



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA**

**ESTIMACIÓN DE MEDIDAS DE RIESGO FINANCIERO. UN ENFOQUE
BASADO EN EL FENÓMENO DE CONCENTRACIÓN DE MEDIDAS
EMPÍRICAS.**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

HÉCTOR CRISTIAN OLIVERO QUINTEROS

**PROFESOR GUÍA:
SR. JOAQUÍN FONTBONA TORRES**

**MIEMBROS DE LA COMISIÓN:
SR. RENE ALEJANDRO JOFRÉ CÁCERES
SR. JAIME RICARDO SAN MARTÍN ARISTEGUI**

**SANTIAGO DE CHILE
DICIEMBRE 2010**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL MATEMÁTICO
POR: HÉCTOR OLIVERO QUINTEROS
FECHA: 22/12/2010
PROF. GUÍA: SR. JOAQUÍN FONTBONA TORRES

ESTIMACIÓN DE MEDIDAS DE RIESGO FINANCIERO. UN ENFOQUE BASADO EN EL FENÓMENO DE CONCENTRACIÓN DE MEDIDAS EMPÍRICAS

El objetivo principal de esta memoria es construir, a partir de información histórica, intervalos de confianza no asintóticos y no paramétricos para una familia de medidas de riesgo financiero. El enfoque utilizado se basa en teoremas recientes sobre concentración de procesos empíricos y busca llegar a resultados explícitos, que eventualmente puedan ser aplicados en la práctica financiera.

Un segundo objetivo metodológico, es estudiar la aplicabilidad de resultados teóricos conocidos de concentración de procesos empíricos a problemas de estimación en general, y en particular al problema de estimación de medidas de riesgo financiero a partir de información histórica.

Esta memoria comienza con una introducción a las medidas de riesgo financiero. Luego se revisan la convergencia de leyes de probabilidad y diferentes métricas en el espacio de leyes de probabilidad. Después se estudian las desigualdades de Talagrand y en especial se estudian caracterizaciones integrales para ellas. Con estos elementos se presentan resultados conocidos recientemente de concentración para procesos empíricos, poniendo énfasis en determinar todos los parámetros que aparecen en ellos a partir de los datos del problema.

En la última parte de esta memoria se aplican los resultados presentados previamente, y se obtienen nuevos resultados que permiten construir intervalos de confianza no asintóticos y no paramétricos para distintas medidas de riesgo financiero y activos que toman valores en todo \mathbb{R} . Más precisamente, se obtiene un resultado para el *Value at Risk*, dos resultados para medidas de riesgo espectrales y dos resultados para el *Expected Shortfall*.

Los resultados obtenidos para el *Expected Shortfall* extienden a situaciones más generales resultados recientes para activos con leyes a soporte compacto, y son por lo tanto, una innovación en el tema estudiado.

Agradecimientos

Escribo estas líneas para agradecer a todos aquellos que me acompañaron en el proceso que finaliza con esta memoria. No sé si lo hubiera logrado sin ustedes.

A mis padres, que me dieron el ser y han hecho de mi lo que soy, que me enseñaron con su ejemplo que el trabajo todo lo vence y que con su amor me han dado la fuerza para seguir adelante, para perseguir mis sueños, para no rendirme nunca. Esta memoria está dedicada a ellos.

A mis hermanas Paola y Javiera, por su apoyo constante, por siempre creer en mi, y sobre todo por quererme tal cual soy. Las quiero hasta el infinito y esta memoria también está dedicada a ustedes.

A mi sobrina Catita, simplemente por existir, por cambiarle el olor a mis mañanas y llenar nuestra casa de su ruidosa alegría.

A mi querida Eliana, por todo lo que me entregaste en los años que recorrimos juntos, por toda la felicidad que me regalaste sin pedir nada o casi nada, que no es lo mismo pero es igual.

A mis amigos de la vida Paula Raquel y Andrés Itukun, por ser tan buenos amigos y haber estado conmigo cada vez que los necesité. La amistad de cada uno de ustedes es un privilegio para mi.

A mis amigos del DIM, los que han hecho que este paso de casi cinco años por el departamento haya sido un tiempo extraordinario. Andrés Iroume, Julio, Omar, Victor, Tomás González, Jorge, Guido, Jerónimo, Félix, Claudio, Cristobal, Thomas, Francisco, Tomás Spencer, Constanza y Gonzalo. De cada uno de ustedes aprendí algo.

A mis auxiliares, especialmente a mis auxiliares DIM, que desde los lejanos tiempos de primer año fueron siempre un modelo para mi. Especial mención para Mauricio Duarte y Roberto Amaru Cortez.

A todos los funcionarios del DIM, por su eterna buena disposición y por todas las veces que me ayudaron. Especialmente quiero agradecer a Eterin Jaña, Silvia Mariano y Oscar Mori.

A todos mis profesores de la Universidad, en especial a mis profesores del DIM, los que influyeron determinadamente en mi elección de especialidad. En especial quiero agradecer a los profesores Pablo Dartnell, Manuel del Pino, Roberto Cominetti, Nancy Lacourly y por supuesto al profesor Carlos Conca, a quien además, siempre le agradeceré haber apoyado mi postulación al doctorado que empiezo el próximo año.

A los profesores de mi comisión evaluadora Alejandro Jofré y Jaime San Martín, por haber aceptado participar en esta memoria a pesar de lo ajustado de los plazos. También quiero agradecer al profesor Jofré su apoyo en mi postulación a la escuela de verano en UBC a la que asistí este año.

Para terminar, quiero agradecer a mi profesor guía Joaquín Fontbona, no sólo por sus correcciones y sugerencias en esta memoria (sin las cuales sería al menos ilegible), si no también por el apoyo constante durante los últimos tres años de la carrera, por haberme permitido ser su auxiliar, por haberme apoyado para ir a UBC, por respaldarme en la postulación al doctorado y un largo etcétera. Dado que siempre me quedaré corto en palabras de agradecimientos para usted, sólo me queda plagiar a Cerati y decir GRACIAS TOTALES!

Índice general

1. Introducción	1
2. Medidas de Riesgo Financiero	4
2.1. Funciones Cuantil	4
2.1.1. Definiciones y Propiedades Básicas	4
2.1.2. Integrales de cuantiles	7
2.2. ¿Cómo medir el riesgo financiero?	10
2.3. Medidas Coherentes de Riesgo	12
2.3.1. Expected Shortfall: El ejemplo central de medida coherente de riesgo	17
2.3.2. Medidas de Riesgo Espectrales	19
2.3.3. Espectros de riesgo como medidas cuantitativas de la aversión al riesgo.	25
3. Convergencia y Métricas en el Espacio de Medidas de Probabilidad	27
3.1. Convergencia débil de medidas de probabilidad.	27
3.2. Métricas en el Espacio de Medidas de Probabilidad	30
3.2.1. Distancia de Lévy-Prokhorov d_P	30
3.2.2. Distancia Dual de Funciones Lipschitz Acotadas BL	31
3.2.3. Distancias de Wasserstein W_p	31
3.2.4. Relaciones entre las distintas Métricas	36
4. Desigualdades de Talagrand $T_p(\lambda)$	38
4.1. Definiciones	38
4.2. Criterio Integral para T_1	40
4.2.1. Criterio de Djelleout, Guillin y Wu.	42
4.2.2. Criterio de Bolley, Gozlan y Villani	46

5. Concentración de la Medida Empírica	50
5.1. Concentración de la Medida Empírica con respecto a W	51
5.1.1. Teorema de Bolley, Guillin y Villani	51
5.1.2. Teorema de Gozlan y Léonard	66
5.2. Concentración de la medida empírica con respecto a d_P	78
5.2.1. Concentración exponencial para la distancia de Lévy.	83
6. Estimación de Medidas de Riesgo Financiero	85
6.1. Estimadores históricos para medidas de riesgo espectrales	85
6.2. Convergencia de Estimadores de Riesgo	87
6.2.1. Convergencia de $V@R_\alpha^N$	87
6.2.2. Convergencia de $\hat{\rho}_\phi^N$	89
6.3. Intervalos de confianza no asintóticos	90
6.3.1. Intervalos de confianza en muestras finitas para el $V@R_\alpha$	90
6.3.2. Intervalos de confianza en muestras finitas para ρ_ϕ	92
6.3.3. Teorema de Brown, Gao y Wang de concentración para $CV@R_\alpha^N(X)$	94
6.3.4. Extensión del resultado de Brown, Gao y Wang a v.a. con valores en \mathbb{R}	98
6.3.5. Una nueva estimación para las desviaciones de $\widehat{ES}_\alpha^N(X)$	102
7. Conclusión	105
7.1. Análisis de resultados obtenidos	105
7.2. Trabajo Futuro	107
Anexos	109
A. Resultados pendientes en la demostración del Teorema 5.1.2	110
A.1. Teorema de Sion para Minimax	110
A.2. Entropía	114
A.2.1. Representación Dual de la Entropía	114
A.2.2. Entropía Métrica en $(\mathcal{P}(\mathcal{X}), W_p(\cdot, \cdot))$	115
Bibliografía	116

Capítulo 1

Introducción

Medir el riesgo asociado a un activo financiero no es una idea nueva, nació al menos hace 50 años, cuando Harry Markowitz, en su trabajo acerca de selección óptima de portafolios [31], puso en evidencia la necesidad de cuantificar el riesgo asociado a una posición financiera. Sin embargo, fue necesario que pasara el tiempo y recién a finales de los años 90 surgió una teoría axiomática del riesgo, en los trabajos de Artzner et al. [4] y [5].

En esta teoría axiomática del riesgo financiero, el mercado se modela como un conjunto de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde Ω representa los distintos estados del mercado y cada variable aleatoria X representa las ganancias o pérdidas debidas al mantenimiento de una posición financiera. En este contexto, una medida de riesgo financiero es una función que asigna un número real a una variable aleatoria. Si esta asignación satisface ciertas propiedades, que estudiaremos más adelante, se dice que la medida de riesgo es coherente. Uno de los ejemplos más importantes de este tipo de medidas, es el *expected shortfall*, que para una variable aleatoria X denotamos como $ES_\alpha(X)$.

Generalmente, cuando se define una medida de riesgo, se hace suponiendo un conocimiento exhaustivo del mercado, es decir, se asume que se conocen todos los posibles escenarios del mercado y la probabilidad de ocurrencia de cada uno. Sin embargo, en la práctica no se dispone de tal información, lo que durante los últimos años ha motivado a diversos autores ha estudiar el impacto de esta incertidumbre en por ejemplo, modelos de precios para derivados complejos [13] o en la optimización de portafolios [12] y [36]. Una característica común de los trabajos en esta área es buscar resultados cuantitativos y no sólo cualitativos. El trabajo desarrollado en esta memoria va en la misma línea de los artículos anteriores, aunque en alguna medida con objetivos más modestos. A lo largo de esta memoria buscaremos cuantificar el error que se comete al estimar el riesgo de un activo financiero a través de información histórica.

Un hecho clave en el desarrollo que presentaremos, que ya había sido notado, por ejemplo, por Pflug y Wozabal en [36], es que si X es un activo financiero con ley de probabilidad μ , del que tenemos una muestra de tamaño N y llamamos $\hat{\mu}^N$ a la ley empírica de dicha muestra,

entonces bajo condiciones muy generales

$$|\widehat{ES}_\alpha^N(X) - ES_\alpha(X)| \leq \frac{1}{\alpha} W(\mu, \hat{\mu}^N),$$

donde $\widehat{ES}_\alpha^N(X)$ corresponde al estimador empírico o histórico de $ES_\alpha(X)$ y W es una distancia apropiada en el espacio de medidas de probabilidad. Lo que hace a esta desigualdad especialmente atractiva, es el hecho de que como veremos en el capítulo 5, existen resultados que permiten acotar $\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon)$ para todo $\varepsilon > 0$. Y luego, a partir de dichos resultados es posible construir intervalos de confianza para $ES_\alpha(X)$. Resultados similares a los que encontraremos ya fueron establecidos por Brown en [10] y mejorados por Wang y Gao en [21], pero sólo en el caso de que X toma valores en un intervalo compacto. En esta memoria se levanta esta hipótesis y se establecen resultados para X con valores en \mathbb{R} , y además, para una familia general de medidas de riesgo, entre las que se cuenta el ES_α .

En términos generales, esta memoria se desarrolla de la siguiente manera: primero se introducen las medidas de riesgo coherente que serán el contexto teórico en el que trabajaremos. Luego se desarrollan las herramientas matemáticas que utilizaremos y por último se aplican estas herramientas al problema específico de estimación de medidas de riesgo.

Para terminar esta introducción, una breve descripción de los capítulos que siguen.

En el capítulo 2 se estudia el contexto teórico para medidas de riesgo financiero. Para ello, es necesario introducir primero las funciones cuantil y mostrar algunas de sus propiedades más importantes. Luego, con las funciones cuantil se construyen las medidas de riesgo espectrales, y finalmente se caracterizan a las medidas coherentes que pertenecen a esta familia.

En el capítulo 3 primero se estudia la convergencia de sucesiones de medidas de probabilidad y se introduce una clase especial de sucesiones, denominadas medidas empíricas. Después se introducen distintas métricas para el espacio de medidas de probabilidad y se estudia su relación con la convergencia ya mencionada.

En el capítulo 4 se introducen las desigualdades de Talagrand, que corresponden a desigualdades funcionales para medidas de probabilidad, y están relacionadas con el fenómeno de concentración de dichas medidas.

En el capítulo 5 se estudia el comportamiento de la cantidad $\mathbb{P}(d(\mu, \hat{\mu}^N) > \varepsilon)$, donde d corresponde a alguna de las métricas definidas en el capítulo 3. En particular, se presentarán resultados dados por Bolley, Guillin y Villani en [8], por Dudley en [17] y por Gozlan y Léonard en [23]. Estos resultados establecen que existe $N_0 \in \mathbb{N}$, tal que para todo $N \geq N_0$

$$\mathbb{P}(d(\mu, \hat{\mu}^N) > \varepsilon) \leq f(N, \varepsilon),$$

donde f es una función explícita que para cada ε tiende a cero cuando N crece a infinito. Esta última desigualdad entrega un control cuantitativo y no asintótico para la probabilidad de tener desvíos mayores a un tamaño dado.

El capítulo 6 está dividido en tres secciones. En la primera se introducen los estimadores históricos y se argumenta a favor de su uso en la práctica financiera. En la segunda se discute

el comportamiento asintótico de estos estimadores, y en particular se presentan resultados ya conocidos acerca de su convergencia casi segura. Finalmente, en la tercera sección de este capítulo, se construyen intervalos de confianza no asintóticos para distintas medidas de riesgo en torno a sus estimadores históricos. Estos resultados son el producto final de esta memoria y al menos uno de ellos, representa una verdadera innovación en el tema.

En el capítulo 7, que corresponde a la conclusión de esta memoria, se hace un repaso de los principales resultados obtenidos, se discuten sus alcances y limitaciones, y se plantean algunas direcciones para el trabajo futuro.

Capítulo 2

Medidas de Riesgo Financiero

Antes de comenzar con el estudio de las medidas de riesgo financiero, se introducen las funciones cuantil, las que como se verá en el desarrollo del capítulo, son un bloque fundamental en las construcciones que se realizarán.

2.1. Funciones Cuantil

En esta sección se presentará un conjunto de resultados conocidos para las funciones cuantil. La mayoría de ellos se puede encontrar en el Apéndice A.3 de [20]. Mientras que la Proposición 2.1.8 se puede encontrar en [2].

2.1.1. Definiciones y Propiedades Básicas

En lo que sigue, por función distribución acumulada de una variable aleatoria real X , definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, siempre nos referiremos a la función

$$F_X(x) = \mathbb{P}(X \leq x).$$

Definición 2.1.1. *Sea X una variable aleatoria real y F_X su función de distribución acumulada. Se define respectivamente el **cuantil inferior y superior** de F_X de orden α como:*

$$q_X^-(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\} \text{ y}$$

$$q_X^+(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) > \alpha\}.$$

Notemos que las funciones:

$$\alpha \rightarrow q_X^+(\alpha) \text{ y}$$

$$\alpha \rightarrow q_X^-(\alpha),$$

2.1. Funciones Cuantil

son crecientes y por lo tanto medibles. Además, para todo $\alpha \in (0, 1)$, $q_X^+(\alpha) \geq q_X^-(\alpha)$. Sin embargo, $q_X^+(\alpha) = q_X^-(\alpha)$ para casi todo $\alpha \in (0, 1)$, pues si $q_X^+(\alpha) > q_X^-(\alpha)$, entonces $q_X^-(\cdot)$ tiene una discontinuidad en α , y como $q_X^-(\cdot)$ es una función creciente y medible en un intervalo acotado, puede tener a lo más una cantidad numerable de discontinuidades.

Proposición 2.1.2. *Para toda variable aleatoria real X la función $q_X^-(\cdot)$ es continua por la izquierda y $q_X^+(\cdot)$ es continua por la derecha.*

Demostración: Sea X una variable aleatoria real, queremos probar que si $\{\alpha_n\}_{m \geq 1}$ es una sucesión que converge a α por la izquierda, entonces

$$\lim_{n \rightarrow \infty} q_X^-(\alpha_n) = q_X^-(\alpha).$$

Como $q_X^-(\cdot)$ es creciente, basta con probar que para todo $\varepsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$. Entonces

$$q_X^-(\alpha_n) \geq q_X^-(\alpha) - \varepsilon.$$

Sea $\varepsilon > 0$, por la definición de $q_X^-(\alpha)$ tenemos que $F_X(q_X^-(\alpha) - \varepsilon) < \alpha$. Luego, como $\{\alpha_n\}$ converge a α , existe un $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$,

$$F_X(q_X^-(\alpha) - \varepsilon) < \alpha_n < \alpha.$$

Así se tiene,

$$\forall n \geq n_0 : q_X^-(\alpha) - \varepsilon < q_X^-(\alpha_n),$$

que es lo que queríamos demostrar. La prueba de que $q_X^+(\cdot)$ es continua por la derecha es similar.

q.e.d.

Definición 2.1.3. *Sea $F : (a, b) \rightarrow \mathbb{R}$ una función creciente (no necesariamente estrictamente creciente). Sea*

$$c := \lim_{x \searrow a} F(x) \quad d := \lim_{x \nearrow b} F(x).$$

Una función $q : (c, d) \rightarrow (a, b)$ es una inversa generalizada para F si

$$F(q(s)-) \leq s \leq F(q(s)+) \quad \forall s \in (c, d).$$

Proposición 2.1.4. *Las funciones cuantil inferior y superior de una variable aleatoria X son inversas generalizadas para F_X .*

Demostración: Empecemos por notar que F_X está definida en $(-\infty, \infty)$ y toma valores en $[0, 1]$. Dado un $s \in [0, 1]$, probemos que

$$F_X(q_X^-(s)-) \leq s \leq F_X(q_X^-(s)+).$$

Notemos que de la definición de $q_X^-(s)$, tenemos que si $x \geq q_X^-(s)$ entonces $F(x) \geq s$, de donde es directo que $F_X(q_X^-(s)+) \geq s$. De la misma manera, si $x < q_X^-(s)$ entonces $F(x) < s$, de donde se sigue que $F_X(q_X^-(s)-) \leq s$. La demostración para $q_X^+(\cdot)$ es similar.

2.1. Funciones Cuantil

q.e.d.

Lema 2.1.5. *Sea X una variable aleatoria real, entonces F_X es una inversa generalizada para las funciones cuantil inferior y cuantil superior.*

Demostración: Probemos que F_X es una inversa generalizada para $q_X^-(\cdot)$, es decir

$$q_X^-(F_X(x)-) \leq x \leq q_X^-(F_X(x)+), \forall x \in \mathbb{R}.$$

Por la definición del cuantil inferior, si $s < F_X(x)$ entonces $q_X^-(s) \leq x$, de donde es directo que $q_X^-(F_X(x)-) \leq x$. Por otro lado, si $s > F_X(x)$, nuevamente por definición de $q_X^-(\cdot)$ se tiene que $q_X^-(s) > x$ de donde es directo que $x \leq q_X^-(F_X(x)+)$. La demostración para $q_X^+(\cdot)$ es similar.

q.e.d.

Lema 2.1.6. *Consideremos una variable aleatoria X con función distribución F_X y U , una variable aleatoria Uniforme(0,1) no necesariamente definida en el mismo espacio de probabilidad que X . Si $q_X^-(u)$ es la función cuantil inferior de F_X entonces $Z = q_X^-(U)$ tiene la misma distribución que X , ie. $F_Z = F_X$.*

Demostración: Para empezar notemos que como $F_X(\cdot)$ y $q_x^-(\cdot)$ son inversas generalizadas, satisfacen

$$q_X^-(F_X(s)) = q_X^-(F_X(s)-) \leq s \leq F_X(q_X^-(s)+) = F_X(q_X^-(s)),$$

de donde concluimos que

$$(0, F_X(t)) \subset \{s \in (0, 1) : q_X^-(s) \leq t\} \subset (0, F_X(t)].$$

Entonces,

$$\mathbb{P}(U \in (0, F_X(t))) \leq \mathbb{P}(U \in \{s \in (0, 1) : q_X^-(s) \leq t\}) \leq \mathbb{P}(U \in (0, F_X(t)]),$$

y como U tiene distribución uniforme en $[0, 1]$, se sigue que

$$\mathbb{P}(U \in \{s \in (0, 1) : q_X^-(s) \leq t\}) = F_X(t).$$

Pero

$$F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(q_X^-(U) \leq t) = \mathbb{P}(U \in \{s \in (0, 1) : q_X^-(s) \leq t\}) = F_X(t),$$

lo que concluye la demostración.

q.e.d.

2.1.2. Integrales de cuantiles

Definición 2.1.7. *Se define el promedio de la cola izquierda de nivel α como*

$$\bar{x}_{(\alpha)} = \frac{1}{\alpha} \int_0^{\alpha} q_X^-(u) du.$$

Proposición 2.1.8. *Sean X e Y dos variables aleatorias tales que $\mathbb{E}(X^-) < \infty$ y $\mathbb{E}(Y^-) < \infty$. Entonces, si $Z = X + Y$, tenemos que*

$$\bar{z}_{(\alpha)} \geq \bar{x}_{(\alpha)} + \bar{y}_{(\alpha)}.$$

Demostración: Comencemos por definir:

$$\mathbb{1}_{\{X \leq x\}}^{(\alpha)}(t) = \begin{cases} \mathbb{1}_{\{X \leq x\}}(t) & \text{si } \mathbb{P}(X = x) = 0, \\ \mathbb{1}_{\{X \leq x\}}(t) + \frac{\alpha - \mathbb{P}(X \leq x)}{\mathbb{P}(X = x)} \mathbb{1}_{\{X = x\}}(t) & \text{si } \mathbb{P}(X = x) > 0. \end{cases}$$

Ahora notemos que:

1. $\forall t, \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}(t) \in [0, 1]$.
2. $\mathbb{E}(\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}) = \alpha$.
3. $\alpha^{-1} \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}) = \bar{x}_{(\alpha)}$.

Probemos estas tres afirmaciones:

1. Notemos que esto es directo si $\mathbb{P}(X = q_X^-(\alpha)) = 0$, pues en este caso la función $\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}(\cdot)$ corresponde a una indicatriz. Si $\mathbb{P}(X = q_X^-(\alpha)) > 0$, el único punto en que la cota indicada no es trivial es cuando evaluamos en $t = q_X^-(\alpha)$. Notemos que en ese punto

$$\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}(q_X^-(\alpha)) = 1 + \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} = 1 - \frac{\mathbb{P}(X \leq q_X^-(\alpha)) - \alpha}{\mathbb{P}(X = q_X^-(\alpha))}.$$

Para probar que esta cantidad está entre $[0, 1]$ basta con notar que el cuociente que aparece a la derecha corresponde al cuociente entre el trazo negro y el trazo verde de la figura 2.1.1, que como se aprecia en la misma figura, tiene un valor entre 0 y 1. Con esto es directo que $\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}(q_X^-(\alpha)) \in [0, 1]$.

2. Separemos por casos:

a) Si $\mathbb{P}(X = q_X^-(\alpha)) = 0$, entonces

$$\mathbb{E}(\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}) = \mathbb{E}(\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) = \mathbb{P}(X \leq q_X^-(\alpha)) = \alpha.$$

2.1. Funciones Cuantil

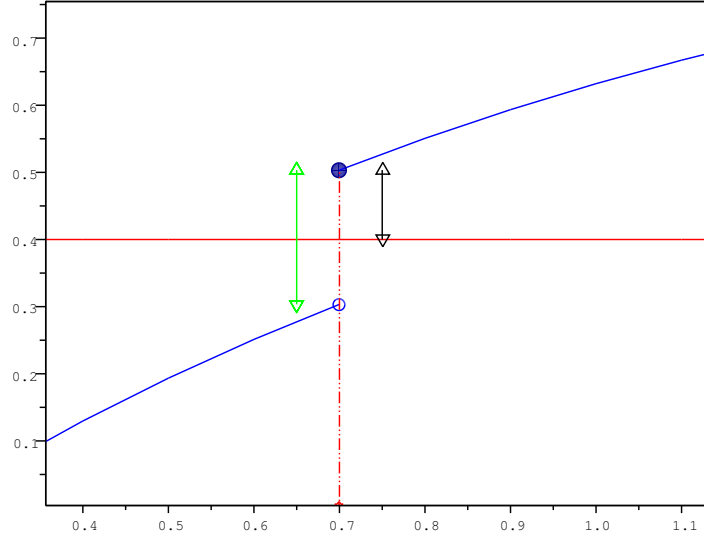


Figura 2.1.1: **Función de Distribución Discontinua.** En azul la función distribución de X . La línea continua roja representa el nivel α , mientras que la línea punteada roja muestra $q_x^-(\alpha)$. El trazo verde corresponde a $\mathbb{P}(X = q_x^-(\alpha))$ y el trazo negro corresponde a $\mathbb{P}(X \leq q_x^-(\alpha)) - \alpha$.

- b) Si $\mathbb{P}(X = q_X^-(\alpha)) > 0$, es importante notar que $\mathbb{P}(X \leq q_X^-(\alpha)) > \alpha$. Con eso en mente, se obtiene

$$\begin{aligned} \mathbb{E}(\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}) &= \mathbb{E}(\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) + \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} \mathbb{E}(\mathbb{1}_{\{X = q_X^-(\alpha)\}}) \\ &= \mathbb{P}(X \leq q_X^-(\alpha)) + \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} \mathbb{P}(X = q_X^-(\alpha)) \\ &= \alpha. \end{aligned}$$

3. Nuevamente separamos en casos:

- a) Si $\mathbb{P}(X = q_X^-(\alpha)) = 0$, recordemos que entonces $\mathbb{P}(X \leq q_X^-(\alpha)) = \alpha$. Luego se tiene

$$\begin{aligned} \bar{x}_{(\alpha)} &= \frac{1}{\alpha} \left[\mathbb{E} \left(X \mathbb{1}_{X \leq q_X^-(\alpha)} \right) + q_X^-(\alpha) (\alpha - \mathbb{P}(X \leq q_X^-(\alpha))) \right] \\ &= \frac{1}{\alpha} \left[\mathbb{E} \left(X \mathbb{1}_{X \leq q_X^-(\alpha)} \right) \right] \\ &= \frac{1}{\alpha} \left[\mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right) \right]. \end{aligned}$$

2.1. Funciones Cuantil

b) Si $\mathbb{P}(X = q_X^-(\alpha)) > 0$, entonces

$$\begin{aligned}
 \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)}) &= \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} + X \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} \mathbb{1}_{\{X = q_X^-(\alpha)\}}) \\
 &= \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} + q_X^-(\alpha) \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} \mathbb{1}_{\{X = q_X^-(\alpha)\}}) \\
 &= \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) + q_X^-(\alpha) \frac{\alpha - \mathbb{P}(X \leq q_X^-(\alpha))}{\mathbb{P}(X = q_X^-(\alpha))} \mathbb{P}(X = q_X^-(\alpha)) \\
 &= \mathbb{E}(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) + q_X^-(\alpha) (\alpha - \mathbb{P}(X \leq q_X^-(\alpha))) \\
 &= \alpha \bar{x}(\alpha).
 \end{aligned}$$

Ahora podemos demostrar que $\bar{z}(\alpha) \geq \bar{x}(\alpha) + \bar{y}(\alpha)$. En efecto

$$\begin{aligned}
 \alpha(\bar{z}(\alpha) - \bar{x}(\alpha) - \bar{y}(\alpha)) &= \mathbb{E} \left(Z \mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} - Y \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}^{(\alpha)} \right) \\
 &= \mathbb{E} \left(X \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right) + Y \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}^{(\alpha)} \right) \right).
 \end{aligned}$$

Ahora notemos que si $X < q_X^-(\alpha)$ entonces

$$\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} = \mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - 1 \leq 0,$$

pues $\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} \in [0, 1]$. Luego

$$X \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right) \geq q_X^-(\alpha) \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right).$$

Por otro lado si $X > q_X^-(\alpha)$ entonces $\mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} = 0$, y se tiene

$$\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} = \mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} \geq 0.$$

Entonces

$$X \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right) \geq q_X^-(\alpha) \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right),$$

de donde concluimos que

$$X \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right) \geq q_X^-(\alpha) \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}^{(\alpha)} \right).$$

Análogamente, se obtiene

$$Y \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}^{(\alpha)} \right) \geq q_Y^-(\alpha) \left(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}}^{(\alpha)} - \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}^{(\alpha)} \right),$$

2.2. ¿Cómo medir el riesgo financiero?

de donde se sigue que

$$\begin{aligned}
 \alpha(\bar{z}_{(\alpha)} - \bar{x}_{(\alpha)} - \bar{y}_{(\alpha)}) &= \mathbb{E}(Z\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}} - X\mathbb{1}_{\{X \leq q_X^-(\alpha)\}} - Y\mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}) \\
 &= \mathbb{E}(X(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) + Y(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}} - \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}})) \\
 &\geq q_X^-(\alpha)\mathbb{E}(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}} - \mathbb{1}_{\{X \leq q_X^-(\alpha)\}}) + q_Y^-(\alpha)\mathbb{E}(\mathbb{1}_{\{Z \leq q_Z^-(\alpha)\}} - \mathbb{1}_{\{Y \leq q_Y^-(\alpha)\}}) \\
 &= q_X^-(\alpha)(\alpha - \alpha) + q_Y^-(\alpha)(\alpha - \alpha) \\
 &= 0.
 \end{aligned}$$

q.e.d.

2.2. ¿Cómo medir el riesgo financiero?

Sin duda esta pregunta es antigua y a lo largo del desarrollo de las finanzas matemáticas ha tenido distintas respuestas. Una respuesta, que es la de Markowitz, es utilizar la varianza de los retornos de los activos. Sin embargo, la varianza como medida de riesgo tiene un problema: “castiga” tanto movimientos a la baja como movimientos al alza en el precio de un activo, como se ve en el ejemplo que aparece en la tabla 2.2.1. En esta tabla aparecen los retornos de dos activos en un mundo con dos posibles estados de la naturaleza que son equiprobables. Notemos que sin importar el estado de la naturaleza el activo B tiene retornos mejores o iguales que el activo A , lo que indica que el activo B es menos riesgoso que el activo A . Sin embargo si calculamos la varianza de estos activos nos encontraremos de que el activo B tiene mayor varianza, lo que equivocadamente nos lleva a concluir que el activo B es más riesgoso.

	Probabilidad	Activo A	Activo B
Estado 1	0,5	-1,0	0,0
Estado 2	0,5	5,0	10,0
Media		2	5
Varianza		9	25

Cuadro 2.2.1: Ejemplo para mostrar las falencias la varianza como medida de riesgo.

Una alternativa a la varianza como medida de riesgo que se hace cargo de que los movimientos que hay que castigar son los movimientos a la baja es el $V@R_\alpha$.

Definición 2.2.1. Sea X la variable aleatoria que representa los retornos de un activo dado, se define el **Value at Risk de nivel alfa** del activo X como

$$V@R_\alpha(X) = -q_X^-(\alpha).$$

2.2. ¿Cómo medir el riesgo financiero?

El $V@R_\alpha$ de un activo representa una cota para las pérdidas que podría tener el activo para un nivel de confianza α . Por ejemplo, si para $\alpha = 5\%$ el $V@R_\alpha$ de un activo es igual a \$1 000 000, esto significa que con un 95 % de probabilidad las pérdidas del activo no superarán dicha cantidad.

Podemos volver al ejemplo de la tabla (2.2.1) y encontramos que en este caso para $\alpha = 5\%$ se tiene que:

$$0 = V@R_\alpha(\text{Activo } B) \leq V@R_\alpha(\text{Activo } A) = 1.$$

Resultado que está de acuerdo con la intuición de que el Activo A es más riesgoso que el activo B.

Observación 2.2.2: Algunos autores definen el $V@R_\alpha(X) = -q_X^+(\alpha)$. Es importante notar que las dos definiciones no son equivalentes, a menos que $q_X^-(\cdot)$ sea continua en α .

A pesar de que el $V@R_\alpha$ es sin duda una de las medidas de riesgo más populares en Chile y el mundo, tiene el defecto fundamental de no tomar en cuenta los efectos de diversificación de una cartera. El efecto de diversificación de una cartera puede entenderse de la siguiente manera: si en una cartera tengo dos activos, en el peor de los casos se presenta simultáneamente el peor escenario de cada activo, luego las pérdidas totales no pueden ser mayores a la suma de las pérdidas de cada activo. De esta manera, el riesgo de una cartera siempre es menor o igual a la suma de los riesgos de sus componentes. Con el siguiente ejemplo podemos ver que el $V@R_\alpha$ no satisface siempre esta propiedad.

Ejemplo 2.2.3: Consideremos un mercado con dos activos A y B , donde existen cuatro estados de la naturaleza, que se describen en la tabla 2.2.2. A partir A y B , se elabora el activo $C = A + B$.

	Probabilidad	Activo A	Activo B	Activo C
Estado 1	0,02	-20	-12	-32
Estado 2	0,02	-10	-1	-11
Estado 3	0,02	-1	-2	-3
Estado 4	0,94	2	1	3

Cuadro 2.2.2: Ejemplo para mostrar la no subaditividad del $V@R_\alpha$.

2.3. Medidas Coherentes de Riesgo

En este caso, tenemos para $\alpha = 5\%$:

$$\begin{aligned}V@R_\alpha(\text{Activo } A) &= 1, \\V@R_\alpha(\text{Activo } B) &= 1, \\V@R_\alpha(\text{Activo } C) &= 3.\end{aligned}$$

Luego, tenemos que $V@R_\alpha(\text{Activo } C) \geq V@R_\alpha(\text{Activo } A) + V@R_\alpha(\text{Activo } B)$. Lo que nos muestra que $V@R_\alpha$ no es siempre subaditivo.

2.3. Medidas Coherentes de Riesgo

En esta sección se dará un conjunto de axiomas que definen a las medidas coherentes de riesgo. Se darán ejemplos de ellas y se estudiará una familia particular: las medidas espectrales. Las ideas y desarrollos que siguen fueron extraídos de [1] y del Capítulo 4 de [20].

Para lo que sigue, la única suposición que haremos es que existe un espacio medible (Ω, \mathcal{F}) que representa los escenarios del mercado y consideraremos:

$$L^0 = \{X : X \text{ es v.a. en } (\Omega, \mathcal{F})\},$$

donde $X \in L^0$ representa una ganancia o pérdida debida al mantenimiento de un portafolio hasta un horizonte especificado.

Definición 2.3.1. *Sea $V \subset L^0$ un cono convexo que contiene a las constantes. Una medida de riesgo monetaria ρ , corresponde a una función $\rho : V \rightarrow \mathbb{R}$ tal que:*

1. ρ es monótona decreciente, es decir

$$(\forall X, Y \in V) \quad X(\omega) \leq Y(\omega), \forall \omega \in \Omega \Rightarrow \rho(Y) \leq \rho(X).$$

2. ρ es invariante por traslaciones, lo que significa

$$(\forall X \in V)(\forall c \in \mathbb{R}) \quad \rho(X + c) = \rho(X) - c.$$

Ejemplo 2.3.2: El $V@R_\alpha$ es una medida monetaria de riesgo.

Las medidas de riesgo monetarias pueden interpretarse como requerimientos de capital para una posición dada. Por ejemplo, si para un agente de inversiones su cartera tiene un $V@R_\alpha = \$1\,000\,000$, entonces la entidad reguladora correspondiente podría exigirle al agente tener en una cuenta de respaldo la totalidad o un porcentaje del $V@R_\alpha$ de su cartera, en orden de estar segura de que el agente podrá responder a sus clientes en la eventualidad de que la cartera pierda su valor. Lo mismo es cierto si se utiliza otra medida monetaria de riesgo distinta al $V@R_\alpha$.

2.3. Medidas Coherentes de Riesgo

Definición 2.3.3. Una medida monetaria de riesgo ρ se dirá *distribución-dependiente*¹ si $F_X = F_Y \Rightarrow \rho(X) = \rho(Y)$.

Ejemplo 2.3.4: Partiendo de la definición de $q_{(\cdot)}^-(\alpha)$ es trivial que si $F_X = F_Y$ entonces $V@R_\alpha(X) = V@R_\alpha(Y)$. Luego el $V@R_\alpha(\cdot)$ es una medida distribución dependiente.

Definición 2.3.5. Si ρ es una medida de riesgo monetaria, una posición financiera X se dice **aceptable** con respecto a ρ si $\rho(X) \leq 0$. Se define el conjunto de aceptación de ρ como

$$A_\rho = \{X \in V : \rho(X) \leq 0\}.$$

Si sabemos que ρ es una medida de riesgo monetaria ¿Qué propiedades satisface A_ρ ? Recíprocamente, si consideramos un conjunto $\mathcal{A} \subset L^0$ de aceptación y definimos el funcional

$$\begin{aligned} \rho_{\mathcal{A}} : L^0 &\rightarrow \mathbb{R} \cup \{-\infty, +\infty\} \\ \rho_{\mathcal{A}}(X) &= \inf\{t \in \mathbb{R} : X + t\mathbf{1} \in \mathcal{A}\}, \end{aligned}$$

donde $\mathbf{1}$ representa a la variable aleatoria constante igual a 1. Es natural preguntarse cuáles son las condiciones a imponer sobre \mathcal{A} , para que $\rho_{\mathcal{A}}$ sea una medida de riesgo monetaria. Las respuestas a estas preguntas las contestaremos un poco más adelante.

Ahora pasamos a definir las medidas coherentes de riesgo.

Definición 2.3.6. (Axiomas de Coherencia) Una medida de riesgo monetaria ρ , definida como antes, es **coherente** si:

1. ρ es monótona decreciente

$$(\forall X, Y \in V) \quad X(\omega) \leq Y(\omega), \forall \omega \in \Omega \Rightarrow \rho(Y) \leq \rho(X).$$

2. ρ es invariante por traslaciones

$$(\forall X \in V)(\forall c \in \mathbb{R}) \quad \rho(X + c) = \rho(X) - c.$$

3. ρ es homogénea positiva

$$(\forall X \in V)(\forall \lambda \in \mathbb{R}_+) \quad \rho(\lambda X) = \lambda \rho(X).$$

4. ρ es subaditiva

$$(\forall X, Y \in V) \quad \rho(X + Y) \leq \rho(X) + \rho(Y).$$

Cada uno de estos axiomas, o bien sintetiza un principio financiero (axiomas 1 y 4) o bien una hipótesis de como asumimos se comporta el riesgo financiero (axiomas 2y 3). Veamos cada axioma por separado:

¹Esta notación fue introducida por Cont et al. en [14] para reemplazar el termino “invariante en ley” usado por otros autores, que según los primero, puede llevar a confusión.

2.3. Medidas Coherentes de Riesgo

1. El principio capturado por este axioma es simple. Si un activo, bajo cualquier escenario es mejor que otro, entonces es menos riesgoso.
2. Notemos que si ρ satisface este axioma, entonces $\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0$. Es decir, $\rho(X)$ es exactamente el mínimo valor que hay que agregar a X para convertirla en una posición aceptable. Este axioma refuerza la idea de $\rho(X)$ como un requerimiento de capital.
3. Este axioma es difícil de interpretar y no está exento de críticas, ya que suponer que el riesgo de una posición crece linealmente con su tamaño tiende a subestimar el riesgo de liquidez que tiene una posición muy grande en un activo.
4. Este axioma captura un principio financiero fundamental. Desde el popular refrán que dice que no hay que poner todos los huevos en la misma canasta, el concepto de que diversificar disminuye el riesgo está presente. En la dirección contraria, si debido a la forma en que se cuantifica el riesgo, la suma del riesgo de las partes fuera menor que el riesgo del todo, las grandes compañías podrían dividirse en un grupo de pequeñas filiales y de este modo reducir el riesgo que reportan a las agencias supervisoras.

Observación 2.3.7: Sin perder mucha generalidad, se puede suponer la condición de normalización

$$\rho(0) = 0.$$

Notemos que si ρ satisface el tercer axioma de coherencia, esta condición equivale a $\rho(0) < \infty$. La interpretación de esta normalización es que no hay riesgo en no hacer nada.

Observación 2.3.8: Notemos que si ρ satisface los axiomas tres y cuatro de coherencia, entonces ρ es una función convexa. Si bien la recíproca no es cierta, si ρ es una función convexa entonces también captura la idea de que al diversificar el riesgo no aumenta. Consideremos por ejemplo que de nuestra inversión total destinamos un porcentaje λ a un activo X y el resto a un activo Y . Entonces si ρ es convexa tendremos que

$$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y),$$

es decir, el riesgo no se incrementa al diversificar. Supongamos que ρ además de convexa está normalizada. Entonces:

$$\begin{aligned} \rho(\lambda X) &\leq \lambda\rho(X) && \text{si } \lambda \in [0, 1], \\ \rho(\lambda X) &\geq \lambda\rho(X) && \text{si } \lambda \geq 1. \end{aligned}$$

pues si $\lambda \in [0, 1]$ se puede considerar la combinación convexa entre X y 0 y obtener la primera desigualdad, mientras que si $\lambda \geq 1$ entonces $\lambda' = 1/\lambda \in [0, 1]$ y si consideramos $X' = \lambda X$ podemos obtener la segunda desigualdad a partir de la primera. Estas dos desigualdades capturan mejor como aumenta o disminuye el riesgo de liquidez de una posición al cambiar su tamaño.

2.3. Medidas Coherentes de Riesgo

Las ideas contenidas en la última observación nos llevan a considerar las medidas de riesgo convexas.

Definición 2.3.9. *Una medida convexa de riesgo es una medida de riesgo monetaria convexa.*

La relación entre una medida monetaria ρ y su conjunto de aceptación A_ρ y la relación entre un conjunto dado y la medida construida a partir de él, están caracterizadas por las siguientes proposiciones, cuyas demostraciones se pueden encontrar en el Capítulo 4 de [20].

Proposición 2.3.10. *Supongamos que $\rho : V \subset L^0 \rightarrow \mathbb{R}$ es una medida monetaria de riesgo con conjunto de aceptación A_ρ , entonces se tiene:*

a) A_ρ es no vacío y satisface

$$\inf\{m \in \mathbb{R} : m\mathbf{1} \in A_\rho\} > -\infty, \quad (2.3.1)$$

$$X \in A_\rho, Y \in V, Y \geq X \Rightarrow Y \in A_\rho. \quad (2.3.2)$$

Más aún, A_ρ satisface la siguiente propiedad de clausura: Si $X \in A_\rho$ y $Y \in V$ entonces

$$\{\lambda \in [0, 1] : \lambda X + (1 - \lambda)Y \in A_\rho\} \text{ es cerrado en } [0, 1]. \quad (2.3.3)$$

b) ρ se puede recuperar a partir de A_ρ como

$$\rho(X) = \inf\{m \in \mathbb{R} : X + m\mathbf{1} \in A_\rho\}.$$

c) ρ es una medida de riesgo convexa si y sólo si A_ρ es convexo.

d) ρ es homogénea positiva si y sólo si A_ρ es un cono. En particular, ρ es coherente si y sólo si A_ρ es un cono convexo.

Proposición 2.3.11. *Asumamos que \mathcal{A} es un subconjunto no vacío de L^0 que satisface (2.3.1) y (2.3.2). Entonces el funcional*

$$\rho_{\mathcal{A}}(X) = \inf\{t \in \mathbb{R} : X + t\mathbf{1} \in \mathcal{A}\},$$

satisface las siguientes propiedades:

a) $\rho_{\mathcal{A}}$ es una medida monetaria de riesgo.

b) Si \mathcal{A} es convexo, entonces $\rho_{\mathcal{A}}$ es una medida de riesgo convexa.

c) Si \mathcal{A} es un cono, entonces $\rho_{\mathcal{A}}$ es homogénea positiva. En particular si \mathcal{A} es un cono convexo, entonces $\rho_{\mathcal{A}}$ es coherente.

d) \mathcal{A} es un subconjunto de $A_{\rho_{\mathcal{A}}}$ y si \mathcal{A} satisface la propiedad de clausura (2.3.3) entonces $A_{\rho_{\mathcal{A}}} = \mathcal{A}$.

2.3. Medidas Coherentes de Riesgo

Ejemplo 2.3.12: Consideremos la siguiente medida de riesgo:

$$\rho_{max}(X) = - \inf_{\omega \in \Omega} X(\omega).$$

Notemos que $\rho_{max}(X) \leq 0$ si y sólo si $X \geq 0$, es decir el conjunto de aceptación de ρ_{max} corresponde al cono convexo de las variables aleatorias positivas. Luego, de la proposición 2.3.10, tenemos que ρ_{max} es una medida coherente de riesgo. Notemos que si ρ es cualquier otra medida monetaria de riesgo normalizada, entonces:

$$\rho(X) \leq \rho(\inf_{\omega \in \Omega} X(\omega)) = - \inf_{\omega \in \Omega} X(\omega) = \rho_{max}(X),$$

de donde es claro que ρ_{max} es la medida de riesgo más conservadora que se puede considerar.

Ejemplo 2.3.13: Supongamos que dotamos a (Ω, \mathcal{F}) de una medida de probabilidad \mathbb{P} . Consideremos el conjunto:

$$A_\lambda = \{X \in L^0 : \mathbb{P}(X \leq 0) < \lambda\},$$

llamemos ρ_λ a ρ_{A_λ} . Entonces

$$\begin{aligned} \rho_\lambda(X) &= \inf\{m \in \mathbb{R} : \mathbb{P}(X + m \leq 0) < \lambda\} \\ &= \inf\{-m \in \mathbb{R} : \mathbb{P}(X \leq m) < \lambda\} \\ &= -\sup\{m \in \mathbb{R} : \mathbb{P}(X \leq m) < \lambda\} \\ &= -\inf\{m \in \mathbb{R} : \mathbb{P}(X \leq m) \geq \lambda\} \\ &= -q_X^-(\lambda) \\ &= V@R_\lambda. \end{aligned}$$

Recordemos que del ejemplo 2.2.3 sabemos que el $V@R_\lambda$ no es subaditivo, luego como sí se tiene que es homogéneo positivo, entonces $V@R_\lambda$ no puede ser convexo, luego por la proposición 2.3.11 tenemos que A_λ no puede ser un conjunto convexo.

Gracias al análisis convexo, existe una rica teoría de representaciones duales para las medidas de riesgo convexas que se puede encontrar en el Capítulo 4 de [20] y en el Capítulo 1 de [39].

A continuación, y antes de pasar a las medidas espectrales, estudiemos un ejemplo de una medida coherente de riesgo, que será el bloque fundamental para la construcción de las medidas espectrales.

2.3.1. Expected Shortfall: El ejemplo central de medida coherente de riesgo

Antes de definir el *expected shortfall* veamos, con un ejemplo, la motivación para introducir esta medida de riesgo. Una vez más supongamos un mercado con dos activos A y B y cuatro estados de la naturaleza, que se resumen en la tabla 2.3.1. En este ejemplo se ve explícitamente como el $V@R_\alpha$ ignora el tamaño de las pérdidas posibles en el peor $\alpha\%$ de los casos, pues si bien ambos activos tienen el mismo $V@R_\alpha$, es claro que el activo A tiene pérdidas mucho más severas que el activo B en el peor $\alpha\%$ de los casos.

	Probabilidad	Activo A	Activo B
Estado 1	0,02	-20	-2
Estado 2	0,02	-10	-3
Estado 3	0,02	-1	-1
Estado 4	0,94	2	1
$V@R_\alpha$		1	1

Cuadro 2.3.1: Ejemplo para mostrar como el $V@R_\alpha$ puede subestimar las pérdidas de una posición. En este ejemplo se considera $\alpha = 0,05$

Para solucionar esta falencia conceptual del $V@R_\alpha$ distintos autores han definido diversas medidas de riesgo que se basan en el promedio de las pérdidas más allá del $V@R_\alpha$. Si bien dichas medidas son parecidas y coinciden bajo determinados supuestos acerca de la distribución de X , para lo que haremos nos quedamos con el *expected shortfall*. Para una discusión acerca de otras medidas similares se puede consultar [2].

Definición 2.3.14. Se define el **Expected Shortfall** de nivel alfa para un activo X como:

$$ES_\alpha(X) = -\frac{1}{\alpha} \left[\mathbb{E} \left(X \mathbb{1}_{X \leq q_X^-(\alpha)} \right) + q_X^-(\alpha) (\alpha - \mathbb{P}(X \leq q_X^-(\alpha))) \right].$$

El ES_α es simplemente el promedio de las pérdidas que son más severas que el $V@R_\alpha$. Notemos que en caso de que la función distribución de X tenga una discontinuidad en $q_X^-(\alpha)$, es decir en caso de que $\mathbb{P}(X = q_X^-(\alpha)) > 0$, se tiene que $F_X(q_X^-(\alpha)) > \alpha$, como se aprecia en la figura (2.3.1). Luego el segundo término de la derecha corresponde a una corrección para tomar en cuenta este “sobrepeso” que se le está dando al punto $q_X^-(\alpha)$.

Para probar que el ES_α es una medida coherente de riesgo, nos será muy útil la siguiente representación.

Proposición 2.3.15. Si X es una variable real definida en $(\Omega, \mathcal{F}, \mathbb{P})$ con $\mathbb{E}(X^-) < \infty$ y $\alpha \in (0, 1)$ es fijo, entonces:

$$ES_\alpha(X) = -\bar{x}(\alpha) = -\frac{1}{\alpha} \int_0^\alpha q_X^-(u) du.$$

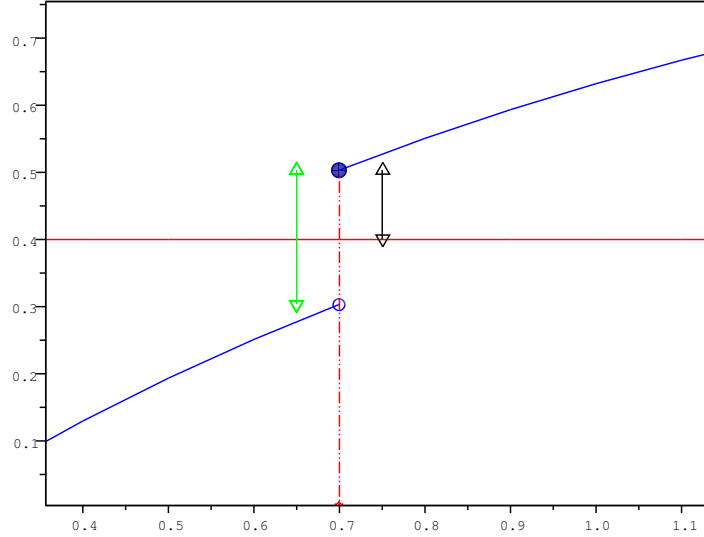


Figura 2.3.1: **Función de Distribución Discontinua.** En azul la función distribución de X . La línea continua roja representa el nivel α , mientras que la línea punteada roja muestra $q_X^-(\alpha)$. El trazo verde corresponde a $\mathbb{P}(X = q_X^-(\alpha))$ y el trazo negro corresponde a $\mathbb{P}(X \leq q_X^-(\alpha)) - \alpha$.

Antes de probar esta proposición, notemos que a partir de esta representación para el ES_α , es claro que esta medida no depende de la definición de cuantil que se considere. En efecto, basta recordar que si fijamos X , entonces $q_X^-(\cdot) = q_X^+(\cdot)$ c.t.p en $[0, 1]$.

Demostración: Consideremos una variable aleatoria U Uniforme(0, 1). Por el lema 2.1.6, sabemos que $q_X^-(U)$ tiene la misma distribución que X . Notemos además que como la función $u \rightarrow q_X^-(u)$ es no decreciente, se tiene lo siguiente:

$$\{U \leq \alpha\} \subset \{q_X^-(U) \leq q_X^-(\alpha)\}, \quad (2.3.4)$$

$$\{U > \alpha\} \cap \{q_X^-(U) \leq q_X^-(\alpha)\} = \{U > \alpha\} \cap \{q_X^-(U) = q_X^-(\alpha)\}. \quad (2.3.5)$$

Ahora notemos que:

$$\begin{aligned} \int_0^\alpha q_X^-(u) du &= \mathbb{E} \left(q_X^-(U) \mathbb{1}_{\{U \leq \alpha\}} \right) \\ &= \mathbb{E} \left(q_X^-(U) \mathbb{1}_{\{q_X^-(U) \leq q_X^-(\alpha)\}} \right) - \mathbb{E} \left(q_X^-(U) \mathbb{1}_{\{U > \alpha\} \cap \{q_X^-(U) \leq q_X^-(\alpha)\}} \right) \\ &= \mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - \mathbb{E} \left(q_X^-(U) \mathbb{1}_{\{U > \alpha\} \cap \{q_X^-(U) \leq q_X^-(\alpha)\}} \right), \end{aligned}$$

2.3. Medidas Coherentes de Riesgo

donde la última igualdad se obtiene porque X y $q_X^-(U)$ tienen la misma distribución. Ahora notemos que de la ecuación (2.3.5)

$$\begin{aligned}
 \int_0^\alpha q_X^-(u) du &= \mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - \mathbb{E} \left(q_X^-(U) \mathbb{1}_{\{U > \alpha\} \cap \{q_X^-(U) = q_X^-(\alpha)\}} \right) \\
 &= \mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - q_X^-(\alpha) \mathbb{P} \left(U > \alpha, q_X^-(U) = q_X^-(\alpha) \right) \\
 &= \mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - q_X^-(\alpha) \mathbb{P} \left(\alpha < U < F_X(q_X^-(\alpha)) \right) \\
 &= \mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - q_X^-(\alpha) \left(F_X(q_X^-(\alpha)) - \alpha \right).
 \end{aligned}$$

La última igualdad se tiene porque U tiene una distribución uniforme en el intervalo $[0, 1]$. Finalmente dividiendo por α la última ecuación

$$\frac{1}{\alpha} \int_0^\alpha q_X^-(u) du = \frac{1}{\alpha} \left(\mathbb{E} \left(X \mathbb{1}_{\{X \leq q_X^-(\alpha)\}} \right) - q_X^-(\alpha) \left(\mathbb{P}(X \leq q_X^-(\alpha)) - \alpha \right) \right).$$

q.e.d.

La última proposición nos permite probar que $ES_\alpha(\cdot)$ es una medida coherente de riesgo. En efecto,

1. El ES_α es monótono: Si $X \leq Y$, entonces $q_X^-(\alpha) \leq q_Y^-(\alpha)$, de donde es directo $\bar{x}_{(\alpha)} \leq \bar{y}_{(\alpha)}$. Esto a su vez implica que $ES_\alpha(Y) \leq ES_\alpha(X)$.
2. El ES_α es invariante por traslaciones: Sea $Y = X + c$, entonces como $q_Y^-(\alpha) = q_X^-(\alpha) + c$, se tiene que $\bar{y}_{(\alpha)} = \bar{x}_{(\alpha)} + c$. Con lo que se concluye que $ES_\alpha(X + c) = ES_\alpha(X) - c$.
3. El ES_α es homogéneo positivo: Sea $Y = \lambda X$, $\lambda > 0$. Entonces como $q_Y^-(\alpha) = \lambda q_X^-(\alpha)$, se tiene que $\bar{y}_{(\alpha)} = \lambda \bar{x}_{(\alpha)}$. Con esto que se concluye que $ES_\alpha(\lambda X) = \lambda ES_\alpha(X)$.
4. El ES_α es subaditivo: La prueba de la subaditividad del $ES_\alpha(\cdot)$ se desprende directamente de la proposición 2.1.8 y la representación que se probó en la proposición anterior.

2.3.2. Medidas de Riesgo Espectrales

Una medida de riesgo espectral es una medida de la forma:

$$\rho_\phi(X) = - \int_0^1 q_X^-(u) \phi(u) du, \tag{2.3.6}$$

donde ϕ es un elemento de L^1 tal que $\int_0^1 \phi = 1$. Notemos que ρ_ϕ depende de X sólo a través de $q_X^-(\cdot)$, luego todas las medidas de riesgo espectrales son distribución dependientes. Además, como $q_X^-(\cdot) = q_X^+(\cdot)$ c.t.p. en $[0, 1]$, tenemos que las medidas de riesgo espectrales no

2.3. Medidas Coherentes de Riesgo

dependen de la definición de cuantil que se use.

Notemos que el ES_α es una medida de riesgo espectral con $\phi = \frac{1}{\alpha} \mathbb{1}_{[0, \alpha]}$ y abusando un poco del lenguaje, podríamos decir que el $V@R_\alpha$ también corresponde a una medida de riesgo espectral con $\phi = \delta_\alpha$. A continuación caracterizaremos las medidas espectrales que son coherentes. Para esto pasaremos por varias etapas previas, la primera de ellas corresponde a aprender a generar nuevas medidas de riesgo coherentes a partir de una familia de medidas coherentes.

Proposición 2.3.16. *Sea ρ_i una medida de riesgo coherente para $i = 1, \dots, n$. Entonces cualquier combinación de la forma $\rho = \sum_{i=1}^n \beta_i \rho_i$ con $\beta_i \geq 0$ y $\sum_i \beta_i = 1$ es una medida de riesgo coherente. Análogamente, si ρ_β es una familia paramétrica de medidas de riesgo para $\beta \in [a, b]$ y μ es una medida de probabilidad en $[a, b]$ entonces $\rho = \int_a^b \rho_\beta d\mu(\beta)$ es una medida de riesgo coherente.*

Demostración: Dado que la demostración de ambas afirmaciones es muy similar sólo haremos la del caso continuo. Consideremos entonces: una familia paramétrica ρ_β para $\beta \in [a, b]$, μ una medida de probabilidad en $[a, b]$ y $\rho = \int_a^b \rho_\beta d\mu(\beta)$ y probemos que ρ satisface las condiciones de una medida de riesgo coherente. En lo que sigue X y Y representan variables aleatorias reales, c una constante real y λ una constante positiva.

1. ρ es monotonamente decreciente. Supongamos que $X \leq Y$. Entonces

$$\rho_\beta(Y) \leq \rho_\beta(X) \quad \forall \beta \in [a, b],$$

de donde es directo que $\rho(Y) \leq \rho(X)$.

2. ρ es invariante por traslaciones. En efecto,

$$\rho(X + c) = \int_a^b \rho_\beta(X + c) d\mu(\beta) = \int_a^b (\rho_\beta(X) - c) d\mu(\beta) = \rho(X) - c.$$

3. ρ es homogénea positiva. En efecto,

$$\rho(\lambda X) = \int_a^b \rho_\beta(\lambda X) d\mu(\beta) = \int_a^b (\lambda \rho_\beta(X)) d\mu(\beta) = \lambda \rho(X).$$

4. ρ es subaditiva. Notemos que $\forall \alpha \in [a, b]$, $\rho_\alpha(X + Y) \leq \rho_\alpha(X) + \rho_\alpha(Y)$. Integrando a ambos lados de la desigualdad se obtiene lo deseado.

Notemos que extendiendo la definición de ES_α al caso $\alpha = 0$ como

$$ES_0(X) = -\text{ess ínf}\{X\},$$

2.3. Medidas Coherentes de Riesgo

tenemos una familia paramétrica de medidas de riesgo coherentes para $\alpha \in [0, 1]$. A partir de esto, usando la proposición anterior, tenemos que para cualquier medida de probabilidad μ en $[0, 1]$,

$$\rho_\mu(X) = \int_0^1 ES_\alpha(X) d\mu(\alpha),$$

es una medida coherente de riesgo. Sin embargo, nos restringiremos a las medidas m que además satisfacen la condición de normalización

$$\int_0^1 \alpha m(d\alpha) = 1,$$

es decir la medida αm es también una medida de probabilidad. Pues en este caso, la medida

$$\rho_m(X) = \int_0^1 \alpha ES_\alpha(X) dm(\alpha),$$

es una medida coherente de riesgo y también una medida espectral. En efecto,

$$\begin{aligned} \rho_m &= \int_0^1 \alpha ES_\alpha(X) dm(\alpha) \\ &= \int_0^1 \left(- \int_0^\alpha q_X^-(u) du \right) m(d\alpha) \\ &= - \int_0^1 q_X^-(u) \int_u^1 m(d\alpha) du \\ &= - \int_0^1 q_X^-(u) \phi(u) du, \end{aligned}$$

donde ϕ corresponde al espectro $\phi = \int_u^1 m(d\alpha)$. Este satisface:

$$\begin{aligned} \phi(u) &\geq 0 \quad \forall u \in [0, 1], \\ \phi(\cdot) &\text{ es decreciente y} \\ \int_0^1 \phi(u) du &= 1. \end{aligned}$$

La última condición viene de la condición de normalización impuesta para m . En efecto,

$$\int_0^1 \phi(u) du = \int_0^1 \int_u^1 dm(\alpha) du = \int_0^1 \left(\int_0^\alpha du \right) dm(\alpha) = \int_0^1 \alpha dm(\alpha) = 1.$$

Supongamos ahora que tenemos una función ϕ que satisface las condiciones anteriores, es decir, positividad, decrecimiento y normalización ¿podemos construir una medida coherente de riesgo a partir de ϕ ?

2.3. Medidas Coherentes de Riesgo

Antes de responder la última pregunta, notemos que más que conocer ϕ punto a punto, necesitamos conocer sus propiedades como elemento de $L^1([0, 1])$. Sin embargo, las nociones de positividad y decrecimiento en el sentido clásico no son directamente transferibles a $\phi \in L^1([0, 1])$. Por esto consideremos las siguientes definiciones.

Definición 2.3.17. Sea ϕ un elemento de $L^1([a, b], dx)$.

1. Diremos que ϕ es **positivo** si $\forall I \subset [a, b]$

$$\int_I \phi \geq 0.$$

2. Diremos que ϕ es **decreciente** si $\forall q \in [a, b]$, y $\forall \varepsilon > 0$, tal que $[q - \varepsilon, q + \varepsilon] \subset [a, b]$

$$\int_{q-\varepsilon}^q \phi \geq \int_q^{q+\varepsilon} \phi.$$

Definición 2.3.18. Un elemento $\phi \in L^1([0, 1])$ se dice espectro de riesgo admisible si:

1. ϕ es positivo.

2. ϕ es decreciente.

3. $\|\phi\| = 1$.

Teorema 2.3.19. Sea ρ_ϕ definida como

$$\rho_\phi(X) = - \int_0^1 q_X^-(u) \phi(u) du,$$

con $\phi \in L^1$. Entonces ρ_ϕ es una medida coherente de riesgo si y sólo si ϕ es un espectro admisible.

Demostración: Probemos primero que si ρ_ϕ es una medida de riesgo coherente entonces ϕ es un espectro admisible. Para esto probaremos la contrareciproca, es decir que si ϕ no es admisible, entonces ρ_ϕ no es medida coherente.

Probemos primero que si ϕ no es positiva entonces ρ_ϕ no es monótona. Supongamos que existe $I = [q_1, q_2] \subset (0, 1)$ tal que

$$\int_I \phi(p) dp < 0.$$

Consideremos un mercado con tres posibles escenarios, representado por un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Supongamos que tenemos dos activos en este mercado, X e Y . Supongamos además que \mathbb{P} está definida como se ve en la siguiente tabla:

2.3. Medidas Coherentes de Riesgo

ω	$\mathbb{P}(\omega)$	$X(\omega)$	$Y(\omega)$
ω_1	q_1	X_1	$Y_1 = X_1$
ω_2	$q_2 - q_1$	X_2	$Y_2 = X_2 + a$
ω_3	$1 - q_2$	X_3	$Y_3 = X_3$

Supongamos también que se satisface $X_1 < X_2 < X_3$ y $Y_1 < Y_2 < Y_3$ y $a > 0$. Podemos calcular entonces la función cuantil de de X e Y :

p	$q_X^-(p)$	$q_Y^-(p)$
$p \in (0, q_1]$	X_1	Y_1
$p \in (q_1, q_2]$	X_2	Y_2
$p \in (q_2, 1]$	X_3	Y_3

A partir de la tabla anterior es fácil calcular:

$$\rho_\phi(Y) - \rho_\phi(X) = - \int_0^1 (q_Y^-(p) - q_X^-(p)) \phi(p) dp = -a \int_I \phi(p) dp > 0.$$

Es decir, $X < Y$ y $\rho_\phi(X) < \rho_\phi(Y)$. Luego, ρ_ϕ no es monotonamente decreciente.

Veamos ahora que si ϕ no es decreciente, entonces ρ_ϕ no es subaditiva. Para esto supongamos que $\exists q \in (0, 1)$ y ε tal que $[q - \varepsilon, q + \varepsilon] \subset (0, 1)$ y

$$\int_{q-\varepsilon}^q \phi(p) dp < \int_q^{q+\varepsilon} \phi(p) dp.$$

Consideremos ahora un mercado con cuatro posibles escenarios, representado por un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Supongamos que tenemos dos activos en este mercado, X e Y . Supongamos además que \mathbb{P} está definida como se ve en la siguiente tabla:

ω	$\mathbb{P}(\omega)$	$X(\omega)$	$Y(\omega)$	$Z(\omega)$
ω_1	$q - \varepsilon$	X_1	Y_1	$Z_1 = X_1 + Y_1$
ω_2	ε	X_2	Y_3	$Z_2 = X_2 + Y_3$
ω_3	ε	X_3	Y_2	$Z_3 = X_3 + Y_2$
ω_4	$1 - q - \varepsilon$	X_4	Y_4	$Z_4 = X_4 + Y_4$

Supongamos también que se satisface $X_1 < X_2 < X_3 < X_4$, $Y_1 < Y_2 < Y_3 < Y_4$, $Z_1 < Z_2 < Z_3 < Z_4$, y además

$$Z_2 = X_2 + Y_3 < X_3 + Y_2 = Z_3.$$

Con esto es podemos calcular las funciones cuantiles de X, Y y Z :

2.3. Medidas Coherentes de Riesgo

p	$q_X^-(p)$	$q_Y^-(p)$	$q_Z^-(p)$
$p \in (0, q - \varepsilon] \equiv I_1$	X_1	Y_1	Z_1
$p \in (q - \varepsilon, q] \equiv I_2$	X_2	Y_2	Z_2
$p \in (q, q + \varepsilon] \equiv I_3$	X_3	Y_3	Z_3
$p \in (q + \varepsilon, 1] \equiv I_4$	X_4	Y_4	Z_4

Luego,

$$\begin{aligned}
 \rho_\phi(Z) - \rho_\phi(X) - \rho_\phi(Y) &= - \int_0^1 (q_Z^-(p) - q_X^-(p) - q_Y^-(p)) \phi(p) dp \\
 &= - \sum_{i=1}^4 \int_{I_i} (Z_i - X_i - Y_i) \phi(p) dp \\
 &= - \left(\int_{I_2} (Z_2 - X_2 - Y_2) \phi(p) dp + \int_{I_3} (Z_3 - X_3 - Y_3) \phi(p) dp \right) \\
 &= - \left((X_2 + Y_3 - X_2 - Y_2) \int_{I_2} \phi(p) dp \right. \\
 &\quad \left. + (X_3 + Y_2 - X_3 - Y_3) \int_{I_3} \phi(p) dp \right) \\
 &= -(Y_3 - Y_2) \left(\int_{I_2} \phi(p) dp - \int_{I_3} \phi(p) dp \right) \\
 &> 0.
 \end{aligned}$$

Es decir, $\rho_\phi(X + Y) > \rho_\phi(X) + \rho_\phi(Y)$. Así, si ϕ no es decreciente, ρ_ϕ no puede ser subaditiva.

Finalmente veamos que si $\|\phi\| = 1$ entonces ρ_ϕ es invariante a traslaciones. Para esto recordemos que si $a \in \mathbb{R}$, X es una variable aleatoria y consideramos $Y = X + a$ entonces $q_Y^-(u) = q_X^-(u) + a \forall u \in [0, 1]$. Luego tenemos que

$$\rho_\phi(Y) = - \int_0^1 q_Y^-(u) \phi(u) du = - \int_0^1 (q_X^-(u) + a) \phi(u) du = \rho_\phi(X) - a \int_0^1 \phi(u) du,$$

con lo que concluimos que

$$\rho_\phi(X + a) = \rho_\phi(X) - a \iff \int_0^1 \phi(u) du = 1.$$

Probemos ahora que si ϕ es un espectro admisible, entonces ρ_ϕ es una medida coherente. Empecemos por notar que si ϕ es un espectro admisible podemos encontrar un representante positivo, decreciente, continuo a la derecha y que cumpla $\phi(1) = 0$. De este modo, si llamamos μ a la medida de Lebesgue-Stieltjes inducida por la función $\phi(0) - \phi(\cdot)$, tenemos por definición

$$\phi(p) = \int_p^1 \mu(d\alpha).$$

Entonces

$$\begin{aligned}
 \rho_\phi(X) &= - \int_0^1 q_{\bar{X}}(p) \phi(p) dp \\
 &= - \int_0^1 q_{\bar{X}}(p) \int_p^1 \mu(d\alpha) dp \\
 &= - \int_0^1 \int_p^1 q_{\bar{X}}(p) \mu(d\alpha) dp \\
 &= - \int_0^1 \int_0^\alpha q_{\bar{X}}(p) dp \mu(d\alpha) \\
 &= \int_0^1 ES_\alpha(X) \alpha \mu(d\alpha).
 \end{aligned}$$

Así, basta con probar que $\int_0^1 \alpha \mu(d\alpha) = 1$ y por la proposición (2.3.16) tendríamos que ρ_ϕ es una medida coherente de riesgo. En efecto,

$$\int_0^1 \alpha \mu(d\alpha) = \int_0^1 \int_0^\alpha dp \mu(d\alpha) = \int_0^1 \int_p^1 \mu(d\alpha) dp = \int_0^1 \phi(p) dp = 1.$$

q.e.d.

Con esto tenemos que las medidas espectrales coherentes están identificadas con los espectros admisibles.

2.3.3. Espectros de riesgo como medidas cuantitativas de la aversión al riesgo.

Notemos que ES_α lo podemos representar como una medida espectral tomando como espectro a $\phi = \frac{1}{\alpha} \mathbb{1}_{[0,\alpha]}$ lo que nos muestra que esta medida de riesgo promedia los peores resultados de un activo X ponderando de igual manera a todos los resultados que caen en la cola izquierda de nivel α . Por otro lado, las medidas de riesgo espectrales promedian las pérdidas con distintas ponderaciones según el cuantil al que pertenecen. Notemos que las condiciones de admisibilidad nos dicen que la suma de los ponderadores es uno, que se ponderan las pérdidas con pesos mayores o iguales que cero y que los peores casos se ponderan con pesos mayores, condiciones que vistas así parecen naturales. Pero, ¿por qué se querría ponderar los resultados con distintos pesos? Respondamos a esta pregunta con un ejemplo.

Ejemplo 2.3.20: Consideremos dos activos A y B en un mundo con seis estados de la naturaleza. El comportamiento de los activos se resumen en la tabla 2.3.2. Los dos activos que se consideran tienen el mismo retorno esperado y si medimos el riesgo con el $ES_{0,05}$

2.3. Medidas Coherentes de Riesgo

	Probabilidad	Activo A	Activo B
Estado 1	0,01	-1000	-210
Estado 2	0,01	-4	-205
Estado 3	0,01	-3	-204
Estado 4	0,01	-2	-203
Estado 5	0,01	-1	188
Estado 6	0,95	500	500
Retorno Medio		464,9	464,9
$ES_{0,05}$		-202	-202

Cuadro 2.3.2: Ejemplo de como dos activos con el mismo retorno medio y el mismo $ES_{0,05}$ son distintos para individuos con distinto grado de aversión al riesgo.

ambos activos parecen tener el mismo riesgo.

Sin embargo, ¿son los dos activos equivalentes?. La respuesta a esta pregunta es “depende”. Depende del grado de aversión al riesgo que tenga el inversionista que esta analizando estos activos. En el caso de que el inversionista sea muy averso al riesgo, el activo A le parecerá más riesgoso, ya que hay un escenario donde las pérdidas del activo son mucho mayores a las del activo B. Para este tipo de inversionista tiene más sentido considerar ρ_ϕ , una medida de riesgo, donde ϕ puede ser, por ejemplo,

$$\phi(p) = \frac{1}{2\sqrt{\alpha p}} \mathbb{1}_{[0,\alpha]}(p).$$

Notemos que en este caso $\rho_\phi(A) = 448,7$ mientras que $\rho_\phi(B) = 205,1$. Luego esta medida captura de mejor forma la aversión al riesgo de un inversionista conservador.

Con el ejemplo anterior, vemos que a través del espectro se puede capturar la actitud hacia el riesgo de los inversores, que no es en todos igual. Es esta flexibilidad de las medidas espectrales lo que las hace interesantes.

Capítulo 3

Convergencia y Métricas en el Espacio de Medidas de Probabilidad

Como ya se mencionó en la introducción, el objetivo de esta memoria es establecer cotas cuantitativas para

$$\mathbb{P}(|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)| > \varepsilon),$$

o más en general para

$$\mathbb{P}(|\rho_\phi(X) - \widehat{\rho}_\phi^N(X)| > \varepsilon), \tag{3.0.1}$$

donde ρ_ϕ es una medida de riesgo espectral y $\widehat{\rho}_\phi^N$ es su estimador histórico, que corresponde a aplicar la medida espectral ρ_ϕ a la distribución empírica de X , como se detallará en el capítulo 6.

Para establecer cotas como (3.0.1), nuestra principal herramienta serán los resultados de concentración que estudiaremos en el capítulo 5. Para poder establecer dichos resultados, es preciso contar con tres objetos matemáticos:

- Las medidas empíricas.
- Diferentes distancias para medidas de probabilidad.
- Desigualdades de Talagrand.

En este capítulo introduciremos los dos primeros puntos de esta lista. El tercero, lo dejamos para el capítulo siguiente.

3.1. Convergencia débil de medidas de probabilidad.

En lo que sigue $(\mathcal{X}, d, \mathcal{B})$ será un espacio polaco medible, es decir (\mathcal{X}, d) es un espacio métrico polaco dotado con su sigma álgebra de Borel \mathcal{B} . Llamamos $\mathcal{P}(\mathcal{X})$ al conjunto de

3.1. Convergencia débil de medidas de probabilidad.

todas las medidas de probabilidad en \mathcal{X} . Notemos que $\mathcal{P}(\mathcal{X})$ es un subconjunto convexo de $(\mathcal{C}_b(\mathcal{X}))^*$, el dual topológico de las funciones continuas y acotadas definidas en \mathcal{X} .

Se dota a $\mathcal{P}(\mathcal{X})$ de la topología débil-* inducida por $\mathcal{C}_b(\mathcal{X})$, es decir si $\{\mu_k\}_{k \in \mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ es una sucesión de medidas de probabilidad y $\mu \in \mathcal{P}(\mathcal{X})$, decimos que $\{\mu_k\}_{k \in \mathbb{N}}$ converge débilmente a μ , lo que denotamos $\mu_k \rightharpoonup \mu$, si

$$(\forall \phi \in \mathcal{C}_b(\mathcal{X})) \quad \int_{\mathcal{X}} \phi(x) d\mu_k(x) \xrightarrow{k} \int_{\mathcal{X}} \phi(x) d\mu(x).$$

Existen varias caracterizaciones equivalentes para la convergencia débil de medidas de probabilidad, las cuales quedan resumidas en el Teorema de Portmanteau, cuya demostración aparece en el Capítulo 1 de [6].

Teorema 3.1.1. (Portmanteau) *Consideremos $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ y $\mu \in \mathcal{P}(\mathcal{X})$, entonces son equivalentes:*

1. $\mu_n \rightharpoonup \mu$.
2. $\lim_n \int_{\mathcal{X}} f d\mu_n = \int_{\mathcal{X}} f d\mu$ para toda $f : \mathcal{X} \rightarrow \mathbb{R}$ acotada y uniformemente continua.
3. $\lim_n \int_{\mathcal{X}} f d\mu_n = \int_{\mathcal{X}} f d\mu$ para toda $f : \mathcal{X} \rightarrow \mathbb{R}$ acotada y Lipschitz.
4. $\limsup_n \mu_n(F) \leq \mu(F)$ para todo cerrado F .
5. $\liminf_n \mu_n(G) \geq \mu(G)$ para todo abierto G .
6. $\lim_n \mu_n(A) = \mu(A)$ para todo A tal que $\mu(\partial A) = 0$, donde ∂A es la frontera de A .

Recordemos que una medida de probabilidad μ la podemos asociar a una variable aleatoria X , definida en algún $(\Omega, \mathcal{F}, \mathbb{P})$, tal que para todo $A \subset \mathcal{X}$ medible, $\mu(A) = \mathbb{P}(X \in A)$. En el caso de variables aleatorias reales, es más usual trabajar con funciones de distribución que con leyes. Una consecuencia del teorema anterior y cuya demostración aparece en el Capítulo 2 de [26] es el siguiente lema, que caracteriza la convergencia débil de medidas de probabilidad en \mathbb{R} .

Lema 3.1.2. *Si $\{\mu_n\}_{n \geq 1}$ es una sucesión de medidas de probabilidad en \mathbb{R} y μ es otra medida de probabilidad en \mathbb{R} , entonces*

$$\mu_n \rightharpoonup \mu \Leftrightarrow F_n(x) \rightarrow F(x) \text{ para todo } x \text{ punto de continuidad de } F,$$

donde F_n es la función de distribución asociada a μ_n y F es la función de distribución asociada a μ .

En lo que sigue nos serán de interés ciertas sucesiones de medidas de probabilidad, las que llamamos medidas empíricas y que definimos a continuación.

3.1. Convergencia débil de medidas de probabilidad.

Definición 3.1.3. Sea $\mu \in \mathcal{P}(\mathcal{X})$. Consideremos una muestra de tamaño N de μ , es decir: X_1, \dots, X_N variables aleatorias con valores en \mathcal{X} independientes entre si e idénticamente distribuidas según μ . La medida empírica de la muestra corresponde a

$$\hat{\mu}^N(A) = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}(A) \quad \forall A \in \mathcal{B},$$

donde para todo $x \in \mathcal{X}$, δ_x denota la masa de Dirac en x .

Notemos ahora que por la ley fuerte de los grandes números, para cada $\phi \in \mathcal{C}_b(\mathcal{X})$ se tiene casi seguramente que

$$\int_{\mathcal{X}} \phi(x) d\hat{\mu}^N(x) = \frac{1}{N} \sum_{i=1}^N \phi(X_i) \xrightarrow{N} \int_{\mathcal{X}} \phi(x) d\mu(x).$$

Sin embargo, no es inmediato deducir que casi seguramente $\hat{\mu}^N \rightarrow \mu$, pues el conjunto donde la integral converge depende de ϕ . No obstante lo anterior, se tiene el siguiente teorema, cuya demostración se puede encontrar en el Capítulo 11 de [18].

Teorema 3.1.4. [Varadarajan] Sea (\mathcal{X}, d) un espacio métrico separable y μ una medida de probabilidad de Borel en \mathcal{X} . Entonces, casi seguramente $\hat{\mu}^N \rightarrow \mu$.

Una consecuencia de este teorema es el siguiente resultado que se debe a V.I. Glivenko y F.P. Cantelli, cuya demostración se encuentra en el Capítulo 11 de [18].

Teorema 3.1.5. [Glivenko-Cantelli] Sea μ una ley de probabilidad en \mathbb{R} con función distribución F . Sea F_N la función distribución empírica de una muestra $\{X_1, \dots, X_N\}$ i.i.d. con ley μ . Es decir,

$$F_N(x) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_i \leq x\}}.$$

Entonces, casi seguramente, $F_N(\cdot) \rightarrow F$ uniformemente en \mathbb{R} cuando $N \rightarrow \infty$.

Si bien estos teoremas nos dan la convergencia casi segura de la medida empírica, nada nos dicen acerca de la velocidad de dicha convergencia. Más adelante se estudiará esta velocidad bajo distintos supuestos sobre la medida límite.

A continuación se introducirán tres métricas para el espacio $\mathcal{P}(\mathcal{X})$: la distancia de Lévy-Prokhorov, la distancia dual de funciones Lipschitz acotadas y la distancia de Wasserstein. Las dos primeras son compatibles con la topología de la convergencia débil, mientras que la tercera induce una topología ligeramente más fuerte.

Por su importancia en el trabajo realizado durante esta memoria se le dará un especial énfasis a la distancia Wasserstein.

3.2. Métricas en el Espacio de Medidas de Probabilidad

3.2.1. Distancia de Lévy-Prokhorov d_P

Comencemos por definir la distancia de Lévy, que se puede considerar un caso particular de la distancia de Lévy-Prokhorov para medidas definidas en \mathbb{R} .

Definición 3.2.1. *La distancia de Lévy $d_L(\cdot, \cdot)$ entre dos medidas de probabilidad en \mathbb{R} , las que identificamos con sus funciones de distribución F y G , está dada por:*

$$d_L(F, G) := \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}.$$

Gráficamente $\sqrt{2}d_L(F, G)$ corresponde a la máxima distancia entre los gráficos de F y G medida en la dirección del vector $(-1, 1)$.

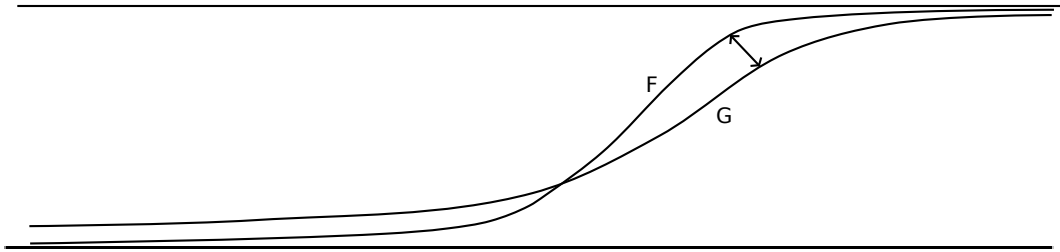


Figura 3.2.1: La distancia de Lévy corresponde a la máxima distancia entre F y G medida en la dirección del vector $(-1, 1)$ y escalada por el factor $\sqrt{2}$.

La característica más importante de la distancia de Lévy es que sirve para metrizar la convergencia débil de medidas de probabilidad en \mathbb{R} . Para ver la demostración de este resultado se puede consultar el Capítulo 2 de [26].

Antes de definir la distancia de Lévy-Prokhorov, necesitamos definir para $A \subset \mathcal{X}$ el ε -engrosamiento de A , que notamos A^ε , y está dado por:

$$A^\varepsilon = \{x \in \mathcal{X} : d(x, y) \leq \varepsilon \text{ para algún } y \in A\}.$$

Definición 3.2.2. *Sean $\mu, \nu \in \mathcal{P}(\mathcal{X})$, se define la distancia de Lévy-Prokhorov como:*

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(F) \leq \nu(F^\varepsilon) + \varepsilon \text{ para todo cerrado } F \subset X\}.$$

Así como en el caso real, para \mathcal{X} un espacio polaco general, también se tiene que la distancia de Lévy-Prokhorov metriza la convergencia débil. La prueba de este resultado también se puede encontrar en el Capítulo 2 de [26].

3.2. Métricas en el Espacio de Medidas de Probabilidad

Para terminar veamos la relación entre la distancia de Lévy y la distancia de Lévy-Prokhorov. Sean μ, ν medidas de probabilidad en \mathbb{R} y sean F y G sus funciones de distribución, entonces:

$$d_L(F, G) = \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\},$$

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(F) \leq \nu(F^\varepsilon) + \varepsilon \text{ para todo cerrado } F \subset S\}.$$

Luego, si $\varepsilon > 0$ es tal que para todo F cerrado se tiene que $\mu(F) \leq \nu(F^\varepsilon) + \varepsilon$ en particular tenemos que $\mu((-\infty, x]) \leq \nu((-\infty, x + \varepsilon]) + \varepsilon$ es decir $F(x) \leq G(x + \varepsilon) + \varepsilon$ para todo x en \mathbb{R} , de lo que concluimos

$$d_L(\mu, \nu) \leq d_P(\mu, \nu).$$

3.2.2. Distancia Dual de Funciones Lipschitz Acotadas BL

Comencemos por definir $L_b(\mathcal{X}, d)$, el conjunto de todas las funciones Lipschitz acotadas en \mathcal{X} . Dotemos a $L_b(\mathcal{X}, d)$ de la norma $\|f\|_{L_b} = \|f\|_\infty + \|f\|_L$ donde:

$$\|f\|_\infty = \sup_{\mathcal{X}} |f(x)| \quad \text{y} \quad \|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}.$$

Trivialmente $L_b(\mathcal{X}, d) \subset C_b(\mathcal{X})$, pero cuando \mathcal{X} es compacto se tiene que esta inclusión es densa para la norma $\|\cdot\|_\infty$.

Definición 3.2.3. Sean $\mu, \nu \in \mathcal{P}(\mathcal{X})$, se define la distancia de las funciones Lipschitz acotadas como:

$$BL(\mu, \nu) = \sup_{\|f\|_{L_b} \leq 1} \left\{ \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f d\nu \right\}.$$

Se puede probar que esta distancia también metriza la convergencia débil de medidas de probabilidad. Ver por ejemplo, el ya mencionado Capítulo 2 de [26], o el Capítulo 11 de [18].

3.2.3. Distancias de Wasserstein W_p

En lo que sigue se hará una breve discusión acerca del problema de transporte óptimo de masa y de cómo este induce una noción de distancia sobre el espacio de probabilidades. La discusión histórica que sigue es un resumen de las ideas que aparecen en el Capítulo 3 de [43].

El problema de Monge - Kantorovich

Supongamos que estamos en la playa construyendo un castillo de arena y que para ello necesitamos tomar arena desde posiciones x 's y llevarla hasta posiciones y 's. Si consideramos

3.2. Métricas en el Espacio de Medidas de Probabilidad

que existe un costo asociado al traslado de la arena desde una posición a otra, naturalmente surge la pregunta acerca de si la transferencia de arena puede realizarse de manera óptima. Otra versión del mismo problema es la de repartir bienes desde un número determinado de bodegas, cada una con oferta dada, a un número determinado de tiendas, cada una con una demanda específica. Cada tienda puede ser atendida por cualquier bodega, sin embargo, es natural buscar una manera óptima de realizar el despacho.

En 1781 el matemático francés Gaspard Monge publicó sus investigaciones acerca del problema de transportar la arena, en las cuales consideraba que el costo asociado a mover un grano de arena, estaba dado por el producto entre su masa y la distancia entre sus posiciones inicial y final. Un punto importante en la formulación de Monge es que los granos que comparten una posición desde donde son extraídos no pueden separarse durante su traslado, es decir que todo lo que está en la posición x debe ir a parar a sólo una posición y . Esta condición nos dice que las formas de mover el material, a las que llamaremos planes de transporte, serán funciones deterministas de la posición inicial. Con todo esto en mente, este problema lo podemos modelar suponiendo que el espacio desde donde sacaremos la arena corresponde a un conjunto \mathcal{X} y la distribución de la arena en este conjunto está dada por una medida finita μ , que sin pérdida de generalidad podemos suponer que es de probabilidad. Por otro lado, el espacio donde queremos construir el castillo corresponde a un conjunto \mathcal{Y} , donde la demanda por arena está dada por ν , que es otra medida de probabilidad. Por último, el plan de transporte lo modelamos como una función medible $T : \mathcal{X} \rightarrow \mathcal{Y}$ que para cada x indica el destino del material contenido en dicha posición. De esta forma, si $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ representa una función de costo, que en particular podría ser una distancia, el problema de Monge se puede escribir como

$$\inf \left\{ \int_{\mathcal{X}} c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\}, \quad (\text{PM})$$

donde $T_{\#}\mu$ es la medida inducida por T en \mathcal{Y} a partir de μ , es decir $T_{\#}\mu(A) = \mu(T^{-1}(A))$ para todo $A \subset \mathcal{Y}$ medible.

Es fácil notar que el problema (PM) puede estar mal puesto. Por ejemplo, si μ es una masa de Dirac y ν no lo es, no existirá un plan de transporte T tal que $T_{\#}\mu = \nu$, ya que en \mathcal{X} la masa está concentrada en un sólo punto, luego para satisfacer cualquier distribución de masa en \mathcal{Y} que no esté concentrada en un punto, será necesario separar masa que comparte una posición en \mathcal{X} , lo que viola una de las restricciones del problema.

En 1942, en [27], el matemático ruso Leonid Vitaliyevich Kantorovich propone y da condiciones de existencia para la solución del problema, que él llamó, “de traslación de masas”¹. Dicho problema aborda la misma pregunta que el problema de Monge, pero permite que masa que comparte posición en el conjunto de origen \mathcal{X} pueda separarse en el conjunto de destino \mathcal{Y} . Luego, en esta formulación, los planes de transporte ya no serán funciones medibles, si no *couplings* que definimos a continuación.

¹Originalmente publicado en ruso, la referencia es a la traducción al inglés realizada por A. N. Sobolevskii.

3.2. Métricas en el Espacio de Medidas de Probabilidad

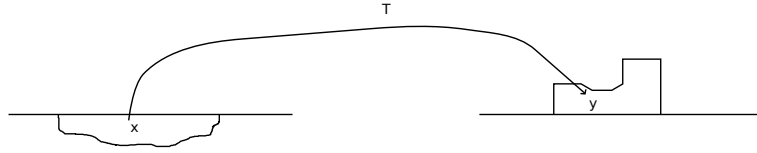


Figura 3.2.2: El problema propuesto por Monge en 1781: Trasladar de manera óptima el material desde el conjunto de la izquierda al conjunto de la derecha.

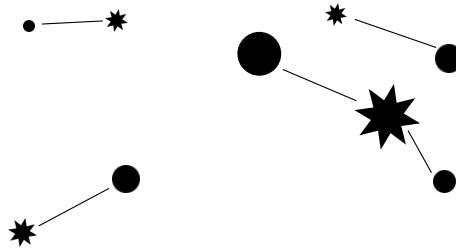


Figura 3.2.3: Otra versión del problema de Monge. Los círculos representan bodegas y las estrellas representan tiendas. El tamaño de las figuras representa el tamaño de su oferta o demanda. Las líneas entre figuras representan una forma de trasladar bienes.

Definición 3.2.4. Sean \mathcal{X} e \mathcal{Y} espacios métricos polacos, diremos que una medida de probabilidad π en $\mathcal{X} \times \mathcal{Y}$ es un *coupling* entre μ y ν medidas de probabilidad en \mathcal{X} e \mathcal{Y} respectivamente, si para todo $A \subset \mathcal{X}$ medible se tiene que $\pi(A \times \mathcal{Y}) = \mu(A)$ y para todo $B \subset \mathcal{Y}$ medible se tiene que $\pi(\mathcal{X} \times B) = \nu(B)$. Llamaremos $\Pi_{\mu, \nu}$ al conjunto de todos los *couplings* entre μ y ν . Notemos que $\Pi_{\mu, \nu}$ es siempre no vacío ya que $\mu \otimes \nu \in \Pi_{\mu, \nu}$.

Con esta definición el problema de Kantorovich, también llamado de Monge-Kantorovich se escribe como:

$$\inf \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) : \pi \in \Pi_{\mu, \nu} \right\}. \quad (\text{PMK})$$

Notemos que en (PMK) el funcional que se quiere minimizar es lineal en π y además el conjunto $\Pi_{\mu, \nu}$ es convexo, lo que hace que este problema sea mucho más fácil de resolver que (PM), en donde el funcional a minimizar depende no linealmente de los planes de transporte. A continuación se enuncia el teorema que caracteriza la existencia de soluciones para (PMK). Su demostración se puede encontrar en el Capítulo 4 de [43].

Teorema 3.2.5. Sea \mathcal{X} , \mathcal{Y} dos espacios métricos polacos y μ , ν medidas de probabilidad en \mathcal{X} e \mathcal{Y} respectivamente. Sean $a : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ y $b : \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ dos funciones semicontinuas superiores tales que $a \in L^1(\mu)$ y $b \in L^1(\nu)$. Sea $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ una función semicontinua inferior que satisface $c(x, y) \geq a(x) + b(y)$ para todo $x \in \mathcal{X}$ e $y \in \mathcal{Y}$.

3.2. Métricas en el Espacio de Medidas de Probabilidad

Entonces existe un coupling de μ y ν , π^* tal que:

$$\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi^*(x, y) = \inf \left\{ \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) : \pi \in \Pi_{\mu, \nu} \right\}.$$

Observación 3.2.6: Notemos que (PMK) se puede formular en términos de variables aleatorias como:

$$\inf \{ \mathbb{E}(c(X, Y)) : \text{ley}(X) = \mu \text{ y } \text{ley}(Y) = \nu \}. \quad (\text{PMK}')$$

A los pares de variables aleatorias (X, Y) que participan de este ínfimo se les llamará *coupling* de μ y ν y, abusando un poco de la notación, escribiremos $(X, Y) \in \Pi_{\mu, \nu}$.

Para lo que queremos hacer, el caso que nos interesa es cuando $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. En este caso, los planes de transporte optimal quedan caracterizados por el siguiente teorema que aparece en el Capítulo 6 de [3].

Teorema 3.2.7. Sean μ y ν medidas de probabilidad en \mathbb{R} con p -ésimo momento finito y sea $c(x, y) = h(x - y)$, con $h \geq 0$ convexa y $h(u) \leq A + Bu^p$. Entonces:

1. Se tiene que

$$\min \left\{ \int_{\mathbb{R}^2} c(x, y) d\pi(x, y) : \pi \in \Pi_{\mu, \nu} \right\} = \int_0^1 c(q_\mu^-(s), q_\nu^-(s)) ds.$$

2. Si μ no tiene átomos, entonces $q_\nu^-(F_\mu)$ es un plan de transporte optimal y es el único si h es estrictamente convexa.

El espacio de Wasserstein

En lo que sigue, consideraremos (\mathcal{X}, d) un espacio métrico polaco y notaremos como $\mathcal{P}(\mathcal{X})$ al conjunto de todas las medidas de probabilidad sobre \mathcal{X} . Para ver la demostración de los resultados que aquí se enuncian, se puede consultar el Capítulo 6 de [43].

Definición 3.2.8. Sea $p \in [1, \infty)$, la distancia de Wasserstein entre μ y ν de orden p , $W_p(\mu, \nu)$ se define como:

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi_{\mu, \nu}} \left(\int \int d(x, y)^p d\pi(x, y) \right)^{1/p}.$$

Se puede probar que $W_p(\cdot, \cdot)$ es una función simétrica, positiva, que es igual a cero si y sólo si ambos argumentos son iguales y que satisface la desigualdad triangular. Sin embargo, con el grado de generalidad que está definida, es posible que $W_p(\mu, \nu) = +\infty$. Para evitar esto, restringimos $W_p(\cdot, \cdot)$ al subconjunto de $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ donde toma valores finitos, lo que nos lleva a la definición 3.2.12, que se da más abajo.

Observación 3.2.9: Notemos que

$$W_p(\mu, \nu) = \inf_{(X,Y) \in \Pi_{\mu,\nu}} (\mathbb{E}(d(X, Y)^p))^{1/p},$$

y entonces, si queremos acotar superiormente $W_p(\mu, \nu)$, basta con acotar $\mathbb{E}(d(X, Y)^p)$ para algún $(X, Y) \in \Pi_{\mu,\nu}$.

Observación 3.2.10: En lo que sigue, para simplificar la notación, llamaremos W a W_1 .

Para W se tiene la siguiente caracterización dual, que corresponde a un caso particular de la dualidad de Kantorovich para el problema de transporte optimal. La prueba de esta caracterización se puede encontrar en el Capítulo 11 de [18], mientras que la prueba del teorema general se puede encontrar el Capítulo 5 de [43].

Teorema 3.2.11. Fórmula de Kantorovich-Rubinstein Si $\mu, \nu \in \mathcal{P}(\mathcal{X})$, entonces

$$W(\mu, \nu) = \inf_{(X,Y) \in \Pi_{\mu,\nu}} \mathbb{E}(d(X, Y)) = \sup_{\|\psi\|_L \leq 1} \left\{ \int_{\mathcal{X}} \psi(x) d\mu(x) - \int_{\mathcal{X}} \psi(y) d\nu(y) \right\},$$

siempre que el supremo de la derecha sea finito.

Notemos que a partir de este teorema, tenemos trivialmente la siguiente relación entre la distancia dual de funciones Lipschitz acotadas y W ,

$$\forall \mu, \nu \in \mathcal{P}(\mathcal{X}), \quad BL(\mu, \nu) \leq W(\mu, \nu).$$

Definición 3.2.12. Se define el espacio de Wasserstein de orden p , como

$$\mathcal{P}_p(\mathcal{X}) = \left\{ \mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(x_0, x)^p d\mu(x) < +\infty \right\},$$

donde $x_0 \in \mathcal{X}$.

Observación 3.2.13: Notemos que en esta definición, x_0 es arbitrario, pero gracias a la desigualdad triangular, el espacio $\mathcal{P}_p(\mathcal{X})$ no depende de la elección de x_0 .

El siguiente resultado está contenido en el Teorema 6.18 y la Observación 6.19 que aparecen en [43], donde se puede encontrar su demostración.

Teorema 3.2.14. $(\mathcal{P}_p(\mathcal{X}), W_p)$ es un espacio métrico polaco. Además si \mathcal{X} es compacto, $\mathcal{P}_p(\mathcal{X})$ también lo será.

Observación 3.2.15: Gracias a la desigualdad de Hölder, se tiene que $p \leq q \Rightarrow W_p \leq W_q$. Luego, W es la métrica más débil.

3.2. Métricas en el Espacio de Medidas de Probabilidad

En cuanto a la topología generada por la distancia de Wasserstein, esta resulta ser más fuerte que la topología de la convergencia débil. Sin embargo, se tiene el siguiente resultado que relaciona ambas convergencias.

Teorema 3.2.16. *Sea $\{\mu_k\}_{k \in \mathbb{N}}$ una sucesión de medidas de probabilidad en $\mathcal{P}_p(\mathcal{X})$ y sea $\mu \in \mathcal{P}_p(\mathcal{X})$. Entonces*

$$W_p(\mu_k, \mu) \xrightarrow{k} 0.$$

Si y sólo si

$$\mu_k \rightharpoonup \mu,$$

y para algún x_0 (y luego para todo x_0), se tiene que

$$\int_{\mathcal{X}} d(x, x_0)^p d\mu_k(x) \xrightarrow{k} \int_{\mathcal{X}} d(x, x_0)^p d\mu(x).$$

Observación 3.2.17: Gracias al teorema de Varadarajan 3.1.4, tenemos que si $\hat{\mu}^N$ es la medida empírica de una muestra i.i.d. de tamaño N de variables aleatorias con ley μ , entonces casi seguramente

$$\hat{\mu}^N \rightharpoonup \mu.$$

Por otro lado, gracias a la Ley Fuerte de los Grandes Números, también se tiene que casi seguramente que

$$\int_{\mathcal{X}} d(x, x_0)^p d\hat{\mu}^N(x) \xrightarrow{N \rightarrow \infty} \int_{\mathcal{X}} d(x, x_0)^p d\mu(x),$$

siempre que el lado derecho sea finito. En resumen, si $\mu \in \mathcal{P}_p(\mathcal{X})$, gracias al teorema 3.2.16, se tiene que casi seguramente

$$W_p(\hat{\mu}^N, \mu) \xrightarrow{N \rightarrow \infty} 0.$$

Observación 3.2.18: Nos interesa explicitar el caso $\mathcal{X} = \mathbb{R}$ y $p \in [1, +\infty)$. Notemos que gracias al teorema 3.2.7 tenemos que para $\mu, \nu \in \mathcal{P}(\mathbb{R})$,

$$W_p^p(\mu, \nu) = \int_0^1 |q_\nu^-(s) - q_\mu^-(s)|^p ds.$$

3.2.4. Relaciones entre las distintas Métricas

Relación entre BL y d_P :

En el Capítulo 2 de [26] aparece el siguiente resultado que relaciona la distancia de las funciones Lipschitz acotadas y la distancia de Lévy-Prokhorov.

Teorema 3.2.19. *Sea \mathcal{X} un espacio polaco medible. Si $\mu, \nu \in \mathcal{P}(\mathcal{X})$ entonces*

$$d_P(\mu, \nu)^2 \leq BL(\mu, \nu) \leq 2d_P(\mu, \nu).$$

3.2. Métricas en el Espacio de Medidas de Probabilidad

Relación entre BL y W :

En la sección anterior, cuando se enunció la fórmula de Kantorovich-Rubinstein 3.2.11, se mencionó que para $\mu, \nu \in \mathcal{P}(\mathcal{X})$ se tiene que

$$BL(\mu, \nu) \leq W(\mu, \nu).$$

En el caso de que \mathcal{X} tenga diámetro finito D , se tiene que $W(\mu, \nu) \leq (1 + D)BL(\mu, \nu)$. En efecto, sea ϕ una función Lipschitz. Se define

$$f_\phi(x) = \phi(x) - \phi(x_0),$$

entonces se tiene que $\|f_\phi\|_\infty \leq D\|\phi\|_L$ y $\|f_\phi\|_L = \|\phi\|_L$, con lo que se concluye

$$\|f_\phi\|_{L_b} \leq (1 + D)\|\phi\|_L.$$

Luego,

$$\begin{aligned} \frac{1}{1 + D} \sup_{\|\phi\|_L \leq 1} \left\{ \int_{\mathcal{X}} \phi d\mu - \int_{\mathcal{X}} \phi d\nu \right\} &= \sup_{\|\phi\|_L \leq 1} \left\{ \int_{\mathcal{X}} \frac{\phi}{1 + D} d\mu - \int_{\mathcal{X}} \frac{\phi}{1 + D} d\nu \right\} \\ &= \sup_{\|\phi\|_L \leq \frac{1}{1 + D}} \left\{ \int_{\mathcal{X}} \phi d\mu - \int_{\mathcal{X}} \phi d\nu \right\} \\ &= \sup_{\|\phi\|_L \leq \frac{1}{1 + D}} \left\{ \int_{\mathcal{X}} f_\phi d\mu - \int_{\mathcal{X}} f_\phi d\nu \right\} \\ &\leq \sup_{\|f\|_{L_b} \leq 1} \left\{ \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f d\nu \right\}. \end{aligned}$$

En resumen, si \mathcal{X} tiene diámetro finito D , entonces para $\mu, \nu \in \mathcal{P}(\mathcal{X})$ se tiene que

$$BL(\mu, \nu) \leq W(\mu, \nu) \leq (1 + D)BL(\mu, \nu).$$

Relación entre las distintas métricas en el caso real:

A partir de todas las relaciones anteriores, en el caso $\mathcal{X} = \mathbb{R}$ tenemos que si μ y ν son medidas de probabilidad, entonces

$$d_L(\mu, \nu)^2 \leq d_P(\mu, \nu)^2 \leq BL(\mu, \nu) \leq W(\mu, \nu). \quad (3.2.1)$$

Si suponemos que μ, ν tienen soporte compacto de diámetro D , entonces también se tiene:

$$W(\mu, \nu) \leq (1 + D)BL(\mu, \nu) \leq 2(1 + D)d_P(\mu, \nu). \quad (3.2.2)$$

Estas desigualdades nos permiten deducir las siguientes cotas

$$\mathbb{P}(d_L(\mu, \hat{\mu}^N) > \varepsilon) \leq \mathbb{P}(d_P(\mu, \hat{\mu}^N)^2 > \varepsilon^2) \leq \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon^2) \quad \text{y} \quad (3.2.3)$$

$$\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq \mathbb{P}(BL(\mu, \hat{\mu}^N) > \varepsilon/(D + 1)) \leq \mathbb{P}(d_P(\mu, \hat{\mu}^N) > \varepsilon/2(D + 1)), \quad (3.2.4)$$

las que serán útiles para traducir resultados de concentración obtenidos para una métrica en resultados de concentración para otra métrica.

Capítulo 4

Desigualdades de Talagrand $T_p(\lambda)$

Las desigualdades de Talagrand son el último objeto que necesitamos introducir antes de pasar a los resultados de concentración del capítulo 5. Son desigualdades funcionales que relacionan la distancia de Wasserstein con la entropía relativa entre medidas de probabilidad. En este capítulo se introducen estas desigualdades y se dan algunos ejemplos de medidas que las satisfacen. Después se presenta un primer resultado, dado por Marton, que relaciona las desigualdades de Talagrand con el fenómeno de concentración de medidas. Y por último, se enuncian dos criterios integrales que nos sirven para caracterizar estas desigualdades en un caso particular pero importante.

4.1. Definiciones

Sea \mathcal{X} un espacio polaco medible. Se define el funcional de entropía relativa de ν con respecto a μ como

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log \left(\frac{d\nu}{d\mu} \right) d\mu & \nu \ll \mu, \\ +\infty & \sim. \end{cases}$$

Definición 4.1.1. Sea $\lambda > 0$, $p \geq 1$ y $\mu \in \mathcal{P}(\mathcal{X})$. Se dice que μ satisface la desigualdad de Talagrand o de transporte $T_p(\lambda)$, si para toda $\nu \in \mathcal{P}(\mathcal{X})$

$$W_p(\nu, \mu) \leq \sqrt{\frac{2}{\lambda} H(\nu|\mu)}. \quad (4.1.1)$$

Se dice que μ satisface una desigualdad T_p , si satisface $T_p(\lambda)$ para algún $\lambda > 0$.

Recordemos además que si $q \leq p$, entonces $W_q(\nu, \mu) \leq W_p(\nu, \mu)$. Luego, si μ satisface $T_p(\lambda)$, también satisface $T_{p'}(\lambda)$ para todo $p' \in [1, p]$. Notemos también que si μ satisface $T_p(\lambda)$, entonces también satisface $T_p(\lambda')$ para $\lambda' \leq \lambda$.

4.1. Definiciones

Ejemplo 4.1.2: En [41] Talagrand muestra que la ley normal multivariada estándar $\mathcal{N}_n(0, I_n)$ satisface $T_2(1)$ en \mathbb{R}^n dotado de la métrica euclidiana, lo que en particular nos dice que $\mathcal{N}_n(0, I_n)$ satisface también $T_1(1)$.

Ejemplo 4.1.3: Sean $\mu, \nu \in \mathcal{P}(\mathcal{X})$ y sea $\|\mu - \nu\|_{TV}$ su distancia en variación total. Si $\nu \ll \mu$, entonces

$$\|\mu - \nu\|_{TV} = \int_{\mathcal{X}} |f - 1| d\mu,$$

donde f es la derivada de Radon-Nicodým de ν con respecto a μ .

Por otro lado, para todo $x > 0$,

$$3(x - 1)^2 \leq (4 + 2x)(x \log(x) - x + 1).$$

Gracias a esta desigualdad, tenemos que

$$\begin{aligned} \|\mu - \nu\|_{TV} &= \frac{1}{\sqrt{3}} \int_{\mathcal{X}} \sqrt{3} |f - 1| d\mu \\ &\leq \frac{1}{\sqrt{3}} \int_{\mathcal{X}} \sqrt{4 + 2f} \sqrt{f \log(f) - f + 1} d\mu \\ &\leq \frac{1}{\sqrt{3}} \left(\int_{\mathcal{X}} 4 + 2f d\mu \right)^{1/2} \left(\int_{\mathcal{X}} f \log(f) - f + 1 d\mu \right)^{1/2} \\ &= \sqrt{2} \sqrt{H(\nu|\mu) - 1 + 1}. \end{aligned}$$

En resumen, hemos probado la desigualdad de Csiszár-Kullback-Pinsker

$$(\forall \mu, \nu \in \mathcal{P}(\mathcal{X})) \quad \|\mu - \nu\|_{TV} \leq \sqrt{2H(\nu|\mu)}. \quad (\text{CKP})$$

Por otro lado, se puede probar que si en \mathcal{X} se considera la distancia trivial $d_0(x, y) = \mathbb{1}_{x \neq y}$, entonces

$$W(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_{TV}.$$

Luego, la desigualdad (CKP) la podemos interpretar como que toda medida boreliana μ satisface $T_1(4)$ en (\mathcal{X}, d_0) , pues si ν es otra medida de probabilidad,

$$W(\mu, \nu) \leq \sqrt{\frac{1}{2} H(\nu|\mu)}.$$

Si bien estos ejemplos nos muestran que existen medidas que satisfacen una desigualdad de transporte, en general es difícil probar que una medida satisface alguna. Al final de este capítulo estudiaremos criterios integrales que caracterizan a las medidas que satisfacen $T_1(\lambda)$.

El siguiente resultado fue establecido por K. Marton en [32], y es el primer resultado que nos muestra la relación entre las desigualdades de transporte y el fenómeno de concentración de medidas de probabilidad:

4.2. Criterio Integral para T_1 .

Proposición 4.1.4. *Si μ es una medida de probabilidad que satisface $T_1(\lambda)$ entonces para todo $A \subset \mathcal{X}$ medible, tal que $\mu(A) \geq 1/2$ se satisface que*

$$\forall r \geq r_0, \quad \mu(A_r) \geq 1 - e^{-\frac{\lambda(r-r_0)^2}{2}},$$

donde $r_0 = \sqrt{2 \log(2)/\lambda}$ y $A_r = \{x \in \mathcal{X} : d(x, A) \leq r\}$.

Demostración: Consideremos $A \subset \mathcal{X}$ y definamos μ_A a través de su derivada de Radon-Nikodym $d\mu_A/d\mu := \mathbb{1}_A/\mu(A)$. Para $B = \mathcal{X} \setminus A_r$ definimos μ_B de manera análoga. Notemos que para todo $x \in A$ e $y \in B$ se tiene que $d(x, y) \geq r$. Luego, para todo *coupling* (X, Y) de μ_A con μ_B tenemos que $E(d(X, Y)) \geq r$, y entonces $r \leq W(\mu_A, \mu_B)$. Por lo tanto, gracias a la desigualdad triangular y a que μ satisface $T_1(\lambda)$, tenemos que

$$\begin{aligned} r &\leq W(\mu_A, \mu_B) \\ &\leq W(\mu_A, \mu) + W(\mu, \mu_B) \\ &\leq \sqrt{\frac{2}{\lambda} H(\mu_A|\mu)} + \sqrt{\frac{2}{\lambda} H(\mu_B|\mu)} \\ &= \sqrt{\frac{2}{\lambda} \log(1/\mu(A))} + \sqrt{\frac{2}{\lambda} \log(1/\mu(B))}. \end{aligned}$$

Entonces,

$$\begin{aligned} (r - r_0)^2 &\leq \frac{2}{\lambda} \log(1/\mu(B)) + \frac{2}{\lambda} \left(\log(2) + 2\sqrt{\log(\mu(A)) \log(1 - \mu(A))} \right. \\ &\quad \left. - 2\sqrt{\log(\mu(A)) \log(1/2)} - 2\sqrt{\log(1 - \mu(A)) \log(1/2)} - \log(\mu(A)) \right). \end{aligned}$$

Luego, si se tiene que $\mu(A) \geq 1/2$ el término entre paréntesis del lado derecho es negativo y entonces

$$(r - r_0)^2 \leq \frac{2}{\lambda} \log(1/\mu(B)),$$

de donde se concluye el resultado.

q.e.d.

Es este vínculo entre el fenómeno de concentración de medidas de probabilidad y las desigualdades de transporte, el que ha impulsado el estudio de estas últimas. Para un estudio detallado de este tema se pueden consultar [24] y [33].

4.2. Criterio Integral para T_1 .

El objetivo de esta sección es presentar caracterizaciones integrales para $T_1(\lambda)$. La importancia práctica de estas caracterizaciones es enorme, ya que probar que una medida μ

4.2. Criterio Integral para T_1 .

satisface (4.1.1), es mucho más difícil que probar que

$$\int_{\mathcal{X}} f_{\alpha} d\mu < +\infty, \quad (4.2.1)$$

para f_{α} en una familia de funciones apropiada y α un parámetro.

Los resultados que mostraremos relacionan, para una familia $\{f_{\alpha}\}_{\alpha}$ particular, los valores de α para los cuales se tiene (4.2.1) con los valores de λ para los cuales μ satisface $T_1(\lambda)$. Antes de pasar a dichos resultados, veamos que si μ satisface $T_1(\lambda)$ entonces tiene momentos exponenciales cuadráticos finitos.

Lema 4.2.1. *Sea (\mathcal{X}, d) un espacio métrico medible y μ una medida de probabilidad que satisface $T_1(\lambda)$ en (X, d) y $\alpha \in (0, \lambda/2)$, entonces*

$$E_{\alpha} = \int_{\mathcal{X}} e^{\alpha d^2(x, x_0)} d\mu(x) < +\infty, \text{ para algún } x_0 \in \mathcal{X}.$$

Para probar este, haremos uso de un resultado encontrado por Bobkov y Götze en [7] y que enunciamos a continuación.

Teorema 4.2.2. [Bobkov y Götze. (1999)] *μ satisface $T_1(\lambda)$ en (\mathcal{X}, d) si y sólo si para toda función Lipschitz $f : \mathcal{X} \rightarrow \mathbb{R}$, f es μ -integrable y*

$$\int_{\mathcal{X}} e^{\alpha(f - \langle f, \mu \rangle)} d\mu \leq \exp\left(\frac{\alpha^2}{2\lambda} \|f\|_L^2\right) \quad \forall \alpha \in \mathbb{R}, \quad (4.2.2)$$

donde $\langle f, \mu \rangle = \int_{\mathcal{X}} f d\mu$ y $\|\cdot\|_L$ es la seminorma para funciones Lipschitz definida en el capítulo 3.

Demostración: del lema 4.2.1. Sea γ a la ley de una variable aleatoria normal estándar en \mathbb{R} , entonces

$$\int_{\mathbb{R}} e^{ts} d\gamma(s) = e^{t^2/2}.$$

En particular, tenemos que para $x_0, x \in \mathcal{X}$ se tiene que

$$e^{\alpha d(x, x_0)^2} = \int_{\mathbb{R}} e^{\sqrt{2\alpha} d(x, x_0) s} d\gamma(s).$$

Y entonces, integrando a ambos lados de la igualdad con respecto a μ

$$\int_{\mathcal{X}} e^{d(x, x_0)^2} d\mu(x) = \int_{\mathcal{X}} \int_{\mathbb{R}} e^{\sqrt{2\alpha} d(x, x_0) s} d\gamma(s) d\mu(x) = \int_{\mathbb{R}} \int_{\mathcal{X}} e^{\sqrt{2\alpha} s d(x, x_0)} d\mu(x) d\gamma(s),$$

donde la segunda igualdad se tiene gracias al teorema de Tonelli. Entonces,

$$\int_{\mathcal{X}} e^{\alpha d(x, x_0)^2} d\mu(x) = \int_{\mathbb{R}} \int_{\mathcal{X}} e^{\sqrt{2\alpha} s (d(x, x_0) - \langle d(\cdot, x_0), \mu \rangle + \langle d(\cdot, x_0), \mu \rangle)} d\mu(x) d\gamma(s),$$

4.2. Criterio Integral para T_1 .

con lo que tenemos

$$\int_{\mathcal{X}} e^{\alpha d(x,x_0)^2} d\mu(x) = e^{\alpha \langle d(\cdot, x_0), \mu \rangle^2} \int_{\mathbb{R}} \int_{\mathcal{X}} e^{\sqrt{2\alpha s} (d(x,x_0) - \langle d(\cdot, x_0), \mu \rangle)} d\mu(x) d\gamma(s).$$

Gracias al teorema anterior, podemos acotar la integral con respecto a μ y obtener

$$\int_{\mathcal{X}} e^{\alpha d(x,x_0)^2} d\mu(x) \leq e^{\alpha \langle d(\cdot, x_0), \mu \rangle^2} \int_{\mathbb{R}} e^{\frac{\alpha s^2}{\lambda}} d\gamma(s).$$

Pero la integral de la derecha es menor que infinito si y sólo si $\alpha/\lambda < 1/2$. Es decir si $\alpha < \lambda/2$. En resumen, si μ satisface $T_1(\lambda)$ entonces tiene momento exponencial cuadrático E_α finito para $\alpha \in (0, \lambda/2)$.

q.e.d.

Los criterios que se dan a continuación, nos muestran que la implicancia recíproca también es cierta.

4.2.1. Criterio de Djelleout, Guillin y Wu.

El resultado principal de esta sección corresponde a explicitar la cota para λ dada en el Teorema 2.3 de [16]. Si bien esta cota resulta ser bastante gruesa, tiene la ventaja que de que el criterio que encontramos es sumamente fácil de usar. Comenzamos por enunciar el resultado de Djelleout, Guillin y Wu.

Teorema 4.2.3. [Djelleout, Guillin y Wu. (2004)] *Una medida de probabilidad μ en (\mathcal{X}, d) satisface T_1 para algún $\lambda > 0$ si y sólo si para algún $\alpha > 0$ se tiene que*

$$\int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) d\mu(y) < +\infty. \quad (4.2.3)$$

Además se satisface que

$$\frac{1}{\lambda} \leq \frac{2}{\alpha} \sup_{k \geq 1} \left(\frac{(k!)^2}{(2k)!} \right)^{\frac{1}{k}} \left[\int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) d\mu(y) \right]^{\frac{1}{k}}. \quad (4.2.4)$$

Observación 4.2.4: La desigualdad (4.2.4) nos dice que si se tiene (4.2.3) para α , entonces μ satisface $T_1(\lambda)$ para algún λ en $[f(\alpha), \infty)$ donde

$$f(\alpha) = \left(\frac{2}{\alpha} \sup_{k \geq 1} \left(\frac{(k!)^2}{(2k)!} \right)^{\frac{1}{k}} \left[\int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) d\mu(y) \right]^{\frac{1}{k}} \right)^{-1}. \quad (4.2.5)$$

Pero si μ satisface la desigualdad $T_p(\lambda)$ también satisface $T_p(\lambda')$ $\forall \lambda' \in (0, \lambda)$, y entonces tenemos que si μ cumple (4.2.3) para α , se tiene que μ satisface $T_1(f(\alpha))$.

4.2. Criterio Integral para T_1 .

Observación 4.2.5: Para tener la condición (4.2.3) es suficiente que

$$\int_{\mathcal{X}} e^{2\alpha d^2(x,x_0)} d\mu(x) < +\infty \quad \text{Para algún } x_0 \in \mathcal{X}. \quad (4.2.6)$$

En efecto,

$$\begin{aligned} \int_E \int_E e^{\alpha d^2(x,y)} d\mu(x) d\mu(y) &\leq \int_E \int_E e^{2\alpha(d^2(x,x_0)+d^2(y,x_0))} d\mu(x) d\mu(y) \\ &= \left(\int_E e^{2\alpha d^2(x,x_0)} d\mu(x) \right)^2. \end{aligned}$$

Así podemos concluir que si

$$\int_E e^{\beta d^2(x,x_0)} d\mu(x) < \infty,$$

entonces μ satisface $T_1(f(\beta/2))$.

Observación 4.2.6: Sea

$$a_k = \left(\frac{(k!)^2}{(2k)!} C(\alpha) \right)^{\frac{1}{k}},$$

con

$$C(\alpha) = \int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) d\mu(y).$$

Notemos que si queremos encontrar una expresión explícita para $f(\alpha)$ debemos calcular $\sup_{k \geq 1} a_k$. En lo que sigue calcularemos este supremo, pero primero veamos una recurrencia que satisface a_k . Se tiene

$$\begin{aligned} \frac{a_k^k}{a_{k-1}^{k-1}} &= \frac{\frac{C(\alpha)k!^2}{(2k)!}}{\frac{C(\alpha)(k-1)!^2}{(2k-2)!}} \\ &= \frac{k}{4k-2}, \end{aligned}$$

y entonces

$$a_k^k = \frac{k}{4k-2} a_{k-1}^{k-1},$$

de donde

$$\begin{aligned} a_k &= \left(\frac{k}{(4k-2)a_{k-1}} \right)^{\frac{1}{k}} a_{k-1} \\ &= \underbrace{\left(\frac{k}{4k-2} \left(\frac{(2k-2)!}{C(\alpha)(k-1)!^2} \right)^{\frac{1}{k-1}} \right)^{\frac{1}{k}}}_{m_k} a_{k-1}. \end{aligned}$$

4.2. Criterio Integral para T_1 .

Luego, se tiene que $a_k < a_{k-1}$ si y sólo si $m_k < 1$. Estudiemos la sucesión $\{m_k\}_{k \geq 2}$.

$$\begin{aligned}
 & m_k < 1 \\
 \Leftrightarrow & \left(\frac{k}{4k-2} \left(\frac{(2k-2)!}{C(\alpha)(k-1)!^2} \right)^{\frac{1}{k-1}} \right)^{\frac{1}{k}} < 1 \\
 \Leftrightarrow & \left(\frac{(2k-2)!}{C(\alpha)(k-1)!^2} \right)^{\frac{1}{k-1}} < \frac{4k-2}{k} \\
 \Leftrightarrow & \underbrace{\left(\frac{k}{4k-2} \right)^{k-1} \frac{(2k-2)!}{(k-1)!^2}}_{d_k} < C(\alpha).
 \end{aligned}$$

Así tenemos que $a_k < a_{k-1}$ si sólo si $d_k < C(\alpha)$.

Por otro lado, se tiene que $\{d_k\}_{k \geq 2}$ es decreciente, pues

$$\frac{d_{k+1}}{d_k} = \left(\frac{k+1}{4k+2} \frac{4k-2}{k} \right)^k \text{ y } \frac{k+1}{4k+2} < \frac{k}{4k-2}.$$

Resumiendo: $\{d_k\}_{k \geq 2}$ es decreciente, $d_2 = 2/3$ y $C(\alpha) \geq 1$. Con esto concluimos que $d_k < C(\alpha)$ para todo $k \geq 2$, lo que sumado a todo el análisis previo, nos dice que $\{a_k\}_{k \geq 1}$ es decreciente y por lo tanto

$$\sup_{k \geq 1} a_k = a_1 = \frac{C(\alpha)}{2}. \quad (4.2.7)$$

Con esto finalmente,

$$f(\alpha) = \frac{\alpha}{C(\alpha)}.$$

A partir de la observación anterior, nos damos cuenta de que podemos reescribir el teorema 4.2.3 de la siguiente manera.

Teorema 4.2.7. *Si μ es una medida de probabilidad en (\mathcal{X}, d) tal que*

$$C(\alpha) = \int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) d\mu(y) < +\infty, \quad (4.2.8)$$

entonces μ satisface $T_1(\alpha/C(\alpha))$.

Antes de ver un ejemplo numérico del uso de este teorema, veamos un corolario que puede ser útil.

Corolario 4.2.8. *Supongamos que \mathcal{X} tiene diámetro finito D . Entonces toda medida de probabilidad μ en \mathcal{X} satisface $T_1(1/eD^2)$.*

4.2. Criterio Integral para T_1 .

Demostración: Sabemos que μ satisface $T_1(\alpha/C(\alpha))$ con

$$C(\alpha) = \int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) d\mu(y),$$

pero en este caso

$$C(\alpha) = \int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) d\mu(y) \leq \int_{\mathcal{X}} \int_{\mathcal{X}} e^{\alpha D^2} d\mu(x) d\mu(y) = e^{\alpha D^2},$$

de donde concluimos que $g(\alpha) := \alpha e^{-\alpha D^2} \leq \alpha/C(\alpha)$. Luego, tenemos que para todo $\alpha > 0$, μ satisface $T_1(g(\alpha))$. En particular podemos tomar α^* que maximiza $g(\alpha)$. Para encontrar α^* calculamos

$$g'(\alpha) = e^{-\alpha D^2} - \alpha D^2 e^{-\alpha D^2},$$

igualando a 0 se obtiene $\alpha^* = 1/D^2$ y $g(\alpha^*) = 1/eD^2$, de donde se concluye el resultado del lema.

Ejemplo 4.2.9: Consideremos $X = \mathbb{R}$ y $d = |\cdot|$ y

$$d\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

es decir, μ es la ley inducida por una variable aleatoria normal de media 0 y varianza 1. En este caso, tenemos que si $0 < \alpha < 1/4$

$$C(\alpha) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{\alpha(x-y)^2} d\mu(x) d\mu(y) = \frac{1}{\sqrt{1-4\alpha}} < +\infty,$$

de donde tenemos

$$f(\alpha) = \alpha \sqrt{1-4\alpha}.$$

En el gráfico 4.2.1 encontramos $f(\alpha)$ para α en $(0, 1/4)$, notemos que para todos los puntos (α, λ) que pertenecen al gráfico μ satisface $T_1(\lambda)$. En particular podemos tomar el valor de α que maximiza $f(\alpha)$, que en este caso podemos encontrar analíticamente calculando $f'(\alpha)$.

$$f'(\alpha) = \sqrt{1-4\alpha} - \frac{2\alpha}{\sqrt{1-4\alpha}}.$$

Y entonces $f'(\alpha^*) = 0$ si $\alpha^* = 1/6$ y $f(1/6) = 0,09623$. Luego la ley normal estándar satisface $T_1(0,09623)$.

Sin embargo, es conocido que la ley normal estándar satisface $T_1(1)$. Lo que nos indica que la cota que nos da este teorema podría no ser muy eficiente.

4.2. Criterio Integral para T_1 .

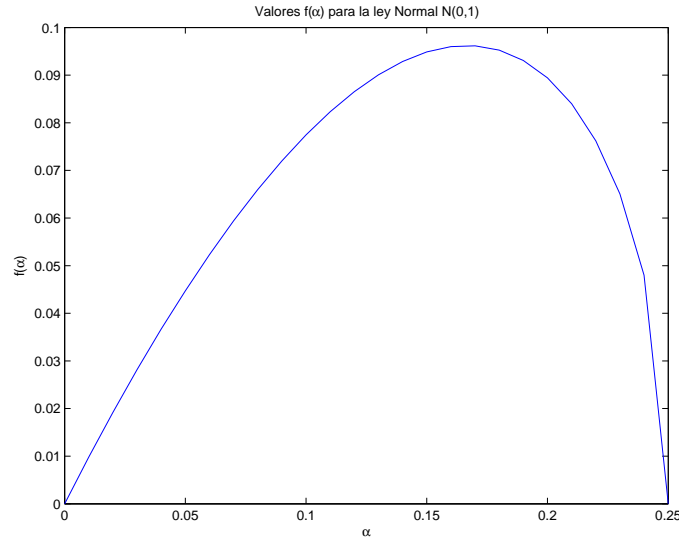


Figura 4.2.1: Valores de $f(\alpha)$ para ley $\mathcal{N}(0, 1)$

4.2.2. Criterio de Bolley, Gozlan y Villani

Como se mostró con el ejemplo anterior, el criterio que aparece en el teorema 4.2.7 puede ser muy poco eficiente. Existe otro criterio integral para T_1 , explicitado por N. Gozlan en [22], a partir de un resultado de F. Bolley y C. Villani que aparece en [9]. Este criterio resulta ser más eficiente, aunque tampoco es óptimo.

Lema 4.2.10. [Bolley, Gozlan y Villani. (2006)] Si μ es una medida de probabilidad en (\mathcal{X}, d) tal que para algún $y \in \mathcal{X}$ y algún $\alpha \in \mathbb{R}$

$$\int_{\mathcal{X}} e^{\alpha d^2(x,y)} d\mu(x) < +\infty, \quad (4.2.9)$$

entonces μ satisface $T_1(\lambda)$ para

$$\lambda = 2 \left(\inf_{y \in \mathcal{X}, \alpha > 0} \left\{ \frac{1}{2\alpha} \left(1 + \log \int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) \right) \right\} \right)^{-1}.$$

Al igual que antes, como corolario de este criterio, podemos obtener un resultado para el caso en que \mathcal{X} tiene diámetro finito.

Corolario 4.2.11. Supongamos que \mathcal{X} tiene diámetro finito D . Entonces toda medida de probabilidad μ en \mathcal{X} satisface $T_1(4/D^2)$.

4.2. Criterio Integral para T_1 .

Demostración: Comencemos por notar que como \mathcal{X} tiene diámetro finito D , entonces

$$e^{\alpha d(x,y)^2} \leq e^{\alpha D^2},$$

de donde tenemos que

$$\inf_{y \in \mathcal{X}, \alpha > 0} \left\{ \frac{1}{2\alpha} \left(1 + \log \int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) \right) \right\} \leq \inf_{y \in \mathcal{X}, \alpha > 0} \left\{ \frac{1}{2\alpha} \left(1 + \log \int_{\mathcal{X}} e^{\alpha D^2} d\mu(x) \right) \right\}.$$

Pero notemos que

$$\inf_{y \in \mathcal{X}, \alpha > 0} \left\{ \frac{1}{2\alpha} \left(1 + \log \int_{\mathcal{X}} e^{\alpha D^2} d\mu(x) \right) \right\} = \inf_{\alpha > 0} \left\{ \frac{1}{2\alpha} + \frac{D^2}{2} \right\} = \frac{D^2}{2},$$

y entonces

$$2 \left(\inf_{y \in \mathcal{X}, \alpha > 0} \left\{ \frac{1}{2\alpha} \left(1 + \log \int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) \right) \right\} \right)^{-1} \geq \frac{4}{D^2}.$$

Pero recordemos que si μ satisface $T_1(\lambda)$, entonces también satisface $T_1(\lambda')$ para todo $\lambda' \leq \lambda$. Luego, aplicando el lema 4.2.10 se obtiene el resultado.

q.e.d.

Notemos que la cota que nos da este corolario es del orden de 10 veces mejor que la dada por el corolario 4.2.8. Veamos ahora un ejemplo numérico del uso del lema 4.2.10.

Ejemplo 4.2.12: Consideremos una vez más la ley normal estándar en \mathbb{R} . En este caso tenemos que

$$\int_{\mathbb{R}} e^{\alpha(x-y)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{e^{\frac{\alpha y^2}{1-2\alpha}}}{\sqrt{1-2\alpha}}.$$

Entonces nuestro problema de optimización a resolver esta dado por

$$\inf_{y \in \mathbb{R}, \alpha \in (0, 1/2)} \left\{ \frac{1}{2\alpha} + \frac{y^2}{2-4\alpha} - \frac{1}{4\alpha} \log(1-2\alpha) \right\}. \quad (4.2.10)$$

En este caso podemos encontrar el mínimo de la función calculando el gradiente e igualando a cero, con lo que se encuentran las siguientes condiciones para (y^*, α^*) , el punto donde se alcanza el óptimo:

$$\begin{aligned} y^* &= 0, \\ \log(1-2\alpha^*) &= \frac{2-6\alpha^*}{1-2\alpha^*}. \end{aligned}$$

La segunda ecuación la podemos resolver numéricamente utilizando Matlab y encontramos

$$\inf_{y \in \mathbb{R}, \alpha \in (0, 1/2)} \left\{ \frac{1}{2\alpha} + \frac{y^2}{2-4\alpha} - \frac{1}{4\alpha} \log(1-2\alpha) \right\} \approx 0,38901816,$$

de donde deducimos que la ley normal estándar en \mathbb{R} satisface $T_1(\lambda)$ para $\lambda \approx 0,88785474$, valor que es mucho más cercano al valor óptimo conocido que es 1.

4.2. Criterio Integral para T_1 .

En defensa del teorema 4.2.7 podemos decir que si bien la cota que da puede ser bastante gruesa, puede ser útil cuando el problema de optimización que hay que resolver para encontrar la cota dada por 4.2.10 es difícil. Para terminar veamos otro ejemplo para el lema 4.2.10.

Ejemplo 4.2.13: Consideremos en \mathbb{R} la ley Weibull simetrizada de parámetros $(a, 2)$, cuya densidad esta dada por

$$f(x) = \frac{|x|}{a^2} e^{-\frac{x^2}{a^2}}.$$

La función densidad de esta distribución se puede observar en el siguiente gráfico para distintos valores del parámetro a .

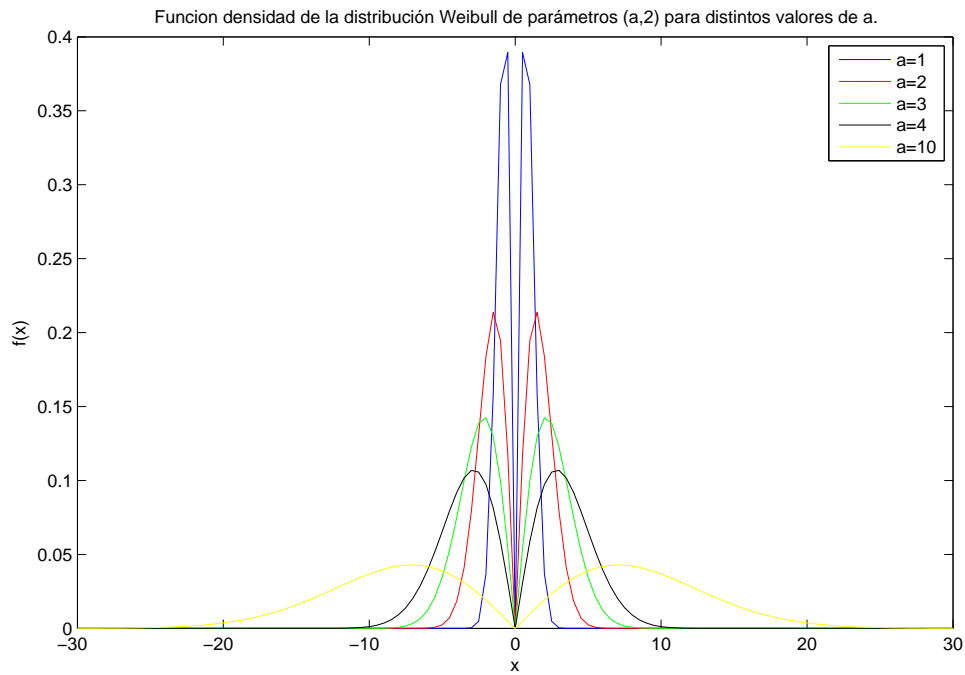


Figura 4.2.2: Densidad de la ley Weibull simetrizada de parámetros $(a, 2)$ para distintos valores de a .

Con esto se tiene que

$$\int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) = \int_{\mathbb{R}} e^{\alpha(x-y)^2} \frac{|x|}{a^2} e^{-\frac{x^2}{a^2}} dx.$$

Notemos que el lado derecho es finito si $\alpha < 1/a^2$. Para estos valores de α se puede calcular

$$\int_{\mathcal{X}} e^{\alpha d(x,y)^2} d\mu(x) = \frac{e^{\alpha y^2}}{1 - \alpha a^2} + y \frac{\alpha e^{C y^2} \sqrt{2\pi\sigma^2}}{1 - \alpha a^2} [2\Phi(y\mu/\sigma) - 1],$$

4.2. Criterio Integral para T_1 .

donde Φ corresponde a la función distribución acumulada de la ley normal estándar,

$$C = \frac{\alpha}{1 - \alpha a^2}, \quad \sigma = \frac{a}{\sqrt{2(1 - \alpha a^2)}} \quad \text{y} \quad \mu = \frac{\alpha a^2}{1 - \alpha a^2}.$$

Luego el problema de optimización a resolver es

$$\inf_{y \in \mathbb{R}, \alpha \in (0, 1/a^2)} \left\{ \frac{1}{2\alpha} \left(1 + \log \left[\frac{e^{\alpha y^2}}{1 - \alpha a^2} + y \frac{\alpha e^{C y^2} \sqrt{2\pi\sigma^2}}{1 - \alpha a^2} [2\Phi(y\mu/\sigma) - 1] \right] \right) \right\}.$$

Para resolver numéricamente este problema es necesario fijar un valor para a , consideraremos $a = 4$, lo que corresponde a la función en color negro en el gráfico 4.2.2. Utilizando Matlab encontramos

$$\inf_{y \in \mathbb{R}, \alpha \in (0, 1/16)} \left\{ \frac{1}{2\alpha} \left(1 + \log \left[\frac{e^{\alpha y^2}}{1 - \alpha a^2} + y \frac{\alpha e^{C y^2} \sqrt{2\pi\sigma^2}}{1 - \alpha a^2} [2\Phi(y\mu/\sigma) - 1] \right] \right) \right\} \approx 25,16954576.$$

Con lo que concluimos que la distribución de Weibull(4, 2) satisface $T_1(0,07946111)$.

Observación 4.2.14: La desigualdad T_2 ha sido relacionada con otras desigualdades funcionales, ver por ejemplo [35]. Sin embargo, por ahora no hay resultados que ligen T_2 con condiciones de integrabilidad.

Capítulo 5

Concentración de la Medida Empírica

En este capítulo se estudiará, para distintas métricas en el espacio de probabilidades, la concentración de la medida empírica $\hat{\mu}^N$ en torno a la ley de la muestra μ . Más precisamente, se estudiarán condiciones para tener desigualdades del tipo

$$\mathbb{P}(d(\mu, \hat{\mu}^N) > \varepsilon) \leq f(N, \varepsilon),$$

con $\varepsilon > 0$, $N \geq N_0$ y f una función que tiende a cero con N .

La motivación para buscar desigualdades como esta es la siguiente: supongamos que θ es un parámetro que desconocemos de μ , o el valor de una función de μ , y que el error de estimación de θ a partir de una muestra de tamaño N , es una función Lipschitz con respecto a la distancia entre μ y $\hat{\mu}^N$. Es decir,

$$|\theta - \hat{\theta}^N| \leq Ld(\mu, \hat{\mu}^N), \text{ para cierta constante } L.$$

Entonces, tenemos que para todo $N \geq N_0$

$$\mathbb{P}(|\theta - \hat{\theta}^N| > \varepsilon) \leq \mathbb{P}(d(\mu, \hat{\mu}^N) > \varepsilon/L) \leq f(N, \varepsilon/L),$$

lo que nos permite construir un intervalo de confianza para θ con N finito, sin conocer ni hacer hipótesis acerca de la distribución de θ . Nuestro objetivo final es aplicar esta metodología a las medidas de riesgo espectrales, ya que gracias a la observación 3.2.18 y a la desigualdad de Hölder, más adelante probaremos que

$$|\hat{\rho}_\phi^N(X) - \rho_\phi(X)| \leq \|\phi\|_{L^q} W_p(\mu, \hat{\mu}^N).$$

Sin embargo, para llevar esta idea a la práctica es fundamental contar con resultados explícitos, donde cada constante que aparezca esté claramente identificada. Este capítulo se ha desarrollado con este principio, y se han rehecho todas las demostraciones necesarias para lograr que cada teorema que se presenta lo cumpla.

5.1. Concentración de la Medida Empírica con respecto a W

Para lo que sigue consideramos μ una medida de probabilidad en \mathbb{R} y $\hat{\mu}^N$ la medida empírica generada a partir de $\{X_1, \dots, X_N\}$ una muestra i.i.d. de μ .

5.1.1. Teorema de Bolley, Guillin y Villani

En [8] se prueba el siguiente resultado de concentración de la media empírica al rededor de la ley de la muestra, para la distancia de Wasserstein de orden p .

Teorema 5.1.1. [Bolley, Guillin y Villani. (2007)] *Sea $p \in [1, 2]$ y sea μ una medida de probabilidad en \mathbb{R}^d que satisface $T_p(\lambda)$. Entonces para todo $\lambda' < \lambda$ y $d' > d$, existe una constante N_0 que depende sólo de λ' , d' y de un momento exponencial cuadrático de μ , tal que para todo $\varepsilon > 0$ y $N \geq N_0 \max\{\varepsilon^{-(d'+2)}, 1\}$,*

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) > \varepsilon) \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \quad (5.1.1)$$

donde

$$\gamma_p = \begin{cases} 1 & p \in [1, 2) \\ 3 - 2\sqrt{2} & p = 2. \end{cases}$$

Como ya se comento en el inicio de este capítulo, para utilizar el resultado anterior necesitamos conocer explícitamente N_0 , lo que no está hecho en la demostración dada por los autores del teorema. Después de un arduo trabajo se calculó el valor de N_0 para el caso $p = 1$, que es el que más nos interesa, y se encontró que la cantidad de observaciones necesarias para que el teorema fuera cierto era muy grande y que además, al calcular el valor numérico del lado derecho de (5.1.1) evaluado en dicho N , era prácticamente cero. Es decir, la cota tenía un comportamiento que en la práctica es binario: antes de un umbral la cota no es cierta, después de dicho umbral el lado derecho de la desigualdad se anula.

Una de las explicaciones para el comportamiento binario de la desigualdad propuesta por el teorema anterior, es que para lograr la cota que se propone es necesario hacer varias elecciones arbitrarias de parámetros y utilizar desigualdades que finalmente resultan poco finas. Una forma de evitar dichas elecciones de parámetros es quedarse con un resultado que está implícito en la demostración original del teorema, y que si bien es menos elegante, se comporta mucho mejor en la práctica.

Teorema 5.1.2. *Sea μ una medida de probabilidad en \mathbb{R} y $\lambda > 0$ tal que μ satisface $T_1(\lambda)$,*

5.1. Concentración de la Medida Empírica con respecto a W

entonces para todo $\lambda_1 < \lambda$, $N \in \mathbb{N}$, $\eta \in (0, 1)$ y $\alpha_1 < \alpha < \lambda/2$,

$$\begin{aligned} \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq & \left(\frac{16eR}{\delta_1 \varepsilon} \right)^{\left(\frac{2R}{\delta_1 \varepsilon} \right)} \exp \left(-N \left[\frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right) \right] \right) + \\ & \exp \left(-N \left[\frac{\alpha_1 R}{2} (1 - \eta) \varepsilon - e^{(\alpha_1 - \alpha) R^2} E_\alpha \right] \right), \end{aligned} \quad (5.1.2)$$

con δ_1 lo suficientemente pequeño y R lo suficientemente grande para que

$$\zeta = (\eta - \delta_1) \varepsilon - 2E_\alpha R e^{-\alpha R^2} > 0, \quad (C1)$$

$$\frac{\alpha_1 R}{2} (1 - \eta) \varepsilon - e^{(\alpha_1 - \alpha) R^2} E_\alpha > 0, \quad (C2)$$

$$\frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right) > 0, \quad (C3)$$

$$R \geq \frac{1}{\sqrt{2\alpha}} \quad (C4)$$

y

$$R^2 \geq -\frac{1}{\alpha} \log \left(\frac{0,7968}{E_\alpha} \right). \quad (C5)$$

Para hacer más fácil de seguir la demostración de este resultado, se probaran varios lemas intermedios que juntos prueban el teorema que acabamos de enunciar.

Para lo que sigue consideremos $R > 0$, $B_R = [-R, R] \subset \mathbb{R}$ y $\mu \in \mathcal{P}_1(\mathbb{R})$ que satisface $T_1(\lambda)$. Definamos

$$\mu_R = \frac{\mathbb{1}_{B_R} \mu}{\mu(B_R)}. \quad (5.1.3)$$

Consideremos el conjunto de variables aleatorias $\{X^k\}_{k=1}^N$ que son i.i.d. y que tiene ley μ y el conjunto de variables $\{Y^k\}_{k=1}^N$ también i.i.d. que se distribuye según μ_R . Supongamos además, que ambos conjuntos de variables son independientes entre sí. Definamos

$$X_R^k = \begin{cases} X^k & \text{si } |X^k| \leq R \\ Y^k & \text{si } |X^k| > R. \end{cases} \quad (5.1.4)$$

Por último llamamos $\hat{\mu}^N$ a la medida empírica asociada a μ y $\hat{\mu}_R^N$ a la medida empírica asociada a μ_R .

Lema 5.1.3. X_R^k tiene ley μ_R .

Demostración: En efecto, si $A \subset \mathbb{R}$ es un conjunto medible, entonces

$$\begin{aligned}
 \mathbb{P}(X_R^k \in A) &= \mathbb{P}(X_R^k \in A / |X^k| \leq R) \mathbb{P}(|X^k| \leq R) + \mathbb{P}(X_R^k \in A / |X^k| > R) \mathbb{P}(|X^k| > R) \\
 &= \mathbb{P}(X^k \in A / |X^k| \leq R) \mu(B_R) + \mathbb{P}(Y^k \in A / |X^k| > R) \mu(B_R^c) \\
 &= \mu_R(A) \mu(B_R) + \mathbb{P}(Y^k \in A) \mu(B_R^c) \\
 &= \mu_R(A) \mu(B_R) + \mu_R(A) \mu(B_R^c) \\
 &= \mu_R(A),
 \end{aligned}$$

donde la primera igualdad se tiene por la regla probabilidades totales, la segunda por definición y la tercera por independencia.

q. e. d.

Lema 5.1.4. Para todo $\alpha < \lambda/2$ y $R \geq 1/\sqrt{2\alpha}$ tenemos que

$$W(\mu, \mu_R) \leq 2Re^{-\alpha R^2} E_\alpha.$$

Demostración: Consideremos X^1 y X_R^1 definidas como antes. Del lema anterior sabemos que X_R^1 tiene ley μ_R . Luego, por definición de la distancia de Wasserstein

$$\begin{aligned}
 W(\mu, \mu_R) &\leq \mathbb{E}(|X^1 - X_R^1|) \\
 &= \mathbb{E}(|X^1 - Y^1| \mathbb{1}_{\{|X^1| > R\}}) \\
 &\leq 2\mathbb{E}(|X^1| \mathbb{1}_{\{|X^1| > R\}}).
 \end{aligned}$$

Es decir

$$W(\mu, \mu_R) \leq 2 \int_{\{|X^1| > R\}} |x| d\mu(x). \quad (5.1.5)$$

Por otro lado, sabemos que si μ satisface $T_1(\lambda)$, entonces para $\alpha < \lambda/2$, tenemos que

$$E_\alpha = \int_{\mathbb{R}} e^{\alpha x^2} d\mu(x) < \infty.$$

Pero, si $R \geq 1/\sqrt{2\alpha}$ se tiene que

$$\frac{r}{e^{\alpha r^2}} \leq \frac{R}{e^{\alpha R^2}},$$

y entonces

$$\int_{\{|X^1| > R\}} |x| d\mu(x) = \int_{\{|X^1| > R\}} |x| \frac{e^{\alpha x^2}}{e^{\alpha x^2}} d\mu(x) \leq \frac{R}{e^{\alpha R^2}} \int_{\{|X^1| > R\}} e^{\alpha x^2} d\mu(x) \leq \frac{R}{e^{\alpha R^2}} E_\alpha.$$

Introduciendo está última desigualdad en (5.1.5) obtenemos

$$W(\mu, \mu_R) \leq 2Re^{-\alpha R^2} E_\alpha.$$

5.1. Concentración de la Medida Empírica con respecto a W

q.e.d.

Lema 5.1.5. *Para todo $\alpha_1 < \alpha$, $\theta > 0$, $\varepsilon > 0$ y $R \geq 2\theta/\alpha_1$ se tiene que*

$$\mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) \leq \exp\left(-N \left[\theta\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha\right]\right). \quad (5.1.6)$$

Demostración: Recordemos que las medidas empíricas están dadas por

$$\hat{\mu}^N = \frac{1}{N} \sum_{k=1}^N \delta_{X^k} \quad \text{y} \quad \hat{\mu}_R^N = \frac{1}{N} \sum_{k=1}^N \delta_{X_R^k}.$$

Entonces, un posible *coupling* entre $\hat{\mu}^N$ y $\hat{\mu}_R^N$ está dado por $\pi = \frac{1}{N} \sum_{k=1}^N \delta_{(X^k, X_R^k)}$. Luego, por la definición de la distancia de Wasserstein se tiene que

$$W(\hat{\mu}^N, \hat{\mu}_R^N) \leq \frac{1}{N} \sum_{k=1}^N |X_R^k - X^k|.$$

Pero $|X_R^k - X^k| \leq 2|X^k| \mathbb{1}_{\{|X^k| > R\}} =: Z^k$, y entonces para todo $\theta > 0$, se tiene que

$$\begin{aligned} \mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) &\leq \mathbb{P}\left(\frac{1}{N} \sum_{k=1}^N Z^k > \varepsilon\right) \\ &= \mathbb{P}\left(\sum_{k=1}^N \theta(Z^k - \varepsilon) > 0\right) \\ &= \mathbb{P}\left(\exp\left(\sum_{k=1}^N \theta(Z^k - \varepsilon)\right) > 1\right) \\ &\leq \mathbb{E}\left(\exp\left(\sum_{k=1}^N \theta(Z^k - \varepsilon)\right)\right), \end{aligned}$$

donde la última desigualdad corresponde a la desigualdad de Markov y el parámetro θ absorbe el factor $1/N$. Ahora notamos que las variables $\{Z^k\}_k$ son i.i.d., con lo que se tiene que

$$\begin{aligned} \mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) &\leq \left(\mathbb{E}[\exp\{\theta(Z^1 - \varepsilon)\}]\right)^N \\ &= \left(\exp(-\theta\varepsilon) \mathbb{E}[\exp\{\theta Z^1\}]\right)^N \\ &= \exp\left(-N [\theta\varepsilon - \log \mathbb{E} \exp(\theta Z^1)]\right), \end{aligned}$$

para todo $\theta > 0$. Acotemos ahora el término $\mathbb{E} \exp(\theta Z^1)$. Para ello, notemos primero que para todo $\alpha_1 < \alpha$ si $R \geq 2\theta/\alpha_1$, entonces se tiene

$$2\theta R \leq \alpha_1 R^2,$$

5.1. Concentración de la Medida Empírica con respecto a W

y entonces

$$\begin{aligned}
\mathbb{E} \exp(\theta Z^1) &\leq \mathbb{E} \exp(\alpha_1 |X^1|^2 \mathbb{1}_{\{|X^1|>R\}}) \\
&= \int_{|x|\leq R} \exp(\alpha_1 x^2 \mathbb{1}_{\{|x|>R\}}) d\mu(x) + \int_{|x|>R} \exp(\alpha_1 x^2 \mathbb{1}_{\{|x|>R\}}) d\mu(x) \\
&\leq 1 + \int_{|x|>R} e^{\alpha_1 x^2} d\mu(x) \\
&\leq 1 + e^{(\alpha_1 - \alpha)R^2} \int_{|x|>R} e^{\alpha x^2} d\mu(x) \\
&\leq 1 + e^{(\alpha_1 - \alpha)R^2} E_\alpha.
\end{aligned}$$

Con esto concluimos que

$$\log \mathbb{E} \exp(\theta Z^1) \leq e^{(\alpha_1 - \alpha)R^2} E_\alpha,$$

de donde finalmente, si $R \geq 2\theta/\alpha_1$, se tiene que

$$\mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) \leq \exp\left(-N \left[\theta\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha\right]\right). \quad (5.1.7)$$

q.e.d.

Lema 5.1.6. *Sea ν con soporte en B_R y absolutamente continua con respecto a μ . Entonces*

$$H(\nu|\mu_R) - H(\nu|\mu) \geq -2E_\alpha e^{-\alpha R^2}. \quad (5.1.8)$$

Demostración: Es directo que $\nu \ll \mu_R$. Por otro lado

$$\begin{aligned}
H(\nu|\mu) &= \int_{B_R} \frac{\partial \nu}{\partial \mu} \log\left(\frac{\partial \nu}{\partial \mu}\right) d\mu \\
&= \int_{B_R} \frac{\partial \nu}{\partial \mu} \log\left(\frac{\partial \nu}{\partial \mu_R} \frac{\partial \mu_R}{\partial \mu}\right) d\mu \\
&= \int_{B_R} \frac{\partial \nu}{\partial \mu} \log\left(\frac{\partial \nu}{\partial \mu_R}\right) d\mu + \int_{B_R} \frac{\partial \nu}{\partial \mu} \log\left(\frac{\partial \mu_R}{\partial \mu}\right) d\mu \\
&= \int_{B_R} \frac{\partial \nu}{\partial \mu_R} \log\left(\frac{\partial \nu}{\partial \mu_R}\right) \frac{\partial \mu_R}{\partial \mu} d\mu + \int_{B_R} \frac{\partial \nu}{\partial \mu} \log\left(\frac{\mathbb{1}_{B_R}}{\mu(B_R)}\right) d\mu \\
&= H(\nu|\mu_R) - \log(\mu(B_R)) \int_{B_R} \frac{\partial \nu}{\partial \mu} d\mu.
\end{aligned}$$

Es decir,

$$H(\nu|\mu_R) - H(\nu|\mu) = \log(\mu(B_R)).$$

Por otro lado notemos que

$$\mu(B_R^c) = \int_{|x|>R} 1 d\mu x \leq \int_{|x|>R} e^{\alpha(x^2 - R^2)} d\mu(x) \leq e^{-\alpha R^2} E_\alpha,$$

5.1. Concentración de la Medida Empírica con respecto a W

y entonces se tiene

$$\mu(B_R) \geq 1 - E_\alpha e^{-\alpha R^2}.$$

Además si $x < 0,7968$ entonces $\log(1 - x) \geq -2x$. Entonces si

$$R^2 \geq -\frac{1}{\alpha} \log\left(\frac{0,7968}{E_\alpha}\right),$$

se tiene

$$\log(\mu(B_R)) \geq -2E_\alpha e^{-\alpha R^2},$$

con lo que finalmente concluimos

$$H(\nu|\mu_R) - H(\nu|\mu) \geq -2E_\alpha e^{-\alpha R^2}.$$

q.e.d.

Lema 5.1.7. *Para todo $x, y \in \mathbb{R}$ y $a \in (0, 1)$ se tiene que:*

$$(x - y)^2 \geq ax^2 - \left(\frac{a}{1-a}\right)y^2. \quad (5.1.9)$$

Demostración: Consideremos la función $f(z) = (1 - a)z^2 - 2z + 1/(1 - a)$. Notemos que

$$f'(z) = 2(1 - a)z - 2.$$

Luego, el mínimo de $f(z)$ se alcanza en $z^* = 1/(1 - a)$ y toma el valor

$$f(z^*) = \frac{1 - a}{(1 - a)^2} - \frac{2}{1 - a} + \frac{1}{1 - a} = 0.$$

Entonces $f(z) \geq 0$ para todo $z \in \mathbb{R}$. Si $y \neq 0$ entonces podemos tomar $z = x/y$ y encontramos

$$(1 - a) \left(\frac{x}{y}\right)^2 - 2\frac{x}{y} + \frac{1}{1 - a} \geq 0,$$

de donde multiplicando por y^2 y reordenando términos se obtiene la desigualdad deseada. El caso $y = 0$ es trivial.

q.e.d.

Lema 5.1.8. *Sea μ que satisface $T_1(\lambda)$ y $\lambda_1 < \lambda$. Entonces para toda ν con soporte en B_R*

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} W^2(\mu_R, \nu) - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1\lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right). \quad (5.1.10)$$

5.1. Concentración de la Medida Empírica con respecto a W

Demostración: Como μ satisface $T_1(\lambda)$, gracias a la desigualdad triangular se tiene que

$$H(\nu|\mu) \geq \frac{\lambda}{2} W^2(\mu, \nu) \geq \frac{\lambda}{2} (W(\mu_R, \nu) - W(\mu_R, \mu))^2.$$

Introduciendo esto en (5.1.8) encontramos que

$$H(\nu|\mu_R) \geq \frac{\lambda}{2} (W(\mu_R, \nu) - W(\mu_R, \mu))^2 - 2E_\alpha e^{-\alpha R^2}. \quad (5.1.11)$$

Gracias al lema anterior para todo $a \in (0, 1)$ se tiene que

$$H(\nu|\mu_R) \geq \frac{\lambda}{2} a W^2(\mu_R, \nu) - \frac{\lambda a}{2(1-a)} W^2(\mu, \mu_R) - 2E_\alpha e^{-\alpha R^2}.$$

Pero del lema 5.1.5, $W^2(\mu, \mu_R) \leq 4E_\alpha^2 R^2 e^{-2\alpha R^2} \leq 4E_\alpha^2 R^2 e^{-\alpha R^2}$, lo que agregado a lo anterior, nos dice que $\forall \lambda_1 < \lambda$

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} W^2(\mu_R, \nu) - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right).$$

q.e.d.

A continuación, dado \mathcal{B} un boreliano en $\mathcal{P}(B_R)$, queremos encontrar una estimación para $\mathbb{P}(\hat{\mu}_R^N \in \mathcal{B})$.

Lema 5.1.9. *Si \mathcal{B} es convexo y compacto, entonces*

$$\mathbb{P}(\hat{\mu}_R^N \in \mathcal{B}) \leq \exp \left(-N \inf_{\nu \in \mathcal{B}} H(\nu|\mu_R) \right). \quad (5.1.12)$$

Demostración: Comencemos por notar que para toda $\phi : B_R \rightarrow \mathbb{R}$ continua y acotada se tiene

$$\mathbb{P}(\hat{\mu}_R^N \in \mathcal{B}) \leq \mathbb{E} \left(\exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right).$$

En efecto,

$$\begin{aligned} \mathbb{E} \left(\exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right) &= \mathbb{E} \left(\mathbb{1}_{\{\hat{\mu}_R^N \in \mathcal{B}\}} \exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right) \\ &+ \mathbb{E} \left(\mathbb{1}_{\{\hat{\mu}_R^N \in \mathcal{B}^c\}} \exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right) \\ &\geq \mathbb{E} \left(\mathbb{1}_{\{\hat{\mu}_R^N \in \mathcal{B}\}} \exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right) \\ &\geq \mathbb{E} \left(\mathbb{1}_{\{\hat{\mu}_R^N \in \mathcal{B}\}} \right), \end{aligned}$$

5.1. Concentración de la Medida Empírica con respecto a W

donde la primera desigualdad se obtiene porque el integrando es positivo y la segunda porque

$$\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \leq 0 \text{ si } \hat{\mu}_R^N \in \mathcal{B}.$$

Por otro lado, tenemos que $\{X_R^k\}$ es una muestra i.i.d. con ley μ_R , de donde

$$\begin{aligned} \mathbb{P}(\hat{\mu}_R^N \in \mathcal{B}) &\leq \mathbb{E} \left(\exp \left[-N \left(\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \int_{B_R} \phi d\hat{\mu}_R^N \right) \right] \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu \right) \mathbb{E} \left(\exp \left(N \int_{B_R} \phi d\hat{\mu}_R^N \right) \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu \right) \mathbb{E} \left(\exp \left(\sum_{k=0}^N \phi(X_R^k) \right) \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu \right) \mathbb{E} \left(\exp \left(\phi(X_R^1) \right) \right)^N \\ &= \exp \left(-N \left[\inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] \right). \end{aligned}$$

Así, se ha probado que para cualquier función ϕ , continua y acotada en B_R , se tiene que

$$-\frac{\log \mathbb{P}(\hat{\mu}_R^N \in \mathcal{B})}{N} \geq \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right].$$

Luego, podemos tomar supremo sobre el conjunto de funciones continuas y acotadas en B_R , al que denotamos como $\mathcal{C}_b(B_R)$, y encontramos que

$$\mathbb{P}(\hat{\mu}_R^N \in \mathcal{B}) \leq \exp \left(-N \sup_{\phi \in \mathcal{C}_b(B_R)} \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] \right).$$

Ahora notemos que si definimos:

$$\begin{aligned} I : \mathcal{B} \times \mathcal{C}_b(B_R) &\rightarrow \mathbb{R} \\ (\nu, \phi) &\rightarrow \int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R, \end{aligned}$$

I es lineal continuo en ν y cóncavo con respecto a ϕ .¹ Luego, si asumimos que \mathcal{B} es convexo y compacto, entonces por el teorema de minimax de Sion que aparece en el anexo A.1 se tiene que

$$\sup_{\phi \in \mathcal{C}_b(B_R)} \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] = \inf_{\nu \in \mathcal{B}} \sup_{\phi \in \mathcal{C}_b(B_R)} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right].$$

¹Las propiedades de I se discuten en el anexo A.2.1.

5.1. Concentración de la Medida Empírica con respecto a W

Por último notamos que el supremo que aparece en el lado derecho de esta última igualdad es justamente la representación dual de la entropía, que aparece en el anexo A.2.1. Con lo que concluimos

$$\mathbb{P}(\hat{\mu}_R^N \in \mathcal{B}) \leq \exp \left(-N \inf_{\nu \in \mathcal{B}} H(\nu | \mu_R) \right).$$

q.e.d.

Demostración: del Teorema 5.1.2. Comencemos por notar que para R dado y $\eta \in (0, 1)$

$$\begin{aligned} \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) &\leq \mathbb{P}(W(\mu, \mu_R) + W(\mu_R, \hat{\mu}_R^N) + W(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon) \\ &\leq \mathbb{P}(W(\mu, \mu_R) + W(\mu_R, \hat{\mu}_R^N) > \eta\varepsilon) \\ &\quad + \mathbb{P}(W(\hat{\mu}_R^N, \hat{\mu}^N) > (1 - \eta)\varepsilon). \end{aligned}$$

Gracias al lema 5.1.4, tenemos que si $R \geq 1/\sqrt{2\alpha}$ entonces

$$W(\mu, \mu_R) \leq 2Re^{-\alpha R^2} E_\alpha,$$

mientras que gracias al lema 5.1.5 tenemos que

$$\mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) \leq \exp \left(-N \left[\theta\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha \right] \right),$$

para $R \geq 2\theta/\alpha_1$. Entonces,

$$\begin{aligned} \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) &\leq \mathbb{P}(W(\mu_R, \hat{\mu}_R^N) > \eta\varepsilon - 2Re^{-\alpha R^2} E_\alpha) \\ &\quad \exp \left(-N \left[\theta(1 - \eta)\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha \right] \right). \end{aligned} \tag{5.1.13}$$

Sea \mathcal{A} un subconjunto medible de $\mathcal{P}(B_R)$ y $\delta > 0$. Definimos \mathcal{A}_δ , el engrosamiento de tamaño δ de \mathcal{A} , como

$$\mathcal{A}_\delta = \{ \nu \in \mathcal{P}(B_R) : (\exists \nu_a \in \mathcal{A}) W(\nu, \nu_a) \leq \delta \}.$$

Sea $\mathcal{N}^{\mathcal{A}}$ el número mínimo de bolas de tamaño $\delta/2$ necesarias para recubrir \mathcal{A} . Llamemos $(B_i)_{i=1}^{\mathcal{N}^{\mathcal{A}}}$ al conjunto de dichas bolas. Notemos cada uno de los conjuntos B_i es convexo y compacto. En efecto, del teorema 3.2.14, sabemos que $\mathcal{P}_p(B_R)$ es un espacio métrico polaco y compacto. Además, gracias al Corolario 6.11 de [43], sabemos que W_p es una función continua en $\mathcal{P}_p(B_R) \times \mathcal{P}_p(B_R)$. Luego, cada B_i es un subconjunto cerrado de un espacio métrico compacto, y por lo tanto compacto. Para probar la convexidad de B_i , basta llamar μ_i al centro de la bola y tomar $\nu_1, \nu_2 \in B_i$. Sean π_1 y π_2 los *couplings* óptimos de (μ_i, ν_1) y (μ_i, ν_2) respectivamente. Entonces, para todo $\gamma \in (0, 1)$, tenemos que $\pi = \gamma\pi_1 + (1 - \gamma)\pi_2$ es

5.1. Concentración de la Medida Empírica con respecto a W

un *coupling* para $(\mu_i, \gamma\nu_1 + (1 - \gamma)\nu_2)$ y entonces

$$\begin{aligned}
 W(\mu_i, \gamma\nu_1 + (1 - \gamma)\nu_2) &\leq \int_{B_R \times B_R} |x - y| d\pi(x, y) \\
 &= \gamma \int_{B_R \times B_R} |x - y| d\pi_1(x, y) + (1 - \gamma) \int_{B_R \times B_R} |x - y| d\pi_2(x, y) \\
 &= \gamma W(\mu_i, \nu_1) + (1 - \gamma) W(\mu_i, \nu_2) \\
 &\leq \delta/2,
 \end{aligned}$$

de donde se concluye que B_i es un conjunto convexo. Luego, podemos aplicar (5.1.12) a cada B_i y obtenemos

$$\begin{aligned}
 \mathbb{P}(\hat{\mu}_R^N \in \mathcal{A}) &\leq \mathbb{P}\left(\hat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}^{\mathcal{A}}} B_i\right) \\
 &\leq \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \mathbb{P}(\hat{\mu}_R^N \in B_i) \\
 &\leq \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \exp\left(-N \inf_{\nu \in B_i} H(\nu|\mu_R)\right) \\
 &\leq \mathcal{N}^{\mathcal{A}} \exp\left(-N \inf_{\nu \in \mathcal{A}_\delta} H(\nu|\mu_R)\right),
 \end{aligned}$$

pues para todo i , $B_i \subset \mathcal{A}_\delta$.

Consideremos ahora el conjunto

$$\mathcal{A} = \left\{ \nu \in \mathcal{P}(B_R) : W(\nu, \mu_R) \geq \eta\varepsilon - 2E_\alpha R e^{-\alpha R^2} \right\}.$$

Entonces, por desigualdad triangular, se tiene para toda medida $\nu \in \mathcal{A}_\delta$

$$W(\nu, \mu_R) \geq \eta\varepsilon - 2E_\alpha R e^{-\alpha R^2} - \delta,$$

pero como $W(\cdot, \cdot) \geq 0$, tenemos que

$$W(\nu, \mu_R) \geq \max\left(\eta\varepsilon - 2E_\alpha R e^{-\alpha R^2} - \delta, 0\right) =: \zeta.$$

Gracias al lema 5.1.8, sabemos que μ_R satisface una desigualdad de Talagrand modificada,

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} W^2(\mu_R, \nu) - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right),$$

y entonces

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right).$$

5.1. Concentración de la Medida Empírica con respecto a W

Notemos que esta cota es independiente de ν , luego tenemos que

$$\inf_{\nu \in \mathcal{A}_\delta} H(\nu | \mu_R) \geq \frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right),$$

y, por lo tanto,

$$\begin{aligned} \mathbb{P}(\hat{\mu}_R^N \in \mathcal{A}) &= \mathbb{P}\left(W(\hat{\mu}_R^N, \mu_R) \geq \eta\varepsilon - 2E_\alpha R e^{-\alpha R^2}\right) \\ &\leq \mathcal{N}^{\mathcal{A}} \exp\left(-N \left[\frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right) \right]\right). \end{aligned}$$

Ahora supongamos que $\delta = \delta_1 \varepsilon$ con δ_1 lo suficientemente pequeño de modo que

$$\zeta = \eta\varepsilon - 2E_\alpha R e^{-\alpha R^2} - \delta_1 \varepsilon = (\eta - \delta_1)\varepsilon - 2E_\alpha R e^{-\alpha R^2}.$$

Gracias al teorema A.2.3, tenemos que para todo $R > 0$ y $0 < \delta < R$

$$\mathcal{N}^{\mathcal{A}} \leq \left(\frac{16eR}{\delta} \right)^{\left(\frac{2R}{\delta} \right)}.$$

Con esto finalmente vemos que

$$\begin{aligned} \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) &\leq \left(\frac{16eR}{\delta_1 \varepsilon} \right)^{\left(\frac{2R}{\delta_1 \varepsilon} \right)} \exp\left(-N \left[\frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right) \right]\right) + \\ &\quad \exp\left(-N \left[\frac{\alpha_1 R}{2} (1 - \eta)\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha \right]\right), \end{aligned}$$

si δ_1 lo suficientemente pequeño y R lo suficientemente grande para que

$$\zeta = (\eta - \delta_1)\varepsilon - 2E_\alpha R e^{-\alpha R^2} > 0, \tag{C1}$$

$$\frac{\alpha_1 R}{2} (1 - \eta)\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha > 0, \tag{C2}$$

$$\frac{\lambda_1}{2} \zeta^2 - E_\alpha e^{-\alpha R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\alpha R^2 + 2 \right) > 0, \tag{C3}$$

$$R \geq \frac{1}{\sqrt{2\alpha}} \tag{C4}$$

y

$$R^2 \geq -\frac{1}{\alpha} \log\left(\frac{0,7968}{E_\alpha}\right). \tag{C5}$$

q. e. d.

Aplicación del teorema 5.1.2 en la práctica

Para poder aplicar el teorema anterior queda pendiente encontrar una forma sistemática de calcular R y δ_1 , de modo que satisfagan las condiciones (C1)-(C5). Además notemos que si bien el teorema es cierto para todo $\eta \in (0, 1)$, $\lambda_1 < \lambda$ y $\alpha_1 < \alpha < \lambda/2$, nos interesa escoger estos parámetros de forma que el N_0 mínimo necesario para hacer la cota no trivial sea lo menor posible.

Comencemos por notar que N_0 crece con R , el que debido a las condiciones (C4) y (C5) explota cuando α se acerca a cero. Por ello fijamos α tan grande como sea posible, lo que a su vez también fija el valor de E_α . Resta fijar η , δ_1 , λ_1 y α_1 . Para ello seguimos el siguiente procedimiento:

1. Definimos R_0 como el mínimo R que satisface (C4) y (C5).
2. Se discretiza el intervalo $(0, 1)$ generando un vector $(\eta_1, \dots, \eta_{N_1})$.
3. Para $k = 2, \dots, N_1$:
 - a) Para $\eta = \eta_k$ se fija R_1 como el mínimo R que satisface (C2). Notemos que con todos los otros parámetros fijos, independientemente del signo que tenga el lado izquierdo de (C2), α_1 se puede fijar de modo que dicha cantidad sea máxima. Se escoge de esta manera.
 - b) Para $j = 1, \dots, k - 1$:
 - 1) Para $\eta = \eta_k$ y $\delta_1 = \eta_j$ se calcula R_2 como el mínimo R que satisface las condiciones (C1) y (C3) simultáneamente. Notemos que al igual que α_1 en (C2), el parámetro λ_1 se puede escoger de manera que el lado izquierdo de (C3) sea máximo cuando todos los otros parámetros están fijos.
 - 2) Definimos $R_{k,j} = \max\{R_0, R_1, R_2\}$ y calculamos $N_{k,j}$, el mínimo N que hace la cota del teorema significativa cuando se considera $\eta = \eta_k$, $\delta_1 = \eta_j$, $R = R_{k,j}$ y los parámetros α_1 y λ_1 escogidos de manera óptima como se mencionó antes.
4. Fijamos los parámetros $\eta = \eta_{k^*}$ y $\delta_1 = \eta_{j^*}$ donde

$$N_{k^*j^*} = \min\{N_{k,j} : k = 2, \dots, N_1, j = 1, \dots, k - 1\}.$$

Para desarrollar los ejemplos que se presentan a continuación el procedimiento anterior fue programado en Matlab.

Ejemplo 5.1.10: En este ejemplo estudiaremos la cota obtenida con el teorema 5.1.2 cuando μ es igual a la ley de una variable aleatoria normal estándar $\mathcal{N}(0, 1)$. En este caso tenemos que μ satisface $T_1(1)$ y que $E_\alpha = 1/\sqrt{1 - 2\alpha}$ y es finito para $\alpha \in [0, 1/2)$. Para empezar, en la tabla 5.1.1 aparece el comportamiento de $W(\mu, \hat{\mu}^N)$ para distintos N . Para cada número de observaciones se generaron 100 muestras y se estimó la distancia promedio, el máximo de la distancia y la probabilidad de que $W(\mu, \hat{\mu}^N) > \varepsilon$ para distintos valores de ε .

5.1. Concentración de la Medida Empírica con respecto a W

N	prom $\{W(\mu, \hat{\mu}^N)\}$	máx $W(\mu, \hat{\mu}^N)$	$\hat{\mathbb{P}}^N(W(\mu, \hat{\mu}^N) > \varepsilon)$				
			$\varepsilon = 0,01$	$\varepsilon = 0,05$	$\varepsilon = 0,1$	$\varepsilon = 0,5$	$\varepsilon = 1$
10	0,4125	0,8327	1,0000	1,0000	1,0000	0,2000	0,0000
20	0,2788	0,5756	1,0000	1,0000	1,0000	0,0200	0,0000
50	0,1781	0,4235	1,0000	1,0000	0,9700	0,0000	0,0000
100	0,1213	0,3616	1,0000	1,0000	0,6600	0,0000	0,0000
200	0,0925	0,2039	1,0000	0,9900	0,2900	0,0000	0,0000
1000	0,0405	0,0819	1,0000	0,2200	0,0000	0,0000	0,0000
2000	0,0302	0,0696	1,0000	0,0700	0,0000	0,0000	0,0000
5000	0,0178	0,0383	0,9300	0,0000	0,0000	0,0000	0,0000
10000	0,0126	0,0308	0,6600	0,0000	0,0000	0,0000	0,0000
50000	0,0056	0,0131	0,0200	0,0000	0,0000	0,0000	0,0000
100000	0,0041	0,0102	0,0100	0,0000	0,0000	0,0000	0,0000
500000	0,0019	0,0059	0,0000	0,0000	0,0000	0,0000	0,0000

Cuadro 5.1.1: Comportamiento de $W(\mu, \hat{\mu}^N)$ para distintos N . Para cada número de observaciones se generaron 100 muestras y se calcularon los estadísticos que aparecen en las columnas.

La distancia de Wasserstein entre la medida empírica de cada muestra y la ley de la muestra, se calculó numéricamente haciendo uso de la fórmula integral dada en la observación 3.2.18, para lo cual primero se calculó la función cuantil empírica, también de forma numérica.

El teorema 5.1.2 nos dice que si se satisfacen ciertas condiciones en los parámetros que en él aparecen, entonces

$$\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq e^{a-Nb} + e^{-Nc},$$

donde a, b, c son constantes que dependen entre otras cosas de ε . Notemos que si bien esta cota es válida para todo N , es significativa a partir de un N_0 tal que el lado derecho sea menor o igual que 1. En la tabla 5.1.2 aparece el valor de a, b, c, ζ y N_0 para distintos valores de ε . Notemos que b, c y ζ son las cantidades a las que se les exigía ser positivas para que el teorema fuera cierto.

ε	a	b	c	ζ	N_0
0,01	$4,521e + 04$	$5,067e - 06$	$5,905e - 05$	$6,400e - 03$	$8,922e + 09$
0,05	$7,118e + 03$	$1,157e - 04$	$2,486e - 04$	$3,200e - 02$	$6,151e + 07$
0,1	$3,166e + 03$	$4,422e - 04$	$4,523e - 04$	$6,400e - 02$	$7,161e + 06$
0,5	$4,474e + 02$	$9,420e - 03$	$1,679e - 03$	$3,150e - 01$	$4,749e + 04$
1	$1,901e + 02$	$3,512e - 02$	$2,758e - 03$	$6,300e - 01$	$5,412e + 03$

Cuadro 5.1.2: Valores para a, b, c, ζ y N_0 para distintos valores de ε utilizando el teorema 5.1.2 .

5.1. Concentración de la Medida Empírica con respecto a W

Comparando los resultados de las tablas 5.1.1 y 5.1.2, nos damos cuenta de que el teorema es bastante ineficiente, ya que por ejemplo, para $\mathbb{P}(W(\mu, \hat{\mu}^N) > 1) \leq 1$ se requiere del orden de 5400 datos, sin embargo para $N = 10$ observamos que la estimación de $\mathbb{P}(W(\mu, \hat{\mu}^N) > 1)$ que aparece en la tabla 5.1.1 es 0. Uno de los posibles caminos que exploramos intentando de mejorar el resultado fue considerar el corolario 4.2.11, el cual aplicado a nuestro caso nos dice que μ_R satisface $T_1(1/R^2)$, si utilizamos este resultado en la demostración del teorema 5.1.2 en lugar del lema 5.1.8, obtenemos el siguiente teorema.

Teorema 5.1.11. *Sea μ una medida de probabilidad en \mathbb{R} y $\lambda > 0$ tal que μ satisface $T_1(\lambda)$, es decir*

$$W(\nu, \mu) \leq \sqrt{\frac{2}{\lambda} H(\nu|\mu)},$$

entonces para todo $N \in \mathbb{N}$, $\eta \in (0, 1)$ y $\alpha_1 < \alpha < \lambda/2$

$$\begin{aligned} \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq & \exp\left(\left(\frac{2R}{\delta_1 \varepsilon}\right) \log\left(\frac{16eR}{\delta_1 \varepsilon}\right) - \frac{N}{2R^2} \zeta^2\right) + \\ & \exp\left(-N \left[\frac{\alpha_1 R}{2}(1-\eta)\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha\right]\right), \end{aligned} \quad (5.1.14)$$

con δ_1 lo suficientemente pequeño y R lo suficientemente grande para que

$$\zeta = (\eta - \delta_1)\varepsilon - 2E_\alpha R e^{-\alpha R^2} > 0,$$

$$\frac{\alpha_1 R}{2}(1-\eta)\varepsilon - e^{(\alpha_1 - \alpha)R^2} E_\alpha > 0,$$

$$R \geq \frac{1}{\sqrt{2\alpha}}$$

y

$$R^2 \geq -\frac{1}{\alpha} \log\left(\frac{0,7968}{E_\alpha}\right).$$

Demostración: Es exactamente la misma que la del teorema 5.1.2, salvo que cuando se introduce el resultado del lema 5.1.8, ahora hay que usar el corolario 4.2.11.

q.e.d.

Ejemplo 5.1.12: En este ejemplo estudiaremos la cota obtenida con el teorema 5.1.11 al igual que antes consideramos μ igual a la ley de una variable aleatoria normal estándar $\mathcal{N}(0, 1)$.

Comparando los resultados de las tablas 5.1.2 y 5.1.3, encontramos que los resultado empeoran al utilizar el teorema 5.1.11. La razón para esto pareciera ser que el criterio integral para encontrar λ tal que μ_R satisface $T_1(\lambda)$ es muy poco eficiente y resulta ser mejor utilizar el lema 5.1.8.

5.1. Concentración de la Medida Empírica con respecto a W

ε	a	b	c	ζ	N_0
0,01	$3,769e + 04$	$5,943e - 07$	$1,440e - 05$	$6,300e - 03$	$6,341e + 10$
0,05	$6,076e + 03$	$1,657e - 05$	$7,171e - 05$	$3,150e - 02$	$3,668e + 08$
0,1	$2,738e + 03$	$6,973e - 05$	$1,431e - 04$	$6,300e - 02$	$3,927e + 07$
0,5	$4,019e + 02$	$1,922e - 03$	$7,113e - 04$	$3,100e - 01$	$2,091e + 05$
1	$1,722e + 02$	$8,071e - 03$	$1,888e - 03$	$6,100e - 01$	$2,134e + 04$

Cuadro 5.1.3: Valores para a, b, c, ζ y N_0 para distintos valores de ε utilizando el teorema 5.1.11.

Ejemplo 5.1.13: Consideremos en \mathbb{R} la ley Weibull simetrizada de parámetros $(4, 2)$, a la que llamaremos μ . Sabemos que μ satisface $T_1(0,07946111)$. Además

$$E_\alpha = \int_{\mathcal{X}} e^{\alpha x^2} d\mu(x) = \frac{1}{1 - 16\alpha},$$

con esto podemos aplicar el teorema 5.1.2 y encontrar las estimaciones que aparecen en las tablas 5.1.4 y 5.1.5. Al igual que en el ejemplo para la variable normal estándar podemos apreciar que el teorema es ineficiente en el sentido de que requiere de mucha más información que la que es realmente necesaria para entregar una cota significativa.

ε	a	b	c	ζ	N_0
0,01	$1,516e + 05$	$1,730e - 07$	$2,298e - 05$	$6,400e - 03$	$8,760e + 11$
0,05	$2,448e + 04$	$4,002e - 06$	$1,009e - 04$	$3,200e - 02$	$6,116e + 09$
0,1	$1,104e + 04$	$1,541e - 05$	$1,886e - 04$	$6,400e - 02$	$7,165e + 08$
0,5	$1,630e + 03$	$3,370e - 04$	$7,770e - 04$	$3,150e - 01$	$4,837e + 06$
1	$7,113e + 02$	$1,279e - 03$	$1,389e - 03$	$6,300e - 01$	$5,562e + 05$

Cuadro 5.1.4: Valores para a, b, c, ζ y N_0 para distintos valores de ε utilizando el teorema 5.1.2.

N	$\text{prom}\{W(\mu, \hat{\mu}^N)\}$	$\text{máx } W(\mu, \hat{\mu}^N)$	$\varepsilon = 0,01$	$\varepsilon = 0,05$	$\varepsilon = 0,1$	$\varepsilon = 0,5$	$\varepsilon = 1$
10	1,4257	4,7966	1,0000	1,0000	1,0000	1,0000	0,7200
100	0,4635	1,1656	1,0000	1,0000	1,0000	0,3100	0,0300
200	0,3281	0,8411	1,0000	1,0000	1,0000	0,0900	0,0000
5000	0,0674	0,1376	1,0000	0,7100	0,1000	0,0000	0,0000

Cuadro 5.1.5: Comportamiento de $W(\mu, \hat{\mu}^N)$ para distintos N . Para cada número de observaciones se generaron 100 muestras y se calcularon los estadísticos que aparecen en las columnas.

5.1.2. Teorema de Gozlan y Léonard

En esta sección estudiaremos tres resultados obtenidos por Gozlan y Léonard en [23], que permiten obtener desigualdades de concentración para medidas empíricas. El argumento utilizado por los autores es distinto a los presentados por Bolley et al. en [8] y pasa por las propiedades de tensorización de las desigualdades de transporte. Gracias a lo explícito de los resultados de Gozlan y Léonard, no es necesario rehacer las demostraciones de dichos resultados para poder usarlos en el contexto que queremos.

Por último, la versión original de los resultados que presentaremos aborda una clase más grande de desigualdades de transporte. Sin embargo, nos quedamos con una versión un poco más particular, pero que incluye todas las aplicaciones que nos interesan.

Lema 5.1.14. [Gozlan y Léonard. (2007)] *Consideremos (\mathcal{X}, d) un espacio métrico polaco. Sea $p \geq 1$ y $\mu \in \mathcal{P}_p(\mathcal{X})$ que satisface $T_p(\lambda)$. Consideremos $Z : \mathcal{X}^N \rightarrow \mathbb{R}$ una función $N^{-1/p}$ -Lipschitz con respecto a la métrica*

$$(x, y) \rightarrow \left(\sum_{i=1}^N d^p(x_i, y_i) \right)^{1/p}.$$

Entonces, para todo $t \geq 0$

$$\mu^{\otimes N} \left(Z(X_1, \dots, X_N) \geq \int_{\mathcal{X}^N} Z(x_1, \dots, x_N) d\mu^{\otimes N} + t \right) \leq \exp \left(-\frac{\lambda N}{2} t^2 \right)$$

Teorema 5.1.15. [Gozlan y Léonard. (2007)] *Consideremos (\mathcal{X}, d) un espacio métrico polaco y sea \mathcal{G} un conjunto de funciones 1-Lipschitz, μ una medida de probabilidad en \mathcal{X} y $\{X_1, \dots, X_N\}$ una muestra i.i.d. de la ley μ . Llamamos proceso empírico a la variable aleatoria*

$$Z_N^{\mathcal{G}} = \sup_{\phi \in \mathcal{G}} \left\{ \left| \frac{1}{N} \sum_{i=1}^N \phi(X_i) - \int_{\mathcal{X}} \phi d\mu \right| \right\}.$$

Si $\mu \in \mathcal{P}_1(\mathcal{X})$ y satisface $T_1(\lambda)$, entonces

$$\mathbb{P}(Z_N^{\mathcal{G}} \geq \mathbb{E}(Z_N^{\mathcal{G}}) + t) \leq \exp \left(-\frac{\lambda N}{2} t^2 \right), \quad \forall t \geq 0.$$

Teorema 5.1.16. [Gozlan y Léonard. (2007)] *Consideremos (\mathcal{X}, d) un espacio métrico polaco y μ una medida de probabilidad en \mathcal{X} . Sea $\{X_1, \dots, X_N\}$ una muestra i.i.d. de la ley μ . Si $\mu \in \mathcal{P}_p(\mathcal{X})$ y satisface $T_p(\lambda)$, entonces*

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \mathbb{E}[W_p(\mu, \hat{\mu}^N)] + t) \leq \exp \left(-\frac{\lambda N}{2} t^2 \right), \quad \forall t \geq 0.$$

5.1. Concentración de la Medida Empírica con respecto a W

Los dos teoremas que se acaban de enunciar son consecuencias más o menos directas del lema 5.1.14. En ambos casos la demostración pasa por probar que alguna función es $N^{-1/p}$ -Lipschitz con respecto a la métrica definida en el lema. Por ejemplo, en el caso del teorema 5.1.16, consideremos la función

$$\begin{aligned} L_N &: \mathcal{X}^N \rightarrow \mathbb{R} \\ x &\rightarrow W_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \mu \right). \end{aligned}$$

Gracias a la desigualdad triangular para $W_p(\cdot, \cdot)$, tenemos que

$$|L_N(x) - L_N(y)| \leq W_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \frac{1}{N} \sum_{i=1}^N \delta_{y_i} \right).$$

Por la definición de $W_p(\cdot, \cdot)$, si consideramos el *coupling* $\pi = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}$, entonces tenemos que

$$|L_N(x) - L_N(y)| \leq W_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \frac{1}{N} \sum_{i=1}^N \delta_{y_i} \right) \leq \left(\frac{1}{N} \sum_{i=1}^N d^p(x_i, y_i) \right)^{1/p},$$

con lo que concluimos que L_N es una función $N^{-1/p}$ -Lipschitz para la métrica que estamos considerando, lo que agregado al lema 5.1.14 prueba el teorema 5.1.16.

Los teoremas que acabamos de enunciar los queremos utilizar para acotar $\mathbb{P}(W_p(\mu, \hat{\mu}^N) > \varepsilon)$ y $\mathbb{P}(Z_N^{\mathcal{G}} > \varepsilon)$. Para lograr esto necesitamos contar con estimaciones para $\mathbb{E}(W_p(\mu, \hat{\mu}^N))$ y para $\mathbb{E}(Z_N^{\mathcal{G}})$. Estas estimaciones las encontramos en las próximas dos secciones.

Estimación de $\mathbb{E}(W_2^2(\mu, \hat{\mu}^N))$

En esta sección se estudiará el caso particular para dimensión 1 del Teorema 10.2.1 que aparece en [37] (volumen 2). Dicho resultado establece una cota para la velocidad de convergencia a cero de $\mathbb{E}(W_2^2(\mu, \hat{\mu}^N))$. Con esta cota, utilizando el teorema 5.1.16, se deduce un resultado de concentración para la medida empírica con respecto a W_p , con $p \in [1, 2]$.

Antes de enunciar el teorema principal de esta sección se enuncian tres lemas que nos serán útiles en la demostración de dicho teorema. Comenzamos por una versión de la desigualdad de Carlson, establecida primero para series de números reales en [11], y luego extendida a funciones medibles. En [29] aparecen demostraciones alternativas y también generalizaciones.

Lema 5.1.17. [Carlson] *Sea g una función medible y no negativa en \mathbb{R} , entonces*

$$\int_{\mathbb{R}} g(x) dx \leq \sqrt{\pi \int_{\mathbb{R}} (1 + x^2) g(x)^2 dx}.$$

Demostración:

$$\begin{aligned}
 \left(\int_{\mathbb{R}} g(x) dx \right)^2 &= \left(\int_{\mathbb{R}} \frac{1}{\sqrt{1+x^2}} \sqrt{1+x^2} g(x) dx \right)^2 \\
 &\leq \int_{\mathbb{R}} \frac{dx}{1+x^2} \int_{\mathbb{R}} (1+x^2) g(x)^2 dx \\
 &= \int_{-\pi/2}^{\pi/2} du \int_{\mathbb{R}} (1+x^2) g(x)^2 dx \\
 &= \pi \int_{\mathbb{R}} (1+x^2) g(x)^2 dx \\
 &\quad \text{q.e.d.}
 \end{aligned}$$

Lema 5.1.18. Sean f, g densidades de probabilidad en \mathbb{R} , tales que

$$\int_{\mathbb{R}} x^2 (f(x) + g(x)) dx < \infty.$$

Consideremos μ y ν definidas por $d\mu(x) = f(x)dx$ y $d\nu(x) = g(x)dx$. Entonces

$$W_2^2(\mu, \nu) \leq 3 \int_{\mathbb{R}} x^2 |f(x) - g(x)| dx.$$

Demostración: La prueba de este lema consiste en proponer un *coupling* en particular y demostrar que dicho *coupling* satisface la estimación que se propone. Consideremos π el *coupling* definido para φ una función medible y acotada como

$$\begin{aligned}
 \int_{\mathbb{R}^2} \varphi(x, y) d\pi(x, y) &= \frac{1}{1-A} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x, y) (f - f \wedge g)(x) (g - f \wedge g)(y) dx dy \\
 &\quad + \int_{\mathbb{R}} \varphi(x, x) f \wedge g(x) dx,
 \end{aligned}$$

donde $A = \int_{\mathbb{R}} f \wedge g(x) dx$. Para ver que efectivamente $\pi \in \Pi_{\mu, \nu}$, basta con probar que para todo $B \subset \mathbb{R}$ medible se tiene que $\pi(B \times \mathbb{R}) = \mu(B)$ y que $\pi(\mathbb{R} \times B) = \nu(B)$. Para ello podemos considerar $\varphi(x, y) = \mathbb{1}_{B \times \mathbb{R}}(x, y) = \mathbb{1}_B(x)$ y entonces

$$\begin{aligned}
 \pi(B \times \mathbb{R}) &= \int_{\mathbb{R}^2} \mathbb{1}_B(x) d\pi(x, y) \\
 &= \frac{1}{1-A} \int_{\mathbb{R}} \mathbb{1}_B(x) (f - f \wedge g)(x) dx \int_{\mathbb{R}} (g - f \wedge g)(y) dy \\
 &\quad + \int_{\mathbb{R}} \mathbb{1}_B(x) f \wedge g(x) dx \\
 &= \int_{\mathbb{R}} \mathbb{1}_B(x) (f - f \wedge g)(x) dx + \int_{\mathbb{R}} \mathbb{1}_B(x) f \wedge g(x) dx \\
 &= \mu(B).
 \end{aligned}$$

5.1. Concentración de la Medida Empírica con respecto a W

De manera similar se prueba que $\pi(\mathbb{R} \times B) = \nu(B)$. Ahora que ya sabemos que $\pi \in \Pi_{\mu, \nu}$, tenemos que

$$W_2^2(\mu, \nu) \leq \int_{\mathbb{R}^2} (x - y)^2 d\pi(x, y).$$

Ahora basta con probar que el lado derecho de la última desigualdad satisface la cota deseada. En efecto

$$\begin{aligned} \int_{\mathbb{R}^2} (x - y)^2 d\pi(x, y) &= \frac{1}{1 - A} \int_{\mathbb{R}} \int_{\mathbb{R}} (x - y)^2 (f - f \wedge g)(x) (g - f \wedge g)(y) dx dy \\ &= \frac{1}{1 - A} \int_{\mathbb{R}} \int_{\mathbb{R}} (x^2 - 2xy + y^2) (f - f \wedge g)(x) (g - f \wedge g)(y) dx dy \\ &= \frac{1}{1 - A} \int_{\mathbb{R}} (g - f \wedge g)(y) dy \int_{\mathbb{R}} x^2 (f - f \wedge g)(x) dx \\ &\quad + \frac{1}{1 - A} \int_{\mathbb{R}} (f - f \wedge g)(x) dx \int_{\mathbb{R}} y^2 (g - f \wedge g)(y) dy \\ &\quad - \frac{2}{1 - A} \int_{\mathbb{R}} y (g - f \wedge g)(y) dy \int_{\mathbb{R}} x (f - f \wedge g)(x) dx \\ &= \int_{\mathbb{R}} x^2 (f - f \wedge g)(x) dx + \int_{\mathbb{R}} y^2 (g - f \wedge g)(y) dy \\ &\quad - \frac{2}{1 - A} \int_{\mathbb{R}} y (g - f \wedge g)(y) dy \int_{\mathbb{R}} x (f - f \wedge g)(x) dx. \end{aligned}$$

Ahora recordamos que $|x - y| = x + y - 2x \wedge y$, donde $x \wedge y$ es el mínimo entre x e y . Con esto tenemos que

$$\begin{aligned} \int_{\mathbb{R}^2} (x - y)^2 d\pi(x, y) &= \int_{\mathbb{R}} x^2 |f - g|(x) dx \\ &\quad - \frac{2}{1 - A} \int_{\mathbb{R}} y (g - f \wedge g)(y) dy \int_{\mathbb{R}} x (f - f \wedge g)(x) dx. \end{aligned}$$

Por otro lado, gracias a la desigualdad de Hölder

$$\begin{aligned} \left| \int_{\mathbb{R}} x (f - f \wedge g)(x) dx \right| &\leq \left(\int_{\mathbb{R}} x^2 |f - f \wedge g|(x) dx \right)^{1/2} \left(\int_{\mathbb{R}} |f - f \wedge g|(x) dx \right)^{1/2} \\ &\leq \left(\int_{\mathbb{R}} x^2 |f - g|(x) dx \right)^{1/2} (1 - A)^{1/2}, \end{aligned}$$

5.1. Concentración de la Medida Empírica con respecto a W

con lo que finalmente

$$\begin{aligned}
 \int_{\mathbb{R}^2} (x-y)^2 d\pi(x,y) &\leq \int_{\mathbb{R}} x^2 |f-g|(x) dx \\
 &+ \frac{2}{1-A} \left| \int_{\mathbb{R}} y(g-f \wedge g)(y) dy \int_{\mathbb{R}} x(f-f \wedge g)(x) dx \right| \\
 &\leq \int_{\mathbb{R}} x^2 |f-g|(x) dx + 2 \int_{\mathbb{R}} x^2 |f-g|(x) dx \\
 &= 3 \int_{\mathbb{R}} x^2 |f-g|(x) dx.
 \end{aligned}$$

q.e.d.

Llamemos Φ_σ a la medida de probabilidad de una variable aleatoria normal de media cero y varianza σ^2 y sea ϕ_σ su densidad. Sea μ una medida de probabilidad en \mathbb{R} , definimos $\mu^\sigma = \Phi * \mu$ la convolución de ϕ con μ . Notemos que μ^σ tiene densidad dada por $\phi_\sigma * \mu$.

Lema 5.1.19. *Si $\mu \in \mathcal{P}_2(\mathbb{R})$ entonces $W_2^2(\mu^\sigma, \mu) \leq \sigma^2$.*

Demostración: Sean X e Y dos v.a. con leyes μ y Φ_σ respectivamente. Entonces el par $(X, X+Y)$ es un *coupling* de μ con μ^σ . Luego, tenemos que

$$W_2^2(\mu, \mu^\sigma) \leq \mathbb{E}((X+Y-X)^2) = \mathbb{E}(Y^2) = \sigma.$$

q.e.d.

Teorema 5.1.20. *Supongamos que μ tiene momento de orden 6 finito, es decir*

$$M_6 := \int_{\mathbb{R}} x^6 d\mu(x) < \infty.$$

Entonces, para todo $N \in \mathbb{N}$,

$$\mathbb{E}(W_2^2(\hat{\mu}^N, \mu)) \leq \frac{D}{N^{2/5}},$$

donde

$$D = 6 \left(\sqrt{4\pi} (61 + 32M_6) \right)^{1/2}.$$

Demostración: Consideremos X_1, \dots, X_N una muestra i.i.d. de ley μ y llamemos $\hat{\mu}^N$ a la medida empírica. Notemos que gracias a la desigualdad triangular y al lema 5.1.19, tenemos que

$$\begin{aligned}
 W_2^2(\mu, \hat{\mu}^N) &\leq 2 (W_2^2(\hat{\mu}^N, (\hat{\mu}^N)^\sigma) + W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma) + W_2^2(\mu^\sigma, \mu^\sigma)) \\
 &\leq 2 (\sigma^2 + W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma) + \sigma^2) \\
 &\leq 2 (2\sigma^2 + W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma)),
 \end{aligned}$$

5.1. Concentración de la Medida Empírica con respecto a W

de donde tenemos

$$\mathbb{E}(W_2^2(\mu, \hat{\mu}^N)) \leq 4\sigma^2 + 2\mathbb{E}(W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma)). \quad (5.1.15)$$

Por otro lado, gracias al lema 5.1.18 se tiene que:

$$W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma) \leq 3 \int_{\mathbb{R}} x^2 |g^\sigma(x) - g_N^\sigma(x)| dx,$$

donde hemos llamado g^σ a la densidad de μ^σ y g_N^σ a la densidad de $(\hat{\mu}^N)^\sigma$. Ahora podemos aplicar el lema de Carlson a la función $g = x^2 |g^\sigma(x) - g_N^\sigma(x)|$ y obtenemos

$$W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma) \leq 3 \sqrt{\pi \int_{\mathbb{R}} (1+x^2)x^4 (g^\sigma(x) - g_N^\sigma(x))^2 dx} \quad (5.1.16)$$

$$\leq 3 \sqrt{2\pi \int_{\mathbb{R}} (1+x^6)(g^\sigma(x) - g_N^\sigma(x))^2 dx}. \quad (5.1.17)$$

Notemos que g_N^σ es una variable aleatoria en el espacio de las densidades en \mathbb{R} . En efecto

$$g_N^\sigma(x) = \frac{1}{N} \sum_{i=1}^N \phi_\sigma(x - X_i).$$

Además, se tiene que

$$\mathbb{E}(g_N^\sigma(x)) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(\phi_\sigma(x - X_i)) = \mathbb{E}(\phi_\sigma(x - X_1)) = \int_{\mathbb{R}} \phi_\sigma(x - y) d\mu(y) = g^\sigma(x),$$

y también

$$\mathbb{E}((g_N^\sigma(x) - g^\sigma(x))^2) = \text{Var}(g_N^\sigma(x)) = \frac{1}{N} \sum_{i=1}^N \text{Var}(\phi_\sigma(x - X_i)) = \frac{\text{Var}(\phi_\sigma(x - X_1))}{N}.$$

Pero, por otro lado

$$\text{Var}(\phi_\sigma(x - X_1)) \leq \mathbb{E}(\phi_\sigma(x - X_1)^2) = \int_{\mathbb{R}} \phi_\sigma^2(x - y) d\mu(y),$$

y entonces

$$\mathbb{E}((g_N^\sigma(x) - g^\sigma(x))^2) \leq \frac{1}{N} \int_{\mathbb{R}} \phi_\sigma^2(x - y) d\mu(y).$$

Luego si calculamos esperanzas en (5.1.17), aplicamos la desigualdad de Jensen y el teorema de Tonelli para intercambiar la esperanza y la integral, tenemos que

$$\mathbb{E}(W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma)) \leq 3 \sqrt{2\pi \int_{\mathbb{R}} (1+x^6) \mathbb{E}((g^\sigma(x) - g_N^\sigma(x))^2) dx}.$$

5.1. Concentración de la Medida Empírica con respecto a W

Incorporando los cálculos anteriores, esto se transforma en

$$\mathbb{E}(W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma)) \leq 3\sqrt{2\pi \int_{\mathbb{R}} (1+x^6) \frac{1}{N} \int_{\mathbb{R}} \phi_\sigma^2(x-y) d\mu(y) dx}.$$

Ahora notemos que

$$\begin{aligned} \phi_\sigma^2(x-y) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \right)^2 \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \left(\frac{1}{\sqrt{2\pi(\sigma/\sqrt{2})^2}} \exp\left(-\frac{(x-y)^2}{2(\sigma/\sqrt{2})^2}\right) \right) \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \phi_{\sigma/\sqrt{2}}(x-y), \end{aligned}$$

con lo que tenemos

$$\int_{\mathbb{R}} (1+x^6) \int_{\mathbb{R}} \phi_\sigma^2(x-y) d\mu(y) dx = \frac{1}{\sqrt{4\pi\sigma^2}} \int_{\mathbb{R}} (1+x^6) \int_{\mathbb{R}} \phi_{\sigma/\sqrt{2}}(x-y) d\mu(y) dx.$$

Pero

$$\begin{aligned} \int_{\mathbb{R}} (1+x^6) \int_{\mathbb{R}} \phi_{\sigma/\sqrt{2}}(x-y) d\mu(y) dx &= 1 + \int_{\mathbb{R}} \int_{\mathbb{R}} (x+y)^6 \phi_{\sigma/\sqrt{2}}(x) dx d\mu(y) \\ &\leq 1 + 2^5 \int_{\mathbb{R}} \int_{\mathbb{R}} (x^6 + y^6) \phi_{\sigma/\sqrt{2}}(x) dx d\mu(y) \\ &= 1 + 32 \left(\int_{\mathbb{R}} x^6 \phi_{\sigma/\sqrt{2}}(x) dx + \int_{\mathbb{R}} y^6 d\mu(y) \right) \\ &= 1 + 32 \left(\frac{15}{8} \sigma^6 + M_6 \right). \end{aligned}$$

Fijando $\sigma < 1$ se tiene que

$$\mathbb{E}(W_2^2(\mu^\sigma, (\hat{\mu}^N)^\sigma)) \leq 3\sqrt{\frac{2\pi}{N} \frac{1}{\sqrt{4\pi\sigma^2}} (61 + 32M_6)} = 3(\sqrt{\pi}(61 + 32M_6))^{1/2} \frac{1}{\sqrt{N\sigma}}.$$

Incorporando la desigualdad anterior a (5.1.15), concluimos que

$$\mathbb{E}(W_2^2(\mu, \hat{\mu}^N)) \leq 4\sigma^2 + 6(\sqrt{\pi}(61 + 32M_6))^{1/2} \frac{1}{\sqrt{N\sigma}}.$$

En particular para $\sigma = 1/N^{1/5}$ tenemos

$$\mathbb{E}(W_2^2(\mu, \hat{\mu}^N)) \leq 4\frac{1}{N^{2/5}} + 6(\sqrt{\pi}(61 + 32M_6))^{1/2} \frac{1}{\sqrt{N^{4/5}}} \leq \frac{D}{N^{2/5}},$$

donde

$$D = 6(\sqrt{\pi}(61 + 32M_6))^{1/2}.$$

q.e.d.

Directamente del teorema anterior y del teorema 5.1.16, se puede obtener el siguiente resultado, que corresponde al Corolario 8 de [23], aunque en este artículo no se da el valor explícito de la constante D .

Corolario 5.1.21. *Sea μ una medida de probabilidad en \mathbb{R} que satisface $T_p(\lambda)$ para algún $p \in [1, 2]$. Entonces, para todo $p' \in [1, p]$:*

$$\mathbb{P}(W_{p'}(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \exp\left(-\frac{N\lambda}{2} \left(\varepsilon - \frac{D^{1/2}}{N^{1/5}}\right)^2\right)$$

para todo $\varepsilon > 0$, $N \geq D^{5/2}/\varepsilon^5$ y

$$D = 6 \left(\sqrt{\pi} (61 + 32M_6)\right)^{1/2}.$$

Demostración: Consideremos μ una medida de probabilidad en \mathbb{R} que satisface $T_p(\lambda)$ para $p \in [1, 2]$. Sea $p' \in [1, p]$, por la desigualdad que satisfacen las distancias de Wasserstein, tenemos las siguientes desigualdades

$$\mathbb{P}(W_{p'}(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \quad (5.1.18)$$

$$\mathbb{E}(W_{p'}(\mu, \hat{\mu}^N)) \leq \mathbb{E}(W_2(\mu, \hat{\mu}^N)). \quad (5.1.19)$$

Además, gracias a la desigualdad de Jensen y al teorema anterior, tenemos que

$$\mathbb{E}\left(\sqrt{W_2^2(\hat{\mu}^N, \mu)}\right) \leq \sqrt{\mathbb{E}(W_2^2(\mu, \hat{\mu}^N))} \leq \sqrt{\frac{D}{N^{2/5}}} = \frac{D^{1/2}}{N^{1/5}}.$$

Luego, gracias a (5.1.19) tenemos que

$$\mathbb{E}(W_p(\mu, \hat{\mu}^N)) \leq \frac{D^{1/2}}{N^{1/5}}. \quad (5.1.20)$$

Como μ satisface $T_p(\lambda)$, gracias al teorema 5.1.16, tenemos que

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \exp\left(-\frac{\lambda N}{2} (\varepsilon - \mathbb{E}[W_p(\mu, \hat{\mu}^N)])^2\right), \quad \forall \varepsilon \geq \mathbb{E}(W_p(\mu, \hat{\mu}^N)).$$

Notemos que para $\varepsilon \geq D^{1/2}/N^{1/5}$, gracias a (5.1.20), tenemos que

$$\exp\left(-\frac{\lambda N}{2} (\varepsilon - \mathbb{E}[W_p(\mu, \hat{\mu}^N)])^2\right) \leq \exp\left(-\frac{\lambda N}{2} \left(\varepsilon - \frac{D^{1/2}}{N^{1/5}}\right)^2\right).$$

Luego, para $\varepsilon \geq D^{1/2}/N^{1/5}$, tenemos que

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \exp\left(-\frac{\lambda N}{2} \left(\varepsilon - \frac{D^{1/2}}{N^{1/5}}\right)^2\right).$$

Introduciendo esta última desigualdad en (5.1.19) se concluye el resultado.

q.e.d.

Una pregunta bastante natural es que dado que tenemos un control para la esperanza de $W_2(\mu, \hat{\mu}^N)$ podríamos obtener una cota para $\mathbb{P}(W_2(\mu, \hat{\mu}^N) > \varepsilon)$ simplemente aplicando la desigualdad de Markov. Eso lo hacemos a continuación y después mostramos que el resultado obtenido por Gozlan y Léonard es mucho mejor.

Corolario 5.1.22. *Para todo $p \in [1, 2]$ y todo $N \geq 1$ se tiene que:*

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \frac{D^{1/2}}{\varepsilon N^{1/5}}.$$

Demostración: Como vimos en la demostración del corolario anterior, para $p \in [1, 2]$ se tiene que

$$\mathbb{E}(W_2(\hat{\mu}^N, \mu)) \leq \frac{D^{1/2}}{N^{1/5}}$$

y

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \mathbb{P}(W_2(\mu, \hat{\mu}^N) \geq \varepsilon).$$

Aplicando la desigualdad de Markov, concluimos que

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \mathbb{P}(W_2(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \frac{D^{1/2}}{\varepsilon N^{1/5}}.$$

q.e.d.

Observación 5.1.23: Notemos que para ambos corolarios, el número crítico de observaciones necesarias para que la cota que proponen sea válida o significativa, está dado por $D^{5/2}/\varepsilon^5$. Sin embargo, como lo muestra el siguiente ejemplo, la cota dada por la desigualdad de Markov decae a cero mucho más lento.

Ejemplo 5.1.24: Consideremos μ igual a la ley normal estándar $\mathcal{N}(0, 1)$. En este caso es conocido que μ satisface $T_2(1)$ y que $M_6 = 15$. En el gráfico (5.1.1) se muestra el comportamiento de las cotas dadas por los corolarios previos para $\varepsilon = 0,1$. Se observa claramente como la cota de Gozlan y Léonard decae mucho más rápido a cero.

Considerando $p' = 1$, en el resultado de Gozlan y Léonard, podemos comparar para distintos valores de ε los resultados que nos da este corolario con los resultados obtenidos con el teorema de Bolley et al., en su versión modificada que se mostró anteriormente. En la tabla 5.1.6 aparece dicha comparación en la que se aprecia que el resultado de Bolley et al., funciona mejor en la práctica. Sin embargo, a favor del resultado de Gozlan y Léonard podemos argumentar que en la medida de que se cuente con mejores estimaciones para $\mathbb{E}(W_2(\mu, \hat{\mu}^N))$, la cota propuesta por estos autores puede mejorar tanto como mejore la estimación de esta última cantidad.

5.1. Concentración de la Medida Empírica con respecto a W

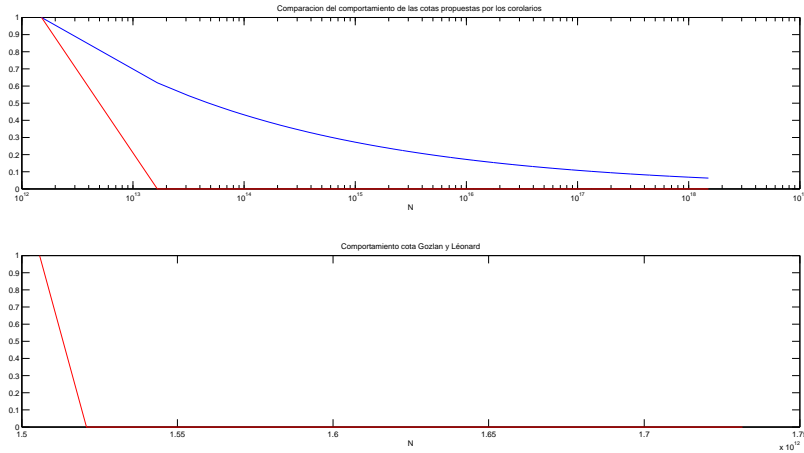


Figura 5.1.1: En rojo la cota dada por Gozlan y Léonard, en azul la cota dada por la desigualdad de Markov. En el gráfico superior se utilizó escala logarítmica en el eje de las x .

ε	N_0	
	Cota Bolley et al.	Cota Gozlan y Léonard
0,01	$8,922e + 09$	$4,705e + 15$
0,05	$6,151e + 07$	$1,506e + 12$
0,1	$7,161e + 06$	$4,705e + 10$
0,5	$4,749e + 04$	$1,506e + 7$
1	$5,412e + 03$	$4,705e + 5$

Cuadro 5.1.6: Comparación de los distintos N_0 's obtenidos. Podemos ver la versión que se presentó del resultado de Bolley et al. es la mejor alternativa para todo valor de ε .

Estimación de $\mathbb{E}(Z_N^{\mathcal{G}})$

Recordemos que para (\mathcal{X}, d) un espacio métrico polaco y \mathcal{G} un conjunto de funciones 1-Lipschitz, μ una medida de probabilidad en \mathcal{X} y $\{X_1, \dots, X_N\}$ una muestra i.i.d. con ley μ . El proceso empírico asociado a μ corresponde a

$$Z_N^{\mathcal{G}} = \sup_{\phi \in \mathcal{G}} \left\{ \left| \frac{1}{N} \sum_{i=1}^N \phi(X_i) - \int_{\mathcal{X}} \phi d\mu \right| \right\}.$$

En esta sección consideraremos un proceso empírico en particular, el cual corresponde al caso $\mathcal{G} = \{f \in L_b(\mathcal{X}, d) : \|f\|_{L_b} \leq 1\}$, donde L_b corresponde al conjunto de funciones

5.1. Concentración de la Medida Empírica con respecto a W

Lipschitz acotadas definido en el capítulo 3. Notemos que con esta elección tenemos

$$Z_N^G = BL(\mu, \hat{\mu}^N).$$

En [17], el autor nos propone el siguiente teorema para acotar la esperanza de $\mathbb{E}(BL(\mu, \hat{\mu}^N))$.

Teorema 5.1.25. *Sea (\mathcal{X}, d) un espacio métrico separable. Llamamos $N(\mu, \varepsilon, \delta)$ al mínimo número de conjuntos de diámetro menor o igual a 2ε necesarios para cubrir \mathcal{X} salvo un conjunto de medida δ . Suponga que para algún real $k > 2$ existe una constante K de modo que*

$$N(\mu, \varepsilon, \varepsilon^{k/(k-2)}) \leq K\varepsilon^{-k}, \quad (5.1.21)$$

cuando $0 < \varepsilon \leq 1$. Entonces para todo $N \in \mathbb{N}$

$$\mathbb{E}(BL(\mu, \hat{\mu}^N)) \leq \mathcal{M}(k, K)N^{-1/k},$$

donde

$$\mathcal{M}(k, K) = 1 + \frac{2}{9^{k/(k+2)} - 3^{k/(k+2)}} + 27K^{1/2} \left(1 + \frac{81}{3^{(k-2)/2} - 1} \right).$$

Gracias a este teorema y al teorema 5.1.15, podemos encontrar el siguiente corolario.

Corolario 5.1.26. *Sea μ una medida de probabilidad en $\mathcal{P}_1(\mathcal{X})$ que satisface $T_1(\lambda)$. Entonces, para todo $\varepsilon > 0$*

$$\mathbb{P}(BL(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \exp \left(-\frac{N\lambda}{2} \left(\varepsilon - \frac{\mathcal{M}(k, K)}{N^{1/k}} \right)^2 \right)$$

para todo $\varepsilon > 0$, $N \geq \mathcal{M}(k, K)^k / \varepsilon^k$ y $\mathcal{M}(k, K)$ es la constante del teorema 5.1.25.

Para poder utilizar este teorema, nos hace falta un criterio que nos permita obtener los valores de k y K en (5.1.21) de modo de poder calcular $\mathcal{M}(k, K)$. A continuación presentamos un criterio dado por Dudley en [17].

Proposición 5.1.27. *Consideremos $\mathcal{X} = \mathbb{R}$ y sea μ una medida de probabilidad en \mathbb{R} y $k > 2$. Supongamos que para $\beta = k/(k-1)(k-2)$ se tiene*

$$M_\beta := \int_{\mathbb{R}} |x|^\beta d\mu(x) < \infty,$$

entonces

$$N(\mu, \varepsilon, \varepsilon^{k/(k-2)}) \leq \frac{M_\beta^{1/\beta}}{\varepsilon^k}.$$

5.1. Concentración de la Medida Empírica con respecto a W

Demostración: Sea $\varepsilon \in (0, 1]$ y $\gamma > 1$. Escojamos $r > 0$ tal que

$$\mu([-r, r]^c) \leq \varepsilon^{k/(k-2)} \leq \mu([-r/\gamma, r/\gamma]^c),$$

lo que podemos lograr fijando

$$r = \sup\{y : \mu([-y/\gamma, y/\gamma]^c) \geq \varepsilon^{k/(k-2)}\}.$$

Entonces

$$M_\beta = \int_{\mathbb{R}} |x|^\beta d\mu(x) \geq \int_{[-r/\gamma, r/\gamma]^c} |x|^\beta d\mu(x) \geq \int_{[-r/\gamma, r/\gamma]^c} \left(\frac{r}{\gamma}\right)^\beta d\mu(x) \geq \left(\frac{r}{\gamma}\right)^\beta \varepsilon^{k/(k-2)}.$$

Por otro lado el intervalo $[-r, r]$ lo podemos recubrir con r/ε bolas cerradas de radio ε . Luego

$$N(\mu, \varepsilon, \varepsilon^{k/(k-2)}) \leq \frac{r}{\varepsilon} \leq \frac{\gamma M_\beta^{1/\beta} / \varepsilon^{k/\beta(k-2)}}{\varepsilon} = \frac{\gamma M_\beta^{1/\beta}}{\varepsilon^k}.$$

Para terminar basta notar que el lado izquierdo de la última desigualdad no depende de γ , por lo que podemos tomar $\gamma \searrow 1$ y concluir.

q.e.d.

Observación 5.1.28: Es útil conocer k en función de β para la última proposición. Ya que en la mayoría de los casos lo que conocemos a priori son los momentos de μ . Es fácil probar que

$$k(\beta) = \frac{3 + 1/\beta + \sqrt{1 + 6\beta + 1/\beta^2}}{2}.$$

Ejemplo 5.1.29: Consideremos una vez más μ , la ley normal estándar en \mathbb{R} . En este caso M_β es finito para todo β , luego podemos buscar aquel β que minimice $\mathcal{M}(k(\beta), M_\beta^{1/\beta})^{k(\beta)}$. Heurísticamente, $\beta = 6$ parece un buen candidato. Con el encontramos que

$$k(\beta) = 2,2953,$$

$$\mathcal{M}(k(\beta), M_\beta^{1/\beta}) = 1,5597e + 04.$$

En la tabla 5.1.7 aparece N_0 , el valor mínimo de N para que el corolario 5.1.26 sea válido. Recordemos que $BL(\mu, \hat{\mu}^N) \leq W(\mu, \hat{\mu}^N)$, luego tenemos

$$\mathbb{P}(BL(\mu, \hat{\mu}^N) > \varepsilon) \leq \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon),$$

con lo que se tiene que los resultados que encontramos antes para concentración de la medida empírica con respecto a la distancia de Wasserstein son también válidos para la distancia BL . En la tabla 5.1.7 también aparece el mínimo N necesario para que fuera válido el teorema 5.1.2, de modo de poder comparar que resultado es más eficiente. Nuevamente encontramos que el teorema de Bolley et al. es mucho mejor que el corolario encontrado con el resultado Gozlan y Léonard y la cota para $\mathbb{E}(BL(\mu, \hat{\mu}^N))$ dada por Dudley. Aunque, si se contara con una cota más fina que la de Dudley probablemente se podría mejorar el resultado del corolario 5.1.26.

ε	N_0	
	Cota Gozlan y Léonard	Cota Bolley et al.
0,01	$1,641e + 14$	$8,922e + 09$
0,05	$4,080e + 12$	$6,151e + 07$
0,1	$8,313e + 11$	$7,161e + 06$
0,5	$2,067e + 10$	$4,749e + 04$
1	$4,211e + 09$	$5,412e + 03$

Cuadro 5.1.7: Comparación del comportamiento N_0 para distintos valores de ε .

5.2. Concentración de la medida empírica con respecto a la distancia de Lévy-Prokhorov

En esta sección estudiaremos un resultado de concentración de la medida empírica con respecto a la distancia de Prokhorov propuesto por Dudley en [17].

En lo que sigue (\mathcal{X}, d) es un espacio métrico separable, μ una medida de probabilidad en \mathcal{X} y $\hat{\mu}^N$ la medida empírica asociada a una muestra i.i.d. de μ de tamaño N .

Proposición 5.2.1. *Sea T un subconjunto medible de \mathcal{X} y sea $\{S_j\}_{j=1}^m$ una partición medible de T , entonces:*

$$\mathbb{E} \left(\sum_{j=1}^m (\mu - \hat{\mu}^N)(S_j)^2 \right) \leq \frac{\mu(T)}{N}, \quad (5.2.1)$$

$$\mathbb{E} \left(\sum_{j=1}^m |(\mu - \hat{\mu}^N)(S_j)| \right) \leq \left(\frac{m\mu(T)}{N} \right)^{1/2}. \quad (5.2.2)$$

Demostración: Empecemos por notar que para A , un subconjunto medible de \mathcal{X} , se tiene que

$$\begin{aligned} \mathbb{E}(\hat{\mu}^N(A)) &= \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_i}(A) \right) \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{P}(X_i \in A) \\ &= \frac{1}{N} \sum_{i=1}^N \mu(A) \\ &= \mu(A). \end{aligned}$$

Así también

$$\begin{aligned}
 \mathbb{E}(\hat{\mu}^N(A)^2) &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N\delta_{X_i}(A)\right)^2 \\
 &= \mathbb{E}\left(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\delta_{X_i}(A)\delta_{X_j}(A)\right) \\
 &= \frac{1}{N^2}\mathbb{E}\left(\sum_{i\neq j}\delta_{X_i}(A)\delta_{X_j}(A) + \sum_{i=1}^N\delta_{X_i}(A)\right) \\
 &= \frac{1}{N^2}\sum_{i\neq j}\mathbb{P}(X_i\in A)\mathbb{P}(X_j\in A) + \frac{1}{N^2}\sum_{i=1}^N\mathbb{P}(X_i\in A) \\
 &= \frac{1}{N^2}\sum_{i\neq j}\mu(A)^2 + \frac{1}{N^2}\sum_{i=1}^N\mu(A) \\
 &= \frac{(N-1)\mu(A)^2}{N} + \frac{\mu(A)}{N},
 \end{aligned}$$

de donde concluimos

$$\mathbb{E}((\mu - \hat{\mu}^N)(A)^2) = \mathbb{E}(\hat{\mu}^N(A) - \mathbb{E}(\hat{\mu}^N(A)))^2 = \frac{\mu(A) - \mu(A)^2}{N}.$$

Entonces

$$\begin{aligned}
 \mathbb{E}\left(\sum_{j=1}^m(\mu - \hat{\mu}^N)(S_j)^2\right) &= \sum_{j=1}^m\mathbb{E}((\mu - \hat{\mu}^N)(S_j)^2) \\
 &= \sum_{j=1}^m\frac{\mu(S_j) - \mu(S_j)^2}{N} \\
 &= \frac{\mu(T) - \sum_j\mu(S_j)^2}{N} \\
 &\leq \frac{\mu(T)}{N}.
 \end{aligned}$$

Luego, gracias a la desigualdad de Cauchy también tenemos

$$\begin{aligned}
 \mathbb{E}\left(\sum_{j=1}^m|(\mu - \hat{\mu}^N)(S_j)|\right) &= \sum_{j=1}^m\mathbb{E}(|(\mu - \hat{\mu}^N)(S_j)|) \\
 &\leq \left(\sum_{j=1}^m1^2\right)^{1/2}\left(\sum_{j=1}^m\mathbb{E}((\mu - \hat{\mu}^N)(S_j)^2)\right)^{1/2} \\
 &\leq \left(\frac{m\mu(T)}{N}\right)^{1/2}.
 \end{aligned}$$

5.2. Concentración de la medida empírica con respecto a d_P

q.e.d.

Dado un $\delta > 0$, notaremos como $N(\mu, \delta)$ al mínimo número de conjuntos de diámetro menor o igual a 2δ que son necesarios para recubrir a S salvo por un conjunto de μ -medida menor o igual que δ . Por ejemplo, si consideramos $S = \mathbb{R}$ y μ la ley de una normal estándar $\mathcal{N}(0, 1)$, tenemos que $N(\mu, 0,05) \leq 40$, pues con 40 intervalos de largo 0,1 se cubre el intervalo $[-2, 2]$ y el conjunto $\mu([-2, 2]^c) \leq 0,05$. En lo que sigue denotaremos

$$k(\mu) := \limsup_{\delta \searrow 0} \frac{\log N(\mu, \delta)}{\log(1/\delta)}.$$

Teorema 5.2.2. *Para todo $\varepsilon > 0$ existe un $N_0 \in \mathbb{N}$ tal que $\forall N \geq N_0$*

$$\mathbb{E}(d_P(\mu, \hat{\mu}^N)) \leq 3N^{-1/(k(\mu)+2+\varepsilon)}.$$

Demostración: Empecemos por notar que

$$k(\mu) = \limsup_{\delta \searrow 0} \frac{\log N(\mu, \delta)}{\log(1/\delta)} = \limsup_{\delta \searrow 0} \frac{\log N(\mu, \eta)}{\log(1/\eta)},$$

luego para todo $\varepsilon > 0$ tenemos que existe un $\delta_0 > 0$ de modo que

$$k(\mu) + \varepsilon \geq \frac{\log N(\mu, \delta)}{\log(1/\delta)} \quad \forall \delta \in (0, \delta_0).$$

Es decir, para todo $\varepsilon > 0$ existe un δ suficientemente pequeño tal que

$$N(\mu, \delta) \leq \left\lceil \frac{1}{\delta^{k(\mu)+\varepsilon}} \right\rceil =: N_0.$$

Consideremos $N = \lceil 1/\delta^{k(\mu)+2+\varepsilon} \rceil$ y notemos que existe una colección de conjuntos medibles A_1, \dots, A_{N_0} con $\text{diam}(A_i) \leq 2\delta$ y un conjunto A_0 con $\mu(A_0) \leq \delta$ tales que:

$$\mathcal{X} = \bigcup_{j=0}^{N_0} A_j,$$

y notemos que considerando $B_0 = A_0 \setminus \cup_j A_j$, $B_1 = A_1$ y $B_{n+1} = A_{n+1} \setminus \cup_{j=1}^n A_j$, tenemos que la colección $\{B_j\}_{j=1}^{N_0}$ es una partición y satisface

$$\text{diam}(B_j) \leq 2\delta, \quad j = 1, \dots, N_0, \quad \mu(B_0) \leq \delta.$$

5.2. Concentración de la medida empírica con respecto a d_P

Sea $F \subset \mathcal{X}$ un cerrado y sea $J = \{j \geq 1 : B_j \cap F \neq \emptyset\}$ entonces

$$\begin{aligned}
 \hat{\mu}^N(F) &\leq \hat{\mu}^N\left(\bigcup_{j \in J} B_j\right) + \hat{\mu}^N(B_0) \\
 &\leq \hat{\mu}^N\left(\bigcup_{j \in J} B_j\right) + \hat{\mu}^N(B_0) + \mu\left(\bigcup_{j \in J} B_j\right) - \mu\left(\bigcup_{j \in J} B_j\right) + \delta - \mu(B_0) \\
 &\leq \mu\left(\bigcup_{j \in J} B_j\right) + \delta + \sum_{j=0}^{N_0} |(\hat{\mu}^N - \mu)(B_j)| \\
 &\leq \mu(F^{2\delta}) + \delta + \sum_{j=0}^{N_0} |(\hat{\mu}^N - \mu)(B_j)| \\
 &\leq \mu(F^{\delta^*}) + \delta^*,
 \end{aligned}$$

con $\delta^* = 2\delta + \sum_{j=0}^{N_0} |(\hat{\mu}^N - \mu)(B_j)|$. Entonces $d_P(\mu, \hat{\mu}^N) \leq \delta^*$.

Finalmente, gracias a la proposición 5.2.1 tenemos

$$\mathbb{E}(d_P(\mu, \hat{\mu}^N)) \leq 2\delta + \left(\frac{N_0}{N}\right)^{1/2} = 3\delta \leq 3N^{-1/(k(\mu)+2+\varepsilon)}.$$

q.e.d.

Observación 5.2.3: Notemos que N no puede ser muy pequeño, ya que $N = \lfloor 1/\delta^{k(\mu)+2+\varepsilon} \rfloor$ y δ está acotado superiormente.

Corolario 5.2.4. Si para $k, K \in \mathbb{R}_+$ tenemos que $N(\mu, \delta) \leq K\delta^{-k}$ para $\delta \in (0, 1]$, entonces para todo $N \in \mathbb{N}$

$$\mathbb{E}(d_P(\mu, \hat{\mu}^N)) \leq \frac{2 + K^{1/2}}{N^{1/(k+2)}}.$$

Demostración: Consideremos $\delta = 1/N^{1/(k+2)}$ y notemos que $\delta \in (0, 1]$ para todo N . Entonces, por hipótesis se tiene

$$N(\mu, \delta) \leq \frac{K}{\delta^k} = KN^{k/(k+2)}.$$

Siguiendo el mismo razonamiento que antes podemos llegar a que

$$\mathbb{E}(d_P(\mu, \hat{\mu}^N)) \leq \frac{2}{N^{1/(k+2)}} + \left(\frac{KN^{k/(k+2)}}{N}\right)^{1/2} = \frac{2 + K^{1/2}}{N^{1/(k+2)}}.$$

q.e.d.

Analicemos ahora el caso $\mathcal{X} = \mathbb{R}$. Empecemos por establecer condiciones para la existencia de una cota como la que aparece en la hipótesis del corolario anterior.

5.2. Concentración de la medida empírica con respecto a d_P

Proposición 5.2.5. *Supongamos que $\mathcal{X} = \mathbb{R}$, entonces:*

$$M_\beta = \int_{\mathbb{R}} |x|^\beta d\mu(x) < \infty \Rightarrow N(\mu, \varepsilon) \leq \frac{M_\beta^{1/\beta}}{\varepsilon^{1+1/\beta}}.$$

Demostración: Sea $\varepsilon \in (0, 1]$, $\gamma > 1$ y $r > 0$ tal que

$$\mu([-r, r]^c) \leq \varepsilon \leq \mu([-r/\gamma, r/\gamma]^c),$$

para lo cual basta fijar

$$r = \sup\{y : \mu([-y/\gamma, y/\gamma]^c) \geq \varepsilon\}.$$

Entonces

$$M_\beta = \int_{\mathbb{R}} |x|^\beta d\mu(x) \geq \int_{[-r/\gamma, r/\gamma]^c} |x|^\beta d\mu(x) \geq \int_{[-r/\gamma, r/\gamma]^c} \left(\frac{r}{\gamma}\right)^\beta d\mu(x) \geq \left(\frac{r}{\gamma}\right)^\beta \varepsilon.$$

Por otro lado, notemos que el intervalo $[-r, r]$ lo podemos recubrir con r/ε bolas cerradas de radio ε . Luego

$$N(\mu, \varepsilon) \leq \frac{r}{\varepsilon} \leq \frac{\gamma M_\beta^{1/\beta} / \varepsilon^{1/\beta}}{\varepsilon} = \frac{\gamma M_\beta^{1/\beta}}{\varepsilon^{1+1/\beta}}.$$

Para terminar basta notar que el lado izquierdo de la última desigualdad no depende de γ , por lo que podemos tomar $\gamma \searrow 1$ y concluir.

q.e.d.

Del capítulo 3 sabemos que si $\mathcal{X} = \mathbb{R}$ y $\mu, \nu \in \mathbb{P}(\mathbb{R})$ entonces

$$d_L(\mu, \nu) \leq d_P(\mu, \nu),$$

de donde tenemos que $\{d_L(\mu, \hat{\mu}^N) \geq \varepsilon\} \subset \{d_P(\mu, \hat{\mu}^N) \geq \varepsilon\}$ y entonces aplicando la desigualdad de Markov, concluimos que

$$\mathbb{P}(d_L(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \mathbb{P}(d_P(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \frac{\mathbb{E}(d_P(\mu, \hat{\mu}^N))}{\varepsilon}.$$

Con todo lo anterior podemos enunciar el siguiente teorema de concentración para medidas empíricas en \mathbb{R} con respecto a la distancia de Lévy.

Teorema 5.2.6. *Supongamos que μ es una medida de probabilidad en \mathbb{R} tal que:*

$$M_\beta = \int_{\mathbb{R}} |x|^\beta d\mu(x) < \infty,$$

entonces para todo $N \in \mathbb{N}$

$$\mathbb{P}(d_L(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \frac{2 + M_\beta^{1/2\beta}}{\varepsilon N^{1/(3+1/\beta)}}.$$

Observación 5.2.7: Notemos que este teorema deja de ser trivial para

$$N \geq \left(\frac{2 + M_\beta^{1/2\beta}}{\varepsilon} \right)^{3+1/\beta}.$$

Ejemplo 5.2.8: Consideremos μ la ley de una variable aleatoria normal estándar $\mathcal{N}(0, 1)$. En este caso M_β es finito para todo β . Si consideramos $\beta = 12$ y $\varepsilon = 0,045$, tenemos que:

$$\mathbb{P}(d_L(\mu, \hat{\mu}^N) \geq \varepsilon) \leq \frac{77,1149}{N^{0,3243}}. \quad (5.2.3)$$

Notemos que esta última cota es significativa para $N \geq 6,5868e + 05$. En la figura 5.2.1 se observa el comportamiento de $\mathbb{P}(d_L(\mu, \hat{\mu}^N) \geq \varepsilon)$ para los valores mencionados.

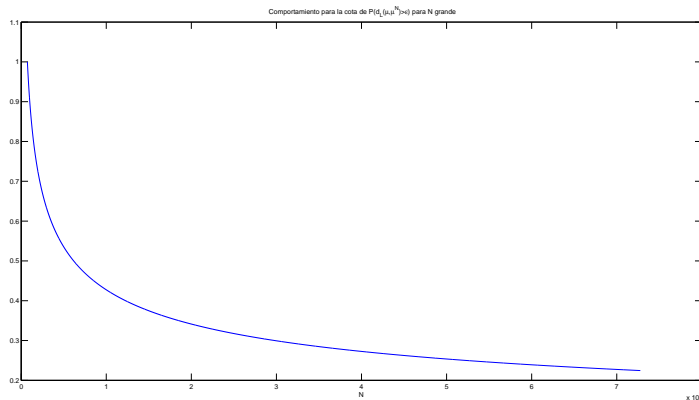


Figura 5.2.1: Comportamiento de la cota para $\mathbb{P}(d_L(\mu, \hat{\mu}^N) \geq \varepsilon)$ cuando N es grande y $\beta = 12$ y $\varepsilon = 0,045$. En el gráfico se observa el lento decaimiento hacia cero.

5.2.1. Concentración exponencial para la distancia de Lévy.

En lo que se acaba de hacer, se mostró que existe una concentración polinomial para la distancia de Lévy. Sin embargo, gracias a la relación que existe entre las distancias de Lévy, de Lévy-Prokhorov y de Wasserstein en \mathbb{R} , los resultados de concentración obtenidos para esta última en el capítulo anterior pueden aplicarse para la distancia de Lévy. En efecto, recordemos que

$$\mathbb{P}(d_L(\mu, \hat{\mu}^N) > \varepsilon) \leq \mathbb{P}(d_P(\mu, \hat{\mu}^N)^2 > \varepsilon^2) \leq \mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon^2).$$

Si bien tenemos varios resultados de concentración para W , nos quedamos con el mejor, que corresponde al teorema 5.1.2, el cual traducimos a este contexto.

5.2. Concentración de la medida empírica con respecto a d_P

Teorema 5.2.9. *Sea μ una medida de probabilidad en \mathbb{R} y $\lambda > 0$ tal que μ satisface $T_1(\lambda)$, entonces para todo $\lambda_1 < \lambda$, $N \in \mathbb{N}$, $\eta \in (0, 1)$ y $\beta_1 < \beta < \lambda/2$*

$$\begin{aligned} \mathbb{P}(d_L(\mu, \hat{\mu}^N) > \varepsilon) \leq & \left(\frac{16eR}{\delta_1 \varepsilon^2} \right)^{\left(\frac{2R}{\delta_1 \varepsilon^2} \right)} \exp \left(-N \left[\frac{\lambda_1}{2} \zeta^2 - E_\beta e^{-\beta R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\beta R^2 + 2 \right) \right] \right) + \\ & \exp \left(-N \left[\frac{\beta_1 R}{2} (1 - \eta) \varepsilon^2 - e^{(\beta_1 - \beta) R^2} E_\beta \right] \right), \end{aligned} \quad (5.2.4)$$

con δ_1 lo suficientemente pequeño y R lo suficientemente grande para que

$$\zeta = (\eta - \delta_1) \varepsilon^2 - 2E_\beta R e^{-\beta R^2} > 0, \quad (C1')$$

$$\frac{\beta_1 R}{2} (1 - \eta) \varepsilon^2 - e^{(\beta_1 - \beta) R^2} E_\beta > 0, \quad (C2')$$

$$\frac{\lambda_1}{2} \zeta^2 - E_\beta e^{-\beta R^2} \left(\frac{2\lambda_1 \lambda}{\lambda - \lambda_1} E_\beta R^2 + 2 \right) > 0, \quad (C3')$$

$$R \geq \frac{1}{\sqrt{2\beta}} \quad (C4')$$

y

$$R^2 \geq -\frac{1}{\beta} \log \left(\frac{0,7968}{E_\beta} \right). \quad (C5')$$

Ejemplo 5.2.10: Consideremos una vez más la ley normal estándar, la que sabemos satisface $T_1(1)$. Utilizando $\varepsilon = 0,045$ como en el ejemplo anterior, encontramos que el N_0 mínimo para que el teorema 5.2.9 sea válido es $N_0 = 1,2231e + 12$, el cual es mucho más grande que el N necesario para que el lado derecho de (5.2.3) sea menor o igual que 1.

El ejemplo anterior nos muestra que este último resultado es en la práctica muy poco útil, sin embargo se incluyó para dejar constancia de todos los resultados que se encontraron.

Capítulo 6

Estimación de Medidas de Riesgo Financiero

El objetivo de este capítulo es construir intervalos de confianza no asintóticos para el $V@R_\alpha$ y las medidas de riesgo espectrales al rededor de sus estimadores históricos. Para lograr este objetivo utilizaremos los resultados de concentración obtenidos en el capítulo anterior.

Con el fin de dar un contexto más completo en cuanto a la estimación de medidas de riesgo, también hemos incluido resultados acerca del comportamiento asintótico de los estimadores, en particular acerca de las condiciones que se deben tener para que el estimador histórico converja casi seguramente cuando el tamaño de la muestra tiende a infinito.

Una hipótesis importante que se asume implícitamente en este capítulo, es que los valores históricos de los retornos de un activo son independientes entre si, lo que en la práctica se sabe que no es cierto. Sin embargo, esta hipótesis se hace sistemáticamente en la literatura. Ver por ejemplo [10], [12], [14], [21].

El capítulo se divide en tres partes. En la primera se da una definición precisa al término estimador histórico y se discuten las razones para centrarnos en ellos. En la segunda parte se estudia el comportamiento asintótico y en la tercera se construyen intervalos de confianza no asintóticos bajo distintas hipótesis.

6.1. Estimadores históricos para medidas de riesgo espectrales

Consideremos una medida de riesgo ρ distribución dependiente. Dada una muestra i.i.d. $\{X_1, X_2, \dots, X_N\}$ de una variable X que representa las ganancias y pérdidas de un activo financiero, el estimador de riesgo histórico $\hat{\rho}^N$ asociado a ρ es el estimador obtenido al aplicar

6.1. Estimadores históricos para medidas de riesgo espectrales

ρ a F_N , la distribución empírica de la muestra, es decir,

$$\widehat{\rho}^N(X) = \rho(F_N). \quad (6.1.1)$$

Si ρ es una medida de riesgo espectral, con espectro ϕ , entonces $\widehat{\rho}^N(X)$ está dado por

$$\widehat{\rho}^N(X) = \rho(F_N) = - \sum_{i=0}^N w_{N,i} X_{(i)}, \quad (6.1.2)$$

donde $X_{(i)}$ es el i -ésimo elemento de $\{X_k\}_{k=1}^N$ ordenada de menor a mayor y $w_{N,i} = \int_{(i-1)/N}^{i/N} \phi(u) du$ para $i = 1, \dots, N$. Esta fórmula se deduce fácilmente si observamos que si $F = F_N$ es una distribución empírica entonces $q_{F_N}^-(\cdot)$ tiene la forma que se observa en la figura 6.1.1.

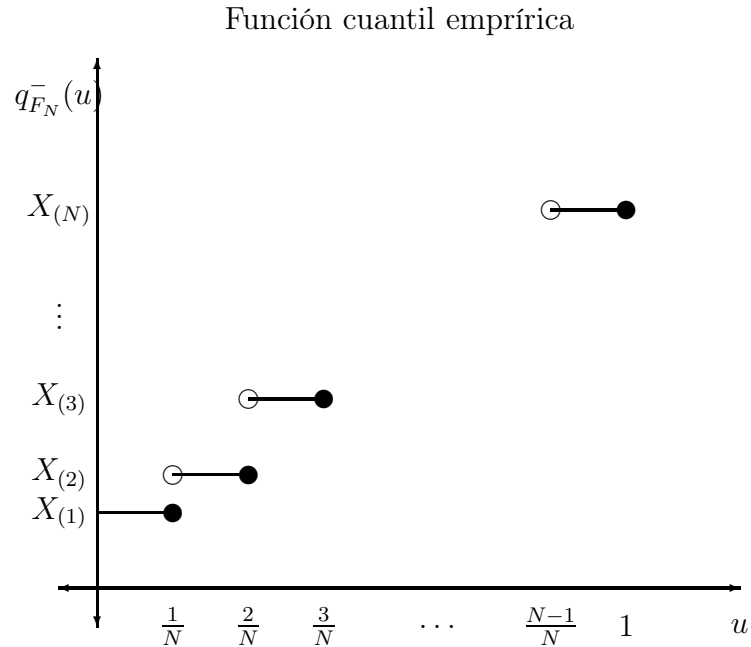


Figura 6.1.1: Función $q_{F_N}^-(u)$ v.s. u para F_N una distribución empírica.

6.2. Convergencia de Estimadores de Riesgo

Ejemplo 6.1.1: Aplicando las ecuaciones (6.1.1) y (6.1.2) podemos obtener fácilmente los estimadores históricos para:

1. **Value at Risk**($V@R_\alpha$):

$$\widehat{V@R}_\alpha^N(X) = -X_{(\lceil N\alpha \rceil)}.$$

2. **Expected Shortfall**(ES_α):

$$\widehat{ES}_\alpha^N(X) = -\frac{1}{N\alpha} \left(\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} + X_{(\lceil N\alpha \rceil)}(N\alpha - \lfloor N\alpha \rfloor) \right).$$

En ambas fórmulas $\lceil x \rceil$ corresponde al cajón superior de x , el menor entero que es mayor o igual que x .

¿Por qué estimadores históricos?

Los estimadores históricos son sólo una de las posibles alternativas para estimar medidas de riesgo, y sin embargo, en esta memoria sólo nos ocupamos de ellos. Es natural preguntarse por qué. La respuesta tiene que ver con los conceptos de robustez y sensibilidad estudiados por Cont et al. en [14], que básicamente apuntan a cuantificar el error de estimación producido por errores en los datos. Una de las conclusiones de dicho artículo es que los estimadores históricos son menos sensibles a estos errores. Fue esta robustez la que nos motivó a concentrarnos en ellos.

La razón subyacente de la robustez de los estimadores históricos, es que son estimadores no paramétricos, por lo que las hipótesis que se requieren para obtener sus propiedades son bastante generales y, en particular, no es necesario suponer que la distribución de la cantidad que se está estimando es alguna en particular.

6.2. Convergencia de Estimadores de Riesgo

6.2.1. Convergencia de $\widehat{V@R}_\alpha^N$

Para estudiar la convergencia de $\widehat{V@R}_\alpha^N$, contamos con dos resultados clásicos de la teoría estadística que aparecen en el Capítulo 21 de [42].

Teorema 6.2.1. *Para toda sucesión de funciones de distribución $F_N(\cdot)$, si $\{q_N^-\}_{N \geq 1}$ son sus funciones cuantil inferior entonces se tiene que $\{F_N\}_{N \geq 1}$ converge puntualmente a F en todo punto de continuidad de F si y sólo si $\{q_N^-\}_{n \geq 1}$ converge puntualmente a q_F^- en todo punto de continuidad de q_F^- .*

6.2. Convergencia de Estimadores de Riesgo

Consideremos una sucesión de variables aleatorias $\{X_n\}_{n \geq 1}$ i.i.d. con función de distribución F . Sea $\{F_N\}_{N \geq 1}$ la sucesión de funciones de distribución empíricas asociada a las primeras N variables y $q_N^-(\cdot)$ las funciones cuantil inferior de dichas muestras. Gracias al teorema de Glivenko-Cantelli 3.1.5, tenemos que casi seguramente F_N converge uniformemente a F . En particular, F_N converge puntualmente a F en todos los puntos de continuidad de F . Esto sumado al teorema anterior, nos dice que casi seguramente, la función cuantil empírica $q_N^-(\cdot)$ converge puntualmente a $q_F^-(\cdot)$ en todos los puntos de continuidad de esta última. En otras palabras, para todo $\alpha \in [0, 1]$ tal que α es un punto de continuidad de $q_F^-(\alpha)$, $q_N^-(\alpha)$ es un estimador consistente para $q_F^-(\alpha)$ en el sentido usual de la teoría de estimadores. (Ver por ejemplo [30]).

También contamos con un resultado acerca de la distribución asintótica de $q_N^-(\alpha)$, el cual enunciamos a continuación.

Teorema 6.2.2. *Sea $0 < \alpha < 1$. Si F es una función distribución y es diferenciable en $q_F^-(\alpha)$ con derivada positiva $f(q_F^-(\alpha))$, si llamamos $q_N^-(\alpha)$ a la función cuantil inferior empírica, entonces $\sqrt{n}(q_N^-(\alpha) - q_F^-(\alpha))$ es asintóticamente normal con media 0 y varianza $\alpha(1 - \alpha)/f^2(q_F^-(\alpha))$.*

Observación 6.2.3: Este resultado nos permite construir intervalos de confianza asintóticos para los cuantiles, pero tiene el defecto de que requiere que conozcamos la densidad de X en $q_F^-(\alpha)$, lo que en la práctica es desconocido.

Podemos reescribir los teoremas anteriores ahora en un lenguaje financiero, con lo que obtenemos el siguiente teorema.

Teorema 6.2.4. *Consideremos una muestra X_1, \dots, X_N i.i.d de un activo financiero X con función distribución F . Entonces:*

1. *Para todo α tal que $V@R_\alpha(X)$ es continuo en α se tiene que casi seguramente*

$$\widehat{V@R}_\alpha^N(X) \rightarrow V@R_\alpha(X).$$

2. *Para todo α tal que F es diferenciable en $V@R_\alpha(X)$ con derivada positiva. Se tiene que $\sqrt{N} \left(\widehat{V@R}_\alpha^N(X) - V@R_\alpha(X) \right)$ es asintóticamente normal con media 0 y varianza $\alpha(1 - \alpha)/(F'(V@R_\alpha(X)))^2$.*

Demostración: Aplicación directa de los teoremas 6.2.1 y 6.2.2.

Observación 6.2.5: Recordemos que $V@R_\alpha(X)$ tiene una discontinuidad en α si

$$|F_X^{-1}(\{\alpha\})| > 1,$$

6.2. Convergencia de Estimadores de Riesgo

es decir, si existe un intervalo donde F_X es constante e igual a α . Luego, si F_X es diferenciable en $V@R_\alpha(X)$ con derivada positiva, entonces $V@R_\alpha(X)$ es continuo en α . Por otro lado, si

$$\mathbb{P}(X = V@R_\alpha(X)) > 0,$$

entonces F_X no será diferenciable en $V@R_\alpha(X)$ y no será posible utilizar el resultado anterior para construir un intervalo de confianza al rededor de este punto.

Observación 6.2.6: Como ya lo notaron Acerbi y Tasche en [2], en términos de convergencia, el punto 1 del teorema anterior es el mejor resultado que se puede obtener. Esto, en el sentido de que si $q_X^-(\alpha) < q_X^+(\alpha)$, entonces la sucesión $\{X_{(\lfloor N\alpha \rfloor + 1)}\}_{N \geq 1}$ no converge con probabilidad 1. Para ver esto podemos citar el Teorema 1 de [19], en el cual se prueba que:

$$\mathbb{P}(X_{(\lfloor N\alpha \rfloor + 1)} \leq q_X^-(\alpha) \text{ infinitas veces}) = \mathbb{P}(X_{(\lfloor N\alpha \rfloor + 1)} \geq q_X^+(\alpha) \text{ infinitas veces}) = 1.$$

6.2.2. Convergencia de $\hat{\rho}_\phi^N$

Recordando las observaciones 3.2.17 y 3.2.18, podemos enunciar el siguiente teorema para la convergencia casi segura de $\hat{\rho}_\phi^N(X)$ a $\rho_\phi(X)$.

Teorema 6.2.7. *Sea X un activo financiero con ley μ . Si para $p \in [0, \infty)$, $\mathbb{E}(|X|^p) < \infty$, entonces para toda $\phi \in L^q((0, 1), dx)$ con $1/p + 1/q = 1$ se tiene con probabilidad 1 que*

$$\hat{\rho}_\phi^N(X) \rightarrow \rho_\phi(X).$$

Demostración: Comencemos por notar que

$$\hat{\rho}_\phi^N(X) = - \int_0^1 \phi(s) q_N^-(s) ds,$$

donde $q_N^-(\cdot)$ es la función cuantil asociada a la medida empírica de la muestra $\hat{\mu}^N$. Gracias a la observación 3.2.18, tenemos que

$$W_p^p(\mu, \hat{\mu}^N) = \int_0^1 |q_\mu^-(s) - q_N^-(s)|^p ds.$$

A partir de esta igualdad y de la desigualdad de Hölder, tenemos que

$$\begin{aligned} |\hat{\rho}_\phi^N(X) - \rho_\phi(X)| &= \left| \int_0^1 \phi(s) q_N^-(s) ds - \int_0^1 \phi(s) q_\mu^-(s) ds \right| \\ &\leq \|\phi\|_{L^q} \left(\int_0^1 |q_\mu^-(s) - q_N^-(s)|^p ds \right)^{1/p} \\ &= \|\phi\|_{L^q} W_p(\mu, \hat{\mu}^N). \end{aligned}$$

Para concluir, recordemos que en la observación 3.2.17, se probó que $W_p(\mu, \hat{\mu}^N)$ converge a cero para toda μ que tiene momento p -ésimo finito.

q.e.d.

A partir del teorema anterior podemos deducir directamente el siguiente corolario para el estimador histórico del *expected shortfall*.

Corolario 6.2.8. *Sea X un activo financiero. Si $\mathbb{E}(|X|) < \infty$, entonces para todo $\alpha \in (0, 1)$ se tiene que con probabilidad 1*

$$\widehat{ES}_\alpha^N(X) \rightarrow ES_\alpha(X).$$

6.3. Intervalos de confianza no asintóticos

6.3.1. Intervalos de confianza en muestras finitas para el $V@R_\alpha$

Para encontrar intervalos de confianza en muestras finitas para el $V@R_\alpha$, primero necesitamos establecer el siguiente teorema, que nos dice que la probabilidad de que el cuantil empírico de la muestra esté lejos del verdadero valor, está acotada por la probabilidad de que la distancia de Lévy entre la función distribución empírica de la muestra y la real sea grande.

Teorema 6.3.1. *Consideremos F una función distribución y $\{X_1, \dots, X_N\}$ una muestra i.i.d. de F . Si F_N es la función distribución empírica de la muestra y $q_N^-(\cdot)$ es la función cuantil empírica, entonces*

$$\mathbb{P}(q_F^-(\alpha) < q_N^-(\alpha - \delta) - \delta) \leq \mathbb{P}(d_L(F, F_N) \geq \delta). \quad (6.3.1)$$

Si además F es continua en $q_F^-(\alpha)$. Entonces

$$\mathbb{P}(q_F^-(\alpha) \notin [q_N^-(\alpha - \delta) - \delta, q_N^+(\alpha + \delta) + \delta]) \leq \mathbb{P}(d_L(F, F_N) \geq \delta). \quad (6.3.2)$$

Demostración: Supongamos que

$$d_L(F, F_N) < \delta,$$

entonces para todo $x \in \mathbb{R}$

$$\begin{aligned} F(x) &\leq F_N(x + \delta) + \delta \quad \text{y} \\ F_N(x - \delta) &\leq F(x) + \delta. \end{aligned}$$

En particular podemos considerar $x = q_F^-(\alpha)$ y entonces tenemos

$$F(q_F^-(\alpha)) \leq F_N(q_F^-(\alpha) + \delta) + \delta \quad \text{y} \quad (6.3.3)$$

$$F_N(q_F^-(\alpha) - \delta) \leq F(q_F^-(\alpha)) + \delta. \quad (6.3.4)$$

Recordemos que si g^- es una inversa generalizada de g , con g continua por la derecha y g^- continua por la izquierda, entonces tenemos que $g(g^-(s)) \geq s$ y $g^-(g(s)) \leq s$. Aplicando

6.3. Intervalos de confianza no asintóticos

estas desigualdades, primero con F en el rol de g y luego con F_N en el rol de g , se obtiene a partir de (6.3.3), que

$$q_N^-(\alpha - \delta) - \delta \leq q_F^-(\alpha), \quad (6.3.5)$$

donde $q_N^-(\cdot)$ es la función cuantil empírica. Recapitulando, hemos probado que

$$d_L(F, F_N) < \delta \Rightarrow q_N^-(\alpha - \delta) - \delta \leq q_F^-(\alpha),$$

lo que es equivalente a

$$q_N^-(\alpha - \delta) - \delta > q_F^-(\alpha) \Rightarrow d_L(F, F_N) \geq \delta,$$

de donde (6.3.1) se sigue trivialmente.

Para probar (6.3.2), supongamos que F es continua en $q_F^-(\alpha)$, lo que implica que

$$F(q_F^-(\alpha)) = \alpha.$$

Entonces,

$$F_N(q_F^-(\alpha) - \delta) \leq \alpha + \delta, \quad (6.3.6)$$

$$\alpha - \delta \leq F_N(q_F^-(\alpha) + \delta). \quad (6.3.7)$$

Con estas desigualdades y argumentos muy similares a los utilizados para obtener (6.3.5), encontramos que

$$q_N^-(\alpha - \delta) - \delta \leq q_F^-(\alpha) \quad y$$

$$q_F^-(\alpha) \leq q_N^+(\alpha + \delta) + \delta.$$

En resumen, tenemos que

$$d_L(F, F_N) < \delta \Rightarrow q_F^-(\alpha) \in [q_N^-(\alpha - \delta) - \delta, q_N^+(\alpha + \delta) + \delta],$$

o, lo que es equivalente,

$$q_F^-(\alpha) \notin [q_N^-(\alpha - \delta) - \delta, q_N^+(\alpha + \delta) + \delta] \Rightarrow d_L(F, F_N) \geq \delta,$$

de donde concluimos directamente (6.3.2).

q.e.d.

Gracias a este resultado y el teorema 5.2.6, se tiene el siguiente resultado para el $V@R_\alpha(X)$.

Teorema 6.3.2. *Sea X_1, \dots, X_N una muestra i.i.d, de un activo financiero X con función distribución F , tal que para $\beta > 1$*

$$M_\beta := \mathbb{E}(|X|^\beta) < \infty.$$

6.3. Intervalos de confianza no asintóticos

Entonces para todo $N \in \mathbb{N}$ y $\delta \in (0, \alpha)$ se tiene que

$$\mathbb{P}(V@R_\alpha(X) > \widehat{V@R}_{\alpha-\delta}^N(X) + \delta) \leq \frac{2 + M_\beta^{1/2\beta}}{\delta N^{1/(3+1/\beta)}}. \quad (6.3.8)$$

Si además F es continua en $V@R_\alpha$ entonces

$$\mathbb{P}(V@R_\alpha(X) \notin [\widehat{V@R}_{(\alpha+\delta)_+}^N(X) - \delta, \widehat{V@R}_{\alpha-\delta}^N(X) + \delta]) \leq \frac{2 + M_\beta^{1/2\beta}}{\delta N^{1/(3+1/\beta)}}. \quad (6.3.9)$$

Demostración: Recordemos que bajo las hipótesis que se tienen, el teorema 5.2.6 nos dice que

$$\mathbb{P}(d_L(F, F_N) \geq \varepsilon) \leq \frac{2 + M_\beta^{1/2\beta}}{\varepsilon N^{1/(3+1/\beta)}}.$$

Luego, como $V@R_\alpha(X) = -q_X^-(\alpha)$ y $\widehat{V@R}_{\alpha-\delta}^N(X) = -q_N^-(\alpha - \delta)$, gracias a (6.3.1) se tiene que

$$\mathbb{P}(V@R_\alpha(X) > \widehat{V@R}_{\alpha-\delta}^N(X) + \delta) \leq \mathbb{P}(d_L(F, F_N) \geq \delta).$$

Con esto se concluye (6.3.8).

Para obtener (6.3.9) los argumentos son muy similares. La única sutileza está en notar que para toda función distribución F , y para $x \in (0, 1)$ se tiene $q_F^-(x+) = q_F^+(x)$. En efecto, para todo $x' > x$, tenemos que $q_F^-(x') \leq q_F^+(x')$, de donde es directo que $q_F^-(x+) \leq q_F^+(x+) = q_F^+(x)$, porque la función cuantil superior es continua por la derecha. Para la otra desigualdad basta notar que para $x' > x$, $q_F^-(x') \geq q_F^+(x)$, luego tomando $x' \rightarrow x$ tenemos $q_F^-(x+) \geq q_F^+(x)$. Con esto $q_N^-(x+) = q_N^+(x)$, lo que nos permite reescribir (6.3.2) en términos de $q_N^-(\cdot)$ y concluir igual que antes.

q. e. d.

Observación 6.3.3: Notemos que el teorema deja de ser trivial para

$$N \geq \left(\frac{2 + M_\beta^{1/2\beta}}{\delta} \right)^{3+1/\beta}.$$

Este número puede ser bastante grande. Por ejemplo, cuando consideramos que la ley de la muestra es la de una variable aleatoria normal estándar, $\beta = 12$ y $\delta = 0,045$, tenemos que $N \approx 658\,690$.

6.3.2. Intervalos de confianza en muestras finitas para ρ_ϕ

Como ya vimos en el teorema 6.2.7, es posible acotar el error de estimación histórico $|\rho_\phi(X) - \widehat{\rho}_\phi^N(X)|$ por la distancia de Wasserstein entre la ley de X y la ley empírica de una

6.3. Intervalos de confianza no asintóticos

muestra i.i.d. de X . Específicamente se mostró que si para $p \in [1, \infty)$ $\mathbb{E}(|X|^p) < \infty$, entonces para todo espectro admisible $\phi \in L^q((0, 1), dx)$ con $1/p + 1/q = 1$, tenemos

$$|\widehat{\rho}_\phi^N(X) - \rho_\phi(X)| \leq \|\phi\|_{L^q} W_p(\mu, \hat{\mu}^N).$$

Por otro lado, en la sección 5.1, vimos que para $N \in \mathbb{N}$ y $\varepsilon > 0$, se tiene

$$\mathbb{P}(W_p(\mu, \hat{\mu}^N) > \varepsilon) \leq r(N, \varepsilon),$$

para una función adecuada r , decreciente a cero y que explicitamos en dicha sección. Juntando ambas desigualdades, concluimos que

$$\mathbb{P}(|\widehat{\rho}_\phi^N(X) - \rho_\phi(X)| > \varepsilon) \leq r(N, \varepsilon/\|\phi\|_{L^q}).$$

El teorema 5.1.2 y el corolario 5.1.21 nos dicen, para distintas hipótesis, cuál es la función r que debemos considerar y nos permiten establecer los resultados que siguen.

Teorema 6.3.4. *Sea X un activo financiero con ley de probabilidad μ , que suponemos satisface $T_1(\lambda)$. Llamemos $E_\beta = \int_{\mathbb{R}} e^{\beta x^2} d\mu(x)$. Entonces para todo $\lambda_1 < \lambda$, $N \in \mathbb{N}$, $\eta \in (0, 1)$, $\beta_1 < \beta < \lambda/2$ y $\phi \in L^\infty((0, 1), dx)$, espectro de riesgo admisible, se tiene que*

$$\begin{aligned} \mathbb{P}\left(|\widehat{\rho}_\phi^N(X) - \rho_\phi(X)| > \varepsilon\right) &\leq \left(\frac{16eR\|\phi\|_\infty}{\delta_1\varepsilon}\right)^{\left(\frac{2R\|\phi\|_\infty}{\delta_1\varepsilon}\right)} \exp\left(-N\left[\frac{\lambda_1}{2}\zeta^2 - E_\beta e^{-\beta R^2}\left(\frac{2\lambda_1\lambda}{\lambda - \lambda_1}E_\beta R^2 + 2\right)\right]\right) \\ &\quad + \exp\left(-N\left[\frac{\beta_1 R}{2\|\phi\|_\infty}(1 - \eta)\varepsilon - e^{(\beta_1 - \beta)R^2}E_\beta\right]\right), \end{aligned} \tag{6.3.10}$$

con δ_1 lo suficientemente pequeño y R lo suficientemente grande para que

$$\zeta = (\eta - \delta_1)\frac{\varepsilon}{\|\phi\|_\infty} - 2E_\beta R e^{-\beta R^2} > 0, \tag{C1}$$

$$\frac{\beta_1 R}{2\|\phi\|_\infty}(1 - \eta)\varepsilon - e^{(\beta_1 - \beta)R^2}E_\beta > 0, \tag{C2}$$

$$\frac{\lambda_1}{2}\zeta^2 - E_\beta e^{-\beta R^2}\left(\frac{2\lambda_1\lambda}{\lambda - \lambda_1}E_\beta R^2 + 2\right) > 0, \tag{C3}$$

$$R \geq \frac{1}{\sqrt{2\beta}} \tag{C4}$$

y

$$R^2 \geq -\frac{1}{\beta} \log\left(\frac{0,7968}{E_\beta}\right). \tag{C5}$$

Teorema 6.3.5. *Sea X un activo financiero con ley μ , que suponemos satisface $T_p(\lambda)$ para algún $p \in [1, 2]$. Entonces, para todo espectro admisible $\phi \in L^q((0, 1), dx)$, con $q \geq p/(p-1)$, se tiene que*

$$\mathbb{P}(|\widehat{\rho}_\phi^N(X) - \rho_\phi(X)| \geq \varepsilon) \leq \exp\left(-\frac{N\lambda}{2} \left(\frac{\varepsilon}{\|\phi\|_{L^q}} - \frac{D^{1/2}}{N^{1/5}}\right)^2\right)$$

para todo $\varepsilon > 0$ y $N \geq D^{5/2}\|\phi\|_{L^q}^5/\varepsilon^5$, donde

$$D = 6(\sqrt{\pi}(61 + 32M_6))^{1/2}.$$

Observación 6.3.6: Notemos que en el caso de que μ satisfaga $T_2(\lambda)$, este teorema nos permite controlar el error de estimación para espectros admisibles en $L^2((0, 1), dx)$.

Observación 6.3.7: En los últimos dos teoremas, el valor de N_0 , la cantidad mínima de observaciones para que las desigualdades que proponen dichos teoremas sean válidas o informativas, puede ser muy grande. Esto se puede apreciar en la siguiente tabla, donde aparecen el valor de N_0 para estos dos teoremas cuando se considera que la muestra tiene ley normal estándar.

ε	N_0 Teorema 6.3.4	N_0 Teorema 6.3.5
0,01	$8,997e + 13$	$1,5057e + 22$
0,05	$6,396e + 11$	$4,8183e + 18$
0,1	$7,565e + 10$	$1,5057e + 17$
0,5	$5,260e + 08$	$4,8183e + 13$
1	$6,151e + 07$	$1,5057e + 12$

Cuadro 6.3.1: Comparación de los valores obtenidos para N_0 por los teoremas 6.3.4 y 6.3.5.

En la práctica, una desigualdad que requiera de más de 61 000 000 de datos es inutilizable. Estos resultados nos llevaron buscar otros caminos para establecer desigualdades que requirieran menos observaciones para ser válidas. En las secciones que siguen se presenta el resultado de dicha búsqueda.

6.3.3. Teorema de Brown, Gao y Wang de concentración para

$$\widehat{CV@R_\alpha}^N(X)$$

Antes de presentar los resultados de esta sección, introducimos el **Conditional Value at Risk** de un activo.

Definición 6.3.8. *Sea X un activo tal que $\mathbb{E}(X^-) < \infty$ y $\alpha \in [0, 1]$. Llamamos Conditional Value at Risk de nivel α de X a la cantidad*

$$CV@R_\alpha(X) = \mathbb{E}(-X | -X \geq V@R_\alpha(X)).$$

6.3. Intervalos de confianza no asintóticos

El Corolario 4.49 que aparece en el Capítulo 4 de [20], nos dice que si

$$P(X \leq q_X^-(\alpha)) = \alpha,$$

entonces $CV@R_\alpha(X) = ES_\alpha(X)$. Notemos que la condición anterior se tiene si la distribución de X es continua.

Un resultado que hasta ahora no hemos mencionado, pero que es muy útil en el contexto de optimización de portafolios, es la siguiente representación dual para el *Expected Shortfall* dada por Rockafellar y Uryasev en [38]. La versión que se presenta a continuación, es un caso particular a la dada por estos autores y corresponde a la que aparece en el Capítulo 4 de [20] (Lema 4.46).

Lema 6.3.9. *Para todo $\alpha \in (0, 1)$ y X tal que $\mathbb{E}(X^-) < \infty$, se tiene que*

$$ES_\alpha(X) = \inf_{s \in \mathbb{R}} \left\{ \frac{\mathbb{E}(s - X)^+}{\alpha} - s \right\}.$$

En lo que sigue, a menos que se indique explícitamente lo contrario, supondremos que X tiene una función distribución continua. En este caso

$$CV@R_\alpha(X) = \inf_{s \in \mathbb{R}} \left\{ \frac{\mathbb{E}(s - X)^+}{\alpha} - s \right\}.$$

Utilizando esta representación, D. Brown propone en [10], el siguiente estimador histórico para el $CV@R_\alpha(X)$:

$$\widehat{CV@R_\alpha}^N(X) = \min_{s \in \mathbb{R}} \left\{ \frac{1}{N\alpha} \sum_{i=1}^N (s - X_i)^+ - s \right\},$$

y prueba el resultado de concentración para este estimador al rededor de $CV@R_\alpha(X)$ cuando X toma valores en un intervalo compacto.

Teorema 6.3.10. [D. Brown.(2007)]. *Sea X una variable aleatoria con soporte contenido en el intervalo $[a, b]$, con a y b números reales. Entonces, para todo $\varepsilon \geq 0$,*

$$\mathbb{P}(\widehat{CV@R_\alpha}^N(X) \leq CV@R_\alpha(X) - \varepsilon) \leq 3e^{-\frac{\alpha}{5} \left(\frac{\varepsilon}{b-a}\right)^2 N}$$

y

$$\mathbb{P}(\widehat{CV@R_\alpha}^N(X) \geq CV@R_\alpha(X) + \varepsilon) \leq e^{-2 \left(\frac{\alpha\varepsilon}{b-a}\right)^2 N}.$$

El resultado de Brown se basa en una desigualdad clásica de concentración propuesta por Mc Diarmid en [34]. Esta desigualdad se puede encontrar también en el Capítulo 5 de [33].

6.3. Intervalos de confianza no asintóticos

Observación 6.3.11: Notemos que la primera desigualdad de Brown acota la probabilidad de estar subestimando el $CV@R_\alpha(X)$, mientras que la segunda acota la probabilidad de estar sobreestimando el $CV@R_\alpha(X)$. La importancia de cada una de estas desigualdades depende del agente que las esté utilizando. Para una entidad reguladora será más importante acotar la probabilidad de subestimar el $CV@R_\alpha(X)$, mientras que para un inversor será más grave sobreestimar el $CV@R_\alpha(X)$, ya que una sobreestimación del riesgo de un activo puede llevarlo a construir un portafolio subóptimo.

A continuación veremos que el problema de minimización que define a $\widehat{CV@R}_\alpha^N(X)$ tiene solución explícita. Llamemos

$$f(s) = \frac{1}{N\alpha} \sum_{i=1}^N (s - X_i)^+ - s,$$

y $s^* = X_{(\lceil N\alpha \rceil)}$, donde $\lceil x \rceil$ es el cajón superior de x . Entonces, el mínimo de f se alcanza en s^* . En efecto, notemos que para $h > 0$ suficientemente pequeño se tiene que

$$\begin{aligned} f(s^* + h) - f(s^*) &= \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil - 1} (X_{(\lceil N\alpha \rceil)} + h - X_{(i)})^+ + \frac{1}{N\alpha} \sum_{i=\lceil N\alpha \rceil}^N (X_{(\lceil N\alpha \rceil)} + h - X_{(i)})^+ \\ &\quad - X_{(\lceil N\alpha \rceil)} - h \\ &\quad - \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil - 1} (X_{(\lceil N\alpha \rceil)} - X_{(i)})^+ - \frac{1}{N\alpha} \sum_{i=\lceil N\alpha \rceil}^N (X_{(\lceil N\alpha \rceil)} - X_{(i)})^+ \\ &\quad + X_{(\lceil N\alpha \rceil)} \\ &= \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil} (X_{(\lceil N\alpha \rceil)} - X_{(i)}) + \frac{h(\lceil N\alpha \rceil - 1)}{N\alpha} + \frac{h}{N\alpha} - h \\ &\quad - \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil} (X_{(\lceil N\alpha \rceil)} - X_{(i)}) \\ &= h \frac{\lceil N\alpha \rceil - N\alpha}{N\alpha} \\ &\geq 0, \end{aligned}$$

donde se usó el hecho de que los valores de la muestra son todos distintos, pues la ley de la

6.3. Intervalos de confianza no asintóticos

muestra no tiene átomos. Por otro lado,

$$\begin{aligned}
 f(s^* - h) - f(s^*) &= \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil - 1} (X_{(\lceil N\alpha \rceil)} - h - X_{(i)})^+ - X_{(\lceil N\alpha \rceil)} + h \\
 &\quad - \frac{1}{N\alpha} \sum_{i=1}^{\lceil N\alpha \rceil - 1} (X_{(\lceil N\alpha \rceil)} - X_{(i)})^+ + X_{(\lceil N\alpha \rceil)} \\
 &= h \frac{N\alpha + 1 - \lceil N\alpha \rceil}{N\alpha} \\
 &\geq 0.
 \end{aligned}$$

A partir de estos cálculos se concluye que s^* es un mínimo local para f . Pero como f es una función convexa, se sigue que en realidad s^* es un mínimo global. Con esto, se tiene que

$$\widehat{CV@R}_\alpha^N(X) = \min_{s \in \mathbb{R}} \left\{ \frac{1}{N\alpha} \sum_{i=1}^N (s - X_i)^+ - s \right\} = f(X_{(\lceil N\alpha \rceil)}).$$

Es decir

$$\widehat{CV@R}_\alpha^N(X) = -\frac{1}{N\alpha} \left(\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} + (N\alpha - \lfloor N\alpha \rfloor) X_{(\lceil N\alpha \rceil)} \right).$$

Utilizando la expresión anterior para el estimador histórico del $CV@R_\alpha(X)$, F. Gao e Y. Wang, encontraron el siguiente resultado que corresponde al Teorema 3.1 de [21].

Teorema 6.3.12. [F. Gao e Y. Wang.(2010)]. *Sea X una variable aleatoria con soporte contenido en el intervalo $[a, b]$, con a y b números reales. Entonces, para todo $\varepsilon \geq 0$,*

$$\mathbb{P}(\widehat{CV@R}_\alpha^N(X) \leq CV@R_\alpha(X) - \varepsilon) \leq 3e^{-\frac{\alpha}{5} \left(\frac{\varepsilon}{b-a}\right)^2 N}$$

y

$$\mathbb{P}(\widehat{CV@R}_\alpha^N(X) \geq CV@R_\alpha(X) + \varepsilon) \leq 3e^{-\frac{\alpha}{11} \left(\frac{\varepsilon}{b-a}\right)^2 N}.$$

Observación 6.3.13: La diferencia entre el resultado de Brown y el resultado de Gao y Wang, es la desigualdad para la probabilidad de sobreestimar el $CV@R_\alpha(X)$. En el caso de que α sea pequeño la cota de Gao y Wang es mejor, por el contrario si α no es muy pequeño, es mejor el resultado de Brown. Más precisamente, la desigualdad propuesta por Brown es más ajustada que la propuesta por Gao y Wang, si

$$\log(3) \geq \frac{N\alpha\varepsilon^2}{(b-a)^2} \left(\frac{1}{11} - 2\alpha \right),$$

lo que se cumple para todo N si $\alpha \geq 1/22$. Si por el contrario, $\alpha < 1/22$, la cota de Brown será más ajustada para

$$N \leq \frac{11(b-a)^2}{\alpha\varepsilon^2(1-22\alpha)} \log(3),$$

mientras que después de este umbral, la cota de Gao y Wang será mejor.

Notemos que para $b - a = 1$, $\varepsilon = 0,05$ y $\alpha = 0,01$, tenemos que el lado derecho de la última desigualdad vale aproximadamente 619 730. Resultados similares se pueden obtener para otros valores de $\alpha \in (0, 1/22)$. Esto nos indica que para muestras pequeñas la desigualdad de Brown es competitiva con la desigualdad de Gao y Wang para cualquier valor de α .

Si bien estos resultados son atractivos por su simplicidad, y porque dados α , ε , a y b , no es necesario tener un N muy grande para que las cotas sean significativas, tienen una limitación importante al ser válidos sólo para variables aleatorias que toman valores en un intervalo compacto. Notemos que en la práctica, la hipótesis de que a y b son conocidos, es una hipótesis difícil de sostener. Si se conoce el valor de la máxima pérdida que puede tomar un activo, inmediatamente se conoce el máximo riesgo que se enfrenta, sin siquiera necesitar una muestra. Por otro lado, en los artículos mencionados de Brown y de Gao y Wang, no se especifica un criterio de truncamiento para medidas que tomen valores en \mathbb{R} , y notemos que establecer un criterio para ello no es fácil, ya que el $CV@R_\alpha$ de un activo, depende justamente de la cola izquierda de la distribución de un activo, que es lo que se quiere truncar.

En la siguiente presentamos una forma de combinar los resultados anteriores con algunas ideas de concentración de la medida empírica que vimos en el capítulo anterior, de modo de extender los teoremas de Brown, Gao y Wang a variables aleatorias con valores en \mathbb{R} . Pero antes, explicitemos un corolario de los resultados anteriores para variables aleatorias con distribución continua y que toman valores en el intervalo $[-R, R]$. Notemos que en este caso el *conditional value at risk* y el *expected shortfall* coinciden.

Corolario 6.3.14. *Sea X una variable aleatoria que toma valores en el intervalo $[-R, R]$ y cuya función distribución es continua. Entonces*

$$\mathbb{P} \left(\left| ES_\alpha(X) - \widehat{ES}_\alpha^N(X) \right| \geq \varepsilon \right) \leq \Psi(\alpha, \varepsilon, R, N),$$

donde

$$\Psi(\alpha, \varepsilon, R, N) = 3 \exp \left(-\frac{N\alpha\varepsilon^2}{20R^2} \right) + \min \left\{ 3 \exp \left(-\frac{N\alpha\varepsilon^2}{44R^2} \right), \exp \left(-\frac{N\alpha^2\varepsilon^2}{2R^2} \right) \right\}. \quad (6.3.11)$$

6.3.4. Extensión del resultado de Brown, Gao y Wang a variables aleatorias con valores en \mathbb{R} .

El objetivo de esta sección es presentar el siguiente teorema, donde se extiende el resultado de Brown, Gao y Wang a variables aleatorias con valores en \mathbb{R} . Además, a través de un ejemplo, se compara el desempeño de esta extensión con el desempeño del teorema (6.3.4).

Teorema 6.3.15. *Sea X un activo financiero con ley μ . Supongamos que μ no tiene átomos y satisface $T_1(\lambda)$. Consideremos $\eta \in (0, 1)$ y sean β y β_1 tales que $\beta_1 < \beta < \lambda/2$. Llamemos*

6.3. Intervalos de confianza no asintóticos

$E_\beta = \int_{\mathbb{R}} e^{\beta x^2} d\mu(x)$, y sea $R > 0$ que satisface las siguientes condiciones:

$$\eta\varepsilon - \frac{2}{\alpha} R e^{-\beta R^2} E_\beta > 0, \quad (\text{D1})$$

$$\frac{\beta_1 R (1 - \eta) \alpha \varepsilon}{2} - e^{(\beta_1 - \beta) R^2} E_\beta > 0 \text{ y} \quad (\text{D2})$$

$$R \geq \frac{1}{\sqrt{2\beta}}. \quad (\text{D3})$$

Entonces, para todo $N \in \mathbb{N}$ se tiene que

$$\begin{aligned} \mathbb{P} \left(\left| ES_\alpha(X) - \widehat{ES}_\alpha^N(X) \right| \geq \varepsilon \right) &\leq \Psi \left(\alpha, \left(\eta\varepsilon - \frac{2}{\alpha} R e^{-\beta R^2} E_\beta \right), R, N \right) \\ &\quad + \exp \left(-N \left[\frac{\beta_1 R (1 - \eta) \alpha \varepsilon}{2} - e^{(\beta_1 - \beta) R^2} E_\beta \right] \right), \end{aligned}$$

donde Ψ es la función definida en (6.3.11).

Demostación: En esta demostración procederemos como en la demostración del teorema 5.1.2. Consideremos μ_R , $\hat{\mu}^N$ y $\hat{\mu}_R^N$ como en dicho teorema. Sean $q_\mu^-(\cdot)$, $q_R^-(\cdot)$, $q_N^-(\cdot)$ y $q_{R,N}^-(\cdot)$ las funciones cuantil de dichas leyes. Entonces:

$$\begin{aligned} \left| ES_\alpha(X) - \widehat{ES}_\alpha^N(X) \right| &= \left| \frac{1}{\alpha} \int_0^\alpha q_\mu^-(s) ds - \frac{1}{\alpha} \int_0^\alpha q_N^-(s) ds \right| \\ &= \left| \frac{1}{\alpha} \int_0^\alpha q_\mu^-(s) ds - \frac{1}{\alpha} \int_0^\alpha q_R^-(s) ds + \frac{1}{\alpha} \int_0^\alpha q_{R,N}^-(s) ds \right. \\ &\quad \left. - \frac{1}{\alpha} \int_0^\alpha q_N^-(s) ds + \frac{1}{\alpha} \int_0^\alpha q_R^-(s) ds - \frac{1}{\alpha} \int_0^\alpha q_{R,N}^-(s) ds \right| \\ &\leq \frac{1}{\alpha} \int_0^1 |q_\mu^-(s) - q_R^-(s)| ds + \frac{1}{\alpha} \int_0^1 |q_{R,N}^-(s) - q_N^-(s)| ds \\ &\quad + \left| \frac{1}{\alpha} \int_0^\alpha q_R^-(s) ds - \frac{1}{\alpha} \int_0^\alpha q_{R,N}^-(s) ds \right| \\ &= \frac{1}{\alpha} W(\mu, \mu_R) + \frac{1}{\alpha} W(\hat{\mu}^N, \hat{\mu}_R^N) + \left| ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R) \right|. \end{aligned}$$

Gracias al lema 5.1.4, para $\beta < \lambda/2$ y $R \geq 1/\sqrt{2\beta}$ se tiene que

$$W(\mu, \mu_R) \leq 2R e^{-\beta R^2} E_\beta,$$

y entonces se cumple que

$$\left| ES_\alpha(X) - \widehat{ES}_\alpha^N(X) \right| \leq \frac{2}{\alpha} R e^{-\beta R^2} E_\beta + \frac{1}{\alpha} W(\hat{\mu}^N, \hat{\mu}_R^N) + \left| ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R) \right|.$$

6.3. Intervalos de confianza no asintóticos

Por otro lado, gracias al lema 5.1.5, para todo $\beta_1 < \beta$ y $\varepsilon > 0$, se tiene que

$$\mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) > \varepsilon) \leq \exp\left(-N \left[\frac{\beta_1 R \varepsilon}{2} - e^{(\beta_1 - \beta)R^2} E_\beta\right]\right).$$

Luego, para todo $\eta \in (0, 1)$ y $\beta_1 < \beta$,

$$\begin{aligned} \mathbb{P}\left(\left|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)\right| \geq \varepsilon\right) &\leq \mathbb{P}\left(\frac{2}{\alpha}Re^{-\beta R^2}E_\beta + \frac{1}{\alpha}W(\hat{\mu}^N, \hat{\mu}_R^N) + \left|ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R)\right| \geq \varepsilon\right) \\ &\leq \mathbb{P}\left(\left|ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R)\right| \geq \eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right) \\ &\quad + \mathbb{P}(W(\hat{\mu}^N, \hat{\mu}_R^N) \geq (1 - \eta)\alpha\varepsilon) \\ &\leq \mathbb{P}\left(\left|ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R)\right| \geq \eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right) \\ &\quad + \exp\left(-N \left[\frac{\beta_1 R(1 - \eta)\alpha\varepsilon}{2} - e^{(\beta_1 - \beta)R^2}E_\beta\right]\right). \end{aligned}$$

En este punto hay que notar que μ_R es la ley de una variable aleatoria que toma valores en $[-R, R]$ y que no tiene átomos, pues μ no los tiene. Así podemos aplicar el corolario 6.3.14 y encontramos que

$$\mathbb{P}\left(\left|ES_\alpha(\mu_R) - \widehat{ES}_\alpha^N(\mu_R)\right| \geq \eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right) \leq \Psi\left(\alpha, \left(\eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right), R, N\right)$$

Resumiendo, para todo $\beta_1 < \beta < \lambda/2$, $\eta \in (0, 1)$ y $N \in \mathbb{N}$. Se tiene que

$$\begin{aligned} \mathbb{P}\left(\left|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)\right| \geq \varepsilon\right) &\leq \Psi\left(\alpha, \left(\eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right), R, N\right) \\ &\quad + \exp\left(-N \left[\frac{\beta_1 R(1 - \eta)\alpha\varepsilon}{2} - e^{(\beta_1 - \beta)R^2}E_\beta\right]\right). \end{aligned}$$

Con R lo suficientemente grande de modo que satisfaga las condiciones

$$\eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta > 0, \tag{D1}$$

$$\frac{\beta_1 R(1 - \eta)\alpha\varepsilon}{2} - e^{(\beta_1 - \beta)R^2}E_\beta > 0 \text{ y} \tag{D2}$$

$$R \geq \frac{1}{\sqrt{2\beta}}. \tag{D3}$$

q. e. d.

6.3. Intervalos de confianza no asintóticos

Ejemplo 6.3.16: Consideremos $\alpha = 0,05$. En este caso

$$\Psi(\alpha, \varepsilon, R, N) = 3 \exp\left(-\frac{N\varepsilon^2}{400R^2}\right) + \exp\left(\frac{-N\varepsilon^2}{800R^2}\right),$$

luego, si X es una variable aleatoria normal estándar, $\beta_1 < \beta < 1/2$ y $R > 0$ satisface las condiciones (D1), (D2) y (D3) que se indican en el teorema anterior, se tiene

$$\begin{aligned} \mathbb{P}\left(\left|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)\right| \geq \varepsilon\right) &\leq 3 \exp\left(-\frac{N}{400R^2} \left(\eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right)^2\right) \\ &+ \exp\left(\frac{-N}{800R^2} \left(\eta\varepsilon - \frac{2}{\alpha}Re^{-\beta R^2}E_\beta\right)^2\right) \\ &+ \exp\left(-N \left[\frac{\beta_1 R(1-\eta)\varepsilon}{40} - e^{(\beta_1-\beta)R^2}E_\beta\right]\right), \end{aligned}$$

para para todo $\varepsilon > 0$, $\eta \in (0, 1)$ y $N \in \mathbb{N}$. Es decir, para c_1 y c_2 constantes adecuadas tenemos que

$$\mathbb{P}\left(\left|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)\right| \geq \varepsilon\right) \leq 3 \exp(-Nc_1) + \exp\left(\frac{-N}{2}c_1\right) + \exp(-Nc_2).$$

En la tabla 6.3.2 aparece el valor de las distintas constantes c_1 , c_2 y N_0 para distintos valores de epsilon.

ε	c_1	c_2	N_0
0,01	$6,348e - 09$	$4,815e - 07$	$2,628e + 08$
0,05	$1,744e - 07$	$3,596e - 06$	$9,564e + 06$
0,1	$7,296e - 07$	$7,181e - 06$	$2,287e + 06$
0,5	$2,042e - 05$	$3,574e - 05$	$8,482e + 04$
1	$8,611e - 05$	$7,132e - 05$	$2,269e + 04$

Cuadro 6.3.2: Resumen comportamiento N_0 para una variable normal estándar y $\alpha = 0,05$.

Una pregunta interesante que se planteó en el desarrollo de esta memoria, es qué tan buena estrategia es acotar el error de estimación del *expected shortfall* de una variable aleatoria X , por la distancia de Wasserstein entre la ley de X y su aproximación empírica. Es decir, aplicar

$$\left|ES_\alpha(X) - \widehat{ES}_\alpha^N(X)\right| \leq \frac{1}{\alpha}W(\mu_X, \hat{\mu}_X^N). \quad (6.3.12)$$

Para cuantificar que tan gruesa es esta cota, en la tabla 6.3.3, se compara el número mínimo de observaciones necesarias para que las cotas dadas por los teoremas 6.3.4 y 6.3.15 sean significativas.

6.3. Intervalos de confianza no asintóticos

ε	N_0 Teorema 6.3.4	N_0 Teorema 6.3.15
0,01	$8,997e + 13$	$2,628e + 08$
0,05	$6,396e + 11$	$9,564e + 06$
0,1	$7,565e + 10$	$2,287e + 06$
0,5	$5,260e + 08$	$8,482e + 04$
1	$6,151e + 07$	$2,269e + 04$

Cuadro 6.3.3: Comparación de los valores obtenidos para N_0 por los teoremas 6.3.4 y 6.3.15.

Como se observa en la tabla 6.3.3, el teorema 6.3.15 es mucho más eficiente que los que se han presentado previamente. La razón para ello es que la desigualdad (6.3.12) es demasiado gruesa, por lo que para lograr mejores resultados se deberían buscar otras estrategias.

6.3.5. Una nueva estimación para las desviaciones de $\widehat{ES}_\alpha^N(X)$

En esta sección, utilizando la representación dual del ES_α y aplicando el lema 5.1.14, encontraremos una cota para

$$\mathbb{P}(\widehat{ES}_\alpha^N(X) \geq ES_\alpha(X) + \varepsilon),$$

válida para variables aleatorias que toman valores en \mathbb{R} , cuya ley satisface T_1 . Para establecer este resultado, nos basaremos en dos lemas que presentamos a continuación.

Lema 6.3.17. *Sea X un activo financiero, entonces*

$$\mathbb{E}(\widehat{ES}_\alpha^N(X)) \leq ES_\alpha(X).$$

Demostración: Comencemos por recordar que

$$\widehat{ES}_\alpha^N(X) = \frac{1}{\alpha} \min_{s \in \mathbb{R}} \left\{ \frac{1}{N} \sum_{i=1}^N (s - X_i)^+ - \alpha s \right\}.$$

Luego, para todo s

$$\widehat{ES}_\alpha^N(X) \leq \frac{1}{N\alpha} \sum_{i=1}^N (s - X_i)^+ - s$$

y entonces

$$\mathbb{E}(\widehat{ES}_\alpha^N(X)) \leq \frac{1}{N\alpha} \sum_{i=1}^N \mathbb{E}((s - X_i)^+) - s = \frac{\mathbb{E}((s - X)^+)}{\alpha} - s.$$

Tomando ínfimo sobre s , en el lado derecho de la última desigualdad, obtenemos la desigualdad deseada.

6.3. Intervalos de confianza no asintóticos

q.e.d.

Este lema nos dice que en promedio el estimador histórico tiende a subestimar el verdadero valor de $ES_\alpha(X)$. Es importante notar que este lema ya fue utilizado en las demostraciones de Brown y de Gao y Wang.

Ahora un pequeño lema técnico.

Lema 6.3.18. Sea $f : \mathbb{R}^N \rightarrow \mathbb{R}$, definida como

$$f(x) = \min_{s \in \mathbb{R}} \left\{ \frac{1}{N} \sum_{i=1}^N (s - x_i)^+ - \alpha s \right\}.$$

Entonces f es una función $1/N$ -Lipschitz con respecto a la métrica $d_1(x, y) = \sum_{i=1}^N |x_i - y_i|$.

Demostración: Notemos que para todo $s \in \mathbb{R}$ la función $(s - \cdot)^+$ es 1-Lipschitz en $(\mathbb{R}, |\cdot|)$, luego para todo $s \in \mathbb{R}$ tenemos que

$$\left(\frac{1}{N} \sum_{i=1}^N (s - x_i)^+ - \alpha s \right) - \left(\frac{1}{N} \sum_{i=1}^N (s - y_i)^+ - \alpha s \right) = \frac{1}{N} \sum_{i=1}^N (s - x_i)^+ - (s - y_i)^+ \leq \frac{1}{N} \sum_{i=1}^N |x_i - y_i|.$$

Luego para todo s la función

$$f_s(x) = \frac{1}{N} \sum_{i=1}^N (s - x_i)^+ - \alpha s,$$

es $1/N$ -Lipschitz. Ahora notemos que para todo $s' \in \mathbb{R}$ y $x, y \in \mathbb{R}^N$

$$\inf_{s \in \mathbb{R}} f_s(x) - f_{s'}(y) \leq f_{s'}(x) - f_{s'}(y) \leq \frac{1}{N} d(x, y),$$

de donde tenemos

$$\inf_{s \in \mathbb{R}} f_s(x) - \frac{1}{N} d(x, y) \leq f_{s'}(y).$$

Tomando ínfimo sobre s' en la última desigualdad y reordenando términos, obtenemos

$$\inf_{s \in \mathbb{R}} f_s(x) - \inf_{s' \in \mathbb{R}} f_{s'}(y) \leq \frac{1}{N} d(x, y).$$

A partir de esta última desigualdad, intercambiando los roles de x e y , concluimos que f es $1/N$ -Lipschitz.

q.e.d.

Teorema 6.3.19. Sea X un activo financiero, con ley μ que satisface $T_1(\lambda)$. Entonces

$$\mathbb{P} \left(\widehat{ES}_\alpha^N(X) \geq ES_\alpha(X) + \varepsilon \right) \leq \exp \left(\frac{-\lambda N \alpha^2 \varepsilon^2}{2} \right).$$

Demostración: Gracias al lema 6.3.17, tenemos que

$$\mathbb{P}\left(\widehat{ES}_\alpha^N(X) \geq ES_\alpha(X) + \varepsilon\right) \leq \mathbb{P}\left(\widehat{ES}_\alpha^N(X) \geq \mathbb{E}(\widehat{ES}_\alpha^N(X)) + \varepsilon\right).$$

Además,

$$\mathbb{P}\left(\widehat{ES}_\alpha^N(X) \geq \mathbb{E}(\widehat{ES}_\alpha^N(X)) + \varepsilon\right) = \mathbb{P}\left(f(X_1, \dots, X_N) \geq \mathbb{E}(f(X_1, \dots, X_N)) + \alpha\varepsilon\right),$$

donde f es la función definida en el lema 6.3.18. Como μ satisface $T_1(\lambda)$ y f es $1/N$ -Lipschitz, gracias al lema 5.1.14, concluimos que para todo $\varepsilon > 0$,

$$\mathbb{P}\left(f(X_1, \dots, X_N) \geq \mathbb{E}(f(X_1, \dots, X_N)) + \alpha\varepsilon\right) \leq \exp\left(\frac{-\lambda N \alpha^2 \varepsilon^2}{2}\right).$$

q.e.d.

Notemos que gracias al corolario 4.2.11, tenemos que si μ tiene soporte contenido en el intervalo $[a, b]$, entonces μ satisface $T_1(4/(b-a)^2)$. Luego, gracias al último teorema se tiene el siguiente corolario, que corresponde a la cota encontrada por Brown.

Corolario 6.3.20. *Si X es un activo cuya distribución tiene soporte en el intervalo $[a, b]$. Entonces*

$$\mathbb{P}\left(\widehat{ES}_\alpha^N(X) \geq ES_\alpha(X) + \varepsilon\right) \leq \exp\left(\frac{-2N\alpha^2\varepsilon^2}{(b-a)^2}\right). \quad (6.3.13)$$

En la observación 6.3.13, se discutió bajo que condiciones esta cota es mejor a la encontrada posteriormente por Gao y Wang.

Capítulo 7

Conclusión

En esta memoria de ingeniería se ha propuesto y estudiado una metodología para cuantificar el error de estimación de medidas de riesgo espectrales. El producto final de este trabajo de investigación son los resultados del capítulo anterior, en los que se establecen intervalos de confianza no asintóticos para los estimadores históricos de las medidas de riesgo espectrales, que dada la generalidad de las hipótesis con que se construyen, vienen a extender los resultados hasta ahora conocidos. Además de los resultados concretos, en esta memoria se avanzó en comprender como influye la existencia de una representación dual para una medida de riesgo, en el comportamiento de su estimador histórico.

7.1. Análisis de resultados obtenidos

En un principio, aprovechando la desigualdad

$$|\rho_\phi(X) - \widehat{\rho}_\phi^N(X)| \leq \|\phi\|_{L^q} W_p(\mu_X, \hat{\mu}_X^N), \quad (7.1.1)$$

y los resultados de concentración conocidos para $W_p(\mu_X, \hat{\mu}_X^N)$, quisimos establecer resultados generales para $\rho_\phi(X)$, con ϕ un espectro de riesgo admisible, tan general como fuera posible. Si bien este objetivo fue logrado, en el sentido de que establecimos resultados explícitos para ϕ en $L^\infty((0, 1), dx)$ cuando la ley de X satisface $T_1(\lambda)$ (teorema 6.3.4) y para ϕ en $L^2((0, 1), dx)$ cuando la ley de X satisface $T_2(\lambda)$ (teorema 6.3.5), nos hubiera gustado que la cantidad de datos necesaria para que las desigualdades que proponen los teoremas fueran significativas fuera menor, de modo que los resultados que encontramos fueran realmente utilizables en la práctica financiera. Por ejemplo, el número de observaciones necesarias para utilizar el teorema 6.3.4 es del orden 61 510 000. Es decir, necesitaríamos más de 200 000 años de información histórica para poder aplicar este resultado. Sin embargo, un punto a favor de estos desarrollos es que son inmediatamente aplicables a cualquier estimador que satisfaga la desigualdad (7.1.1). Por esto, en un contexto donde haya una gran cantidad de datos disponibles, como al trabajar con simulaciones, nuestros resultados pueden ser directamente aplicados.

7.1. Análisis de resultados obtenidos

A lo largo del desarrollo de esta memoria, pudimos identificar dos razones para que los resultados mencionados en el párrafo anterior no funcionen todo lo bien que nos gustaría. La primera es que en algunos casos la cota (7.1.1) es demasiado gruesa. Por ejemplo, en el caso del ES_α , tenemos que $\phi = 1/\alpha \mathbb{1}_{[0,\alpha]}$ y entonces

$$\|\phi\|_{L^\infty} W(\mu_X, \hat{\mu}_X^N) - |ES_\alpha(X) - \widehat{ES}_\alpha^N(X)| \geq \frac{1}{\alpha} \int_\alpha^1 |q_X^-(s) - q_N^-(s)| ds,$$

donde se observa que a medida que α se acerca a 0, la cota se hace más y más gruesa. La segunda razón por la que la metodología anterior no es óptima, es que en los resultados de concentración utilizados $\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon)$ converge a cero mucho más lento de lo que se observa en experimentos numéricos para algunas distribuciones conocidas.

Para evitar los problemas mencionados en el párrafo anterior, intentamos dos cosas. La primera fue combinar el resultado de Brown, Gao y Wang con algunas ideas de la demostración del teorema de Bolley, Guillin y Villani (teorema 5.1.2) para la concentración de la medida empírica con respecto a la distancia de Wasserstein. Esto produjo el Teorema 6.3.15 que en la práctica se comporta mucho mejor que los resultados encontrados previamente.

El segundo cambio de estrategia, que fue motivado por los trabajos de Brown [10] y de Gao y Wang [21], fue tratar de obtener resultados mejores para la estimación de $ES_\alpha(X)$ aprovechando su representación dual. El producto de esta estrategia es el Teorema 6.3.19, que generaliza el resultado de Brown, Gao y Wang a variables aleatorias que toman valores en \mathbb{R} , es significativo para todo N , y en el caso de que se utilice para variables que toman valores en un intervalo compacto, recupera el resultado de Brown. Estos resultados nos indican que tratar de establecer teoremas generales pasando por acotar el error de estimación por la distancia de Wasserstein, quizás no sea el mejor camino.

Dada la importancia que tiene el $V@R_\alpha$ como medida de riesgo, e inspirados en la desigualdad (7.1.1), buscamos una desigualdad similar para esta medida de riesgo. Encontramos que era posible acotar el error de estimación del $V@R_\alpha(X)$, controlando la distancia de Lévy entre la ley de la muestra y la ley empírica de la muestra, lo que sumado a un resultado de concentración para la distancia de Prokhorov, nos permitió establecer el teorema 6.3.2 que entrega intervalos de confianza no asintóticos para el $V@R_\alpha(X)$ bajo distintas hipótesis para la ley de X .

Un punto importante que queda pendiente en esta memoria, es establecer tests estadísticos que permitan contrastar la hipótesis: “La ley de la variable aleatoria X satisface $T_1(\lambda)$ ”. En la sección 4.2 se presentó un resultado que dice que una ley μ satisface $T_1(\lambda)$ si y sólo si para algún $\beta > 0$,

$$\int_{\mathcal{X}} \int_{\mathcal{X}} e^{\beta d^2(x,y)} d\mu(x) d\mu(y) < +\infty.$$

Sin embargo, no es claro a priori como se pueden realizar tests sobre esta condición, pues para una muestra finita el momento exponencial cuadrático empírico será finito para todo β .

Para terminar, podemos resumir nuestras conclusiones en los siguientes puntos

- Se pueden establecer resultados muy generales, pero en la práctica pueden ser poco útiles por la gran cantidad de observaciones necesarias para que sean válidos.
- La desigualdad (7.1.1) puede ser extremadamente gruesa, pero podría ser útil para medidas espectrales cuyo espectro tenga un soporte más grande en $(0, 1)$.
- Los resultados de concentración para las medidas empíricas pueden ser engañosos, no siempre un decaimiento exponencial nos asegura un buen comportamiento en la práctica.
- Existen resultados de concentración para funcionales de la medida empírica que son aplicables a estimadores de medidas de riesgo.

7.2. Trabajo Futuro

En cuanto a trabajo futuro, existen varias direcciones que se pueden tomar. Una dirección apunta a levantar la hipótesis de independencia entre las variables que componen la muestra con que se construye la medida empírica. Para esto, sería posible considerar el caso Markoviano, para el cuál desigualdades de concentración similares a las establecidas por Gozlan y Léonard fueron desarrolladas por Guillin et al. en [25].

Otra dirección es tratar de mejorar las cotas para $\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon)$. En esta memoria presentamos desigualdades de la forma

$$\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq A_1 e^{-b_1 N} + A_2 e^{-b_2 N},$$

y encontramos que las constantes A_i resultaban ser muy grandes, mientras que las tasas de decaimiento b_i resultaban ser muy pequeñas. Tal vez, si abandonamos las cotas exponenciales, podamos encontrar desigualdades de la forma

$$\mathbb{P}(W(\mu, \hat{\mu}^N) > \varepsilon) \leq C_1 N^{-q_1} + C_2 N^{-q_2},$$

donde las constantes C_i no son tan grandes como las constantes A_i , de modo que para muestras pequeñas esta cota es más ajustada que la cota exponencial.

Una tercera posibilidad, que parece bastante prometedora en el corto plazo, es tratar de extender el Teorema de Brown, Gao y Wang. Recordemos que ellos descubrieron las siguientes cotas para variables que toman valores en un intervalo compacto

$$\mathbb{P}\left(\widehat{CV@R_\alpha}^N(X) \leq CV@R_\alpha(X) - \varepsilon\right) \leq 3e^{-\frac{\alpha}{5}\left(\frac{\varepsilon}{b-a}\right)^2 N}$$

y

$$\mathbb{P}\left(\widehat{CV@R_\alpha}^N(X) \geq CV@R_\alpha(X) + \varepsilon\right) \leq \min\left\{3e^{-\frac{\alpha}{11}\left(\frac{\varepsilon}{b-a}\right)^2 N}, e^{\left(\frac{-2N\alpha^2\varepsilon^2}{(b-a)^2}\right)}\right\}.$$

7.2. Trabajo Futuro

Con el teorema 6.3.19, ya hemos extendido la segunda cota a variables aleatorias con valores en \mathbb{R} , por lo que sólo nos falta extender la primera desigualdad. Esto sería una verdadera innovación, ya que hasta ahora el mejor resultado disponible en la literatura es el de Brown, Gao y Wang.

Por último, en la sección 5.1.2 se presentaron tres resultados de Gozlan y Léonard para la concentración de funcionales de la medida empírica. Como se mencionó en dicha sección, los resultados presentados por los autores en [23] son más generales. En particular, permiten considerar otras hipótesis de integrabilidad sobre la ley de la muestra y obtener otras cotas de concentración. Esta flexibilidad en los resultados de Gozlan y Leonard plantean la posibilidad de extender los resultados encontrados a medidas que satisfagan otras condiciones de integrabilidad.

Anexos

Apéndice A

Resultados pendientes en la demostración del Teorema 5.1.2

A.1. Teorema de Sion para Minimax

En este anexo se enuncia y prueba un teorema de minimax dado por Maurice Sion en [40]. La demostración que se presenta no corresponde a la dada en el artículo mencionado, si no a la dada por Hidetoshi Komiya en [28]. Antes de enunciar y probar el teorema se procederá a dar algunas definiciones y a probar dos lemas técnicos que nos serán útiles en la demostración del teorema principal.

Definición A.1.1. Consideremos X un espacio vectorial topológico y $f : X \rightarrow \mathbb{R}$ diremos que:

- f es cuasi-concava en X si para todo $\lambda \in \mathbb{R}$ el conjunto $S_\lambda = \{x \in X : f(x) \leq \lambda\}$ es un conjunto convexo.
- f es cuasi-concava en X si para todo $\lambda \in \mathbb{R}$ el conjunto $S^\lambda = \{x \in X : f(x) \geq \lambda\}$ es un conjunto convexo.

Proposición A.1.2. Sea X un espacio vectorial topológico. Entonces $f : X \rightarrow \mathbb{R}$ es cuasi-concava si y sólo si

$$\forall x, y \in X, \forall \gamma \in [0, 1], \quad f(\gamma x + (1 - \gamma)y) \geq \min\{f(x), f(y)\}.$$

Demostración: (\Rightarrow) Sin pérdida de generalidad podemos asumir que $f(x) \leq f(y)$, de donde se tiene que $y \in S^{f(x)}$, pero como este conjunto es convexo y trivialmente $x \in S^{f(x)}$, tenemos que $\forall \gamma \in [0, 1]$ el punto $\gamma x + (1 - \gamma)y \in S^{f(x)}$. Es decir,

$$f(\gamma x + (1 - \gamma)y) \geq f(x) = \min\{f(x), f(y)\}.$$

A.1. Teorema de Sion para Minimax

(\Leftarrow) Sea $\lambda \in \mathbb{R}$ y $x, y \in S^\lambda$ se quiere probar que $\forall \gamma \in [0, 1] \gamma x + (1 - \gamma)y \in S^\lambda$. Notemos que

$$f(\gamma x + (1 - \gamma)y) \geq \min\{f(x), f(y)\} \geq \lambda,$$

luego $\gamma x + (1 - \gamma)y \in S^\lambda$.

q.e.d.

En lo que sigue se considerará X es un subconjunto conexo y compacto de un espacio vectorial topológico y que Y es un espacio vectorial topológico. Se considerará $f : X \times Y \rightarrow \mathbb{R}$ que satisface:

1. $f(x, \cdot)$ es semicontinua superior (s.c.s.) y cuasi-concava en $Y, \forall x \in X$.
2. $f(\cdot, y)$ es semicontinua inferior (s.c.i.) y cuasi-convexa en $X, \forall y \in Y$.

A estas condiciones nos referiremos como las hipótesis H.

Lema A.1.3. *Si X, Y y f satisfacen las hipótesis H, entonces $\forall y_1, y_2 \in Y$ y $\forall \alpha \in \mathbb{R}$ que satisface*

$$\alpha < \inf_{x \in X} \max\{f(x, y_1), f(x, y_2)\},$$

existe $y_0 \in Y$ que satisface:

$$\alpha < \inf_{x \in X} f(x, y_0).$$

Demostración: Por contradicción supongamos que $\exists \alpha \in \mathbb{R}$ tal que

$$\alpha < \inf_{x \in X} \max\{f(x, y_1), f(x, y_2)\} =: c(y_1, y_2),$$

pero que para todo $y \in Y$

$$\alpha \geq \inf_{x \in X} f(x, y).$$

Sea β tal que $\alpha < \beta < c(y_1, y_2)$. Denotemos por $[y_1, y_2] = \{\gamma y_1 + (1 - \gamma)y_2 : \gamma \in [0, 1]\}$. Definimos para $z \in [y_1, y_2]$ los conjuntos:

$$Cz = \{x \in X : f(x, z) \leq \alpha\} \quad \text{y} \quad C'z = \{x \in X : f(x, z) \leq \beta\}.$$

Sea $A = C'y_1$ y $B = C'y_2$. Entonces se tiene lo siguiente:

1. $Cz \subset C'z$, pues $\alpha < \beta$.
2. $Cz, C'z$ son cerrados, pues $f(\cdot, z)$ es s.c.i.
3. $Cz, C'z$ son no vacíos, ya que por hipótesis de contradicción $\forall y \in Y \quad \alpha \geq \inf_{x \in X} f(x, y)$.
4. $Cz, C'z$ son convexos pues $f(\cdot, z)$ es cuasi-convexa.

A.1. Teorema de Sion para Minimax

5. $Cz, C'z$ son conexos, directo de lo anterior.
6. $A \cap B = \emptyset$. Si $x_0 \in A \cap B$, entonces $\max\{f(x_0, y_1), f(x_0, y_2)\} \leq \beta$ y luego $\beta > c(y_1, y_2)$, pero por construcción $\beta < c(y_1, y_2)$.
7. $C'z \subset A \cup B$. En efecto si $x \in C'z$ entonces como $f(x, \cdot)$ es cuasi concava se tiene que $\min\{f(x, y_1), f(x, y_2)\} \leq f(x, z) \leq \beta$ luego $f(x, y_1) \leq \beta \vee f(x, y_2) \leq \beta$, es decir $x \in A \vee x \in B$, luego $x \in A \cup B$.

Como A y B son conjuntos disjuntos y cerrados, y $C'z \subset A \cup B$ es conexo, se tiene que

$$Cz \subset C'z \subset A \quad \vee \quad Cz \subset C'z \subset B.$$

Además, si $Cz \subset A$ (resp. B), entonces $C'z \subset A$ (resp. B).

Sean:

$$I = \{z \in [y_1, y_2] : Cz \subset A\} \quad \text{y} \quad J = \{z \in [y_1, y_2] : Cz \subset B\}.$$

Entonces $I \cap J = \emptyset$, $I \cup J = [y_1, y_2]$. Luego si probamos que I y J son conjuntos cerrados tendremos una contradicción con la conexidad de $[y_1, y_2]$ y la prueba estaría completa. Probemos que I es cerrado:

Sea $\{z_n\} \subset I$ una sucesión convergente. Como $[y_1, y_2]$ es cerrado tenemos que $z_n \rightarrow z \in [y_1, y_2]$. Queremos probar que $z \in I$, para esto tomaremos $x \in Cz$ y probaremos que $x \in A$, luego $Cz \subset A$ y entonces por definición $z \in I$. Como $x \in Cz$, entonces $f(x, z) \leq \alpha$ y entonces como $f(x, \cdot)$ es s.c.s. se tiene que $\limsup_n f(x, z_n) \leq \alpha < \beta$. Entonces $\exists m \in \mathbb{N}$ tal que $f(x, z_m) \leq \beta$ es decir $x \in C'z_m \subset A$ pues $z_m \in I$, luego $x \in A \Rightarrow Cz \subset A \Rightarrow z \in I$. Luego I es cerrado. Para J es totalmente análogo.

q.e.d.

Lema A.1.4. Si X, Y y f satisfacen las hipótesis H , entonces para todo conjunto finito $\{y_1, \dots, y_n\} \subset Y$ y $\forall \alpha \in \mathbb{R}$ tal que

$$\alpha < \inf_{x \in X} \max_{1 \leq i \leq n} f(x, y_i),$$

entonces existe $y_0 \in Y$ tal que:

$$\alpha < \inf_{x \in X} f(x, y_0).$$

Demostración: Por inducción en n . Notemos que para $n = 1$ el lema es trivial, basta considerar $y_0 = y_1$. Supongamos ahora que el resultado es cierto para $n - 1$ y probemos que se tiene también para n . Sea $X' = \{x \in X : f(x, y_n) \leq \alpha\}$ como $f(\cdot, y_n)$ es cuasi-concava y s.c.i. en X tenemos que X' es un subconjunto compacto y convexo de X . Podemos además suponer que X' es no vacío, ya que si lo fuera basta con considerar $y_0 = y_n$ y se tendría el resultado.

Notemos que:

$$\alpha < \inf_{x \in X} \max_{1 \leq i \leq n} f(x, y_i) \leq \inf_{x \in X'} \max_{1 \leq i \leq n} f(x, y_i) = \inf_{x \in X'} \max_{1 \leq i \leq n-1} f(x, y_i),$$

A.1. Teorema de Sion para Minimax

donde la primera desigualdad se tiene de la hipótesis, la segunda porque se está tomando ínfimo sobre un conjunto más pequeño y la última igualdad se tiene porque para $x \in X'$, $f(x, y_n) \leq \alpha$ mientras que $\max_{1 \leq i \leq n} f(x, y_i) > \alpha$, y entonces el máximo se alcanza en y_{i^*} con $i^* \in 1, \dots, n-1$. Luego tenemos que

$$\alpha < \inf_{x \in X'} \max_{1 \leq i \leq n-1} f(x, y_i).$$

Ahora notemos que si consideramos la restricción de f a $X' \times Y$, entonces X', Y y f satisfacen las hipótesis H . Luego podemos aplicar la hipótesis de inducción y encontramos que existe $y'_0 \in Y$ tal que

$$\alpha < \inf_{x \in X'} f(x, y'_0).$$

Pero notemos que en $X \setminus X'$ se tiene que $f(x, y_n) > \alpha$, luego tenemos que

$$\alpha < \inf_{x \in X} \max\{f(x, y'_0), f(x, y_n)\}.$$

Finalmente, por el lema anterior existe $y_0 \in Y$ tal que

$$\alpha < \inf_{x \in X} f(x, y_0).$$

q.e.d.

Teorema A.1.5. *Si X, Y y f satisfacen las hipótesis H , entonces*

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

Demostración: Dado que la desigualdad

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \geq \sup_{y \in Y} \inf_{x \in X} f(x, y),$$

es siempre cierta, sólo es necesario probar la desigualdad en la otra dirección. Para esto sea α tal que

$$\alpha < \inf_{x \in X} \sup_{y \in Y} f(x, y). \tag{A.1.1}$$

Definamos como $X_y = \{x \in X : f(x, y) \leq \alpha\}$. Entonces $\forall y \in Y$, X_y es convexo y compacto, pues $f(\cdot, y)$ es cuasi-convexa y s.c.i.. Notemos además que como α satisface (A.1.1) se tiene que $\bigcap_{y \in Y} X_y = \emptyset$. Pero X es compacto, luego existen $\{y_1, \dots, y_n\} \in Y$ tales que

$$\bigcap_{i=1}^n X_{y_i} = \emptyset,$$

lo que es equivalente a que

$$\alpha < \inf_{x \in X} \max_{1 \leq i \leq n} f(x, y_i),$$

A.2. Entropía

y entonces por el lema anterior existe $y_0 \in Y$ de modo que

$$\alpha < \inf_{x \in X} f(x, y_0) \leq \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

Para concluir consideremos ahora $\{\alpha_n\}_n \subset \mathbb{R}$ una sucesión creciente que satisface

$$\alpha_n \nearrow \inf_{x \in X} \sup_{y \in Y} f(x, y),$$

entonces $\forall n \in \mathbb{N}$ tenemos que

$$\alpha_n < \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

Con lo que finalmente concluimos que

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \leq \sup_{y \in Y} \inf_{x \in X} f(x, y)$$

q.e.d.

A.2. Entropía

A.2.1. Representación Dual de la Entropía

Sea \mathcal{X} un espacio métrico polaco. Recordemos que $\mathcal{P}(\mathcal{X})$ es el espacio de medidas de probabilidad en \mathcal{X} y $C_b(\mathcal{X})$ es el conjunto de funciones continuas y acotadas en \mathcal{X} .

Lema A.2.1. *Sea $\mu \in \mathcal{P}(\mathcal{X})$. Se define: $\Lambda : C_b(\mathcal{X}) \rightarrow \mathbb{R}$ como*

$$\Lambda(\phi) = \log \left(\int_S e^{\phi} d\mu \right)$$

Entonces

$$\Lambda(\phi) = \sup_{\nu \in \mathcal{P}(\mathcal{X})} \{ \langle \nu, \phi \rangle - H(\nu|\mu) \}. \quad (\text{A.2.1})$$

En particular $\Lambda(\cdot)$ es una función convexa y semicontinua inferior.

Lema A.2.2. *Sea $\mu \in \mathcal{P}(\mathcal{X})$ y llamemos $\mathcal{P}_\mu(\mathcal{X})$ al conjunto de medidas de probabilidad absolutamente continuas con respecto a μ . Entonces para toda $\nu \in \mathcal{P}_\mu(\mathcal{X})$ se tiene que*

$$H(\nu|\mu) = \sup_{\phi \in C_b} \left\{ \int_S \phi d\nu - \Lambda(\phi) \right\}.$$

La demostración de estos resultados se puede encontrar en [15].

A.2.2. Entropía Métrica en $(\mathcal{P}(\mathcal{X}), W_p(\cdot, \cdot))$

Teorema A.2.3. *Sea (E, d) un espacio métrico con diámetro finito D . Para todo $r > 0$, sea $N(E, r)$ como el número mínimo de bolas de radio r necesarias para cubrir E . Entonces, para todo $p \geq 1$ y $\delta \in (0, D)$, el espacio de medidas de probabilidad sobre E , $\mathcal{P}(E)$ puede ser recubierto por $\mathcal{N}_p(E, \delta)$ bolas de radio δ para la distancia W_p , con*

$$\mathcal{N}_p(E, \delta) \leq \left(\frac{8eD}{\delta} \right)^{pN(E, \delta/2)}$$

Demostración: La demostración de este teorema se encuentra en el anexo A de [8].

Bibliografía

- [1] C. Acerbi. Spectral measures of risk : A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26:105–1518, 2002.
- [2] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487–1503, 2002.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, 2nd edition, 2008.
- [4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Thinking coherently. *RISK* 10, 1997.
- [5] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3), 1999.
- [6] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., USA, 1968.
- [7] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic sobolev inequalities. *Journal of Functional Analysis*, 163(1):1 – 28, 1999.
- [8] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2006.
- [9] F. Bolley and C. Villani. Weighted csizár-kullback-pinsker inequalities and applications to transportation inequalities. *Annales de la faculté des sciences de Toulouse*, 14(3):331–352, 2005.
- [10] D. B. Brown. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, 35(6):722 – 730, 2007.
- [11] F. Carlson. Une inégalité. *Ark. Mat. Astr. Fysik*, 1935.
- [12] A. Charpentier and A. Oulidi. Estimating allocations for value-at-risk portfolio optimization. *Mathematical Methods of Operations Research*, 69:395–410, 2009. 10.1007/s00186-008-0244-7.

- [13] R. Cont. Model uncertainty and its impact on the pricing of derivative instrument. *Mathematical Finance*, 16(3):519–547, July 2006.
- [14] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10:593–606, 2010.
- [15] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Springer-Verlag, New York, 2nd edition, 1998.
- [16] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *The Annals of Probability*, 32(3B):2702–2732, 2004.
- [17] R. M. Dudley. The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1968.
- [18] R. M. Dudley. *Real Analysis and Probability*. Number 74 in Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge, Reino Unido, 2nd edition, 2002.
- [19] D. Feldman and H. G. Tucker. Estimation of non-unique quantiles. *The Annals of Mathematical Statistics*, 37(2):451–457, 1966.
- [20] H. Föllmer and A. Shied. *Stochastic Finance: An introduction in Discrete Time*. Walter de Gruyter, 2004.
- [21] F. Gao and Y. Wang. Deviation inequalities for an estimator of the conditional value-at-risk. *Operations Research Letters*, 38(3):236 – 239, 2010.
- [22] N. Gozlan. Integral criteria for transportation-cost inequalities. *Electronic Communications in Probability*, 11:64–77, 2006.
- [23] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probability Theory and Related Fields*, 139(1-2):235–283, 2007.
- [24] N. Gozlan and C. Léonard. Transport inequalities. a survey. 03 2010.
- [25] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for markov processes. *Probability Theory and Related Fields*, 144:669–695, 2009. 10.1007/s00440-008-0159-5.
- [26] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., 2nd edition, 2009.
- [27] L. V. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4), 2006.

- [28] H. Komiya. Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.
- [29] L. Larsson, L. Malingranda, J. Pecaric, and L.E. Persson. *Multiplicative Inequalities of Carlson Type and Interpolation*. World Scientific Publishing Company, 2006.
- [30] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, 2nd edition, 1998.
- [31] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.
- [32] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inf. Theor.*, 32(3):445–446, 1986.
- [33] P. Massart. *Concentration Inequalities and Model Selection*. Number 1896 in Lectures Notes in Mathematics. Springer, 2007.
- [34] C. McDiarmid. On the method of bounded differences. *in: Surveys in Combinatorics*, 1989.
- [35] F. Otto and C. Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173:361–400, 200.
- [36] G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4), 2007.
- [37] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problem*, volume I and II. Springer, 1998.
- [38] R. T. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *Journal of Risk* 2, pages 493–517, 2000.
- [39] B. Rudloff. *Hedging in Incomplete Markets and Testing Compound Hypotheses via Convex Duality*. PhD thesis, Martin Luther University of Halle-Wittenberg, 2006.
- [40] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.
- [41] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric And Functional Analysis*, 6(3):587–600, 1996.
- [42] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.
- [43] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.