



INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



# Introducción elemental al problema de transporte

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ACTUARÍA

PRESENTA

JOAQUÍN SÁNCHEZ GARCÍA

ASESOR: DR. CESAR LUIS GARCÍA GARCÍA

CIUDAD DE MÉXICO.

2016

“Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “Introducción Elemental al problema de transporte”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.”

JOAQUÍN SÁNCHEZ GARCÍA

---

FECHA

---

FIRMA

# Índice general

<b>1. Historia e Introducción</b>	<b>1</b>
1.1. Introducción al problema de transporte . . . . .	1
1.2. Introducción histórica . . . . .	2
1.2.1. ¿Qué hizo Monge? . . . . .	2
1.2.2. ¿Qué hizo Kantorovich? . . . . .	2
1.2.3. El problema de Monge . . . . .	3
1.2.4. El problema de Kantorovich . . . . .	3
1.3. Desambiguación por capítulos . . . . .	4
<b>2. El problema real (<math>\mathbb{R}</math>) de Kantorovich</b>	<b>5</b>
2.1. El caso real . . . . .	5
2.1.1. El enfoque general de la cópula bivariada . . . . .	5
2.1.2. El enfoque en términos de variables aleatorias . . . . .	10
2.2. Comentarios sobre el capítulo . . . . .	14
<b>3. Estructura del espacio de medidas</b>	<b>15</b>
3.1. Norma Variación . . . . .	16
3.2. Distancia Wassersstein . . . . .	17
3.3. Distancia Prokhorov-Levy . . . . .	19
3.3.1. Otras métricas en espacios de medidas . . . . .	20
3.4. Topologías inducidas e identificación del espacio dual . . . . .	21
3.4.1. Lema Urysohn en espacios métricos . . . . .	22
3.4.2. Teorema de Riesz . . . . .	22
3.4.3. Enunciado del Teorema de Riesz . . . . .	23
3.4.4. Teorema de Portmanteau [6] p.16 . . . . .	27
3.4.5. Teorema Prokhorov [6] p. 59 . . . . .	28
3.4.6. Comentarios sobre el capítulo . . . . .	30
<b>4. Formulación y dualidad del problema de Kantorovich</b>	<b>31</b>
4.1. El Problema de Kantorovich . . . . .	31
4.2. Indicadoras Infinitas . . . . .	32
4.3. Transformada de Legendre-Frechet . . . . .	33
4.3.1. Desigualdad de Fenchel-Young . . . . .	33
4.3.2. La conjugada de la norma . . . . .	34

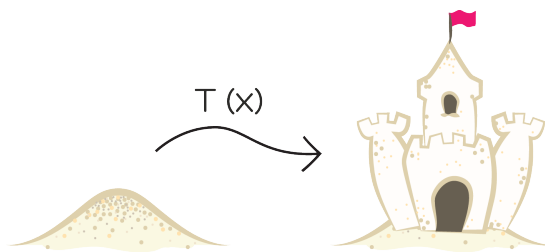
4.3.3.	Déficit esperado . . . . .	34
4.3.4.	El ejemplo fundamental de la transformada . . . . .	35
4.3.5.	La función convexa conjugada es en efecto convexa . . . . .	37
4.3.6.	La función convexa conjugada invierte el orden . . . . .	38
4.3.7.	Interpretación geométrica . . . . .	38
4.4.	Principios minimax . . . . .	39
4.4.1.	Algunas formulaciones de principios minimax . . . . .	39
4.5.	Diagrama de seguimiento . . . . .	45
4.6.	Dualidad del problema de Kantorovich . . . . .	46
4.6.1.	Teorema de dualidad del problema de Kantorovich . . . . .	46
4.6.2.	De la dualidad débil . . . . .	46
4.6.3.	Ideas generales y bosquejo de la prueba . . . . .	47
4.6.4.	Primer paso de la prueba . . . . .	47
4.6.5.	Segunda parte de la prueba: sobre la existencia del minimizador . . . . .	49
4.6.6.	Comentarios sobre el capítulo . . . . .	49
<b>5.</b>	<b>El problema de Monge</b> . . . . .	<b>50</b>
5.1.	Formulación del problema de Monge . . . . .	50
5.2.	Relación entre el problema de Monge y el de Kantorovich . . . . .	51
5.3.	Átomos y masas de Dirac . . . . .	52
5.3.1.	La otra perspectiva de las deltas de Dirac . . . . .	54
5.4.	Ejemplos básicos del problema de Monge . . . . .	54
5.4.1.	Múltiples soluciones . . . . .	54
5.4.2.	Dividir un átomo . . . . .	55
5.4.3.	Juntar dos átomos . . . . .	56
5.5.	La medida <i>push-forward</i> y el cambio de variable . . . . .	56
5.6.	Comentarios sobre el capítulo . . . . .	56
<b>6.</b>	<b>Determinación de cópulas y la solución de Kantorovich para medición de riesgo</b> . . . . .	<b>57</b>
6.1.	Medición y administración del Riesgo . . . . .	57
6.1.1.	Fenómenos de Colas Pesadas . . . . .	57
6.2.	Las cópulas en medición de riesgo . . . . .	57
6.3.	Cópulas de colas pesadas y supervivencia . . . . .	58
6.4.	Uso de la solución de Kantorovich para estimación de riesgo . . . . .	58
6.4.1.	Ejemplo 1 . . . . .	59
6.4.2.	Ejemplo 2 . . . . .	59
6.4.3.	Ejemplo 3 . . . . .	59
6.5.	Sobre la manera en la que se escoge la cópula . . . . .	60
6.5.1.	Enfoque paramétrico . . . . .	60
6.5.2.	Familias paramétricas de cópulas . . . . .	61
6.5.3.	Cópulas arquimedianas . . . . .	61
6.5.4.	Máxima Verosimilitud . . . . .	61
6.5.5.	Enfoque no paramétrico . . . . .	63
6.5.6.	Uso de la cópula no paramétrica . . . . .	65
6.6.	El enfoque mixto . . . . .	65
6.7.	Comentarios sobre la elección de cópulas . . . . .	65
6.8.	Reglas de decisión con la dualidad de Rockafellar . . . . .	66
6.8.1.	Complicaciones . . . . .	66
6.9.	Comentarios sobre el capítulo . . . . .	66

<b>7. El problema finito de Kantorovich</b>	<b>68</b>
7.1. El problema del Transporte natural . . . . .	68
7.2. Factibilidad del Problema de Transporte . . . . .	69
7.3. Algoritmos para el problema del transporte . . . . .	69
7.3.1. Ford-Fulkerson . . . . .	70
7.3.2. Simplex y la esquina noreste . . . . .	71
7.3.3. Caso Natural . . . . .	71
7.4. Conclusiones del capítulo . . . . .	71
<b>Apéndices</b>	<b>76</b>
<b>A. El cambio de variable en medidas</b>	<b>77</b>
<b>B. Axiomas de Separación</b>	<b>78</b>
<b>C. Completo y Totalmente Acotado</b>	<b>79</b>
<b>D. Regularidad automática de las medidas finitas en espacios métricos [21] p. 27</b>	<b>80</b>

# Historia e Introducción

## 1.1. Introducción al problema de transporte

Supongamos que tenemos una pila de arena y queremos construir un castillo. Pensemos que tenemos el diseño del castillo ya preestablecido, es decir, no vamos a reinventar el castillo durante el proceso de construcción. Pensemos también, que la arena está a una distancia significativa del lugar donde queremos construir el castillo, y que llevar la arena al lugar tiene algún costo.



Nos interesa trasladar la arena de manera óptima. La optimalidad radica en la minimización de costos y el objetivo es minimizar el costo total de transporte de arena. Puede ser que los costos dependan únicamente de la distancia recorrida por cada grano de arena o podemos tener costos tan generales como se quieran.

En el problema de traslado de arena surgen inmediatamente muchas preguntas:

- ¿Existen maneras óptimas de llevar la arena?
- ¿Bajo qué condiciones sobre el ambiente se puede asegurar que esa forma es óptima?
- ¿Qué funciones de costos admiten soluciones?
- ¿Se puede encontrar una fórmula que indique cómo se debe llevar la arena?
- ¿Cómo afecta el diseño del castillo la manera en la que transportamos la arena?
- ¿Cómo afecta la estructura de la pila original en el costo total de transporte?

Todas estas preguntas tienen respuestas y se contestarán a través del desarrollo de este trabajo. Nos enfocaremos a pensar en **espacios métricos**, que se supondrán **separables y completos**. El objetivo de esta tesis es hacer una introducción al problema de transporte óptimo, así como presentar algunos de los resultados más importantes en el área.

## 1.2. Introducción histórica

En 1781, Gaspard Monge, inventor de la geometría descriptiva y confidente de Napoleón, planteó el problema de transporte óptimo de manera conjuntista:

*Dados dos subconjuntos  $U, V \subseteq \mathbb{R}^3$  con el mismo volumen, encontrar un mapeo que conserve volumen entre ellos, minimizando a la vez una función de costo  $c(x, y)$ .*

Con esta formulación, se busca una función  $T : U \rightarrow V$  tal que la suma total de costos del tipo  $c(x, Tx)$  sea mínima. El primer conjunto se puede entender como la distribución inicial de masa. El segundo conjunto representa la posición final de cada partícula y  $c(x, y)$  se interpreta como el costo de transportar una unidad de la posición  $x$  a la posición  $y$ .

### 1.2.1. ¿Qué hizo Monge?

Gaspard Monge (1746-1818) fue uno de los grandes científicos de su época. Fue confidente de Napoleón, inclusive acompañándolo a expediciones en Egipto. Fue parte de la primera asamblea de la revolución francesa por lo que en 1989 (bicentenario de dicha revolución) el gobierno francés le rindió homenaje. El planteamiento original de su problema estaba relacionado a la ingeniería civil en la reconstrucción de algunas calles de París. Monge intentó resolver el problema para la función de costos euclidiana, es decir el costo de transporte era únicamente la distancia recorrida. Esta idea es natural en economía y física porque una vez que se tiene una herramienta para transportar el material, el único costo es la distancia que tiene que recorrer.

Pasaron doscientos años hasta que V.N. Sudakov demostró que el mapeo  $T$ , en realidad existe. La Academia de París ofreció un premio por la solución del problema. Leonid Kantorovich recibió el premio Nobel en economía en 1975 por su trabajo relacionado a este problema.

### 1.2.2. ¿Qué hizo Kantorovich?

Leonid Kantorovich (1912-1986) fue un matemático y economista de la Unión Soviética. Dedicó su vida al estudio de la asignación óptima de recursos y al análisis funcional. Junto con Tjalling Koopmans obtuvo un Premio Nobel en 1975. El premio se le otorgó por su trabajo en el Centro Soviético de Investigación Económica, donde mostró que es posible descentralizar las decisiones siempre que exista un sistema racional de precios.

Uno de los grandes logros de este matemático en el contexto de transporte óptimo fue “relajar” el problema de Monge, en el sentido que modeló el problema con una formulación más sencilla. Kantorovich logró que el problema pasara de un atascamiento no lineal a una pregunta soluble. El problema de Kantorovich se define formalmente más adelante, pero intuitivamente, dejó de pensar que las distribuciones eran totalmente ajenas y lo transformó en encontrar una distribución en otro espacio cuyas marginales son las distribuciones originales.



Figura 1.1: Leonid Kantorovich



### 1.2.3. El problema de Monge

En términos formales, el problema de Monge es:

$$\inf_{T_{\#}\mu=\nu} \left\{ \int_X c(x, Tx) d\mu \right\},$$

Donde  $T_{\#}\mu = \nu$  significa que  $T$  es una función de transporte, un *push forward* de  $\mu$  en  $\nu$ .

La definición de función de transporte es que para cualquier conjunto  $A$  en la  $\sigma$ -álgebra del espacio  $Y$ , se tiene que

$$T : X \rightarrow Y \text{ y } \mu(T^{-1}(A)) = \nu(A).$$

En algunos libros, se refieren a la misma definición en términos de medidas; en vez de decir que  $T$  es la función *push forward*, le llaman a  $\nu$  la **medida push forward** inducida por  $T$ .

El primer gran problema de esta formulación salta rápidamente a la vista: el problema no se resolverá con una función lineal en  $T$ , lo cual hace al espacio de búsqueda realmente grande y por tanto el problema se vuelve muy difícil de resolver.

### 1.2.4. El problema de Kantorovich

¿Cómo modificar el problema de Monge para poder obtener soluciones? Kantorovich “relajó” el problema para dar condiciones de optimalidad. El ingenio de Kantorovich fue dejar de buscar una función de transporte, y buscar una medida en el espacio producto que tenga como marginales  $\mu$  y  $\nu$ . Esta idea ha sido altamente explotada en economía, finanzas y actuaría, donde el concepto de cópulas ha avanzado de manera significativa en los últimos años. Entonces el problema de Kantorovich se puede formular como:

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) d\gamma \right\}.$$

Donde  $\Gamma(\mu, \nu) = \{\gamma \in M_1(X \times Y, S) : \pi_x(\gamma) = \mu, \pi_y(\gamma) = \nu\}$ .

En esta definición,  $M_1$  denota el conjunto de medidas de probabilidad definidas en el espacio producto y en este contexto  $\pi_x$  denota la proyección en la primera coordenada. El uso de las marginales se refiere a que si  $A \in S_X$  y  $B \in S_Y$  entonces  $\gamma(A \times Y) = \mu(A)$  y análogamente  $\gamma(X \times B) = \nu(B)$ . En este contexto  $S$  es la  $\sigma$ -álgebra producto,  $S_X$  y  $S_Y$  las correspondientes  $\sigma$ -álgebras en  $X$  y  $Y$ .

Este problema tiene características muy diferentes al anterior en términos de su solución. Sin embargo, permite encontrar una cota inferior del problema de Monge. Es decir, el problema Kantorovich es un relajamiento verdadero del problema de Monge. Todas estas afirmaciones se probarán más adelante.

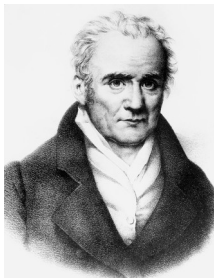


Figura 1.2: Gaspard Monge

### 1.3. Desambiguación por capítulos

- En el siguiente capítulo empezamos con la solución del problema en el caso real. Por lo mismo cuando el espacio son los números reales, la solución de problema de Kantorovich se obtiene a partir del estudio de cópulas. Las cópulas son funciones que están siendo muy utilizadas en diversas áreas de estadística, principalmente porque permiten relacionar de manera analítica dos o más variables aleatorias.
- En el capítulo 3, regresamos a la generalidad de los espacios de Banach. Se estudian los teoremas necesarios de convergencia de probabilidades, para el entendimiento de la formulación dual del problema de Kantorovich. Este capítulo contiene pruebas detalladas de los teoremas de Riesz, Portmanteau y Prokhorov así como el desarrollo de distancias en el espacio de probabilidades.
- En el capítulo 4 determinamos la existencia de la solución al problema de Kantorovich a partir de su formulación dual. A esta formulación se le denomina teorema de dualidad de Kantorovich; sin embargo, para entender en su totalidad esta prueba se debe hacer un repaso de análisis convexo y principios de minimax. En esta última sección se prueba el teorema de dualidad de Rockafellar.
- El capítulo quinto se dedica a mostrar la relación entre el problema de Kantorovich y el de Monge, dando intuición sobre las funciones de transporte. También, en este capítulo se incluyen ejemplos básicos del problema de Monge y conceptos como la delta de Dirac.
- El sexto capítulo adquiere un enfoque más aplicado, se basa en el uso del resultado del capítulo 2 para la ciencia actuarial. Una de las grandes críticas al uso de cópulas es el poco análisis que se suele hacer para determinar la cópula a implementar, por lo mismo este capítulo detalla aspectos prácticos de la determinación de cópulas.
- Uno de los grandes alcances del problema se basa en su generalidad: el problema del transporte óptimo se formula sobre un espacio de medida  $(X, S_X, \mu)$  a otro espacio de medida  $(Y, S_Y, \nu)$ . Sin embargo, muchas sustituciones en estos espacios llevan a problemas diferentes. El problema desde su perspectiva en las probabilidades discretas es interesante por sí mismo y se desarrolla a detalle en el capítulo siete. Hay dos enfoques en su formulación, los cuales se detallan posteriormente. Este problema ha sido estudiado en el área de investigación de operaciones, por lo que existen algoritmos y análisis de la solución.

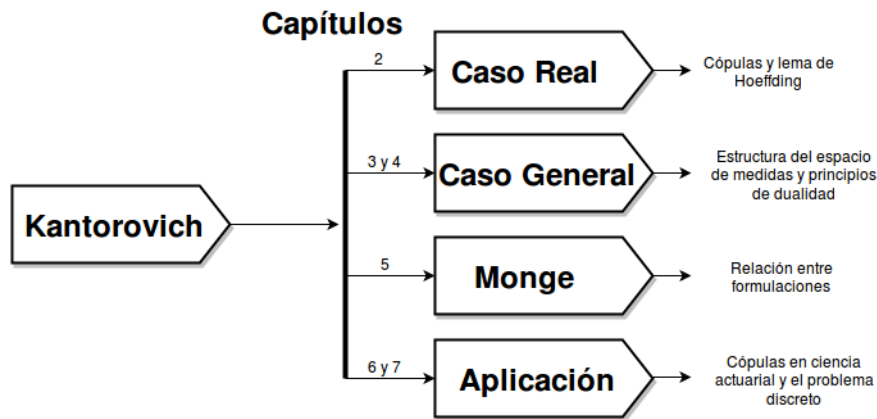


Figura 1.3: Estructura seguida

# El problema real ( $\mathbb{R}$ ) de Kantorovich

## 2.1. El caso real

En este capítulo se desarrolla el problema de Kantorovich en el espacio de los números reales. Es evidente la importancia de este caso por sí mismo. El enfoque en el caso real se basa en el análisis de cópulas.

### 2.1.1. El enfoque general de la cópula bivariada

Las cópulas son funciones bivariadas que retienen toda la información de dependencia entre variables aleatorias. En los últimos años, las cópulas han tenido un desarrollo significativo; son usadas en problemas de finanzas, administración del riesgo, estadística, teoría de valores extremos, probabilidad y seguros entre otros. Sin embargo, hay muchos peligros en el uso **inadvertido** de las cópulas. La teoría de cópulas es inmaculada, el problema viene en lo que se sabe hasta ahora respecto a su implementación. Primeramente, parece no haber un algoritmo óptimo de selección de cópulas, por lo que pueden ser ineficientes. Además, el uso de cópulas imposibilita el análisis estocástico del problema. Para más detalles en las desventajas del uso de cópulas se recomienda [18]. La implementación de cópulas es un problema estadístico abierto a investigación. En el capítulo 6 se desarrollan algunas ideas en esta dirección.

**Definición.** *Cópula bivariada*

Una cópula bivariada es una función  $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$  tal que

1.  $C(x, 0) = C(0, x) = 0$  y  $C(1, x) = C(x, 1) = x, \forall x \in [0, 1]$
2.  $C$  es creciente de manera simultánea en ambas coordenadas en el sentido que si  $a \leq b, c \leq d$  entonces

$$V_C([a, b] \times [c, d]) := C(b, d) - C(a, d) - C(b, c) + C(a, c) \geq 0.$$

Las cópulas se utilizan para “acoplar” variables aleatorias sujetas a algún tipo de dependencia. Esto permite la existencia de una infinidad de cópulas.

La segunda condición es para pensar en monotonía en términos de ambas coordenadas.  $V_C$  se puede entender como la medida del área de un rectángulo inducido por  $C$ . En efecto, si se utiliza la cópula  $C(x, y) = xy$  (independencia), la segunda condición induce el área usual. Esta segunda condición se denomina “condición de 2-monotonidad” o “2-creciente” [19] o “Schur-creciente”. Esta condición puede generalizarse para cópulas en el “n-cubo”. Intuitivamente, la condición permite que el avance en una dirección no decrezca el valor de la cópula con respecto a las demás direcciones. Más adelante, en 2.1.2, notaremos que esta condición es suficiente para que las cópulas sean funciones **uniformemente continuas**.

**Lemma 2.1.1.** *Toda cópula es  $\|\cdot\|_1$ -Lipschitz con constante uno.*

*Demostración.*  $|C(u_1, v_1) - C(u_2, v_2)| \leq |C(u_1, v_1) - C(u_2, v_1)| + |C(u_2, v_1) - C(u_2, v_2)|$ .

Analicemos el sumando de la izquierda.

Pensemos el caso  $u_2 \leq u_1$ . Por la propiedad de “2-monotonicidad” tenemos:

$$C(u_1, 1) + C(u_2, v_1) - C(u_1, v_1) - C(u_2, 1) \geq 0 \Rightarrow u_1 - u_2 \geq C(u_1, v_1) - C(u_2, v_1).$$

Considerando ambos casos obtenemos la desigualdad en valor absoluto y para el segundo sumando usamos la misma prueba pero con  $v_1 - v_2$ . De donde deducimos

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_1 - u_2| + |v_1 - v_2| = \|(u_1, v_1) - (u_2, v_2)\|_1. \quad \blacksquare$$

**Proposición 2.1.2.** *Las cópulas son uniformemente continuas.*

*Demostración.* Como todas las normas son equivalentes en dimensión finita, también  $C$  es  $\|\cdot\|_2$ -Lipschitz pero con constante  $\sqrt{2}$ . Dado  $\epsilon$  positivo, consideramos  $\delta < \epsilon/\sqrt{2}$  y obtenemos:

$$\|(u_1, v_1) - (u_2, v_2)\|_2 < \delta \Rightarrow |C(u_1, v_1) - C(u_2, v_2)| < \epsilon,$$

donde  $\delta$  no depende de los valores de  $(u_1, v_1)$  ni  $(u_2, v_2)$  por lo que se concluye la continuidad uniforme.  $\blacksquare$

**Teorema 2.1.3.** *Método de construcción de medidas exteriores [11] pp. 148-149.*

*Sea  $X$  un conjunto y sea  $\mathcal{C}$  una familia de subconjuntos que cubren a  $X$  y contenga al vacío. Sea  $V : \mathcal{C} \rightarrow [0, \infty]$  una función, entonces hay una única medida exterior  $\mu^*$  tal que*

1.  $\mu^*(A) \leq V(A) \quad \forall A \in \mathcal{C}$
2. Si  $\nu^*$  es otra medida exterior en  $\mathcal{P}(X)$  tal que  $\nu^*(A) \leq V(A) \quad \forall A \in \mathcal{C}$   
 $\Rightarrow \nu^*(B) \leq \mu^*(B) \quad \forall B \in \mathcal{P}(X)$ .

*Demostración.* Para probar unicidad es suficiente notar que si existieran dos medidas finitas cumpliendo (1) y (2) entonces se tendrían ambas desigualdades y por lo tanto se concluiría la igualdad. Por lo tanto enfoquémonos en probar la existencia.

La estrategia de prueba es un método constructivo muy utilizado en teoría de la medida, se repite este argumento para que los capítulos sean tan autocontenidos como se pueda y porque haremos el argumento sobre el conjunto  $\mathcal{C}$  que cubre a  $X$ . Esta construcción suele hacerse sobre álgebras pero esta generalización nos ayudará a probar el teorema.

Sea  $B \in \mathcal{P}(X)$  definimos

$$\mu^*(B) = \inf \left\{ \sum_{A \in D} V(A) \right\},$$

donde el ínfimo se considera sobre todas las posibles cubiertas numerables  $D$  de  $B$ , por elementos de  $\mathcal{C}$ . Como  $\mathcal{C}$  es una familia arbitraria de conjuntos que cubre a  $X$ , esta puede ser no numerable. Por lo que puede haber subconjuntos de  $X$  que no tengan cubiertas numerables. Sin embargo, para este caso la definición también funciona bajo la convención de que el ínfimo del conjunto vacío es  $\infty$ . Por ende, si un conjunto no tiene ninguna cubierta numerable por elementos de  $\mathcal{C}$  entonces la medida exterior de ese conjunto sería infinito.

Veamos que  $\mu^*$  es una medida exterior:

$\mu^*(\emptyset) = 0$ , debido a que el vacío cubre al vacío y la suma vacía es cero.

Si  $E \subseteq F$ , entonces cualquier  $\mathcal{C}$ -cubierta numerable de  $F$  es  $\mathcal{C}$ -cubierta numerable de  $E$ . Por lo que el conjunto en el que se considera el ínfimo es más grande para  $E$ , ya que pueden existir  $\mathcal{C}$ -cubiertas numerables que cubran a  $E$  pero no a  $F$ , y así se deduce que  $\mu^*(E) \leq \mu^*(F)$ .

Sea  $A_1, A_2, \dots$  una sucesión de subconjuntos de  $X$ , queremos probar la  $\sigma$ -subaditividad de  $\mu^*$ ; es decir,

$$\mu^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

Si  $\sum_{k=1}^{\infty} \mu^*(A_k) = \infty$ , entonces la desigualdad se cumple de manera trivial pues el lado derecho es infinito.

Suponemos entonces que todos los conjuntos de la sucesión son de medida finita. Sea  $\epsilon > 0$ , para cada  $n$  en los naturales, se escoge una  $\mathcal{C}$ -cubierta numerable, digamos  $D_n$  (por definición de ínfimo) tal que

$$\sum_{A \in D_n} V(A) < \mu^*(A_n) + 2^{-n} \epsilon$$

Definimos  $D = \bigcup_{n=1}^{\infty} D_n$ , como  $D_n$  cubre a  $A_n$ ,  $D$  cubre a la unión y necesariamente

$$\mu^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{A \in D} V(A) \leq \sum_{n=1}^{\infty} \sum_{A \in D_n} V(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \sum_{n=1}^{\infty} 2^{-n} \epsilon = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon$$

La primera desigualdad se debe a que  $D$  es una cubierta de la unión por lo que es uno de los elementos en el conjunto sobre el que se considera el ínfimo. La segunda desigualdad se debe a que si hay un conjunto que está en distintas cubiertas  $D_i$  y  $D_j$ , ese se repite al sumar ordenadamente. La tercera desigualdad se obtiene a partir de la propiedad que le exigimos a las cubiertas y la última igualdad es sólo calcular el valor de la serie. Como  $\epsilon$  es arbitrario se obtiene la  $\sigma$ -subaditividad.

Se concluye que  $\mu^*$  es una medida exterior.

Para probar la propiedad (1.), observamos que si  $A \in \mathcal{C}$ , el conjunto  $D^* = \{A\}$  es una  $\mathcal{C}$ -cubierta numerable de  $A$ . Entonces como la medida exterior es el ínfimo sobre las  $\mathcal{C}$ -cubiertas numerables,

$$\mu^*(A) \leq \sum_{B \in D^*} V(B) = V(A).$$

Para probar la propiedad (2.), sea  $B \in \mathcal{P}(X)$ , si  $\mu^*(B) = \infty$  la desigualdad es clara, por lo que nos remitimos al caso  $\mu^*(B) < \infty$ .

Sea  $B_n$  cualquier  $\mathcal{C}$ -cubierta numerable de  $B$ , entonces como  $\nu^*$  es una medida exterior

$$\nu^*(B) \leq \nu^* \left( \bigcup_{n=1}^{\infty} B_n \right) \leq \sum_{n=1}^{\infty} \nu^*(B_n) \leq \sum_{n=1}^{\infty} V(A_n),$$

y como el lado derecho es uno de los elementos del conjunto sobre el que se considera el ínfimo, entonces se tiene que

$$\nu^*(B) \leq \mu^*(B).$$

Este lema construye una medida exterior a partir de una función positiva arbitraria definida en un conjunto que cubre a  $X$ . Sea  $A^{\mu^*}$  el conjunto donde  $\mu^*$  es aditiva en el sentido de Caratheodory. Usamos el Teorema de extensión de Caratheodory-Hopf [14] y entonces  $\mu = \mu^*|_{A^{\mu^*}}$  es una medida en la  $\sigma$ -álgebra  $A^{\mu^*}$  ■

**Proposición 2.1.4.** [7] pp.13 Sea  $\mathcal{A}$  un álgebra y  $V : \mathcal{A} \rightarrow [0, \infty)$  una función **aditiva**. Supongamos que existe una familia de conjuntos compactos  $\mathcal{K}$  tal que para todo  $\epsilon > 0$  existen  $A_\epsilon \in \mathcal{A}$  y  $K_\epsilon \in \mathcal{K}$  tales que  $A_\epsilon \subseteq K_\epsilon \subseteq A$  y se tiene  $V(A \setminus A_\epsilon) < \epsilon$ . Entonces  $V$  es  $\sigma$ -aditiva.

*Demostración.* Sea  $A_1, A_2, \dots$  una sucesión disjunta de conjuntos en  $\mathcal{A}$ . Como  $V$  es aditiva tenemos que

$$\sum_{i=1}^N V(A_i) \leq V\left(\bigcup_{i=1}^N A_i\right) \quad \forall N \in \mathbb{N}$$

$$\sum_{i=1}^{\infty} V(A_i) \leq V\left(\bigcup_{i=1}^{\infty} A_i\right)$$

Sea  $N$  un número natural, definimos:  $F_k = \bigcup_{i=N+k}^{\infty} A_i$ . Claramente la sucesión de conjuntos  $(F_k)$  es

decreciente y además  $\bigcap_{k=1}^{\infty} F_k = \emptyset$ , probemos que  $V(F_k) \rightarrow 0$ . Sea  $\epsilon > 0$  dado, entonces por hipótesis existen  $A_{\epsilon, n} \in \mathcal{A}$  y  $K_{\epsilon, n} \in \mathcal{K}$  tales que  $A_{\epsilon, n} \subseteq K_{\epsilon, n} \subseteq F_n$  con  $V(F_n \setminus A_{\epsilon, n}) < \epsilon 2^{-n}$ .

Como  $\bigcap_{n=1}^{\infty} K_{\epsilon, n} \subseteq \bigcap_{n=1}^{\infty} F_n = \emptyset$  y los  $K_{\epsilon, n}$  son conjuntos compactos, existe  $M$  tal que  $\bigcap_{n=1}^M K_{\epsilon, n} = \emptyset$ .

Observamos que  $\bigcap_{n=1}^N F_n \subseteq \bigcup_{n=1}^N (F_n \setminus A_{\epsilon, n})$  y de ahí que  $V\left(\bigcup_{n=1}^N F_n\right) \leq \sum_{n=1}^{\infty} V(F_n \setminus A_{\epsilon, n}) \leq \sum_{n=1}^N \epsilon 2^{-n} \leq \epsilon$ .

Finalmente para todo  $\epsilon_2 > 0$  tenemos que  $V\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} V(A_n) + \epsilon_2$ . ■

**Corolario 2.1.5.** Toda cópula bivariada induce una medida.

Así pues, dada una cópula  $C$ , la función  $V_C$  definida anteriormente es una función positiva sobre el conjunto  $\mathcal{C} = \mathcal{A}\{[a, b] \times [c, d] : a \leq b, c \leq d, a, b, c, d \in [0, 1]\}$  que es un álgebra de subconjuntos de  $[0, 1] \times [0, 1]$ , por el método de construcción de medidas anterior  $V_C$  define una medida, pero definir esta medida puede estar definida en un conjunto en un conjunto inútil por lo que debemos probar que  $V_C$  es una casi-medida para asegurar  $\mathbb{B}_{\mathbb{R}^2} \subseteq \mathcal{A}^{\mu^*}$ .

Es claro que  $V_C$  es aditiva pero para ver que es  $\sigma$ -aditiva usaremos la proposición anterior. Como la cópula es uniformemente continua ver dado  $\epsilon > 0$  sea  $\delta > 0$  tal que si  $\|(u, v) - (u_2, v_2)\| < \delta$  entonces  $|C(u, v) - C(u_2, v_2)| < \frac{\epsilon}{6}$ . Para usar la notación del ejercicio anterior, escribimos:

$$A = [a, b] \times [c, d]$$

$$K_\epsilon = [a, b - \delta] \times [c, d - \delta]$$

$$A_\epsilon = [a, b - \delta] \times [c, d - \delta]$$

De este modo,  $V_C(A \setminus A_\epsilon) = C(b - \delta, d) + C(a, d - \delta) - C(a, d) - C(b - \delta, d - \delta) + C(b - \delta, d - \delta) + C(b, d) - C(b - \delta, d) - C(b, d - \delta) + C(b, d - \delta) + C(b - \delta, c) - C(b - \delta, d - \delta) - C(b, c) < 6(\epsilon/6) = \epsilon$ .

$$\therefore V_C(A \setminus A_\epsilon) < \epsilon.$$

Por la proposición anterior obtenemos que  $V_C$  es una casi-medida por lo que aplica para la segunda parte del Teorema de Caratheodory-Hopf y así se asegura que las medidas generadas por las cópulas son de utilidad al intentar medir conjuntos. ■

Parece ser demasiada construcción para argumentar la existencia de medidas inducidas por cópulas pero es de fundamental importancia hacerlo así. Recordemos que queremos resolver el problema de Kantorovich, que involucra medidas en el espacio producto, utilizaremos la teoría de cópulas para generar desigualdades convenientes y por tanto es bueno tener el camino para transitar entre cópulas y medidas. Todo este desarrollo nos da una fórmula para generar una medida a partir de una cópula bivariada. Tenemos que considerar la restricción a los conjuntos Caratheodory-medibles de

$$\mu_C^*(A) = \inf \left\{ \sum_{i=1}^{\infty} C(b_i, d_i) + C(a_i, c_i) - C(a_i, d_i) - C(b_i, c_i) \right\},$$

donde el ínfimo se toma sobre todas las cubiertas de  $A$  de la forma  $\bigcup_{i=1}^{\infty} [a_i, b_i] \times [c_i, d_i]$ .

**Observación 2.1.6.** ¿Qué pasa si sustituímos  $C(u, v) = uv$ ? La condición de 2-monotonidad se convierte en

$$C(b, d) + C(a, c) - C(b, c) - C(a, d) = bd + ac - bc - ad = (b - a)(d - c),$$

Por lo que el problema de encontrar la medida inducida por  $C$  es el llamado “problema difícil de la medida en  $\mathbb{R}^2$ ” si se utiliza como  $\sigma$ -álgebra la potencia de los números reales y se pide  $\sigma$  aditividad. La construcción que se hizo es una de las posibles generalizaciones de la medida de Lebesgue-Stieltjes en dos dimensiones.

**Observación 2.1.7.** Más adelante se identificará cada cópula en términos de una función de distribución. Esta idea será más sencilla que la construcción que hemos hecho sobre medidas. El objetivo de esta sección es hacer la construcción en el mismo contexto que Kantorovich para ver directamente el uso de las cópulas. También es importante notar que a pesar de que las cópulas están definidas en  $[0, 1]$  en la mayoría de los casos se pueden evaluar en todos los números reales.

El uso de cópulas resulta fundamental en el análisis del problema de Kantorovich en  $\mathbb{R}$  por lo que sería ideal poder encontrar cotas superiores e inferiores.

**Proposición 2.1.8.** *Cotas de Frechét-Hoeffding*  
Si  $C$  es una cópula bivariada, entonces

$$\max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\}$$

*Demostración.* La prueba es sencilla usando la condición de 2-monotonidad. Para la cota inferior:

$$\begin{aligned} C(1, 1) + C(u, v) - C(1, v) - C(u, 1) &\geq 0 \\ 1 + C(u, v) - u - v &\geq 0 \\ C(u, v) &\geq u + v - 1. \end{aligned}$$

De lo anterior junto con  $C(u, v) \geq 0$ , se obtiene la primera desigualdad. Para la segunda desigualdad se hace una prueba casi idéntica:

$$\begin{aligned} C(u, 1) + C(0, v) - C(0, 1) - C(u, v) &\geq 0 \\ u + 0 - 0 - C(u, v) &\geq 0 \\ u &\geq C(u, v). \end{aligned}$$

La prueba para  $v$  es análoga, por lo que  $C(u, v) \leq u, C(u, v) \leq v \Rightarrow C(u, v) \leq \min\{u, v\}$ . ■

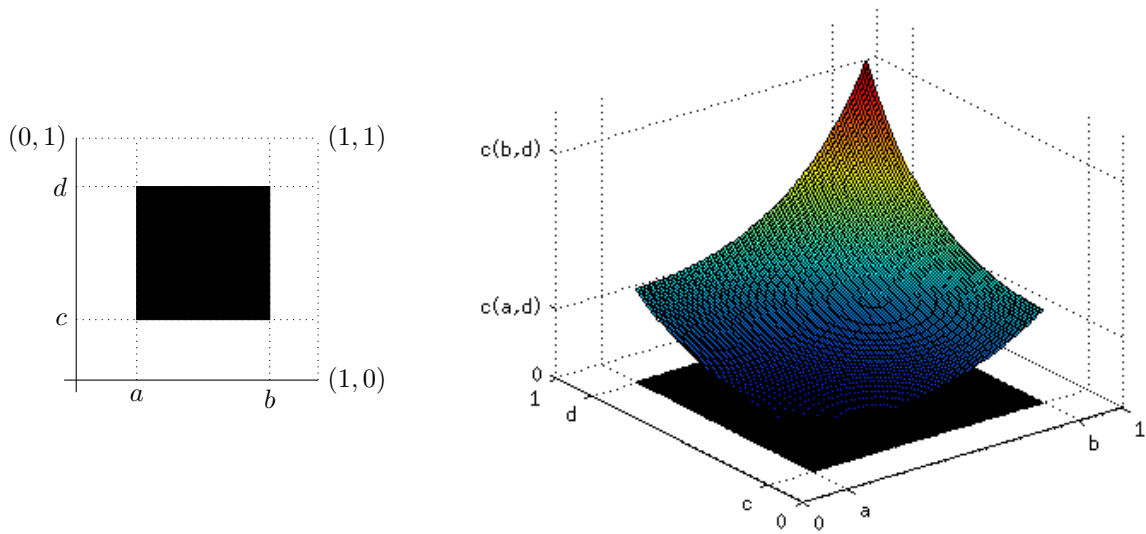


Figura 2.1: Área de la superficie inducida por una cópula

### 2.1.2. El enfoque en términos de variables aleatorias

Como se mencionó al principio, una cópula bivariada sirve para “acoplar” variables aleatorias. Esto se debe a que las cópulas agregan la relación de dependencia que se quiera. Pero cuando tenemos variables aleatorias, tenemos directamente distribuciones asociadas a cada una de ellas.

En la práctica, cuando se usan cópulas, se consideran las distribuciones de las variables como conocidas y se busca una cópula que modele la relación de interdependencia. Es, por tanto, fundamental estudiar el enfoque de cópulas en términos de las distribuciones.

**Proposición 2.1.9.** Sean  $F_1$  y  $F_2$  distribuciones de probabilidad, entonces para cualquier cópula  $C$ , la composición  $C(F_1, F_2)$  es una distribución bivariada.

*Demostración.* Las funciones de distribución para variables aleatorias tienen como contradominio  $[0, 1]$  por lo que la evaluación  $x \rightarrow F(x)$  nos arroja algo que podemos evaluar en la cópula. Así debido a la continuidad de  $C$ , calculamos los límites a infinito y por la continuidad de  $C$ :

$$\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} C(F_1(x), F_2(y)) = C\left(\lim_{x \rightarrow \infty} F_1(x), \lim_{y \rightarrow \infty} F_2(y)\right) = C(1, 1) = 1$$

$$\text{y } \lim_{\substack{x \rightarrow -\infty \\ y \rightarrow -\infty}} C(F_1(x), F_2(y)) = C\left(\lim_{x \rightarrow -\infty} F_1(x), \lim_{y \rightarrow -\infty} F_2(y)\right) = C(0, 0) = 0$$

Además la condición de 2-monotonicidad implica la condición que se les pide a las funciones de distribución bivariada (generalización de continuas por la derecha), por lo que se deduce que la cópula evaluada en las marginales es una distribución. ■



**Teorema.** (Abel Sklar) [19] p.18

Si  $H(x, y)$  es una distribución conjunta de probabilidad para las variables  $\mathbf{X} \stackrel{d}{=} F_1, \mathbf{Y} \stackrel{d}{=} F_2$  entonces existe una cópula  $C$  tal que

$$H(x, y) = C(F_1(x), F_2(y))$$

*Demostración.* El objetivo es construir una cópula,  $C$ , cuya evaluación, como en la proposición anterior, resulte igual a  $H$ . Las cópulas están definidas en  $[0, 1] \times [0, 1]$ , sin embargo  $\text{ran}(F_1) \times \text{ran}(F_2)$  en general no es el espacio completo. Procedemos a definir una cópula sobre  $\text{ran}(F_1) \times \text{ran}(F_2)$ , y luego extenderla de manera continua al conjunto deseado.

Sin pérdida de generalidad podemos suponer que  $\mathbf{X}, \mathbf{Y}$  están definidas en todo  $\mathbb{R}$  por simplicidad, ya que en otro caso los límites cambian pero la prueba es análoga; queremos aprovechar la información del mapeo:  $(x, y) \rightarrow (F_1(x), F_2(y))$ , sin embargo pueden ocurrir varias cosas:

1. Pueden existir varios puntos con imagen idéntica, a pesar de que las funciones de distribución son no-decrescentes esto no impide que dos elementos del dominio tengan asignado el mismo valor.
2. Puede ocurrir que las funciones  $F_1, F_2$  tengan saltos (*i.e.* correspondan a variables aleatorias discretas).

Los puntos en  $\text{ran}(F_1) \times \text{ran}(F_2)$  los podemos escribir de la forma  $(F_1(x), F_2(y))$ . A un punto con la forma  $(F_1(x), F_2(y))$  le asignamos el valor  $H(x, y)$ .

Esto es,  $C(F_1(x), F_2(y)) := H(x, y)$ .

El primer caso se resuelve al probar que la extensión continua de  $C$  es una cópula porque en virtud de la proposición (1.6) tenemos:

$$|C(F_1(x), F_2(y)) - C(F_1(u), F_2(v))| \leq |F_1(x) - F_1(u)| + |F_2(y) - F_2(v)| = 0,$$

por lo que la cópula estaría bien definida.

En el segundo caso, usamos cualquier extensión continua sobre los puntos que no estén en el rango de la función.

Así, usamos el Teorema de extensión de Tietze y obtenemos  $C$  en  $[0, 1] \times [0, 1]$  una función continua.

La propiedad de 2-monotonicidad se obtiene de la definición de  $H$  y para la otra condición tenemos:

$$\begin{aligned} C(1, a) &= H(\lim_{x \rightarrow \infty} F_1(x), a) = a = H(a, \lim_{y \rightarrow \infty} F_2(y)) = C(a, 1) \\ C(0, a) &= H(\lim_{x \rightarrow -\infty} F_1(x), a) = 0 = H(a, \lim_{y \rightarrow -\infty} F_2(y)) = C(a, 0). \end{aligned}$$

Observamos que si  $F_1$  y  $F_2$  son continuas, no se da la patología 2, por lo que no es necesario usar el teorema de extensión en ningún punto del interior de  $[0, 1] \times [0, 1]$ . Por lo anterior, la fórmula que usamos determina  $C$  de manera única. ■

## Ejemplos de Cópulas

Existen una infinidad de cópulas. En el capítulo 6 se detallan algunas familias, sin embargo aquí se presentan algunos ejemplos:

1. Cópula de independencia.

$$C(u, v) = uv.$$

2. Cópula de Frechét (o máxima)

$$C(u, v) = \min\{u, v\}.$$

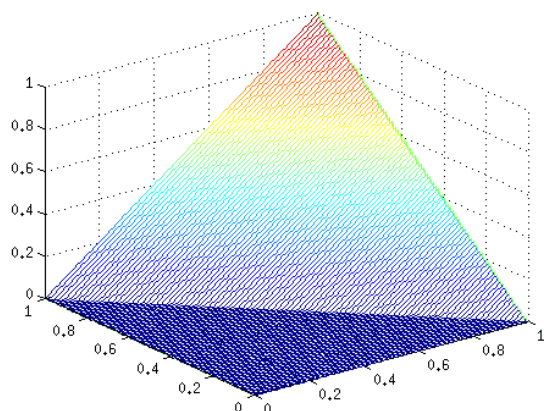
3. Cópula mínima.

$$C(u, v) = \max\{u + v - 1, 0\}.$$

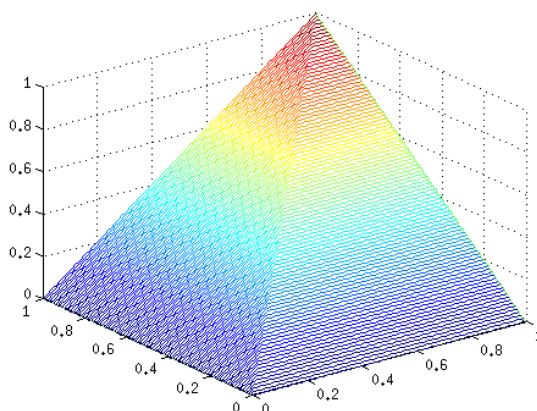
4. Cópulas de Clayton - Son la familia de cópulas que para algún  $\theta > 0$  se pueden escribir:

$$C_\theta(u, v) = (\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{\frac{1}{\theta}}.$$

De hecho, las cópulas de independencia y de Fréchet se pueden obtener como límites en  $\theta$  de la cópula de Clayton (a 0 y a  $\infty$  respectivamente).



(a) Cópula mínima



(b) Cópula máxima

Figura 2.2: Gráfica de la cópula mínima y cópula máxima

**¿Qué hemos logrado hasta este momento?** Usaremos todas estas herramientas para el problema de Kantorovich en  $\mathbb{R}$ ; hemos construido una medida a partir de cualquier cópula (Corolario 2.1.2) y hemos encontrado una manera de relacionar cópulas con distribuciones (Proposición 2.1.9). El objetivo será atacar el problema de Kantorovich desde el punto de vista de las variables aleatorias y usar sus distribuciones para encontrar cópulas y medidas óptimas.

Para resolver el problema de Kantorovich, sólo nos falta un resultado.

**Definición.** *Condiciones de Monge*

Si  $c(x, y)$  es una función bivariada que satisface:

1.  $-c$  es 2-monótona.
2.  $c$  es continua por la derecha en ambas coordenadas.

se dice que  $c$  es una función de costos que satisface las condiciones de Monge.

En lugar de hacer el desarrollo sobre funciones que satisfacen las condiciones de Monge, restringimos el trabajo sobre las cópulas, con el propósito de utilizar todo lo que hemos desarrollado en términos de cópulas. Las condiciones de Monge son las propiedades mínimas que se les exigen a las funciones de costos para tener soluciones. La definición puede hacerse con  $c$  o con  $-c$  siempre y cuando se tenga una convención sobre considerar costos positivos o negativos.

**Lemma 2.1.10.** (*Lema de Hoeffding*)

Sean  $\mathbf{X}, \mathbf{Y}$  variables aleatorias con distribución conjunta  $F$  y con distribuciones marginales  $F_1, F_2$  respectivamente.

Pensemos que  $\mathbf{E}[\mathbf{XY}] < \infty, \mathbf{E}[\mathbf{X}^2] < \infty, \mathbf{E}[\mathbf{Y}^2] < \infty$  entonces se tiene que

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = \int \int (F(x, y) - F_1(x)F_2(y)) dx dy$$

**Lemma 2.1.11.** (*Lema de Hoeffding generalizado*)

Sea  $F$  una distribución en  $\mathbb{R}^2$  con marginales  $F_1$  y  $F_2$ . Pensemos que los vectores aleatorios  $(\mathbf{X}_1, \mathbf{Y}_1) \stackrel{d}{=} F, \mathbf{X}_2 \stackrel{d}{=} F_1, \mathbf{Y}_2 \stackrel{d}{=} F_2$  y pensemos que  $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)$  son independientes. Consideremos una cópula  $C$ . Además supongamos que  $\mathbf{E}[C(\mathbf{X}_1, \mathbf{Y}_1)], \mathbf{E}[C(\mathbf{X}_2, \mathbf{Y}_2)] < \infty$  entonces se tiene que:

$$\mathbf{E}[C(\mathbf{X}_2, \mathbf{Y}_2)] - \mathbf{E}[C(\mathbf{X}_1, \mathbf{Y}_1)] = \int (F(x, y) - F_1(x)F_2(y)) d\mu_C^*(x, y).$$

*Demostración.* La prueba que se detalla a continuación es muy interesante. Esta demostración es análoga a la prueba que se utiliza para probar el lema de Hoeffding (sin generalizar). La veracidad de la prueba radica en el Teorema de Fubini.

Escribamos  $\mathbb{V}_1 = (\mathbf{X}_1, \mathbf{Y}_1), \mathbb{V}_2 = (\mathbf{X}_2, \mathbf{Y}_2)$ .

Definimos  $\phi_{\mathbb{V}_1, \mathbb{V}_2}(u, v) = \left( \chi_{(-\infty, \mathbf{X}_1]}(u) - \chi_{(-\infty, \mathbf{X}_2]}(u) \right) \left( \chi_{(-\infty, \mathbf{Y}_1]}(v) - \chi_{(-\infty, \mathbf{Y}_2]}(v) \right)$ .

Por ser producto de diferencias de indicatoras,  $\phi$  sólo toma valores en  $\{-1, 0, 1\}$ , analicemos los casos:

1.  $\phi = 1$  cuando ambos son positivos o ambos son negativos, es decir si simultáneamente  $u \in (-\infty, \mathbf{X}_1] \wedge u \notin (-\infty, \mathbf{X}_2]$  con  $v \in (-\infty, \mathbf{Y}_1] \wedge v \notin (-\infty, \mathbf{Y}_2]$ .  
Pero esta condición es sólo posible si  $\mathbf{X}_2 < \mathbf{X}_1$  y  $\mathbf{Y}_2 < \mathbf{Y}_1 \Rightarrow (u, v) \in (\mathbf{X}_2, \mathbf{X}_1] \times (\mathbf{Y}_2, \mathbf{Y}_1]$ .
2.  $\phi = 0$  cuando alguno de los dos factores es cero, esto se da si  $u < \min\{\mathbf{X}_1, \mathbf{X}_2\}$  o  $v < \min\{\mathbf{Y}_1, \mathbf{Y}_2\}$  o  $u > \max\{\mathbf{X}_1, \mathbf{X}_2\}$  o  $v > \max\{\mathbf{Y}_1, \mathbf{Y}_2\}$ .
3.  $\phi = -1$  se da en el caso contrario a  $\phi = 1$  respecto a las desigualdades de  $\mathbf{X}_1$  con  $\mathbf{X}_2$  y  $\mathbf{Y}_1$  con  $\mathbf{Y}_2$ .

Entonces  $\phi \neq 0$  sólo en el rectángulo  $[\min\{\mathbf{X}_1, \mathbf{X}_2\}, \max\{\mathbf{X}_1, \mathbf{X}_2\}] \times [\min\{\mathbf{Y}_1, \mathbf{Y}_2\}, \max\{\mathbf{Y}_1, \mathbf{Y}_2\}]$  y  $\phi$  alterna sus signos dependiendo de cuáles son máximos y mínimos.

Como  $\mu_C^*$  está definida sobre rectángulos, y como  $\phi$  alterna signos:

$$C(\mathbf{X}_1, \mathbf{Y}_1) + C(\mathbf{X}_2, \mathbf{Y}_2) - C(\mathbf{X}_1, \mathbf{Y}_2) - C(\mathbf{X}_2, \mathbf{Y}_1) = \int \phi d\mu_C^*(x, y).$$

$$\mathbf{E}[C(\mathbf{X}_1, \mathbf{Y}_1) + C(\mathbf{X}_2, \mathbf{Y}_2) - C(\mathbf{X}_1, \mathbf{Y}_2) - C(\mathbf{X}_2, \mathbf{Y}_1)] = \mathbf{E} \left[ \int \phi d\mu_C^* \right].$$

Aplicamos el teorema de Fubini en el lado derecho para intercambiar la esperanza con la integral y en el lado izquierdo usamos la linealidad del operador esperanza.

Recordemos que  $\mathbf{E}[\chi_A] = \mathbb{P}(\mathbf{X} \in A)$ . Además  $\chi_A \chi_B = \chi_{A \cap B}$  y desarrollemos la esperanza de  $\phi$ .

$$\begin{aligned} \phi_{\mathbb{V}_1, \mathbb{V}_2}(u, v) &= \left( \chi_{(-\infty, \mathbf{X}_1]}(u) - \chi_{(-\infty, \mathbf{X}_2]}(u) \right) \left( \chi_{(-\infty, \mathbf{Y}_1]}(v) - \chi_{(-\infty, \mathbf{Y}_2]}(v) \right) \\ &= \chi_{(-\infty, \mathbf{X}_1] \times (-\infty, \mathbf{Y}_1]}(u, v) + \chi_{(-\infty, \mathbf{X}_2] \times (-\infty, \mathbf{Y}_2]}(u, v) \\ &\quad - \chi_{(-\infty, \mathbf{X}_1] \times (-\infty, \mathbf{Y}_2]}(u, v) - \chi_{(-\infty, \mathbf{X}_2] \times (-\infty, \mathbf{Y}_1]}(u, v) \end{aligned}$$

Además,  $\chi_{(-\infty, \mathbf{z}]}(v) = 1 \Leftrightarrow \mathbf{Z} \geq v \Rightarrow \mathbf{E}[\chi_{(-\infty, \mathbf{z}]}(v)] = \mathbb{P}(\mathbf{Z} \geq v)$ , usando esta propiedad podemos calcular la esperanza directamente, y usando independencia:

$$\begin{aligned} \mathbf{E}[\phi(u, v)] &= \mathbb{P}(\mathbf{X}_1 \geq u, \mathbf{Y}_1 \geq v) + \mathbb{P}(\mathbf{X}_2 \geq u, \mathbf{Y}_2 \geq v) - \mathbb{P}(\mathbf{X}_1 \geq u, \mathbf{Y}_2 \geq v) - \mathbb{P}(\mathbf{X}_2 \geq u, \mathbf{Y}_1 \geq v) \\ &= 1 - F(u, v) + 1 - F(u, v) - 1 + F_1(u)F_2(v) - 1 + F_1(u)F_2(v) = -2(F(u, v) - F_1(u)F_2(v)). \end{aligned}$$

Sustituimos este resultado en la ecuación a la que le aplicamos el teorema de Fubini y obtenemos:

$$\mathbf{E}[C(\mathbf{X}_2, \mathbf{Y}_2) - C(\mathbf{X}_1, \mathbf{Y}_2)] = \int (F(u, v) - F_1(u)F_2(v)) d\mu_C^*(u, v). \quad \blacksquare$$

**Condiciones para poder usar Fubini:** En virtud de poder usar el Teorema de Fubini tenemos que asegurar que la esperanza de  $|\phi|$  sea finita.

$$\begin{aligned} \mathbf{E}\left[\int |\phi| d\mu_C^*\right] &= \mathbf{E}[|C(\mathbf{X}_1, \mathbf{Y}_1) + C(\mathbf{X}_2, \mathbf{Y}_2) - C(\mathbf{X}_1, \mathbf{Y}_2) - C(\mathbf{X}_2, \mathbf{Y}_1)|] \\ &\leq 4 \max\{\mathbf{E}[|C(\mathbf{X}_1, \mathbf{Y}_1)|], \mathbf{E}[|C(\mathbf{X}_2, \mathbf{Y}_2)|]\} < \infty. \end{aligned}$$

**Observación 2.1.12.** Al sustituir, en el Lema de Hoeffding generalizado, la cópula de independencia  $c(u, v) = uv$  se obtiene el Lema de Hoeffding. Por lo que mostramos que la proposición anterior sí es su generalización en sentido estricto.

**Observación 2.1.13.** Sea  $\mathbf{U}$  una variable aleatoria uniforme en  $(0, 1)$ , entonces si  $F$  es una función de distribución,  $F^{-1}(\mathbf{U}) \stackrel{d}{=} F$ . Donde  $F^{-1}$  puede entenderse en el sentido generalizado.

**Teorema.** (Solución al problema de Kantorovich en  $\mathbb{R}$ ).

Mismas hipótesis como el lema anterior

$$\int_0^1 c(F_1^{-1}(u), F_2^{-1}(u)) du \leq \mathbf{E}[c(\mathbf{X}_1, \mathbf{Y}_1)] \leq \int_0^1 c(F_1^{-1}(u), F_2^{-1}(1-u)) du$$

*Demostración.* El teorema es la conclusión directa del teorema de cambio de variable para medidas que demostramos en el apéndice. Esta observación se debe a Kolmogorov. Y es la que nos permite observar la fórmula explícita de la solución de Kantorovich en  $\mathbb{R}$ .  $\blacksquare$

Hemos encontrado el valor de la integral en la solución, si nuestro objetivo fuera encontrar la función de probabilidad sólo hay que usar la técnica del principio del capítulo para describirla. Nótese que dependiendo si los costos se consideraron como negativos o no, la solución sería el lado derecho o el lado izquierdo respectivamente.

## 2.2. Comentarios sobre el capítulo

La solución del problema de Kantorovich en  $\mathbb{R}$  nos indujo al estudio de cópulas. Las cópulas están siendo utilizadas en finanzas y administración de riesgos, sin embargo, una de las grandes dificultades en términos de la aplicabilidad de cópulas es que no existe un algoritmo **óptimo** de selección de cópulas. En general, no se sabe que cópula escoger, por lo mismo en el capítulo 6 se detallan algunas ideas de la teoría de selección de cópulas. En el capítulo 6 se hace este desarrollo.

No obstante, en el desarrollo de este capítulo no se usó la elección de ninguna cópula, sino que se obtuvo a partir de la función de costos que se considera dada. En el caso en el que no conociéramos una función de costos, buscar una función que sea óptima en algún sentido es el mismo problema de escoger cópulas.

# Capítulo 3

## Estructura del espacio de medidas

El problema del transporte relaciona dos medidas en función de la optimización de una función de costos. Por lo que es necesario tener conocimiento sobre la estructura del espacio de medidas. El problema de Monge y el de Kantorovich nos dan relaciones entre medidas.

El problema de Monge relaciona dos medidas a partir de una función de transporte. Nos preguntamos cuál es la mejor manera de “llevar” (en algún sentido) una medida a la otra. Un transporte óptimo nos indica el camino ideal para transportar cada unidad de masa desde un acomodo a otro.

El problema de Kantorovich relaciona dos medidas a partir de otra medida en el espacio producto. En el caso de medidas de probabilidad, nos indica cuál es la mejor manera de acomodar dos probabilidades como marginales en una probabilidad, según cada función de costos.

Ambos casos nos ilustran la necesidad de medir separación entre medidas. El problema del transporte mide esta separación en función de los costos, por lo que es evidente la obligación de contestar las preguntas:

- ¿Cómo podemos medir distancias entre medidas?
- ¿Hay distintas maneras de hacerlo?
- ¿Qué estructura adquiere el espacio de medidas?

Más formalmente, para el estudio del problema de Kantorovich se usará la idea del problema dual. El problema dual requiere del entendimiento del espacio dual topológico. El espacio dual topológico definido como el espacio de funcionales lineales nos permite relacionar espacios normados de manera directa. En el problema de Kantorovich se usa el Teorema de Riesz (que se probará en este capítulo) para identificar el espacio dual topológico del espacio de funciones continuas de  $X$  con soporte compacto con el espacio de medidas de Borel. Por lo que para entender la prueba de dualidad del siguiente capítulo, necesitamos:

1. Darle una norma al espacio de medidas definidas sobre un espacio métrico  $(X, d)$
2. Entender el teorema de Riesz para identificar las funcionales lineales
3. Identificar la topología débil\* en el dual del espacio de medidas

El objetivo de este capítulo es responder estas preguntas. A lo largo del capítulo, se desarrollarán métricas en el espacio de medidas y luego se hace un breve resumen de las topologías que inducen.

### 3.1. Norma Variación

El objetivo de esta sección es dar una norma en el espacio de medidas de probabilidad, pero para el enfoque de la norma variación se necesita hacer el desarrollo a partir de medidas con signo. Queremos definir una norma en el conjunto de medidas con signo, sin embargo hay que hacer la observación que el espacio no constituye un espacio vectorial. Las medidas con signo no pueden tomar ambos valores  $\infty$  y  $-\infty$  al mismo tiempo, por lo que la suma de medidas con signo no es necesariamente una medida con signo. Para resolver este problema, se consideran únicamente medidas con signo finitas. Como nuestro tema principal son las medidas de probabilidad no tenemos problema al suponer esto. A partir de este momento, las **particiones** se considerarán **medibles**.

**Definición.** Sea  $\mu$  una medida finita con signo en un espacio medible  $(X, S)$ .

Sean  $\mu_+$  y  $\mu_-$  respectivamente, la variación positiva y negativa de la descomposición de Jordan (Jordan-Hahn) de  $\mu$ . Es decir  $\mu = \mu_+ - \mu_-$

Se define la variación total de  $\mu$  como  $|\mu| = \mu_+ + \mu_-$

A la cantidad  $|\mu|(X)$  se le llama norma de variación.

**Lemma 3.1.1.** Para toda medida con signo  $\mu$ ,  $|\mu|(X) = \max \left\{ \sum_{i=1}^{\infty} |\mu(E_i)| : \{E_i\} \text{ es una partición numerable de } X \right\}$

*Demostración.* Sea  $(A|B)$  una descomposición de Hahn para  $\mu$  entonces,

$$\begin{aligned} \sum_{i=1}^{\infty} |\mu(E_i)| &= \sum_{i=1}^{\infty} |\mu(E_i \cap A) + \mu(E_i \cap B)| \leq \sum_{i=1}^{\infty} |\mu(E_i \cap A)| + \sum_{i=1}^{\infty} |\mu(E_i \cap B)| \\ &= \sum_{i=1}^{\infty} \mu_+(E_i) + \sum_{i=1}^{\infty} \mu_-(E_i) = \sum_{i=1}^{\infty} |\mu|(E_i) = |\mu|(X) \end{aligned}$$

Y de hecho el máximo se alcanza para la partición disjunta  $A \cup B$  dado que,

$$|\mu|(X) = \mu_+(X) + \mu_-(X) = \mu_+(A) + \mu_-(B) = |\mu(A)| + |\mu(B)|$$

**La norma variación define una norma en el espacio de medidas con signo finitas**

#### 1. No negatividad

Si  $\mu$  es una medida con signo en  $(X, S)$ , tanto  $\mu_+$  como  $\mu_-$  son medidas (positivas) por lo que  $\mu_+(X) + \mu_-(X) > 0$ . Además,  $|\mu|(X) = 0 \Leftrightarrow \mu_+(X) + \mu_-(X) = 0 \Leftrightarrow \mu_+(X) = 0 = \mu_-(X)$ .

Por monotonía  $\mu_+(A) = 0 = \mu_-(A) \forall A \in S \Leftrightarrow \mu \equiv 0$ .

#### 2. Homogeneidad

Sea  $k > 0$  entonces  $k\mu = k\mu_+ - k\mu_-$  con  $k\mu_+$  y  $k\mu_-$  medidas positivas, por lo que  $|k\mu|(X) = k\mu_+(X) + k\mu_-(X) = k(\mu_+(X) + \mu_-(X)) = k|\mu|(X)$ .

En el caso contrario, si  $k < 0$  entonces  $k\mu = k\mu_+ - k\mu_- = (-k)\mu_- - (-k)\mu_+$  donde  $(-k)\mu_-$  y  $(-k)\mu_+$  son medidas positivas. Además, como  $k < 0 \Rightarrow |k| = -k$  por lo que podemos escribir la variación total de  $k\mu(X) = |k|\mu_-(X) + |k|\mu_+(X)$ . Por lo que se concluye el caso general,  $|k\mu|(X) = |k||\mu|(X)$

#### 3. Desigualdad triangular

Sea  $(A|B)$  una descomposición de Hahn para  $\mu + \nu$  en virtud del Lema (3.1.1) tenemos,

$$|\mu + \nu|(X) = |\mu(A) + \nu(A)| + |\mu(B) + \nu(B)| \leq |\mu(A)| + |\mu(B)| + |\nu(A)| + |\nu(B)| \leq |\mu|(X) + |\nu|(X)$$

Como toda norma induce una distancia, en particular en el espacio de probabilidades obtenemos la distancia:

$$\|\mu - \nu\| = |\mu - \nu|(X)$$

## 3.2. Distancia Wasserstein

El concepto de distancia de Wasserstein es fundamental en el desarrollo del problema de transporte, porque la definición está formulada en términos del problema del Kantorovich. Sea  $X$  un espacio métrico con una distancia  $d$ . Para  $p \in [1, \infty)$  sea  $P_p(X)$  el conjunto de medidas de probabilidad con momentos finitos de orden  $p$ . Definimos:

$$W_p(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{X \times X} d(x, y)^p d\gamma \right\} \right)^{\frac{1}{p}}.$$

Es decir, la distancia entre  $\mu$  y  $\nu$  es el mínimo costo de transporte según la función de costos  $c(x, y) = d(x, y)^p$ .

1. **No negatividad.** Como  $d(x, y)$  es no negativa,  $d(x, y)^p$  es no negativa y por lo tanto la integral respecto a  $\gamma$  es siempre no negativa. Al considerar el ínfimo el resultado es un valor no negativo, es decir:

$$W_p^p(\mu, \nu) \geq 0.$$

2. **Identidad de Indiscernibles** En el capítulo 5 se prueba que el problema de Kantorovich es la relajación del problema de Monge por lo que:

$$W_p^p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{X \times X} d(x, y)^p d\gamma \right\} \leq \inf_{T_{\#}\mu = \nu} \left\{ \int d(x, Tx)^p d\mu \right\}.$$

Si queremos analizar  $W_p^p(\mu, \mu)$ , podemos pensar en la función de transporte  $T(x) = x$  entonces el problema de Monge tiene solución y se obtiene que

$$d(x, Tx) = d(x, x) = 0 \Rightarrow W_p^p(\mu, \mu) = 0.$$

Para el recíproco, supongamos  $W_p^p(\mu, \nu) = 0$  entonces

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{X \times X} d(x, y)^p d\gamma \right\} = 0,$$

pero en el capítulo 4 probaremos que existe un plan óptimo  $\pi \in \Gamma(\mu, \nu)$  tal que el ínfimo se convierte en mínimo. Entonces para  $\pi$

$$\int_{X \times X} d(x, y)^p d\pi = 0.$$

Pero como  $d$  es una métrica en  $X$ , sólo toma valores distintos de cero fuera de la diagonal  $D = \{(x, x) : x \in X\}$  por lo que, esta condición se cumple si  $\pi$  es cualquier probabilidad en el espacio producto que esté concentrada totalmente en la diagonal. La diagonal  $D$  es **medible** debido a que  $X$  es un espacio métrico separable: porque  $D = d^{-1}(\{0\})$  que es una función continua y por lo mismo  $D$  es cerrado y de ahí que esté en la  $\sigma$ -álgebra. Por lo mismo si  $f$  es cualquier función  $\mu$  integrable con dominio en  $X$  entonces

$$\int_X f(x) d\mu(x) = \int_{X \times X} f(x) d\pi(x, y) = \int_D f(x) d\pi(x, y) = \int_X f(y) d\nu(y).$$

En particular el resultado se tiene para indicadoras y entonces  $\mu = \nu$ .

### 3. Simetría

Si  $\gamma \in \Gamma(\mu, \nu)$  entonces podemos pensar en el espacio “al revés”, en vez de considerar  $X \times Y$  analizamos  $Y \times X$ . Entonces  $\gamma \in \Gamma(\nu, \mu)$  por lo que el costo total en el problema de Kantorovich es totalmente simétrico, no así en el problema de Monge. Esta discusión se extenderá en el capítulo 5.

4. **Desigualdad triangular** Sean  $\mu_1, \mu_2, \mu_3$  medidas de probabilidad y pensemos que  $\gamma_{1,2} \in \Gamma(\mu_1, \mu_2)$ ,  $\gamma_{2,3} \in \Gamma(\mu_2, \mu_3)$  y sean  $X_1, X_2, X_3$  los soportes de las medidas respectivamente. Supongamos que existe  $\pi$  una medida definida en  $X_1 \times X_2 \times X_3$  tal que sus proyecciones satisfagan:  $\pi_{X_1, X_2}(\pi) = \gamma_{1,2}$ ,  $\pi_{X_2, X_3}(\pi) = \gamma_{2,3}$ . Y denotemos  $\pi_{1,3} = \pi_{X_1, X_3}(\pi)$

$$W_p^p(\mu_1, \mu_3) \leq \int_{X_1 \times X_3} d(x_1, x_3)^p d\pi_{1,3}(x_1, x_3) = \int_{X_1 \times X_2 \times X_3} d(x_1, x_3)^p d\pi(x_1, x_2, x_3)$$

Donde el primer paso se debe a la definición de ínfimo. La igualdad se debe a que  $\pi_{1,2}$  es la marginal de  $\pi$  en el espacio  $X_1, X_3$  y la variable  $x_2$  no aparece en el argumento.

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left( \int_{X_1 \times X_2 \times X_3} (d(x_1, x_2) + d(x_2, x_3))^p d\pi(x_1, x_2, x_3) \right)^{1/p} \\ &\leq \left( \int_{X_1 \times X_2 \times X_3} d(x_1, x_2) d\pi(x_1, x_2, x_3) \right)^{1/p} + \left( \int_{X_1 \times X_2 \times X_3} d(x_2, x_3) d\pi(x_1, x_2, x_3) \right)^{1/p} \end{aligned}$$

En el primer paso, usamos la desigualdad del triángulo para  $d$  y la monotonía de la integral. Para el segundo paso se usó la desigualdad de Minkowski.

$$W_p(\mu_1, \mu_2) \leq \left( \int_{X_1 \times X_2} d(x_1, x_2) d\gamma_{1,2} \right)^{1/p} + \left( \int_{X_2 \times X_3} d(x_2, x_3) d\gamma_{2,3} \right)^{1/p}$$

Se usó el hecho de que  $\gamma_{1,2}$  y  $\gamma_{2,3}$  son marginales de  $\pi$  por lo que tomando ínfimos dos veces del lado derecho, se obtiene el resultado deseado:

$$W_p(\mu_1, \mu_3) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$$

Por lo tanto,  $W_p(\cdot, \cdot)$  define una métrica en el espacio de medidas de probabilidad. ■

Empezamos a entender un poco más, la riqueza de la estructura del espacio de medidas de probabilidad.

**Observación 3.2.1.** Definimos la métrica de Wasserstein para  $p \in [1, \infty)$  pero puede hacerse de manera análoga para  $p \in (0, 1)$ , pero en vez de utilizar  $d$  en la métrica de  $X$ , utilizando  $\bar{d} = \min\{d, 1\}$  para asegurar la desigualdad del triángulo.

**Observación 3.2.2.** La métrica de Wasserstein también se puede definir para  $p = 0$  pero usando la métrica discreta  $d(x, y) = \chi_{\{x=y\}}$



### 3.3. Distancia Prokhorov-Levy

**Definición.** En un espacio métrico  $(X, d)$  se define la distancia de un punto  $x$ , a un conjunto  $A$  no vacío, como

$$d(x, A) = \inf\{d(x, y) : y \in A\}.$$

**Observación 3.3.1.** La función  $d_A : X \rightarrow [0, \infty)$  dada por  $d_A(x) = d(x, A)$  es una función uniformemente continua pues

$$|d(x, A) - d(z, A)| \leq d(x, z)$$

La distancia Prokhorov-Levy es muy importante en teoría de convergencia de medidas por la topología que induce. Además, se puede dar una representación en términos de funciones de distribución de variables aleatorias. Para un subconjunto  $A$  de  $X$ , se define el conjunto  $A^\epsilon$  como el conjunto de puntos  $\epsilon$ -ceranos a  $A$ .

**Definición.** Conjunto  $\epsilon$ -cerano.

Sea  $A$  un conjunto, si  $d$  es la métrica en  $X$ , entonces el conjunto de puntos  $\epsilon$ -ceranos a  $A$  es:

$$A^\epsilon = \{y \in X : d(y, A) \leq \epsilon\}.$$

**Observación 3.3.2.**  $A^\epsilon$  es medible debido a que es la imagen inversa de el conjunto cerrado  $[0, \epsilon]$  bajo la función continua  $d_A$

**Observación 3.3.3.** Todo punto en  $\bar{A}$  satisface que  $d(x, A) = 0$ . Entonces podemos escribir a la cerradura de  $A$  como una intersección numerable, usando la propiedad arquimediana tenemos:

$$\bar{A} = \bigcap_{n=1}^{\infty} A^{1/n}.$$

*Demostración.*  $x \in \bar{A} \Leftrightarrow d(x, A) = 0 < \frac{1}{n} \forall n \in \mathbb{N} \Leftrightarrow x \in A^{1/n} \forall n \in \mathbb{N} \Leftrightarrow x \in \bigcap_{n=1}^{\infty} A^{1/n}$ .

Además, se satisface que  $A^{\epsilon_1} \subseteq A^{\epsilon_2}$  siempre que  $\epsilon_1 \leq \epsilon_2$ .

En virtud de los conjuntos  $\epsilon$ -ceranos definimos la distancia Prokhorov-Levy entre medidas de probabilidad regulares (Apéndice D):

$$d_P(\mu, \nu) = \inf \left\{ \epsilon > 0 : \mu(A) - \nu(A^\epsilon) \leq \epsilon \text{ y } \nu(A) - \mu(A^\epsilon) \leq \epsilon \forall A \text{ cerrado} \right\}.$$

#### 1. No negatividad

La no negatividad de  $d$  se obtiene observando únicamente el posible rango de valores de  $\epsilon$  que se pueden considerar.

El ínfimo se considera sobre valores  $\epsilon > 0$  por lo que  $d_P(\mu, \nu) \geq 0$ .

#### 2. Simetría

La simetría se tiene directamente, debido a que el orden en la condición sobre las medidas no es importante, es decir

$$[\mu(A) - \nu(A^\epsilon) \leq \epsilon \text{ y } \nu(A) - \mu(A^\epsilon) \leq \epsilon] \Leftrightarrow [\nu(A) - \mu(A^\epsilon) \leq \epsilon \text{ y } \mu(A) - \nu(A^\epsilon) \leq \epsilon].$$

### 3. Identidad de indiscernibles

Si  $d_P(\mu, \nu) = 0$  y sea  $A$  cualquier conjunto cerrado entonces  $\forall n \in \mathbb{N}, \mu(A) - \nu(A^{1/n}) \leq \frac{1}{n}$  y también

$$\nu(A) - \mu(A^{1/n}) \leq \frac{1}{n}$$

Consideramos límites en ambas condiciones y tenemos

$$\mu(A) - \lim \nu(A^{1/n}) = \mu(A) - \nu(A) \leq 0 \text{ y análogamente } \nu(A) - \mu(A) \leq 0.$$

Así pues,  $\mu(A) = \nu(A)$  para todo  $A$  cerrado. Como la  $\sigma$ -álgebra de Borel se puede generar a partir de los cerrados, se obtiene por regularidad que  $\mu = \nu$ .

Para la otra implicación, sea  $A$  un conjunto cerrado y  $\epsilon > 0$ , escribimos

$$A^\epsilon = A \cup \{y \in X \setminus A : d(y, A) \leq \epsilon\}.$$

Como la unión es ajena, y  $\mu$  es aditiva  $\mu(A^\epsilon) = \mu(A) + \mu(\{y \in X \setminus A : d(y, A) \leq \epsilon\})$

$$\Rightarrow \mu(A) - \mu(A^\epsilon) = -\mu(\{y \in X \setminus A : d(y, A) \leq \epsilon\}) < 0 < \epsilon$$

De donde se deduce que cualquier  $\epsilon > 0$  pertenece al conjunto sobre el que se considera el ínfimo. Por lo tanto  $d_P(\mu, \mu) = 0$

### 4. Desigualdad triangular

Si  $A$  es un conjunto cerrado, y  $\epsilon, \delta > 0$  consideremos tres medidas de probabilidad regulares tales que

$$\begin{aligned} \mu_1(A) &\leq \mu_2(A^\epsilon) + \epsilon \text{ y } \mu_2(A) \leq \mu_1(A^\epsilon) + \epsilon. \\ \mu_2(A) &\leq \mu_3(A^\delta) + \delta \text{ y } \mu_3(A) \leq \mu_2(A^\delta) + \delta. \end{aligned}$$

Para analizar la desigualdad triangular hagamos una observación.

**Observación 3.3.4.**  $(A^\epsilon)^\delta \subseteq A^{\epsilon+\delta}$

*Demostración.* Sea  $x \in (A^\epsilon)^\delta$  entonces  $d(x, A^\epsilon) \leq \delta$

$$\Rightarrow d(x, A) \leq d(x, A^\epsilon) + d(A^\epsilon, A) \leq \delta + \epsilon.$$

Con la observación, evidentemente ambos conjuntos son Borel-medibles por lo que a partir de la monotonía de las medidas tenemos que  $\mu((A^\epsilon)^\delta) \leq \mu(A^{\delta+\epsilon})$ .

Usando la observación encontramos el resultado:

$$\begin{aligned} \mu_1(A) &\leq \mu_2(A^\epsilon) + \epsilon \leq \mu_3((A^\epsilon)^\delta) + \epsilon + \delta \leq \mu_3(A^{\delta+\epsilon}) + \delta + \epsilon, \\ \mu_3(A) &\leq \mu_2(A^\delta) + \delta \leq \mu_1((A^\delta)^\epsilon) + \epsilon + \delta \leq \mu_1(A^{\delta+\epsilon}) + \delta + \epsilon. \end{aligned}$$

Como  $\epsilon$  y  $\delta$  eran arbitrarios, al considerar ínfimos, se concluye la desigualdad:

$$d_P(\mu_1, \mu_3) \leq d_P(\mu_1, \mu_2) + d_P(\mu_2, \mu_3).$$

■

### 3.3.1. Otras métricas en espacios de medidas

Hemos probado con rigor que la norma de variación es una norma en el espacio de medidas con signo, que la métrica de Wasserstein es una métrica en el espacio de probabilidades con  $p$ -momentos finitos y que la métrica de Prokhorov-Levy es una métrica sobre las medidas definidas en la  $\sigma$ -álgebra de Borel.

Sin embargo, éstas no son las únicas maneras de definir distancias en espacios de probabilidades, en esta sección se mencionan (sin demostración) algunas otras:

### Kantorovich-Rubinshtein

Esta métrica se verá identificada con la de Wasserstein cuando hayamos estudiado el Teorema de Riesz:

$$\|\mu\|_0 = \sup \left\{ \int f d\mu : f \in Lip_1(X), \sup_{x \in X} |f(x)| \leq 1 \right\}$$

donde  $Lip_1(X)$  es el conjunto de funciones de Lipschitz con dominio  $X$  y con constante 1.

### Sobre funciones de Lipschitz acotadas

$$\|\mu\|_{BL} = \sup \left\{ \int f d\mu : f \in BL(X), \|f\|_{bl} \leq 1 \right\}$$

donde BL es el conjunto de funciones de Lipschitz acotadas con la norma:

$$\|f\|_{bl} = \sup_{x \in X} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}$$

Esta norma resulta equivalente a la anterior, pero se desarrolla así para simplificar las pruebas de convergencia y completitud de BL

### Con densidades

Si  $\mu$  y  $\nu$  son absolutamente continuas respecto a una medida común  $\lambda$ , sean  $f, g$  sus derivadas de Radon-Nikodým respectivamente y  $S$  denota soporte, se pueden definir desviaciones y distancias de manera sencilla:

1. **Distancia de Hellinger:**  $d_H(\mu, \nu) = \int (\sqrt{f} - \sqrt{g})^2 d\lambda$

Así como esta distancia, se pueden definir muchas más usando las densidades y la integral. Por ejemplo, si en vez de haber considerado las raíces de las densidades, hubiesemos considerado las densidades tendríamos otra distancia.

2. **Desviación Ji-cuadrada:**  $d_{\chi^2}(\mu|\nu) = \int_{S_\mu \cup S_\nu} \frac{(f - g)^2}{g} d\lambda$

## 3.4. Topologías inducidas e identificación del espacio dual

Habiendo ya estudiado la estructura de espacio normado que obtiene el espacio de medidas, es necesario ahora identificar las topologías que inducen esas normas y distancias.

En análisis funcional (y en la vida en general) es de mucha utilidad reconocer los espacios duales topológicos para tener otra manera de atacar problemas. En general es de mucha ayuda contar con herramientas del espacio dual y aún de más ayuda es poder caracterizarlo. Por ejemplo, en programación lineal se estudia el problema dual para analizar la factibilidad del problema primal. En análisis funcional, otro ejemplo se da cuando se analiza la propiedad de Schur en algún espacio es imperioso hacer desarrollos en el espacio dual. El problema del transporte de Kantorovich se resuelve a partir del entendimiento del espacio dual del espacio de medidas, la caracterización de este espacio es el Teorema de Riesz de este capítulo.

### 3.4.1. Lema Urysohn en espacios métricos

El lema de Urysohn es un resultado muy importante en topología y establece que un espacio satisface el axioma  $T_4$  (Normal de Hausdorff- Cualesquiera dos conjuntos cerrados ajenos tienen vecindades abiertas ajenas) si existe una función continua que separa conjuntos cerrados disjuntos.

El lema de Urysohn se utiliza en las pruebas estándar del Teorema de Riesz. El enfoque de esta tesis no es sobre espacios topológicos generales sino medidas a partir de espacios métricos, probaremos el caso particular del lema de Urysohn para espacios métricos. La prueba es significativamente más sencilla en espacios métricos pues utiliza la información de la métrica definida en  $X$ .

**Lemma 3.4.1.** (*Urysohn en espacios métricos*)

Sean  $U, V$  dos subconjuntos cerrados y ajenos de un espacio métrico  $(X, d)$ , entonces existe una función continua que los separa.

Más aún, existe una función continua que toma el valor de 1 en  $U$  y 0 en  $V$ . Si  $U$  es compacto y  $V$  cerrado.

*Demostración.* Definimos  $f(x) = \frac{d(x, V)}{d(x, U) + d(x, V)}$

La función que asigna  $x \rightarrow d_A(x)$  es continua para todo conjunto  $A$ . Entonces  $f$  es suma y división de continuas. El denominador no se anula ya que  $V \cap U = \emptyset$  por lo que  $f$  es una función continua bien definida. Además,  $x \in \bar{V} \Leftrightarrow d(x, V) = 0$  entonces  $f(x) = 0$  y si  $x \in \bar{U}$  entonces  $d(x, U) = 0 \Rightarrow f(x) = 1$  ■

**Definición.** Un espacio se llama normal si siempre que  $A, B$  sean dos conjuntos cerrados disjuntos en  $X$  existen abiertos disjuntos  $U, V$  tales que  $A \subseteq U$  y  $B \subseteq V$ .

Un espacio normal  $T_1$  se denomina  $T_4$  o normal de Hausdorff. Para más detalles ver apéndice B.

En la prueba del Teorema de Riesz se utiliza el lema de Urysohn que vale en espacios normales. Además se utilizan abiertos que separan conjuntos, para que estas afirmaciones estén bien justificadas necesitamos asegurar que los espacios métricos son espacios normales de Hausdorff. (Satisface axioma  $T_4$  de separación).

**Observación 3.4.2.** Con la prueba del lema de Urysohn se tiene directamente que todo espacio métrico es normal de Hausdorff.

### 3.4.2. Teorema de Riesz

En esta sección el objetivo es identificar el espacio dual de las funciones continuas con soporte compacto con el espacio de medidas. Para fines del problema de transporte es importante identificar el dual del espacio de medidas, sin embargo este puede ser muy exótico. El doble dual  $E^{**}$  siempre contiene al menos al espacio original  $E$ . Esto se verifica vía la inmersión canónica dada por  $\langle x, f \rangle$ .

Por lo mencionado, nos enfocamos a analizar el predual del espacio de medidas.

**Observación 3.4.3.** Si el lector tiene **conocimiento previo** del Teorema de Riesz, se le recomienda **saltarse esta sección**, pues es el desarrollo de la prueba estándar del teorema. Se sigue y se detalla la prueba de [15] que esencialmente es la misma que la que se puede encontrar en [24].

Se escribe la prueba con el fin de tener un capítulo totalmente auto-contenido.

Sea  $\mathcal{L} = \{f : X \rightarrow \mathbb{R} : f \text{ es continua y con soporte compacto}\}$  (Recordemos que  $X$  es un espacio métrico)

La prueba se hará por partes, misma estrategia que consideran [24] y [15], pero se detallará cada paso, haciendo la prueba más extensa. La prueba se divide en 4 subteoremas.

Para lo que precede  $\Lambda$  será una funcional lineal no negativa en  $\mathcal{L}$  y definimos para todo compacto  $C$

$$\lambda(C) = \inf \left\{ \Lambda(f) : f \in \mathcal{L}_+, \chi_C \leq f \right\}$$

### 3.4.3. Enunciado del Teorema de Riesz

**Teorema.** (Riesz) *Se identifica al espacio dual topológico de las funciones continuas con soporte compacto sobre un espacio métrico como las medidas regulares de Borel.*

#### Procedimiento de prueba

El procedimiento que se sigue es el mismo que en [15] p. 246.

1.  $\lambda$  es una casi-medida regular.  
Si  $\mu$  es su extensión sobre la  $\sigma$ -álgebra de Borel entonces  $\mu(U) \leq \Lambda(f)$  para cualquier abierto  $\chi_U \leq f$
2.  $\int f d\mu \leq \Lambda(f)$
3. Para todo compacto  $C$  y  $\epsilon > 0$  existe  $f_0 \in \mathcal{L}_+$ ,  $f_0 \leq 1$ ,  $\chi_C \leq f_0$  tal que  $\Lambda(f_0) \leq \int f_0 d\mu + \epsilon$
4. Para cualquier funcional  $\Lambda$  existe una medida de Borel tal que  $\Lambda(f) = \int f d\mu$

Por lo mismo, en las siguientes páginas se detalla la prueba por secciones. Para facilitar la lectura se le denomina afirmación a cada una de las proposiciones numeradas. Es decir la primera afirmación corresponde al número 1 de la lista, la segunda afirmación al segundo elemento de la lista, etcétera.

Las ideas de esta prueba se pueden encontrar también en [4] y en [10] pero con menos detalle.



Figura 3.1: Frigyes Riesz (1880-1956)

### Primera afirmación

*Demostración.* Queremos ver que  $\lambda$  definida en la página anterior es una casi medida.

$\Lambda$  es una funcional positiva y el ínfimo se toma sobre un subconjunto de  $\mathcal{L}_+$  y por definición  $\lambda(C) \geq 0$

Para ver que  $\lambda$  es finita, si  $C$  es un compacto, sea  $U$  un conjunto abierto que contenga  $C$ , consideramos  $f$  como en el lema de Urysohn, es decir  $f \in \mathcal{L}_+$  tal que  $f(x) = 1 \ \forall x \in C$  y  $f(x) = 0 \ \forall x \in X \setminus U$

Entonces se tiene que  $\chi_C \leq f$  y por lo tanto  $\Lambda(f)$  es uno de los elementos de los que se toma el ínfimo.

Como  $\Lambda$  es una funcional,  $\Lambda(f)$  es finita.

$$\lambda(C) \leq \Lambda(f) < \infty$$

Si  $D, C$  son compactos, tales que  $D \subseteq C$  entonces cualquier función  $f \in \mathcal{L}_+$  tal que  $\chi_C \leq f$  también cumple que  $\chi_D \leq f$  ya que  $\chi_D \leq \chi_C$

Por lo que todos los elementos del conjunto que constituye  $\lambda(C)$  son elementos del conjunto con el que se calcula  $\lambda(D)$ .

Si  $C, D$  son compactos y  $\chi_C \leq f, \chi_D \leq g$ , con  $f, g \in \mathcal{L}_+$  entonces  $\chi_{C \cup D} \leq f + g$  con  $f + g \in \mathcal{L}_+$  por lo que

$$\lambda(C \cup D) \leq \Lambda(f + g) = \Lambda(f) + \Lambda(g)$$

Considerando ínfimos dos veces en el lado derecho, se concluye que  $\lambda$  es subaditiva.

$$\lambda(C \cup D) \leq \lambda(C) + \lambda(D)$$

Pero queremos probar que  $\lambda$  es una casi-medida, no que es subaditiva. Como las casi-medidas son aditivas nos falta probar superaditividad. Si  $C, D$  son dos conjuntos compactos disjuntos, como todo espacio métrico es normal de Hausdorff existen dos conjuntos abiertos  $U, V$  ajenos, conteniendo cada uno a los compactos  $C, D$ .

Sea  $f$  tal que  $f(x) = 1 \ \forall x \in C, f(x) = 0 \ \forall x \in X \setminus U$

Sea  $g$  tal que  $g(x) = 1 \ \forall x \in D, g(x) = 0 \ \forall x \in X \setminus V$

Si  $h \in \mathcal{L}_+$  tal que  $\chi_{C \cup D} \leq h$  entonces  $hf \in \mathcal{L}_+$  y  $\chi_C \leq hf$ , y  $hg \in \mathcal{L}_+$  con  $\chi_D \leq hg$

$$\lambda(C) + \lambda(D) \leq \Lambda(hf) + \Lambda(hg) = \Lambda(h(f + g)) \leq \Lambda(h)$$

Consideramos el ínfimo del lado derecho y tenemos  $\lambda(C) + \lambda(D) \leq \lambda(C \cup D)$

Por lo que para  $C, D$  compactos ajenos,  $\lambda(C \cup D) = \lambda(C) + \lambda(D)$

Por lo que se concluye que  $\lambda$  es una casi-medida. **Procedemos a probar que  $\lambda$  es regular.**

Por definición de ínfimo, para todo  $\epsilon > 0$  existe  $f_\epsilon \in \mathcal{L}_+$ ,  $\chi_C \leq f_\epsilon$  tal que  $\Lambda(f_\epsilon) \leq \lambda(C) + \frac{\epsilon}{2}$

Sea  $\beta \in (0, 1)$  y definimos  $A_{\epsilon, \beta} = \{x : f_\epsilon(x) \geq \beta\}$  entonces

$$C \subseteq \{x : f_\epsilon(x) \geq 1\} \subseteq \{x : f_\epsilon(x) > \beta\} \subseteq A_{\epsilon, \beta}$$

Pero si  $\chi_C \leq f$  entonces  $\chi_{A_{\epsilon, \beta}} \leq \frac{1}{\beta} f_\epsilon$  entonces  $\lambda(A_{\epsilon, \beta}) \leq \frac{1}{\beta} \Lambda(f_\epsilon) \leq \frac{1}{\beta} \left( \lambda(C) + \frac{\epsilon}{2} \right)$

Como  $\beta$  es cualquier valor, se escoge  $\beta$  de manera que  $\lambda(A_{\beta, \epsilon}) \leq \lambda(C) + \epsilon$

Sea  $\mu$  la medida de Borel inducida por  $\lambda$ . Si  $U$  es un abierto y  $f \in \mathcal{L}_+$  tal que  $\chi_U \leq f$  entonces

$$\mu(U) = \sup_C \{\lambda(C)\} \leq \Lambda(f)$$

## Segunda afirmación

La prueba de esta afirmación es sin duda la construcción más interesante. Se hace una descomposición de funciones que permite ver desigualdades en términos de  $\mu$ . Tanto la integral como la funcional son lineales en  $f$  por lo que se puede probar el resultado para las funciones en  $0 \leq f \leq 1$  sin pérdida de generalidad. Para  $n \in \mathbb{N}$  fija,  $i \in \{1, \dots, n\}$  definimos

$$f_i(x) = \begin{cases} 0 & f(x) \leq \frac{i-1}{n} \\ \frac{f(x) - \frac{i-1}{n}}{1/n} & \frac{i-1}{n} < f(x) \leq \frac{i}{n} \\ 1 & \frac{i}{n} < f(x) \end{cases}$$

La definición de  $f_i$  en cada  $x$  no depende únicamente de  $x$  sino de la ubicación de  $f(x)$ . Se está haciendo una partición de la imagen y nos fijamos donde cae cada  $f(x)$ . Al restarle  $(i-1)/n$  estamos analizando la diferencia de  $f$  con el piso de cada escalón.

La definición de  $f_i$  es casi una indicadora, parte la unidad en  $n$  subintervalos de misma longitud, cada  $f_i$  indica si  $f$  rebasó el nivel  $i$ .

Cada  $f_i \in \mathcal{L}_+$  porque  $f_i = \min\{\max\{nf - (i-1), 0\}, 1\}$

Además, si  $x \in \text{Dom}(f) \Rightarrow f(x) \in \left(\frac{j-1}{n}, \frac{j}{n}\right]$  para alguna  $j$  en  $\{1, \dots, n\}$  por ser estos una partición de la unidad.

$$\begin{aligned} \Rightarrow \frac{1}{n} \sum_{i=1}^n f_i(x) &= \frac{1}{n} \left( \sum_{i < j} f_i(x) + nf(x) - (j-1) + \sum_{i > j} f_i(x) \right) \\ \Rightarrow \frac{1}{n} \sum_{i=1}^n f_i(x) &= \frac{1}{n} \left( \sum_{i < j} 1 + nf(x) - (j-1) \right) = \frac{nf(x)}{n} = f(x) \end{aligned}$$

Es decir, para toda  $x$ ,  $f(x)$  es el promedio de todas las  $f_i$  en  $x$ .

Definimos  $U_i = \left\{ x : f(x) > \frac{i}{n} \right\} \forall i \in \{0, 1, \dots, n\}$

Si  $x \in U_i \Rightarrow f(x) > \frac{i}{n} \Rightarrow f_i(x) = 1$  por lo que  $\chi_{U_i} \leq f_i$  y por lo anterior  $\mu(U_i) \leq \Lambda(f_i)$

$$\Lambda(f) = \Lambda\left(\frac{1}{n} \sum_{i=1}^n f_i\right) = \frac{1}{n} \sum_{i=1}^n \Lambda(f_i) \geq \frac{1}{n} \sum_{i=1}^n \mu(U_i)$$

Es importante notar que  $\emptyset = U_n \subset U_{n-1} \subseteq \dots \subseteq U_1 \subseteq U_0$

Recordemos que si  $B \subseteq A \Rightarrow A = B \cup (A \cap (X \setminus B)) = B \cup A \setminus B \Rightarrow \mu(A) - \mu(B) = \mu(A \setminus B)$  siempre que  $\mu(B) < \infty$

Esto nos dice que  $\mu(U_i \setminus U_{i+1}) = \mu(U_i) - \mu(U_{i+1})$

$$\frac{1}{n} \sum_{i=1}^n \mu(U_i) = \frac{1}{n} \sum_{i=1}^n (i - (i-1))\mu(U_i) = \frac{1}{n} \sum_{i=1}^{n-1} i [\mu(U_i \setminus U_{i+1})]$$

Por otro lado, veamos el valor de la suma para poder usarlo como “cero” (sumándolo y restándolo) en la desigualdad de arriba.

$$\sum_{i=1}^{n-1} \mu(U_i \setminus U_{i+1}) = \sum_{i=1}^n [\mu(U_i) - \mu(U_{i+1})] = \mu(U_1) - \mu(U_2) + \mu(U_2) - \mu(U_3) + \cdots - \mu(U_n) = \mu(U_1)$$

Usamos estos últimos dos resultados en la estimación de  $\Lambda(f)$  y encontramos que

$$\Lambda(f) \geq \sum_{i=1}^{n-1} \left[ \frac{i+1}{n} \mu(U_i \setminus U_{i+1}) \right] - \frac{1}{n} \mu(U_1)$$

Observemos que  $U_i \setminus U_{i+1} = \left\{ x : \frac{i}{n} \leq f(x) < \frac{i+1}{n} \right\}$  por lo que usando la monotonía de la integral tenemos:

$$\begin{aligned} \int_{U_i \setminus U_{i+1}} f d\mu &\leq \frac{i+1}{n} \mu(U_i \setminus U_{i+1}) \\ \Rightarrow \sum_{i=1}^{n-1} \int_{U_i \setminus U_{i+1}} f d\mu - \frac{1}{n} \mu(U_1) &\leq \sum_{i=1}^{n-1} \frac{i+1}{n} \mu(U_i \setminus U_{i+1}) - \frac{1}{n} \mu(U_1) \leq \Lambda(f) \end{aligned}$$

Por linealidad de la integral podemos juntar los sumandos del lado izquierdo y obtenemos la integral de  $f$  sobre  $U_1$ . Además como  $U_1 \subset U_0$  encontramos

$$\int f d\mu - \frac{1}{n} \mu(U_0) \leq \Lambda(f)$$

Pero como  $\mu(U_0)$  no depende de  $n$ ,  $\mu(U_0) < \infty$  y  $n$  es arbitraria, entonces

$$\int f d\mu \leq \Lambda(f)$$

### Tercera afirmación

Sea, por definición de ínfimo,  $g_0 \in \mathcal{L}_+$  con  $\chi_C \leq g_0$  tal que  $\lambda(g_0) \leq \lambda(C) + \epsilon$ .

Sea  $f_0 = \min\{g_0, 1\}$  entonces  $\Lambda(f_0) \leq \Lambda(g_0) \leq \lambda(C) + \epsilon \leq \int f_0 d\mu + \epsilon$ .

### Cuarta afirmación

Si  $\Lambda$  es una funcional, construimos  $\mu$  como en los pasos anteriores. Sea  $f \in \mathcal{L}$ . He aquí la diferencia,  $f$  no necesariamente es positiva como en los pasos anteriores. Consideremos  $C$  cualquier compacto tal que  $\{x : f(x) \neq 0\} \subseteq C$  y para este conjunto sea  $f_0$  como la del resultado anterior, entonces como  $\chi_C \leq f_0 \Rightarrow f f_0 = f$

Como  $f$  se anula fuera de un compacto, alcanza máximo y mínimo, por lo que existe  $c$  tal que  $|f| \leq c$  además  $(f+c)f_0 \in \mathcal{L}_+$

$$\begin{aligned} \int f d\mu + c \int f_0 d\mu &\leq \Lambda(f) + c\Lambda(f_0) \\ \Rightarrow \int f d\mu + c \left[ \int f_0 d\mu - \Lambda(f_0) \right] &\leq \Lambda(f) \\ \Rightarrow \int f d\mu - c\epsilon &\leq \Lambda(f) \Rightarrow \int f d\mu \leq \Lambda(f) \end{aligned}$$

Por lo que el resultado de la segunda afirmación es válido para todo  $f \in \mathcal{L}$ , pero cuando se aplica a  $-f$  entonces se obtiene la desigualdad contraria y por lo tanto la igualdad deseada. La unicidad de  $\mu$  se deduce a partir de su regularidad. ■



El teorema de Riesz identifica al espacio dual vectorial del espacio vectorial de funciones continuas con soporte compacto con el espacio de medidas. Por lo que el espacio dual del espacio de medidas debe incluir al menos a las funciones continuas con soporte compacto. Entonces una sucesión  $\{\mu_n\}$  converge débilmente a una medida  $\mu$  (denotado  $\mu_n \Rightarrow \mu$ ) si satisface que para toda función continua  $f$  con soporte compacto.

$$\int f d\mu_n \rightarrow \int f d\mu$$

### 3.4.4. Teorema de Portmanteau [6] p.16

Si  $(X, d)$  es un espacio métrico y  $\{\mu_n\}_{n \in \mathbb{N}}, \mu$  son medidas de probabilidad en la  $\sigma$ -álgebra de Borel de  $X$  entonces son equivalentes:

1.  $\mu_n \Rightarrow \mu$
2.  $\lim_n \int f d\mu_n = \int f d\mu$  para cualquier función  $f$  uniformemente continua
3.  $\limsup_n \mu_n(F) \leq \mu(F)$  para todo cerrado  $F$
4.  $\liminf_n \mu_n(G) \geq \mu(G)$  para todo abierto  $G$
5.  $\lim_n \mu_n(A) = \mu(A)$  para todo conjunto  $A$  tal que  $\mu(\partial(A)) = 0$

*Demostración.* (1.  $\rightarrow$  2.) se deduce de que si la función es uniformemente continua, consideramos un soporte compacto para que cumpla las condiciones del teorema de Riesz.

(2.  $\rightarrow$  3.) Para cualquier conjunto  $F$  y  $\epsilon > 0$  consideremos  $f_{\epsilon, F}(x) = \max\left\{1 - \frac{d(x, F)}{\epsilon}, 0\right\}$

Notamos  $|f(x) - f(y)| \leq \frac{|d(x, F) - d(y, F)|}{\epsilon} \leq \frac{d(x, y)}{\epsilon}$  por lo que cada  $f_{\epsilon, F}$  es uniformemente continua y por lo tanto la podemos usar en 2.

Notemos también que si  $F$  es un conjunto cerrado entonces  $\chi_F \leq f_{\epsilon, F}$  pues si  $x \in F \Rightarrow d(x, F) = 0 \Rightarrow f_{\epsilon, F}(x) = \max\{1, 0\} = 1$  y también  $f_{\epsilon, F} \leq \chi_{F^\epsilon}$  pues si  $d(x, F) = c < \epsilon \Rightarrow f_{\epsilon, F}(x) = \max\{1 - \frac{c}{\epsilon}, 0\} = 1 - \frac{c}{\epsilon} < 1$  por lo que concluimos

$$\limsup_n \mu_n(F) \leq \lim_n \int f_{\epsilon, F} d\mu_n = \int f_{\epsilon, F} d\mu \leq \mu(F^\epsilon)$$

Si hacemos  $\epsilon \rightarrow 0$  obtenemos el resultado deseado.

(3.  $\rightarrow$  4.) Si  $G$  es un abierto, utilizamos el resultado anterior para  $X \setminus G$  y obtenemos el resultado directamente pues  $\liminf_n \mu_n(G) = 1 - \limsup_n \mu_n(X \setminus G)$ .

(3. + 4.  $\rightarrow$  5.) Si  $A$  es un conjunto tal que  $\mu(\partial(A)) = 0$  entonces

$$\mu(A) = \mu(A^\circ) \leq \liminf_n \mu_n(A^\circ) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A)$$

(5.  $\rightarrow$  1.) Sin pérdida de generalidad, se puede suponer  $0 < f < 1$  si  $A_t = \{x \in X : f(x) > t\}$

$$\begin{aligned} \int f d\mu &= \int_0^1 \mu(A_t) dt \\ \Rightarrow \int f d\mu &= \int_0^1 \mu(A_t) dt = \int_0^1 \lim_n \mu_n(A_t) dt = \lim_n \int_0^1 \mu_n(A_t) dt = \int f d\mu_n \end{aligned}$$

■

### 3.4.5. Teorema Prohkorov [6] p. 59

**Definición.** Una familia  $P$  de medidas de probabilidad en un espacio  $X$  se denomina **tirante** (tight) si para todo  $\epsilon > 0$  existe un conjunto compacto  $K_\epsilon \subseteq X$  tal que

$$\sup_{\mu \in P} \left\{ \mu(X \setminus K_\epsilon) \right\} \leq \epsilon$$

**Teorema.** (Prohkorov)

En un espacio métrico completo y separable

Una familia de probabilidades del espacio es tirante si y sólo si es débil compacta por sucesiones.

*Demostración.* ( $\Rightarrow$ ) Seguimos la prueba clásica que aparece en [6] pero detallamos mucho más algunos pasos.

Sea  $\{\mu_n\}_{n \in \mathbb{N}}$  una sucesión de medidas de probabilidad en la familia tirante  $\Pi$ .

El objetivo es construir una subsucesión  $\{\mu_{n_i}\}$  y una medida de probabilidad  $\mu$  tal que

$$\mu_{n_i} \Rightarrow \mu$$

Consideremos conjuntos compactos  $K_1 \subseteq K_2 \subseteq K_3 \cdots$  tales que  $\mu(K_u) \geq 1 - \frac{1}{u}$

Para cada  $\epsilon > 0$  el conjunto  $\bigcup_{k \in K_u} B_\epsilon(k)$  es una cubierta abierta del conjunto  $K_u$ . Por compacidad de  $K_u$

existen  $\{k_{1,\epsilon}, k_{2,\epsilon}, \dots, k_{n,\epsilon}\}$  tales que  $\bigcup_{i=1}^n B_\epsilon(k_{i,\epsilon})$  es una cubierta de  $K_u$ ,  $\bigcup_{i=1}^{\infty} \bigcup_{m=1}^{n_{1/i}} k_{m,1/i}$  es un conjunto denso

numerable en  $K_u$  por lo que el conjunto  $\bigcup_u K_u$  es separable.

Escribamos  $B = \{b_1, b_2, \dots, b_n\}$  para denotar el conjunto denso numerable en esta unión.

Si consideramos  $A = \{B_d(b_k) : b_k \in B, d \in \mathbb{Q}\}$  entonces  $A$  es una familia numerable de subconjuntos de  $X$  por ser tanto  $B$  como  $\mathbb{Q}$  conjuntos numerables. Por lo que podemos reescribir a  $A = \{A_1, A_2, \dots\}$ .

Si  $G$  es un conjunto abierto y  $x \in G \cap \bigcup_u K_u \Rightarrow \exists m$  tal que  $x \in A_m, \bar{A}_m \subseteq G$

Vamos a definir una clase de subconjuntos de  $X$ , a partir de la cual definiremos una casi-medida y posteriormente una medida.

$$\mathcal{H} = \{\emptyset, \bar{A}_m \cap K_i : m, i \in \mathbb{N}\}$$

Si  $F \subseteq H$  para algún  $H \in \mathcal{H}$  y  $G$  es un abierto tal que  $F \subseteq G$  entonces para cada  $x \in F$  existe  $A_x \in A$  tal que  $\bar{A}_x \subseteq G$ . Como  $F$  es subconjunto de  $H$ ,  $F$  es compacto por lo que de la cubierta  $\bigcup_{x \in F} A_x$  podemos

extraer un número finito de conjuntos  $A_1, A_2, \dots, A_n$  tales que  $F \subseteq G$  pero como  $F \subseteq \bigcup_{i=1}^n A_i$  y  $F \subseteq K_m$

para alguna  $m$  natural, consideramos  $H_0 = \bigcup_{i=1}^n (A_i \cap \bar{K}_m)$  satisface que  $F \subseteq H_0 \subseteq G$ .

Es decir si tenemos un conjunto cerrado  $F$  contenido en algún elemento  $H$  de el conjunto  $\mathcal{H}$  y un conjunto abierto  $G$  tal que  $F \subseteq G$  entonces podemos separar a los conjuntos por elementos de  $\mathcal{H}$  o bien: **Podemos discernir cerrados y abiertos por  $\mathcal{H}$**

Notemos que  $\mathcal{H}$  es numerable pues  $A$  es numerable y sólo tenemos una sucesión de elementos del tipo  $K_i$  entonces escribamos  $\mathcal{H} = \{H_1, H_2, H_3, \dots\}$  para poder hacer un argumento diagonal.

La sucesión  $\{\mu_n(H_1)\}$  es una sucesión de elementos en  $[0, 1]$  por lo que por Bolzano-Weierstrass admite una subsucesión convergente  $\{\mu_{n_1,i}(H_1)\}$ . Aplicamos la misma idea a la sucesión  $\{\mu_{1,n}(H_2)\}$  y obtenemos una subsucesión  $\{\mu_{n_2,k}\}$  para la cual hay convergencia en la evaluación en  $H_1$  y  $H_2$ .

Procedemos inductivamente y obtenemos una subsucesión  $\{\mu_{n_i}\}$  tal que  $\{\mu_{n_i}(H_k)\}$  converge para toda  $k \in \mathbb{N}$ .

Definimos  $\alpha(H) = \lim_i \mu_{n_i}(H)$ ,  $\alpha$  está bien definida pues la convergencia en cada  $H$  está asegurada por el argumento diagonal.

Si  $H_1 \subseteq H_2$  entonces  $\mu_{n_i}(H_1) \leq \mu_{n_i}(H_2) \forall i \in \mathbb{N}$

$$\therefore H_1 \subseteq H_2 \Rightarrow \alpha(H_1) \leq \alpha(H_2)$$

Si  $H_1 \cap H_2 = \emptyset \Rightarrow \mu_{n_i}(H_1 \cup H_2) = \mu_{n_i}(H_1) + \mu_{n_i}(H_2)$  y por lo mismo se obtiene

$$\begin{aligned} \lim_i \mu_{n_i}(H_1 \cup H_2) &= \lim_i \mu_{n_i}(H_1) + \lim_i \mu_{n_i}(H_2) \\ H_1 \cap H_2 = \emptyset &\Rightarrow \alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2) \\ \therefore \alpha(H_1 \cup H_2) &\leq \alpha(H_1) + \alpha(H_2) \quad \forall H_1, H_2 \in \mathcal{H} \end{aligned}$$

Vamos a aprovechar la propiedad de que los conjuntos en  $A$  aproximan a los abiertos en el sentido que podemos encontrar elementos de  $A$  cuya cerradura sigue contenida en cualquier abierto. Entonces consideremos

$$\beta(G) = \sup_{H \subseteq G} \alpha(H)$$

Si  $G_1, G_2$  son dos conjuntos abiertos y  $H \in \mathcal{H}$  es un conjunto tal que  $H \subseteq G_1 \cup G_2$  definimos

$$\begin{aligned} F_1 &= \{x \in H : d(x, X \setminus G_1) \geq d(x, X \setminus G_2)\} \\ F_2 &= \{x \in H : d(x, X \setminus G_2) \geq d(x, X \setminus G_1)\} \end{aligned}$$

Si  $x \in F_1$  y  $x \notin G_1$  como  $H \subseteq G_1 \cup G_2$  entonces  $x \in G_2$  por lo que  $d(x, X \setminus G_2) > 0$  y  $d(x, X \setminus G_1) = 0$  lo cual es una contradicción por la definición de  $F_1$ .

Por lo que  $F_1 \subseteq G_1$  y análogamente  $F_2 \subseteq G_2$ .

Como  $\mathcal{H}$  nos permite discernir abiertos de cerrados, existen  $H_{0,1}, H_{0,2} \in \mathcal{H}$  tales que

$$\begin{aligned} F_1 &\subseteq H_{0,1} \subseteq G_1 \\ F_2 &\subseteq H_{0,2} \subseteq G_2 \end{aligned}$$

Así,  $\beta(G_1 \cup G_2) = \sup_{H \subseteq G_1 \cup G_2} \{\alpha(H)\} \leq \alpha(H_{0,1}) + \alpha(H_{0,2}) \leq \beta(G_1) + \beta(G_2)$ .

O bien,  $\beta$  es **subaditividad en conjuntos abiertos**. Si  $H \in \mathcal{H}$  tiene una cubierta numerable por abiertos  $H \subseteq \bigcup_{n=1}^{\infty} G_n$  como  $H$  es compacto, existe una colección finita  $\{G_{n_1}, G_{n_2}, \dots, G_{n_m}\}$  de los  $G_n$  cuya unión cubre a  $H$ , por lo tanto:

$$\alpha(H) \leq \beta \left( \bigcup_{i=1}^m G_{n_i} \right) \leq \sum_{i=1}^m \beta(G_{n_i}) \leq \sum_{n=1}^{\infty} \beta(G_n)$$

Considerando supremos sobre  $\mathcal{H}$  se deduce que  $\beta$  es  **$\sigma$ -subaditividad** Y si  $S$  denota la  $\sigma$ -álgebra del espacio en la que está definida  $\Pi$ , para  $M \in S$  definimos

$$\gamma(M) = \inf_{M \subseteq G} \beta(G)$$

$\gamma$  es una medida exterior.

La prueba se replica usando la que ya se hizo en el capítulo 2 por lo que no se repite aquí.

Consideremos  $\mu$  la restricción de  $\gamma$  a  $S$  es decir,  $(\mu = \gamma|_S)$ .

Entonces si  $G$  es un abierto  $\mu(G) = \gamma(G) = \beta(G)$  y por lo mismo  $\mu(G) = \sup_{H \subseteq G} \alpha(H)$

Entonces  $\lim_i \mu_{n_i}(H) = \alpha(H) \leq \lim_i \inf \mu_{n_i}(G) \Rightarrow \mu(G) \leq \lim_i \inf \mu_{n_i}(G)$  por lo que por el teorema de Portmanteau nos permite concluir

$$\mu_{n_i} \Rightarrow \mu$$

*Demostración.* ( $\Leftarrow$ ) Si  $G_1 \subseteq G_2 \subseteq \dots$  es una sucesión de conjuntos abiertos tal que  $X = \bigcup_{n=1}^{\infty} G_n$  entonces

para cada  $\epsilon$  existe  $n$  tal que  $\mu(G_n) \geq 1 - \epsilon$  para toda medida  $\mu \in \Pi$

Porque si tal  $n$  no existiera, consideramos cualquier sucesión de elementos en  $\Pi$  tales que  $\mu_i(G_n) < 1 - \epsilon$ . Por hipótesis existe una subsucesión que converge a una probabilidad  $\mu$  y tendríamos que  $\mu(G_n) \leq 1 - \epsilon$  para toda  $n$  y entonces  $\mu$  no sería una probabilidad pues  $\mu(X) < 1$  como esto es un absurdo, se concluye que tal  $n$  existe.

Si  $k$  es un natural, como  $X$  es separable existen puntos  $k_1, k_2, \dots$  tales que las bolas  $B_{1/k}(k_1), B_{1/k}(k_2), \dots$  cubren a  $X$ . Para cada  $k$  existe  $n_k$  tal que  $\mu(\bigcup_{n < n_k} B_{1/k}(k_n)) > 1 - \frac{\epsilon}{2^k}$ . Si  $K$  es la cerradura de  $\bigcap_k \bigcup_{n < n_k} B_{1/k}(k_n)$  es compacto por completez (ver Apéndice C) entonces

$$\mu(K) = \mu(X \setminus K) \geq 1 - \mu\left(X \setminus \bigcup_{k=1}^{\infty} \bigcup_{n < n_k} B_{1/k}(k_n)\right) > 1 - \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = 1 - \epsilon \quad \forall \mu \in \Pi$$

■

### 3.4.6. Comentarios sobre el capítulo

La topología inducida por la métrica de Wasserstein y la métrica de Prohkorov-Levy metrizan la convergencia débil. Para probar que la métrica de Wasserstein metriza la convergencia débil se utiliza el Teorema de representación de Kantorovich-Rubenstein, que dice que la métrica de Wasserstein con  $p = 1$  se puede representar en términos de integrales en diferencias de medidas.

También se puede mostrar que las métricas de Wasserstein inducen la misma topología y en ese sentido son equivalentes. Inclusive, la métrica de Wasserstein está determinada por la función distancia  $d$ , pero en términos de su topología esta se puede obtener si se sustituye  $d$  por una métrica acotada, como por ejemplo  $\min\{d, 1\}$ .

La métrica de Prohkorov Levy también metriza la convergencia débil y aún así no son las únicas maneras de hacerlo.

En el Teorema de Prohkorov se colocaron las hipótesis de separabilidad y completez desde el principio, sin embargo, este enfoque no es el más utilizado, ya que esas hipótesis sólo se utilizan en la segunda parte. En la literatura de convergencias de probabilidades se suele separar el Teorema de Prohkorov en dos partes, “La parte directa” y el regreso. A pesar de que la construcción es mucho más compleja, se le llama “parte directa” a la implicación que se probó primero en este capítulo ya que no necesita de esas hipótesis complementarias. Las métricas inducidas por densidades son de fundamental importancia en problemas estadísticos. A pesar de que sólo se expuso una, existen muchas más. No se debe pensar que las demás son de menor importancia, sino que aparecen naturalmente al resolver diferentes problemas. En este capítulo se prepararon las bases para poder desarrollar la teoría de dualidad del problema de Kantorovich, con el objetivo de poder entender todos los pasos en las siguientes pruebas, es fundamental tener las nociones de convergencia en el espacio de medidas y su convergencia débil.

# Formulación y dualidad del problema de Kantorovich

El problema que formuló Kantorovich se reduce a relacionar dos distribuciones en términos de otra distribución. Busca la mejor manera de acoplar variables aleatorias de distintos grupos, de tal manera que cuando nos fijamos en direcciones específicas, las distribuciones coincidan con las originales.

## 4.1. El Problema de Kantorovich

Si  $(X, S_X, \mu), (Y, S_Y, \nu)$  son dos espacios de medida con  $X$  y  $Y$  compactos y si  $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$  es una función continua no negativa, el objetivo es encontrar:

$$\inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c(x, y) d\pi(x, y) \right\}$$

Sea  $\Gamma(\mu, \nu)$  el conjunto de probabilidades  $\pi$ , en el espacio producto  $X \times Y$  tales que sus marginales coinciden con  $\mu$  y  $\nu$ , es decir:

$$\Gamma(\mu, \nu) = \left\{ \pi \in M(X \times Y) : \pi_x(\pi) = \mu, \pi_y(\pi) = \nu \right\}$$

donde  $M$  indica el conjunto de todas las medidas de probabilidad de un espacio,  $\pi_x, \pi_y$  denotan las proyecciones en  $X$  y en  $Y$  respectivamente.

Equivalentemente

$$\Gamma(\mu, \nu) = \left\{ \pi \in M(X \times Y) : \pi(A \times Y) = \mu(A), \pi(X \times B) = \nu(B) \quad \forall A \in S_X, B \in S_Y \right\}$$

donde  $S_X$  y  $S_Y$  son las  $\sigma$ -álgebras de los conjuntos medibles en los espacios  $X, Y$  respectivamente.

El objetivo de esta sección es detallar la prueba de la formulación dual del problema de Kantorovich. Este proceso de prueba, como se verá más adelante, asegura la existencia de la solución al problema de Kantorovich en condiciones muy generales. La formulación dual es una técnica muy usada en programación. En programación lineal existen teoremas de dualidad que dan la forma explícita del problema. En programación cuadrática se formula el problema dual a partir de las condiciones de Karush-Kuhn-Tucker.

Las formulaciones duales son de fundamental interés en el desarrollo de problemas de optimización como el que nos compete.

En el Teorema de dualidad del final de este capítulo capítulo, se requiere estudiar límites en conjuntos que dependen de medidas. Para esto, es importante considerar casos extremos en las medidas, como son las deltas de Dirac. Muchos resultados en el problema de Monge exigen que las medidas no acumulen masa en conjuntos determinados, para formalizar esta noción se usa la idea de átomo de una medida.

## 4.2. Indicadoras Infinitas

En muchos textos, es común no usar las indicadoras como las hemos definido previamente, sino considerar una clase especial de ellas. Las indicadoras pueden pensarse como “focos” que se prenden o se apagan dependiendo de dónde se ubica el elemento. Por lo mismo, muchos autores prefieren usar el valor infinito en vez del 1, este enfoque tiene algunas ventajas y algunas desventajas como por ejemplo, hacer que la indicadora sea una función convexa.

**Definición.** *Indicadora infinita* Para un conjunto  $A \subseteq X$  definimos la indicadora infinita  $\chi_A^\infty : X \rightarrow \bar{\mathbb{R}}$  según la regla:

$$\chi_A^\infty(x) = \begin{cases} \infty & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Debe ser claro al lector que podría no definirse la indicadora infinita y hacer la multiplicación  $\infty \chi_A$  y a pesar de que sería equivalente, se escribe esta sección para enfatizar la importancia de esta clase de funciones. Una de las grandes ventajas de definir esta clase de funciones es poder cambiar el espacio sobre el que se tomarán supremos. Por ejemplo supongamos que queremos calcular:

$$\sup_{x \in A} f(x)$$

donde  $f$  es una función finita. Entonces notemos que  $\chi_{X \setminus A}^\infty$  se prende fuera de  $A$  y se apaga en  $A$ . Pero cuando esta función se prende toma un valor tan grande que superaría cualquier valor de  $f$ , por lo que podríamos usarla para reescribir el valor que queremos calcular:

$$\sup_{x \in A} f(x) = \sup_{x \in X} \left\{ f(x) - \chi_{X \setminus A}^\infty(x) \right\}$$

esta técnica sirve para cambiar el dominio sobre que se calcula el supremo, pasamos de calcularlo sobre  $A$  a calcularlo sobre  $X$ . En el Teorema de Dualidad de Kantorovich será importante tener esta herramienta para pasar a dominios más grandes.

Sin embargo, esta formulación también tiene sus desventajas, por ejemplo, el uso de indicadoras permite escribir todas las operaciones de conjuntos como operaciones de indicadoras. Así, se puede escribir:  $\chi_{A \cup B} = \chi_A + \chi_B - \chi_A \chi_B$

Lamentablemente, esta idea no se puede replicar con las indicadoras infinitas, debido a que la operación puede no estar bien definida. Si intentáramos usar la misma idea pero con indicadoras infinitas nos encontraríamos con que si  $x \in A \cap B$  entonces  $\chi_A^\infty(x) = \chi_B^\infty(x) = \infty$ , por lo que la evaluación  $\chi_A^\infty + \chi_B^\infty - \chi_A^\infty \chi_B^\infty$  no está definida en toda  $x$ , pues la operación “ $\infty - \infty$ ” no está definida.

Con estos razonamientos, el problema físico de encontrar las deltas de Dirac como funciones integrables es equivalente a buscar “el otro enfoque” respecto a las indicadoras infinitas, como hicimos en la sección anterior.

### 4.3. Transformada de Legendre-Frechet

Sea  $X$  un espacio normado y sea  $X^*$  su dual topológico. Y denotemos por  $\langle \cdot, \cdot \rangle$  a la evaluación de las funcionales lineales en los puntos del espacio.

**Definición.** Transformada de Legendre-Frechet / Función convexa conjugada.

Sea  $f : X \rightarrow \mathbb{R}$  una función, definimos  $f^* : X^* \rightarrow \bar{\mathbb{R}}$  por la siguiente regla de correspondencia:

$$f^*(x^*) = \sup_{x \in X} \left\{ \langle x^*, x \rangle - f(x) \right\}$$

A  $f^*$  se le llama transformada de Legendre-Frechet de  $f$  o bien la función convexa conjugada de  $f$ .

Obsérvese que no hay problema con la definición pues el contradominio son los reales extendidos, por lo que  $f^*$  puede tomar el valor  $\infty$

#### 4.3.1. Desigualdad de Fenchel-Young

De la definición de supremo se obtiene

$$\begin{aligned} \langle x^*, x \rangle - f(x) &\leq f^*(x^*) \quad \forall x \in X \\ \Rightarrow \langle x^*, x \rangle &\leq f(x) + f^*(x^*) \quad \forall x \in X, x^* \in X^* \end{aligned}$$

A esta última desigualdad se le llama Desigualdad de Fenchel-Young, debido a que esta desigualdad es una generalización de la desigualdad de Hölder-Young. Comprobemos que, en efecto, se puede obtener la desigualdad de Hölder-Young en  $\mathbb{R}$ . ■

#### Ejemplo: La desigualdad de Hölder-Young

Si  $X = \mathbb{R}$  y  $h : X \rightarrow \mathbb{R}$  es una funcional lineal, entonces para todo  $x, y \in \mathbb{R}$  y para todo  $t \in (0, 1)$  se tiene que  $h(tx + (1-t)y) = th(x) + (1-t)h(y)$ , por lo que  $h$  es una función afín y podemos escribir  $h(x) = \beta x$ ,  $\beta \in \mathbb{R}$  por lo que  $X^* = \mathbb{R}$  también.

Sea  $q \in (1, \infty)$  y sea  $p$  su exponente conjugado es decir  $\frac{1}{p} + \frac{1}{q} = 1$  definamos  $f(x) = \frac{x^q}{q}$ .

El objetivo es encontrar la forma explícita de  $f^*(x^*)$

Pero como acabamos de ver que el dual de los reales son los reales, escojamos  $a \in \mathbb{R}, a > 0$ , y definamos  $x^* = a$ , entonces  $\langle x^*, x \rangle = ax$ .

Por definición

$$f^*(a) = \sup_{x \in \mathbb{R}} \left\{ ax - \frac{x^q}{q} \right\}$$

Pero aprovechemos nuestras técnicas de cálculo para encontrar máximos de funciones.

Sea  $L(x) = ax - \frac{x^q}{q}$  entonces  $\frac{dL}{dx}(x_{opt}) = a - (x_{opt})^{q-1} = 0 \Leftrightarrow a^{1/(q-1)} = x_{opt}$

Por lo que el óptimo se alcanza en  $a^{p-1}$ , y podemos verificar a que corresponde usando la prueba de la segunda derivada:

$$\frac{d^2L}{dx^2}(a^{p-1}) = -(q-1)a^{(p-1)(q-2)} < 0 \text{ ya que } q > 1.$$

Por lo mismo, el valor  $a^{p-1}$  corresponde a un máximo de  $L$  y entonces encontremos el máximo

$$L(a^{p-1}) = aa^{p-1} - \frac{a^p}{q} = a^p \left(1 - \frac{1}{q}\right) = \frac{a^p}{p}$$

Como el valor máximo se alcanza, podemos escribir  $f^*(a) = \frac{a^p}{p}$  y sustituyendo este resultado en la desigualdad de Fenchel-Young encontramos:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

que es la desigualdad de Hölder-Young.

Sin embargo, esta manera de deducir la desigualdad es equivalente a encontrar este supremo con técnicas de cálculo. Este ejemplo se presenta para notar la aplicabilidad de la desigualdad, sin embargo su verdadera fuerza radica en espacios en los que no podemos usar cálculo (diferenciabilidad) para encontrar cotas superiores.

### 4.3.2. La conjugada de la norma

En el mismo contexto de espacios normados y sus duales, buscamos encontrar la función convexa conjugada de la norma del espacio.

**Proposición 4.3.1.** Sea  $f(x) = \|x\|$  entonces  $f^*(x^*) = \chi_{X^* \setminus B_{X^*}}^\infty(x^*)$

*Demostración.* Si  $\|x^*\|_* > 1$  por definición de la norma existe  $x \in B_X$  tal que  $\langle x^*, x \rangle > 1$ . Para cada  $n \in \mathbb{N}$  definamos  $x_n = nx$ . Entonces  $\langle x^*, x_n \rangle - \|x_n\| = n(\langle x^*, x \rangle - \|x\|)$  por lo que  $f^*(x^*) = \infty$

Si  $\|x^*\|_* \leq 1$  entonces  $\langle x^*, x \rangle \leq \|x\| \|x^*\|_* \leq \|x\| \Rightarrow \langle x^*, x \rangle - \|x\| \leq 0$  para toda  $x \in X$  por lo que el supremo se alcanzaría en 0 si hubiera un punto en el conjunto que nos diera este valor. En efecto,  $x = 0$  hace que  $\langle x^*, 0 \rangle - \|0\| = 0$ . De ahí que  $f^*(x^*) = 0$ .

Juntando estos dos casos:

$$f^*(x^*) = \chi_{X^* \setminus B_{X^*}}^\infty(x^*) = \begin{cases} \infty & \text{si } \|x^*\|_* > 1 \\ 0 & \text{si } \|x^*\|_* \leq 1 \end{cases}$$

■

### 4.3.3. Déficit esperado

En problemas actuariales, es de fundamental importancia analizar funciones de las colas de distribuciones, porque muchos de los eventos asegurable son de colas muy pesadas. Por lo mismo, y con el propósito de encontrar medidas de riesgo como el V@R (value at risk) es de interés analizar el mínimo de cantidad esperada por pagar.

Sea  $\mathbf{X} \stackrel{d}{=} F$  y con función de densidad  $f$ .

Así, definamos

$$g(x) = \int_{-\infty}^x F(t)dt,$$

el objetivo de definir esta función es encontrar su función convexa conjugada pues corresponde a lo que queremos acotar. Usando el teorema fundamental del cálculo encontramos que el punto donde se maximiza  $g^*(x^*)$  es  $x = F^{-1}(x^*)$  el percentil  $x^*$  si  $F$  no fuera invertible, se considera la inversa generalizada.

$$\Rightarrow g^*(x^*) = x^* F^{-1}(x^*) - \int_{-\infty}^{F^{-1}(x^*)} F(t)dt$$



Aún más, si resolvemos la integral por partes podemos simplificar la expresión:

$$g^*(x^*) = \int_{-\infty}^{F^{-1}(x^*)} tf(t)dt = \mathbf{E}[\text{máx}\{F^{-1}(x^*) - \mathbf{X}, 0\}]$$

Por lo que la función convexa conjugada se puede reescribir como el pago promedio sujeto a un límite. Entonces aplicamos la desigualdad de Frenchel Young para tener una estimación:

$$x^*x \leq \int_{-\infty}^x F(t)dt + \mathbf{E}[\text{máx}\{F^{-1}(x^*) - \mathbf{X}, 0\}]$$

Por lo que despejando podemos encontrar una cota inferior para el pago esperado debajo de un límite:

$$x^*x - \int_{-\infty}^x F(t)dt \leq \mathbf{E}[\text{máx}\{F^{-1}(x^*) - \mathbf{X}, 0\}]$$

pero en la práctica, esta cota sólo puede tener uso si el lado izquierdo es un valor positivo, caso en el cual se tiene una estimación de la mínima cantidad de pagos por hacer.

#### 4.3.4. El ejemplo fundamental de la transformada

El objetivo de este capítulo es probar una versión general del teorema de dualidad del problema de Kantorovich. En el teorema se usan dos funciones y sus funciones convexas conjugadas a las que se les aplica el principio de minimax de Rockafellar. Para evitar confusiones en la prueba y aprovechando la práctica que hemos adquirido en el cálculo de funciones convexas conjugadas, calculemos las transformadas en esta sección. Usaremos la misma notación de la sección 1 de este capítulo (la formulación del problema de Kantorovich). Pero aquí pensemos que  $c$  no es cualquier función sino una **función continua no negativa acotada**.

Sea  $E = C_b(X \times Y)$  el espacio vectorial de funciones continuas con soporte compacto que tienen como dominio  $X \times Y$  y contradominio  $\mathbb{R}$ .

Con el teorema de Riesz se probó, en el capítulo anterior,  $E^* = M(X \times Y)$ , es el conjunto de medidas de Borel dotado de la norma variación.

Sean  $\Theta, \Xi : C_b(X \times Y) \rightarrow \mathbb{R} \cup \{\infty\}$  funciones dadas por:

1. 
$$\Theta(u) = \chi_{E \setminus \{u \geq -c\}}^\infty(u)$$

- 2.

$$\Xi(u) = \begin{cases} \int_X \phi d\mu + \int_Y \psi d\nu & \text{si } u(x, y) = \phi(x) + \psi(y) \\ \infty & \text{en otro caso} \end{cases}$$

Usando el teorema de Riesz calculemos las transformadas de Legendre-Freché de estas dos funciones. Sea  $\pi$  una medida de probabilidad de Borel en  $X \times Y$

$$\Theta^*(-\pi) = \sup_{f \in C_b} \left\{ - \int_{X \times Y} f d\pi - \chi_{E \setminus \{f \geq -c\}}^\infty(f) \right\}$$

Pero como ya vimos, las indicadoras infinitas nos permiten acotar el espacio de búsqueda sobre el cual se considera un supremo, por lo que podemos reescribir

$$\Theta^*(-\pi) = \sup_{f \in C_b} \left\{ - \int_{X \times Y} f d\pi : f(x, y) \geq -c(x, y) \right\}$$

Pero si  $f \in C_b$  entonces  $-f \in C_b$  y podemos entonces escribir el supremo usando  $-f$

$$\Theta^*(-\pi) = \sup_{f \in C_b} \left\{ \int_{X \times Y} f d\pi : f(x, y) \leq c(x, y) \right\}$$

Si existe  $(a_1, a_2)$  tal que  $\pi((a_1, a_2) \times Y) < 0$  o  $\pi(X \times (a_1, a_2)) < 0$ . Para  $\epsilon > 0$  consideremos

$$-v_\epsilon(x, y) = \begin{cases} 0 & \text{si } x \notin [a_1 - \epsilon, a_2 + \epsilon] \\ (1-t)a_1 & \text{si } x = t(a_1 - \epsilon) + (1-t)a_1 \\ 1 & \text{si } x \in (a_1, a_2) \\ ta_2 & \text{si } x = ta_2 + (1-t)(a_2 + \epsilon) \end{cases}$$

Así,  $v_\epsilon$  es una función negativa y por propiedad arquimediana se puede escoger  $\epsilon$  tan pequeño como se requiera para que  $\int v_\epsilon d\pi > 0$

Pero entonces para cualquier  $n \in \mathbb{N}$  la función  $v_n = nv_\epsilon$  es integrable y satisface que  $v_n \leq 0 \leq c$  pero  $\int v_n d\pi \rightarrow \infty$  y por lo mismo si  $\pi$  es negativa entonces  $\Theta^*(-\pi) = \infty$

Por otro lado si  $\pi$  es una medida no negativa,  $f_n = \min\{c, n\} \in C_b$ , que cumple  $f_n \leq c$  tenemos que el supremo es  $\int_{X \times Y} c(x, y) d\pi$

Juntamos estos dos casos para concluir:

$$\Theta^*(-\pi) = \begin{cases} \int_{X \times Y} c(x, y) d\pi & \text{si } \pi \in M_+(X \times Y) \\ \infty & \text{si } \pi \notin M_+(X \times Y) \end{cases}$$

Ahora calculemos la transformada de Legendre-Frechet de  $\Xi$ , pero primero tenemos que probar que  $\Xi$  es una función bien definida. Porque  $\phi$  y  $\psi$  son funciones que se determinan dentro de la definición de  $\Xi$  por lo que hay que ver que si existen dos pares  $(\phi_1, \psi_1)$  y  $(\phi_2, \psi_2)$  tales que  $\phi_1(x) + \psi_1(y) = u(x, y) = \phi_2(x) + \psi_2(y)$ . Consideremos sólo los puntos donde las funciones son finitas:

$$\begin{aligned} &\Rightarrow \phi_1(x) - \phi_2(x) = \psi_2(y) - \psi_1(y) \quad \forall y \in Y \\ &\Rightarrow \psi_1(y) - \psi_2(y) = \phi_2(x) - \phi_1(x) \quad \forall x \in X \\ &\Rightarrow \phi_1 - \phi_2 = c \text{ y } \psi_1 - \psi_2 = -c \text{ para alguna constante } c \\ &\Rightarrow \int \phi_1(x) d\mu + \int \psi_1(y) d\nu = \int \phi_2(x) d\mu + \int \psi_2(y) d\nu \end{aligned}$$

Por lo tanto,  $\Xi$  está bien definida, pues no depende de la elección de las funciones  $(\phi, \psi)$

Procedemos a calcular la transformada de  $\Xi$

$$\Xi^*(\pi) = \sup_{f \in C_b} \left\{ \int f d\pi - \Xi(f) \right\}$$

Sin embargo,  $\Xi$  toma el valor extendido infinito si no podemos escribir a  $f$  como suma de funciones (una dependiendo de  $x$  y otra de  $y$ ), por lo que, como aparece restando en el argumento del supremo, el supremo se resume a calcular sobre las funciones que cuya suma es  $f$ , pues en otro caso el valor es  $-\infty$

$$\begin{aligned} \Xi^*(\pi) &= \sup_{f \in C_b} \left\{ \int \phi(x) + \psi(y) d\pi - \int \phi(x) d\mu - \int \psi(y) d\nu : f(x, y) = \phi(x) + \psi(y) \right\} \\ &= \sup_{f \in C_b} \left\{ \int_X \phi(x) d\pi + \int_Y \psi(y) d\pi - \int \phi(x) d\mu - \int \psi(y) d\nu : f(x, y) = \phi(x) + \psi(y) \right\} \end{aligned}$$

Recordemos la definición de  $\Gamma(\mu, \nu) = \{\pi \in M(X \times Y) : \pi_x(\pi) = \mu, \pi_y(\pi) = \nu\}$ .

Y notemos que si  $\pi \in \Gamma(\mu, \nu)$  entonces las integrales coinciden y el valor del supremo es cero. Por otro lado si  $\pi \notin \Gamma(\mu, \nu)$  entonces las integrales pueden no ser iguales. Sin pérdida de generalidad, podemos pensar que la diferencia de estas integrales es positiva porque en caso contrario consideramos  $-f$  que también está en el conjunto. Si la diferencia es estrictamente positiva, entonces consideramos  $f_n = nf$  y por lo mismo  $(f_n)$  es una sucesión de elementos en  $C_b$  y  $\Xi^*(f_n) \rightarrow \infty$  a medida de que crece  $n$ . De este modo concluimos que  $\Xi^*(\pi)$  es cero para  $\pi \in \Gamma(\mu, \nu)$  y es infinito para  $\pi \notin \Gamma(\mu, \nu)$ . O bien, usando la notación que hemos implementado:

$$\Xi^* = \chi_{E^* \setminus \Gamma(\mu, \nu)}^\infty$$

Así, hemos encontrado las funciones convexas conjugadas de  $\Theta$  y  $\Xi$  elementos que serán fundamentales en la prueba de dualidad de Kantorovich. En esa prueba (que se presenta al final de este capítulo) se hace referencia a este desarrollo. Dada la importancia de este ejemplo en la prueba, hacemos un resumen de los resultados:

### Resumen

Sean  $\Theta$  y  $\Xi$  dos funciones definidas en el espacio de funciones continuas acotadas y dadas por las fórmulas:

1. 
$$\Theta(u) = \chi_{E \setminus \{u \geq -c\}}^\infty(u)$$

- 2.

$$\Xi(u) = \begin{cases} \int_X \phi d\mu + \int_Y \psi d\nu & \text{si } u(x, y) = \phi(x) + \psi(y) \\ \infty & \text{en otro caso} \end{cases}$$

Entonces la conjugada convexa de  $\Theta$  está dada por la regla:

$$\Theta^*(-\pi) = \begin{cases} \int_{X \times Y} c(x, y) d\pi & \text{si } \pi \in M_+(X \times Y) \\ \infty & \text{si } \pi \notin M_+(X \times Y) \end{cases}$$

Además  $\Xi$  está bien definida y su función convexa conjugada es:

$$\Xi^*(\pi) = \chi_{E^* \setminus \Gamma(\mu, \nu)}^\infty(\pi)$$

Algo que esta prueba hace evidente es el hecho que  $\Theta$  y  $\Xi$  tienen alguna relación en su estructura. Pues,  $\Theta$  es una función indicadora infinita y su convexa conjugada se parece a una indicadora infinita pero difiere por una integral en términos de  $\pi$ , por otro lado  $\Xi$  se parece a una indicadora infinita pero difiere en una integral y su transformada si es una indicadora infinita.

### 4.3.5. La función convexa conjugada es en efecto convexa

Sea  $f$  una función cualquiera,  $f^*$  su convexa conjugada y consideremos  $x_1^*, x_2^* \in X^*$  y  $t \in (0, 1)$  entonces

$$\begin{aligned} f^*(tx_1^* + (1-t)x_2^*) &= \sup_{x \in X} \left\{ \langle tx_1^* + (1-t)x_2^*, x \rangle - f(x) \right\} = \sup_{x \in X} \left\{ t\langle x_1^*, x \rangle - tf(x) + (1-t)\langle x_2^*, x \rangle - (1-t)f(x) \right\} \\ &\leq t \sup_{x \in X} \left\{ \langle x_1^*, x \rangle - f(x) \right\} + (1-t) \sup_{y \in X} \left\{ \langle x_2^*, y \rangle - f(y) \right\} = tf^*(x_1^*) + (1-t)f^*(x_2^*) \\ f^*(tx_1^* + (1-t)x_2^*) &\leq tf^*(x_1^*) + (1-t)f^*(x_2^*). \end{aligned}$$

■

Es interesante notar que la transformada de Legendre-Frechet es convexa sin importar como es la función original. Esta propiedad nos otorga una herramienta importante para usar los conceptos y teoremas de funciones convexas para analizar propiedades específicas de la función original.

### 4.3.6. La función convexa conjugada invierte el orden

**Proposición 4.3.2.** Sean  $f, g$  dos funciones finitas con dominio en  $X$  y contradominio  $\mathbb{R}$ ,

$$f \leq g \Rightarrow g^* \leq f^*.$$

*Demostración.* Supongamos que  $f(x) \leq g(x) \quad \forall x \in X$  entonces :

$$\langle x^*, x \rangle - g(x) \leq \langle x^*, x \rangle - f(x).$$

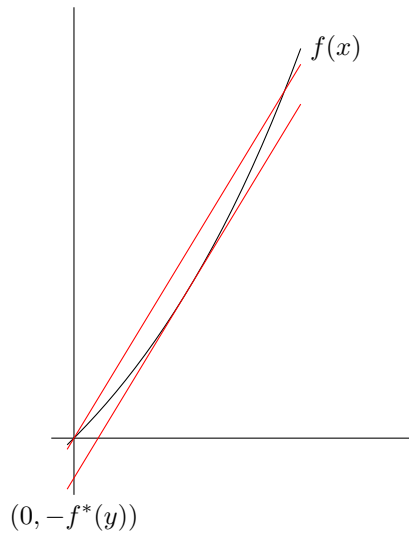
$$\Rightarrow g^*(x^*) = \sup_{x \in X} \left\{ \langle x^*, x \rangle - g(x) \right\} \leq \sup_{x \in X} \left\{ \langle x^*, x \rangle - f(x) \right\} = f^*(x^*).$$

■

Esta última propiedad ilustra la interpretación geométrica en el sentido que si la gráfica de la función  $f$  se encuentra por debajo de la función  $g$  entonces la separación que va a haber respecto a alguna recta será mayor.

### 4.3.7. Interpretación geométrica

La función convexa conjugada asigna a cada funcional lineal  $x^*$  la máxima diferencia entre la funcional lineal y la función. En el caso real, las funcionales lineales se pueden dibujar como rectas y la función conjugada es la máxima separación entre la función y la recta que pasa por el origen.



Esta interpretación se hace cuando las funciones se consideran sobre el espacio de los números reales, sin embargo como la función convexa conjugada está definida para cualquier espacio normado, la interpretación sólo se extiende en el sentido intuitivo.

## 4.4. Principios minimax

Llegamos al último preámbulo para poder estudiar el teorema de dualidad de Kantorovich, necesitamos hacer intercambios de supremos e ínfimos. No es lo mismo calcular el ínfimo de un conjunto de supremos que calcular el supremo de un conjunto de ínfimos. Estas ideas han sido plenamente desarrolladas en términos de análisis real con los conceptos de  $\liminf$  y  $\limsup$  y en teoría de la medida con los límites inferiores y superiores de conjuntos.

Por ejemplo sean  $A, B \in \mathcal{P}(X)$  dos conjuntos y  $f : X \times X \rightarrow \mathbb{R}$  una función bivariada se satisface la siguiente desigualdad:

$$\sup_{w \in B} \inf_{u \in A} f(u, w) \leq \inf_{u \in A} \sup_{w \in B} f(u, w)$$

*Demostración.* Sean  $u \in A, w \in B$  y entonces:

$$\inf_{u \in A} f(u, w) \leq f(u, w) \quad \forall u, \forall w$$

$$\sup_{w \in B} \inf_{u \in A} f(u, w) \leq f(u, w) \leq \sup_{w \in B} f(u, w) \quad \forall u, \forall w$$

$$\sup_{w \in B} \inf_{u \in A} f(u, w) \leq \inf_{u \in A} f(u, w) \leq \inf_{u \in A} \sup_{w \in B} f(u, w) \quad \forall w$$

$$\sup_{w \in B} \inf_{u \in A} f(u, w) \leq \inf_{u \in A} \sup_{w \in B} f(u, w).$$

Nos interesa saber bajo que condiciones de la función  $f$  o de los conjuntos  $A, B$  se obtiene la igualdad. Para recordar esta desigualdad se puede dar una nemotecnia que se basa en la esencia de la desigualdad: **El más grande de los pequeños debe ser más chico que el más pequeño de los grandes.** Por lo que esta desigualdad resulta natural.

Se denominan “principios minimax” a las formulaciones que permiten recuperar la igualdad de esta desigualdad en algún contexto. Los principios de minimax se han presentado en diversas áreas de las matemáticas y constituyen una herramienta fundamental para el análisis.

### 4.4.1. Algunas formulaciones de principios minimax

En esta sección se presentan, sin demostración, varias formulaciones de principios de minimax en diversas áreas. Finalmente, se prueba la versión de Rockafellar, que se usa para detallar el argumento formal del problema de dualidad de Kantorovich.

#### 1. En teoría de juegos (Lema de Von Neumann)

Una relación de preferencias para un individuo en un conjunto de consecuencias  $C$  es una relación binaria, reflexiva y completamente transitiva en  $C$ . Se puede pensar la relación de preferencias como un orden en el conjunto de posibles acciones  $A$ .

Un juego se denomina estrictamente competitivo si siempre que un individuo prefiera una acción sobre la otra, su rival hubiera escogido las acciones al revés.

Sea  $\{(1, 2), A_i, u_i\}$  un juego estratégico estrictamente competitivo. Es decir, tenemos dos individuos en el juego y tienen preferencias  $u_1, u_2$  y conjuntos de posibles acciones  $A_1, A_2$ . Una acción  $a \in A_1$  se denomina **maxminimizador** para el jugador 1 si

$$\min_{y \in A_2} u_1(a, y) \geq \min_{y \in A_2} u_1(x, y) \quad \forall x \in A$$

Análogamente se definen los maxminimizadores para el jugador 2.

**Lemma 4.4.1.** *En un juego estratégico estrictamente competitivo, tenemos*

$$\max_{v \in A_2} \min_{u \in A_1} u_1(u, v) = - \min_{v \in A_2} \max_{u \in A_1} u_2(u, v).$$

Nótese que esta proposición es diferente a los planteamientos que habíamos realizado, ya que se considera el mínimo sobre el conjunto que antes consideraba el máximo y viceversa pero fundamentalmente son las mismas ideas.

**Proposición 4.4.2.** *Para un juego estratégico estrictamente competitivo un punto es maxminimizador para ambos jugadores si y sólo si es un equilibrio de Nash. [20]*

El objetivo en los problemas de minimax es resolver el conflicto: **Maximizar sobre un conjunto mientras se minimiza sobre el otro**, con este enfoque es evidente que el problema en teoría de juegos califica como principio de minimax. Pues se minimiza en términos de la utilidad de un jugador mientras maximiza el otro jugador.

2. **Funciones semicontinuas** (Análisis Convexo) Sea  $L : A \times B \rightarrow \mathbb{R}$  una función bivariada. Un punto  $(x_0, y_0)$  se denomina punto silla si cumple que

$$L(x_0, y) \leq L(x_0, y_0) \leq L(x, y_0) \quad \forall x \in A, y \in B$$

**Proposición 4.4.3.** *L tiene un punto silla en  $A \times B$  si y sólo si*

$$\max_{p \in B} \inf_{u \in A} L(u, v) = \min_{u \in A} \sup_{p \in B} L(u, v).$$

**Proposición 4.4.4.** *Si además se agregan las hipótesis :*

- (H.1)  $A \subseteq V$  cerrado convexo no vacío.  
 (H.2)  $B \subseteq Z$  cerrado convexo no vacío.  
 (H.3)  $L(x, \cdot)$  es cóncava y semicontinua superior  $\forall x \in A$   
 (H.4)  $L(\cdot, y)$  es convexa y semicontinua inferior  $\forall y \in B$

*entonces el conjunto de puntos silla es un conjunto convexo y si las funciones son estrictamente convexas o estrictamente cóncavas respectivamente, entonces el punto silla es único.*

Lo que nosotros queremos es poder intercambiar ínfimos y supremos, esto es válido sólo en virtud de la existencia de puntos silla. Estas proposiciones sin embargo, no son suficientes para asegurar la existencia de puntos sillas, se debe considerar otra restricción.

**Proposición 4.4.5.** *Si además de las hipótesis  $(H,1) \rightarrow (H,4)$  se supone que los conjuntos  $A$  y  $B$  son acotados entonces la función  $L$  tiene al menos un punto silla.*

Las pruebas de estos teoremas no son difíciles de entender ni requieren de muchos conocimientos previos. Se invita al lector a estudiarlas en [12], no se incluyen aquí pues vamos probar la formulación de Rockafellar que será suficiente para para el entendimiento del argumento formal de la dualidad de Kantorovich.

3. **Decisión en Estadística Bayesiana** Si  $(X, S, \mathbb{P})$  es un espacio de medida y  $\Theta \subseteq \mathbb{R}$  es el conjunto de posibles valores de un parámetro  $\theta$ , empecemos con un enfoque frecuentista en el que  $\theta$  no tiene una distribución previa.  
 Una función de pérdida  $L : \mathbb{R} \times \mathbb{R}$  es una función que compara valores. Las funciones de pérdida más usuales son:

- Pérdida cuadrática:  $L(\theta, a) = (\theta - a)^2$
- Pérdida absoluta:  $L(\theta, a) = |\theta - a|$
- Pérdida cero-uno:  $L(\theta, a) = \chi_a(\theta)$

Una función  $d : X \rightarrow \mathbb{R}$  se denomina regla de decisión y para cada regla de decisión se define el riesgo de la regla:

$$R(\theta, d(x)) = \int_X L(\theta, d(x))p(x|\theta)dx.$$

Las reglas de decisión se pueden interpretar como la consecuencia del experimento que resultó en  $x$ . El objetivo es tener una noción de optimización de decisiones en función de la pérdida.

Se dice que una regla de decisión  $d_1$  es más eficaz (mejor) respecto al riesgo  $R$  que otra función de decisión  $d_2$  si

$$R(\theta, d_1) \leq R(\theta, d_2).$$

Las reglas de decisión se pueden interpretar como planes a seguir en determinados momentos. Para el experimento  $w$  la regla de decisión nos indica el valor numérico asociado  $d(w)$ . La pérdida mide la separación entre esta decisión y el valor real del parámetro; sin embargo, el riesgo depende del valor de  $\theta$  que no podemos controlar por lo que para tener estimaciones cautelosas el objetivo es encontrar una regla de decisión tal que

$$\sup_{\theta \in \Theta} \inf_d R(\theta, d) = \inf_d \sup_{\theta \in \Theta} R(\theta, d).$$

Con esta formulación es evidente que nos encontramos en un problema de minimax.

Hay que notar el conjunto sobre el cual se considera el ínfimo. En principio no se considera ninguna restricción por lo que el problema es encontrar una función cualquiera. Un problema de tal tipo no se puede resolver debido a la infinidad de posibles decisiones, por lo que generalmente se agregan hipótesis sobre la función.

La última formulación de minimax que se presenta es el teorema de dualidad de Rockafellar. Esta perspectiva radica en el uso de la transformada de Legendre-Fenchel.

#### 4. Dualidad de Rockafellar

**Teorema.** *Sea  $X$  un espacio normado y  $X^*$  su dual topológico.*

*Sean  $\Theta$  y  $\Xi$  dos funciones convexas con dominio  $X$  y contradominio  $\mathbb{R} \cup \{\infty\}$ , sean  $\Theta^*$  y  $\Xi^*$  sus funciones convexas conjugadas respectivamente.*

*Supongamos que existe  $z_0$  tal que  $\Theta(z_0) < \infty$ , también  $\Xi(z_0) < \infty$  y además  $\Theta$  es continua en  $z_0$  entonces*

$$\inf_{x \in X} \left\{ \Theta(x) + \Xi(x) \right\} = \max_{x^* \in X^*} \left\{ -\Theta^*(-x^*) - \Xi^*(x^*) \right\}.$$

*Demostración.* Seguimos la prueba de [9] pero la detallamos mucho más. Usaremos un resultado de análisis convexo: Si  $C$  es convexo no vacío entonces  $C^\circ$  es convexo y además:  $\bar{C} = \bar{C}^\circ$ , para este hecho ver [23] p.44 . Para simplificar la notación y el entendimiento de la prueba definamos

$$a = \inf_{x \in X} \left\{ \Theta(x) + \Xi(x) \right\}$$

$$b = \sup_{x^* \in X^*} \left\{ -\Theta^*(-x^*) - \Xi^*(x^*) \right\}.$$

entonces  $a$  corresponde al lado izquierdo en la igualdad de la conclusión del teorema. Por otro lado,  $b$  no es el lado derecho, por ser un supremo. Hay que ver que existe una funcional que alcanza este valor para que el supremo sea un máximo.

Queremos establecer la igualdad entre  $a$  y  $b$ .

Primeramente comenzamos por establecer una de las desigualdades:  $b \leq a$ .

Empezamos usando las definiciones de las transformadas como supremos, entonces para todo  $x \in X$  se tiene

$$\begin{aligned} -\langle x^*, x \rangle - \Theta(x) &\leq \Theta^*(-x^*) & \text{y} & & \langle x^*, x \rangle - \Xi(x) &\leq \Xi^*(x^*). \\ \Rightarrow -\Theta^*(-x^*) &\leq \langle x^*, x \rangle + \Theta(x) & \text{y} & & -\Xi^*(x^*) &\leq -\langle x^*, x \rangle + \Xi(x), \end{aligned}$$

Sumando las desigualdades notamos,

$$-\Theta(-x^*) - \Xi^*(x^*) \leq \Theta(x) + \Xi(x).$$

Con el ínfimo del lado derecho,

$$-\Theta(-x^*) - \Xi^*(x^*) \leq a.$$

Finalmente, usando el supremo en el lado izquierdo,

$$b \leq a.$$



Ahora bien, si  $a = -\infty$  entonces  $b = -\infty$  y por lo mismo  $a = b$ . Asumamos que  $a \in \mathbb{R}$ .

Sea  $C = \text{epi}(\Theta) = \{(x, \lambda) \in X \times \mathbb{R} : \Theta(x) \leq \lambda\}$  probemos  $C^\circ \neq \emptyset$ .

**Afirmamos que**  $(z_0, \Theta(z_0) + 1) \in C^\circ$ .

Pero  $C \subseteq X \times \mathbb{R}$ , por lo que necesitamos conocer la topología de este espacio, sin embargo podemos usar la norma infinito como norma:

$$\|(x, \lambda)\|_\infty = \max\{\|x\|, |\lambda|\}.$$

Por lo que afirmar que  $(z_0, \Theta(z_0) + 1) \in C^\circ$  es afirmar que existe una bola inducida por la norma infinito centrada en este punto y totalmente contenida en  $C$ , de hecho es de radio  $\delta$ .

Sea  $\epsilon \in (0, 1)$  como  $\Theta$  es continua en  $z_0$  existe  $\delta_\epsilon$  tal que

$$\|z - z_0\| < \delta_\epsilon \Rightarrow |\Theta(z) - \Theta(z_0)| < \epsilon.$$

Ponemos  $0 < \delta < \min\{\delta_\epsilon, 1 - \epsilon\}$  entonces  $\|z - z_0\| < \delta \Rightarrow |\Theta(z) - \Theta(z_0)| < \epsilon$ .

Sea  $(z, \mu) \in B_\delta^{\|\cdot\|_\infty}((z_0, \Theta(z_0) + 1))$  el objetivo es probar que  $\Theta(z) \leq \mu$ .

Por definición de la norma infinito tenemos que  $\|z - z_0\| < \delta$  y  $|\Theta(z_0) + 1 - \mu| < \delta$ .

De la segunda condición podemos reescribir a  $\mu$  en términos de  $\Theta(z_0) + 1$  y la diferencia. Es decir, la segunda condición nos permite escribir:

$$\Theta(z_0) + 1 = \mu + r \text{ con } |r| < \delta.$$

$$\Rightarrow \Theta(z_0) + \epsilon = \mu + (r - 1) + \epsilon.$$

Si  $r < 0$  entonces  $\Theta(z_0) + \epsilon < \Theta(z_0) + 1 < \mu$  pero si  $r > 0$  entonces  $1 - r > 1 - \delta$ .

$$\Theta(z_0) + \epsilon \leq \mu \Leftrightarrow (r - 1) + \epsilon \leq 0 \Leftrightarrow \epsilon \leq 1 - r.$$

Pero esta condición está asegurada pues  $\delta < 1 - \epsilon \Rightarrow \epsilon < 1 - \delta < 1 - r$ .

$$\Theta(z_0) + \epsilon \leq \mu.$$

Usamos la desigualdad del triángulo inversa en la condición:  $|\Theta(z_0) - \Theta(z)| < \epsilon$ .

$$|\Theta(z)| < |\Theta(z_0)| + \epsilon = \Theta(z_0) + \epsilon \leq \mu.$$

$$\Theta(z) \leq \mu.$$

Así,  $(z, \mu) \in C$  por lo que  $(z_0, \Theta(z_0) + 1) \in C^\circ$ . Esto nos indica que el interior de  $C$  es no vacío, cosa que necesitamos para el uso del Teorema de Hahn-Banach geométrico. Notamos que la hipótesis de continuidad es la que nos permite concluir que  $C^\circ \neq \emptyset$ .

Además como  $C$  es convexo y no vacío se tiene que  $\bar{C}$  es convexo y además  $\bar{C}^\circ = \bar{C}$  ver [23] p.44 .

Sea  $A = C^\circ$  y sea  $B = \{(x, \lambda) \in X \times \mathbb{R} : \lambda \leq a - \Xi(x)\}$  comprobemos que  $A \cap B = \emptyset$ .

Consideremos  $(x, \lambda) \in A$ , por definición  $\Theta(x) < \lambda$  y por definición de  $a$ ,  $a \leq \Theta(x) + \Xi(x)$

$$\Rightarrow a - \Xi(x) \leq \Theta(x) < \lambda \Rightarrow (x, \lambda) \notin B.$$

Como  $A$  y  $B$  son conjuntos convexos,  $A$  es abierto, no vacío y  $A \cap B = \emptyset$  aplicamos el **Teorema de Hahn-Banach en su forma geométrica**, por lo que existe un hiperplano cerrado de la forma  $[\Phi = \alpha]$  que separa a  $\bar{A} = C$  y a  $B$  en el sentido:

$$\begin{aligned}\Phi((x, \lambda)) &\geq \alpha \quad \forall (x, \lambda) \in C \\ \Phi((x, \lambda)) &\leq \alpha \quad \forall (x, \lambda) \in B.\end{aligned}$$

Pero  $\Phi$  es una funcional lineal distinta de cero sobre  $X \times \mathbb{R}$  entonces podemos reescribir:

$$\Phi((x, \lambda)) = \langle f, x \rangle + k\lambda \text{ donde } f \in X^* \text{ y } k \in \mathbb{R}.$$

**Queremos concluir que  $k > 0$ .**

Si  $(x, \lambda) \in C$  entonces  $(x, n\lambda) \in C, n \in \mathbb{N}$  debido a que  $\lambda \leq n\lambda$ . Como  $\alpha$  es un número real fijo, si fijamos un punto  $x$  para el que  $\langle f, x \rangle < \infty$  y si  $k < 0$  entonces por la propiedad arquimedea existe  $n$  tal que  $\langle f, x \rangle + k(n\lambda) < \alpha$  contradiciendo la conclusión del teorema de Hahn Banach. Por lo mismo  $k < 0$  es imposible.

$$k \geq 0.$$

Supongamos por reducción al absurdo que  $k = 0$  entonces :

$$\langle f, x \rangle \geq \alpha \quad \forall x \text{ tal que } \Theta(x) < \infty.$$

$$\langle f, x \rangle \leq \alpha \quad \forall x \text{ tal que } \Xi(x) < \infty.$$

Por hipótesis, en  $z_0$  ambas funciones son valores reales por lo que  $\langle f, z_0 \rangle \leq \alpha$  y también  $\langle f, z_0 \rangle \geq \alpha$ , que es equivalente a  $\langle f, z_0 \rangle = \alpha$ .

Como  $\Theta$  es continua en  $z_0$  existe  $\epsilon > 0$  tal que si  $z \in B_\epsilon^{\|\cdot\|}(z_0) \Rightarrow \Theta(z_0) < \infty$  pero en un espacio vectorial tenemos que  $B_\epsilon^{\|\cdot\|} = z_0 + \epsilon B_1^{\|\cdot\|}(0)$

$$\begin{aligned}\langle f, z_0 + \epsilon z \rangle &\geq \alpha \quad \forall z \in B_1^{\|\cdot\|}(0) \\ \langle f, z_0 \rangle + \epsilon \langle f, z \rangle &\geq \alpha \quad \forall z \in B_1^{\|\cdot\|}(0) \\ \epsilon \langle f, z \rangle &\geq 0 \quad \forall z \in B_1^{\|\cdot\|}(0).\end{aligned}$$

Pero si  $z \in B_1^{\|\cdot\|}(0)$  entonces  $-z \in B_1^{\|\cdot\|}(0)$  por lo que  $\langle f, z_0 \rangle \geq 0$  y  $-\langle f, z_0 \rangle \geq 0$  y por lo mismo  $\langle f, z \rangle = 0 \quad \forall z \in B_1^{\|\cdot\|}(0) \Rightarrow \|f\|_* = 0 \Rightarrow \Phi = 0$  lo cual es imposible.

$$\therefore k > 0$$

**$b$  se alcanza en  $x^* = -f/k$  y satisface la otra desigualdad  $b \geq a$**  En la condición que define a  $B$  se acepta la igualdad, es decir si ponemos  $\lambda = a - \Xi(x)$  el teorema nos permite concluir que si  $\Xi(x) < \infty$  entonces

$$\begin{aligned}\frac{\langle f, x \rangle}{k} + a - \Xi(x) &\leq \frac{\alpha}{k} \\ \frac{\langle f, x \rangle}{k} - \Xi(x) &\leq \frac{\alpha}{k} - a \\ \Xi^* \left( \frac{f}{k} \right) &= \sup_{x \in X} \left\{ \frac{\langle f, x \rangle}{k} - \Xi(x) \right\} \leq \frac{\alpha}{k} - a\end{aligned}$$

$$-\Xi^* \left( \frac{f}{k} \right) \geq a - \frac{\alpha}{k}.$$

Procedemos a evaluar la función conjugada convexa de  $\Theta$  en  $-f/k$

$$\Theta^* \left( -\frac{f}{k} \right) \leq - \left( \frac{\langle f, x \rangle}{k} + \Theta(x) \right) \leq \frac{-\alpha}{k} \Rightarrow -\Theta^* \left( -\frac{f}{k} \right) \geq \frac{\alpha}{k}$$

Combinemos estos dos resultados para obtener la conclusión del teorema:

$$-\Theta^* \left( -\frac{f}{k} \right) - \Xi^* \left( \frac{f}{k} \right) \geq -\frac{\alpha}{k} + a - \frac{\alpha}{k} = a$$

Finalmente  $b$  es el supremo del conjunto definido por la operación que aparece en el lado izquierdo de la desigualdad y por lo tanto:

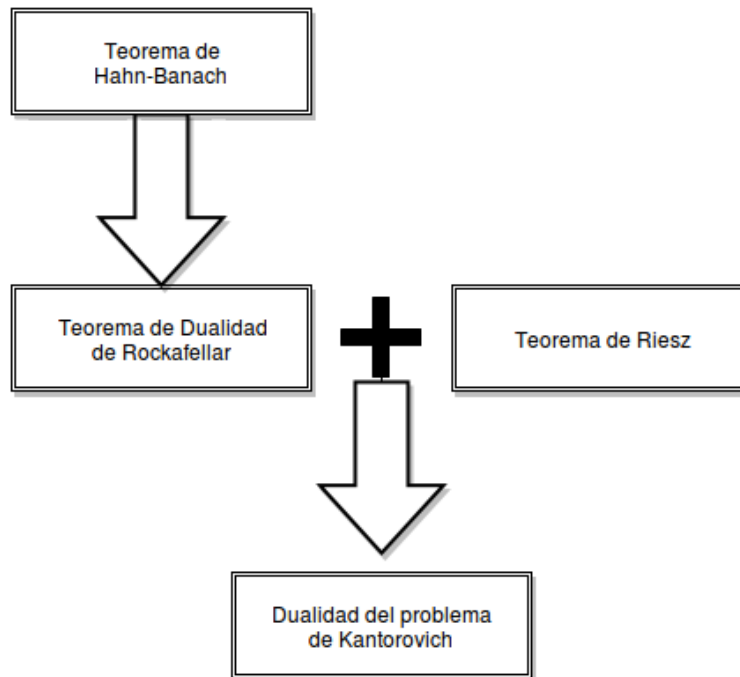
$$-\Theta^* \left( -\frac{f}{k} \right) - \Xi^* \left( \frac{f}{k} \right) = b = a.$$

Por lo tanto hemos encontrado que  $b$  y  $a$  son iguales y  $b$  de hecho se alcanza con la funcional  $\frac{f}{k}$  por lo que reescribimos el resultado con las definiciones de  $a$  y  $b$ :

$$\inf_{x \in X} \left\{ \Theta(x) + \Xi(x) \right\} = \max_{x \in X^*} \left\{ -\Theta(-x^*) - \Xi(x^*) \right\}.$$

■

## 4.5. Diagrama de seguimiento



## 4.6. Dualidad del problema de Kantorovich

El primer objetivo de este capítulo es encontrar el dual del problema de Kantorovich. Es decir queremos encontrar un problema de maximización cuya solución sea la solución del problema de transporte con la formulación de Kantorovich. Encontrar dicho dual, no es nada sencillo. Inclusive, casi todos los desarrollos que hemos hecho han sido parte de la preparación de las herramientas necesarias para entenderlo en su totalidad. Los teoremas de Prokhorov, Riesz y Rockafellar así como los conceptos de convexas conjugadas e indicadoras infinitas son fundamentales para el entendimiento de estas ideas.

En algunos libros, como en la referencia por excelencia [28], se acostumbra poner dos pruebas. Una prueba detallada pero no absolutamente justificada y la prueba rigurosa. La prueba rigurosa se basa en el teorema de Rockafellar que probamos en la sección pasada y por lo mismo no es necesario detenerse en la primera prueba.

### 4.6.1. Teorema de dualidad del problema de Kantorovich

**Teorema.** Sean  $X$  y  $Y$  espacios compactos.

Sea  $\mu$  una medida de probabilidad definida en  $X$  y  $\nu$  una medida de probabilidad definida en  $Y$ .

Sea  $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$  una función continua no negativa.

Sea  $\Gamma(\mu, \nu)$  el conjunto de medidas de probabilidad de Borel en  $X \times Y$  tales que su proyección en  $X$  sea  $\mu$  y su proyección en  $Y$  sea  $\nu$

Sea  $\Phi_c$  el conjunto de pares de funciones medibles  $(\phi, \psi) \in L_1(\mu) \times L_1(\nu)$  tales que

$$\phi(x) + \psi(y) \leq c(x, y)$$

$\mu$ -casi donde quiera en  $x$  y  $\nu$ -casi dondequiera en  $y$ .

Entonces

$$\inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c(x, y) d\pi \right\} = \sup_{(\phi, \psi) \in \Phi_c} \left\{ \int \phi d\mu + \int \psi d\nu \right\}$$

Además, el problema de Kantorovich admite un minimizador.

### 4.6.2. De la dualidad débil

**Teorema.** (Dualidad débil)

Bajo las hipótesis del teorema anterior, se tiene:

$$\sup_{\substack{(\phi, \psi) \in \Phi_c \\ \phi \in C_b(X), \psi \in C_b(Y)}} \left\{ \int \phi d\mu + \int \psi d\nu \right\} \leq \sup_{(\phi, \psi) \in \Phi_c} \left\{ \int \phi d\mu + \int \psi d\nu \right\} \leq \inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c(x, y) d\pi \right\}.$$

*Demostración.* La desigualdad de la izquierda se obtiene directamente pues al tomar el supremo sobre un subconjunto este no puede ser mayor que el supremo del conjunto donde está contenido.

Para la segunda desigualdad, sean  $\phi, \psi \in \Phi_c$ . Por definición de  $\Phi_c$  se tiene que

$$\phi(x) + \psi(y) \leq c(x, y)$$

para todo  $x$  excepto en un conjunto  $\mu$ -nulo  $N_1$  y todo  $y$  excepto en un conjunto  $\nu$ -nulo  $N_2$ .

Si  $\pi \in \Gamma(\mu, \nu)$  entonces  $N_1 \times N_2$  está en la  $\sigma$ -álgebra producto y observamos  $\pi(N_1 \times N_2) \leq \pi(N_1 \times Y) = \mu(N_1) = 0 = \nu(N_2)$  entonces como  $N_1 \times N_2$  es un conjunto  $\pi$ -nulo, podemos integrar ambos lados de la desigualdad con respecto a  $\pi$  sin preocuparnos por los conjuntos nulos en sus respectivos espacios.

$$\int \phi(x)d\mu + \int \psi(y)d\nu = \int (\phi + \psi)d\pi \leq \int c(x, y)d\pi.$$

Finalmente, considerando el ínfimo del lado derecho y el supremo del lado izquierdo se concluye la desigualdad. ■

### 4.6.3. Ideas generales y bosquejo de la prueba

Antes de hacer la prueba del teorema de dualidad hagamos un bosquejo del procedimiento, para la prueba del teorema general (5.1) de dualidad se seguirán 2 pasos:

1. Probar el resultado para espacios compactos  $X, Y$ . En esta parte se usa el teorema de Riesz y el principio de minimax de Rockafellar. Esta parte de la prueba será sencilla para nosotros porque ya hicimos gran parte del trabajo: **ya demostramos la dualidad de Rockafellar** y además en el “Ejemplo fundamental de la transformada” **ya encontramos las funciones convexas conjugadas**.
2. En la segunda parte se usa el teorema de Prokhorov para justificar la existencia del minimizador sin apelar a las ideas del teorema de Hahn-Banach.

**Observación 4.6.1.** *El teorema que presentamos en este documento está lejos de ser el más general. Incluso en [28] se hace un paso más para considerar  $c$  sólo semicontinua inferiormente y  $X, Y$  espacios polacos. Además existe literatura sobre el caso en espacios únicamente topológicos e inclusive en espacios más generales. De hecho, existen versiones de estas ideas en grupos. [2]*

### 4.6.4. Primer paso de la prueba

Con las hipótesis del teorema, consideremos que  $X, Y$  son espacios compactos.

Si  $E = C_b(X \times Y)$  entonces por el teorema de Riesz probado en el capítulo anterior tenemos que las funcionales lineales positivas son las medidas de Borel no negativas en  $X \times Y$  dotado de la norma variación. Exactamente como hicimos en **El ejemplo fundamental de la transformada** Sean  $\Theta$  y  $\Xi$  dos funciones definidas en el espacio de funciones continuas acotadas y dadas por las fórmulas:

1.  $\Theta(u) = \chi_{E \setminus \{u \geq -c\}}^\infty(u)$

- 2.

$$\Xi(u) = \begin{cases} \int_X \phi d\mu + \int_Y \psi d\nu & \text{si } u(x, y) = \phi(x) + \psi(y) \\ \infty & \text{en otro caso} \end{cases}$$

Entonces **como ya demostramos en esa sección** (4.2.4), la conjugada convexa de  $\Theta$  está dada por la regla:

$$\Theta^*(-\pi) = \begin{cases} \int_{X \times Y} c(x, y)d\pi & \text{si } \pi \in M_+(X \times Y) \\ \infty & \text{si } \pi \notin M_+(X \times Y) \end{cases}$$

Además  $\Xi$  está bien definida y su función convexa conjugada es:

$$\Xi^*(\pi) = \chi_{E^* \setminus \Gamma(\mu, \nu)}^\infty(\pi).$$

Hay que verificar que podemos usar el teorema de Rockafellar, por lo que necesitamos ver que  $\Theta$  y  $\Xi$  son funciones convexas, que existe un punto  $z_0$  tal que  $\Theta(z_0) < \infty$  y  $\Xi(z_0) < \infty$  y además que  $\Theta$  sea continua en  $z_0$ .

La convexidad de las funciones se obtiene directamente de sus definiciones. Sea  $z_0$  la función constante e igual a uno.

Como  $c$  es positiva,  $-c \leq 0 \leq z_0$  y si  $0 < \epsilon < 1$  y  $f$  está en la bola de norma infinito centrada en  $z_0$  con radio  $\epsilon$  entonces  $-c \leq f$  y además  $\Theta(z_0) = 0 = \Theta(f)$ , por lo que  $\Theta$  es continua y finita en  $z_0$ . Además  $z_0 = (1/2) + (1/2)$  que podemos usar como funciones en la definición de  $\Xi$  y por lo mismo  $\Xi(z_0) = 1 < \infty$  por lo que se satisfacen las hipótesis del teorema de dualidad de Rockafellar. Aplicamos este teorema y la conclusión dice:

$$\inf_{f \in C_b} \left\{ \Theta(f) + \Xi(f) \right\} = \max_{\pi \in M(X \times Y)} \left\{ -\Theta^*(-\pi) - \Xi^*(\pi) \right\}.$$

Calculemos ambos lados de la igualdad usando las reglas de correspondencia encontradas:

- Para el lado izquierdo queremos encontrar el ínfimo de  $\Theta + \Xi$ . No consideramos los conjuntos donde alguna de ellas es infinito porque eso no corresponderá a óptimos.

$$\Rightarrow \inf_{f \in C_b} \left\{ \Theta(f) + \Xi(f) \right\} = \inf_{f \in C_b} \left\{ \int \phi(x) d\mu + \int \psi(y) d\nu : f \geq -c, f(x, y) = \phi(x) + \psi(y) \right\}.$$

Como  $-f \in C_b$  podemos reescribir estas condiciones en términos de  $-f$ . Además podemos poner las dos restricciones en una sola pues si  $-f(x, y) = -(\phi(x) + \psi(y))$  y  $-f \geq -c(x, y)$  entonces podemos escribir sólo  $\phi(x) + \psi(y) \leq c(x, y)$

$$\begin{aligned} \Rightarrow \inf_{f \in C_b} \left\{ \Theta(f) + \Xi(f) \right\} &= \inf_{(\phi, \psi) \in C_b} \left\{ - \left( \int \phi(x) d\mu + \int \psi(y) d\nu \right) : \phi(x) + \psi(y) \leq c \right\} \\ &= - \sup_{\substack{(\phi, \psi) \in \Phi_c \\ \phi \in C_b(X), \psi \in C_b(Y)}} \left\{ \int \phi(x) d\mu + \int \psi(y) d\nu \right\}. \end{aligned}$$

Por lo que observamos que el lado izquierdo de la conclusión del Teorema de Rockafellar coincide con el lado izquierdo en la desigualdad de dualidad débil.

- Para el lado derecho, queremos encontrar un máximo por lo que los valores en los que  $-\Theta^*$  y  $-\Xi$  resultan en  $-\infty$ , usando esta condición se obtiene lo siguiente:

$$\begin{aligned} \max_{\pi \in M(X \times Y)} \left\{ -\Theta^*(-\pi) - \Xi^*(\pi) \right\} &= \max_{\pi \in \Gamma(\mu, \nu)} \left\{ - \int_{X \times Y} c(x, y) d\pi \right\} \\ &= - \min_{\pi \in \Gamma(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) d\pi \right\}. \end{aligned}$$

Juntando estos dos resultados y quitando el signo menos, el teorema de Rockafellar nos dice:

$$\sup_{\substack{(\phi, \psi) \in \Phi_c \\ \phi \in C_b(X), \psi \in C_b(Y)}} \left\{ \int \phi(x) d\mu + \int \psi(y) d\nu \right\} = \min_{\pi \in \Gamma(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) d\pi \right\}.$$

De este modo, usando dualidad débil, se obtiene el resultado. ■

Notemos que como el lado derecho es un mínimo y no un ínfimo esto expresa que el problema de Kantorovich admite un minimizador.

#### 4.6.5. Segunda parte de la prueba: sobre la existencia del minimizador

A pesar de que se ha concluido el resultado del teorema (5.1), es de gran ayuda analizar la existencia del minimizador a partir de la estructura topológica del conjunto sobre el que se toma el supremo. El teorema resulta en un mínimo debido al teorema de Hahn-Banach. Pero para poder hacer este argumento usamos que  $X, Y$  eran compactos para que las funciones acotadas tuvieran como dual al espacio de medidas. El objetivo de esta sección no es probar el teorema de dualidad cuando  $X$  y  $Y$  no son compactos sino dar los argumentos de la existencia del minimizador en términos de la estructura de  $\Gamma(\mu, \nu)$ .

**Proposición 4.6.2.**  $\mu$  y  $\nu$  medidas de Borel de probabilidad en espacios  $X, Y$  métricos separables entonces  $\mu$  y  $\nu$  son tirantes en el sentido que para todo  $\epsilon > 0$  existen compactos  $K_1, K_2$  tales que

$$\mu(X \setminus K_1) < \epsilon \text{ y } \nu(Y \setminus K_2) < \epsilon.$$

*Demostración.* Por la prueba en el apéndice D es directo pues ambas medidas son regulares. Entonces  $\mu(X) = \sup\{\mu(K) : K \subseteq X, K \text{ compacto}\}$  y por definición de supremo, para  $\epsilon > 0$  dado existe  $K_\epsilon$  con la propiedad deseada. La prueba es análoga para  $\nu$ . ■

**Proposición 4.6.3.** El teorema de Kantorovich admite minimizador.

*Demostración.* Sean  $K_1$  y  $K_2$  conjuntos compactos como en la proposición anterior. Entonces  $K_1 \times K_2$  es un conjunto compacto en  $X \times Y$  y si  $\pi \in \Gamma(\mu, \nu)$  se observa

$$\pi((X \times Y) \setminus K_1 \times K_2) \leq \mu(X \setminus K_1) + \nu(Y \setminus K_2) < 2\epsilon$$

Por lo que si se obtienen  $K_1$  y  $K_2$  para  $\frac{\epsilon}{2}$ , la familia  $\Gamma(\mu, \nu)$  es tirante, cumpliendo  $K_1 \times K_2$  su definición. Se deduce usando el teorema de Prokhorov que  $\Gamma(\mu, \nu)$  es débilmente compacta por sucesiones. Por lo mismo, para obtener el resultado necesitamos una sucesión débilmente convergente en  $\Gamma(\mu, \nu)$  para que el teorema nos otorgue el límite.

Sea  $(\pi_n)_{n=1}^\infty$  una sucesión de elementos en  $\Gamma(\mu, \nu)$  tal que

$$\int c(x, y) d\pi_n \rightarrow \inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c(x, y) d\pi \right\}.$$

Por el teorema de Prokhorov existe  $\pi^* \in \Gamma(\mu, \nu)$  para la cual se alcanza el ínfimo, es decir  $\pi^*$  corresponde al minimizador del problema de Kantorovich. ■

#### 4.6.6. Comentarios sobre el capítulo

Es importante tener asegurada la existencia del minimizador en el problema de Kantorovich, puesto que podría ocurrir que el ínfimo existiera sin alcanzarse. Por lo mismo, el último resultado es fundamental para la teoría de transporte.

En este capítulo se hizo el desarrollo en la formulación con sólo dos probabilidades,  $\mu$  y  $\nu$ , sin embargo, éste no es el caso más general. El desarrollo se puede hacer para espacios productos de dimensión arbitraria, pero muchos de los argumentos se complican. Por ejemplo, argumentar que  $\Gamma$  es una familia tirante en otras dimensiones apela al teorema de Tychonoff. El enfoque de usar únicamente dos probabilidades es excelente para dar las ideas generales, inclusive es preferible entender en su totalidad el caso de dos probabilidades y luego intentar pasar a la generalidad que estudiar el caso general directamente, pues resulta muy técnico y mucho más difícil de probar y de entender.

## El problema de Monge

### 5.1. Formulación del problema de Monge

El problema de Gaspard Monge plantea encontrar la mejor función de transporte entre dos medidas con respecto a una función de costos  $c$ .

El objetivo es encontrar una función de transporte, que nos indique de que manera tenemos que mover cada partícula para minimizar el costo total.

La formulación de Monge es mucho más intuitiva que la de Kantorovich, debido a que es más natural pensar en una regla que nos diga qué hacer con cada punto, que pensar en la distribución conjunta de los puntos. Sin embargo, a pesar de ser más intuitivo, el problema de Monge es mucho más difícil de resolver porque la condición que debe cumplir la función  $T$  es altamente no trivial.

**Definición.** (*Función de Transporte*)

Sean  $(X, S_X, \mu)$  y  $(Y, S_Y, \nu)$  dos espacios de medida. Decimos que una función  $(S_X - S_Y)$ -medible,  $T : X \rightarrow Y$ , es una función de transporte que transporta a  $\mu$  en  $\nu$  si cumple  $\forall A \in S_Y$

$$\mu(T^{-1}(A)) = \nu(A).$$

Si  $T$  es una función que transporta a  $\mu$  en  $\nu$  denotamos  $T_{\#}\mu = \nu$

**Definición.** (*Problema de Monge*)

Sean  $(X, S_X, \mu), (Y, S_Y, \nu)$  espacios de probabilidades y  $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$  una función continua no negativa. Entonces el objetivo es encontrar

$$\inf_{T_{\#}\mu = \nu} \left\{ \int c(x, Tx) d\mu \right\}.$$

Además de calcular este ínfimo, nos interesa encontrar (cuando exista) una función de transporte  $T$  que alcance este valor.

Fácilmente, se observa una similitud entre el problema de Monge y el de Kantorovich. Sin embargo no es evidente cómo se relacionan estos dos problemas. En esta sección se detallará la relación entre ambos. Además, se harán explícitos ejemplos sencillos donde los problemas difieren y se observarán las condiciones que nos permiten pasar de uno a otro.



## 5.2. Relación entre el problema de Monge y el de Kantorovich

**Teorema.** *El problema de Kantorovich es una relajación del problema de Monge.*

*Demostración.* Sea  $T$  tal que  $T_{\#}\mu = \nu$  y definamos para todo  $A$  en la  $\sigma$ -álgebra producto

$$\gamma_T(A) = \mu(\{x \in X : (x, Tx) \in A\})$$

¿Qué pasaría si  $\gamma_T \in \Gamma(\mu, \nu)$ ?

Esto concluiría que a cada elemento del conjunto sobre el que se considera el ínfimo en el problema de Monge, corresponde un elemento en el conjunto en sobre el que se considera el ínfimo de Kantorovich. Por ende, el problema de Kantorovich sería una relajación del problema de Monge.

Probemos que en efecto,  $\gamma_T \in \Gamma(\mu, \nu)$

1.  $\gamma_T$  es una medida de probabilidad en el espacio producto.

$$\gamma_T(\emptyset) = \mu(\emptyset) = 0, \gamma_T(X \times Y) = \mu(\{x : (x, Tx) \in X \times Y\}) = \mu(X) = 1$$

Si  $E \subseteq F \in S_{X \times Y}$  entonces  $(x, Tx) \in E \Rightarrow (x, Tx) \in F$

$$\Rightarrow \{x : (x, Tx) \in E\} \subseteq \{x : (x, Tx) \in F\} \Rightarrow \gamma_T(E) \leq \gamma_T(F)$$

Si  $(E_k)$  es una sucesión disjunta en  $S_{X \times Y}$  entonces no existe  $x$  tal que  $(x, Tx)$  esté en más de un  $E_i$  y para hacer la prueba más clara notemos

$$x \in \bigcup_{i=1}^{\infty} \{x : (x, Tx) \in E_i\} \Leftrightarrow (x, Tx) \in E_k \text{ p.a. } k \in \mathbb{N} \Leftrightarrow x \in \left\{ x : (x, Tx) \in \bigcup_{i=1}^{\infty} E_i \right\}.$$

Por lo que podemos usar esta igualdad de conjuntos y la  $\sigma$ -aditividad de  $\mu$  para concluir la  $\sigma$ -aditividad de  $\gamma_T$

$$\gamma_T \left( \bigcup_{i=1}^{\infty} E_i \right) = \mu \left( \bigcup_{i=1}^{\infty} \{x : (x, Tx) \in E_i\} \right) = \sum_{i=1}^{\infty} \mu(\{x : (x, Tx) \in E_i\}) = \sum_{i=1}^{\infty} \gamma_T(E_i)$$

2.  $\pi_x(\gamma_T) = \mu$

$$\gamma_T(A \times Y) = \mu(\{x : (x, Tx) \in A \times Y\}) = \mu(\{x : x \in A\}) = \mu(A)$$

3.  $\pi_y(\gamma_T) = \nu$

$$\gamma_T(X \times B) = \mu(\{x : (x, Tx) \in X \times B\}) = \mu(\{x : Tx \in B\}) = \mu(T^{-1}(B)) = \nu(B).$$

■

Por lo mismo, queda probado que el problema de Kantorovich es una relajación del problema de Monge. Los resultados sobre la existencia de soluciones del problema de Kantorovich no son aplicables a la solución de Monge porque el problema de Kantorovich permite muchos más elementos que el de Monge. Dada una función de probabilidad en el espacio producto, digamos  $\pi$ , no necesariamente se puede recuperar una función de transporte tal que  $\nu$  sea la medida *push-forward* de  $\mu$  respecto a esa función. Esto es ejemplificado por los casos de las deltas de Dirac de este capítulo.

Las diferencias básicas entre estos problemas radican en la estructura de las medidas. Para esto, es importante considerar casos extremos en medidas, como son las deltas de Dirac. Muchos resultados en el problema de Monge exigen que las medidas no acumulen masa en conjuntos determinados, para formalizar esta noción se usa la idea de átomo de una medida.

### 5.3. Átomos y masas de Dirac

La palabra átomo proviene del griego y se refiere a lo indivisible. En teoría de la medida se intenta replicar el concepto de indivisibilidad en términos de medidas.

**Definición. Átomo**

En un espacio de medida  $(X, S_X, \mu)$  un conjunto  $A \in S_X$  se denomina átomo si

1.  $\mu(A) > 0$
2.  $E \in S_X, E \subsetneq A \Rightarrow \mu(E) = 0$

Es equivalente decir que  $A$  es un átomo de  $\mu$ , a decir que es un átomo sobre  $(S_X, \mu)$  sin embargo en la literatura se encuentran ambas expresiones.

**Definición. Puramente Atómica**

Una medida  $\mu$  en un espacio medible  $(X, S_X)$  se denomina puramente atómica si existe un conjunto contable  $C \in S_X$  tal que

$$\mu(X \setminus C) = 0.$$

**Observación 5.3.1.** Los conceptos de puramente atómica y átomo no son equivalentes. En principio, uno se refiere a medidas y el otro a conjuntos. En la definición de átomo no se habla de cardinalidad de conjuntos, un átomo puede ser no numerable. También puede existir un conjunto  $C$  que cumpla la definición para una medida atómica sin ser un átomo, pues pueden sobrarle elementos. Ambos conceptos nos ayudan a pensar en la concentración de masa.

**Observación 5.3.2.** Si  $X$  es un espacio métrico y  $S_X$  es al menos tan fina como la  $\sigma$ -álgebra de Borel, y  $\mu$  es una medida puramente atómica en  $(X, S_X)$ , entonces  $\mu$  tiene al menos un átomo.

*Demostración.* Si  $\mu$  es puramente atómica, existe un conjunto contable  $C$  tal que  $\mu(X \setminus C) = 0$  por lo que  $\mu(C) = 1$ , como  $C$  es contable lo podemos escribir  $C = \{c_1, c_2, \dots\}$ , o bien,  $C = \bigcup_{k=1}^{\infty} \{c_k\}$  y por  $\sigma$ -

subaditividad  $1 = \mu(C) \leq \sum_{n=1}^{\infty} \mu(\{c_n\})$  por lo que algún  $\{c_n\}$  es un átomo.

En muchos problemas de física, así como en ecuaciones diferenciales es importante definir funciones que no hagan nada por mucho tiempo, “exploten” en algún punto y continúen sin hacer nada posteriormente. Esta necesidad genera el desarrollo de la función delta de Dirac. Se busca una función con el valor  $\infty$  en un sólo punto y 0 en cualquier otro lugar, que sea medible y que se pueda pensar como una función de densidad, es decir, buscan una función  $f$  tal que

$$f(x) = \begin{cases} \infty & \text{si } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \quad \text{y que cumpla que } \int_{-\infty}^{\infty} f(x) dx = 1$$

**Ninguna función que conozcamos puede satisfacer estas propiedades.**

Por lo que existen varios enfoques:

1. Teoría de distribuciones ó funciones generalizadas.
2. Aproximación con sucesiones de funciones.
3. Aproximación con funciones y medidas.

El primer enfoque requiere de una construcción propia, suele considerarse como un capítulo extra en algunos libros de análisis. El segundo enfoque es relativamente sencillo, consideremos la siguiente sucesión de funciones:

$$f_n(x) = \begin{cases} \frac{n}{2} & \text{si } |x| \leq \frac{1}{n} \\ 0 & \text{otro caso} \end{cases}$$

De este modo  $f_n \rightarrow f$  puntualmente y  $\int f_n(t)dt = 1$

**pero no hay convergencia a la integral del límite**, pues la función límite no es una función Riemann-integrable.

Por estas razones, adoptaremos el tercer enfoque, que se detallará a continuación.

**Definición.** *Medida de Dirac*

Sea  $(X, S)$  un espacio medible, y  $x_0$  un punto en  $X$ . Definimos la medida de Dirac en  $x_0$  como la función conjuntista  $\delta_{x_0} : S \rightarrow \mathbb{R}$  dada por la siguiente regla de correspondencia:

$$\delta_{x_0}(A) = \begin{cases} 1 & \text{si } x_0 \in A \\ 0 & \text{si } x_0 \notin A \end{cases}$$

**Observación 5.3.3.** *La delta de Dirac que hemos definido, es el símil en teoría de la medida de las funciones que se consideraron en la introducción. Sin embargo, no son exactamente lo que se buscaba, son medidas. Por lo que ni siquiera podemos pensar en la evaluación en el punto  $x_0$ , ya que el dominio de la función es una  $\sigma$ -álgebra de conjuntos.*

**Proposición 5.3.4.** *Para todo  $x_0 \in X$ ,  $\delta_{x_0}$  es una medida en el espacio medible  $(X, S)$*

*Demostración.* Es evidente y se omite.

**Observación 5.3.5.** *Por definición, una delta de Dirac  $\delta_{x_0}$ , es una medida puramente atómica, pues  $C = \{x_0\}$  satisface la definición siempre que la  $\sigma$ -álgebra sea al menos tan fina como la de Borel, con respecto a la distancia del espacio, para poder discernir al conjunto  $\{x_0\}$  y también  $\{x_0\}$  es un átomo.*

**Observación 5.3.6.** *En cualquier espacio medible que permita medir a los unitarios podemos construir medidas de probabilidad discretas.*

*A partir de cualquier conjunto numerable de puntos  $A$  podemos construir una medida de Dirac compuesta como:*

$$\delta_A(B) = \sum_{a \in A} c_a \delta_a(B) \quad \forall B \in S_X$$

*bajo la restricción de que las constantes  $c_\alpha$  sean positivas y  $\sum_{a \in A} c_a = 1$*

**Observación 5.3.7.** *De la definición de las medidas de probabilidad discretas se nota que son puramente atómicas y tienen como átomos al menos una unión de singuletes.*

### 5.3.1. La otra perspectiva de las deltas de Dirac

Hasta el momento, hemos desarrollado las ideas de deltas de Dirac fijando un punto  $x_0$  y considerando todos los posibles conjuntos en la  $\sigma$ -álgebra.

Hemos pensado en las funciones con el enfoque  $\delta_{x_0}(\cdot) : S \rightarrow \mathbb{R}$ .

Pero ¿qué pasa si nos fijamos en la delta de Dirac desde la otra perspectiva?. Es decir pensemos que  $A \in S$  está fijo y ahora lo que movemos es el punto.

Es decir, nos preguntamos: ¿quién es  $F(a)$  si definimos  $F : X \rightarrow \mathbb{R}$  dada por  $F(a) = \delta_a(A)$  ?

$$F(a) = \delta_a(A) = \chi_A(a)$$

Por lo que analizar las deltas de Dirac es darle la vuelta al problema de analizar indicadoras. Esta técnica usa medidas para estudiar indicadoras desde otro enfoque.

El objetivo de tener bien definidas las medidas discretas en espacios generales, es que facilitan los cálculos de supremos e ínfimos. Estas medidas son casos extremos, por lo que cuando se consideran supremos sobre todas las medidas de probabilidad, es usual que una delta de Dirac nos permita concluir si el ínfimo es finito o no lo es.

En el problema de Monge nos interesa saber cuándo podemos encontrar una función que transporte de manera óptima, la existencia de tal función no se puede asegurar en el caso de medidas atómicas como veremos en las siguientes secciones.

## 5.4. Ejemplos básicos del problema de Monge

El objetivo de estos ejemplos es notar que a pesar de que consideremos espacios “generosos” en el sentido de aportarnos una estructura topológica deseable, podemos dar ejemplos donde el problema no tenga solución.

### 5.4.1. Múltiples soluciones

Sea  $X = Y = \mathbb{R}^2$ ,  $S_X = S_Y = \mathbb{B}_{\mathbb{R}^2}$  y sean  $x_1, x_2, x_3, x_4 \in \mathbb{R}^2$  definamos

$$\mu = \frac{1}{2}(\delta_{x_1} + \delta_{x_2}) \text{ y } \nu = \frac{1}{2}(\delta_{x_3} + \delta_{x_4})$$

Es decir concentramos la masa de  $\mu$  en dos puntos  $x_1$  y  $x_2$  y toda la masa de  $\nu$  en  $x_3$  y  $x_4$ . Nótese que todavía no imponemos ninguna función de costos, pero las ideas empiezan a ser evidentes. La existencia del transporte óptimo dependerá de la función de costos y de la ubicación de los puntos.

Empecemos con los vértices de un cuadrado centrado en cero:

$$\hat{x}_1 = (-1, 1), \hat{x}_2 = (1, -1), \hat{x}_3 = (-1, -1), \hat{x}_4 = (1, 1)$$

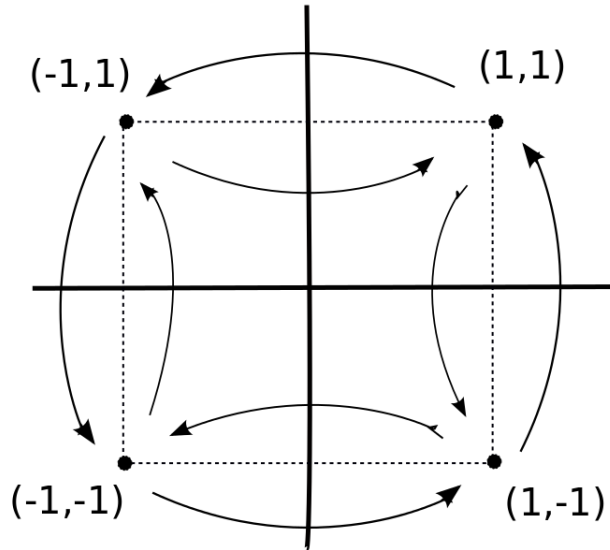
Nos preguntamos por las posibles funciones de transporte. Estas tendrían que cumplir

$$\mu(T^{-1}(\hat{x}_3)) = \frac{1}{2} = \mu(T^{-1}(\hat{x}_4)).$$

Pero como  $\mu$  sólo asigna probabilidades positivas a los singuletes  $\{\hat{x}_1\}$  y  $\{\hat{x}_2\}$  entonces sólo existen dos funciones de transporte:

- La rotación hacia la izquierda de 90 grados  $\left(\frac{\pi}{2}\right)$
- La rotación hacia la derecha de 90 grados  $\left(\frac{-\pi}{2}\right)$

Y si  $c(\hat{x}, \hat{y}) = \|\hat{x} - \hat{y}\|_2$  entonces  $\int c(\hat{x}, T\hat{x})d\mu = \|(1, 1) - (-1, 1)\|_2 = 1$  para ambas funciones de transporte. Por lo que **ambas funciones son óptimas** y de hecho son todas las posibles funciones de transporte. Como la función de costos sólo pondera la distancia recorrida, no hay diferencia entre ambas soluciones. Esto no debe ser de nuestro agrado pues nos ilustra que no podemos obtener una mejor solución, inclusive en un caso muy básico. Además, este caso se puede generalizar de manera rápida, porque lo único que necesitamos es considerar puntos equidistantes. En dimensiones mayores, esta idea se puede replicar.



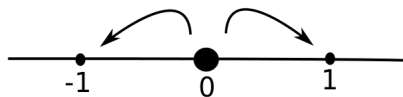
### 5.4.2. Dividir un átomo

Sea  $X = Y = \mathbb{R}$  y  $S_X = S_Y = \mathbb{B}_{\mathbb{R}}$  definamos  $\mu = \delta_0$  y  $\nu = \frac{1}{2}(\delta_1 + \delta_{-1})$ .

En este caso, estamos intentando transportar la masa concentrada en un punto a una función con dos puntos de masa de probabilidad. Si nos preguntamos por las funciones de transporte entre  $\mu$  y  $\nu$  observamos que se tienen que cumplir

$$\frac{1}{2} = \mu(T^{-1}(-1)) = \mu(T^{-1}(1)).$$

pero no existe ningún conjunto al que  $\mu$  asigne probabilidad  $1/2$ . Por lo que  $T$  no puede ser definida. Encontramos que no existen funciones de transporte entre  $\mu$  y  $\nu$  porque **no se puede dividir el átomo**.

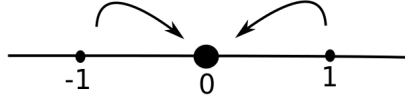


### 5.4.3. Juntar dos átomos

Consideremos el caso del ejemplo anterior, pero al revés. Intentemos resolver el problema del transporte en  $X = Y = \mathbb{R}$  y  $S_X = S_Y = \mathbb{B}_{\mathbb{R}}$  y consideremos  $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$  y  $\nu = \delta_0$

La condición de transporte es  $1 = \mu(T^{-1}(-1)) = \mu(T^{-1}(1))$ . Esta condición sí se puede cumplir.

Sea  $T = \chi_{\{-1\} \cup \{1\}}$  entonces  $T$  satisface la condición del transporte, sin embargo  $T$  podría no ser óptima, ni siquiera hemos definido la función de costos  $c$ . Concluimos que en general, no sabemos si hay un óptimo en este problema pero al menos siempre hay un conjunto sobre el cual considerar el ínfimo.



Estos problemas ilustran 3 tipos de patologías atómicas:

- En el primer caso vimos que podemos tener más de una solución del problema, no necesariamente hay unicidad. Inclusive puede pasar que todas las funciones de transporte integren a lo mismo con la función de costos con respecto a la medida de  $\mu$
- En el segundo caso observamos que si  $\mu$  y  $\nu$  son puramente atómicas, puede no existir alguna función de transporte. Puede ser que no podamos encontrar funciones de transporte.
- En el tercer ejemplo observamos que a pesar de tener bien definidas las funciones de transporte, la función de costos puede hacer que no se alcance el ínfimo o que se alcance en muchas funciones.

## 5.5. La medida *push-forward* y el cambio de variable

Si  $T$  es una función medible, entonces como se ve en el apéndice A,  $\mu \circ T^{-1}$  es una medida. De hecho, si en  $T$  transporta  $\mu$  en  $\nu$  por definición tenemos que  $\nu = \mu \circ T^{-1}$ . Utilizando la conclusión de ese Teorema encontramos que el argumento que define el ínfimo en el supremo del problema de Monge se puede escribir:

$$\int_X c(x, Tx) d\mu(x) = \int_Y c(T^{-1}y, y) d\nu(y)$$

Este resultado nos muestra cierta simetría del problema, sin embargo como ya vimos en los ejemplos anteriores, la simetría no es total. Pero nos ayuda a identificar ideas como la que tuvo Kantorovich. Kantorovich pensó en que esta simetría se podía explotar, y por esto en el problema de Kantorovich no se observa diferencia entre plantear el problema para  $(\mu, \nu)$  o para  $(\nu, \mu)$ .

Intuitivamente se debe a que no nos costaría lo mismo llevar todas las unidades de  $X$  a  $Y$  que llevarlas de  $Y$  a  $X$  pues, no siempre tendríamos los mecanismos suficientes para regresar cada unidad. Esto nos obliga a preguntarnos cuál es la relación entre ambos problemas.

## 5.6. Comentarios sobre el capítulo

También hay ciertos resultados para el problema de Monge pero en general no son tan concluyentes como los del problema de Kantorovich. Algunos textos indican que imponer que las medidas no sean puramente atómicas permite avanzar en el desarrollo de la teoría del problema de Monge.

# Determinación de cópulas y la solución de Kantorovich para medición de riesgo

## 6.1. Medición y administración del Riesgo

Uno de los grandes intereses de la ciencia actuarial es poder medir de manera cuantitativa el riesgo. Analizar los posibles resultados de alguna inversión o portafolio para poder decidir si se incurre o no en dicho riesgo. El objetivo de la medición cuantitativa de riesgos es proveer, a quien tomará la decisión, de herramientas útiles que faciliten el entendimiento del comportamiento de determinada cartera. El objetivo de esta sección es generar herramientas que ayuden en la toma de decisiones sobre algunos riesgos. La estadística inferencial juega un papel fundamental en la construcción de medidas de riesgo debido a que la experiencia propia de los datos informa en alto grado frecuencia y severidad de ocurrencias.

### 6.1.1. Fenómenos de Colas Pesadas

Un caso de particular interés para la ciencia actuarial, son aquellos fenómenos que concentran probabilidad en la cola de sus densidades, es decir pueden ocurrir eventos muy extremos. Los incendios, terremotos y otros desastres naturales son ejemplos de fenómenos de colas pesadas, porque existen casos donde la severidad es increíblemente alta. Estos fenómenos son la base de los seguros y los reaseguros. Dado el interés propio que generan este tipo de fenómenos, cada vez que se presenta una medida de riesgo, se debe hacer la diferencia entre su implementación en fenómenos “normales” y su implementación en colas pesadas. Inclusive, hay medidas de riesgo que parecen funcionar bien pero no son de ninguna ayuda en el estudio de colas pesadas.

## 6.2. Las cópulas en medición de riesgo

Supongamos que tenemos un portafolio con dos tipos de riesgos con distribuciones distintas  $R_1 \stackrel{d}{=} F_1$  y  $R_2 \stackrel{d}{=} F_2$ . Queremos analizar el riesgo total de este portafolio. Teóricamente deberíamos considerar todos los posibles escenarios, sin embargo esto resulta imposible. Por lo mismo, en la práctica, se utilizan medidas de riesgo que consideran promedios. Queremos analizar la pérdida esperada que puede generar nuestro portafolio.

No es lógico suponer que los dos riesgos son independientes, pues en general no lo son. Por ejemplo, si estamos asegurando niños y adultos ante enfermedades graves, una epidemia afectaría a ambos grupos de manera

similar. Por tanto, es necesario tener medidas que identifiquen la relación entre ambos tipos de riesgos. Las cópulas, como mencionamos en el segundo capítulo, son funciones que nos permiten modelar la interdependencia entre variables aleatorias.

Supongamos que para nuestro problema, de alguna manera (a detallar en las siguientes secciones) determinamos una cópula apropiada. ¿Cómo podríamos analizar el riesgo del portafolio a partir de esta cópula? Podemos encontrar la realización promedio de los riesgos a partir de la cópula, es decir considerar

$$\mathbf{E}[c(R_1, R_2)].$$

Este indicador nos daría una estimación de lo que esperaríamos que sucediera en términos de severidad. Este estimador sería bueno siempre y cuando la elección de la cópula también lo fuera. En el enfoque de costos se permite evaluar a la función  $c$  en puntos que no están en  $[0, 1] \times [0, 1]$ .

### 6.3. Cópulas de colas pesadas y supervivencia

Distintos tipos de cópulas llevan a distintos modelos. Se debe conocer bien el tipo de datos que se busca modelar; como se mencionó anteriormente, en muchos problemas actuariales se necesita la idea de colas pesadas. En cópulas, esta idea se replica al considerar cópulas que concentren probabilidad en  $(0, 0)$  o en  $(1, 1)$ .

Aquellas cópulas que concentran probabilidad cerca de  $(1, 1)$  se les denomina cópulas de cola pesada derecha y aquellas que lo hacen en  $(0, 0)$  se denominan cópulas de cola pesada izquierda.

Un ejemplo de una cópula de cola pesada derecha es

$$C(u, v) = u + v - 1 + [(1 - u)^{1/\theta} + (1 - v)^{1/\theta} - 1]^{-\theta}$$

para un valor  $\theta > 0$ .

En la ciencia actuarial es común trabajar con funciones de supervivencia en vez de funciones de distribución. Recordemos  $S(t) = \mathbb{P}(T > t) = 1 - F(t)$ . Del mismo modo, se puede trabajar con cópulas en términos de supervivencia.

Pensemos que tenemos dos variables aleatorias dependientes  $T_1$  y  $T_2$  y que nos interesa calcular la supervivencia conjunta  $\mathbb{P}(T_1 > t_1, T_2 > t_2)$ , entonces usando el Teorema de Sklar (Teo. Sklar) podemos escribir:

$$S(t_1, t_2) = 1 - F_1(t_1) - F_2(t_2) + F(t_1, t_2) = S(t_1) + S(t_2) - 1 + \bar{C}(1 - S(t_1), 1 - S(t_2)).$$

Así, para cualquier cópula podemos escribir la cópula de supervivencia asociada como:

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) = P(U_1 > u_1, U_2 > u_2).$$

### 6.4. Uso de la solución de Kantorovich para estimación de riesgo

En el capítulo 2 concluimos la desigualdad,

$$\int_0^1 c(F_1^{-1}(u), F_2^{-1}(u)) du \leq \mathbf{E}[c(R_1, R_2)] \leq \int_0^1 c(F_1^{-1}(u), F_2^{-1}(1 - u)) du.$$

Aprovecharemos esta conclusión para generar medidas de riesgo útiles. Podemos entonces, encontrar un intervalo para la severidad esperada usando ambos lados de esta desigualdad. Mientras más grande sea este intervalo, nos dará indicios de mayor variabilidad y mientras más pequeño sea, de menor variabilidad.

Este intervalo no plantea sustituir ninguna de las otras medidas de riesgo, sino aportar información distinta para quien tome la decisión.

En general las integrales de cópulas pueden no tener forma cerrada, sin embargo este no es ningún problema porque siempre podemos usar el método de Monte Carlo para obtener los valores numéricos.



### 6.4.1. Ejemplo 1

**Uniforme-Uniforme-Clayton** Supongamos que tenemos evidencia de que las variables aleatorias son uniformes y las ajustamos con una cópula de Clayton. Un ejemplo en ciencia actuarial es el supuesto “DUM”-Distribución Uniforme de Muertes en períodos de un año. Sabemos que muertes de hombres y mujeres no son independientes por lo que tiene sentido pensar en una cópula que las relacione.

Entonces  $F_1^{-1}(u) = u$  y también  $F_2^{-1}(u) = u$  y por lo mismo, podemos calcular el lado izquierdo de la desigualdad

$$\int_0^1 c(F_1^{-1}(u), F_2^{-1}(u)) du = \int_0^1 \{u^{-\theta} + u^{-\theta}\}^{-\frac{1}{\theta}} du = 2^{-\frac{1}{\theta}} \int_0^1 u du = 2^{-1-\frac{1}{\theta}} = \frac{1}{2^{1+\frac{1}{\theta}}}.$$

Esto muestra la utilidad del resultado, porque el cálculo de la cota inferior fue bastante rápido y nos da una herramienta útil para la toma de decisiones. Así, **encontramos un valor mínimo para el costo esperado de manera eficiente.**

### 6.4.2. Ejemplo 2

**Pareto-Pareto-Independiente** Pensemos que  $F_1(x) = 1 - \left(\frac{\theta_1}{x}\right)^{\alpha_1}$ ,  $x > \theta_1$  y  $F_2(x) = 1 - \left(\frac{\theta_2}{x}\right)^{\alpha_2}$   $x > \theta_2$  y podemos encontrar la cota inferior de manera rápida:

$$\int_0^1 c(F_1^{-1}(u), F_2^{-1}(u)) du = \theta_1 \theta_2 \int_{\max\{\theta_1, \theta_2\}}^1 u^{-\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right)} du = \theta_1 \theta_2 \left(1 - (\max\{\theta_1, \theta_2\})^{1-\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right)}\right).$$

Lo cual es una estimación rápida para el mínimo costo esperado y no fue difícil de calcular. Además, nos permite ver que si  $\theta_1$  es muy grande y  $\alpha_1$  también con todo lo demás constante, entonces este valor será muy grande. De manera contraria, si estos valores son pequeños, ésta cota inferior es pequeña. Por lo general, cuando se aplican estos modelos, se obtienen estimaciones de los valores por lo que se sabe el valor de esta cota. Si este valor supera lo que el capital permite perder con la agravación, es un claro indicador para no tomar riesgo. Este ejemplo logra **ilustrar que las cotas que mencionamos nos dan información del riesgo en términos de los parámetros.**

### 6.4.3. Ejemplo 3

**Cauchy-Cauchy-Clayton** (Numérico)

El objetivo de este ejemplo es mostrar que inclusive cuando no se integra de manera analítica, **con herramientas numéricas se puede encontrar información útil para la toma de decisiones.** Ahora consideremos distribuciones Cauchy:

$$F_1^{-1}(x) = \theta_1 \tan \left[ \left( \pi \left( y - \frac{1}{2} \right) \right) \right], F_2^{-1}(x) = \theta_2 \tan \left[ \left( \pi \left( y - \frac{1}{2} \right) \right) \right]$$

Consecuentemente si consideramos la cópula de Clayton encontramos que necesitamos resolver:

$$2 \int_0^1 \theta_1 \theta_2 \tan^2 \left[ \left( \pi \left( u - \frac{1}{2} \right) \right) \right] du.$$

$$\int_0^1 \theta_1 \theta_2 \tan \left[ \left( \pi \left( u - \frac{1}{2} \right) \right) \right] \tan \left[ \left( \pi \left( (1-u) - \frac{1}{2} \right) \right) \right] du.$$

que se pueden resolver de manera analítica pero podemos hacer estimaciones numéricas sencillas con el método de Monte Carlo (100,000) simulaciones y obtenemos el intervalo

```
>> b

b =

1.0e+07 *

0.092567182250000    3.372161837730088
```

Pero este resultado es el de las integrar sin considerar los valores de  $\theta_1$  y  $\theta_2$  . Entonces el intervalo sería

$$[2\theta_1\theta_2 925671,822, \theta_1\theta_2 33721618,377]$$

Pensemos, para fijar ideas, que los resultados que obtuvimos del ajuste de las Cauchy estiman  $\theta_1\theta_2 = 1$  . Entonces tendríamos que la pérdida esperada estaría en el intervalo aproximado,

$$[1, 851, 343,644, 33, 721, 618,377].$$

y sería labor del agente que tomará la decisión del riesgo si es solvente o no, incurrir en estos costos. No se puede decir mucho sobre la longitud del intervalo porque si se está trabajando en miles de millones, podría verse muy pequeño. Por lo mismo hemos dotado al agente que toma la decisión de otra herramienta para evaluar si conviene o no incurrir en determinado costo.

El objetivo de esta sección fue mostrar que las desigualdades anteriores sirven para **generar herramientas de decisión útiles** que se pueden obtener **con métodos analíticos o métodos numéricos**.

## 6.5. Sobre la manera en la que se escoge la cópula

Una de las grandes críticas en el uso de cópulas es la elección de dicha cópula. Ya que cada cópula induce un modelo analítico de dependencia entre variables y, como es bien sabido en estadística esto es casi imposible de perfeccionar.

Podemos usar tres enfoques en esta sección:

- Enfoque paramétrico (basado en modelos)
- Enfoque no paramétrico (basado en datos)
- Enfoque mixto

### 6.5.1. Enfoque paramétrico

El objetivo es encontrar la mejor cópula a partir de un conjunto finito de modelos. Este conjunto está formado por las cópulas que conocemos y puede expandirse a medida que se definan más cópulas. Cada cópula puede estar sujeta a parámetros. Usamos estadística inferencial para determinar cada uno de estos parámetros. Y finalmente consideramos el que opinemos sea mejor ajuste según los mejores parámetros de cada familia.

### 6.5.2. Familias paramétricas de cópulas

Las familias que normalmente se usan para hacer ajustes de cópulas son:

- Familia Gaussiana

Sea  $\Sigma$  una matriz de correlación y  $\Phi_{\Sigma}$  la distribución normal multivariada con matriz de correlación  $\Sigma$ , entonces

$$C_{\Sigma}(u, v) = \Phi_{\Sigma}(\Phi^{-1}(u), \Phi^{-1}(v)),$$

se denomina la cópula Gaussiana con matriz  $\Sigma$ .

- Familia Clayton

Para cualquier valor de  $\theta \in [-1, \infty) \setminus \{0\}$  la cópula de Clayton es

$$C_{\theta} = (\max\{u^{-\theta} + v^{-\theta} - 1, 0\})^{-\frac{1}{\theta}}.$$

Notamos que si  $\theta = -1$  obtenemos la cópula de Fréchet.

- Familia t (t-Cópula)

La  $t$  cópula se define de manera análoga a la cópula Gaussiana pero con la función de distribución de una variable aleatoria  $t$  de Student.

### 6.5.3. Cópulas arquimedianas

Si  $\psi : [0, 1] \rightarrow [0, \infty)$  es una función estrictamente decreciente, convexa, continua y además invertible, entonces la cópula arquimedea inducida por  $\psi$  es la función  $C$  que satisface la ecuación funcional

$$\psi(C(u, v)) = \psi(u) + \psi(v).$$

Equivalentemente, como asumimos invertibilidad podemos despejar a  $C$ . Sin embargo, en general no se asume invertibilidad y se hace uso de la pseudo-inversa de  $\psi$  o inversa generalizada de  $\psi$ .

Varias de las cópulas que definimos anteriormente son cópulas arquimedianas. Por ejemplo, la gaussiana corresponde a  $\psi = \phi^{-1}$  y la cópula de Clayton corresponde a  $\psi(t) = \frac{1}{\theta} \left( \frac{1}{t^{\theta}} - 1 \right)$ .

De este mismo modo, se pueden definir familias paramétricas de cópulas sólo determinando  $\psi$ :

- Familia de Frank

Es la familia de cópulas arquimedianas con  $\psi = -\log \left( \frac{-\exp(-\theta t) - 1}{-\exp(-\theta) - 1} \right)$ .

- Familia de Gumbel

Se define como las cópulas arquimedianas resultantes de  $\psi = (-\log(t))^{\theta}$ .

### 6.5.4. Máxima Verosimilitud

Notemos que la mayoría de las cópulas que definimos dependen de un parámetro  $\theta$ , uno de los procedimientos para la elección de cópulas es usar máxima verosimilitud.

En el enfoque paramétrico se supone la existencia de un modelo que ajusta perfectamente nuestros datos para un parámetro  $\theta$  desconocido.

Supongamos que hemos definido la familia paramétrica de cópulas que vamos a usar para ajustar a nuestros datos, entonces escogemos el parámetro de la cópula usando la técnica de máxima verosimilitud.

$$\hat{\theta} = \arg \max L(\theta; \underline{x}, \underline{y}) = \arg \max \prod_{i=1}^n C((X_i, Y_i)|\theta).$$

donde cada  $(X_i, Y_i)$  es un elemento de la muestra y  $C$  es la cópula que se escogió.

En general, no se pueden encontrar fórmulas analíticas para el valor de  $\hat{\theta}$  por lo que se deben resolver con métodos numéricos.

### Necesidad de métodos numéricos en la cópula de Clayton

Hagamos un ejemplo que muestre **la dificultad de encontrar analíticamente el valor de  $\hat{\theta}$** . Pensemos que queremos ajustar una cópula de Clayton a una muestra  $(u_i, v_i)$ , con  $i = 1, \dots, n$  entonces necesitamos encontrar el argumento máximo de

$$L(\theta|\underline{X}) = \prod_{i=1}^n (u_i^{-\theta} + v_i^{-\theta} - 1)^{-\frac{1}{\theta}} \cdot \chi_{\{u_i^{-\theta} + v_i^{-\theta} + 1 > 0\}}.$$

Del conjunto de indicadores se obtiene  $\min\{u_i^{-\theta} + v_i^{-\theta} - 1\} > 0$ , inclusive desde esta condición ya es evidente la necesidad de los métodos numéricos. Si además se intenta derivar (o derivar la log-verosimilitud) la necesidad se hace más clara; pues se debe resolver:

$$\frac{1}{\theta^2} \left[ \sum_{i=1}^n \log(u_i^{-\theta} + v_i^{-\theta}) \right] - \frac{1}{\theta} \left[ \sum_{i=1}^n \frac{-1}{u_i^{-\theta} + v_i^{-\theta}} (u_i^{-\theta} \log(u_i) + v_i^{-\theta} \log(v_i)) \right] = 0,$$

cuya solución será numérica.

### Ejemplo computacional del enfoque paramétrico

En esta sección se hace un ejemplo de un ajuste de cópula por el método de máxima verosimilitud y **se muestra en que casos los paquetes estadísticos pueden fallar**. En Matlab se pueden ajustar cópulas paramétricas con el comando `copulafit` que hace estimaciones numéricas de los valores de máxima verosimilitud. Procedemos a revisar la precisión de estas estimaciones. Para lo mismo, simulamos una variable aleatoria bivariada normal, con correlaciones  $\rho = -0,75, -0,25, 0,25, 0,75$ . Luego consideramos su función de distribución, la graficamos y posteriormente usamos `copulafit` con la opción `Gaussian` para estimar  $\rho$  obteniendo las siguientes estimaciones:

$\rho$	$\hat{\rho}$
0.00001	-0.0490
0.0001	0.0024
0.001	0.0374
0.01	0.0035
0.75	0.7424
0.25	0.2306
-0.25	-0.218
-0.75	-0.7543

De estas observaciones notamos a primera vista que los parámetros estimados son buenos siempre que la correlación entre las variables sea alta. Esto es una observación que debe hacerse cuando se usan cópulas en finanzas, debido a que, en general, las variables aleatorias relacionadas a precios de mercado suelen tener correlaciones bajas. Eso debería ser un desincentivo para el uso de cópulas.

El comando `copulafit` tiene buenos rendimientos para valores de correlación altos. Esto indica que si se sospecha que ambas variables aleatorias no están muy correlacionadas no conviene usar este comando para aproximar la cópula. Además, debido a la aritmética finita, si la correlación entre las variables es tan

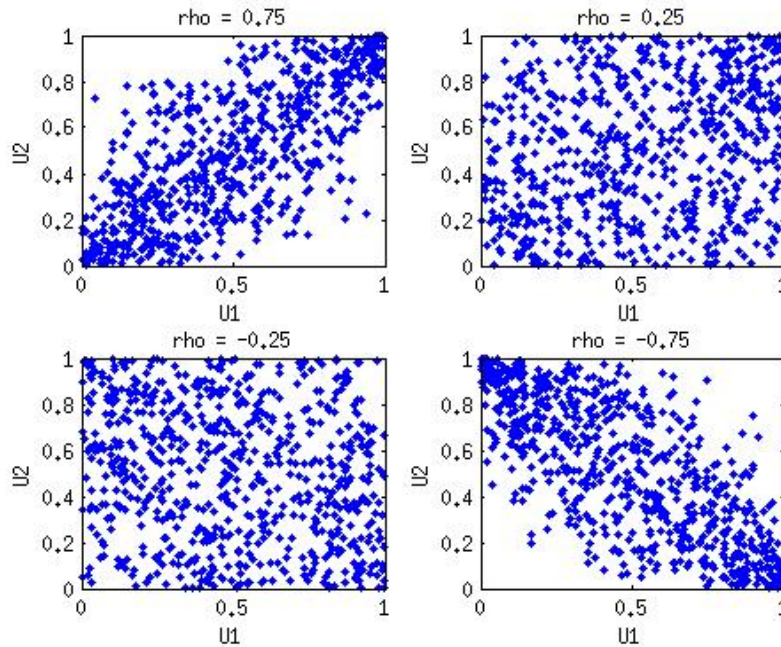


Figura 6.1: Gráfica de las cópulas simuladas

pequeña es mejor usar la cópula de independencia  $C(F_1(x), F_2(y)) = F_1(x)F_2(y)$ .

Los valores estimados  $\hat{\rho}$  que obtuvimos se determinaron posterior a definir que íbamos a ajustar cópulas gaussianas. Esto lo hicimos porque sabíamos de dónde provenían los datos. Sin embargo, en general esta no es información conocida por lo que no es tan fácil determinar que familia paramétrica usar.

De este sencillo ejemplo **se concluye** que cuando la correlación entre variables no es muy alta, **los paquetes estadísticos pueden fallar**.

### 6.5.5. Enfoque no paramétrico

Este enfoque se basa exclusivamente en los datos. Así como se puede utilizar la distribución empírica, usamos la cópula empírica de Deheuvels.

Consideremos una muestra aleatoria de nuestro portafolio  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ . Definimos las funciones de distribución empíricas como:

$$\hat{F}_1(x) = \frac{1}{n} \sum_{i=1}^n \chi_{\{X_i \leq x\}}(x) \text{ y análogamente } \hat{F}_2(y) = \frac{1}{n} \sum_{i=1}^n \chi_{\{Y_i \leq y\}}(y).$$

Usamos las definiciones de distribución empírica para formar la cópula empírica de Deheuvels :

$$\hat{C}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n \chi_{\{\hat{F}_1(X_i) \leq t_1\}}(t_1) \cdot \chi_{\{\hat{F}_2(Y_i) \leq t_2\}}(t_2).$$

**Observación 6.5.1.** *Sobre el denominador.*

*En muchas ocasiones se suele escribir  $n+1$  en el denominador en vez de  $n$ , esto depende de lo que se quiera modelar ya que  $n+1$  no permite que el valor de 1 se alcance en la muestra; idea importante para el desarrollo de la teoría de colas pesadas.*

**Observación 6.5.2.** *Sobre la definición de la cópula empírica.*

Consideremos la muestra ordenada  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \Rightarrow \hat{F}_1(X_{(i)}) = \frac{i}{n}$

entonces si  $\frac{i}{n} \leq t < \frac{i+1}{n}$  podemos calcular la suma

$$\frac{1}{n} \sum_{i=1}^n \chi_{\{\hat{F}_1(X_{(i)}) \leq t\}} = \frac{i}{n}$$

pues hay  $i$  elementos de la muestra cuya distribución empírica está por debajo de  $t$ . Esto nos permite reescribir la cópula:

$$C\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\# \text{ de pares en la muestra tales que } x \leq X_{(i)} \text{ y } y \leq Y_{(j)}}{n}.$$

Este análisis nos hace mucho más clara la definición de la cópula empírica y nos permite hacer pruebas de manera mucho más sencilla.

**Proposición 6.5.3.** *La cópula empírica goza de algunas propiedades de cópula.*

*Demostración.* Los hechos  $C(1, 1) = 1$  y  $C(0, t) = 0 = C(t, 0)$  se concluyen directamente de la definición.

Para probar la 2-monotonía, así como las marginales en cada coordenada, es más sencillo usar la conclusión de la observación.

Al calcular  $C(1, j/n)$  encontramos que es el número de pares en la muestra tales que  $\hat{F}_1(x) \leq 1$  y  $y \leq Y_{(j)}$  que coincide con la definición de la distribución empírica únicamente para  $Y$ . La conclusión en la otra coordenada es idéntica.

Para concluir la 2-monotonía, sean  $a \leq b$  y  $c \leq d$  números en el intervalo  $[0, 1]$  entonces  $\hat{C}(b, d)$  es el número de pares de la muestra con primera coordenada menor que  $b$  y segunda coordenada menor que  $d$  dividido por  $n$ , queremos concluir

$$\hat{C}(b, c) + \hat{C}(a, d) \leq \hat{C}(b, d) + \hat{C}(a, c).$$

Pero si hay un par en la muestra que cuya coordenada en  $x$  sea menor que  $b$  y cuya coordenada en  $y$  sea menor que  $c$ , entonces en particular ese punto tiene coordenada en  $x$  menor que  $b$  y en  $y$  menor que  $d$ . Lo mismo para el segundo término. Además si un par aparece en el cálculo de la primera y la segunda evaluación, entonces ese mismo punto aparece en ambas evaluaciones del lado derecho dado que se tendría

$$(x \leq b \wedge x \leq a) \wedge (y \leq c \wedge y \leq d) \Rightarrow x \leq a \wedge y \leq c,$$

y por lo mismo se deduce la desigualdad deseada; pasando el lado izquierdo restando se tendría que  $\hat{C}$  es 2-monótona.

**Observación 6.5.4.** *En estricto sentido la cópula empírica no es una cópula.*

Esto se debe a que  $C(1, u) \neq u$ , al ser suma ponderada de indicatoras (como  $n$  es entero) hay muchos valores que no se pueden alcanzar. Por ejemplo,  $C\left(1, \frac{\pi}{6}\right) \neq \frac{\pi}{6}$  porque esta suma no puede alcanzar valores irracionales.

En la literatura se suele sustituír la cópula empírica por la “cópula muestral” cuya idea es generar un rejado y definir uniformes continuas en cada celda; sin embargo, para nuestros fines de selección, la cópula empírica es una gran herramienta. Una de las propiedades fundamentales de la cópula empírica es el teorema de Glivenko-Cantelli que asegura que a medida de que la muestra crece, la cópula empírica converge casi seguramente a la cópula.

### 6.5.6. Uso de la cópula no paramétrica

Algo que se debe destacar de la cópula no paramétrica es que su implementación es extremadamente sencilla. Sólo se debe hacer un contador para los casos en los que no se hayan rebasado los valores en los que se quiere calcular la cópula.

#### Ejemplo del enfoque no paramétrico

Usamos la misma técnica del ejemplo de la sección anterior y programamos la cópula empírica. Obtenemos la figura 6.2:

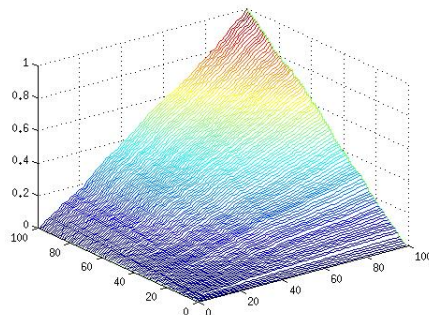


Figura 6.2: Cópula empírica de los datos simulados

## 6.6. El enfoque mixto

Este enfoque plantea no sólo usar el enfoque paramétrico, sino también el no paramétrico. Plantea usar la cópula no paramétrica para encontrar la mejor cópula paramétrica. Es decir, usar los valores de la cópula  $\hat{C}$  para encontrar la mejor estimación  $C_\theta$  en términos de minimizar el error  $\|\hat{C} - C_\theta\|_\infty$ . Este problema se puede plantear con cualquier norma en el espacio de funciones continuas en  $[0, 1]$  o sólomente minimizar el error de aproximación en los puntos de la muestra.

Para el enfoque mixto podemos comparar el dibujo de la cópula paramétrica que hayamos ajustado con el dibujo de la cópula no paramétrica. Podemos usar máxima verosimilitud para obtener la mejor cópula de cada familia y comparar la gráfica de cada cópula con la gráfica de la cópula empírica. Si ambas gráficas son similares, se puede esperar que el ajuste sea bueno. Esto es una manera práctica para ajustar cópulas a datos.

## 6.7. Comentarios sobre la elección de cópulas

Para que la teoría de cópulas termine de ser aceptada en el mundo de la estadística es necesario que se hagan pruebas de bondad de ajustes para las cópulas. Porque, hasta el momento, la elección de la familia o de la técnica es libre y depende únicamente de la persona que quiere hacer el ajuste.

De esta manera escogemos la mejor cópula en algún sentido y con algún enfoque de los planteados. En general no existe una manera óptima de escoger cópulas, pero para fines prácticos los métodos planteados en estas secciones resultan útiles. Las cópulas son de gran aplicación en la ciencia actuarial así como en física y economía pues permiten identificar la relación entre variables aleatorias de manera analítica.

## 6.8. Reglas de decisión con la dualidad de Rockafellar

Una vez analizada la idea detrás de la solución al problema de Kantorovich, podemos dedicarnos a pensar en lo que haríamos en situaciones más generales. Por ejemplo, retomemos el ejemplo del riesgo de la sección 4.3. Si tenemos que encontrar la mejor regla de decisión, podríamos intentar aprovechar el desarrollo teórico para encontrarla. Recordemos que el principio de minimax que resolvimos y demostramos fue el del problema dual de Kantorovich y se hizo a partir del Teorema de Rockafellar.

En el teorema de Rockafellar se utilizó el Teorema de Hahn-Banach en su primera forma geométrica para separar conjuntos convexos. La funcional del hiperplano que general el teorema de Hahn-Banach nos devolvió el valor en el que se alcanza el supremo. Por lo mismo, es lógico pensar que a pesar de no poder encontrar el plano separador de Hahn-Banach, podríamos encontrar un aproximante con máquinas de soporte vectorial (o SMO o cualquier método de clasificación) y recuperar la regla de decisión a partir de ese aproximante. Sin embargo, como la condición en los conjuntos es continua, para términos de aplicaciones necesitaríamos dar  $n$  puntos de cada conjunto y separarlos con el mayor margen. Mientras más puntos demos de cada conjunto mejor será la aproximación al plano separador y por lo tanto mejor la regla de decisión.

**Data:** La función de pérdida

**Result:** Una regla de decisión

Empezar obteniendo  $n$  puntos de  $\text{epi}(\Theta)$  e  $\text{hipo}(\Xi) - a$  como en la formulación de Kantorovich, determinar de algún modo un aproximante de  $a$ ;

**while** la regla de decisión no sea buena **do**

    agregar más puntos a los conjuntos;

    separar ambos conjuntos con MSV;

    recuperar la regla de decisión a partir de la fórmula  $\frac{f}{k}$ ;

**end**

**Algorithm 1:** Primeras ideas para un algoritmo

Esto constituye las ideas fundamentales para aplicar la dualidad de Kantorovich **en su forma general** a los teoremas de decisión en estadística.

### 6.8.1. Complicaciones

- Encontrar el primer aproximante de  $a$  puede ser muy difícil en general.
- No hay manera de determinar cuantos puntos escoger en los conjuntos.
- Una mala elección de puntos (muy cercanos entre ellos y separados del otro conjunto) no llevará a soluciones.
- Recuperar la regla de decisión puede ser altamente no trivial.

**Observación 6.8.1.** *Este método no está acabado ni se ha probado. Sólo se plantean ideas muy generales, este método se seguirá analizando para ver su potencia, rendimiento y aplicabilidad. Sin embargo, este algoritmo pretende relacionar un resultado tan teórico como es Hahn-Banach en su forma geométrica con aplicaciones muy específicas y útiles en finanzas y ciencia actuarial.*

## 6.9. Comentarios sobre el capítulo

En este capítulo se usó el resultado del capítulo 2 para hacer estimaciones que proveen herramientas útiles para la medición cuantitativa de un riesgo.



Del mismo modo, se hizo una breve introducción a la teoría de selección de cópulas. Esta teoría de selección de cópulas tiene muchas variantes e incluso existe experiencia empírica. Por ejemplo, en finanzas, dependiendo de la correlación entre activos, se utilizan distintas familias de cópulas. En ese aspecto, este capítulo presentó tres enfoques. Todos ellos basados principalmente en datos, no en conocimientos previos. Juntos, estos 3 enfoques, presentan una manera organizada y clara de decidir cuál es la cópula que se debe ajustar. Este capítulo mostró que el problema de Kantorovich tiene aplicaciones que son de interés para muchas personas, además de ser una herramienta diferente y bastante útil cuando se necesitan estimaciones rápidas.

# El problema finito de Kantorovich

## 7.1. El problema del Transporte natural

En teoría de la medida ocurre que la abstracción es tal que no deja clara la aplicabilidad y utilidad de ciertos resultados. A pesar de haber usado algunos resultados en ciencia actuarial, podemos simplificar las cosas un poco más para obtener un problema muy conocido y ampliamente estudiado:

¿Qué pasa si en la formulación general **sustituimos**  $X = \mathbb{R}, S = \mathcal{P}(\mathbb{R})$  pero consideramos probabilidades con sólo un número finito de puntos?

Entonces, estamos considerando probabilidades con la forma  $\mu = \{\mu_i\}_{i=1}^n, \nu = \{\nu_j\}_{j=1}^m$  y por lo mismo

$$\inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c(x, y) d\pi \right\} = \inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \sum_{i=1}^n \sum_{j=1}^m c_{ij} \pi_{ij} \right\}$$

En este contexto  $\pi \in \Gamma(\mu, \nu) \Rightarrow \mu_i = \sum_{j=1}^m \pi_{ij}$  y  $\nu_j = \sum_{i=1}^n \pi_{ij} \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\}$

Cambiamos de notación a  $\pi_{ij}$  por  $X_{ij}$ , a  $\mu_i = O_i$  y  $\nu_j = D_j$  por lo que el problema de Kantorovich se puede reescribir como

$$\begin{aligned} & \text{mín} \sum_{i=1}^n \sum_{j=1}^m c_{ij} X_{ij} \\ & \text{s.a.} \sum_{i=1}^n X_{ij} = D_j \\ & \sum_{j=1}^m X_{ij} = O_i \\ & \sum_{i=1}^n \sum_{j=1}^m X_{ij} = 1 \\ & X_{ij} \geq 0. \end{aligned}$$

Este problema es bastante conocido en el área de investigación de operaciones. Si se considera el problema en términos de medidas y no sólo de probabilidades, la restricción de que la doble suma sea uno desaparecería. Este problema se llama el problema de transporte balanceado, y se puede interpretar como sigue: se va a

distribuir un producto, este producto sale de  $m$  lugares y debe llegar a  $n$  lugares.  $X_{ij}$  es la cantidad de producto que se transporta de  $i$  a  $j$  para minimizar costos.  $O_i$  es lo que se oferta a partir de la fuente  $i$  y  $D_j$  es lo que se demanda en el lugar  $j$ .

¿Qué nos dice el teorema de Dualidad de Kantorovich en este contexto?

$$\int \phi d\mu + \int \psi d\nu = \sum_{i=1}^n O_i \phi_i + \sum_{j=1}^m D_j \psi_j.$$

En este contexto  $(\phi, \psi) \in \Phi_c$  es equivalente a pedir  $\phi_i + \psi_j \leq c_{ij}$ . Por lo que el problema dual que obtuvimos coincide con el problema dual que se obtiene con técnicas de programación lineal:

$$\text{máx} \sum_{i=1}^n O_i \phi_i + \sum_{j=1}^m D_j \psi_j$$

$$\text{sujeto a } \phi_i + \psi_j \leq c_{ij}.$$

Así, dediquemos este capítulo al problema de transporte en los naturales. Veamos que el problema es en efecto factible. Sea  $O$  la oferta total, si el problema es balanceado entonces,  $O = D$ , si el problema no es balanceado se define un nuevo lugar de demanda con demanda  $D_{m+1} = O - D$  y con costo asociado cero.

## 7.2. Factibilidad del Problema de Transporte

Definimos  $X_{ij} = \frac{O_i D_j}{O} \Rightarrow \sum_{i=1}^n X_{ij} = D_j, \sum_{j=1}^m X_{ij} = O_i$  y por lo tanto  $X$  es factible.

El problema es acotado porque si  $Y$  es factible entonces  $0 \leq Y_{ij} \leq \min\{O_i, D_j\}$  por lo que la región factible es un poliedro acotado no vacío. Por lo mismo, el problema tiene solución. Pero esta idea ya la teníamos porque era parte de la conclusión de dualidad de Kantorovich.

## 7.3. Algoritmos para el problema del transporte

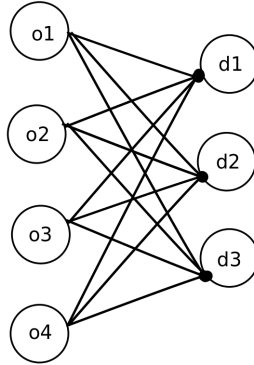
En general se usan dos algoritmos para el problema del transporte:

- Ford-Fulkerson (flujos)
- Simplex y la esquina noreste
- Puntos Interiores

Estos algoritmos se cubren en cursos de investigación de operaciones o de modelado, sin embargo por fines de completitud se presentan sin detalle en esta sección.

### 7.3.1. Ford-Fulkerson

A partir del problema de transporte se puede definir una gráfica bipartita, siendo los vértices de un conjunto las fuentes y los vértices del otro conjunto los lugares que demandan el producto.



#### Resumen del algoritmo

Se comienza con un flujo factible, en cada iteración habrá vértices etiquetados y no etiquetados. Inicialmente el vértice 1 tendrá la etiqueta  $E_1 = (-, \infty)$  indicando que ningún vértice precede al vértice 1 en una trayectoria incrementable y que en principio podemos enviar cualquier cantidad de flujo a partir de ese vértice. En general un vértice etiquetado puede tener etiquetas

$$E_i = (k^+, \delta_i) \text{ o } E_i = (k^-, \delta_i).$$

$\delta_i$  es la mayor cantidad que se puede enviar. En el primer caso se indica que tenemos una trayectoria incrementable que termina en  $i$ , el vértice anterior es  $k$  y el arco  $(i, k)$  es un arco de avance. En el segundo caso, la diferencia es que  $(i, k)$  es de retroceso y no de avance.

#### Algoritmo

**Data:** Una red de flujo con capacidades enteras

**Result:** Un flujo máximo

Empezar con un flujo inicial factible y  $E_1 = (-, \infty)$

**while** no todos los vértices etiquetados han sido explorados **do**

    encontrar un vértice etiquetado pero no explorado y explorarlo:

    Si  $j$  es un vecino exterior de  $i$  no etiquetado y  $f_{ij} < c_{ij}$  poner  $E_j = (i^+, \delta_j)$ ,  $\delta_j = \text{mín}\{\delta_i, c_{ij} - f_{ij}\}$  ;

    Si  $j$  es un vecino interior de  $i$  no etiquetado y  $f_{ij} > 0$  poner  $E_j = (i^-, \delta_j)$ ,  $\delta_j = \text{mín}\{\delta_i, f_{ji}\}$  ;

**end**

Encontrar la trayectoria  $f$ -incrementable y aumentar el flujo en  $\delta$  unidades

#### Algorithm 2: Ford-Fulkerson flujo en redes

Existen ejemplos en los que se muestra que el algoritmo de Ford-Fulkerson no es polinomial en todos los casos, sin embargo si se considera el primer vértice en ser explorado como el primer vértice etiquetado, el algoritmo es del orden  $\mathcal{O}(nm^2)$ .

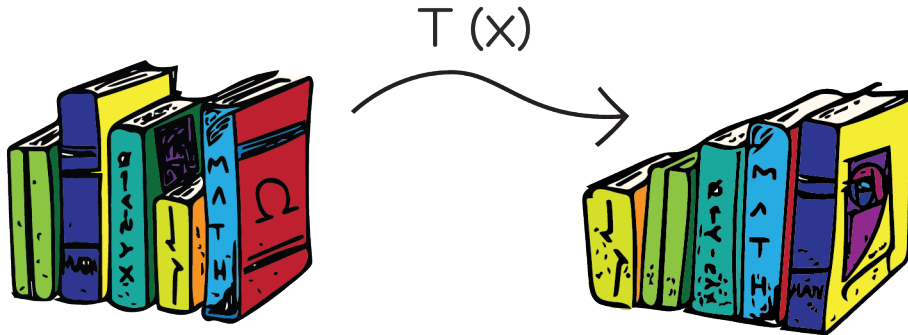
### 7.3.2. Simplex y la esquina noreste

El método de la esquina noreste es un algoritmo que se utiliza para hallar soluciones factibles iniciales. En la matriz de transporte  $\{c_{ij}\}$ , se comienza por la entrada  $(1, 1)$  (la esquina noreste) y se coloca como flujo el mínimo entre la demanda y la oferta. Posteriormente el algoritmo se mueve por fila o por columna dependiendo donde no se considere el mínimo.

Posteriormente se usa el método simplex para encontrar la solución. Es bien conocido que el simplex funciona bien para los ejemplos prácticos, pero hay ejemplos en los que el simplex tiene que visitar todos los vértices para encontrar la solución. Por lo que el algoritmo trabaja en tiempo exponencial. Estos casos son patológicos y no suelen ser encontrados en la práctica.

### 7.3.3. Caso Natural

En la sección anterior consideramos  $X = \mathbb{R}$ . En vez de pensar que la variable puede ser cualquier número real, consideremos  $X = \{1, 2, \dots, n\}$  para algún número natural  $n$ . Entonces, como la variable sólo puede tomar un número finito de valores, el problema se puede interpretar como mover libros. Es decir cada libro tiene asignado un número y un peso. Queremos minimizar el costo total de transporte, es decir, las maneras en las que movemos libros.



## 7.4. Conclusiones del capítulo

El problema del transporte finito es de suma importancia para el desarrollo de algoritmos computacionales e investigación de operaciones. Existen métodos basados en estadística que mejoran las implementaciones que aquí se presentaron. Los métodos se presentaron de manera breve porque son algoritmos que se detallan en cualquier curso de investigación de operaciones. No obstante, el repaso era fundamental para aterrizar ideas y comprender el impacto de los resultados generales.

## Conclusiones Generales

A través de este trabajo, hemos desarrollado todas las herramientas necesarias para entender los principios básicos del problema de transporte óptimo.

Observamos, que cuando nos dedicamos a resolver el caso real, las cópulas fueron la base para poder lograrlo. Vimos la importancia teórica de las cópulas y concluimos la fórmula gracias a ellas. Esta idea es muy interesante porque relacionamos las funciones de costos y las cópulas gracias a la condición de “2-monotonicidad”. Posteriormente aprendimos que si el problema proviene de espacios con estructuras topológicas ricas (léase espacios métricos completos y separables), entonces existe una solución al problema de Kantorovich y se puede encontrar a partir de su formulación dual.

Observamos que el problema formulado por Monge y el formulado por Kantorovich están altamente relacionados. El problema de Kantorovich es la relajación del problema de Monge. Vimos que si las medidas son puramente atómicas, pueden ocurrir diversas situaciones: podemos no tener soluciones, tener una infinidad de ellas o ni si quiera poder plantearlo.

En los dos capítulos consecuentes notamos que el problema de Kantorovich tiene aplicaciones importantes y los desarrollamos de manera breve.

# Bibliografía

- [1] Ambrosio, L. y Gigli, N. *A user's guide to optimal transport*, 2009, Italia.
- [2] Ambrosio, L. y Rigot, S. *Optimal mass transportation in the Heisenberg group*, Journal of Functional Analysis, 2002, Italia y Francia.
- [3] Ball, K. *An elementary introduction to monotone transportation*, London.
- [4] Bass, R. *Real Analysis for Graduate Students*, 2011, California.
- [5] Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, Springer, 1985, New York.
- [6] Billingsley, P. *Convergence of Probability Measures*, John Wiley and sons, 1999, New York.
- [7] Bogachev, V.I. *Measure Theory vol 1,2* , Springer, New York.
- [8] Boyd, S. y Vandenberghe, L. *Convex optimization*, Cambridge University Press, 2009, Glasgow.
- [9] Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011, New York.
- [10] Cohn, D.L. *Measure Theory*, Birkhäuser, 1980, Boston.
- [11] Edgar, G. *Measure, Topology, and Fractal Geometry*, Springer, 2005, New York.
- [12] Ekeland, I. y Temam, R. *Convex Analysis and variational problems*, North-Holland, 1976, New York.
- [13] Gangbo, W. y McCann, R.J. *The Geometry of optimal transportation*, 1996, Estados Unidos.
- [14] Grabinsky, G. *Teoría de la Medida*, Facultad de Ciencias UNAM, 2009, CD. México.
- [15] Halmos, P. *Measure Theory*, Springer, 1947, New York.
- [16] Kantorovich, L.K. y Akilov, G.P. *Functional Analysis*, Pergamon Press Wheaton and Co. Ltd., 1982, U.K.
- [17] Krugman, S. y Panjer, H. y Willmot, G. *Loss Models: From Data to Decisions*, John Wiley and sons, 2012, New Jersey.
- [18] Mikosch, T. *Copulas: Tales and Facts*, 2006 Dinamarca.
- [19] Nelsen, R.B. *An introduction to Copulas*, Springer, 2006, New York.

- [20] Osborne, M.J. y Rubinstein, A. *A course in Game Theory*, MIT Press, 1994, Massachusetts.
- [21] Parthasarathy, K.R. *Probability measures on Metric Spaces*, Academic Press, 1967, London.
- [22] Riesz, F. y Nagy, B.S. *Functional Analysis*, Dover, 1990, New York.
- [23] Rockafellar, T. *Convex Analysis*, Princeton, 1970, New Jersey.
- [24] Rudin, W. *Real and Complex Analysis*, McGraw-Hill, 1970, México.
- [25] Savage, L.M. *Foundations of Statistics*, Dover, 1972, New York.
- [26] Trivedi, P.K. y Zimmer, D.M. *Copula Modeling: An introduction for Practitioners*, NOW, 2007, Neww York.
- [27] Trudinger, N.S. y Wang, X-J. *On the Monge transfer problem*, Centre for Mathematics and Its Applications, Australian National University, 2000.
- [28] Villani, C. *Topics in Optimal Transportation*, American Mathematical Society, 2003, New York.
- [29] Wilard, S. *General Topology*, Dover, 2004, New York.



## Notación

- $(X, d), (Y, \rho)$  espacios métricos.
- $(E^*, \|\cdot\|_*)$  - El espacio dual topológico de un espacio normado  $(E, \|\cdot\|)$ .
- $\mathcal{P}(X)$  - Conjunto potencia de  $X$ .
- $\Lambda$  una funcional lineal.
- $\mathbf{Z} \stackrel{d}{=} F$  -  $\mathbf{Z}$  es variable aleatoria con función de distribución  $F$ .
- $\text{ran}(\phi)$  - Rango de  $\phi$ .
- $\chi_A$  - función indicadora del conjunto  $A$ .
- $M(X, S)$  - El conjunto de medidas definidas en el espacio medible  $(X, S)$ . En caso de obviar  $S$  se considera la  $\sigma$ -álgebra de Borel de  $X$ .
- $M_1(X, S)$  - El conjunto de probabilidades en el espacio medible  $(X, S)$ . En caso de obviar  $S$  se considera la  $\sigma$ -álgebra de Borel de  $X$ .
- $T_{\#}\mu = \nu$  -  $T$  es una función de transporte de  $\mu$  en  $\nu$ .
- $\Gamma(\mu, \nu)$  - El conjunto de probabilidades en el espacio producto con marginales  $\mu$  y  $\nu$  respectivamente.
- $\pi_x(\gamma)$  - La proyección de  $\gamma$  en la coordenada  $x$ .
- $A^\epsilon$  - El conjunto de puntos  $\epsilon$ -ceranos al conjunto  $A$ .
- $A^\circ$  - El interior del conjunto  $A$ .
- $\bar{A}$  - La cerradura del conjunto  $A$ .
- $S_\mu$  - El soporte de  $\mu$  - El complemento del mayor conjunto abierto y  $\mu$ -nulo.
- $d(x, A)$  - La distancia de un punto al conjunto  $A$  definido como el ínfimo de las distancias a puntos del conjunto.
- $\mathcal{L}$  - Funciones continuas con soporte compacto definidas en un espacio  $X$ .
- $\mathcal{L}_+$  - Funciones continuas con soporte compacto y no negativas definidas en un espacio  $X$ .
- $\mu_n \Rightarrow \mu$  - La sucesión de medidas  $\{\mu_n\}$  converge débilmente a la medida  $\mu$ .
- $X_A^\infty$  - Función indicadora infinita del conjunto  $A$ .
- $f$  es  $\|\cdot\|$  - Lipschitz se refiere a que  $|f(x) - f(y)| \leq C\|x - y\|$  para alguna constante  $C$ .

# Apéndices

## El cambio de variable en medidas

**Teorema.** (Del cambio de variable en medidas)

Sean  $(X, S_X, \mu), (Y, S_Y, \nu)$  dos espacios de medida. Sea  $f : X \rightarrow Y$  una función  $(S_X, S_Y)$ -medible. Entonces  $\mu \circ f^{-1}$  es una medida en  $Y$  y una función  $g : Y \rightarrow \mathbb{R}$  es integrable respecto a la medida  $\mu \circ f^{-1}$  cuando la función  $g \circ f$  es integrable respecto a  $\mu$ . Además, se obtiene la siguiente fórmula

$$\int_Y g(y) d\mu \circ f^{-1}(y) = \int_X g(f(x)) d\mu(x). \quad (\text{A.1})$$

*Demostración.* Empezamos por probar que  $\mu \circ f^{-1}$  es una medida. Esto se debe a “lo bondadosa” que es la imagen inversa de una función en cuanto a operaciones de conjuntos.

- $\mu \circ f^{-1}(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$ .
- Si  $E \subseteq F \Rightarrow f^{-1}(E) \subseteq f^{-1}(F) \Rightarrow \mu \circ f^{-1}(E) \leq \mu \circ f^{-1}(F)$ .
- Sea  $(E_k)_{k=1}^{\infty}$  una sucesión disjunta de elementos en  $S_Y$ , entonces no existe  $y \in E_i \cap E_j$  por lo que para todo  $x$  se tiene que  $f(x) \notin E_i \cap E_j$  y necesariamente  $f^{-1}(E_i) \cap f^{-1}(E_j) = \emptyset$   
Por  $\sigma$ -aditividad de  $\mu$  se obtiene la  $\sigma$ -aditividad de  $\mu \circ f^{-1}$  dado que

$$\mu \circ f^{-1} \left( \bigcup_{i=1}^{\infty} E_i \right) = \mu \left( \bigcup_{i=1}^{\infty} f^{-1}(E_i) \right) = \sum_{i=1}^{\infty} \mu(f^{-1}(E_i)) = \sum_{i=1}^{\infty} \mu \circ f^{-1}(E_i).$$

La segunda parte del teorema procede, como es usual en teoría de la medida, por aproximaciones. Se verifica la fórmula para funciones indicadoras, esto la prueba para funciones  $S_Y$ -simples y posteriormente se usa el teorema básico de aproximación y el teorema de convergencia dominada de Lebesgue para concluir la propiedad para cualquier función medible  $g$ .

Procedemos de esta manera, verificamos la fórmula para funciones indicadoras: para una función indicadora  $\chi_A$  el lado izquierdo de (1.1) es  $\mu \circ f^{-1}(A)$  y en el lado derecho de (1.1) la composición nos deja en  $\chi_{f(x) \in A}(x)$  por lo que al evaluar la integral tenemos  $\mu(\{x : f(x) \in A\}) = \mu(f^{-1}(A)) = \mu \circ f^{-1}(A)$  por lo que ambos lados son iguales. Se generaliza a funciones simples de manera directa. Por el teorema básico de aproximación se tiene el resultado para  $g_n = \min\{n, \max\{g, -n\}\}$ .

Notamos que  $g_n \rightarrow g$  y además  $|g_n| \leq g$ . Como  $g$  es integrable respecto a la medida  $\mu \circ f$  por hipótesis, entonces se satisfacen todas las hipótesis del Teorema de la convergencia dominada de Lebesgue. Se concluye el resultado para  $g$ .

## Axiomas de Separación

Una topología  $\tau$  en  $X$  es una familia de subconjuntos de  $X$  que contiene a  $X$ , al vacío, uniones arbitrarias de elementos de la topología y todas las intersecciones finitas de elementos de la topología.

Sin embargo, estas condiciones pueden generar conjuntos inútiles en los que la topología no permite discernir puntos ni conjuntos. Por lo mismo existen varios axiomas de separación para asegurar cierta estructura a los espacios:

- $T_0$ : Un espacio topológico se denomina  $T_0$  si siempre que  $x$  y  $y$  son puntos distintos en  $X$ , existe un conjunto abierto que contiene a uno de los puntos y no al otro.
- $T_1$ : Un espacio topológico se denomina  $T_1$  si siempre que  $x$  y  $y$  son dos puntos distintos en  $X$  existe una vecindad de cada uno que no contiene al otro.
- $T_2$ : Un espacio topológico se denomina Hausdorff o  $T_2$  si siempre que  $x$  y  $y$  son dos puntos distintos en  $X$  existen conjuntos abiertos disjuntos  $U$  y  $V$  tales que  $x \in U$  y  $y \in V$ .
- $T_3$ : Un espacio topológico se llama regular si siempre que  $A$  sea un conjunto cerrado en  $X$  y  $x \in X, x \notin A$  entonces existen conjuntos abiertos disjuntos  $U, V$  tales que  $x \in U, A \subseteq V$ .  
Un espacio topológico se denomina  $T_3$  si es un espacio topológico  $T_1$ -regular.
- $T_{3\frac{1}{2}}$ : Un espacio topológico se denomina Tychonoff si siempre que un punto  $x$  no pertenezca a un cerrado  $F$  existe una función continua  $f : X \rightarrow [0, 1]$  tal que  $f(x) = 0, f(a) = 1 \forall a \in F$ .
- $T_4$ : Un espacio se llama normal si siempre que  $A, B$  sean dos conjuntos cerrados disjuntos en  $X$  existen abiertos disjuntos  $U, V$  tales que  $A \subseteq U$  y  $B \subseteq V$ .  
Un espacio normal  $T_1$  se denomina  $T_4$  o normal de Hausdorff.
- $T_5$ : Un espacio es completamente normal si cualesquiera dos conjuntos pueden ser separados por vecindades.  
Un espacio se denomina  $T_5$  si es completamente normal y  $T_1$ .
- $T_6$ : Un espacio se llama perfectamente normal si cualesquiera dos conjuntos cerrados pueden ser separados por una función continua.  
Un espacio se denomina  $T_6$  si es  $T_1$  y perfectamente normal.

# Apéndice **C**

## Completo y Totalmente Acotado

Sea  $(X, d)$  un espacio métrico separable y completo.

**Definición.** Una cubierta para  $A$  de la forma  $\{B_\epsilon(x_k)\}$  se denomina  $\epsilon$ -red de  $A$ . (En este caso  $B_\epsilon(x_k)$  denota la bola abierta de radio  $\epsilon$  centrada en  $x_k$ ).

**Definición.** Un conjunto  $A \subseteq X$  se denomina totalmente acotado si  $\forall \epsilon > 0$  existe una  $\epsilon$ -red finita de subconjuntos de  $X$ .

**Teorema.** [4] p.213 Para  $A \subseteq X$ ,  $A$  es compacto si y sólo si  $A$  es completo y totalmente acotado.

*Demostración.* ( $\Rightarrow$ ) Si  $A$  es compacto, para cualquier  $\epsilon$ -red se tiene una subcubierta finita y por lo tanto  $A$  es totalmente acotado.

Por otro lado, si  $(x_n)_{n=1}^\infty$  es una sucesión de Cauchy en  $A$  admite una subsucesión convergente digamos  $x_{n_j} \rightarrow x^*$ .

Para  $\epsilon > 0$  dado, por la propiedad de Cauchy existe  $N_\epsilon$  tal que si  $m, n > N$  entonces  $d(x_n, x_m) < \frac{\epsilon}{2}$ .

Usando la convergencia escogemos  $n_j > N_\epsilon$  tal que  $d(x_{n_j}, x^*) < \frac{\epsilon}{2}$  y usando la desigualdad del triángulo encontramos,

$$d(x_n, y) \leq d(x_n, x_{n_j}) + d(x_{n_j}, x^*) < \epsilon$$

por lo que se concluye que la sucesión completa converge y entonces  $A$  es completo.

( $\Leftarrow$ ) Ahora supongamos que  $A$  es totalmente acotado y completo.

Consideremos  $(x_n)$  una sucesión en  $A$ , el objetivo es probar que admite una subsucesión convergente. Si la sucesión solo tuviera un número finito de términos se acaba la prueba por lo que sin pérdida de generalidad podemos suponer que es infinita, denotamos por  $\mathbf{B}$  al conjunto de distintos valores de la sucesión. Como  $A$  está totalmente acotado existen bolas abiertas  $B_{1/2}(y_{11}), \dots, B_{1/2}(y_{n_1})$  que cubren a  $A$ . Como supusimos que  $(x_n)$  tiene una infinidad de términos distintos, existe al menos una bola conteniendo una infinidad de términos. Escogemos alguna con esta propiedad y la llamamos  $\mathbf{B}_1$  y definimos  $C_1 = \mathbf{B} \cap \mathbf{B}_1$ . Como  $A$  es totalmente acotado podemos usar el mismo procedimiento pero para bolas de radio  $1/4$  y definimos  $\mathbf{B}_2$  como alguna de las bolas con una infinidad de términos en  $B$ . Posteriormente ponemos  $C_2 = C_1 \cap \mathbf{B}_2$  procedemos con inducción matemática y obtenemos  $\dots \subseteq C_2 \subseteq C_1$ . Se escoge  $n_i > n_{i-1}$  tal que  $x_{n_i} \in C_i$ .

$(x_{n_i})$  es la subsucesión deseada debido a que es de Cauchy: para  $\epsilon > 0$  dado sea  $N$  tal que  $2^{-N+1} < \epsilon$  y si  $N < i < j$  entonces  $x_{n_j} \in C_j \subseteq C_i$  y además  $d(x_{n_i}, x_{n_j}) < 2^{-i+1} < 2^{-N+1} < \epsilon$  y como  $A$  es completo se tiene que la subsucesión converge.

## Regularidad automática de las medidas finitas en espacios métricos [21] p. 27

**Proposición D.0.1.** *En un espacio métrico todo conjunto cerrado es un  $G_\delta$ .*

*Demostración.* Es evidente de la descomposición:

$$A = \bigcap_{i=1}^{\infty} \left\{ x : d(x, A) < \frac{1}{n} \right\}$$

**Teorema.** *Sea  $(X, d)$  un espacio métrico separable y  $\mu$  una medida finita definida en  $(X, \mathbb{B}_X)$  entonces  $\mu$  es regular.*

*Demostración.* Debemos probar que para  $\epsilon > 0$  dado y cualquier conjunto  $A$  en la  $\sigma$ -álgebra  $\mathbf{S}$  existe un abierto  $U_\epsilon$  y un cerrado  $C_\epsilon$  tal que  $\mu(U_\epsilon \setminus C_\epsilon) < \epsilon$ .

Sea  $\mathbf{S}$  la familia de todos los conjuntos de Borel que cumplen esta propiedad. Por definición  $\mathbf{S} \subseteq \mathbb{B}_X$ .

Probemos que  $\mathbf{S}$  forma una  $\sigma$ -álgebra de subconjuntos de  $\mathcal{P}(X)$ .

1.  $\emptyset$  y  $X$  están en  $\mathbf{S}$  por ser ellos mismos abiertos y cerrados.
2. Si  $A \in \mathbf{S}$  entonces  $C_\epsilon \subseteq A \subseteq U_\epsilon$  por lo que  $X \setminus U_\epsilon \subseteq X \setminus A \subseteq X \setminus C_\epsilon$ . Por lo que  $\mathbf{S}$  es cerrada ante la operación complemento debido a que el primero es un cerrado y el tercero es un abierto por ser complementos de abiertos y cerrados respectivamente.  
Con la propiedad  $\mu((X \setminus C_\epsilon) \setminus (X \setminus U_\epsilon)) = \mu(U_\epsilon \setminus C_\epsilon) < \epsilon$

3. Sea  $A_1, A_2, \dots$  una sucesión disjunta de conjuntos en  $\mathbf{S}$ , para  $\epsilon > 0$  dado escogemos en cada  $n$  conjuntos  $C_n \subseteq A_n \subseteq U_n$  tales que  $\mu(U_n \setminus C_n) < \frac{\epsilon}{3^n}$

Como  $\mu$  es una medida entonces para que la serie  $\sum_{i=1}^{\infty} \mu(C_i)$  pueda converger necesariamente  $\mu(C_k) \rightarrow 0$

por lo que escogemos  $k$  tal que  $\mu\left(\bigcup_{n=1}^{\infty} C_n \setminus \bigcup_{n=1}^k C_n\right) < \frac{\epsilon}{2}$

Ponemos  $U_\epsilon = \bigcup_{n=1}^{\infty} U_n$  y  $C_\epsilon = \bigcup_{n=1}^k C_n$  y así  $C_\epsilon \subseteq \bigcup_{n=1}^{\infty} A_n \subseteq U_\epsilon$  con  $C_\epsilon$  cerrado por ser unión finita de cerrados y  $U_\epsilon$  abierto.

$$\text{Concluimos que } \mu(U_\epsilon \setminus C_\epsilon) \leq \mu\left(U_\epsilon \setminus \bigcup_{n=1}^{\infty} C_n\right) + \mu\left(\bigcup_{n=1}^{\infty} C_n \setminus C_\epsilon\right) < \sum_{n=1}^{\infty} \mu(U_n \setminus C_n) + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Y por lo mismo se obtiene que  $\mathbf{S}$  es una  $\sigma$ -álgebra contenida en  $\mathbb{B}_X$ . Para concluir la prueba del teorema veamos que cualquier cerrado es elemento de  $\mathbf{S}$ .

Sea  $C$  un conjunto cerrado y  $\epsilon > 0$  por la proposición anterior sea  $(U_n)$  una sucesión decreciente de conjuntos tales que  $C = \bigcap_{n=1}^{\infty} C_n$ . Como  $\mu(U_n) \rightarrow \mu(C)$  entonces existe  $U_\epsilon$  tal que  $\mu(U_\epsilon \setminus C) < \epsilon$ .