_n es medida y por la desigualdad anterior se tiene que

$$\begin{aligned}
\frac{1}{n}\log\rho_n(\Gamma) &\leq \frac{1}{n}\log\rho_n\left(\bigcup_{j=1}^{N} B(q_j, r_{q_j})\right) \leq \frac{1}{n}\log\left(\sum_{j=1}^{N}\rho_n(B(q_j, r_{q_j}))\right) \\
&\leq \frac{1}{n}\log N + \delta - \min_{j=1,\ldots,N}\{\langle\lambda_j, q_j\rangle - \Lambda(\lambda_j)\}.
\end{aligned}$$

Por la elección de α_q

$$\limsup_{n\to\infty}\frac{1}{n}\log\rho_n(\Gamma) \leq \delta - \min_{j=1,\ldots,N} I^{\delta}(q_j),$$

de donde tenemos la cota superior para compactos.

Ahora para extender esta cota a conjuntos cerrados en $\mathbb{R}^d$, mostraremos que $\{\rho_n\}$ es una familia de medidas con tensión exponencial y aplicando el Lema 1.7 tendremos el resultado.

Sea $H_M = [-M, M]^d$, como $H_M^c = \bigcup_{j=1}^{d}\{x : |x_j| > M\}$. Tenemos

$$\rho_n(H_M^c) \leq \sum_{j=1}^{d}\rho_n^j([M,\infty)) + \sum_{j=1}^{d}\rho_n^j((-\infty, -M]), \tag{1.27}$$

donde ρ_n^j, $j = 1, 2, \ldots, d$, son las leyes marginales de los vectores S_n/n. Aplicando las desigualdades del caso unidimensional para cualquier $M \geq |\bar{x}|$,

$$\rho_n^j((-\infty, -M]) \leq e^{-n\Lambda_j^*(-M)} \quad ; \quad \rho_n^j([M,\infty)) \leq e^{-n\Lambda_j^*(M)},$$

donde Λ_j^* es la transformada de Legendre–Fenchel de $\log\mathbb{E}[e^{\lambda X_1^j}]$, $j = 1, 2, \ldots, d$; además es fácil ver que $\Lambda_j^*(x)$ tiende a infinito cuando hacemos $|x|$ tender a infinito. Aplicando la cota anterior y la desigualdad (1.27), y haciendo tender n a infinito y después M, tenemos la siguiente identidad:

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\rho_n(H_M^c) = -\infty.$$

Por consecuencia $\{\rho_n\}$ es una familia de medidas con tensión exponencial ya que los cubos H_M son compactos.

Para la cota inferior es suficiente probar que para toda $y \in \mathscr{D}_{\Lambda^*}$ y toda $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n(B(y,\delta)) \geq -\Lambda^*(y).$$

Supongamos primero que $y = \nabla\Lambda(\eta)$ para alguna $\eta \in \mathbb{R}^d$; definamos la medida $\tilde{\rho}$ como

$$\tilde{\rho}(dz) = e^{\langle \eta, z\rangle - \Lambda(\eta)}\rho(dz),$$

sea $\tilde{\rho}_n$ la ley de S_n/n, donde las $\tilde{X}_i$ son i.i.d. con ley $\tilde{\rho}$. Entonces

$$\begin{aligned}
\frac{1}{n}\log \rho_n(B(y,\delta)) &= \Lambda(\eta) - \langle \eta, y\rangle + \frac{1}{n}\int_{\{z\in B(y,\delta)\}} e^{\langle \eta, y-z\rangle}\tilde{\rho}_n(dz) \\
&\geq \Lambda(\eta) - \langle \eta, y\rangle - |\eta|\delta + \frac{1}{n}\log \tilde{\rho}_n(B(y,\delta)).
\end{aligned}$$

Aplicando el Teorema de Convergencia Dominada para intercambiar el orden de diferenciación e integración

$$\mathbb{E}_\eta[X_1] = \frac{1}{\mathbb{E}[e^{\langle \eta, X_1\rangle}]}\int_{\mathbb{R}^d} x e^{\langle \eta, x\rangle}\rho(dx) = \nabla\Lambda(\eta) = y.$$

De esta relación y la Ley Débil de los Grandes Números tenemos que $\lim_{n\to\infty} \tilde{\rho}_n(B(y,\delta)) = 1$ para todo $\delta > 0$. Más aún, como $\Lambda(\eta) - \langle \eta, y\rangle \geq -\Lambda^*(y)$ por la desigualdad anterior se tiene la cota inferior

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n(B(y,\delta)) \geq -\Lambda^*(y) - |\eta|\delta.$$

Entonces

$$\liminf_{n\to\infty} \frac{1}{n} \log \rho_n(B(y,\delta)) \geq \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \rho_n(B(y,\delta)) \geq -\Lambda^*(y).$$

Para extender este resultado para cubrir $y \in \mathscr{D}_{\Lambda^*}$ tal que $y \notin \{\nabla\Lambda(\alpha) : \alpha \in \mathbb{R}^d\}$, regularizaremos ρ: agregaremos a cada X_j una variable aleatoria Normal. Más específicamente fijamos $M < \infty$ y sea ν la ley marginal de los vectores aleatorios i.i.d. $Y_j = X_j + V_j/\sqrt{M}$ donde las V_j son variables aleatorias normales multivariadas $(0, \mathbf{Id})$ en $\mathbb{R}^d$ e independientes de $X_1, X_2, \ldots$. Sea $\Lambda_M(\cdot)$ la función generadora de de momentos logarítmicos de Y_1, mientras que ν_n denota la ley de $S_n^{(M)}/n = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

Recordando que la función generadora de momentos de cada V_j es $e^{|\lambda|^2/2}$, se tiene que

$$\Lambda_M(\lambda) = \Lambda(\lambda) + \frac{1}{2M}|\lambda|^2 \geq \Lambda(\lambda),$$

y entonces

$$\Lambda^*(y) = \sup_{\lambda\in\mathbb{R}^d}\langle\lambda, y\rangle - \Lambda(\lambda) \geq \sup_{\lambda\in\mathbb{R}^d}\langle\lambda, y\rangle - \Lambda_M(\lambda).$$

Ahora, como $\mathscr{D}_\Lambda = \mathbb{R}^d$, se sigue que $\bar{x} = \mathbb{E}[X_1]$ es finita; entonces aplicando la desigualdad de Jensen se tiene que $\Lambda(\alpha) \geq \langle\alpha, \bar{x}\rangle$ para toda $\alpha \in \mathbb{R}^d$. Por lo cual la función

$$g(\alpha) = \langle\alpha, y\rangle - \Lambda_M(\alpha),$$

es finita y diferenciable; además satisface

$$\lim_{r\to\infty} \sup_{|\alpha|>r} g(\alpha) = -\infty.$$

En consecuencia, el supremo de $g(\cdot)$ en $\mathbb{R}^d$ es alcanzado en algún η finito, para el cual se cumple

$$0 = \nabla g(\eta) = y - \nabla\Lambda_M(\eta).$$

Entonces, por la prueba anterior, la cota inferior para $\{\nu_n\}$ se tiene para todo $\delta > 0$, es decir

$$\liminf_{n\to\infty} \frac{1}{n}\log \nu_n(B(y,\delta)) \geq -\Lambda^*(y) > -\infty.$$

Notemos que $S_n^{(M)}/n$ tiene la misma distribución que $S_n/n + V/\sqrt{nM}$ donde V es una normal multivariada $(0, \mathbf{Id})$ y es independiente de S_n/n. Por lo tanto

$$\rho_n(B(y, 2\delta)) \geq \nu_n(B(y,\delta)) - \mathbb{P}(|V| \geq \sqrt{nM}\delta).$$

Por último, como V es una variable normal multivariada estándar se tiene que

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|V| \geq \sqrt{nM}\delta) \leq -\frac{M\delta^2}{2}.$$

Combinando las desigualdades anteriores, haciendo tender n a infinito y después M concluimos la prueba. □

Observación. Es importante notar que para demostrar la cota superior para compactos no utilizamos el hecho de que $\mathscr{D}_\Lambda = \mathbb{R}^d$, sin embargo sí lo hicimos para incluir el caso de cerrados y también para demostrar la cota inferior. Ahora queremos mejorar nuestra demostración.

Debilitaremos la hipótesis sobre el dominio tan sólo pidiendo que $0 \in \mathscr{D}_\Lambda^\circ$.

Proposición 1.13. *Sea $\{X_n\} \in \mathbb{R}^d$ una sucesión de variables i.i.d con ley ρ y sea ρ_n la ley de la media muestral. Además supongamos que $0 \in \mathscr{D}_\Lambda^\circ$. Entonces para cualquier $F \in \mathcal{F}$ se tiene que*

$$\limsup_{n\to\infty} \frac{1}{n} \log \rho_n(F) \leq -\inf_{x\in F} \Lambda^*(x) = -\Lambda^*(F). \tag{1.28}$$

Demostración. La demostración para conjuntos compactos es idéntica a la hecha anteriormente por lo cual no la hacemos aquí. Extendemos la demostración a los conjuntos cerrados, mostrando que ρ_n es una familia con tensión exponencial. Denotamos por $\mathbf{e}_j$ al j–ésimo vector canónico de $\mathbb{R}^d$. En particular como tenemos que $0 \in \mathscr{D}_\Lambda^\circ$, se sigue que existen $\eta_j > 0$ y $\theta_j > 0$ tales que $\Lambda(\theta_j \mathbf{e}_j) < \infty$ y $\Lambda(-\eta_j \mathbf{e}_j) < \infty$, para $j = 1, 2, \ldots, d$. Ahora aplicando la desigualdad de Tchebyshev a las leyes marginales de los vectores S_n/n tenemos que

$$\begin{aligned}
\rho_n^j((-\infty, r]) &\leq \exp\left(-n\eta_j r + \Lambda_n(-n\eta_j \mathbf{e}_j)\right), \\
\rho_n^j([r, \infty)) &\leq \exp\left(-n\theta_j r + \Lambda_n(n\theta_j \mathbf{e}_j)\right).
\end{aligned}$$

de donde

$$\begin{aligned}
\lim_{r\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \rho_n^j((-\infty, -r]) &= -\infty, \\
\lim_{r\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \rho_n^j([r, \infty)) &= -\infty,
\end{aligned}$$

y como $([-r, r])^c = \bigcup_{i=1}^d ((-\infty, -r] \cup [r, \infty))$,

$$\lim_{\rho\to\infty} \limsup_{n\to\infty} \log \mu_n(([-r, r]^d)^c) = -\infty. \tag{1.29}$$

Por lo tanto se tiene que $\{\rho_n\}$ es una familia con tensión exponencial, y por el Lema 1.7 se tiene el resultado. □

Observación. Se puede demostrar con las mismas hipótesis de esta proposición que $\{\rho_n\}$ no sólo cumple la cota superior sino que cumple la PGD con FGD Λ^* (ver por ejemplo [DZ98]).

1.2.3. Ley Condicional Débil de los Grandes Números

La Ley Débil de los Grandes Números nos dice que, dada una sucesión de variables aleatorias $\{X_n\}$ independientes e idénticamente distribuidas con media $\bar{x}$, y $\epsilon > 0$, se tiene que

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \bar{x}\right| > \epsilon\right) = 0.$$

Pero ¿qué pasa si condicionamos a que nuestros promedios estén en un conjunto B tal que $\bar{x} \notin \bar{B}$?. ¿Existirá un punto que juegue el papel de la media?.

Este problema se resuelve por la Ley Condicional Débil de los Grandes Números [Num89], la cual muestra la convergencia en probabilidad de la media muestral a un punto, el cual se conoce como *punto dominante* [Ney83, Ney84].

Lo primero que haremos es discutir un poco sobre el *punto dominante*. Antes de seguir adelante denotemos por

$$\begin{aligned} \mathscr{S} &= \text{la cerradura convexa del soporte de } \rho \\ H^+(\alpha, v) &= \{x \in \mathbb{R}^d : \langle x, \alpha\rangle \geq \langle v, \alpha\rangle\} \end{aligned}$$

Notemos que $H^+(\alpha, v)$ correponde al semiespacio derecho delimitado por el plano $\langle x, \alpha\rangle = \langle v, \alpha\rangle$.

Definición 1.14. Un punto $v_B \in \mathbb{R}^d$es un punto dominante del conjunto B y de la función Λ si

(a) $\nu_B \in \partial B$.

(b) La ecuación en λ, $\nabla\Lambda(\lambda) = \nu_B$ tiene una única solución $\lambda_{\nu_B} \in \mathscr{D}_\Lambda$ y

(c) $B \subseteq H^+(\lambda_{\nu_B}, \nu_B)$.

Puede que no exista un *punto dominante* para un conjunto dado. Por esto, es necesario buscar condiciones que garanticen la existencia del mismo. El siguiente Teorema da condiciones suficientes aunque no necesarias para la existencia del punto dominante.

Teorema 1.15. *Supongamos que $\mathscr{D}_\Lambda$ es un abierto que contiene una vecindad del origen, B es un conjunto convexo tal que $[B \cap \mathscr{S}]^\circ \neq \emptyset$ y $\bar{x} = \int xp(dx) \notin \bar{B}$. Entonces existe un único punto dominante ν_B para Λ y B.*

La demostración de este teorema la podemos consultar en los artículos de Ney [Ney83, Ney84]. En el primero se ve de una manera constructiva y en el segundo se ve que la existencia del punto dominante es una consecuencia de la convexidad de Λ. En el primer artículo también se presenta la siguiente propiedad: Si para un conjunto B existe el punto dominante entonces $\Lambda^*(\nu_B) = \Lambda^*(B)$, y en este caso a $\Lambda^*(\nu_B)$ lo llamaremos la entropía de B, es decir, ν_B *es el punto cuya entropía es igual a la entropía de* B, o el punto donde la entropía es mínima.

Teorema 1.16 (Ley Condicional Débil de los Grandes Números). *Sea $B \subseteq \mathbb{R}^d$ un conjunto boreliano tal que $[B \cap \mathscr{D}_\Lambda]^\circ \neq \emptyset$ y sea $\bar{x} \notin \bar{B}$. Entonces para toda $\epsilon > 0$ existe una constante $I(\epsilon) > 0$ tal que*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \nu_B\right| > \epsilon \,\middle|\, \frac{S_n}{n} \in B\right) \leq e^{-nI(\epsilon)}; \qquad \textit{si } n \textit{ es suficientemente grande.} \tag{1.30}$$

A lo largo de la demostración usaremos la norma $\|x\|_\infty = \text{máx}_{1\leq i\leq d} |x_i|$ que define la misma topología que la métrica euclideana.

Demostración. Dada $\epsilon > 0$; definimos $B^{(\epsilon)} = B \cap \{v \in \mathbb{R}^d : \|v - v_B\|_\infty > \epsilon\}$, y

$$\begin{aligned} B_{i_+} &= B \cap \{v \in \mathbb{R}^d : v_i - (v_B)_i > \epsilon\}, \\ B_{i_-} &= B \cap \{v \in \mathbb{R}^d : v_i - (v_B)_i > -\epsilon\}. \end{aligned}$$

Tenemos que $B^{(\epsilon)} = \bigcup_{i=1}^d B_i$, donde $B_i = B_{i_+} \cup B_{i_-}$; también cada B_i es un convexo. Por otro lado, como $B_i \subset B$ y de la unicidad de v_B se sigue que

$$\Lambda^*(v_{B_i}) = \Lambda^*(B_i) > \Lambda^*(B) = \Lambda^*(v_B); \tag{1.31}$$

por la misma razón se tiene que $\Lambda^*(v_{B^{(\epsilon)}}) > \Lambda^*(v_B)$.

Por último

$$\begin{aligned} \limsup_{n\to\infty} \quad & \frac{1}{n} \log \mathbb{P}\left(\left| \frac{S_n}{n} - v_B \right| > \epsilon \middle| \frac{S_n}{n} \in B \right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \frac{\mathbb{P}\left(\left| \frac{S_n}{n} - v_B \right| > \epsilon, \frac{S_n}{n} \in B \right)}{\mathbb{P}\left(\frac{S_n}{n} \in B \right)} \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \frac{\mathbb{P}\left(\frac{S_n}{n} \in B^{(\epsilon)} \right)}{\mathbb{P}\left(\frac{S_n}{n} \in B \right)} \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in B^{(\epsilon)} \right) - \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \in B \right) \\ &\leq -\Lambda^*(B^{(\epsilon)}) + \Lambda^*(B) < 0. \end{aligned}$$

donde la útima desigualdad se tiene por la cota superior de grandes desviaciones (1.28). □

Decimos que $\{W_n\}_{n\in\mathbb{N}}$, una sucesión de variables aleatorias *converge exponencialmente rápido a la variable aleatoria W_∞ con respecto a las medidas* $\{\mathbb{P}_n\}$, si para cada $\epsilon > 0$ existe un constante $I = I(\epsilon)$, tal que

$$\mathbb{P}_n(|W_n - W_\infty| > \epsilon) < e^{-nI(\epsilon)}$$

si n es suficientemente grande, y lo denotamos por $W_n \overset{\exp}{\to} W_\infty$ (c.r.a $\mathbb{P}_n$). Notemos también que si $\{W_n\}$ converge exponencialmente a W_∞, se tiene que

$$\lim_{n\to\infty} \mathbb{P}_n(|W_n - W_\infty| > \epsilon) = 0. \tag{1.32}$$

esto da lugar al siguiente resultado.

Lema 1.17. *Sea $\{W_n\}$ una sucesión de variables aleatorias uniformemente acotadas que converge exponencialmente rápido a W_∞ con respecto a las medidas $\{\mathbb{P}_n\}$. Entonces*

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[|W_n - W_\infty|] = 0. \tag{1.33}$$

Demostración. Primero notemos que como la sucesión $\{W_n\}$ es uniformemente acotada, existe $K > 0$ tal que

$$|W_n - W_\infty| \leq K;$$

de donde se deduce que $\mathbb{P}_n(|W_n - W_\infty| > \epsilon) = 0$ para $\epsilon > K$. Por otra parte también se tiene que

$$\int_0^\infty \lim_{n\to\infty} \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon = \int_0^K \lim_{n\to\infty} \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon = 0;$$

por (1.32). Como $\mathbb{P}_n$ es una medida de probabilidad, para toda n, podemos intercambiar el límite y la integral (Teorema de Convergencia Dominada)

$$\begin{aligned} 0 = \int_0^K \lim_{n\to\infty} \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon &= \lim_{n\to\infty} \int_0^K \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon \\ &= \lim_{n\to\infty} \int_0^\infty \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon. \end{aligned}$$

Recordemos que si $Y \geq 0$ es una variable aleatoria entonces

$$\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > y)dy.$$

Por lo tanto

$$0 = \lim_{n\to\infty} \int_0^\infty \mathbb{P}_n(|W_n - W_\infty| > \epsilon)d\epsilon = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[|W_n - W_\infty|].$$

□

Podemos ver que la convergencia exponencial es muy parecida a la convergencia en probabilidad (es claro que la convergencia exponencial no es precisamente la convergencia en probabilidad) de una sucesión de variables aleatorias y al igual que en ésta, si tenemos manera de dominar la sucesión tenemos algo parecido a la convergencia en L_1.

Capítulo 2

Teoremas de convergencia para las distribuciones microcanónicas y canónicas.

En este capítulo presentaremos una serie de teoremas y propiedades que nos ayudarán en el trabajo posterior: la aplicación de la teoría a la Mecánica Estadística. Todo el trabajo que haremos en esta sección es consecuencia, principalmente, de la Ley Condicional Débil de Nummelin [Num89].

En la primera sección definiremos que es una *observable*, así como también otra serie de conceptos que usaremos en la parte de las aplicaciones.

En la segunda sección nos enfocaremos en la convergencia de cierto tipo de distribuciones a lo que llamaremos la *distribución canónica*. Cabe destacar que la distribución canónica estará estrechamente relacionada con el concepto de punto dominante [Ney83].

2.1. Distribuciones microcanónicas, canónicas y empíricas

Tomemos $X_1, X_2, \ldots$ una sucesión de variables aletorias independientes e idénticamente distribuidas con ley P, con valores en un espacio de medida $(E \subseteq \mathbb{R}^d, \mathcal{F})$; pensaremos que cada X_i representa el estado de la i–ésima partícula de un sistema. Podemos interpretar a una función de E en $\mathbb{R}^d$ como una observable.

Definición 2.1. Una función $f : E \to \mathbb{R}^d$ medible se le llama observable.

Ahora supongamos que tenemos una observable f tal que $f = (g, u)$ donde $g : E \to \mathbb{R}^{d_1}$ y $u : E \to \mathbb{R}^{d_2}$, $d = d_1 + d_2$. Supongamos que lo que podemos medir es la media muestral de una observable; bajo este hecho obtendremos información acerca de los estados $X_1, X_2, \ldots$ tan sólo ocupando esta medición, es decir

$$\frac{U_n}{n} = \frac{\sum_{i=1}^n u(X_i)}{n} \in C$$

Para ser más exactos, si tenemos un convexo no vacío $C \in \mathbb{R}^{d_1}$ podemos preguntarnos qué pasa con las siguientes probabilidades

$$P_{n,C} = \mathbb{P}(\,\cdot \mid U_n/n \in C).$$

A estas medidas las llamaremos *probabilidades microcanónicas*. Entonces tenemos las distribuciones condicionales inducidas

$$P_{n,C}(dx) = \mathbb{P}(X_1 \in dx \mid U_n/n \in C),$$

que llamaremos las *distribuciones microcanónicas (de los estados)*. Notemos que $P_{n,C}(dx) = \mathbb{P}(X_i \in dx \mid U_n/n \in C)$ para cualquier $i = 2, 3, \ldots, n$ porque las variables $\{X_i\}$ son intercambiables, a consecuencia de ser i.i.d., por lo cual se tiene que

$$P_{n,C}(dx) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(X_i \in dx \mid U_n/n \in C). \tag{2.1}$$

Dado $\beta \in \mathbb{R}^{d_2}$, podemos evaluar la función generadora de momentos de $u(X_1)$ como

$$M(\beta) = \mathbb{E}[\exp(\langle \beta, u(X_1) \rangle)],$$

entonces tenemos que $M : \mathbb{R}^{d_2} \to \mathbb{R}$. Ahora, si M es finita podemos definir la distribución de Gibbs con dirección β como

$$P_\beta(dx) = \frac{1}{Z}\exp(\langle\beta, u(x)\rangle)P(dx).$$

A P_β le llamamos la *distribución canónica* y a $Z = M(\beta)$ la *función de partición*.

También consideraremos la *ditribución empírica*

$$\hat{P}_n(\omega, dx) = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}(dx),$$

donde δ_{X_i} es la Delta de Dirac.

En la siguiente sección veremos, bajo ciertas condiciones de regularidad, que

$$\lim_{n\to\infty}\hat{P}_n(dx) = P_\beta(dx), \qquad \text{para alguna } \beta \in \mathbb{R}^d,$$

exponencialmente con respecto a las probabilidades microncanónicas. Es importante notar que la distribución empírica es una medida aleatoria, es por eso que la convergencia de $\hat{P}$ es con respecto a ciertas medidas de probabilidad. Después veremos que

$$\lim_{n\to\infty}P_{n,C}(dx) = P_\beta(dx)$$

donde β es la misma que en el caso anterior, y además especificaremos de modo único quien es β.

2.2. Convergencia a la distribución canónica

Sean $X_1, X_2, \ldots$ los estados de las partículas y sea $f = (g, u)$ una observable. Para cada $v \in \mathbb{R}^d$ podemos partir a $v = (v_1, v_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $d = d_1 + d_2$. En esta parte aplicaremos los resultados del capítulo anterior a las variables $\xi_i = f(X_i)$ y a su función generadora de momentos $\Lambda(\lambda) = \log\mathbb{E}[e^{\langle\lambda,\xi_1\rangle}]$, $\lambda \in \mathbb{R}^d$, y también a $u(X_i)$ y su respectiva función generadora de momentos, $\Lambda_U(\beta) = \log \mathbb{E}[e^{\langle \beta, u(X_1)\rangle}]$. Pero antes de hacer esto presentamos una breve discusión sobre un tipo de medidas de Gibbs que nos servirá más adelante; en particular podemos ver que la distribución canónica es una medida de Gibbs.

Recordemos que en el Lema 1.11 vimos que $\Lambda(\cdot)$ es diferenciable, al menos en $\mathscr{D}_\Lambda$ según la observación, por lo cual

$$\frac{\partial \Lambda}{\partial \lambda_i}(\lambda) = \frac{1}{M(\lambda)} \frac{\partial}{\partial \lambda_i} \int e^{\langle \lambda, x\rangle} P(dx) = \frac{1}{M(\lambda)} \int x_i e^{\langle \lambda, x\rangle} P(dx),$$

(donde el intercambio de la derivada y la integral lo podemos hacer por el Teorema de Convergencia Dominada) y así tenemos que

$$\nabla \Lambda(\lambda) = \frac{1}{M(\lambda)} \int x e^{\langle \lambda, x\rangle} P(dx) = \int x e^{\langle \lambda, x\rangle - \Lambda(\lambda)} P(dx).$$

Tomando $P_\lambda(dx) = e^{\langle \lambda, x\rangle - \Lambda(\lambda)} P(dx)$

$$\nabla \Lambda(\lambda) = \int x P_\lambda(dx) = \mathbb{E}_\lambda[X] = m(\lambda),$$

donde la penúltima igualdad se da si X tiene a P_λ como ley . Escribiendo con esta notación el Lema 1.11 tenemos que

$$\Lambda^*(y) = \langle y, m^{-1}(y)\rangle - \Lambda(m^{-1}(y)), \tag{2.2}$$

siempre que exista $m^{-1}(y)$; en particular dado una conjunto B donde el punto dominante exista se tiene que $\lambda_{v_B} = m^{-1}(v_B)$.

Supongamos que el dominio efectivo de Λ_u, $\mathscr{D}_u$, es abierto y que g es una función acotada. Entonces el dominio efectivo de Λ es $\mathscr{D}_\Lambda = \mathbb{R}^{d_1} \times \mathscr{D}_u$, que es un abierto de $\mathbb{R}^d$.

Sea $\alpha \in \mathbb{R}^{d_2}$ tal que $\beta = m_u^{-1}(\alpha)$ exista, definimos $t_\alpha = (0, \beta)(\in \mathbb{R}^d)$ y $v_\alpha = m(t_\alpha)$, entonces

$$\Lambda^*(v_\alpha) = \Lambda_U^*(v_\alpha) = \inf_{v\,:\, v_2 = \alpha} \Lambda^*(v).$$

Ya que $v \in \mathbb{R}^d$, tal que $v_2 = \alpha$, se tiene que

$$\Lambda^*(v) \geq \langle t_\alpha, v\rangle - \Lambda(t_\alpha) = \langle \beta, \alpha\rangle - \Lambda_U(\beta) = \Lambda_U^*(v_\alpha)$$

y se da la igualdad cuando $\nu = \nu_\alpha$.

Ahora tomemos $C \in \mathbb{R}^{d_2}$ un boreliano convexo tal que $[\mathscr{S}_U \cap C]^\circ \neq \emptyset$ y $m_U(0) = \mathbb{E}[U(X_1)] \notin \bar{C}$. Entonces existe α_C un punto dominante para Λ_u y C.

Sea $B = \mathbb{R}^{d_1} \times C$ entonces

$$\Lambda^*(B) = \inf_{\nu\in B} \Lambda^*(\nu) = \inf_{\alpha\in C} \inf_{\nu_1\in\mathbb{R}^{d_1}} \Lambda^*(\nu_1,\alpha) \;\; = \;\; \inf_{\alpha\in C} \Lambda^*(\nu_\alpha) = \inf_{\alpha\in C} \Lambda_U^*(\alpha).$$

Pero el $\inf_{\alpha\in C} \Lambda_u^*(\alpha) = \Lambda_u^*(C) = \Lambda_u^*(\alpha_C)$, por ser α_C el punto dominante. Entonces

$$\nu_B = \nu_{\alpha_C} = m(t_{\alpha_C}) = m(0,\beta),$$

donde $\beta = m_u^{-1}(\alpha_C)$ que sabemos que existe, por lo cual

$$\alpha_C = m_u(\beta) = \mathbb{E}_\beta(u(X_1)) = \int u(x)P_\beta(dx).$$

Por último

$$(\nu_B)_2 = \mathbb{E}_\beta[U(X_1)] = \alpha_C,$$

y

$$(\nu_B)_1 = \mathbb{E}_\beta[g(X_1)] = \int g(x)P_\beta(dx).$$

Teorema 2.2. *Supongamos que $\mathscr{D}_u$ es abierto y no vacío, g es una función acotada, y C es un boreliano convexo tal que $[C \cap \mathscr{S}_u]^\circ \neq \emptyset$ y $m_u(0) = \mathbb{E}[u(X_1)] \notin \bar{C}$. Sea P_β la distribución canónica con $\beta = m^{-1}(\alpha_C)$, α_C el punto dominante para Λ_U y C. Entonces*

$$\frac{G_n}{n} \xrightarrow{exp} \int g(x)P_\beta(dx) = \mathbb{E}_\beta[g(X_1)] \tag{2.3}$$

con respecto a las probabilidades microcanónicas $P_{n,C}(\cdot) = \mathbb{P}\left(\cdot \mid \frac{U_n}{n} \in C\right)$, donde $U_n = \sum_{i=1}^n u(X_i)$ y $G_n = \sum_{i=1}^n g(X_i)$.

Demostración. Si $\frac{U_n}{n} \in C$ se sigue que $\frac{F_n}{n} = \frac{1}{n}\sum_{i=1}^{n}\xi_i \in B = \mathbb{R}^{d_1} \times C$. Entonces por la LCDGN (Teo. 1.16), tenemos que $\frac{F_n}{n}$ converge exponencialmente a ν_B, el punto dominante de B, con respecto a $\mathbb{P}(\cdot \mid \frac{F_n}{n} \in B)$. Es decir $\frac{(G_n,U_n)}{n}$ converge exponencialmente a ν_B con respecto a $\mathbb{P}(\cdot \mid \frac{G_n}{n} \in \mathbb{R}^{d_1}, \frac{U_n}{n} \in C)$, de donde $\frac{G_n}{n}$ converge exponencialmente a $(\nu_B)_1$ y $\frac{U_n}{n}$ converge exponencialmente a $(\nu_B)_2$, ambos respecto a $\mathbb{P}(\cdot \mid \frac{G_n}{n} \in \mathbb{R}^{d_1}, \frac{U_n}{n} \in C)$.

Por lo tanto $(\nu_B)_2 = \alpha_C$ y entonces $\frac{G_n}{n}$ converge exponencialmente a $\mathbb{E}_\beta[g(X_1)]$ con respecto a $\mathbb{P}\left(\cdot \mid \frac{U_n}{n} \in C\right)$, donde $\beta = m_u^{-1}(\alpha_C)$. □

Definición 2.3. Sean $\mu, \mu_1, \mu_2, \ldots$ medidas de probabilidad en $(\mathcal{X}, \mathcal{F})$. Decimos que la sucesión $\{\mu_n\}$ *b*–converge a μ si la sucesión $\{\int g\, d\mu_n\}$ converge a $\int g\, d\mu$ conforme n tiende a infinito, para toda $g : \mathcal{X} \to \mathbb{R}^d$ medible y acotada, y lo denotamos por $\mu_n \xrightarrow{b} \mu$

Definición 2.4. Sean $\pi_1, \pi_2, \ldots$ medidas aleatorias de probabilidad en $(\mathcal{X}, \mathcal{F})$. Decimos que la sucesión $\{\pi_n\}$ converge exponencialmente a la medida π (no aleatoria) si la sucesión de variables aleatorias $\int g\, d\pi_n$ converge exponencialmente a $\int g\, d\pi$ con respecto a $\{\mathbb{P}_n\}$, para toda $g : \mathcal{X} \to \mathbb{R}^d$ medible y acotada.

Recordemos que en la sección anterior definimos las distribuciones empíricas de una sucesión de variables aleatorias, $\hat{P}_n(dx) = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(dx)$, las cuales son medidas de probabilidad aleatorias y discretas. Dada $g : E \to \mathbb{R}^d$ $\mathcal{F}-$ medible se tiene que

$$\int g(x)\, \hat{P}_n(dx) = \frac{\sum_{i=1}^{n} g(X_i)}{n} = \frac{G_n}{n}.$$

Entonces por el Teorema 2.2 sabemos que bajo ciertas condiciones sobre la función g y el conjunto C,

$$\int g(x)\, \hat{P}_n(dx) = \frac{G_n}{n} \xrightarrow{\exp} \int g(x)\, P_\beta(dx) \qquad (\text{c.r.a } P_{n,C}(\cdot)),$$

donde $\beta = m_u^{-1}(\alpha_C)$ y α_C el punto dominante para Λ_u y C. Por esta breve discusión tenemos el siguiente Corolario:

Corolario 2.5. *Supongamos que $\mathscr{D}_u$ es abierto y no vacío, C un boreliano convexo tal que $[C \cap \mathscr{S}_u]^\circ \neq \emptyset$ y $m_u(0) = \mathbb{E}[u(X_1)] \notin \bar{C}$. Entonces $\{\hat{P}_n\}$ converge exponencialmente a P_β donde $\beta = m_u^{-1}(\alpha_C)$ con respecto a $P_{n,C}(\cdot)$.*

La demostración es consecuencia inmediata del Teorema 2.2 y la discusión antes hecha. Por último, veremos qué pasa con las distribuciones microcanónicas $P_{n,C}(dx)$. Esto lo haremos en el siguiente resultado:

Corolario 2.6. *Supongamos que $\mathscr{D}_u$ es abierto y C es un boreliano convexo. Entonces las distribuciones microcánonicas $P_{n,C}(dx) = \mathbb{P}\left(X_1 \in dx \,\middle|\, \frac{U_n}{n} \in C\right)$ b–convergen a la distribución canónica P_β con $\beta = m_u^{-1}(\alpha_C)$.*

Demostración. Por el Corolario 2.5 se tiene que

$$\int g(x)\,\hat{P}_n(dx) \stackrel{\exp}{\longrightarrow} \int g(x)\,P_\beta(dx) \qquad (\text{c.r.a } P_{n,C}(\cdot)),$$

para toda $g : E \to \mathbb{R}^d$ acotada. La acotación de g nos implica existe $K > 0$ tal que

$$\left|\int g(x)\,\hat{P}_n(dx) - \int g(x)\,P_\beta(dx)\right| \leq K.$$

Por lo tanto, por el Lema 1.17 se tiene que

$$\mathbb{E}_{P_{n,C}}\left[\left|\int g(x)\,\hat{P}_n(dx) - \int g(x)\,P_\beta(dx)\right|\right] \to 0 \qquad \text{si } n \to \infty,$$

de donde

$$\begin{aligned} 0 \;\leq\; & \lim_{n\to\infty}\left|\mathbb{E}_{P_{n,C}}\left[\int g(x)\,\hat{P}_n(dx) - \int g(x)\,P_\beta(dx)\right]\right| \\ & \leq \lim_{n\to\infty}\mathbb{E}_{P_{n,C}}\left[\left|\int g(x)\,\hat{P}_n(dx) - \int g(x)\,P_\beta(dx)\right|\right] = 0. \end{aligned}$$

Por último notemos que

$$\begin{aligned} \mathbb{E}_{P_{n,C}}\left[\int g(x)\,\hat{P}_n(dx)\right] &= \mathbb{E}_{P_{n,C}}\left[\frac{\sum_{i=1}^n G(X_i)}{n}\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{P_{n,C}}\left[G(X_i)\right] \\ &= \frac{1}{n}\sum_{i=1}^n \int g(x_i)P_{n,C}(dx_i) = \int g(x)P_{n,C}(dx), \end{aligned}$$

donde la última igualdad se da por la relación (2.1). □

Resumiendo, lo que hemos visto: si tenemos $u : E \to \mathbb{R}^d$ tal que $\mathscr{D}_u$ es abierto y no vacío, y C un boreliano convexo, entonces

(a) $G_n/n \xrightarrow{\text{exp}} \int g(x)P_\beta(dx) = \mathbb{E}_\beta[g(X_1)]$ con respecto a $P_{n,C}$, para toda función $g : E \to \mathbb{R}^p$ acotada

(b) $\hat{P}_n(dx) \xrightarrow{\text{exp}} P_\beta(dx)$ con respecto a $P_{n,C}$

(c) $P_{n,C}(dx) \xrightarrow{\text{b}} P_\beta(dx)$

donde β está definido de manera única como $\beta = m_u^{-1}(\alpha_C)$ y α_C el punto dominante de (C, Λ).

Capítulo 3

Distribución canónica y gran canónica para el gas ideal.

En este capítulo aplicaremos los resultados de grandes desviaciones antes expuestos para analizar un modelo básico de la mecánica estadística. Esta rama de la física aplica la teoría de la probabilidad para estudiar propiedades de sistemas en equilibrio que consisten de un gran número de partículas. Aquí se presentará un modelo muy simple llamado el gas ideal. Este modelo, en el cual se supone que no hay interacciones entre las partículas, es el equivalente físico a que las variables aleatorias sean i.i.d..

La descripción macroscópica de un sistema físico como el gas ideal lo da la termodinámica. La termodinámica reúne las propiedades del gas en términos de sus variables macroscópicas como presión, volumen, temperatura y energía interna. Pero esta teoría no tiene en cuenta que el gas está hecho de pequeñas moléculas. La mecánica estadística tiene como objetivo obtener las propiedades del gas a partir de una distribución de probabilidad que describe el comportamiento microscópico de éste.

3.1. Distribución canónica para el gas ideal.

El modelo del gas ideal es una idealización del comportamiento límite de gases suficientemente diluídos. Un imagen microscópica del gas ideal es la siguiente: *es un gas hecho de moléculas que no interactúan entre si y cuya energía es sólo cinética.* Nosotros estaremos interesados en calcular la distribución canónica de este sistema, la cual contiene información del comportamiento de las moléculas que conforman el gas.
Lo primero que haremos es considerar un sistema físico estacionario con un gran número de partículas indistinguibles con estados $X_1, X_2, \ldots, X_n$. Supongamos que tenemos una observable u escalar y no negativa. Intepretaremos $u(X_i)$ como la energía de la i-ésima partícula. Supongamos que tenemos una observación del promedio de la energía del sistema, es decir $U_n/n \in (\alpha_1, \alpha_2)$ que supondremos contenido en $(\mathscr{S}_u)^\circ \subseteq (0, \infty)$.
Consideraremos a $E \subseteq \mathbb{R}^6$ el espacio de estados y P la tomaremos (esencialmente) como la medida de Lebesgue, ya que aunque ésta no es una medida de probabilidad, tiene muchas de las características que necesitamos como la invarianza temporal (Teorema de Liouville) y espacial. Además, si nos restringimos a un conjunto de medida finita tendremos en cierto sentido una distribución uniforme. También observemos que independientemente de la medida P, las medidas de Gibbs P_η son medidas de probabilidad.
Notemos que la función Λ_u^* es decreciente en $(\mathscr{S}_u)^\circ$, ya que si tomamos $\alpha \in (\mathscr{S}_u)^\circ$ tenemos que

$$\begin{aligned}
(\Lambda_u^*)'(\alpha) &= (\alpha m_u^{-1}(\alpha))' - (\Lambda_u(m_u^{-1}(\alpha)))' \\
&= m_u^{-1}(\alpha) + \alpha (m_u^{-1})'(\alpha) - m_u(m_u^{-1}(\alpha))(m_u^{-1})'(\alpha) \\
&= m_u^{-1}(\alpha) \leq 0
\end{aligned}$$

porque m_u define un homeomorfismo entre $\mathscr{D}_u$ y $(\mathscr{S}_u)^\circ$ (ver por ejemplo [Roc97], Teorema 26.5).

Sea $C = (\alpha_1, \alpha_2)$ como arriba y además $m_u(0) = \int_E u(x)P(dx) \notin \bar{C}$. Entonces sabemos que C tiene un punto dominante y por el cálculo anterior es $\alpha_C = \alpha_2$. Por comodidad denotaremos $\alpha_C = \alpha$ y $C = (\alpha - \delta, \alpha)$.

Entonces si $U_n/n \in C$, la distribución empírica converge exponencialmente a la distribución canónica

$$P_\beta(dx) = \frac{1}{Z} e^{\beta u(x)} P(dx),$$

donde $\beta = m_u^{-1}(\alpha)$ y $Z = M_u(\beta) = e^{\Lambda_u(\beta)}$. También se tiene que las distribuciones micro-canónicas $P_{n,C}$ b–convergen a la distribución canónica. Aunque en el problema el número de partículas es fijo, se tiene una convergencia en el sentido de que el número de partículas es muy grande y, como la convergencia en ambos casos es exponencial, se tiene algo muy parecido al límite.

Una de la cantidades físicas importantes es la entropía; esta la obtenemos como $s(\alpha) = -\Lambda_u^*(\alpha)$, con lo cual

$$ds = d(\Lambda_u^*) = \beta d\alpha.$$

Concluimos que la temperatura absoluta correspondiente al nivel de energía α será

$$T_\alpha = -\frac{1}{\beta} \qquad \beta = m_u^{-1}(\alpha).$$

Para el gas ideal tenemos que $E = \Gamma \times \mathbb{R}^3 \subsetneq \mathbb{R}^6$, donde $\Gamma \subset \mathbb{R}^3$ lo interpretamos como un contenedor, es por eso que supondremos que $|\Gamma| < \infty$; también tendremos que

$$P(dx) = \frac{dq\,dp}{|\Gamma|}.$$

Como habíamos dicho anteriormente, en el gas ideal no hay interación entre las partículas y estás sólo poseen energía cinética por lo cual $u(x) = u(q, p) = |p|^2/2$. Por lo tanto tenemos que

$$\begin{aligned}
\Lambda_u(\beta) &= \log \int_E e^{\frac{\beta|p|^2}{2}} P(dx) = \log \iint_{\Gamma\times\mathbb{R}^3} e^{\frac{\beta|p|^2}{2}} \frac{dp\,dq}{|\Gamma|} \\
&= \log \int_{\mathbb{R}^3} e^{\frac{\beta|p|^2}{2}} dp = \log \int_0^{\infty} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} r^2 e^{\frac{\beta r^2}{2}} \cos\phi \, d\theta d\phi dr \\
&= \log 4\pi \int_0^{\infty} r^2 e^{\frac{\beta r^2}{2}} dr = \log\left(\frac{-\beta}{2\pi}\right)^{-3/2} = -\frac{3}{2}\log\left(\frac{-\beta}{2\pi}\right)
\end{aligned}$$

y entonces tenemos que

$$m_u(\beta) = \Lambda_u'(\beta) = -\frac{3}{2\beta}$$

para todo $\beta \in \mathscr{D}_u = (-\infty, 0)$. Ahora, si tomamos $\alpha \in (\mathscr{S})^{\circ} = (0, \infty)$

$$\beta = \beta_\alpha = -\frac{3}{2\alpha}$$

por la fórmula (2.2) se sigue que

$$\begin{aligned}
\Lambda_u^*(\alpha) &= \alpha\left(-\frac{3}{2\alpha}\right) + \frac{3}{2}\log\left(\frac{3}{4\pi\alpha}\right) = -\frac{3}{2}\left(1 - \log\left(\frac{3}{4\pi\alpha}\right)\right) \\
&= -\frac{3}{2}\log\left(\frac{4\pi e\,\alpha}{3}\right)
\end{aligned}$$

de donde se concluye que la función de entropía para el gas ideal está dada por

$$s(\alpha) = -\Lambda_u^*(\alpha) = \frac{3}{2}\log\left(\frac{4\pi e\,\alpha}{3}\right).$$

Y por la discusión antes hecha, si tenemos que $U_n/n \in C = (\alpha - \delta, \alpha)$, tenemos que las distribuciones microcanónicas *b*-convergen a

$$P_\beta(dq, dp) = \frac{1}{Z}\exp\left(-\frac{|p|^2}{2T}\right)\frac{dq\,dp}{|\Gamma|}$$

donde $T = -\frac{1}{\beta}$ con $\beta = -\frac{3}{2\alpha}$, que se le conoce como *distribución de Maxwell–Boltzmann.*

3.2. Convergencia a la distribución gran canónica y el gas ideal.

En la sección anterior resolvimos el problema de encontrar la distribución canónica para el gas ideal, que describe un sistema de partículas en equilibrio en el cual sólo puede haber cambio de energía y el *número de partículas está fijo*. Otra posible situación es que se puedan intercambiar partículas con el entorno, es decir tenemos un sistema abierto. Por ejemplo, si estudiamos sólo una pequeña región de un gas, quisiéramos poder obtener información acerca de su temperatura, presión, etcétera, y si este gas fuera lo suficientemente homogéneo, estas características se preservarían para todo el sistema.

Supongamos que tenemos un sistema de partículas estacionario en $\mathbb{R}^3$. Además, supongamos que cada partícula está en la posición $Q_i(\in \mathbb{R}^3)$ la cual es aleatoria y en el estado X_i (en un espacio de estados E), también supondremos que $\{Q_n\}$ y $\{X_n\}$ son dos sucesiones independientes, es decir $(X_{n_1}, X_{n_2}, \ldots, X_{n_k}) \amalg (Q_{m_1}, Q_{m_1}, \ldots, Q_{m_j})$ para todo k, j; además supondremos que $X_1, X_2, \ldots$ son variables aleatorias i.i.d. con Ley P. Dado un conjunto $\Gamma \in \mathcal{B}(\mathbb{R}^3)$ denotaremos por

$$N_\Gamma = \sum_{i:Q_i \in \Gamma} 1,$$

es decir el número de partículas en Γ.

Supondremos además que

(a) N_Γ tiene una distribución Poisson de parámetro $\nu_0|\Gamma|$, siempre que $|\Gamma| < \infty$.

(b) Si $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ son conjuntos disjuntos tales que $|\Gamma_i| < \infty$ para todo i, entonces $N_{\Gamma_1}, N_{\Gamma_2}, \ldots, N_{\Gamma_n}$ son independientes.

De forma abreviada $\{Q_n\}$ da lugar a un *proceso Poisson espacial de parámetro* ν_0.

En lo siguiente diremos que Γ es un contenedor si $|\Gamma| < \infty$ y para no complicar más adelante la notación, denotaremos a $N_\Gamma = N$ y a los estados de las partículas en Γ como $X_1, X_2, \ldots, X_N$. Llamaremos a $\mathbf{X} = (X_1, X_2, \ldots, X_N; N)$ el gran estado del contenedor unitario de Γ ($|\Gamma| = 1$). Entonces $\mathbf{X}$ toma valores en $\mathbf{E} = \{0\} \cup (\bigcup_{n=1}^{\infty} E^{\times n} \times \{n\})$, donde $\mathbf{X} = 0$ si $N = 0$; también podemos calcular la distribución de $\mathbf{X}$ la cual está dada por

$$\mathbf{P}(d\mathbf{x}) = \mathbf{P}(dx_1, dx_2, \ldots, dx_n; n) = \frac{\nu_0^n}{n!} e^{-\nu_0} P(dx_1)P(dx_2)\ldots P(dx_n).$$

Sea $u : E \to \mathbb{R}$ la energía observada. A lo largo de este desarrollo nos interesa estudiar tanto la energía de las partículas como su número, por lo que definiremos la gran energía como una función $\mathbf{u} : \mathbf{E} \to \mathbb{R} \times (\mathbb{N} \cup \{0\})$ dada por

$$\mathbf{u}(\mathbf{x}) = \mathbf{u}(x_1, x_2, \ldots, x_n; n) = \Big(\sum_{i=1}^{n} u(x_i); n\Big).$$

Tomemos un contenedor Γ muy grande y sea $\Gamma_1, \Gamma_2, \ldots$ una partición de éste en contenedores unitarios; para evitar cualquier indefinición tomaremos siempre $|\Gamma| \in \mathbb{N}$. Denotaremos al estado gran canónico en Γ_i por $\mathbf{X}_i$. Por las propidades del Proceso Poisson espacial y la independencia entre las variables $X_1, X_2, \ldots$ podemos concluir que las variables $\mathbf{X}_1, \mathbf{X}_2, \ldots$ son independientes; además podemos ver que tienen la misma distribución $\mathbf{P}$ ya que $|\Gamma_i| = 1$ para toda i.

Nos gustaría poder aplicar los resultados anteriores de convergencia a los contenedores Γ_i; recordemos que anteriormente los habíamos usado sólo para el número de partículas. Lo primero que haremos es calcular la función $\Lambda_{\mathbf{u}}$, que llamaremos la función de energía libre

$$\begin{aligned}
\Lambda_{\mathbf{u}}(\beta, \lambda) &= \log \mathbb{E}[\exp\{\langle(\beta, \lambda), \mathbf{u}(\mathbf{X}_1)\rangle\}] = \log \mathbb{E}\left[\exp\left\{\beta \sum_{i=1}^{N} u(X_i) + \lambda N\right\}\right] \\
&= \log \mathbb{E}\left[\mathbb{E}\left[\exp\left\{\beta \sum_{i=1}^{N} u(X_i) + \lambda N\right\}\middle| N\right]\right] \\
&= \log \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}\left[\exp\left\{\beta \sum_{i=1}^{n} u(X_i) + \lambda n\right\}\right] \\
&= \log \sum_{n=0}^{\infty} \frac{\nu_0^n}{n!} e^{-\nu_0}[M_u(\beta)e^{\lambda}]^n = -\nu_0 + \nu_0 M_u(\beta)e^{\lambda}, \qquad (3.1)
\end{aligned}$$

donde $M_u(\beta) = \mathbb{E}[e^{\beta u(X_1)}]$.

Si suponemos que $\mathscr{D}_u$ es abierto, tenemos que $\mathscr{D}_{\mathbf{u}} = \mathscr{D}_u \times \mathbb{R}$ es también abierto. Por lo cual podemos derivar (3.1), y obtenemos

$$m_{\mathbf{u}}(\beta, \lambda) = \nabla\Lambda_{\mathbf{u}}(\beta, \lambda) = (\nu_0 M_u'(\beta)e^{\lambda}, \nu_0 M_u(\beta)e^{\lambda}).$$

Entonces para $(\beta, \lambda) \in \mathscr{D}_{\mathbf{u}}$ tenemos que $m_{\mathbf{u}}(\beta, \lambda) = \mathbb{E}_{\mathbf{P}_{(\beta,\lambda)}}[\mathbf{X}_1]$ donde

$$\mathbf{P}(\beta, \lambda) = \exp\{-\Lambda_{\mathbf{u}}(\beta, \lambda) + \langle(\beta, \lambda), \mathbf{u}(\mathbf{x})\rangle\}\mathbf{P}(d\mathbf{x}),$$

que es la distribución gran canónica; como en ocasiones anteriores demostraremos la convergencia a ésta.

Tomemos $(\alpha, \nu) \in \mathrm{ri}(\mathscr{S}_{\mathbf{u}})$ y consideremos la ecuación $m_{\mathbf{u}}(\beta, \lambda) = (\alpha, \nu)$ y resolvamos para (β, λ); entonces

$$\nu_0 M_u'(\beta)e^{\lambda} = \alpha \qquad \nu_0 M_u(\beta)e^{\lambda} = \nu \tag{3.2}$$

de donde se tiene que

$$\frac{\alpha}{\nu} = \frac{M_u'(\beta)}{M_u(\beta)} = \Lambda_u'(\beta) = m_u(\beta)$$

por lo tanto, tenemos que $\beta = \Lambda_u(\alpha/\beta)$ y por (3.2) tenemos que

$$\lambda = \log(\nu/\nu_0) - \Lambda_u(\beta) = \log(\nu/\nu_0) - \Lambda_u(m_u^{-1}(\alpha/\nu)).$$

Calculamos la función conjugada de $\Lambda_{\mathbf{u}}$ para $(\alpha, \nu) \in \mathrm{ri}(\mathscr{S})$, la cual llamaremos la función de gran entropía

$$\begin{aligned}
\Lambda_{\mathbf{u}}(\alpha, \nu) &= \langle(\alpha, \nu), (\beta, \lambda)\rangle - \Lambda_{\mathbf{u}}(m_{\mathbf{u}}^{-1}(\beta, \lambda)) \\
&= -\nu_0 + \nu_0 M_u(\beta)\exp\{\log(\nu/\nu_0) - \Lambda_u(\beta)\} \\
&= \alpha\beta + \lambda\nu + \nu_0 - \nu.
\end{aligned}$$

Ahora bien, si suponemos que hicimos una observación de la gran media

$$\frac{\mathbf{U}_\Gamma}{|\Gamma|} = \frac{\sum_{i=1}^n \mathbf{u}(\mathbf{X}_i)}{n} = \left(\sum_{i:Q_i\in\Gamma} \frac{u(X_i)}{|\Gamma|}, \frac{N}{|\Gamma|}\right) \qquad (|\Gamma| = n),$$

es decir, tenemos una observación de la energía media y la densidad de partículas en el contenedor Γ. Supongamos que esta observación de $\mathbf{U}_n/|\Gamma|$ está en $C \subseteq \mathbb{R}^2$ un conjunto convexo, por ejemplo un rectángulo $(\alpha_1, \alpha_2) \times (\nu_1, \nu_2)$. Supongamos además que $(C \cap \mathscr{S}_{\mathbf{u}})^\circ \neq \emptyset$ y que $\mathbb{E}[\mathbf{u}(\mathbf{X}_1)] = (\nu_0 \mathbb{E}[u(X_1)], \nu_0) \notin \bar{C}$. Ahora, si consideramos una función $g : \mathbf{E} \to \mathbb{R}^d$ medible y acotada, y consideramos la gran media asociada a g

$$\frac{\mathbf{G}_\Gamma}{|\Gamma|} = \frac{\sum_{i=1}^n g(\mathbf{X}_i)}{n} = \frac{\sum_{\{i: Q_i \in \Gamma\}} g(X_i)}{|\Gamma|}$$

bajo las condiciones anteriores y el Teorema 2.2 tenemos que

$$\frac{\mathbf{G}_\Gamma}{|\Gamma|} \xrightarrow{\exp} \int g(\mathbf{x}) \mathbf{P}_{(\beta,\lambda)}(d\mathbf{x})$$

conforme $|\Gamma|$ tiende a infinito, con respecto a $\mathbb{P}_{\Omega,C} = \mathbb{P}(\cdot \mid \mathbf{U}_\Gamma/|\Gamma|)$, donde $\mathbf{P}_{(\beta,\lambda)}$ es la distribución gran canónica de $(\beta, \lambda) = m_{\mathbf{u}}^{-1}(\alpha, \nu)$ y (α, ν) es el punto dominante de C y $\Lambda_{\mathbf{u}}$.
Si definimos

$$\hat{\mathbf{P}}_\Gamma(d\mathbf{x}) = \frac{1}{|\Gamma|} \sum_{i=1}^{|\Gamma|} \delta_{\mathbf{X}_i}(d\mathbf{x})$$

tenemos que bajo las mismas condiciones de arriba sobre $\mathbf{U}_{|\Gamma|}/|\Gamma|$ podemos aplicar el Corolario 2.5 con lo cual

$$\hat{\mathbf{P}}_\Gamma \xrightarrow{\exp} \mathbf{P}_{(\beta,\lambda)} \qquad \text{si } |\Gamma| \to \infty$$

con respecto a $\mathbb{P}_{\Gamma,C}$, y $\mathbf{P}_{(\beta,\lambda)}$ como arriba.
Por último podemos considerar las distribuciones microcanónicas

$$\mathbf{P}_{\Gamma,C} = \mathbf{P}\left(\mathbf{X}_1 \in \cdot \,\middle|\, \frac{\mathbf{U}_\Gamma}{|\Gamma|} \in C\right)$$

las cuales bajo las condiciones sobre C de arriba y el Corolario 2.6 se tiene que b–convergen a $\mathbf{P}_{(\beta,\lambda)}$.

Por último consideremos la distribución gran canónica

$$\begin{aligned}
\mathbf{P}_{(\beta,\lambda)}(d\mathbf{x}) &= \exp\{-\Lambda_{\mathbf{u}}(\beta,\lambda) + \langle(\beta,\lambda),\mathbf{u}(\mathbf{x})\rangle\}\mathbf{P}(d\mathbf{x}) \\
&= \exp\left\{-\Lambda_{\mathbf{u}}(\beta,\lambda) + \sum_{i=1}^{n}\beta u(x_i) + \lambda n\right\}\mathbf{P}(dx_1, dx_2, \ldots, dx_n; n) \\
&= e^{\nu_0}e^{-(n\nu_0 M_u(\beta)e^{\lambda})}[M_u(\beta)e^{\lambda}]^n e^{-\nu_0}\frac{\nu_0^n}{n!}P(dx_1)\ldots P(dx_n) \\
&= \frac{\nu^n}{n!}e^{\lambda}P_\beta(dx_1)P_\beta(dx_2)\ldots P_\beta(dx_n)
\end{aligned}$$

donde P_β es la distribución canónica con $\beta = m_u^{-1}(\alpha/\nu)$. Entonces podemos ver que en el límite el número de partículas en el contendor unitario tienen una distribución Poisson de parámetro ν, y dado este número, los estados de las partículas en el contenedor son i.i.d. y se distribuyen de acuerdo con la distribución canónica P_β; en este caso podemos interpretar a α como la energía media y a ν como la densidad de las partículas observadas en el contenedor.

Lo siguiente que haremos es considerar el problema anterior para el gas ideal, es decir $u(q,p) = |p|^2/2$. Denotemos por q_i a la posición de la i–ésima partícula y por p_i al momento de la misma. Supongamos que las ubicaciones de las partículas $q_1, q_2, \ldots$ dan lugar a un proceso Poisson espacial con intensidad ν_0 y además que son independientes de $p_1, p_2, \ldots$ las cuales son variables inpendientes con la medida de Lebesgue como distribución (*veáse sección anterior*).

Supongamos que observamos la gran energía, es decir, medimos la energía media de nuestro sistema y también la densidad de partículas, y tenemos que

$$\frac{U_\Gamma}{|\Gamma|} \in C = (\alpha_1, \alpha_2) \times (\nu_1, \nu_1).$$

Sea $(\alpha,\nu) \in (\mathscr{S}_{\mathbf{u}})^\circ$ el punto dominante de C y $\Lambda_{\mathbf{u}}^*$; entonces, por lo hecho anteriormente

$$\nabla\Lambda_{\mathbf{u}}^*(\alpha,\nu) = (\beta,\lambda)$$

de modo que

$$\frac{\partial\Lambda_{\mathbf{u}}^*}{\partial\alpha}(\alpha,\nu) = \beta = m_u^{-1}(\alpha/\nu) = -\frac{3\nu}{2\alpha} < 0$$

y

$$\frac{\partial \Lambda_u^*}{\partial \nu}(\alpha, \nu) = \lambda = \log(\nu/\nu_0) + \frac{3}{2}\log\left(\frac{3\nu}{4\pi\alpha}\right)$$

donde esta derivada es positiva si $\nu > \nu(\alpha) = \nu_0^{2/5}(4\pi\alpha/3)$, es negativa si $\nu < \nu(\alpha)$ y es igual a cero si $\nu = \nu(\alpha)$.

Por lo tanto, (α, ν) el punto dominante de $C = (\alpha_1, \alpha_2) \times (\nu_1, \nu_2)$, es

$$\alpha = \alpha_2$$

y

$$\nu = \begin{cases} \nu_1 & \text{si } \nu_1 > \nu(\alpha) \\ \nu(\alpha) & \text{si } \nu_1 \leq \nu(\alpha) \leq \nu_2 \\ \nu_2 & \text{si } \nu_2 < \nu(\alpha) \end{cases}.$$

Entonces, tomando (α, ν) como corresponda y haciendo $\beta = -\frac{3\nu}{2\alpha}$ tenemos que la distribución P_β es igual a

$$P_\beta(dp) = \frac{1}{Z}e^{-|p|^2/2T} \qquad \text{,donde } T = -\frac{1}{\beta} = \frac{2\alpha}{3\nu}$$

con $Z = (2\pi T)^{3/2}$, y por lo tanto la distribución gran canónica está dada por la fórmula

$$\mathbf{P}_{(\beta,\lambda)}(dp_1, dp_2, \ldots, dp_n; n) = \frac{\nu^n}{n!}e^{-\nu}\left(\frac{1}{Z}\right)^n \exp\left\{-\sum_{i=1}^{n}|p_i|^2/2T\right\}dp_1 \ldots dp_n$$

Ahora bien supongamos que tenemos un contenedor Γ_1 con volumen V, entonces la distribución gran canónica del estado $\mathbf{X}$ en Γ_1 se obtiene reemplazando a ν en la fórmula anterior por νV.

Si consideramos la pareja $(\mathbf{Q}, \mathbf{X}) = (Q_1, Q_2, \ldots, Q_N; X_1), X_2, \ldots, X_N; N)$ y tratamos de calcular la distribución condicionada a que $\mathbf{U}_\Gamma/\Gamma$ esté en C, se tiene que

$$\begin{aligned}
&\mathbf{P}(dq_1, dq_2, \ldots, dq_n; dp_1, dp_2, \ldots, dp_n; n \,|\, \mathbf{U}_\Gamma/\Gamma \in C) \\
&= \mathbf{P}(dp_1, dp_2, \ldots, dp_n; n \,|\, \mathbf{U}_\Gamma/\Gamma \in C)(dq_1/V)(dq_2/V)\ldots(dq_n/V) \\
&\xrightarrow{b}= (\nu^n/n!)\exp(-\nu V)(1/Z)^n \\
&\quad \times \exp\left\{-\sum_{i=1}^{n}|p_i|^2/2T\right\} dp_1\ldots dp_n(dq_1/V)(dq_2/V)\ldots(dq_n/V)
\end{aligned}$$

para $p_i \in \mathbb{R}^3$ y $q_i \in \Gamma_1$, $i = 1, 2, \ldots, n$ y $n = 0, 1, 2, \ldots$, donde la primera igualdad se da por el siguiente Lema:

Lema 3.1. *Sea N un proceso de Poisson espacial de intensidad ν. Si $\Theta \subseteq \Gamma$ entonces*

$$\mathbb{P}(N_\Theta = 1 \,|\, N_\Gamma = 1) = \frac{|\Theta|}{|\Gamma|}.$$

Demostración.

$$\begin{aligned}
\mathbb{P}(N_\Theta = 1 \,|\, N_\Gamma = 1) &= \frac{\mathbb{P}(N_\Theta = 1 \cap N_\Gamma = 1)}{N_\Gamma = 1} = \frac{\mathbb{P}(N_\Theta = 1 \cap N_{\Gamma\cap\Theta^c} = 0)}{N_\Gamma = 1} \\
&= \frac{\nu|\Theta|e^{-\nu|\Theta|}e^{-\nu|\Gamma\cap\Theta^c|}}{\nu|\Gamma|e^{|-\nu\Gamma|}} = \frac{|\Theta|}{|\Gamma|}.
\end{aligned}$$

□

Bibliografía

[DZ98] A. Dembo; O. Zeitouni. *Large deviations techniques and applications*, volumen 38 de *Applications of Mathematics (New York)*. Springer-Verlag, New York, segunda edición, 1998.

[Ell85] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*, volumen 271 de *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[ET76] I. Ekeland; R. Temam. *Convex analysis and variational problems*. North-Holland Publishing Co., Amsterdam, 1976. Traducido del francés, Studies in Mathematics and its Applications, Vol. 1.

[Fol99] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[LN90] T. Lehtonen; E. Nummelin. Level I theory of large deviations in the ideal gas. *Internat. J. Theoret. Phys.*, 29(6):621–635, 1990.

[ML79] A. Martin-Löf. *Statistical Mechanics an the Foundation of Thermodynamics*, volumen 101 de *Lecture Notes in Physics*. Springer, Berlin, 1979.

[Ney83] P. Ney. Dominating points and the asymptotics of large deviations for random walk on $\mathbf{R}^d$. *Ann. Probab.*, 11(1):158–167, 1983.

[Ney84] P. Ney. Convexity and large deviations. *Ann. Probab.*, 12(3):903–906, 1984.

[Num89] E. Nummelin. A conditional weak law of large numbers. En *Stability problems for stochastic models (Sukhumi, 1987)*, volumen 1412 de *Lecture Notes in Math.*, páginas 259–262. Springer, Berlin, 1989.

[Roc97] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reimpresión de original de 1970, Princeton Paperbacks.