



---

**Universidad de Valladolid**

Facultad de Ciencias

## **TRABAJO FIN DE GRADO**

Grado en Matemáticas

**Órdenes entre distribuciones de probabilidad**

*Autor: Alejandro Granado González*

*Tutor: Carlos Matrán Bea*



*A los que alguna vez  
creyeron en mí*

*Valladolid, diciembre de 2020*



# Resumen

Una cuestión que tenemos muy presente en el día a día es **comparar**. Lo hacemos por ejemplo al tratar de ver cuál de entre dos cantidades es más grande, cuál de entre dos cualidades es mejor... Si se introducen los términos estadísticos adecuados, modelando las distribuciones de los fenómenos observados en una cierta población, el **orden estocástico** entre dos distribuciones de probabilidad nos permite dar una respuesta global a estas comparaciones, más adecuada en muchas situaciones que las más simples y habituales que se limitan a comparar características como la media, la mediana o la varianza. Este concepto también se utiliza en ámbitos más específicos como en economía, medicina o sociología, permitiéndonos además resolver cuestiones más propias de la estadística descriptiva. El estudio de este tipo de comparación se remonta a Mann y Whitney ([24], 1947) y a Lehmann ([21], 1955).

La motivación de este Trabajo Fin de Grado consiste en mostrar algunas de las definiciones de orden estocástico existentes, cada una de las cuales nos permite resolver problemas como el de estudiar si una característica es mayor en una determinada población que en otra o el de comparar su variabilidad. Por ejemplo, la afirmación de que los hombres son más altos que las mujeres puede analizarse a través del orden estocástico e involucra una comparación del estatus de cada individuo de un grupo visto como si fuera individuo del otro grupo. Si para cada uno de ellos su estatus mejora (tal como decir que un hombre de una cierta altura parecería tener una altura mayor entre las mujeres) estamos entonces ante la presencia de orden estocástico, de modo que obtenemos así un enfoque novedoso que supone una relación global entre las distribuciones implicadas.

El gran inconveniente que presenta es que en algunas ocasiones el concepto de orden estocástico es demasiado rígido, puesto que supone una relación entre las distribuciones que o bien existe o bien no. Por esta razón en este trabajo se estudiarán también algunas versiones más débiles que nos permitirán resolver el problema de una manera aproximada así como adaptarlo a escenarios apropiados según cada caso que consideremos. Esta perspectiva nos da además la posibilidad de determinar un nivel de desviación respecto a la relación de orden que será fundamental al tratar con muestras. Para mostrar su utilidad se incluye un análisis de la evolución del envejecimiento en España a lo largo de los últimos años así como su evolución prevista en los próximos, en el cual se comparan los datos de edad de la población durante varios periodos de tiempo.

**Palabras clave:** comparar, orden estocástico, índices de desviación, comparación de medias, órdenes estocásticos convexos, mayorización.



# Índice general

Resumen	I
Notación	V
Introducción	1
<b>1. Orden estocástico usual</b>	<b>5</b>
1.1. Caracterizaciones del orden estocástico usual . . . . .	6
1.2. Orden estocástico usual en familias de localización y escala . . . . .	10
1.3. Consecuencias del orden estocástico usual . . . . .	13
1.4. Condiciones suficientes . . . . .	15
1.4.1. Orden del cociente de riesgo . . . . .	15
1.4.2. Orden de razón de verosimilitud . . . . .	18
1.5. Orden estocástico usual multivariante . . . . .	20
<b>2. Aproximaciones al orden estocástico usual</b>	<b>23</b>
2.1. Índices de desviación . . . . .	24
2.1.1. Precedencia estocástica y el índice $\rho(F, G)$ . . . . .	25
2.1.2. Orden estocástico restringido y el índice $\gamma(F, G)$ . . . . .	26
2.1.3. Casi orden estocástico y el índice $\alpha(F, G)$ . . . . .	28
2.1.4. Modelos de contaminación, recorte de probabilidades y el índice $\pi(F, G)$ . . . . .	30
2.2. $\gamma(F, G)$ y $\pi(F, G)$ en los modelos de localización y escala . . . . .	36
2.2.1. $\gamma(F, G)$ en modelos LS . . . . .	36
2.2.2. $\pi(F, G)$ en modelos LS . . . . .	37
2.3. Orden estocástico usual <i>versus</i> comparación de medias . . . . .	39
2.4. Contrastando el orden estocástico usual . . . . .	40
2.5. Ejemplo: Comparación de la altura entre hombres y mujeres . . . . .	42
2.5.1. Caso 1 . . . . .	43
2.5.2. Caso 2 . . . . .	44
2.6. Ejemplo real: Comparación de las edades de dos poblaciones para determinar si una está mas envejecida que la otra . . . . .	47
2.6.1. Evolución del envejecimiento en España . . . . .	52

<b>3. Órdenes estocásticos convexos</b>	<b>55</b>
3.1. Caracterizaciones de los órdenes estocásticos convexos . . . . .	58
3.2. El teorema de Strassen . . . . .	64
3.3. Órdenes estocásticos convexos en familias de localización y escala . . . . .	67
3.4. Consecuencias de los órdenes estocásticos convexos . . . . .	68
3.5. Contrastando el orden estocástico convexo . . . . .	70
3.6. Órdenes convexos multivariantes . . . . .	72
<b>4. Mayorización</b>	<b>75</b>
4.1. Principio de transferencia . . . . .	77
4.2. Relación entre la mayorización y los órdenes estocásticos convexos . . . . .	80
4.3. Consecuencias de la mayorización . . . . .	82
4.3.1. Midiendo la aleatoriedad . . . . .	83
4.3.2. Dualidad con el concepto de entropía . . . . .	84
<b>A. Resultados de teoría de la probabilidad</b>	<b>87</b>
A.1. Función cuantil . . . . .	87
A.2. Momentos de una variable aleatoria . . . . .	88
A.3. Convergencias . . . . .	89
<b>B. Algunos resultados de análisis</b>	<b>91</b>
B.1. Reordenamientos decrecientes . . . . .	91
B.2. El teorema de Radon-Nikodym . . . . .	93
<b>C. Programas en R</b>	<b>95</b>
C.1. <code>gamma()</code> . . . . .	95
C.2. <code>alpha()</code> . . . . .	96
<b>Bibliografía</b>	<b>101</b>

# Notación

El trabajo está formado por capítulos que se dividen en secciones y subsecciones.

Las definiciones, lemas, proposiciones, teoremas, corolarios, observaciones y ejemplos se enumeran progresivamente según capítulo, sección y un dígito adicional que los ordena.

El símbolo  $\square$  denota el fin de una demostración.

Los vectores se representarán en negrita a menos que se escriban de forma explícita, indicando sus componentes.

Otra notación que se utilizará es la siguiente:

## Conjuntos

- $\mathbb{N}$  Conjunto de los números naturales (sin incluir el 0).
- $\mathbb{Z}$  Conjunto de los números enteros.
- $\mathbb{Z}_+$  Conjunto de los números enteros no negativos (incluye el 0).
- $\mathbb{R}$  Conjunto de los números reales.
- $\mathbb{R}_+$  Conjunto de los números reales no negativos.

## General

- $A = (a_{ij})_{i \in I, j \in J}$  Matriz con entradas  $a_{ij}$  con  $i \in I, j \in J$ .
- $a_{ij}$  Entrada de la matriz  $A$  correspondiente a la fila  $i$  y columna  $j$ .
- $I$  Matriz identidad (con dimensiones adecuadas al contexto).

## Distribuciones

- $X \sim F$  La variable aleatoria  $X$  tiene distribución  $F$ .
- $X \stackrel{d}{=} Y$  Las variables aleatorias  $X$  e  $Y$  están igualmente distribuidas.
- $B(n, p)$  Distribución binomial de parámetros  $n$  y  $p$ .
- $\varepsilon(\lambda)$  Distribución exponencial de parámetro  $\lambda > 0$ .
- $U(a, b)$  Distribución uniforme continua en el intervalo  $(a, b)$ ,  $a < b$ .
- $W(\lambda, k)$  Distribución de Weibull con parámetro de escala  $\lambda > 0$  y parámetro de forma  $k > 0$ .

- $N(\mu, \sigma)$  Distribución normal de media  $\mu$  y desviación típica  $\sigma > 0$ .  
 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  Distribución normal multivariante con vector de medias  $\boldsymbol{\mu}$  y matriz de covarianzas  $\boldsymbol{\Sigma}$ .

## Funciones

- $P$  Función de probabilidad.  
 $I_A$  Función indicatriz o indicadora del conjunto  $A$ .  
 $E[X]$  o  $EX$  Esperanza matemática de la variable aleatoria  $X$ .  
 $E[X^k]$  Momento de orden  $k$  centrado en el origen de la variable aleatoria  $X$ .  
 $Var(X)$  Varianza de la variable aleatoria  $X$ .  
 $\sigma_X$  Desviación típica de la variable aleatoria  $X$ .  
 $F^{-1}$  Función cuantil de una distribución de probabilidad con función de distribución  $F$ .  
 $\bar{F}$  Función de supervivencia asociada a la función de distribución  $F$ .  
 $\pi_X$  Integral de función de supervivencia asociada a la variable aleatoria  $X$ .

## Acrónimos

- v.a. (v.as.) Variable(s) Aleatoria(s).  
i.i.d. Independiente e idénticamente distribuido.  
c.s. Casi seguro.

## Órdenes

- $\leq_{st}$  Orden estocástico usual.  
 $\leq_{hr}$  Orden del cociente de riesgo.  
 $\leq_{lr}$  Orden de razón de verosimilitud.  
 $\leq_{cx}$  Orden estocástico convexo.  
 $\leq_{icx}$  Orden estocástico convexo creciente.  
 $\leq_{icv}$  Orden estocástico cóncavo creciente.  
 $\leq_D$  Orden de peligrosidad.  
 $\leq_M$  Mayorización.  
 $\leq_{rand}$  Orden de aleatoriedad.  
 $\leq_{uo}$  Orden del octante superior.  
 $\leq_{lo}$  Orden del octante inferior.  
 $\leq_{sp}$  Precedencia estocástica.

# Introducción

Consideremos un espacio probabilístico  $(\Omega, \sigma, P)$  el cual está formado por un conjunto  $\Omega$  al que denominaremos *espacio muestral*, una  $\sigma$ -álgebra  $\sigma$  definida sobre  $\Omega$  formada por elementos a los que llamaremos *sucesos* y una función de probabilidad  $P : \sigma \rightarrow [0, 1]$  que asigna a cada suceso una probabilidad. En los casos en los que trabajemos sobre la recta real, sobre esta se considerará la  $\sigma$ -álgebra de Borel que denotaremos por  $\beta$ .

Sea  $X : \Omega \rightarrow \mathbb{R}$  una *variable aleatoria*, esto es, una función  $\sigma|\beta$ -medible definida sobre  $\Omega$  con valores reales. Se define la *ley* o *distribución de probabilidad* de  $X$ ,  $P_X$ , como  $P_X(A) = P(X \in A)$ , para todo  $A \in \beta$ . A partir de esta queda determinada su *función de distribución*  $F_X$  como

$$F_X(t) = P_X((-\infty, t]) = P(X \leq t),$$

para todo  $t \in \mathbb{R}$ . Se define también su *función de supervivencia*  $\bar{F}_X$  como

$$\bar{F}_X(t) = P(X > t) = 1 - F_X(t).$$

Un *vector aleatorio*  $\mathbf{X} = (X_1, \dots, X_n)$  es una colección de  $n$  variables aleatorias unidimensionales definido sobre un espacio de probabilidad  $(\Omega, \sigma, P)$ , esto es, se trata de una función  $\sigma|\beta^n$ -medible donde  $\beta^n$  denota a la  $\sigma$ -álgebra de Borel sobre  $\mathbb{R}^n$ . El espacio  $\mathbb{R}^n$  se considera con su orden parcial usual, es decir,  $\mathbf{x} \leq \mathbf{y}$  si  $x_i \leq y_i$  para cada  $i = 1, \dots, n$ .

La distribución del vector  $\mathbf{X}$  es por tanto una medida de probabilidad sobre  $(\mathbb{R}^n, \beta^n)$  y queda determinada por su *función de distribución*

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Se define también su *función de supervivencia* como

$$\bar{F}_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} > \mathbf{x}) = P(X_1 > x_1, \dots, X_n > x_n).$$

A las funciones de distribución  $F_i(x) = P(X_i \leq x)$  para  $i = 1, \dots, n$  se les llama *marginales*.

Un *orden estocástico* es básicamente una relación de orden parcial  $\preceq$  definida sobre el conjunto de todas las funciones de distribución de variables aleatorias con valores en  $\mathbb{R}$  (o sobre un subconjunto

suyo adecuado). Este concepto se puede extender también al conjunto de todas las funciones de distribución de vectores aleatorios con valores en  $\mathbb{R}^n$ , como veremos para cada orden que estudiaremos en este trabajo.

## Estructura de la memoria

En el primer capítulo se tratará el concepto del *orden estocástico usual*, el cual se centra en comparar dos poblaciones (más en concreto, sus distribuciones) para determinar si una cierta característica puede considerarse mayor en una población que en la otra. Se ofrecen algunas caracterizaciones equivalentes, algunas condiciones suficientes y su extensión al caso multivariante. A modo de ejemplo se propone también la forma de comparar entre distribuciones pertenecientes a la misma familia de localización y escala, un hecho que resulta habitual al trabajar con el orden estocástico en la práctica. La bibliografía básica que se ha tomado es el libro de Müller y Stoyan [27] en el cual se basa la estructura que seguiremos, así como el libro de Shaked y Shanthikumar [30]. Se han elegido también algunos resultados encontrados en el libro de Belzunce, Martínez-Riquelme y Mulero [7] y en el de Levy [23].

En el segundo capítulo se mostrarán algunas relajaciones del orden estocástico usual a través del estudio de diversos índices que miden el grado de desviación respecto a la relación que lo define. En él también se confronta el orden estocástico frente a la comparación de medias como procedimientos para asegurar el aumento o la mejora de una propiedad en una población, así como el método que se sigue habitualmente para medir el grado de confianza que nos permite asegurar el orden cuando se trabaja con muestras. Como bibliografía se han tomado los artículos de Álvarez-Esteban et al. [1] (2014) y [2] (2017), el artículo de del Barrio et al. [6] (2018) y los problemas planteados en el artículo de Berger [8] y en el de Mann y Whitney [24]. Al final del capítulo se incluyen varios ejemplos en los que se muestra cómo se puede utilizar el orden estocástico (y en concreto, los índices de desviación) para analizar algunos problemas más relativos al ámbito de la estadística descriptiva.

El tercer capítulo se dedica al estudio de los *órdenes estocásticos convexos*, los cuales se utilizan para comparar la variabilidad que existe entre los valores que toman las variables aleatorias que se tomen en cuestión. Veremos que esta definición, a pesar de ser sencilla, en la práctica no resulta útil, por lo que se recurre a otras caracterizaciones equivalentes más fáciles de comprobar pero cuya demostración es compleja. Asimismo se aportarán algunos resultados relacionados con la ley fuerte de los grandes números, extenderemos el concepto al caso multivariante y veremos también cómo se aplica al caso de las familias de localización y escala. Se proponen también algunos métodos e índices que permiten realizar contrastes de manera similar a lo tratado en el capítulo 2 para el orden usual. Para el contenido sobre órdenes convexos se ha consultado el libro de Szekli [32], el de Müller y Stoyan [27] y el de Shaked y Shanthikumar [30], y para los índices los artículos de Berrendero y Cárcamo [9] (2011) y [10] (2012).

En el cuarto y último capítulo se considera el concepto de *mayorización* aplicado al caso de vectores aleatorios como forma de medir la dispersión que existe entre sus componentes. Este se puede

usar también para comparar la heterogeneidad de distribuciones de probabilidad sobre un conjunto de  $n$  puntos, permitiendo su comparación con el objetivo de ver cuál de ellas es menos aleatoria. Se muestra además la relación que existe entre la mayorización y los órdenes estocásticos convexos, de forma que la primera se podría interpretar como una versión discreta de los segundos. Este capítulo abarca también la dualidad existente entre la mayorización y la idea de *entropía*, habitual en la teoría de la información y utilizada como forma de medir la aleatoriedad. Finalmente se aportan algunos ejemplos de su uso. Como bibliografía se han tomado el libro de Marshall y Olkin [25], el de Bhatia [11] y los artículos de Hickey [19] (1982) y [20] (1983).

Con el objetivo de que la memoria sea lo más autocontenida posible, al final del texto se han añadido tres apéndices: el primero recoge algunos contenidos sobre teoría de la probabilidad que facilitan la exposición y explicación de los órdenes estocásticos aquí considerados. El segundo contiene algunos resultados de análisis y teoría de la medida que son necesarios para entender algunos aspectos que se exponen a lo largo del trabajo. La razón por la que se han incluido estos y no otros es o bien para facilitar la exposición de algunas partes o porque son contenidos no triviales que no se han tratado a lo largo del grado y que merece la pena entender. El tercer apéndice incluye los códigos de dos programas elaborados para calcular los índices de desviación expuestos en el capítulo 2.

En este trabajo se han considerado los órdenes estocásticos que mayor relevancia han tenido desde los inicios de esta teoría. Muchos de ellos se han desarrollado motivados por las diferentes áreas de aplicación, e intentando siempre dar explicación a problemas del mundo real. Existen muchos más órdenes definidos entre distribuciones que atienden a otras características, como la dispersión o el no depender de la localización (véase el *orden dispersivo*), o que utilizan otras herramientas como transformadas funcionales (véase el *orden de la transformada de Laplace*), entre otros. Todos ellos son también interesantes y tienen numerosas aplicaciones. Si se desea ampliar sobre el tema, puede encontrarse más información en el libro de Müller y Stoyan [27] (con aplicaciones a la teoría de riesgos), en el de Shaked y Shanthikumar [30] o en el de Szekli [32] (con aplicaciones a cadenas de Markov y teoría de colas), y desde un punto de vista más aplicado en el libro de Belzunde, Martínez-Riquelme y Mulero [7] o en el de Levy [23] (con aplicaciones a la economía y a las finanzas).

**Convención:** en lo que sigue, dadas variables aleatorias  $X$  e  $Y$  con distribuciones  $P_X$  y  $P_Y$  y funciones de distribución  $F_X$  y  $F_Y$  para las que se verifica  $X \preceq Y$  para un orden estocástico  $\preceq$  dado, se utilizará también la notación  $P_X \preceq P_Y$  y  $F_X \preceq F_Y$  para indicar este mismo hecho.



# Capítulo 1

## Orden estocástico usual

Para introducir el concepto consideremos el siguiente ejemplo: supongamos que  $X$  e  $Y$  son dos variables aleatorias que describen los beneficios que se obtienen de dos inversiones distintas. Podemos asegurar que  $Y$  nos reporta una mayor cantidad de beneficios que  $X$  si para cualquier valor  $t$  que fijemos la probabilidad de que una inversión nos devuelva un beneficio mayor o igual que  $t$  es siempre mayor para  $Y$  que para  $X$ . Esto da sentido a la siguiente definición:

**Definición 1.0.1.** Dadas dos variables aleatorias  $X$  e  $Y$ , se dice que  $X$  es menor que  $Y$  respecto al orden estocástico usual (escrito  $X \leq_{st} Y$ ) si

$$F_X(t) \geq F_Y(t) \text{ para todo } t \in \mathbb{R}. \quad (1.1)$$

En algunas ocasiones como en el ejemplo anterior es más sencillo estudiar el orden estocástico usual utilizando la función de supervivencia  $\bar{F}_X$ , ya que  $X \leq_{st} Y$  es equivalente a

$$\bar{F}_X(t) \leq \bar{F}_Y(t) \text{ para todo } t \in \mathbb{R}. \quad (1.2)$$

Al orden estocástico usual también se le denomina en algunos manuales *primer orden estocástico*, especialmente en los que se aplica este concepto a la economía y las finanzas como en el libro de Levy [23]. Es muy fácil ver que la relación  $\leq_{st}$  se trata de una relación de orden parcial.

**Observación 1.0.2.** La definición anterior es sencilla de interpretar: si  $F_X(t) \geq F_Y(t)$  para todo  $t$  real (esto es,  $P(X \leq t) \geq P(Y \leq t)$  para todo  $t$ ) entonces  $X$  toma valores más pequeños con mayor probabilidad que  $Y$ , y a la vista de  $\bar{F}_X(t) \leq \bar{F}_Y(t)$  para todo  $t$  real (es decir,  $P(X > t) \leq P(Y > t)$  para todo  $t$ ) se tiene que  $X$  toma valores más grandes con menor probabilidad que  $Y$ . **El orden estocástico usual se centra básicamente en determinar cuál de entre dos distribuciones de probabilidad es la dominante**, permitiéndonos comparar dos variables que describen una característica de dos poblaciones para determinar si puede considerarse mayor en una que en la otra. Aplicándolo al ejemplo del inicio, si para  $X$  e  $Y$  se tiene que  $X \leq_{st} Y$ , podemos afirmar que el proyecto asociado a  $X$  nos devolverá con mayor probabilidad menos beneficios que el asociado a  $Y$ .

**Observación 1.0.3.** La condición que define el orden estocástico usual implica que las funciones de distribución (o funciones de supervivencia, según se considere) no se pueden cruzar, ya que esto no nos permitiría asegurar la dominación.

## 1.1. Caracterizaciones del orden estocástico usual

Al comparar dos poblaciones para determinar si una es más grande que la otra (como puede suceder con la altura entre hombres y mujeres, el tiempo de vida ante dos tratamientos, los beneficios económicos de dos proyectos distintos...), la forma intuitiva que tenemos de hacerlo es comparar individuo a individuo. Si extraemos dos muestras de las poblaciones, ordenamos sus valores de menor a mayor (obteniendo así estatus equiparables) y los vamos comparando dos a dos de modo que siempre el valor de la primera es menor que el de la segunda, entonces podremos afirmar que una población es estocásticamente más pequeña que la otra. Esta idea se refleja en la siguiente caracterización del orden estocástico usual **en términos de la función cuantil**, extraída del libro de Belzunce, Martínez-Riquelme y Mulero [7]. Haremos uso de algunas de sus propiedades, las cuales vienen resumidas en la proposición A.1.1 del apéndice A situado al final del texto.

**Proposición 1.1.1.** Sean  $X, Y$  variables aleatorias con funciones cuantil  $F^{-1}$  y  $G^{-1}$  respectivamente. Entonces  $X \leq_{st} Y$  si y solo si

$$F^{-1}(p) \leq G^{-1}(p) \text{ para todo } p \in (0, 1). \quad (1.3)$$

*Demostración.* Sea  $p \in (0, 1)$ . Si  $X \leq_{st} Y$  entonces para cada  $t$  se tiene  $F(t) \geq G(t)$ . Utilizando la propiedad A.1.1.2 se tiene que

$$[F^{-1}(p), +\infty) = \{t \in \mathbb{R} \mid F(t) \geq p\} \supseteq \{t \in \mathbb{R} \mid G(t) \geq p\} = [G^{-1}(p), +\infty),$$

luego  $F^{-1}(p) \leq G^{-1}(p)$  para todo  $p \in (0, 1)$ .

Recíprocamente, supongamos que  $F^{-1}(p) \leq G^{-1}(p)$  para todo  $p \in (0, 1)$  y tomemos una variable aleatoria  $U$  distribuida uniformemente en el intervalo  $(0, 1)$ . A partir de nuestra hipótesis se tiene que  $\{F^{-1}(U) \leq t\} \supseteq \{G^{-1}(U) \leq t\}$ , para todo  $t \in \mathbb{R}$ . Por la propiedad A.1.1.6  $F^{-1}(U)$  y  $G^{-1}(U)$  tienen funciones de distribución  $F$  y  $G$  respectivamente. Entonces, utilizando la monotonía de  $P$ ,

$$F(t) = P(F^{-1}(U) \leq t) \geq P(G^{-1}(U) \leq t) = G(t),$$

para todo  $t \in \mathbb{R}$ , lo que equivale a  $X \leq_{st} Y$ . □

El resultado anterior nos permite además analizar el problema del orden dentro de un espacio de medida finita, el intervalo  $(0, 1)$  sobre el cual se considera la medida de Lebesgue  $\ell$ . De este modo podemos medir la proporción de casos en los que no se cumple la relación de orden estocástico que hemos estudiado calculando  $\ell(\{F^{-1} > G^{-1}\})$ . Esto nos proporciona una manera de aproximarnos al orden estocástico que estudiaremos con más detalle en el capítulo 2.

El siguiente teorema caracteriza completamente al orden estocástico a través de una **propiedad de representación** y una **caracterización integral para funciones crecientes**. Puede encontrarse en el libro de Müller y Stoyan [27].

**Teorema 1.1.2.** Sean  $X, Y$  dos variables aleatorias con funciones de distribución  $F$  y  $G$  respectivamente. Las siguientes afirmaciones son equivalentes:

1.  $X \leq_{st} Y$ ;
2. Existen un espacio de probabilidad  $(\Omega, \mathcal{A}, P)$  y variables aleatorias  $\hat{X}$  e  $\hat{Y}$  definidas sobre él tales que  $X \stackrel{d}{=} \hat{X}$ ,  $Y \stackrel{d}{=} \hat{Y}$  y  $P(\hat{X} \leq \hat{Y}) = 1$  (esto es,  $\hat{X}(\omega) \leq \hat{Y}(\omega)$  casi seguro para  $\omega \in \Omega$ ).
3.  $\phi(X) \leq_{st} \phi(Y)$  para toda función real  $\phi$  creciente.
4.  $E[\phi(X)] \leq E[\phi(Y)]$  para toda función real  $\phi$  creciente siempre que las esperanzas anteriores sean finitas.

Además, si dada una función  $\phi$  la desigualdad en 4. se mantiene para todas las variables  $X, Y$  que verifican  $X \leq_{st} Y$ , entonces  $\phi$  ha de ser creciente.

*Demostración.* En primer lugar probaremos la equivalencia entre 1 y 2. Supongamos que  $X \leq_{st} Y$  y consideremos una v.a.  $U$  distribuida uniformemente en el intervalo  $(0, 1)$ . Por la propiedad A.1.1.6 de la función cuantil, se tiene que  $\hat{X} := F^{-1}(U) \stackrel{d}{=} X$ ,  $\hat{Y} := G^{-1}(U) \stackrel{d}{=} Y$  y por la proposición 1.1.1 se concluye que  $P[F^{-1}(U) \leq G^{-1}(U)] = 1$ . Recíprocamente, supongamos ahora que existen dos variables aleatorias  $\hat{X}$  e  $\hat{Y}$  definidas sobre el mismo espacio de probabilidad tales que  $X \stackrel{d}{=} \hat{X}$ ,  $Y \stackrel{d}{=} \hat{Y}$  y  $P(\hat{X} \leq \hat{Y}) = 1$ . En vista de esto, para todo  $x \in \mathbb{R}$  sucede que  $P(x < \hat{X} \mid \hat{X} \leq \hat{Y}) \leq P(x < \hat{Y} \mid \hat{X} \leq \hat{Y})$ , lo que equivale por la probabilidad condicionada a que  $P(\hat{X} > x, \hat{X} \leq \hat{Y}) \leq P(\hat{Y} > x, \hat{X} \leq \hat{Y})$ . Del hecho de que  $X \leq Y$  c.s. se sigue que  $\bar{F}(x) = P(\hat{X} > x) \leq P(\hat{Y} > x) = \bar{G}(x)$  para todo  $x$  real, lo cual equivale a  $X \leq_{st} Y$  por 1.2.

Probemos ahora la equivalencia con 3. Si  $X \leq_{st} Y$  y consideramos las variables  $\hat{X}$  e  $\hat{Y}$  descritas anteriormente, dada una función  $\phi$  creciente sucede que  $\phi(\hat{X}) \stackrel{d}{=} \phi(X)$ ,  $\phi(\hat{Y}) \stackrel{d}{=} \phi(Y)$ , y al ser igualmente distribuidas  $P[\phi(\hat{X}) \leq \phi(\hat{Y})] = 1$  por la equivalencia con 2. Entonces se tiene que  $\phi(X) \leq_{st} \phi(Y)$ . La implicación contraria se tiene tomando la identidad, esto es,  $\phi(x) = x$ .

Finalmente veamos la equivalencia con 4. Supongamos que  $X \leq_{st} Y$ . Por la equivalencia probada con 2, podemos suponer sin pérdida de generalidad que  $X \leq Y$  c.s.. Entonces si  $\phi$  es creciente se mantiene que  $\phi(X) \leq \phi(Y)$  c.s. y por la monotonía de la esperanza se tiene  $E[\phi(X)] \leq E[\phi(Y)]$ . Para el recíproco utilizaremos la propiedad de que  $E(I_A) = P(A)$  para todo  $A \in \sigma$ . Si lo aplicamos a la función

$$f_t(x) = I_{(t, \infty)}(x) = \begin{cases} 1 & \text{si } x > t \\ 0 & \text{en otro caso,} \end{cases}$$

que es en efecto creciente, se tiene que  $\bar{F}(t) = P(X > t) = E f_t(X)$ . Entonces como  $E f_t(X) \leq E f_t(Y)$  sucede que  $\bar{F}(t) \leq \bar{G}(t)$  para todo  $t$ , esto es,  $X \leq_{st} Y$ .

Para probar la última afirmación, supongamos que  $\phi$  no es creciente, luego existen  $x \leq y$  con  $\phi(x) > \phi(y)$  y supongamos que  $P(X = x) = P(Y = y) = 1$ . Entonces  $X \leq_{st} Y$  pero  $E[\phi(X)] > E[\phi(Y)]$ , en contra de 4.  $\square$

Las caracterizaciones realizadas en el teorema anterior no solo tienen interés desde el punto de vista teórico sino también desde el práctico: la propiedad de representación en 1.1.2.2 es muy útil para probar si se verifica la condición de orden estocástico al realizar simulaciones con variables aleatorias. La caracterización dada en 1.1.2.3 muestra que la dominación se conserva por transformaciones crecientes, y en muchas ocasiones resulta más sencillo probar esto (por ejemplo, si el dinero que se gana en un negocio depende de manera creciente del tiempo invertido). De la caracterización en términos integrales en 1.1.2.4 se sigue el siguiente corolario presente en el libro de Müller y Stoyan [27], que muestra que **si las medias no están ordenadas, no existe orden estocástico** y que **diferentes distribuciones con la misma esperanza no se pueden ordenar con respecto al orden  $\leq_{st}$** .

**Corolario 1.1.3.** Sean  $X$  e  $Y$  variables aleatorias con esperanza finita.

1. Si  $X \leq_{st} Y$  entonces  $EX \leq EY$ .
2. Si  $X \leq_{st} Y$  y  $EX = EY$  entonces  $X$  e  $Y$  tienen la misma distribución.

*Demostración.*

1. Se deduce del teorema 1.1.2.4 al considerar  $\phi(x) = x$ .
2. La esperanza de una variable aleatoria se puede representar como

$$EX = \int_0^{\infty} [1 - F_X(t)] dt - \int_{-\infty}^0 F_X(t) dt$$

(véase la proposición A.2.3 en el apéndice). Con lo anterior, se tiene que

$$EY - EX = \int_{-\infty}^{\infty} [F_X(t) - F_Y(t)] dt.$$

Si  $X \leq_{st} Y$  y  $EX = EY$ , entonces el primer miembro de la igualdad anterior es 0, y el segundo miembro es la integral de una función no negativa (pues como  $X \leq_{st} Y$  entonces  $F_X(t) - F_Y(t) \geq 0$  para todo  $t$ ) y continua por la derecha. Que tal integral sea igual a 0 solo es posible si la función del integrando también se anula. Por lo tanto,  $F_X = F_Y$ .

□

Aunque en general no es cierto, hay algunas situaciones especiales en las que la condición de dominio en las variables sí es equivalente al orden en las esperanzas, como sucede si ambas tienen ley exponencial de distinto parámetro (puede encontrarse el enunciado en el libro de Levy [23]):

**Proposición 1.1.4.** Sean  $X$  e  $Y$  variables aleatorias con distribución exponencial de parámetros  $\lambda$  y  $\mu$ . Entonces  $X \leq_{st} Y$  si y solo si  $EX \leq EY$ , si y solo si  $\lambda \geq \mu$ .

*Demostración.* Una de las implicaciones es cierta para cualquier distribución. Veamos la otra: supongamos que  $X \sim \varepsilon(\lambda)$  e  $Y \sim \varepsilon(\mu)$ . Entonces  $X$  tiene función de distribución  $F(x) = 1 - e^{-\lambda x}$  si

$x \geq 0$  e  $Y$  tiene función de distribución  $G(x) = 1 - e^{-\mu x}$  si  $x \geq 0$ . Entonces, para  $x$  fijado,

$$\begin{aligned} G(x) \leq F(x) &\Leftrightarrow 1 - e^{-\mu x} \leq 1 - e^{-\lambda x} \Leftrightarrow -e^{-\mu x} \leq -e^{-\lambda x} \Leftrightarrow \\ &\Leftrightarrow \mu \leq \lambda \Leftrightarrow \frac{1}{\lambda} \leq \frac{1}{\mu} \Leftrightarrow EX \leq EY, \end{aligned}$$

donde hemos utilizado que la función  $e^{-x}$  es estrictamente decreciente para  $x \geq 0$ . □

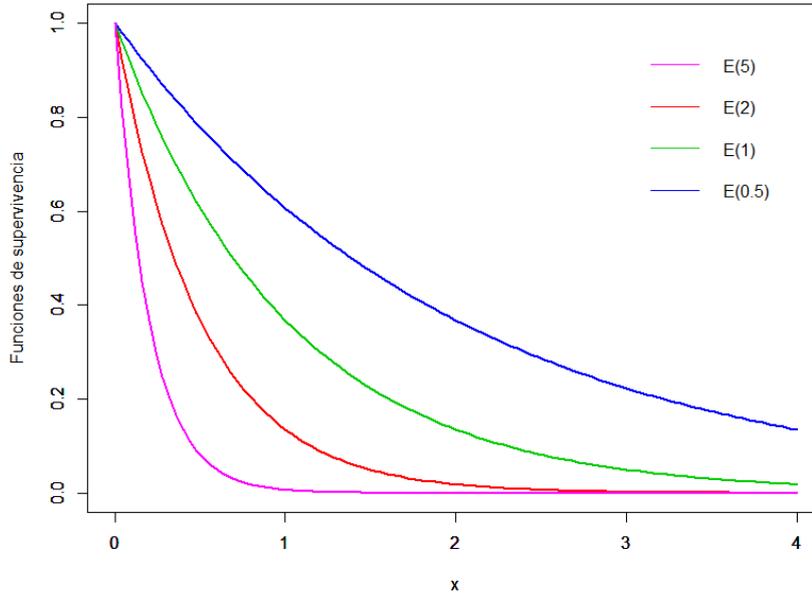


Figura 1.1: Comparación de funciones de supervivencia de variables exponenciales

Si consideramos variables  $U \sim \varepsilon(5)$ ,  $V \sim \varepsilon(2)$ ,  $X \sim \varepsilon(1)$  e  $Y \sim \varepsilon\left(\frac{1}{2}\right)$ , según dice el teorema anterior, como  $\frac{1}{2} < 1 < 2 < 5$ , entonces se tiene que  $U \leq_{st} V \leq_{st} X \leq_{st} Y$ . Esto lo podemos observar en la gráfica de sus funciones de supervivencia en la Figura 1.1. Recordemos que la condición  $X \leq_{st} Y$  equivale a la relación  $\bar{F}_X(t) \leq \bar{F}_Y(t)$  para todo  $t \in \mathbb{R}$  entre las funciones de supervivencia.

Generalizamos lo anterior a la comparación de momentos de orden superior en el siguiente corolario, tomado del libro de Müller y Stoyan [27]:

**Corolario 1.1.5.** Si  $X \leq_{st} Y$ , entonces  $EX^n \leq EY^n$  para  $n = 1, 3, 5, \dots$  siempre que los momentos anteriores existan. En el caso de que  $X$  e  $Y$  sean variables aleatorias no negativas se verifica  $EX^n \leq EY^n$  para  $n = 1, 2, 3, \dots$

*Demostración.* Simplemente basta observar la monotonía de las funciones  $\phi(x) = x^n$ . Para  $n$  impar su gráfica es creciente, luego basta aplicar la propiedad 4 del teorema 1.1.2 para concluir. En el caso de que  $x \geq 0$ , la función  $\phi$  es creciente para todo  $n \in \mathbb{N}$ , lo que prueba la segunda parte del corolario. □

## 1.2. Orden estocástico usual en familias de localización y escala

Cuando se consideran modelos paramétricos, el orden estocástico usual se utiliza frecuentemente entre distribuciones pertenecientes a la misma familia de localización y escala, como es el caso de las distribuciones normal y uniforme continua, entre otras. Por tanto sería deseable disponer de algún criterio que nos permita dar una caracterización del orden estocástico usual en términos de la distribución de referencia y los parámetros implicados.

Sea  $\mathcal{F}_0$  una función de distribución arbitraria sobre la recta real. Se dice que una función de distribución  $F$  pertenece a la *familia de localización y escala basada en  $\mathcal{F}_0$*  (nos referiremos a ella como *familia LS*) si se satisface  $F(x) = \mathcal{F}_0\left(\frac{x-\theta}{\lambda}\right)$  para todo  $x \in \mathbb{R}$  y para algún parámetro de localización  $\theta \in \mathbb{R}$  y algún parámetro de escala  $\lambda > 0$ . Incluiremos esta dependencia de los parámetros en la notación como  $F_{\theta,\lambda}$ . Si utilizamos sus funciones cuantil tendríamos que

$$F_{\theta,\lambda}^{-1}(y) = \lambda\mathcal{F}_0^{-1}(y) + \theta \text{ para todo } y \in (0, 1). \quad (1.4)$$

Dadas dos funciones de distribución  $F_{\theta_1,\lambda_1}$  y  $F_{\theta_2,\lambda_2}$  pertenecientes a la misma familia LS basada en  $\mathcal{F}_0$ , la relación de orden estocástico  $F_{\theta_1,\lambda_1} \leq_{st} F_{\theta_2,\lambda_2}$  se escribiría como

$$(\lambda_1 - \lambda_2)\mathcal{F}_0^{-1}(y) \leq \theta_2 - \theta_1 \text{ para todo } y \in (0, 1). \quad (1.5)$$

Si suponemos que  $\mathcal{F}_0$  es continua y estrictamente creciente (suposición que haremos de ahora en adelante en la familia LS), entonces se tiene 1.5 si y solo si  $\lambda_1 = \lambda_2$  y  $\theta_1 \leq \theta_2$ . Además dos funciones cuantil de la misma familia LS tienen un punto de corte  $y_0$  si y solo si  $\lambda_1 \neq \lambda_2$  y  $\mathcal{F}_0^{-1}(y_0) = \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$ . Por la condición de ser estrictamente creciente deducimos que si ese punto de corte existe, entonces es único. De todo lo anterior (que puede encontrarse en el artículo de Álvarez-Esteban et al. [2]) se sigue el siguiente teorema:

**Teorema 1.2.1.** Sean  $F_{\theta_1,\lambda_1}$  y  $F_{\theta_2,\lambda_2}$  dos funciones de distribución estrictamente crecientes pertenecientes a la misma familia de localización y escala con parámetros de localización  $\theta_1, \theta_2$  y parámetros de escala  $\lambda_1, \lambda_2$ . Si

$$\begin{cases} \lambda_1 = \lambda_2 \\ y \\ \theta_1 \leq \theta_2 \end{cases}$$

entonces  $F_{\theta_1,\lambda_1} \leq_{st} F_{\theta_2,\lambda_2}$ .

En el caso en que los dos primeros momentos sean finitos, la media se suele utilizar como parámetro de localización y la desviación típica como parámetro de escala. Como ejemplo de aplicación, a continuación vamos a ver algunos casos concretos con distribuciones de uso habitual en estadística, como son la distribución normal y la uniforme continua.

**Ejemplo 1.2.2 (Orden estocástico usual en la distribución normal).** La distribución normal  $N(\mu, \sigma)$  es el caso más conocido de familia de localización y escala, donde  $\mu$  (la media) representa el

parámetro de localización y  $\sigma$  (la desviación típica) representa el parámetro de escala, ambos finitos. Dadas dos variables  $X \sim N(\mu_X, \sigma_X)$  e  $Y \sim N(\mu_Y, \sigma_Y)$  podemos utilizar las condiciones suficientes del corolario anterior para determinar la existencia de orden estocástico  $X \leq_{st} Y$ : ha de ser  $\sigma_X = \sigma_Y$  y  $\mu_X \leq \mu_Y$ .

De forma alternativa también podemos llegar a este mismo resultado si utilizamos las funciones de distribución, donde en este caso la función de distribución  $\mathcal{F}_0$  se toma como la asociada a la normal estándar  $N(0, 1)$ , y que se suele denotar como  $\phi$ . Dadas variables  $X \sim N(\mu_X, \sigma_X)$  e  $Y \sim N(\mu_Y, \sigma_Y)$ , queremos ver qué condiciones hemos de imponer sobre estos valores para determinar la existencia de orden estocástico  $X \leq_{st} Y$ . Si estandarizamos sucede que  $Z = \frac{X - \mu_X}{\sigma_X} = \frac{Y - \mu_Y}{\sigma_Y} \sim N(0, 1)$ .

En el caso de la distribución normal, las funciones de distribución son continuas y estrictamente crecientes. Supongamos que existe un valor  $x_0$  de modo que

$$\frac{x_0 - \mu_X}{\sigma_X} = \frac{x_0 - \mu_Y}{\sigma_Y} \equiv z_0. \quad (1.6)$$

Entonces  $x_0 = \mu_X + \sigma_X z_0 = \mu_Y + \sigma_Y z_0$  y  $F(x_0) = G(x_0) = \phi(z_0)$ , luego  $x_0$  sería el punto de corte de ambas en el caso de que exista. La ecuación 1.6 es lineal en  $x_0$  por lo tanto hay **como mucho** un único punto de corte  $x_0$ . Si resolvemos la ecuación para  $x_0$ , obtenemos su expresión:

$$x_0 = \frac{\mu_X \sigma_Y - \mu_Y \sigma_X}{\sigma_Y - \sigma_X}.$$

De esto, podemos deducir que **dos funciones de distribución pertenecientes a la familia de localización y escala  $N(\mu, \sigma)$  se cortan en como mucho un punto**. Además si  $\sigma_X = \sigma_Y$ , es decir, si ambas variables tienen la misma desviación típica, las dos funciones nunca se cruzan. En este caso, deberemos de pedir que  $\mu_X \leq \mu_Y$  para conseguir que la función de distribución de  $X$  quede por encima de la de  $Y$ . De este modo obtenemos que  $X \leq_{st} Y$ . Tenemos el siguiente corolario:

**Corolario 1.2.3.** Sean  $X \sim N(\mu_X, \sigma_X)$  e  $Y \sim N(\mu_Y, \sigma_Y)$  dos variables aleatorias normales. Entonces  $X \leq_{st} Y$  si y solo si  $\sigma_X = \sigma_Y$  y  $\mu_X \leq \mu_Y$ .

Estudiemos algunos ejemplos: en la figura 1.2 tenemos varias comparaciones de las funciones de distribución de distribuciones normales con distintos valores para las medias y las desviaciones típicas. Para que se produzca orden estocástico en el sentido usual se ha de producir la dominación de una de las funciones de distribución por la otra, es decir, gráficamente no se han de cortar y una ha de quedar por encima de la otra. En este caso tenemos condiciones suficientes para que se produzca orden estocástico:

- En los gráficos (a) y (b) las desviaciones típicas son distintas, por lo tanto no puede producirse orden estocástico. Podemos observar en la figura que en ambos casos se produce un corte entre las funciones de distribución, de hecho, en lo anterior hemos dado una fórmula para calcular ese punto:

$$x_a = \frac{0 \cdot 2 - 0 \cdot 0,5}{2 - 0,5} = 0 \text{ y } x_b = \frac{-1 \cdot 3 - 1 \cdot 1}{3 - 1} = -2.$$

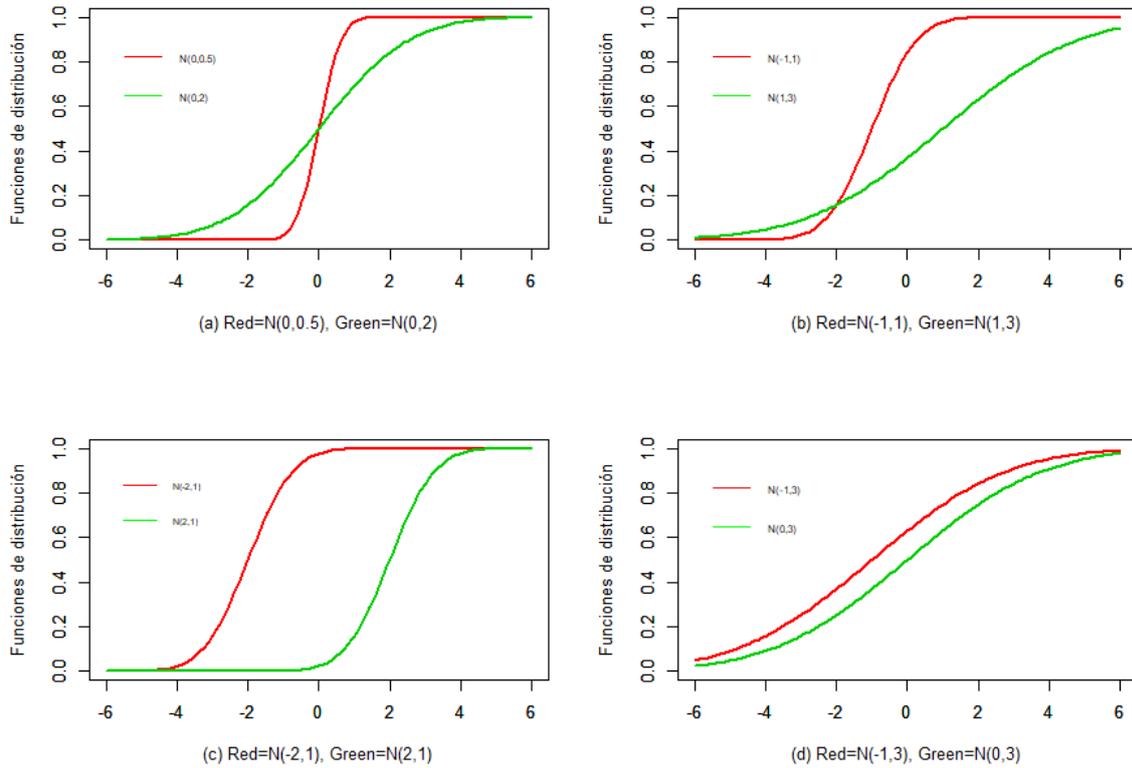


Figura 1.2: Comparación de variables normales según el orden estocástico usual

- En los gráficos (c) y (d) sí existe orden estocástico, ya que en este caso las desviaciones típicas son iguales y las medias están ordenadas de modo que  $\mu_X \leq \mu_Y$ , por lo tanto  $X \leq_{st} Y$ . Esta condición equivale a que  $F_X(t) \geq F_Y(t)$  para todo  $t$  real, y es lo que sucede en la gráfica: la función de distribución de  $X$  (en rojo) domina a la de  $Y$  (en verde) y no se producen cortes.

**Ejemplo 1.2.4 (Orden estocástico usual en la distribución uniforme).** Supongamos que  $X \sim U(a, b)$  e  $Y \sim U(c, d)$  y que tienen como funciones de distribución  $F$  y  $G$  respectivamente. La distribución uniforme continua es otro ejemplo de familia de localización y escala. Sin embargo ahora su función de distribución no es estrictamente creciente, ya que vale 0 hasta el valor  $a$  y 1 a partir del valor  $b$ . No obstante estas zonas no nos interesan para discutir la existencia del orden estocástico, ya que fuera del intervalo  $(\max\{a, c\}, \min\{b, d\})$  no hay problemas para la dominación. Para obtener  $X \leq_{st} Y$  necesitamos que los segmentos que unen  $(a, 0)$  con  $(b, 1)$  por un lado y  $(c, 0)$  con  $(d, 1)$  por el otro no tengan cortes en  $\mathbb{R} \times (0, 1)$ . Esto lo conseguimos si  $a \leq c$  y  $b \leq d$ . Por lo tanto, dos distribuciones uniformes continuas están ordenadas según el orden estocástico usual si están ordenados los extremos superior e inferior que la caracterizan, de modo que tenemos el siguiente corolario:

**Corolario 1.2.5.** Sean  $X \sim U(a, b)$  e  $Y \sim U(c, d)$  dos variables aleatorias uniformes continuas. Si  $a \leq c$  y  $b \leq d$  entonces  $X \leq_{st} Y$ .

Como podemos apreciar en la figura 1.3, en los casos (a) y (b) no se produce orden estocástico, puesto que las funciones de distribución se cortan cuando ambas crecen, y no hay una que domine

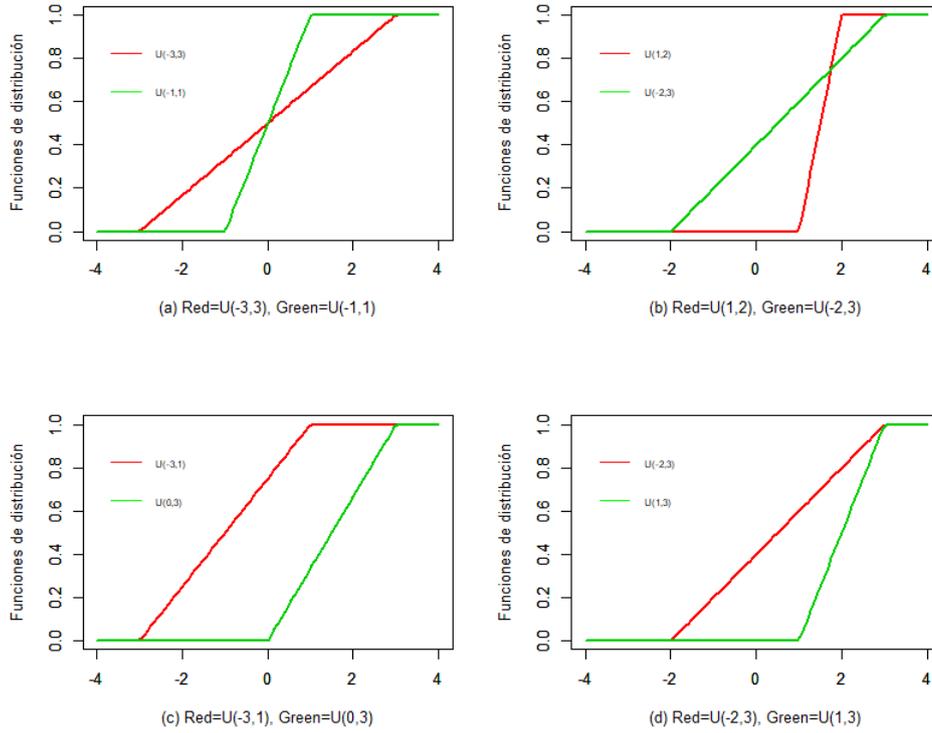


Figura 1.3: Comparación de variables uniformes continuas según el orden estocástico usual

para todo  $x \in \mathbb{R}$ . Esto sucede porque los extremos inferior y superior no están ordenados en el mismo sentido. Ocurre lo contrario en los casos (c) y (d) donde sí que existe orden estocástico, puesto que  $a \leq c$  y  $b \leq d$ .

Esta condición es muy visual si se interpreta el orden estocástico usual como la dominación entre las funciones de distribución. Solo debemos preocuparnos por lo que ocurre en los instantes de crecimiento, no queremos que estos segmentos se corten. Como ambas son funciones continuas, la única forma de que no lo hagan es que los intervalos no estén encajados, de ahí la condición de que los extremos estén ordenados en el mismo sentido.

### 1.3. Consecuencias del orden estocástico usual

Enunciamos a continuación algún resultado más sobre el orden  $\leq_{st}$  de carácter teórico pero con gran importancia práctica para la realización de simulaciones. Para la siguiente proposición conviene tener en cuenta el *teorema de representación de Skorohod*, que nos permite cambiar la convergencia en ley por una convergencia casi seguro. Según este, dada una v.a.  $X$  y una sucesión  $(X_n)_n$  de v.as. tales que  $X_n \xrightarrow{d} X$ , existe un espacio de probabilidad  $(\Omega^*, \sigma^*, P^*)$  tal que existen v.as.  $X_1^*, \dots, X_n^*, \dots, X^*$  definidas sobre él de modo que  $X_n^* \stackrel{d}{=} X_n$ ,  $X^* \stackrel{d}{=} X$  y  $X_n^* \xrightarrow{c.s.} X^*$  (respecto de  $P^*$ ). El enunciado que sigue puede encontrarse en el libro de Müller y Stoyan [27].

**Proposición 1.3.1.** La relación de orden estocástico  $\leq_{st}$  es cerrada para la convergencia en distribución, esto es, si  $F_n \leq_{st} G_n$  para todo  $n \in \mathbb{N}$  y las sucesiones  $(F_n)_n$  y  $(G_n)_n$  convergen en

distribución hacia  $F$  y  $G$  entonces  $F \leq_{st} G$ .

*Demostración.* Si  $X$  es una variable aleatoria con función de distribución  $F$ , siempre sucede que  $X \stackrel{d}{=} F^{-1}$ . Por el teorema de representación de Skorohod se tiene que  $F_n^{-1} \rightarrow_{c.s.} F^{-1}$ , donde  $P^* = \ell$  (la medida de Lebesgue sobre  $(0, 1)$ ), y el análogo utilizando  $G$ . Por hipótesis  $F_n \leq_{st} G_n$  para todo  $n$ , lo que equivale por la caracterización cuantil a que  $F_n^{-1}(p) \leq G_n^{-1}(p)$  para todo  $p \in (0, 1)$  (esto es,  $\ell(\{F_n^{-1} \leq G_n^{-1}\}) = 1$ ). Utilizando la convergencia casi seguro se tiene que  $F^{-1}(p) \leq G^{-1}(p)$  para casi todo  $p \in (0, 1)$ , es decir,  $\ell(\{F^{-1} \leq G^{-1}\}) = 1$  (o que  $F^{-1} \leq G^{-1}$  c.s.). Si utilizamos la propiedad de representación del orden estocástico que aparece en el apartado 2 del teorema 1.1.2, como  $X \stackrel{d}{=} F^{-1}$  e  $Y \stackrel{d}{=} G^{-1}$  lo anterior equivale a que  $X \leq_{st} Y$ .  $\square$

La caracterización del orden estocástico mediante funciones crecientes admite también la siguiente generalización:

**Proposición 1.3.2.** Sean  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$  variables aleatorias independientes tales que  $X_i \leq_{st} Y_i$  para  $i = 1, \dots, n$  y supongamos que  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  es una función creciente. Entonces

$$\psi(X_1, \dots, X_n) \leq_{st} \psi(Y_1, \dots, Y_n).$$

*Demostración.* Para probarlo utilizaremos el Teorema 1.1.2 (la equivalencia entre 1 y 4) y procederemos por inducción.

Para  $n = 1$ , hay que demostrar que  $Ef(\psi(X_1)) \leq Ef(\psi(Y_1))$  para toda función  $f$  creciente. Esto es cierto porque la composición de funciones crecientes, como lo son  $f$  y  $\psi$ , es creciente.

Supongamos que lo anterior es cierto para  $n - 1$  y veamos que se cumple para  $n$ . Definimos las funciones  $g(x) = Ef(\psi(X_1, \dots, X_{n-1}, x))$  y  $h(x) = Ef(\psi(Y_1, \dots, Y_{n-1}, x))$  para todo  $x$  real. Por la hipótesis de inducción, para cada  $x \in \mathbb{R}$  se tiene que  $g(x) \leq h(x)$ , y además  $g$  y  $h$  son funciones crecientes pues así lo son  $f$  y  $\psi$ . Por tanto,

$$Ef(\psi(X_1, \dots, X_n)) = Eg(X_n) \leq Eh(X_n) \leq Eh(Y_n) = Ef(\psi(Y_1, \dots, Y_n)),$$

y aplicando ahora la equivalencia de 1.1.2 pero a la inversa, se tiene que  $\psi(X_1, \dots, X_n) \leq_{st} \psi(Y_1, \dots, Y_n)$ .  $\square$

Este teorema nos permite obtener consecuencias como las siguientes, tomadas del libro de Levy [23]. En él se pueden consultar también algunas aplicaciones de estos resultados a la economía y a la teoría de riesgos.

**Corolario 1.3.3.** Sean  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$  variables aleatorias independientes tales que  $X_i \leq_{st} Y_i$  para  $i = 1, \dots, n$ . Entonces:

1.

$$\sum_{i=1}^n X_i \leq_{st} \sum_{i=1}^n Y_i,$$

2.

$$\max_n X_i \leq_{st} \max_n Y_i,$$

3.

$$\min_n X_i \leq_{st} \min_n Y_i.$$

*Demostración.* Basta considerar las aplicaciones  $\psi_1(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ ,  $\psi_2(x_1, \dots, x_n) = \max_n x_i$  y  $\psi_3(x_1, \dots, x_n) = \min_n x_i$  que son crecientes, y aplicar el teorema precedente.  $\square$

A partir de lo anterior, utilizando varias veces el resultado sobre el mínimo, se tiene el siguiente resultado para los estadísticos de orden:

**Corolario 1.3.4.** Sean  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$  variables aleatorias independientes tales que  $X_i \leq_{st} Y_i$  para  $i = 1, \dots, n$ . Sean  $X_{(1)} \leq \dots \leq X_{(n)}$  y  $Y_{(1)} \leq \dots \leq Y_{(n)}$  los estadísticos ordenados correspondientes a las variables anteriores. Entonces

$$X_{(i)} \leq_{st} Y_{(i)} \text{ para cada } i = 1, \dots, n.$$

## 1.4. Condiciones suficientes

En esta sección vamos a estudiar dos nuevos conceptos de orden que nos servirán como condición suficiente para demostrar la existencia del orden estocástico usual. La bibliografía básica que se ha utilizado han sido el libro de Belzunce, Martínez-Riquelme y Mulero [7] y el de Müller y Stoyan [27].

### 1.4.1. Orden del cociente de riesgo

Consideremos el siguiente ejemplo modificado del libro de Müller y Stoyan [27]: imaginemos que una persona quiere comprar un móvil y que puede elegir entre un iPhone o un Samsung con tiempos de vida  $X$  e  $Y$  respectivamente. Si el precio de ambos es el mismo y sucede que  $X \leq_{st} Y$  (esto es,  $P(X > t) \leq P(Y > t)$ , luego es más probable que  $Y$  funcione durante más tiempo que  $X$ ), parece evidente que el consumidor comprará el teléfono Samsung, asociado a  $Y$ . Sin embargo, supongamos ahora que quiere comprarse un teléfono como los anteriores reutilizado con un año de antigüedad, donde ahora los tiempos de vida son  $X'$  e  $Y'$  de modo que  $P(X' > t) = P(X > 1 + t | X > 1)$  y el análogo para  $Y'$ . ¿Sigue siendo el teléfono Samsung mejor que el iPhone, esto es,  $X' \leq_{st} Y'$ ? No necesariamente. Veamos un ejemplo:

**Ejemplo 1.4.1** ( $X \leq_{st} Y$  no se conserva con la edad). Si el tiempo de vida del iPhone,  $X$ , se distribuye de manera uniforme en  $(0, 5)$  e  $Y$  tiene densidad

$$g(x) = \begin{cases} \frac{1}{10} & \text{si } 0 < x \leq 1 \\ \frac{7}{10} & \text{si } 1 < x \leq 3 \\ \frac{1}{5} & \text{si } 3 < x < 5 \end{cases}$$

entonces se tiene que  $X \leq_{st} Y$  (puede verse la comparación de las funciones de distribución en el gráfico a la izquierda en la figura 1.4, ambas coinciden para  $t > 3$ ). Respecto a los tiempos de vida

pasado un año,  $X'$  se distribuye de manera uniforme en  $(0, 4)$ , e  $Y'$  tiene densidad

$$g'(x) = \begin{cases} \frac{7}{9} & \text{si } 0 < x \leq 2 \\ \frac{2}{9} & \text{si } 2 < x < 4 \end{cases}$$

pues ahora, como mucho, el móvil puede durar cuatro años. En esta ocasión, se tiene que  $X' \geq_{st} Y'$  (puede verse la comparación de sus funciones de distribución en el gráfico a la derecha en la figura 1.4), es decir, el iPhone tiene mayor probabilidad de durar más tiempo que el Samsung. Luego el orden estocástico no se conserva con el envejecimiento.

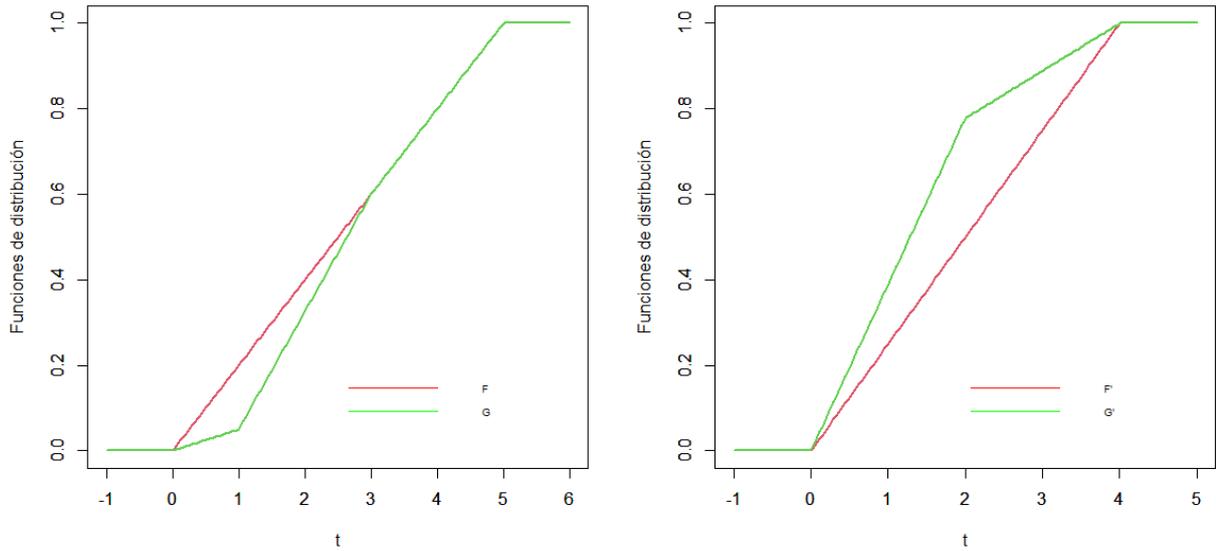


Figura 1.4: Comparación de tiempos de vida sin envejecimiento y con envejecimiento

Podemos pensar en qué condiciones hacen falta imponer para que el orden estocástico usual se mantenga para variables que llevan en funcionamiento un tiempo  $t$ , esto es, cuándo

$$[X|X > t] \leq_{st} [Y|Y > t] \text{ para todo } t.$$

Utilizando la definición de  $\leq_{st}$  lo anterior se puede escribir como

$$P(X > s + t | X > t) \leq P(Y > s + t | Y > t) \text{ para todo } s \geq 0 \text{ y para todo } t,$$

esto es, si  $X$  e  $Y$  tiene funciones de distribución  $F$  y  $G$  respectivamente,

$$\frac{\overline{F}(s+t)}{\overline{F}(t)} \leq \frac{\overline{G}(s+t)}{\overline{G}(t)} \text{ para todo } s \geq 0 \text{ y para todo } t.$$

o, de manera equivalente,

$$\frac{\overline{G}(t)}{\overline{F}(t)} \leq \frac{\overline{G}(s+t)}{\overline{F}(s+t)} \text{ para todo } s \geq 0 \text{ y para todo } t.$$

Tenemos la siguiente definición:

**Definición 1.4.2.** Dadas dos variables aleatorias  $X$  e  $Y$ , se dice que  $X$  es menor que  $Y$  respecto al orden del cociente de riesgo (escrito  $X \leq_{hr} Y$ , del inglés *hazard rate order*) si la función

$$t \mapsto \frac{\bar{G}(t)}{\bar{F}(t)}$$

es creciente.

En algunos textos se hace también uso de la siguiente definición, que involucra a las funciones de distribución en vez de a las funciones de supervivencia:

**Definición 1.4.3.** Dadas dos variables aleatorias  $X$  e  $Y$ , se dice que  $X$  es menor que  $Y$  respecto al orden del cociente de riesgo inverso (escrito  $X \leq_{rh} Y$ ) si la función

$$t \mapsto \frac{G(t)}{F(t)}$$

es creciente.

Se da este nombre a este orden porque, asumiendo que las variables son absolutamente continuas (luego tienen densidad), existe una caracterización equivalente en términos de sus *cocientes de riesgo*, también llamados en algunos libros *tasas de fallo*. El cociente de riesgo  $r_X(t)$  es la tasa infinitesimal de fallo para un tiempo  $t$ , esto es

$$r_X(t) = \lim_{h \rightarrow 0} \frac{P(X \leq t+h | X > t)}{h} = \frac{f(t)}{\bar{F}(t)} = -\frac{d}{dt} \ln(\bar{F}(t))$$

siempre que  $\bar{F}(t) > 0$ , donde  $f$  es la función de densidad de  $X$ . Cuanto mayor sea esta tasa de fallo, menor será  $X$  en el sentido estocástico. De hecho, si  $X$  e  $Y$  son variables aleatorias absolutamente continuas, entonces  $X \leq_{hr} Y$  es equivalente a  $r_X(t) \geq r_Y(t)$  para todo  $t$  real, como se puede ver en el libro de Müller y Stoyan [27].

El orden  $\leq_{hr}$  es una condición más fuerte que el orden  $\leq_{st}$ , como muestra la siguiente proposición:

**Proposición 1.4.4.**  $X \leq_{hr} Y$  implica  $X \leq_{st} Y$ .

*Demostración.*  $X \leq_{hr} Y$  equivale a que  $\bar{G}(s)\bar{F}(t) \leq \bar{F}(s)\bar{G}(t)$  para todo  $s \leq t$ . Las funciones de supervivencia  $\bar{G}(s)$ ,  $\bar{F}(s)$  convergen a 1 si  $s \rightarrow -\infty$ , por lo tanto basta tomar este límite en la desigualdad anterior para obtener que  $\bar{F}(t) \leq \bar{G}(t)$  para todo  $t$  real, esto es,  $X \leq_{st} Y$ .  $\square$

Un ejemplo fácil de aplicación sería entre distribuciones exponenciales, que gozan de la propiedad de pérdida de memoria, de modo que si  $\lambda \leq \mu$  se tiene que  $\varepsilon(\mu) \leq_{hr} \varepsilon(\lambda)$ , lo que implica  $\varepsilon(\mu) \leq_{st} \varepsilon(\lambda)$  como ya vimos.

El orden del cociente de riesgo tiene también interés por si mismo, por ejemplo para probar la fiabilidad de sistemas como se muestra en el libro de Shaked y Shanthikumar [30]. Si  $X_1, \dots, X_n$ ,

$\Lambda_1, \Lambda_2$  son v.a. no negativas e independientes y se considera

$$T_{j,i} = \frac{X_i}{\Lambda_j}, \quad i = 1, \dots, n, \quad j = 1, 2,$$

cada uno de estos valores  $T_{1,i}$  y  $T_{2,i}$  pueden entenderse como los tiempos de vida de  $n$  componentes independientes que trabajan sometidas a una presión  $\Lambda_j$  aleatoria que afecta a la escala de los tiempos de vida originales de las componentes. Si ahora tomamos

$$N_j(t) = \sum_{i=1}^n I_{\{T_{j,i} > t\}}(t), \quad j = 1, 2, \quad t \geq 0,$$

entonces  $N_j(t)$  es el número de componentes que siguen funcionando en un tiempo  $t$  para cada uno de los sistemas  $j = 1, 2$ . En el libro de Shaked y Shanthikumar [30] se prueba que

$$\Lambda_1 \leq_{rh} \Lambda_2 \Rightarrow N_1(t) \geq_{hr} N_2(t) \Rightarrow N_1(t) \geq_{st} N_2(t) \text{ para todo } t \geq 0,$$

lo que quiere decir que si se produce tal dominación entre las tasas de presión  $\Lambda_j$  para el orden del cociente de riesgo inverso, entonces las componentes del sistema 1 tienen mayor duración que las componentes del sistema 2.

#### 1.4.2. Orden de razón de verosimilitud

El orden de razón de verosimilitud tiene también relevancia por ser una condición suficiente para la existencia del orden estocástico usual. En primer lugar daremos la definición para variables con densidad, aunque esta se puede extender a distribuciones generales como veremos.

**Definición 1.4.5.** Sean  $X$  e  $Y$  variables aleatorias. Se dice que  $X$  es menor que  $Y$  respecto al orden de razón de verosimilitud (escrito  $X \leq_{lr} Y$ , del inglés *likelihood ratio order*) si  $X$  e  $Y$  tienen densidades  $f$  y  $g$  respecto a una medida dominante  $\mu$  de modo que

$$f(t)g(s) \leq f(s)g(t) \text{ para todo } s \leq t. \quad (1.7)$$

**Observación 1.4.6.** La ventaja que tiene esta definición respecto a otras menos generales consiste en que de este modo se incluyen los casos de variables aleatorias absolutamente continuas (donde  $\mu$  es la medida de Lebesgue) el caso de distribuciones discretas (donde  $\mu$  es una medida de conteo) y otros casos como mixturas de ambas.

La ecuación que define este orden consiste básicamente en afirmar que el cociente  $\frac{g(t)}{f(t)}$  es creciente para todo  $t$ , pero escrito en la forma anterior se evita el problema de que el denominador se anule (algo que no pasaba con las funciones de distribución).

Si integramos a ambos lados en 1.7 con respecto a  $s \in A$  y  $t \in B$  donde  $A, B \in \beta$  entonces la definición es equivalente (por el teorema de Fubini en el caso continuo) a

$$P(X \in B)P(Y \in A) \leq P(X \in A)P(Y \in B)$$

para todos aquellos conjuntos medibles  $A$  y  $B$  tales que  $A \leq B$  (esto quiere decir que si  $s \in A$  y  $t \in B$  entonces  $s \leq t$ ). Esta definición no depende de las densidades, luego se podría considerar como definición del orden  $\leq_{lr}$  para cualquier distribución de probabilidad sobre  $\mathbb{R}$ .

Utilizando lo anterior se deduce la siguiente propiedad, la cual se asemeja a la que utilizamos para dar significado al orden  $\leq_{hr}$ :

**Proposición 1.4.7.**  $X \leq_{lr} Y$  si y solo si  $[X|X \in B] \leq_{st} [Y|Y \in B]$  para todos los sucesos  $B \in \beta$  tales que  $P(X \in B) > 0$  y  $P(Y \in B) > 0$ .

*Demostración.*  $[X|X \in B] \leq_{st} [Y|Y \in B]$  para todo  $B \in \beta$  quiere decir que

$$P(X > t|X \in B) \leq P(Y > t|Y \in B)$$

para todo  $t$  real. Si  $P(X \in B) > 0$  y  $P(Y \in B) > 0$ , utilizando la definición de probabilidad condicionada se tiene que

$$\frac{P(X \in (t, \infty) \cap B)}{P(X \in B)} \leq \frac{P(Y \in (t, \infty) \cap B)}{P(Y \in B)}.$$

Tomando  $A := (t, \infty) \cap B$  (lo que implica  $A \leq B$ ) se tiene que

$$P(X \in B)P(Y \in A) \leq P(X \in A)P(Y \in B),$$

esto es, que  $X \leq_{lr} Y$ . Para el recíproco basta dar la vuelta al argumento anterior y considerar la probabilidad condicionada, observando que todo conjunto  $A$  que verifica  $A \leq B$  puede escribirse como  $A = (t, \infty) \cap B$  para algún valor  $t$ .  $\square$

La importancia y utilidad de este orden radica en que en la práctica es fácil de comprobar (es ver que una función es creciente), y es una condición suficiente para que exista orden  $\leq_{st}$ .

**Proposición 1.4.8.**  $X \leq_{lr} Y \implies X \leq_{hr} Y \implies X \leq_{st} Y$ .

*Demostración.* Para probarlo vamos a utilizar su caracterización a través de las variables condicionadas. Observemos que  $[X|X \in A] \leq_{st} [Y|Y \in A]$  se verifica para todo  $A \in \beta$  si  $X \leq_{lr} Y$ , pero en particular para los conjuntos de la forma  $A = (-\infty, t]$  (lo que equivale al orden  $X \leq_{hr} Y$ ), y todavía más en particular para  $A = (-\infty, +\infty)$ , esto es, el orden  $X \leq_{st} Y$ .  $\square$

El orden  $\leq_{lr}$  en general no tiene relevancia por si mismo, se utiliza como herramienta para determinar si existe orden  $\leq_{st}$  entre distribuciones. Este nos permite determinar fácilmente qué condiciones debemos imponer a los parámetros de las distribuciones más utilizadas en estadística para que exista orden  $\leq_{st}$ .

**Ejemplo 1.4.9.** Si tenemos dos distribuciones exponenciales  $X \sim \varepsilon(\lambda)$  e  $Y \sim \varepsilon(\mu)$ , para ver si hay orden  $\leq_{lr}$  debemos analizar el cociente de sus densidades, que son de la forma  $f(t) = \lambda e^{-\lambda t}$ . Luego al dividir ambas debemos estudiar la función

$$\frac{g(t)}{f(t)} = \frac{\mu}{\lambda} e^{-(\mu-\lambda)t},$$

la cual es creciente si y solo si  $\lambda - \mu \geq 0$ , esto es,  $\lambda \geq \mu$ . De este modo se tiene que  $X \leq_{lr} Y$  y por tanto  $X \leq_{st} Y$ . Como podemos ver, esto coincide con lo que estudiamos en la sección 1 sobre el orden estocástico usual, concretamente en la proposición 1.1.4.

Lo anterior no solo se puede hacer en el caso continuo, sino que también se puede probar con distribuciones discretas utilizando la función de masa de probabilidad:

**Ejemplo 1.4.10.** Supongamos que tenemos dos distribuciones binomiales  $X \sim B(m, p)$  e  $Y \sim B(n, q)$ . Para estudiar si existe orden  $\leq_{lr}$  debemos considerar el cociente de sus funciones de masa de probabilidad  $p_Y$  entre  $p_X$  y ver que es creciente, o de forma equivalente, ver que

$$p_X(t)p_Y(s) \leq p_X(s)p_Y(t) \text{ si } s \leq t, \text{ } s, t \in \mathbb{Z}_+.$$

Sustituyendo sus valores, hay que ver

$$\binom{m}{t} p^t (1-p)^{m-t} \binom{n}{s} q^s (1-q)^{n-s} \leq \binom{m}{s} p^s (1-p)^{m-s} \binom{n}{t} q^t (1-q)^{n-t}, \text{ } s \leq t.$$

Si operamos y agrupamos, esto es equivalente a

$$\frac{(m-s)!}{(m-t)!} \left( \frac{p}{1-p} \right)^{t-s} \leq \frac{(n-s)!}{(n-t)!} \left( \frac{q}{1-q} \right)^{t-s}, \text{ } s \leq t.$$

Lo anterior es cierto si  $m \leq n$  y  $p \leq q$ , por lo tanto si se dan estas condiciones se puede asegurar que  $X \leq_{lr} Y$  y en consecuencia  $X \leq_{st} Y$ .

## 1.5. Orden estocástico usual multivariante

Como ya hemos visto, el orden estocástico usual  $X \leq_{st} Y$  se puede caracterizar como

1.  $Ef(X) \leq Ef(Y)$  para toda función  $f$  creciente;
2.  $\bar{F}_X(t) \leq \bar{F}_Y(t)$  para todo  $t$  real;
3.  $F_X(t) \geq F_Y(t)$  para todo  $t$  real.

Sin embargo, a diferencia de lo que ocurre en el caso unidimensional, ahora no es cierta la igualdad  $F_{\mathbf{X}}(\mathbf{x}) = 1 - \bar{F}_{\mathbf{X}}(\mathbf{x})$ . Por ejemplo en el caso bivalente se tiene que, para todos  $x_1, x_2$ ,

$$\bar{F}_{\mathbf{X}}(x_1, x_2) = 1 - F_1(x_1) - F_2(x_2) + F_{\mathbf{X}}(x_1, x_2).$$

Por lo tanto en el caso multivariante estas tres condiciones dejan de ser equivalentes. Esto da sentido a la siguiente definición:

**Definición 1.5.1.** Sean  $\mathbf{X}$  e  $\mathbf{Y}$  vectores aleatorios con valores en  $\mathbb{R}^n$ . Se dice que  $\mathbf{X}$  es menor que  $\mathbf{Y}$  respecto a:

1. orden estocástico usual (escrito  $\mathbf{X} \leq_{st} \mathbf{Y}$ ) si  $Ef(\mathbf{X}) \leq Ef(\mathbf{Y})$  para toda función creciente y acotada  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ;

2. *orden del octante superior* (escrito  $\mathbf{X} \leq_{uo} \mathbf{Y}$ ) si  $\bar{F}_{\mathbf{X}}(\mathbf{t}) \leq \bar{F}_{\mathbf{Y}}(\mathbf{t})$  para todo  $\mathbf{t}$ ;

3. *orden del octante inferior* (escrito  $\mathbf{X} \leq_{lo} \mathbf{Y}$ ) si  $F_{\mathbf{X}}(\mathbf{t}) \geq F_{\mathbf{Y}}(\mathbf{t})$  para todo  $\mathbf{t}$ .

El orden del octante superior está generado por las funciones indicatrices  $f(\mathbf{x}) = I_{(\mathbf{t}, \infty)}(\mathbf{x})$ ,  $\mathbf{t} \in \mathbb{R}^n$  y el orden del octante inferior está generado por las funciones  $f(\mathbf{x}) = -I_{(-\infty, \mathbf{t}]}(\mathbf{x})$ ,  $\mathbf{t} \in \mathbb{R}^n$ , donde los conjuntos  $(\mathbf{t}, \infty)$  y  $(-\infty, \mathbf{t}]$  se deben entender como los productos de intervalos  $(t_1, \infty) \times \dots \times (t_n, \infty)$  y  $(-\infty, t_1] \times \dots \times (-\infty, t_n]$  (hiperoctantes). Como estas funciones son crecientes, se tiene el siguiente resultado:

**Proposición 1.5.2.** Si  $\mathbf{X} \leq_{st} \mathbf{Y}$  entonces  $\mathbf{X} \leq_{uo} \mathbf{Y}$  y  $\mathbf{X} \leq_{lo} \mathbf{Y}$ .

El orden  $\leq_{st}$  multivariante se puede caracterizar por lo que se llaman *conjuntos superiores*, que son subconjuntos  $U \subset \mathbb{R}^n$  tales que si  $\mathbf{x} \in U$  y si  $\mathbf{x} \leq \mathbf{y}$  entonces  $\mathbf{y} \in U$ .

**Teorema 1.5.3.** Son equivalentes:

1.  $\mathbf{X} \leq_{st} \mathbf{Y}$ ;
2.  $Ef(\mathbf{X}) \leq Ef(\mathbf{Y})$  para toda función  $f$  creciente y continua;
3.  $P(\mathbf{X} \in U) \leq P(\mathbf{Y} \in U)$  para todo conjunto superior  $U$ ;
4.  $P(\mathbf{X} \in U) \leq P(\mathbf{Y} \in U)$  para todo conjunto superior  $U$  cerrado.

La demostración de este hecho puede encontrarse en el libro de Müller y Stoyan [27]. La parte más interesante es la demostración de (3)  $\Rightarrow$  (1), donde se construye la función  $f$  para asegurar que  $\mathbf{X} \leq_{st} \mathbf{Y}$ .

Muchos de los resultados del caso univariante también son ciertos para el caso multivariante, como es el teorema de representación (que nos permite tratar el orden  $\leq_{st}$  por un orden c.s.) o que el orden se conserva por transformaciones crecientes. Como curiosidad se tiene la siguiente proposición, tomada del artículo de Müller [26]:

**Proposición 1.5.4.** Sean  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $\mathbf{X}' \sim N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$  vectores aleatorios  $n$ -dimensionales normalmente distribuidos. Entonces  $\mathbf{X} \leq_{st} \mathbf{X}'$  si y solo si  $\boldsymbol{\mu} \leq \boldsymbol{\mu}'$  y  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$ .

*Demostración.* La condición suficiente es inmediata a partir del teorema de representación por un orden c.s., puesto que  $\mathbf{X}'$  tiene la misma distribución que  $\mathbf{X} + \boldsymbol{\mu}' - \boldsymbol{\mu}$ . Supongamos que  $\mathbf{X} \leq_{st} \mathbf{X}'$ , lo que implica que  $X_i \leq_{st} X'_i$  para todas las marginales (basta tomar la función  $f(\mathbf{x}) = x_i$  en la definición del orden  $\leq_{st}$  multivariante, que es creciente). Utilizando la caracterización del orden usual para las normales univariantes, lo anterior sucede si y solo si  $\mu_i \leq \mu'_i$  y  $\sigma_{i,i} = \sigma'_{i,i}$ , luego ha de ser  $\boldsymbol{\mu} \leq \boldsymbol{\mu}'$ . Por otro lado, toda combinación lineal de normales tiene distribución normal, de modo que  $X_i + X_j \sim N(\mu_i + \mu_j, \sigma_{i,i}^2 + \sigma_{j,j}^2 + 2\sigma_{i,j})$ . Se tiene entonces que  $X_i + X_j \leq_{st} X'_i + X'_j$  para  $1 \leq i < j \leq n$ . Las condiciones univariantes nos dicen que las desviaciones típicas han de ser iguales, esto es,  $\sigma_{i,i}^2 + \sigma_{j,j}^2 + 2\sigma_{i,j} = \sigma_{i,i}^{\prime 2} + \sigma_{j,j}^{\prime 2} + 2\sigma'_{i,j}$ . Utilizando que  $\sigma_{i,i} = \sigma'_{i,i}$  para todo  $i$  se deduce que también  $\sigma_{i,j} = \sigma'_{i,j}$  para todos  $i < j$ , y por ser  $\boldsymbol{\Sigma}$  simétrica se cumple para todos, luego  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$ .  $\square$



## Capítulo 2

# Aproximaciones al orden estocástico usual

Aunque el orden estocástico es un instrumento muy útil para realizar comparaciones entre distribuciones de probabilidad, la definición que hemos estudiado establece una relación 0 – 1, es decir, que o bien existe o bien no. El siguiente ejemplo nos muestra que, a pesar de lo anterior, el orden estocástico nos aporta más información sobre las distribuciones que otros tipos de comparaciones más simples como el contraste de medias.

**Ejemplo 2.0.1.** Consideremos dos variables aleatorias  $X$  e  $Y$  con distribución de Weibull de modo que  $X \sim W(1, 2)$  e  $Y \sim W\left(\frac{\sqrt{\pi}}{2}, 1\right)$ , muy utilizadas en el análisis de supervivencia (si  $Z \sim W(\lambda, k)$ , a  $\lambda$  se le llama parámetro de escala y a  $k$  parámetro de forma de la distribución). Las funciones de distribución de  $X$  e  $Y$  son, respectivamente,  $F(x) = 1 - \exp(-x^2)$  y  $G(x) = 1 - \exp\left(-\frac{2}{\sqrt{\pi}}x\right)$ .

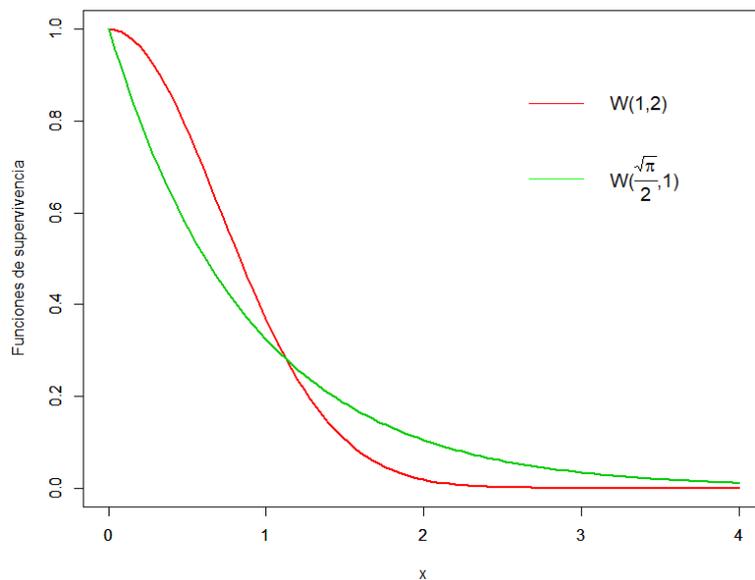


Figura 2.1: Comparación de funciones de supervivencia de variables con distribución de Weibull

En primer lugar, calculemos sus esperanzas, que vienen dadas como  $EZ = \lambda \Gamma\left(1 + \frac{1}{k}\right)$  donde  $\Gamma$  es la *función gamma* de Euler. Así  $EX = \Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$  y  $EY = \frac{\sqrt{\pi}}{2} \Gamma(1) = \frac{\sqrt{\pi}}{2}$ . Observamos por tanto que  $EX = EY$ .

Si representamos las funciones de supervivencia podemos observar que no existe orden estocástico, pues no se produce la condición de dominación y se produce un corte entre ambas en el punto  $x_0 = \frac{2}{\sqrt{\pi}} \approx 1'128$ .

Sin embargo, si por ejemplo  $X$  e  $Y$  representan el tiempo de vida aleatorio de dos dispositivos, o los tiempos de supervivencia de unos pacientes ante dos determinados tratamientos, podemos afirmar que se espera que  $X$  e  $Y$  tengan el mismo tiempo medio de vida, al considerar solamente los valores medios. No obstante, si nos fijamos en la probabilidad de supervivencia a partir de un tiempo fijado  $x \geq 0$ , entonces  $P(X > x) \geq P(Y > x)$  si  $x \in [0, \frac{2}{\sqrt{\pi}}]$  y  $P(X > x) \leq P(Y > x)$  para cada  $x \in [\frac{2}{\sqrt{\pi}}, +\infty)$ . No se produce orden estocástico *en términos absolutos* pero la relación que hay entre ambas distribuciones nos da más información que estudiar sus medias.

En general, la solución que se le ha dado al problema de comparar dos poblaciones (o al comparar una población sometida a dos situaciones distintas, como puede ser el antes y el después ante una perturbación) ha consistido con más frecuencia en la comparación de características de estas, como por ejemplo comparar las medias entre dos muestras extraídas de ambas o la proporción de que se dé un determinado suceso en cada una. El procedimiento general tiene el objetivo de encontrar evidencia para concluir que una característica es mayor/mejor que la otra. Sin embargo, lo que estamos tratando de ver es que se ha producido un aumento o una mejora en la población, y puede ocurrir que un aumento *en media* de una característica sea compatible con un deterioro en gran parte de la población. A diferencia de la comparación de medias, el orden estocástico tiene en cuenta las distribuciones completas. La gran desventaja que presenta es que aún considerando el orden más sencillo de los estudiados en el capítulo 1, el orden estocástico usual  $\leq_{st}$ , se trata de una suposición demasiado fuerte en algunos problemas. Puede incluso suceder que la dominación se mantenga en una gran parte del rango de las variables de interés pero que falle para otra parte, como pasaba en el ejemplo anterior. Por estas razones será conveniente efectuar una relajación del concepto, de modo que en este capítulo veremos algunas herramientas que nos permitirán cuantificar cómo de lejos estamos de la relación de orden estocástico cuando este no existe y que nos servirán posteriormente para comprobar la existencia de orden en la práctica.

## 2.1. Índices de desviación

Si echamos un vistazo a la figura 2.2 en la cual se comparan dos poblaciones normales en las que se produce un incremento en la media pero entre las que no existe orden estocástico, podríamos pensar que el grado de desviación respecto al orden es mucho mayor en la gráfica situada a la derecha que en la de la izquierda. Se podría pensar incluso que *casi* existe orden estocástico, o que en la práctica sí que se mantiene la relación de dominancia entre ambas distribuciones.

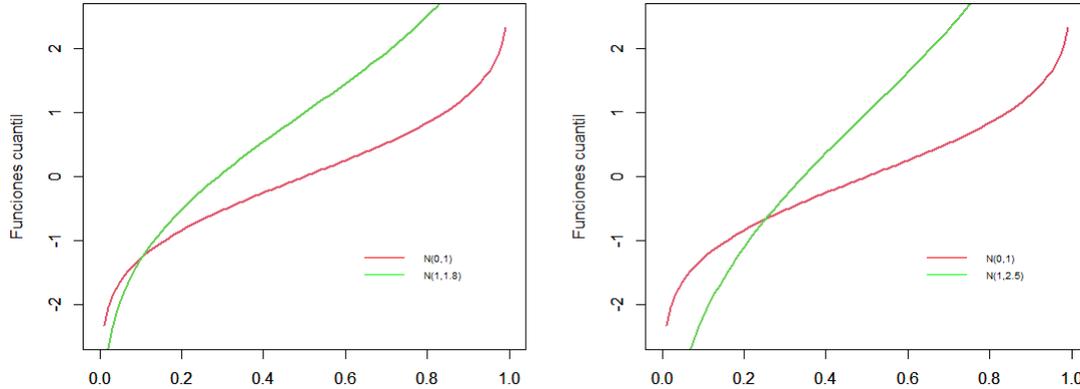


Figura 2.2: Comparación de funciones cuantil de variables normales

En lo que sigue vamos a estudiar algunos índices que nos permitirán determinar el grado de desviación respecto al concepto de orden estudiado, los cuales nos serán también útiles al realizar contrastes de hipótesis para poder determinar si dos poblaciones están ordenadas.

### 2.1.1. Precedencia estocástica y el índice $\rho(F, G)$

Dadas dos variables aleatorias  $X$  e  $Y$ , una forma sencilla de acercarse a la relación de orden estocástico  $X \leq_{st} Y$  consiste en pedir que  $X \leq Y$  con una probabilidad suficientemente grande. En el teorema 1.1.2 del capítulo 1 vimos que  $X \leq_{st} Y$  es equivalente a que existan variables  $\hat{X} =^d X$  e  $\hat{Y} =^d Y$  definidas en el mismo espacio de probabilidad de modo que  $P(\hat{X} \leq \hat{Y}) = 1$ . Por un lado, esto nos permite diseñar procedimientos que nos permiten trabajar con muestras de modo que cualquier valor obtenido a partir de la primera distribución es menor c.s. que otro obtenido de la segunda. Pero, por otro lado, también nos permite realizar aproximaciones al concepto de orden estocástico si pedimos que la probabilidad de que  $X \leq Y$  sea suficientemente grande.

En el artículo de Arcones et al. [3] se introduce un nuevo concepto de orden: si  $X$  e  $Y$  son v.a. independientes, entonces se dice que  $X$  precede estocásticamente a  $Y$  (escrito  $X \leq_{sp} Y$ ) si  $P(X \leq Y) \geq \frac{1}{2}$ . Esto podría interpretarse, por ejemplo, como que con probabilidad mayor que  $1/2$  el tratamiento  $Y$  tiene mejores resultados que el tratamiento  $X$  cuando se prueban en muestras de pacientes independientes.

Se tiene que  $X \leq_{st} Y \Rightarrow X \leq_{sp} Y$ : en efecto, si  $F$  y  $G$  son las funciones de distribución de  $X$  e  $Y$ , entonces tomando  $Z := X - Y$  tenemos que

$$\begin{aligned}
 P(X \leq Y) &= P(Z \leq 0) = \int_{-\infty}^{\infty} P(Z \leq 0 | X = x) dF(x) = \int_{-\infty}^{\infty} P(X - Y \leq 0 | X = x) dF(x) = \\
 &= \int_{-\infty}^{\infty} P(x - Y \leq 0 | X = x) dF(x) = \int_{-\infty}^{\infty} P(x \leq Y) dF(x) = \int_{-\infty}^{\infty} (1 - P(Y > x)) dF(x) = \\
 &= \int_{-\infty}^{\infty} (1 - G(x-)) dF(x) \geq \int_{-\infty}^{\infty} (1 - F(x-)) dF(x) = P(X \leq X') \geq \frac{1}{2},
 \end{aligned}$$

donde  $X'$  es una v.a.i.i.d. que  $X$ .

El orden  $\leq_{sp}$  es por tanto una relación menos restrictiva entre variables aleatorias que el orden estocástico usual, por lo que se puede considerar como una aproximación a este. Por ejemplo, para el estudio de variables aleatorias normales vimos que  $\mu_X \leq \mu_Y$  y  $\sigma_X = \sigma_Y$  eran condiciones suficientes para la existencia de orden  $\leq_{st}$ . En el caso del orden  $\leq_{sp}$  basta solamente con que  $\mu_X \leq \mu_Y$ . Esto hace que el estudio sea en muchas ocasiones más sencillo, pero la precedencia estocástica no nos aporta tanta información como el orden estocástico.

En esta misma línea, en el artículo de del Barrio et al. [6] (2018) se propone el estudio del índice  $\rho(F, G)$  aplicado a variables independientes, que se define como

$$\rho(F, G) := P(X > Y).$$

Desde nuestro punto de vista, para acercarnos al concepto de orden estocástico estaríamos buscando que  $\rho(F, G)$  fuera pequeño, para de este modo poder concluir que con una alta probabilidad  $Y$  toma valores mayores que  $X$  suponiendo que ambas son independientes. Si  $\rho(F, G) = 0$  entonces el soporte de la distribución de  $X$  y el de la de  $Y$  están separados por al menos un punto, lo que quiere decir que cualquier resultado de  $Y$  es siempre mayor que cualquier resultado de  $X$ , de lo que podemos concluir que  $X \leq_{st} Y$ . Esto supone una generalización del orden  $\leq_{sp}$  con la ventaja de que cuantifica el grado de lejanía con la relación de orden estocástico y nos ofrece información en términos de una probabilidad. De esta forma, si tomamos dos muestras de pacientes independientes a los que sometemos a dos tratamientos distintos  $X$  e  $Y$  y obtenemos un valor de  $\rho(F, G)$  pequeño, entonces podemos concluir que los resultados para el tratamiento  $X$  son peores que para el tratamiento  $Y$ . Como estamos interesados en un marco general para medir el grado de discordancia con el orden estocástico (y no solo trabajar en el caso de independencia), en lo que sigue nos centraremos en otros índices que nos permitirán desarrollar la teoría con más generalidad.

### 2.1.2. Orden estocástico restringido y el índice $\gamma(F, G)$

Berger consideró en [8] la idea del *orden estocástico restringido*, que se basa en buscar la relación  $F(x) \geq G(x)$  en un intervalo cerrado arbitrario (aunque fijado), excluyendo de este modo las colas de la distribución de la muestra. Si recordamos el ejemplo 2.0.1 relativo a la comparación de dos variables de Weibull, en él ya vimos que no existe orden estocástico en términos absolutos, pero podríamos relajar este concepto en el sentido anterior según las aplicaciones que estemos considerando restringiéndolo a intervalos concretos (por ejemplo, si estamos comparando dos tratamientos clínicos en enfermos y sabemos que la duración de sus vidas al aplicarlo es siempre mayor que un tiempo fijado  $t_0$ ).

Para estudiar el orden estocástico restringido podríamos utilizar la caracterización del orden estocástico mediante las funciones cuantiles que estudiamos en el capítulo 1, aprovechando de este modo la normalización que estas nos ofrecen al compararse en el intervalo  $(0, 1)$ , que es de medida

finita. El procedimiento consistiría en fijar un intervalo contenido en  $(0, 1)$  y verificar si se tiene que  $F^{-1}(t) \leq G^{-1}(t)$  para todos los valores  $t$  dentro de ese intervalo.

Podemos incluso evitar el hecho de tener que elegir un intervalo arbitrario en el que ver si se cumple lo anterior. Esta idea, que se propone en el artículo de Álvarez-Esteban et al. (2017) [2], consiste en medir la longitud del conjunto donde no se verifica la relación de orden estocástico, esto es, estudiar el índice

$$\gamma(F, G) := \ell\left(t \in (0, 1) : F^{-1}(t) > G^{-1}(t)\right) \quad (2.1)$$

donde  $\ell$  es la medida de Lebesgue en  $(0, 1)$ . Este índice  $\gamma \in [0, 1]$  nos permite medir la desviación respecto al orden estocástico, ya que poder afirmar  $\gamma(F, G) \leq \gamma_0$  para una cantidad  $\gamma_0$  pequeña fijada nos proporciona información sobre cómo de lejos están  $F$  y  $G$  de la relación de orden estocástico. Si observamos la definición resulta evidente que  $F \leq_{st} Y$  si y solo si  $\gamma(F, G) = 0$ .

El índice es también muy fácil de interpretar: por ejemplo, si  $F$  y  $G$  son las funciones de distribución de dos variables  $X$  e  $Y$  que representan el tiempo de vida ante dos determinados tratamientos, poder afirmar que  $\gamma(F, G) = 0,05$  significaría que el 95 % de los pacientes consiguen mejores resultados con el tratamiento asociado a  $Y$  que con el de  $X$ . En la sección 2.2 lo aplicaremos a las familias de localización y escala y discutiremos algunas propiedades adicionales sobre él.

En el caso de que tengamos muestras de dos funciones  $F$  y  $G$  para las que podemos asegurar que  $\ell(\{t : F^{-1}(t) > G^{-1}(t)\}) = 0,05$ , esto significaría que el procedimiento que genera los datos de ambas producirá una proporción del 5 % de valores de  $G$  que son menores que los de  $F$ . De hecho, como se señala en el artículo de del Barrio et al. [6] (2018), en el caso de poblaciones finitas, lo que sucede cuando tomamos muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_n$  de  $F$  y  $G$ , el valor de  $\gamma$  se obtiene como sigue. Primero debemos reordenar ambos conjuntos de datos de manera creciente,  $x_{(1)} \leq \dots \leq x_{(n)}$  e  $y_{(1)} \leq \dots \leq y_{(n)}$ . Tras ello tenemos que comparar los valores  $x_{(i)}$  e  $y_{(i)}$  para todo  $i = 1, \dots, n$  de modo que  $\gamma$  se corresponde con el número de veces que  $x_{(i)} > y_{(i)}$  dividido entre el tamaño muestral,  $n$ . El código del programa `gamma()` realizado en el entorno de programación R puede consultarse en el apéndice C. Este índice nos proporciona un estadístico de rangos ordenados (llamado en algunos textos *estadístico de Galton*) el cual ha sido considerado en contrastes de hipótesis del tipo  $H_0 : F = G$  frente a  $H_a : F \leq_{st} G$  de modo que se rechaza la hipótesis nula para valores pequeños de  $\gamma$  en favor de la alternativa, que asegura la existencia de orden estocástico. La definición que hemos dado de  $\gamma(F, G)$  corresponde a la generalización del caso de poblaciones finitas utilizando las cuantiles.

El punto de vista anterior, si consideramos muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  de  $F$  y  $G$ , nos invita a considerar el estimador *plug-in*  $\gamma(F_n, G_m)$ , donde  $F_n$  y  $G_m$  son las funciones de distribución empíricas. Estas se definen en el contexto de una distribución que asigna probabilidad  $1/n$  (resp.  $1/m$ ) a cada valor  $x_i$ ,  $i = 1, \dots, n$  (resp.  $y_j$ ,  $j = 1, \dots, m$ ) de modo que, por ejemplo,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x), \quad x \in \mathbb{R},$$

y de modo análogo para  $G_m$ . El principio del plug-in está soportado por el *teorema de Glivenko-Cantelli*, el cual afirma que

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$$

casi seguro. Por lo tanto, la función de distribución empírica  $F_n$  es una estimación de la distribución completa  $F$ .

Si dos funciones de distribución coinciden en un intervalo, o incluso si ambas se cruzan un número infinito de veces, entonces  $\gamma(F, G)$  no se puede estimar de manera consistente al utilizar su versión plug-in. Si se evita este hecho, se puede demostrar el siguiente teorema, recogido en el artículo de del Barrio et al. [6] (2018):

**Teorema 2.1.1.** Sean  $F$  y  $G$  dos funciones de distribución tales que  $\ell(F^{-1} = G^{-1}) = 0$ . Entonces el estimador plug-in  $\hat{\gamma}_{n,m} = \gamma(F_n, G_m)$  es consistente para  $\gamma(F, G)$ , esto es

$$\hat{\gamma}_{n,m} \xrightarrow{c.s.} \gamma(F, G) \text{ si } n, m \rightarrow \infty.$$

Este índice, que en la práctica resulta fácil de calcular, se puede utilizar para realizar contrastes que permiten determinar si se produce orden al tratar con muestras (siempre que  $\gamma(F, G)$  sea menor que una cantidad  $\gamma_0$  fijada). Para ello se necesita su distribución asintótica, es decir, lo que ocurre cuando  $n \rightarrow \infty$ . Todo esto lo comentaremos más detenidamente en la sección 2.4.

### 2.1.3. Casi orden estocástico y el índice $\alpha(F, G)$

Aunque esta relación se introduce a través de funciones de utilidad, Leshno y Levy publicaron el artículo “*Preferred by “All” and Preferred by “Most” Decision Makers: Almost Stochastic Dominance*” [22] en el cual extendieron el concepto de orden estocástico al de *casi orden estocástico* para resolver algunos temas de inversiones y toma de decisiones. El título del artículo es bastante informativo, ya que trataron de buscar que la relación de orden estocástico se tenga para un rango suficientemente grande de modo que se pueda asegurar para la mayoría de sucesos que puedan tener lugar. Su idea consiste en medir la contribución relativa de la diferencia entre ambas funciones de distribución  $|F - G|$  sobre el conjunto  $\{F < G\}$  donde no se satisface  $F \leq_{st} G$ , es decir, estamos midiendo la distancia  $L_1$  entre  $F$  y  $G$ .

A partir del concepto de casi orden estocástico expuesto en este artículo se considera el índice  $\alpha(F, G)$  definido como

$$\alpha(F, G) := \frac{\int_{\{G > F\}} (G(x) - F(x)) dx}{\int_{-\infty}^{\infty} |G(x) - F(x)| dx},$$

que podemos utilizar al igual que  $\rho(F, G)$  y  $\gamma(F, G)$  para medir el grado de desviación que hay con la relación de orden estocástico entre dos distribuciones. Observemos que este índice no está relacionado con argumentos probabilísticos, luego se ha de tratar dentro del ámbito del análisis matemático.

Mientras que  $\gamma(F, G)$  está bien definido para cualquier par de funciones de distribución  $F$  y  $G$ ,

el índice  $\alpha(F, G)$  necesita que hagamos algunas hipótesis sobre  $F$  y  $G$ . En el artículo de Leshno y Levy se trabaja con distribuciones acotadas en un cierto intervalo, y se limitan al caso en que  $\alpha(F, G) < 0,5$ . En el artículo de Álvarez-Esteban et al. (2017) [2] se generaliza a distribuciones cualesquiera pero siempre asegurando que  $F$  y  $G$  tienen esperanza finita (esto es, que las variables consideradas con funciones de distribución  $F$  y  $G$  son integrables).

En el contexto de las funciones de utilidad sobre las que se habla en este artículo de Leshno y Levy, este índice  $\alpha(F, G)$  goza de buenas propiedades, aunque tiene un inconveniente. En el teorema 1.1.2 del capítulo 1 vimos que el orden estocástico se conserva a través de funciones crecientes, por lo que sería recomendable que esto también ocurriese para los índices que aquí contemplamos. La propiedad de conservarse por transformaciones crecientes es cierta para  $\gamma(F, G)$ , pero no lo es para  $\alpha(F, G)$ . Si tomamos una función estrictamente creciente  $\varphi$  y transformamos dos variables  $X$  e  $Y$  de modo que  $\varphi(X)$  y  $\varphi(Y)$  tienen funciones de distribución  $F_\varphi$  y  $G_\varphi$ , entonces se verifica que  $\gamma(F, G) = \gamma(F_\varphi, G_\varphi)$ , mientras que no tenemos garantías de que  $\alpha(F, G) = \alpha(F_\varphi, G_\varphi)$ .

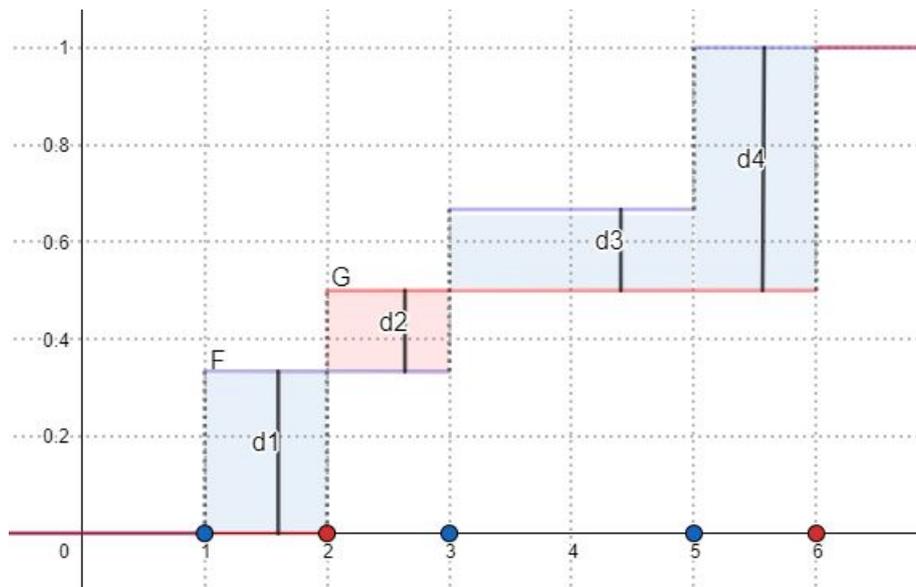


Figura 2.3: Funciones de distribución empíricas para los conjuntos  $\{1,3,5\}$  ( $F$ ) y  $\{2,6\}$  ( $G$ )

Al tratar con muestras, procediendo de manera análoga a como hicimos con el índice anterior, también podemos considerar el estimador *plug-in*  $\alpha(F_n, G_m)$  donde  $F_n$  y  $G_m$  son las funciones de distribución empíricas calculadas a partir de dos muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  tomadas de  $F$  y  $G$  respectivamente. Si tenemos en cuenta que las gráficas de  $F_n$  y  $G_m$  son escalonadas, y que el índice  $\alpha(F_n, G_m)$  se define en términos de la distancia  $L_1$  entre ambas, para calcularlo simplemente tenemos que ir sumando el área de los rectángulos que se van formando entre ambas como podemos ver en la figura 2.3, teniendo en cuenta donde  $G_m$  está por encima de  $F_n$  (contribuye al numerador y denominador) y donde se produce lo contrario (solo contribuye al denominador). Hemos realizado el programa `alpha()` en **R** que nos permite calcularlo sin tener ningún tipo de información adicional sobre las muestras, cuyo código y cuya explicación pueden consultarse en el apéndice C y que

utilizaremos al tratar algunos ejemplos al final del capítulo.

#### 2.1.4. Modelos de contaminación, recorte de probabilidades y el índice $\pi(F, G)$

A continuación vamos a considerar una aproximación alternativa a las anteriores, contemplada en el artículo de Álvarez-Esteban et al. (2014) [1]. En primer lugar, tengamos en cuenta que dadas funciones de distribución  $F$  y  $G$  siempre existe un valor  $\pi \in [0, 1]$  de modo que  $F$  y  $G$  admiten descomposiciones

$$\begin{cases} F &= (1 - \pi)\tilde{F} + \pi R \\ G &= (1 - \pi)\tilde{G} + \pi S \end{cases} \quad (2.2)$$

para determinadas funciones de distribución  $\tilde{F}, \tilde{G}, R, S$  tales que  $\tilde{F} \leq_{st} \tilde{G}$ . La *descomposición en mixturas* anterior puede entenderse como un *modelo de contaminación*, en el sentido de que  $F$  y  $G$  han sido generadas de manera aleatoria a partir de ciertas distribuciones latentes  $\tilde{F}$  y  $\tilde{G}$  (a las cuales en principio no podemos acceder) para las que se verifica  $\tilde{F} \leq_{st} \tilde{G}$  y a las que se les ha introducido un residuo que permite cuantificar cómo de lejos estamos de la relación de orden (podría pensarse en una distribución de Bernoulli de parámetro  $\pi$ ). Si ese valor  $\pi$  que mide el *nivel de contaminación* es lo suficientemente pequeño, entonces podríamos decir que *gran parte* de la distribución  $G$  domina a la de  $F$ . Por lo tanto, el valor  $\pi \in [0, 1]$  más pequeño y compatible con tal descomposición podría tomarse como el grado de desacuerdo con la relación de orden estocástico, ya que mide la proporción de los individuos que quedan fuera. Notemos que si tomamos  $\pi$  suficientemente grande el modelo 2.2 sigue teniendo sentido, incluso si  $\pi = 1$  (que en el fondo sería no tener información), por lo que obviaremos este caso.

Los modelos de contaminación comparten una **dualidad** con la idea de *recortar probabilidades*, como se establece en el artículo de Álvarez-Esteban et al. (2014) [1]. El concepto de *recortar* tiene que ver con reducir la influencia de datos contaminantes al hacer inferencia, técnica propia de la estadística robusta con el propósito de eliminar el *ruido* (es decir, observaciones molestas). Para entender esta noción, veamos en qué consiste en el caso discreto (lo que sucede, por ejemplo, al tratar con muestras): para un valor  $\pi \in [0, 1]$ , *recortar* un conjunto de datos  $\chi = \{x_1, \dots, x_n\}$  con *nivel*  $\pi$  consiste en definir una función peso  $\omega$  sobre  $\chi$  de modo que

$$0 \leq \omega(x_i) \leq 1, \quad i = 1, \dots, n \quad \text{y} \quad S_\omega := \frac{1}{n} \sum_{i=1}^n \omega(x_i) \geq 1 - \pi,$$

donde  $\omega(x_i)$  es la proporción que queda de  $x_i$  después de recortar. A cada una de las observaciones se le asigna una probabilidad recortada  $\tilde{P}_\omega(x_i) = \frac{1}{n} \frac{\omega(x_i)}{S_\omega}$ , esto es, estamos reemplazando la muestra original en la que cada punto tenía peso  $\frac{1}{n}$  por una nueva en la que se les da un peso desigual a los datos, de modo que los puntos de menor influencia tienen ahora probabilidad 0 y el resto tienen peso  $\frac{1}{n(1-\pi)}$ . De este modo, en vez de quitar los puntos *contaminados* estamos incrementando el peso de los puntos que están en un rango bueno (por un factor acotado por  $\frac{1}{1-\pi}$ ) y rebajando la importancia de los datos que están en una zona mala.

Para medidas de probabilidad generales, el concepto ahora es muy sencillo ya que la función de pesos se sustituye por una *función de densidad*. Dada una función de probabilidad  $P$  y un suceso  $B \in \sigma$ , el análogo al proceso de recorte sería reemplazar la probabilidad  $P(B) = \int_B 1 dP$  por la nueva medida

$$\tilde{P}(B) = \int_B g dP \text{ para alguna función } g \text{ con } 0 \leq g \leq \frac{1}{1-\pi} \text{ } P\text{-c.s.} \quad (2.3)$$

A una medida de probabilidad  $\tilde{P}$  como en 2.3 la llamaremos  $\pi$ -recorte de  $P$ . De modo análogo a lo anterior, el papel de la función  $g$  del integrando es el de una *densidad*, que nos permite incrementar o disminuir la influencia de algunas regiones del espacio muestral (teniendo probabilidad de hasta  $1 - \pi$ ) en el modelo.

Al conjunto de los  $\pi$ -recortes de  $P$  lo denotaremos como  $\mathcal{R}_\pi(P)$ . Notemos que si  $\pi = 0$  entonces  $\mathcal{R}_\pi(P) = \{P\}$ . Este método de recortar probabilidades en vez de quitar puntos dota a  $\mathcal{R}_\pi(P)$  de buenas propiedades. En particular, **existe una dualidad entre los problemas de contaminación y los conjuntos de  $\pi$ -recortes**, como recoge el siguiente teorema que aparece en Álvarez-Esteban et al. (2014) [1]. Para su demostración son necesarios algunos conocimientos sobre análisis real relacionados con el famoso teorema de Radon-Nikodym en teoría de la medida, y que recogemos en el apéndice B.2.

**Teorema 2.1.2.** Sean  $P_0, P$  distribuciones de probabilidad sobre  $\mathbb{R}$  con funciones de distribución  $F_0$  y  $F$  respectivamente, y sea  $\pi \in [0, 1)$ . Las siguientes afirmaciones son equivalentes:

1.  $P = (1 - \pi)P_0 + \pi Q$  para alguna medida de probabilidad  $Q$ .
2.  $(1 - \pi)P_0(B) \leq P(B)$  para todo  $B \in \sigma$ .
3.  $P_0 \in \mathcal{R}_\pi(P)$ .
4.  $F(x) = (1 - \pi)F_0(x) + \pi G(x)$  para todo  $x \in \mathbb{R}$  y alguna función de distribución  $G$ .

*Demostración.*

(1)  $\Rightarrow$  (2): si  $P = (1 - \pi)P_0 + \pi Q$ , dado que  $Q$  es una probabilidad (está entre 0 y 1) entonces se tiene que  $P \geq (1 - \pi)P_0$ .

(2)  $\Rightarrow$  (3): por la condición en (2),  $P_0$  es absolutamente continua con respecto a  $P$ , puesto que si  $P(B) = 0$  entonces  $P_0(B) = 0$  para todo  $B \in \sigma$ . Entonces por el teorema de Radon-Nikodym (B.2.3) existe una función de densidad no negativa  $g := \frac{dP_0}{dP}$  tal que  $P_0(B) = \int_B g dP$  para todo  $B \in \sigma$ . Solo nos falta ver que  $g \leq \frac{1}{1-\pi}$   $P$ -c.s.. Razonemos por reducción al absurdo y consideremos el conjunto  $B = \{g > \frac{1}{1-\pi}\}$ . Si  $P(B) > 0$  y  $B_\delta = \{g \geq \frac{\delta}{1-\pi}\}$  entonces  $P(B_\delta) > 0$  para algún  $\delta > 1$ . De este modo tendríamos que

$$(1 - \pi)P_0(B) = (1 - \pi) \int_B g dP \geq (1 - \pi) \frac{\delta}{1 - \pi} P(B) > P(B)$$

lo que contradice que  $P \geq (1 - \pi)P_0$ . Por lo tanto ha de ser  $g \leq \frac{1}{1-\pi}$   $P$ -c.s. y se tiene (3).

(3)  $\Rightarrow$  (1): supongamos que  $P_0 \in \mathcal{R}_\pi(P)$ . Sea  $g$  la densidad de  $P_0$  respecto de  $P$  de modo que para todo  $B \in \sigma$  se tiene  $P_0(B) = \int_B g dP$  y  $0 \leq g \leq \frac{1}{1-\pi}$ . Entonces  $(1 - \pi)P_0(B) \leq \int_B 1 dP = P(B)$ . Definamos la medida (no negativa)  $\tilde{Q}(B) := P(B) - (1 - \pi)P_0(B)$ . Para obtener (1) basta tomar  $Q(B) := \tilde{Q}(B)/\pi$ .

(1)  $\Leftrightarrow$  (4) es trivial, pues por un lado es la relación entre las funciones de probabilidad y por otro lado para las funciones de distribución (basta tomar  $B = (-\infty, x]$  y tener en cuenta que estos conjuntos generan la  $\sigma$ -álgebra de Borel  $\beta$ ).  $\square$

Si identificamos cada probabilidad sobre la recta con su función de distribución y denotamos por  $\mathcal{R}_\pi(F)$  al conjunto de  $\pi$ -recortes de  $F$ , el teorema anterior nos permite concluir que la descomposición en 2.2 existe si y solo si existen  $\tilde{F} \in \mathcal{R}_\pi(F)$  y  $\tilde{G} \in \mathcal{R}_\pi(G)$  de modo que  $\tilde{F} \leq_{st} \tilde{G}$ . Por tanto, sería deseable saber más cosas sobre estas funciones  $\tilde{F}$  y  $\tilde{G}$  ya que nos permitirán ver qué proporción de datos nos impide establecer la condición de orden.

La siguiente proposición tomada del artículo de del Barrio et al. [6] (2018) nos aporta más información sobre de la relación entre los modelos de contaminación y el recorte de probabilidades, en concreto sobre la función latente  $\tilde{F}$ :

**Proposición 2.1.3.** Sean  $\tilde{P}, P$  distribuciones de probabilidad sobre  $\mathbb{R}$  con funciones de distribución  $\tilde{F}$  y  $F$  respectivamente y sea  $\pi \in [0, 1)$ . Se definen las funciones de distribución

$$F^\pi(x) = \max\left(\frac{1}{1-\pi}(F(x) - \pi), 0\right) \quad \text{y} \quad F_\pi(x) = \min\left(\frac{1}{1-\pi}F(x), 1\right).$$

Entonces  $\tilde{P} \in \mathcal{R}_\pi(P)$  es equivalente a las siguientes afirmaciones:

1.  $F(x) = (1 - \pi)\tilde{F}(x) + \pi R(x)$  para todo  $x \in \mathbb{R}$  y una función de distribución  $R$ .
2.  $F_\pi \leq_{st} \tilde{F} \leq_{st} F^\pi$ .

*Demostración.* La equivalencia entre (1) y (2) ya la vimos en el teorema 2.1.2, aunque es importante tenerla en cuenta en lo que sigue. Si consideramos  $F_\pi$ , se tiene que  $\frac{1}{\pi}(F - (1 - \pi)F_\pi)$  es una función de distribución, luego aplicando el teorema 2.1.2 se tiene que  $F_\pi \in \mathcal{R}_\pi(F)$ . Recíprocamente, si tenemos un  $\pi$ -recorte  $\tilde{F}$  de  $F$ , por definición lo podemos expresar como  $\tilde{F}(x) = \int_{-\infty}^x g(t) dF(t)$  para alguna densidad  $g$  con respecto a  $P$  que satisface  $0 \leq g \leq \frac{1}{1-\pi}$ . Pero entonces  $\tilde{F}(x) \leq \min(\frac{1}{1-\pi}F(x), 1) = F_\pi(x)$  para todo  $x$ , esto es,  $F_\pi \leq_{st} \tilde{F}$ . De modo análogo se prueba para  $F^\pi$ .  $\square$

La afirmación 2 quiere decir que el conjunto de  $\pi$ -recortes de una distribución dada sobre  $\mathbb{R}$  tiene un mínimo y un máximo respecto al orden estocástico, caracterizados mediante  $F_\pi$  y  $F^\pi$  (a los que llamaremos *recorte minimal* y *recorte maximal* de  $F$ ). Estas funciones de distribución se obtienen *recortando* la cola derecha (respectivamente izquierda) de la probabilidad con función de distribución  $F$  (quitando una proporción  $\pi$  en ambos casos).

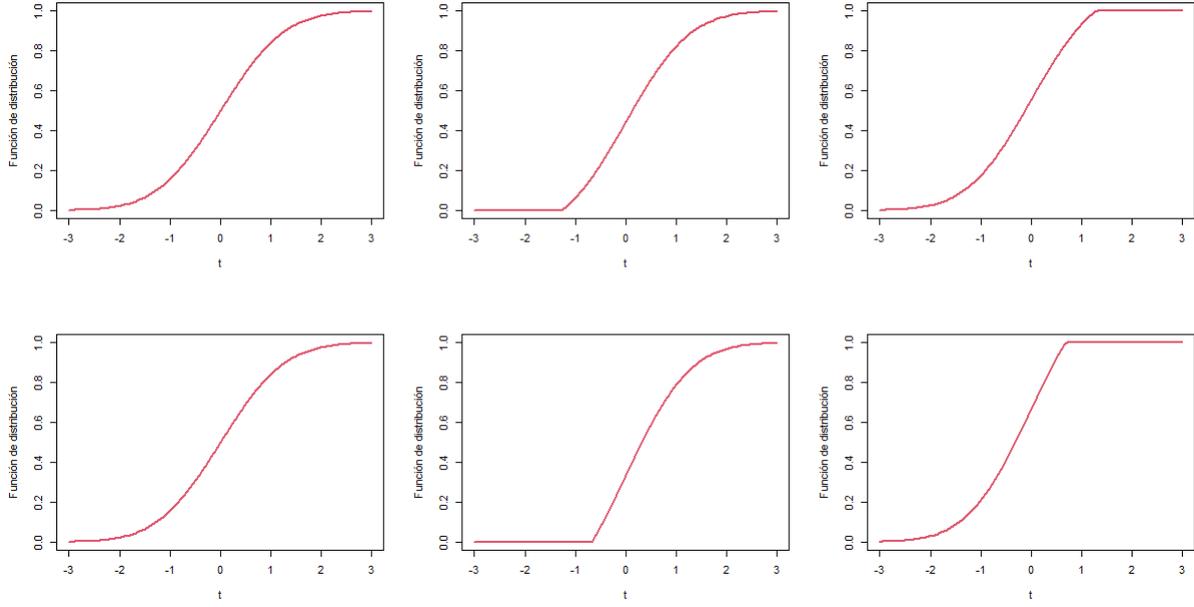


Figura 2.4: Recorte de normales para  $\pi = 0,1$  (arriba) y  $\pi = 0,25$  (abajo)

Echemos un vistazo a la figura 2.4. En la línea de arriba aparece la función de distribución de la normal estándar  $N(0, 1)$  (izquierda) acompañada de los recortes  $F^\pi$  (centro) y  $F_\pi$  (derecha) para  $\pi = 0,1$ , mientras que en la línea de abajo se hace lo mismo para  $\pi = 0,25$ . En la gráfica  $F^\pi$  se da influencia 0 a los valores que están por debajo de  $F^{-1}(\pi)$  (se hace un recorte por la izquierda), mientras que en la de  $F_\pi$  se da influencia 0 a los que están por encima de  $F^{-1}(1 - \pi)$  (se hace un recorte por la derecha). También podemos apreciar que como en las gráficas de arriba el nivel de contaminación  $\pi$  es más bajo, el recorte que se hace es más pequeño que en las de abajo.

A partir de las caracterizaciones anteriores de los recortes se deduce la siguiente proposición, que se encuentra en el artículo de Álvarez-Esteban et al. (2017) [2]:

**Proposición 2.1.4.** Dadas funciones de distribución  $F$  y  $G$  y un valor  $\pi \in [0, 1)$ , entonces existe la descomposición en mixturas dada en 2.2 si y solo si  $F_\pi \leq_{st} G^\pi$ , si y solo si  $\pi \geq \pi(F, G)$ , donde

$$\pi(F, G) := \sup_{x \in \mathbb{R}} (G(x) - F(x)) \quad (2.4)$$

La proposición anterior nos muestra cual es el valor de  $\pi \in [0, 1)$  mínimo para el cual existe el modelo de contaminación 2.2, que nos permite cuantificar la desviación que existe respecto al orden estocástico entre  $F$  y  $G$ . Este teorema nos proporciona además las funciones  $\tilde{F} = F_\pi$  y  $\tilde{G} = G^\pi$  para las que se satisface  $\tilde{F} \leq_{st} \tilde{G}$  en dicho modelo. En la práctica esto quiere decir que si recortamos la cola derecha de  $F$  y la cola izquierda de  $G$  removiendo una proporción de  $\pi$  en ambos casos entonces las distribuciones resultantes  $\tilde{F}$  y  $\tilde{G}$  verifican  $\tilde{F} \leq_{st} \tilde{G}$ .

Obviamente se tiene que  $\pi(F, G) = 0$  si y solo si  $F \leq_{st} G$ , y además  $\pi(F, G) + \pi(G, F) \leq 1$ . Es también importante resaltar que  $\pi(F, G)$  es invariante al aplicar funciones crecientes, igual que

$\gamma(F, G)$ .

Al igual que hicimos con  $\gamma(F, G)$ , si tratamos con muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  de  $F$  y  $G$  también podemos considerar la versión *plug-in* de  $\pi(F, G)$ , esto es,

$$\hat{\pi}_{n,m} := \sup_{x \in \mathbb{R}} (G_m(x) - F_n(x)) = \pi(F_n, G_m).$$

Como podemos ver, este estimador se corresponde con el *estadístico de Kolmogorov-Smirnov* de un lado, que tiene un papel importante en los problemas de dos muestras en el marco de las pruebas de bondad de ajuste no paramétricas. Este función ya viene programada por defecto en **R**: para calcular el índice dadas dos muestras (vectores) **c1** y **c2** de  $F$  y  $G$  basta con utilizar el comando `ks.test(c1, c2, alternative="less")`. El valor de  $\pi(F_n, G_m)$  coincide con el valor del estadístico  $D^-$  que proporciona el test.  $\pi(F_n, G_m)$  es un estimador consistente, como muestra el siguiente teorema:

**Teorema 2.1.5.** Sean  $F$  y  $G$  dos funciones de distribución. Entonces el estimador plug-in  $\hat{\pi}_{n,m} = \pi(F_n, G_m)$  es consistente para  $\pi(F, G)$ , esto es,

$$\hat{\pi}_{n,m} \xrightarrow{c.s.} \pi(F, G) \text{ si } n, m \rightarrow \infty.$$

*Demostración.* Para la demostración utilizaremos el teorema de Glivenko-Cantelli aplicado a los estimadores  $F_n$  y  $G_m$ . Según este, se tiene que

$$\sup_{x \in \mathbb{R}} |F_n(\omega, x) - F(x)| \longrightarrow 0 \text{ c.s.}$$

Fijado un elemento  $\omega$  perteneciente al conjunto donde sucede lo anterior con probabilidad 1, por la convergencia uniforme lo anterior equivale a que para todo  $\varepsilon > 0$  existe  $N \in \mathbb{N}$  tal que  $|F_n(x) - F(x)| < \varepsilon$  para todo  $n \geq N$  y para todo  $x \in \mathbb{R}$ .

De modo análogo para  $G_m$ ,

$$\sup_{x \in \mathbb{R}} |G_m(\omega, x) - G(x)| \longrightarrow 0 \text{ c.s.,}$$

luego para el mismo  $\varepsilon > 0$  anterior podemos tomar un  $N \in \mathbb{N}$  común de forma que se dé lo anterior y además  $|G_m(x) - G(x)| < \varepsilon$  para todo  $m \geq N$  y para todo  $x \in \mathbb{R}$ .

Para demostrar nuestro resultado, observemos que

$$\begin{aligned} G_m(x) - F_n(x) &= G_m(x) - G(x) + G(x) - F(x) + F(x) - F_n(x) \leq \\ &\leq \varepsilon + G(x) - F(x) + \varepsilon = G(x) - F(x) + 2\varepsilon, \end{aligned}$$

luego tomando superiores

$$\sup_{x \in \mathbb{R}}(G_m(x) - F_n(x)) \leq \sup_{x \in \mathbb{R}}(G(x) - F(x)) + 2\varepsilon. \quad (2.5)$$

Por otro lado, tenemos que

$$\begin{aligned} G(x) - F(x) &= G(x) - G_m(x) + G_m(x) - F_n(x) + F_n(x) - F(x) \leq \\ &\leq \varepsilon + G_m(x) - F_n(x) + \varepsilon = G_m(x) - F_n(x) + 2\varepsilon, \end{aligned}$$

luego tomando superiores

$$\sup_{x \in \mathbb{R}}(G(x) - F(x)) \leq \sup_{x \in \mathbb{R}}(G_m(x) - F_n(x)) + 2\varepsilon. \quad (2.6)$$

Juntando 2.5 y 2.6 obtenemos que

$$\sup_{x \in \mathbb{R}}(G(x) - F(x)) - 2\varepsilon \leq \sup_{x \in \mathbb{R}}(G_m(x) - F_n(x)) \leq \sup_{x \in \mathbb{R}}(G(x) - F(x)) + 2\varepsilon$$

de modo que tenemos asegurada la convergencia.  $\square$

Este índice también es fácil de interpretar, ya que si por ejemplo  $F$  e  $G$  son las funciones de distribución de dos variables que representan la esperanza de vida ante dos determinados tratamientos para las que sabemos que  $\pi(F, G) = 0,05$ , entonces podemos afirmar que el nuevo tratamiento clínico mejora la esperanza de vida al quitar el 5% de las observaciones de cada lado. Como comentaremos en la sección 2.4, este índice se suele utilizar también para realizar contrastes que permiten determinar que las poblaciones están ordenadas (asegurando que  $\pi(F, G)$  es menor que un valor  $\pi_0$  fijado), y para lo que es necesario conocer su distribución asintótica.

Observemos que  $\pi(F, G)$  y  $\gamma(F, G)$  son índices que miden el grado de desviación entre dos distribuciones respecto de la relación de orden estocástico, toman valores en  $[0, 1]$  y  $\gamma(F, G) = 0$  si y solo si  $\pi(F, G) = 0$ , lo que equivale a que  $F \leq_{st} G$ . Una cuestión que podríamos plantearnos es si existe alguna relación entre ambos, como se estudia en el artículo de Álvarez-Esteban et al. (2017) [2]. Si tomamos un par  $(X, Y)$  de variables aleatorias con funciones de distribución marginales  $F$  y  $G$  respectivamente, entonces

$$\begin{aligned} G(x) &= P(Y \leq x, X \leq Y) + P(Y \leq x, X > Y) \leq \\ &\leq P(X \leq x, X \leq Y) + P(X > Y) \leq F(x) + P(X > Y) \end{aligned}$$

Si consideramos la representación a través de las funciones cuantil  $(X, Y) =^d (F^{-1}, G^{-1})$ , utilizando que  $P(X > Y) = \ell(F^{-1} > G^{-1}) = \gamma(F, G)$  y la caracterización del índice  $\pi(F, G)$  dada en la proposición 2.1.4, se tiene el siguiente resultado al tomar el superior:

**Proposición 2.1.6.** Sean  $F$  y  $G$  dos funciones de distribución arbitrarias. Entonces

$$\pi(F, G) \leq \gamma(F, G).$$

## 2.2. $\gamma(F, G)$ y $\pi(F, G)$ en los modelos de localización y escala

En el capítulo 1, concretamente en el teorema 1.2.1, ya estudiamos las condiciones que han de cumplir los parámetros de una familia de localización y escala (de aquí en adelante, familia LS). Allí vimos que si  $F_{\theta_1, \lambda_1}$  y  $F_{\theta_2, \lambda_2}$  son dos funciones de distribución continuas y estrictamente crecientes pertenecientes a la misma familia LS basada en  $\mathcal{F}_0$ , donde  $\theta_1, \theta_2$  son los parámetros de localización y  $\lambda_1, \lambda_2$  los parámetros de escala, entonces las condiciones suficientes para que  $F_{\theta_1, \lambda_1} \leq_{st} F_{\theta_2, \lambda_2}$  son  $\lambda_1 = \lambda_2$  y  $\theta_1 \leq \theta_2$ .

### 2.2.1. $\gamma(F, G)$ en modelos LS

En dicho teorema hicimos el estudio a través de las funciones cuantil, lo que nos permite utilizarlo ahora para calcular el índice  $\gamma(F, G)$  en este modelo. Vimos que dos funciones cuantil de la misma familia LS tienen un punto de corte  $y_0$  si y solo si  $\lambda_1 \neq \lambda_2$  y  $\mathcal{F}_0^{-1}(y_0) = \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$ . Por la condición de ser estrictamente creciente se deduce que si ese punto de corte existe, entonces es único. Dependiendo del signo de  $\lambda_1 - \lambda_2$ , esto quiere decir que:

- el conjunto  $\{y : F_{\theta_1, \lambda_1}^{-1}(y) \leq F_{\theta_2, \lambda_2}^{-1}(y)\}$  es  $(0, \mathcal{F}_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})]$  o  $[\mathcal{F}_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}), 1)$ ;
- $\ell(\{F_{\theta_1, \lambda_1}^{-1} > F_{\theta_2, \lambda_2}^{-1}\})$  es  $1 - \mathcal{F}_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})$  o  $\mathcal{F}_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})$ .

Notemos también que

$$\gamma(F_{\theta_1, \lambda_1}, F_{\theta_2, \lambda_2}) = \gamma\left(F_{0,1}, F_{\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}, \lambda_1}\right),$$

luego podemos centrar nuestro análisis en comparar con  $\mathcal{F}_0$  como función de distribución de referencia.

Dadas ahora dos funciones de distribución  $F$  y  $G$  pertenecientes al mismo modelo LS, si queremos garantizar un nivel de correspondencia con la relación de orden estocástico  $F \leq_{st} G$  de por ejemplo el 95 % utilizando el índice  $\gamma(F, G)$ , entonces bastaría considerar el punto de corte entre ambas funciones cuantil y comprobar si  $\ell(\{F^{-1} \leq G^{-1}\}) \geq 0,95$  (lo que equivale a que  $\gamma(F, G) < 0,05$ ).

Por ejemplo, si consideramos el modelo LS normal, tomando  $\mathcal{F}_0 = \phi$  (la función de distribución asociada a la normal estándar  $N(0, 1)$ ) y sustituyendo en la expresión  $\ell(\{F_{\theta_1, \lambda_1}^{-1} > F_{\theta_2, \lambda_2}^{-1}\}) = 1 - \mathcal{F}_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})$ , se tiene que

$$\gamma(N(0, 1), N(\mu, \sigma)) = 1 - \phi\left(\frac{\mu}{|\sigma - 1|}\right) \text{ si } \sigma \neq 1. \quad (2.7)$$

El caso  $\sigma = 1$  ya ha sido estudiado: corresponde a la situación donde las desviaciones típicas son iguales, luego no hay cortes entre ambas. Si  $\mu \geq 0$  tenemos que  $\gamma(N(0, 1), N(\mu, 1)) = 0$  (relación de orden estocástico perfecta, pues  $N(0, 1) \leq_{st} N(\mu, 1)$ ), y si  $\mu < 0$  entonces  $\gamma(N(0, 1), N(\mu, 1)) = 1$  (la dominación sería a la inversa,  $N(\mu, 1) \leq_{st} N(0, 1)$ ). Lo anterior se podría generalizar a cualquier modelo LS reemplazando  $\phi$  por la función de referencia  $\mathcal{F}_0$  que se haya tomado y considerando en vez de  $\mu$  y  $\sigma$  los parámetros de localización y escala  $\theta$  y  $\lambda$  adecuados.

Esto nos sirve para determinar, fijado un nivel de desviación  $\gamma_0$  con la relación de orden, la sensibilidad que tiene el índice  $\gamma(F, G)$  al comparar cualquier distribución  $N(\mu, \sigma)$  con la estándar  $N(0, 1)$  estudiando cuando  $1 - \phi(\frac{\mu}{|\sigma-1|}) < \gamma_0$ , lo que equivale a estudiar los valores  $\mu \in \mathbb{R}$  y  $\sigma > 0$  tales que  $z_{\mu, \sigma} = \frac{\mu}{|\sigma-1|} > \phi^{-1}(1 - \gamma_0)$ .

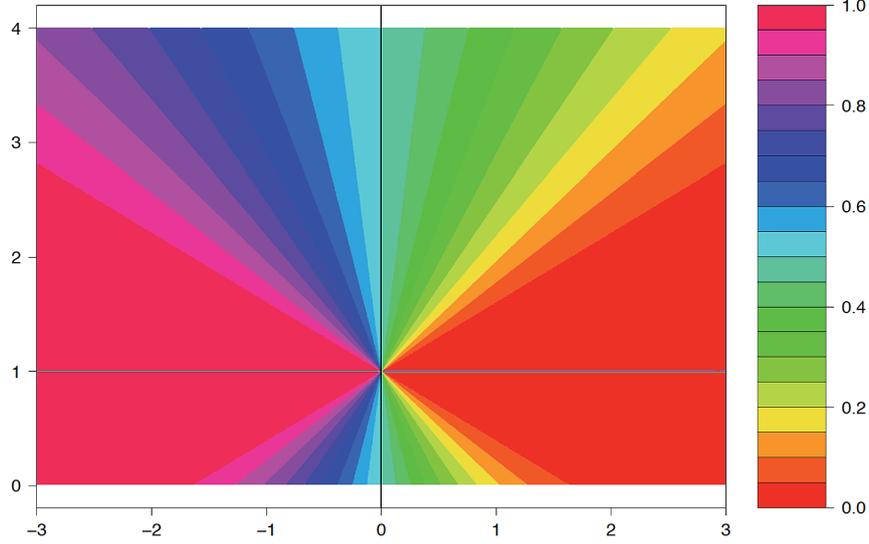


Figura 2.5: Contour-plot de  $\gamma(N(0, 1), N(\mu, \sigma))$  para diferentes valores de  $\mu$  (eje X) y  $\sigma$  (eje Y)

En el artículo de Álvarez-Esteban et al. (2017) [2] encontramos la gráfica de contornos de la figura 2.5, que mide el índice  $\gamma(F, G)$  entre la distribución normal estándar  $N(0, 1)$  y otra normal  $N(\mu, \sigma)$  al hacer variar los parámetros  $\mu$  y  $\sigma > 0$ . Para dibujarla se ha utilizado la igualdad en 2.7. Notemos que  $\gamma(N(0, 1), N(\mu, \sigma))$  es constante a lo largo de los rayos  $\{(\mu, \sigma) : \mu = C|\sigma - 1|, \sigma > 0\}$  para  $C > 0$ , y presenta una singularidad en  $\mu = 0, \sigma = 1$ . También se puede apreciar que según  $\sigma$  aumenta por encima de 1 (esto es, cuando la segunda muestra tiene mayor varianza) es compatible que  $\mu > 0$  (exista un aumento en media) mientras que  $\gamma(N(0, 1), N(\mu, \sigma))$  se acerca a  $\frac{1}{2}$ .

### 2.2.2. $\pi(F, G)$ en modelos LS

En lo que sigue vamos a estudiar el comportamiento de  $\pi(F, G)$  bajo el modelo LS. En primer lugar, puesto que

$$\sup_{x \in \mathbb{R}} (F_{\theta_2, \lambda_2}(x) - F_{\theta_1, \lambda_1}(x)) = \sup_{x \in \mathbb{R}} \left( F_{\frac{\theta_2 - \theta_1}{\lambda_1}, \frac{\lambda_2}{\lambda_1}}(x) - F_{0,1}(x) \right) \quad (2.8)$$

entonces se tiene que

$$\pi(F_{\theta_1, \lambda_1}, F_{\theta_2, \lambda_2}) = \pi \left( F_{0,1}, F_{\frac{\theta_2 - \theta_1}{\lambda_1}, \frac{\lambda_2}{\lambda_1}} \right),$$

por lo que, igual que antes, basta con hacer el análisis respecto a la función de referencia fijada (podemos tomar  $\mathcal{F}_0 = F_{0,1}$ ). A diferencia de la subsección anterior ahora no podemos dar una expresión cerrada y simple para  $\pi(F_{0,1}, F_{\theta, \lambda})$  para cualquier modelo LS, ya que el problema de optimización en 2.8 depende de la función de referencia  $\mathcal{F}_0$  elegida.

En el artículo de Álvarez-Esteban et al. (2017) [2] se muestra que sí es posible obtenerla en el caso del modelo normal (donde  $\mathcal{F}_0 = \phi$ ). Si  $\sigma \neq 1$  y  $\mu \geq 0$  (pues solo nos interesa hacer el análisis con un aumento en la media) entonces se tiene

$$\pi(N(0, 1), N(\mu, \sigma)) = \phi\left(\frac{\bar{x} - \mu}{\sigma}\right) - \phi(\bar{x}),$$

con

$$\bar{x} = \frac{\mu \pm \sqrt{\mu^2 \sigma^2 + 2\sigma^2(\sigma^2 - 1) \log \sigma}}{1 - \sigma^2} \quad (2.9)$$

donde se toma el signo positivo si  $\sigma > 1$  y el negativo si  $\sigma < 1$ . Notemos también que, para los mismos valores de  $\mu$  y  $\sigma$ ,

$$\pi(N(\mu, \sigma), N(0, 1)) = \phi(\bar{x}) - \phi\left(\frac{\bar{x} - \mu}{\sigma}\right),$$

tomando como  $\bar{x}$  la otra solución de 2.9. El resto de casos ya han sido estudiados: si  $\sigma = 1$  y  $\mu \geq 0$  se tiene que  $\pi(N(0, 1), N(\mu, 1)) = 0$  (esto es, la relación de orden estocástico es perfecta) y  $\pi(N(\mu, 1), N(0, 1))$  se alcanza en el único punto de corte de ambas funciones de densidad,  $\bar{x} = \mu/2$ .

A diferencia de la facilidad de cálculo que teníamos para  $\gamma(F, G)$  donde  $F$  es la función de distribución de  $N(0, 1)$  y  $G$  la de  $N(\mu, \sigma)$ , la expresión en 2.9 es ahora más complicada de estudiar, ya que por ejemplo las zonas donde  $\pi(F, G)$  se mantiene constante ya no son rectas.

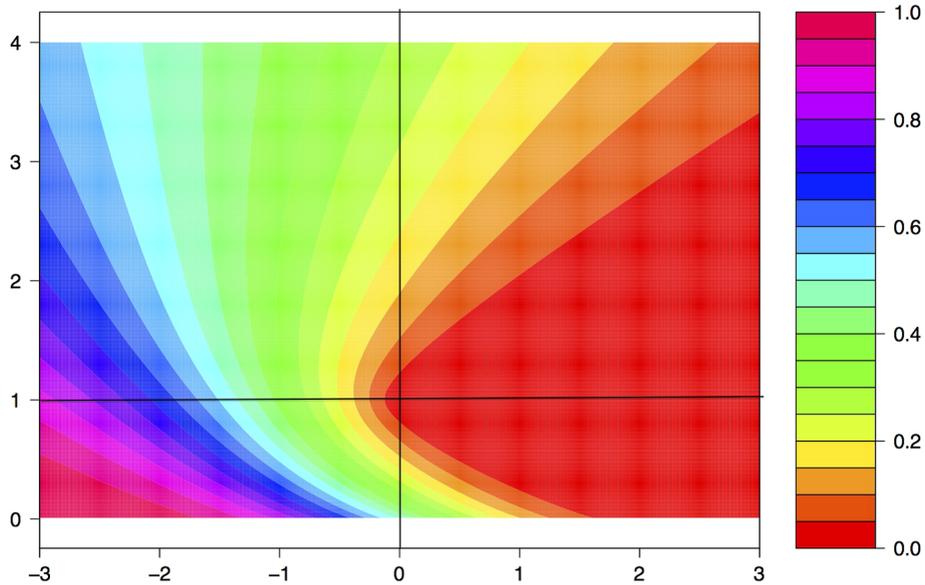


Figura 2.6: Contour-plot de  $\pi(N(0, 1), N(\mu, \sigma))$  para diferentes valores de  $\mu$  (eje X) y  $\sigma$  (eje Y)

Al igual que hicimos con  $\gamma(F, G)$ , también podemos utilizar  $\pi(F, G)$  para realizar un gráfico de nivel que nos compare una distribución  $N(0, 1)$  con cualquier otra distribución  $N(\mu, \sigma)$  al hacer variar sus parámetros para determinar cómo de alejadas están de la relación de orden estocástico,

como se puede apreciar en la figura 2.6 tomada del artículo de Álvarez-Esteban et al. (2017) [2]. De manera similar a como pasaba con  $\gamma(F, G)$  vemos que un aumento en la media es compatible con que haya niveles altos de desacuerdo con la relación de orden estocástico.

### 2.3. Orden estocástico usual *versus* comparación de medias

En este tipo de problemas de comparación de dos poblaciones se suele utilizar el *t-test* que da respuesta al *problema de Behrens-Fisher*, el cual ha sido estudiado en la asignatura de Estadística del grado y que se basa en contrastar la igualdad de medias entre dos poblaciones normales sin hacer ninguna suposición acerca de las varianzas. Supongamos que observamos dos muestras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_m$  independientes tomadas de dos poblaciones normales  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$ , donde las desviaciones típicas  $\sigma_1$  y  $\sigma_2$  se suponen desconocidas y distintas. El problema de Behrens-Fisher considera la *comparación de medias* de estas poblaciones normales a través del estudio de la diferencia entre ambas,  $\Delta = \mu_1 - \mu_2$ , además de tratar de obtener un intervalo de confianza al contrastar  $H_0 : \Delta \leq \Delta_0$  frente a  $H_a : \Delta > \Delta_0$ . No hay un test exacto (puesto que no se puede conseguir un estadístico de distribución libre, esto es, independiente de  $\sigma_1$  y  $\sigma_2$  y asumiendo  $\Delta = \Delta_0$ ), pero sí hay soluciones aproximadas (*test asintóticos*) como es el caso del *t-test*, que rechaza la hipótesis  $H_0$  con un nivel de aproximadamente  $\alpha$ . Esta aproximación es buena para valores grandes de  $n$  y  $m$  pero no lo es para valores moderados, por eso se suele utilizar la *corrección de Welch* que permite hacer una mejor aproximación mediante una modificación de los grados de libertad del *t-test*.

Siguiendo lo que se dice en Álvarez-Esteban et al. (2017) [2], supongamos que las leyes de probabilidad de la variable de interés bajo dos sistemas que queremos comparar son normales, esto es,  $P_i = N(\mu_i, \sigma_i)$ ,  $i = 1, 2$ . Si a través del contraste podemos concluir que  $\mu_1 > \mu_2$  **solamente** podríamos decir que, *en media*, el primer proceso devuelve valores mayores que el segundo. Si invocamos a la ley fuerte de los grandes números, esta nos asegura que si tenemos una muestra suficientemente grande de ambos entonces la media de la muestra obtenida a través de  $P_1$  sería mayor c.s. que la obtenida mediante  $P_2$ . Es también importante resaltar que este hecho no depende de las desviaciones típicas  $\sigma_i$ . La cuestión que nos interesa es: ¿estas situaciones son compatibles con que los valores que nos proporciona el primer proceso son mayores que los del segundo?.

Para analizar si esto sucede, podemos simplificar considerablemente el modelo si previamente ordenamos los elementos de cada población según cual sea su estatus (por ejemplo categorías de edad, precio...) ya que de este modo la posición relativa de cada elemento en su población se encuentra más rápidamente. También podemos normalizar cada una de las poblaciones cambiando las posiciones  $1, \dots, n$  por números entre 0 y 1 para conseguir de este modo estatus comparables que nos lleven a una relación más simple. Si  $F$  y  $G$  son las funciones de distribución de la variable de interés en la primera y segunda población respectivamente, y  $F^{-1}, G^{-1}$  son las funciones cuantil correspondientes, entonces

$$F^{-1}(t) \leq G^{-1}(t) \text{ para todo } t \in (0, 1) \quad (2.10)$$

significaría que la variable de la primera población es más pequeña que la segunda. Pero, ¿esto no

debería parecernos ya familiar?. En efecto, esta es la caracterización del **orden estocástico usual** (a no ser que se diga lo contrario, en este capítulo nos referiremos a él como *orden estocástico*) a través de las funciones cuantil, como muestra la proposición 1.1.1 que vimos en el capítulo 1. Por lo tanto, estaríamos afirmando que  $F \leq_{st} G$ .

Por ejemplo, podemos estar interesados en comparar un nuevo tratamiento clínico para ver si aumenta la duración de la vida de un grupo de pacientes respecto al que ya se utilizaba anteriormente. Si tomamos ciertas medidas del nivel de salud de estos tras someterlos al nuevo tratamiento, una opción podría ser comparar los valores medios del viejo respecto al nuevo. Sin embargo, incluso asumiendo la normalidad implícita en el  $t$ -test, tener evidencia de que la media del nuevo tratamiento es significativamente mayor que la del anterior es compatible con que haya una proporción alta de pacientes en los que la duración de vida disminuya. Observemos por ejemplo la gráfica 2.5 sobre el índice  $\gamma(F, G)$ : es posible que se produzca simultáneamente un incremento en la media ( $\mu > 0$ ) y que el grado de desacuerdo respecto al orden estocástico sea de hasta casi un 50 %, lo que significa que el tratamiento tiene peores efectos para una gran proporción de pacientes. Queda por lo tanto claro que la respuesta a la pregunta “¿es el nuevo tratamiento mejor que el anterior?” debería basarse en **la existencia de orden estocástico**, y en general ante cualquier comparación entre dos poblaciones o para determinar un aumento o mejora en una población (comparar el antes con el después).

## 2.4. Contrastando el orden estocástico usual

Al trabajar con muestras, y actuando de manera similar al problema de Behrens-Fisher visto en la sección anterior, si queremos realizar un contraste de hipótesis para determinar la existencia de orden estocástico entre estas deberíamos probar la hipótesis nula de que no existe orden contra la alternativa que afirma que sí existe, para así poder tener evidencia que lo demuestre (y concluir que el nuevo tratamiento del que hablábamos antes es mejor). Sin embargo, desde el inicio de esta teoría el contraste se ha planteado desde varios puntos de vista, como podemos ver en lo que sigue.

En 1947, Mann y Whitney estudiaron en [24] métodos para asegurar la existencia de orden estocástico *estricto* (esto es,  $F <_{st} G$  si  $F(x) \geq G(x)$  para todo  $x$  y  $F(x_0) > G(x_0)$  para al menos un valor  $x_0$ ). En este artículo trataron de contrastar la hipótesis nula  $H_0 : F = G$  contra la alternativa  $H_a : F <_{st} G$ . A partir de dos muestras de  $n$  valores  $x_i$  pertenecientes a un grupo y  $m$  valores  $y_i$  pertenecientes a otro (ambas ya ordenadas) se fijaron en  $U$ , el número de veces que un valor  $y_i$  es anterior a un valor  $x_i$ , encontraron su distribución y dieron un criterio para contrastar ambas hipótesis (lo que se conoce como *estadístico de Mann y Whitney*). Sin embargo, deberíamos tener cuidado con este método, ya que tanto  $H_0$  como  $H_a$  podrían ser falsas simultáneamente.

Otros autores trataron de estudiar el problema de contrastar  $H_0 : F \leq_{st} G$  contra  $H_a : F \not\leq_{st} G$ . El no rechazar la hipótesis nula consiste en reconocer que no tenemos evidencia en contra de que no exista orden estocástico. Recordemos que el contraste de hipótesis está diseñado para mostrar evidencia de poder rechazar la hipótesis nula, pero que no la rechazamos no quiere decir que tengamos

evidencia de que esta es cierta, es decir, no tenemos garantías de que una variable sea estocásticamente menor que otra. No somos capaces de confirmar el modelo pero, en el mejor de los casos, tendríamos falta de evidencia para rechazarlo. Este enfoque nos podría dar problemas al trabajar posteriormente con los datos.

Por lo tanto, si queremos tener un test para el cual rechazar la hipótesis nula sea interpretable como tener evidencia suficiente de que  $F \leq_{st} G$ , deberíamos estudiar el contraste de hipótesis

$$H_0 : F \not\leq_{st} G \text{ contra } H_a : F \leq_{st} G. \quad (2.11)$$

El artículo [8] de R. L. Berger nos muestra que esto no es práctico, pues en él se demuestra que el test uniformemente más potente de nivel  $\alpha$  para contrastar  $H_0$  contra  $H_a$  es el test “*no data*”, el cual rechaza  $H_0$  con probabilidad  $\alpha$  sin tener en cuenta los datos de los que disponemos. Esto se debe a que es imposible la interpretación del orden estocástico debido a que pequeñas variaciones en las colas de la distribución pueden impedirnos o facilitarnos la relación de dominancia.

Podemos, no obstante, plantearnos el enfoque que se recoge en los artículos de Álvarez-Esteban et al. [1] (2014) y [2] (2017): los índices  $\pi(F, G)$  y  $\gamma(F, G)$  estudiados en la sección 1 de este capítulo miden el grado de desviación respecto al orden estocástico entre  $F$  y  $G$ . Poder afirmar que estos son pequeños quiere decir que estamos cerca de la relación de orden, por lo tanto podemos realizar el contraste

$$\begin{aligned} H_0 : \pi(F, G) \geq \pi_0 \text{ contra} \\ H_a : \pi(F, G) < \pi_0, \end{aligned} \quad (2.12)$$

o

$$\begin{aligned} H_0 : \gamma(F, G) \geq \gamma_0 \text{ contra} \\ H_a : \gamma(F, G) < \gamma_0. \end{aligned} \quad (2.13)$$

Si en 2.12 conseguimos evidencia que nos permita rechazar la hipótesis nula, entonces tendríamos evidencia de que se produce orden estocástico aproximado, salvo por un valor de contaminación  $\pi_0$  pequeño. En 2.13 rechazar la hipótesis nula nos garantiza que el tratamiento asociado a  $G$  tiene mejores resultados que el asociado a  $F$  en una proporción de al menos  $1 - \gamma_0$  de la población. Si tenemos en cuenta la dualidad que existe entre los contrastes unilaterales y las cotas de confianza, estaríamos de este modo interesados en obtener una cota superior de confianza tanto para  $\pi(F, G)$  como para  $\gamma(F, G)$ . Si  $\hat{U} = \hat{U}(X_1, \dots, X_n, Y_1, \dots, Y_m)$  (resp.  $\hat{V}$ ) fuera una cota superior de confianza (asintótica) para  $\pi(F, G)$  (resp. para  $\gamma(F, G)$ ), rechazar  $H_0$  en 2.12 cuando  $\hat{U} < \pi_0$  (resp. cuando  $\hat{V} < \gamma_0$ ) nos proporcionaría un test donde tendríamos controlado, de manera asintótica, el error de tipo I (es decir, rechazar la hipótesis nula siendo esta verdadera para la población). En este sentido, podremos por tanto basar nuestra decisión en estas cotas superiores de confianza.

Para estudiar el contraste 2.12 se utiliza el estimador plug-in  $\pi(F_n, G_m) = \sup_{x \in \mathbb{R}} (G_m(x) - F_n(x))$

de  $\pi(F, G)$ . Como ya vimos en 2.1.5, el estadístico de Kolmogorov-Smirnov anterior es un estimador consistente c.s. de  $\pi(F, G)$ . Su distribución asintótica es conocida y puede encontrarse en el artículo de Álvarez-Esteban (2014) [1], donde puede apreciarse la aparición de un tipo de procesos estocásticos conocidos como *puentes brownianos*. Esta distribución asintótica nos permite dar respuesta al contraste en 2.12, obteniendo de este modo las cotas de confianza de las que hablábamos antes para poder rechazar  $H_0$ . El análogo podría realizarse para estudiar el contraste 2.13 como se muestra en el artículo de Álvarez-Esteban et al. (2017) [2], de modo que la distribución asintótica de  $\gamma(F, G)$  y el estimador plug-in  $\gamma(F_n, G_m)$  nos proporcionan igualmente las cotas para poder rechazar  $H_0$ .

Solo existe un inconveniente fácil de remediar: las distribuciones asintóticas de los índices  $\pi(F, G)$  y  $\gamma(F, G)$  son complicadas de manejar en la práctica, por lo que se suele recurrir a métodos **bootstrap** que nos permiten estimarlas de una manera más sencilla. Estas técnicas, basadas en el remuestreo, nos permiten estudiar el sesgo y el error estándar de los estimadores plug-in y corregirlo cuando es necesario, obteniendo una aproximación que no puede ser mejorada en el límite asintótico.

## 2.5. Ejemplo: Comparación de la altura entre hombres y mujeres

En esta sección vamos a analizar la cuestión “¿Los hombres son más altos que las mujeres?” utilizando los conocimientos hasta ahora estudiados sobre el orden estocástico, con el propósito de mostrar el significado de cada uno de los índices  $\gamma(F, G)$ ,  $\alpha(F, G)$  y  $\pi(F, G)$  que hemos considerado. Para ello tendremos en cuenta el estudio que realizó en 2009 la OCDE (Organización para la Cooperación y el Desarrollo Económicos) y que fue publicado en el libro *Society at a Glance*, en el que se analizan algunos indicadores sociales como la altura a partir de datos procedentes de varios países. Por un lado, a partir de los datos de la talla de 11.389 varones mayores de edad se determinó que la altura de los hombres sigue una distribución aproximadamente normal con media 174,6 cm y con desviación típica 7,6. Por el otro, a partir de los datos de la talla de 11.855 mujeres mayores de edad se determinó que la altura de las mujeres sigue también una distribución aproximadamente normal con media 163,1 cm y con desviación típica 6,5.

Si entendemos esta afirmación desde el punto de vista clásico, esto es, en términos de la altura media de los individuos de la población, y si tomamos un grupo de hombres y otro grupo de mujeres elegidos completamente al azar y lo suficientemente grandes, entonces la media de la altura del grupo de hombres es mayor que la media de la altura del grupo de mujeres. Sin embargo, esta respuesta es compatible con que para un rango determinado la altura de las mujeres sea mayor que la altura de los hombres (como por ejemplo sucede con la población de 13 – 14 años, edad en la cual las mujeres son más altas que los hombres).

La cuestión sobre si los hombres son más altos que las mujeres debería detallarse mejor diciendo que un hombre bajo no sería tan bajo entre las mujeres, un hombre de altura media sería alto entre las mujeres, y un hombre alto sería todavía más alto entre las mujeres. Estas comparaciones conllevan de manera natural una posición relativa o un determinado estatus de cada individuo en

ambas poblaciones, esto es, habría que comprobar el nuevo estatus de cada uno de los elementos de la primera población considerado como uno de la segunda, y si la altura siempre es mayor podríamos concluir que los hombres son más altos que las mujeres.

Para mostrar como funcionan los índices  $\gamma$ ,  $\alpha$  y  $\pi$  aquí estudiados, vamos a considerar el siguiente ejemplo artificial formado por una muestra aleatoria de  $n = 20$  hombres extraída de una distribución  $N(174.6, 7.6)$  y otra muestra aleatoria de  $n = 20$  mujeres extraída de una distribución  $N(163.1, 6.5)$ . En lo que sigue, supongamos que  $X$  es la altura de las mujeres e  $Y$  la altura de los hombres, y que tienen funciones de distribución  $F$  y  $G$  respectivamente.

### 2.5.1. Caso 1

En primer lugar, generamos en R los datos correspondiente a los dos grupos de 20 hombres y 20 mujeres, y los ordenamos de manera creciente.

```

1 > hombres1 <- rnorm(20, 174.6, 7.6) #Generamos 20 muestras de N(174.6, 7.6)
2 > mujeres1 <- rnorm(20, 163.1, 6.5) #Generamos 20 muestras de N(163.1, 6.5)
3 > hombres11 <- sort(hombres1) #Ordenamos los datos de hombres
4 > mujeres11 <- sort(mujeres1) #Ordenamos los datos de mujeres
5 > hombres11;
6 [1] 153.6051 155.1436 165.4309 165.7174 167.4588 168.0684 168.6931 170.5956
7     171.3116 172.7600 174.7697 175.1783 178.4137 178.7812 179.6633 180.5319
8     181.0246 181.9360 184.1520 184.7797
9 > mujeres11;
10 [1] 152.3737 154.7102 156.1184 157.2270 157.9594 159.6087 159.7576 160.0784
11     160.1554 162.2034 163.1062 165.2530 165.9406 166.1446 167.0055 168.9146
12     171.2379 173.5421 175.1484 177.5970

```

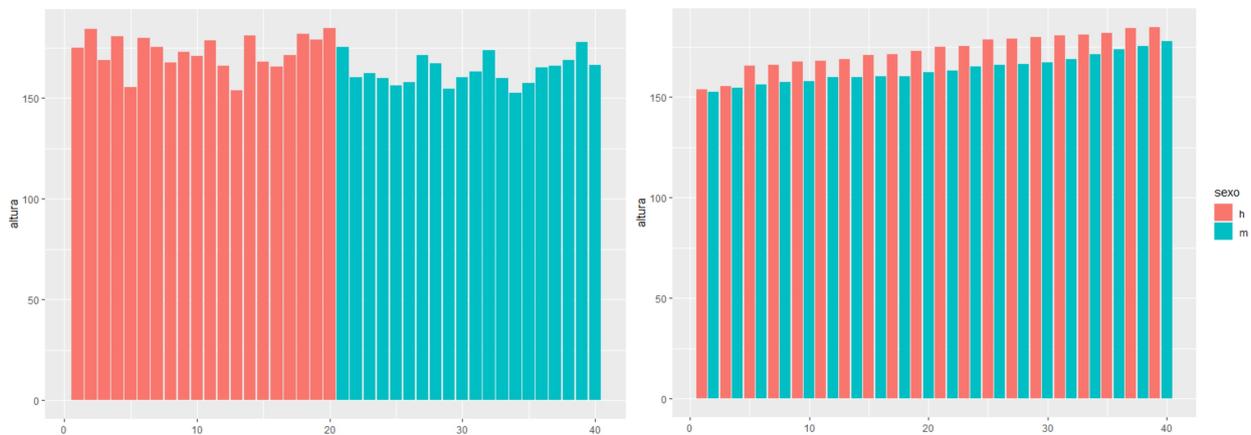


Figura 2.7: Gráficos de barras de las alturas de dos conjuntos de datos de 20 hombres (rojo) y 20 mujeres (azul). A la izquierda están yuxtapuestas las barras de las dos muestras. A la derecha se yuxtaponen las correspondientes barras ordenadas por altura de ambas muestras.

Podemos representar en un gráfico de barras múltiple los datos para poder compararlos con la función `ggplot` de R. En la figura 2.7 se representan estos valores en rojo para hombres y en azul para mujeres. En el gráfico de la izquierda podemos apreciar que las barras rojas son en general más altas que las barras azules. También se puede ver que hay algunas barras azules que son más altas que algunas rojas, y viceversa. Tal y como se presentan los datos sin ordenar, la información que podemos sacar es escasa.

El gráfico de la derecha en 2.7 nos muestra los mismos datos que antes pero ahora están yuxtapuestas la barra roja más pequeña frente a la barra azul más pequeña, la segunda contra la segunda,... y así sucesivamente. Si se considera el estatus de cada individuo como su posición según la altura en su grupo (bien sea hombre o mujer) podemos ver que cada una de las barras rojas en la fila es mayor que la barra azul que está en su mismo estatus. Si ahora normalizamos cambiando las posiciones  $1, \dots, 20$  por números reales en el intervalo  $(0, 1)$ , de modo que la posición relativa de cada individuo sea la proporción de individuos más bajos que él/ella en su grupo (hombre o mujer), obtenemos una manera general de comparar poblaciones o distribuciones a través de la comparación de los valores que toman en esta fila ordenada.

Lo que estamos realizando no es nada nuevo: al normalizar de esta manera lo que obtenemos es la representación en términos de las funciones cuantil (asociadas a las muestras), y a partir de los datos se puede ver que  $F_n^{-1}(t) \leq G_n^{-1}(t)$  para todo  $t \in (0, 1)$ . Podemos apreciar como lo que se muestra en el segundo gráfico, en el que cada barra roja es más alta que la barra azul que le acompaña, corresponde con la idea de que el primer grupo, el de los hombres, es estocásticamente mayor que el segundo, el de las mujeres. Como ya comentábamos en la subsección 2.1.2, lo que hemos hecho es precisamente calcular  $\gamma(F_n, G_n)$ . En este caso resulta que  $\gamma(F_n, G_n) = 0$ , por lo tanto sí podemos decir que existe orden estocástico entre las muestras.

### 2.5.2. Caso 2

Generamos ahora un nuevo conjunto de datos en R con las alturas de 20 hombres y 20 mujeres de la misma manera que hicimos en el caso 1.

```

1 > hombres2 <- rnorm(20, 174.6, 7.6) #Generamos 20 muestras de N(174.6, 7.6)
2 > mujeres2 <- rnorm(20, 163.1, 6.5) #Generamos 20 muestras de N(163.1, 6.5)
3 > hombres22 <- sort(hombres2) #Ordenamos los datos de hombres
4 > mujeres22 <- sort(mujeres2) #Ordenamos los datos de mujeres
5 > hombres22;
6 [1] 155.1983 155.9130 161.0123 167.7437 168.2438 168.8464 169.7357 172.1838
7     173.6051 174.0511 174.1100 174.2214 176.0723 176.2435 176.6069 177.0169
8     177.2520 179.7605 181.2738 182.0849
9 > mujeres22;
10 [1] 149.7428 159.7953 161.1894 161.8890 162.1399 162.2204 162.3946 162.9396
11     163.8284 163.8541 163.9191 164.1164 164.6412 164.6647 165.2938 166.5140
12     166.6471 168.4186 169.9858 170.4675

```

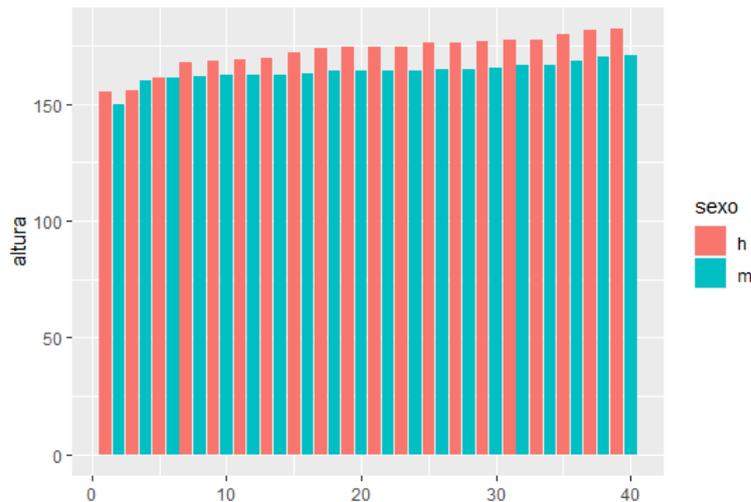


Figura 2.8: Gráficos de barras de las alturas de dos conjuntos de datos de 20 hombres (rojo) y 20 mujeres (azul). En él se yuxtaponen las correspondientes barras ordenadas por altura de ambas muestras.

Echemos un vistazo a la figura 2.8, en el que aparecen ordenados los valores de la altura de hombres y mujeres, y se han yuxtapuesto los dos más pequeños, los segundos dos más pequeños... En el gráfico podemos ver que aunque *en general* en estas nuevas muestras los hombres son más altos que las mujeres (esto es, las barras rojas son más altas que las azules) no ocurre esto para los elementos segundo y tercero. Por lo tanto, la caracterización del orden estocástico a través de las funciones cuantil nos impide concluir que la altura de los hombres es estocásticamente mayor que la de las mujeres.

Podemos, no obstante, medir de manera precisa el grado de desviación con el concepto de orden estocástico utilizando el índice  $\gamma(F_n, G_n)$  como ya comentamos en la subsección 2.1.2, y encontramos que  $\gamma(F, G) = 2/20 = 0,1$ . Utilizando las funciones cuantil correspondientes  $G^{-1}$  asociada al grupo de hombres y  $F^{-1}$  asociada al grupo de mujeres entonces podemos garantizar que si tenemos muestras  $u_1, \dots, u_n$  de una distribución uniforme en el intervalo  $(0, 1)$ , si los transformamos utilizando  $F^{-1}$  y  $G^{-1}$  obtendríamos muestras  $y_1, \dots, y_n$  de  $G$  y  $x_1, \dots, x_n$  de  $F$  que para un valor de  $n$  suficientemente grande verifican que  $\#\{x_i > y_i : i = 1, \dots, n\}/n \approx 0,1$ .

Podemos utilizar también el índice  $\alpha(F_n, G_n)$  estudiado en la sección 2.1.3 para estudiar cómo de lejos estamos de la relación de orden estocástico. Para ello haremos uso del programa `alpha()` que hemos elaborado en R (que no tiene en cuenta ningún tipo de información adicional sobre cómo es la muestra) y cuyo código se ha incluido en el apéndice C.2.

```
1 > alpha(mujeres22, hombres22)
2 [1] 0.02324539
```

Recordemos que el índice  $\alpha(F_n, G_n)$  se definía para medir la contribución de la diferencia entre las gráficas de  $F_n$  y  $G_n$ ,  $|G_n - F_n|$ , en las zonas donde no se satisface la relación de orden estocástico,  $\{G_n > F_n\}$ . Como puede verse en la gráfica de la izquierda en la figura 2.10 que aparece más adelante, el área donde  $G_n > F_n$  es pequeña comparada con el área total que hay entre las dos funciones de distribución. Con el programa `alpha()` obtenemos que  $\alpha(F_n, G_n) \approx 0,023$ , luego los datos se encuentran cerca de la relación de orden estocástico  $F \leq_{st} G$ .

Otra herramienta de la que disponemos para medir el grado de desacuerdo con la relación de orden estocástico es el índice  $\pi(F, G)$  que, como hemos visto en la subsección 2.1.4, mide el nivel de contaminación en los datos. El estimador  $\pi(F_n, G_n)$  se corresponde con el estadístico de Kolmogorov-Smirnov para un lado. Lo podemos calcular en R como sigue:

```

1 > ks.test(mujeres22, hombres22, alternative="less")
2
3       Two-sample Kolmogorov-Smirnov test
4
5 data:  mujeres22 and hombres22
6 D^- = 0.05, p-value = 0.9512
7 alternative hypothesis: the CDF of x lies below that of y

```

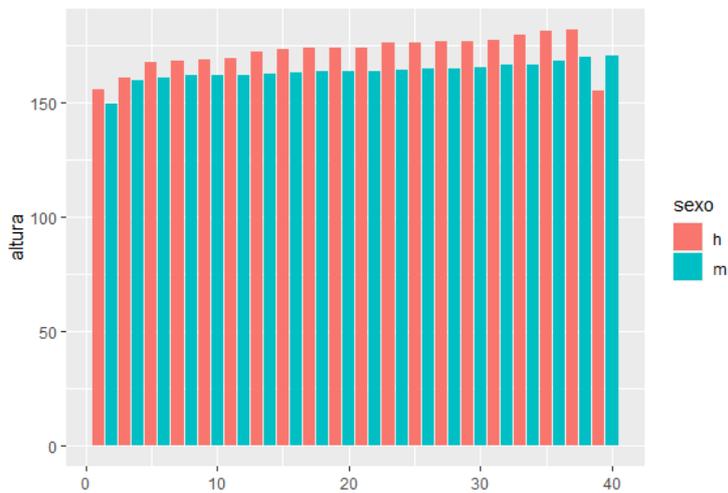


Figura 2.9: Gráficos de barras de las alturas de dos conjuntos de datos de 20 hombres (negro) y 20 mujeres (blanco). La última comparación hace referencia al dato más pequeño en hombres y al más grande en mujeres.

Según lo que hemos estudiado sobre los modelos de contaminación y el recorte de probabilidades, esto quiere decir que si quitamos una proporción de 0,05 de los chicos más bajos y de las chicas más altas, obtendríamos distribuciones recortadas  $\tilde{F}_n$  y  $\tilde{G}_n$  para las que sí se verifica la relación de orden estocástico  $\tilde{F}_n \leq_{st} \tilde{G}_n$ . Esto lo podemos ver en la figura 2.9, donde las dos barras más a la derecha hacen referencia al dato más bajo de altura de los chicos y al dato más alto de altura de las chi-

cas. Si quitamos estos dos, podemos apreciar como sí se mantiene la comparación en el resto de datos.

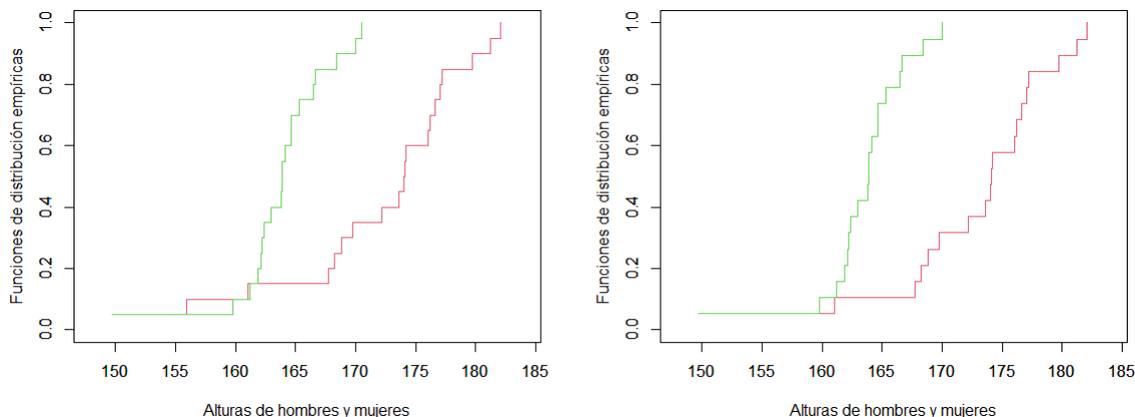


Figura 2.10: Distribuciones empíricas de la altura de mujeres ( $F_n$ , verde) y hombres ( $G_n$ , rojo) antes (izquierda) y después (derecha) del recorte

Podemos aprovechar también este cálculo para ver qué ocurre con las funciones de distribución empíricas. En la figura 2.10 se muestran las funciones de distribución empíricas para la altura de hombres y mujeres antes y después del recorte. En la gráfica de la izquierda podemos ver que no se produce orden estocástico en el sentido estudiado, ya que se producen cortes y una no domina a la otra. Dado que  $\pi(F_n, G_n) = 0,05$  ha de interpretarse como el nivel de contaminación en el modelo, podemos calcular las funciones  $\tilde{F}_n = (F_n)^\pi$  y  $\tilde{G}_n = (G_n)^\pi$  que vimos en la proposición 2.1.4 y que se obtenían quitando el dato más bajo de hombres ( $(F_n)^\pi$ ) y el dato más alto de mujeres ( $(G_n)^\pi$ ). Hecho esto obtenemos el gráfico de la derecha, donde ahora sí podemos apreciar que una función domina a la otra y por tanto que existe orden  $\tilde{F}_n \leq_{st} \tilde{G}_n$ .

## 2.6. Ejemplo real: Comparación de las edades de dos poblaciones para determinar si una está mas envejecida que la otra

Para mostrar la utilidad que nos ofrecen el orden estocástico y, más en concreto, los índices expuestos a lo largo de este capítulo, en esta sección vamos a trabajar con los censos de población de distintos países con el propósito de determinar si la población de un cierto país está más envejecida que la de otro. Para ello utilizaremos los datos que se alojan en la web <http://www.census.gov/data-tools/demo/idb/informationGateway.php>, actualizados a día 10 de diciembre de 2020. En esta página, seleccionamos "Population by Single Year Age Groups" y, según cada caso que consideremos a continuación, el país y el año que hemos elegido. Esto nos genera un conjunto de datos que contiene, entre otras cosas, el censo de población de un cierto país en un cierto año clasificados por edades. No tomaremos en consideración la división por sexos y haremos nuestro análisis sobre el número total de personas que hay de cada edad. Los datos contienen estimaciones del número de habitantes por cada año de edad desde 0 hasta 84. Después se incluye la suma de estas estimaciones en los intervalos  $[85, 89]$ ,  $[90, 94]$ ,  $[95, 99]$  y  $[100, \infty)$ . Si asignamos estos resultados a las edades 87, 92, 97 y 102, al realizar el histograma podemos observar unos picos más altos

al lado derecho correspondientes a estos grupos, lo que no nos permite realizar un análisis uniforme al mezclar grupos de varias edades con grupos que van de año en año. Por esta razón, y de manera análoga al tratamiento de estos mismos datos que se hace en el artículo de del Barrio et al. (2016) [5], hemos decidido fraccionar el número de individuos atribuidos a las edades  $a = 87, 92, 97, 102$  entre las edades  $a - 2, a - 1, a, a + 1, a + 2$  proporcionalmente a los pesos 2.3, 2.0125, 1.725, 1.4375, 1. De este modo se produce un lógico declive en la cola derecha, como puede verse en los histogramas que aparecen en la figura 2.14, y tendremos por tanto edades comprendidas entre los 0 y los 105 años.

Una vez descargado nuestro fichero con extensión `.csv` que contiene los datos del censo, podemos leerlos en R utilizando la librería `readr` y hacer el tratamiento comentado anteriormente como sigue:

```

1 pesos<-c(2.3, 2.0125, 1.725, 1.4375, 1);
2 s<-sum(pesos);
3 pais1<-read_csv(file.choose(), skip=1, col_names=TRUE);
4 edades1<-pais1$'Both Sexes Population' [2:86];
5 for(i in 1:4){
6   for(j in 1:5){
7     edades1[85+5*(i-1)+j]<-(pais1$'Both Sexes Population' [86+i]*pesos[j])%/%s
8   }
9 }
10 n<-sum(edades1);
11 # edades1 contiene los individuos divididos por edad entre 0 y 105
12 # n es el numero total de individuos en la poblacion

```

Supongamos que  $X$  es la edad de un habitante de un cierto país  $A$  en un año concreto e  $Y$  es la edad de un habitante de otro país  $B$  y otro año concretos, de modo que  $n$  es el número de habitantes en  $A$ ,  $m$  es el número de habitantes en  $B$  y  $F_n, G_m$  son las funciones de distribución empíricas obtenidas a partir de los censos por edades de cada país. Notemos aquí que también podemos aplicar el orden estocástico en la población de un determinado país para determinar si está más envejecida a lo largo del tiempo comparando el censo en años distintos, de modo que  $n$  será el número de habitantes en un año  $A$  y  $m$  el número de habitantes en otro año  $B$ . Para ver si una población está más envejecida que otra tenemos que probar que  $X \leq_{st} Y$ . Es posible que no se produzca una dominación estricta en las distribuciones de las edades, por lo cual basaremos nuestra decisión en los índices  $\pi(F_n, G_m)$ ,  $\alpha(F_n, G_m)$  y  $\gamma(F_n, G_m)$  estudiados en la sección 2.1 y que miden el grado de desviación respecto al orden estocástico. Valores pequeños de estos índices nos mostrarán que estamos cerca de la relación de orden estocástico, pudiendo dar un valor objetivo al grado de envejecimiento según el tipo de índice utilizado.

Tras estas aclaraciones, ahora la cuestión fundamental es el cálculo de  $\pi(F_n, G_m)$ ,  $\alpha(F_n, G_m)$  y  $\gamma(F_n, G_m)$  a partir de los censos de población. No parece muy lógico tratar el problema como en el ejemplo de la sección 2.5 tomando para cada individuo el dato de su edad y aplicando las funciones `gamma()`, `alpha()` y `ks.test()`, ya que ahora nuestra población puede llegar a tener

un gran número de individuos (por ejemplo, China tenía 1.339.724.851 habitantes en 2010) y estos programas no tenían en cuenta ningún tipo de información adicional acerca de cómo son las muestras.

Edad	Habitantes	Frecuencia relativa acumulada
0	$x_0$	$p_0 = \frac{x_0}{n}$
1	$x_1$	$p_1 = \frac{x_0+x_1}{n}$
2	$x_2$	$p_2 = \frac{x_0+x_1+x_2}{n}$
$\vdots$	$\vdots$	$\vdots$
105	$x_{105}$	$p_{105} = \frac{x_0+\dots+x_{105}}{n} = 1$
Total	$n$	

Tabla 2.1: Censo de población por edades de un país  $A$  con frecuencias absolutas (habitantes, segunda columna) y frecuencias relativas acumuladas (tercera columna)

Podemos aprovechar las frecuencias absolutas para cada edad que aparecen en nuestro conjunto de datos y realizar el análisis observando como son  $F_n$  y  $G_m$ . Por ejemplo, si denotamos por  $p_0, p_1, \dots, p_{105}$  a las frecuencias relativas acumuladas obtenidas a partir del censo como se muestra en la tabla 2.1,  $F_n$  es una función escalonada que vale 0 desde  $-\infty$  hasta 0,  $p_0$  en el intervalo  $[0,1]$ ,  $p_1$  en  $[1,2], \dots, p_{104}$  en  $[104,105]$  y  $1 = p_{105}$  desde 105 hasta  $+\infty$ . En la figura 2.11 podemos ver un ejemplo ideal para dos supuestas poblaciones de como máximo 3 años, donde las proporciones  $p_0, \dots, p_3$  para  $A$  son  $\frac{1}{8}, \frac{2}{8}, \frac{7}{8}, 1$  y las proporciones  $q_0, \dots, q_3$  para  $B$  son  $\frac{4}{8}, \frac{5}{8}, \frac{6}{8}, 1$ .

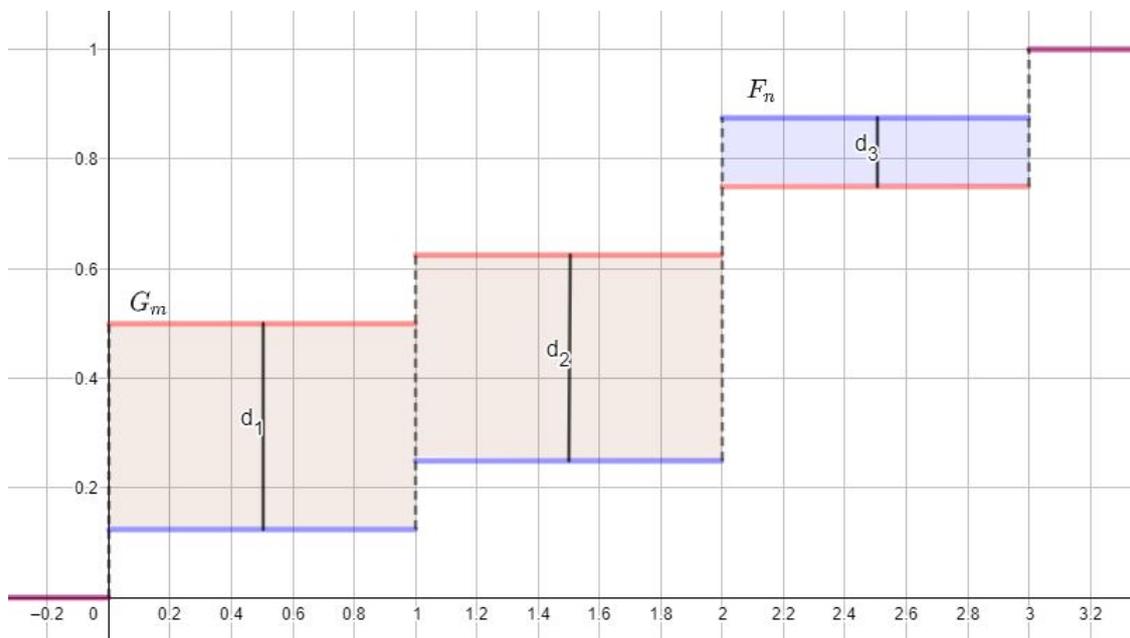


Figura 2.11: Funciones de distribución empíricas para las proporciones  $\frac{1}{8}, \frac{2}{8}, \frac{7}{8}, 1$  ( $F$ ) y  $\frac{4}{8}, \frac{5}{8}, \frac{6}{8}, 1$  ( $G$ )

A partir de lo anterior, el cálculo de  $\pi(F_n, G_m)$  es sencillo: si ahora  $q_0, \dots, q_{105}$  son las frecuencias relativas acumuladas para la otra población  $B$ , basta con calcular todas las diferencias  $d_i := q_i - p_i$  asociadas a cada intervalo  $[i, i + 1]$ ,  $i$  in  $0:104$  y tomar  $\pi(F_n, G_m)$  como el máximo de estas canti-

dades. Dados dos vectores `edades1` y `edades2` que contienen la distribución de población por edades entre 0 y 105,  $\pi(F_n, G_m)$  se puede calcular con el siguiente programa `pi_edades()` en R:

```

1 pi_edades <- function(edades1, edades2){
2   p <- cumsum(edades1) / sum(edades1);
3   q <- cumsum(edades2) / sum(edades2);
4   d <- q - p;
5   pi <- max(d);
6   pi;
7 }

```

El cálculo de  $\alpha(F_n, G_m)$  es igual de simple: al ser ambas  $F_n$  y  $G_m$  funciones escalonadas y tomar los saltos en los mismos puntos, basta con ir sumando los rectángulos de base 1 y altura  $d_i$  (luego área  $d_i$ ) teniendo en cuenta si  $d_i > 0$  ( $G_m$  encima de  $F_n$ , contribuye al numerador y denominador) o si  $d_i \leq 0$  ( $F_n$  encima de  $G_m$ , solo contribuye al denominador cambiando el signo). Para entender bien el procedimiento este procedimiento se recomienda observar el gráfico de la figura 2.11 donde aparecen marcadas estas diferencias  $d_i$  así como las zonas donde  $G_m > F_n$  (rojo, se suma al numerador y denominador) y donde  $F_n > G_m$  (solo suma al denominador). El siguiente programa `alfa_edades()` nos permite calcularlo en R:

```

1 alfa_edades <- function(edades1, edades2){
2   p <- cumsum(edades1) / sum(edades1);
3   q <- cumsum(edades2) / sum(edades2);
4   d <- q - p;
5   num_alfa <- sum(d[d > 0]);
6   denom_alfa <- sum(d[d > 0]) - sum(d[d < 0]);
7   alfa <- num_alfa / denom_alfa;
8   alfa;
9 }

```

Para hallar  $\gamma(F_n, G_m)$  necesitamos utilizar el concepto de *función cuantil empírica*. Se calcula a partir de las frecuencias relativas acumuladas  $p_0, \dots, p_{105}$ : vale 0 desde 0 hasta  $p_0$ , 1 desde  $p_0$  hasta  $p_1, \dots$ , y 105 desde  $p_{104}$  hasta  $1 = p_{105}$ . En el gráfico 2.12 pueden verse las funciones cuantil empíricas planteadas para el ejemplo ideal del que hablamos anteriormente.

Podríamos realizar un programa similar a `alpha()`, pero para el cálculo de  $\gamma(F_n, G_m)$  conviene aprovechar que  $F_n^{-1}$  y  $G_m^{-1}$  son ambas funciones escalonadas y el tamaño del salto es 1. Para hallarlo debemos medir la longitud del intervalo de aquellos valores  $t \in (0, 1)$  donde  $F_n^{-1}(t) > G_m^{-1}(t)$ . Como tenemos el mismo número de puntos de salto, 105, y como estos saltos son de tamaño 1, solo se puede producir un cruce entre  $F_n$  y  $G_m$  cuando al ordenar todas las proporciones  $p_0, \dots, p_{105}$  y  $q_0, \dots, q_{105}$  (llamémoslo  $z_0, \dots, z_{211}$ ) haya dos valores  $z_i$  y  $z_{i+1}$  seguidos de modo que  $z_i = p_j$  y  $z_{i+1} = q_j$  para algún  $j = \dots, 105$ . Si vamos comparando dos a dos todos los valores  $p_j$  y  $q_j$ , cuando la diferencia

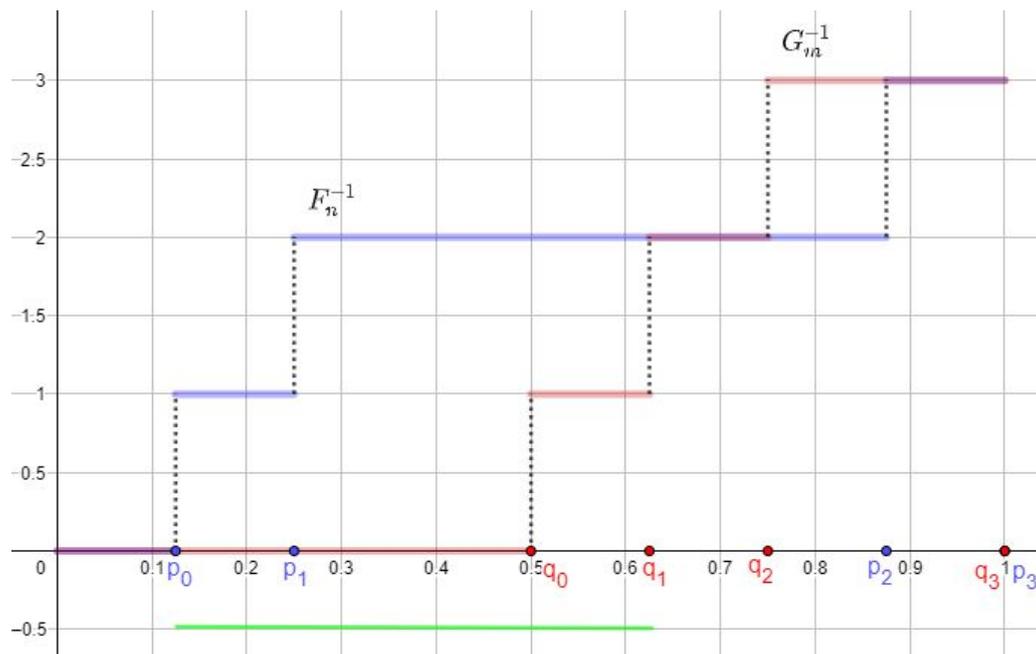


Figura 2.12: Funciones cuantil empíricas para las proporciones  $\frac{1}{8}, \frac{2}{8}, \frac{7}{8}, 1$  ( $F$ ) y  $\frac{4}{8}, \frac{5}{8}, \frac{6}{8}, 1$  ( $G$ )

$d_j = q_j - p_j$  sea positiva esto quiere decir que  $p_j < q_j$ , luego  $F_n^{-1}$  ha saltado antes y está por encima de  $G_m^{-1}$ . Dejará de estar por encima cuando haya algún  $k > j$  de modo que  $p_k > q_j$ . Por lo tanto desde  $p_j$  hasta  $q_k$  tenemos que  $F_n^{-1} > G_m^{-1}$ . Si repetimos este procedimiento  $m$  veces hasta que no encontramos más  $p_j, q_j$  verificando la condición anterior,  $\gamma(F_n, G_m)$  se calcula como la suma de las  $m$  longitudes  $q_k - p_j$  obtenidas en las  $m$  iteraciones anteriores. Lo podemos calcular con la función `gamma_edades()` en R como sigue a continuación:

```

1 gamma_edades <- function(edades1, edades2){
2   gamma=0; i=1;
3   p<-cumsum(edades1)/sum(edades1);
4   q<-cumsum(edades2)/sum(edades2);
5   while(i<=length(edades1)){
6     dif=q[i]-p[i];
7     if(dif>0){
8       inf=p[i]; sup=q[i];
9       while((i<length(edades1))&&(p[i+1]<sup)){
10        i<-i+1;
11      }
12      gamma<-gamma+(q[i]-inf);
13    }
14    i<-i+1;
15  }
16  gamma;
17 }

```

### 2.6.1. Evolución del envejecimiento en España

Podemos aprovechar la información que nos proporcionan los índices anteriores sobre la desviación respecto al orden estocástico para analizar el envejecimiento de la población española a lo largo del tiempo, utilizando a modo de ejemplo el índice  $\gamma(F_n, G_m)$ . En la página de `census.gov` disponemos de los datos por edades de los habitantes en España en cada año desde 1991 hasta la actualidad, y además se incluye una previsión de estos datos año por año hasta 2050. Para realizar nuestro análisis compararemos las poblaciones cada 5 años: 1991, 1995, 2000, 2005, 2010, 2015, 2020. Hemos añadido también las previsiones en los futuros 2030, 2040 y 2050 para ver cómo será la evolución prevista.

Una idea ingenua (a la vista de todo lo comentado en este trabajo, y en especial en la sección 2.3) para ver si el envejecimiento de la población ha sido progresivo consiste en observar la media de edad de la población en cada año y ver si esta ha ido aumentando. Estos datos se muestran en la tabla 2.2. Podemos apreciar que con el paso del tiempo la edad media de la población española ha ido aumentando, y que lo mismo ocurrirá en los años futuros. Esto nos puede dar la sensación de que el envejecimiento en la población ha ido aumentando, pero en base a los comentarios efectuados en la sección 2.3 ya sabemos que la simple comparación de medias no es la herramienta adecuada para demostrarlo sino que tal afirmación se ha de someter al concepto de orden estocástico.

Año	1991	1995	2000	2005	2010	2015	2020	2030	2040	2050
<b>Edad media de la población</b>	33.9	35.6	37.4	38.7	40.1	42.0	43.9	47.2	48.7	49.2

Tabla 2.2: Edad media de la población en los años indicados

Para estudiar la evolución del envejecimiento vamos a utilizar el índice  $\gamma(F_n, G_m)$  que hemos estudiado en este capítulo. Para ello, realizaremos la comparación en cada uno de los periodos 1991-1995, 1995-2000, 2000-2005, 2005-2010, 2010-2015, 2015-2020, 2020-2030, 2030-2040 y 2040-2050, de modo que para cada periodo calcularemos  $\gamma(F_n, G_m) = \text{gamma\_edades}(\text{edades1}, \text{edades2})$  utilizando el programa R, donde `edades1` son los datos de los habitantes por edades al inicio del periodo y `edades2` son los correspondientes al final del periodo.

En la tabla 2.3 aparecen calculados los valores de  $\gamma(F_n, G_m)$  para cada periodo que hemos elegido, donde  $F_n$  y  $G_m$  representan las distribuciones de las edades al inicio y al final del periodo, respectivamente. Estos datos se interpretan mejor utilizando el gráfico de líneas y puntos que aparece en la figura 2.13, y que muestra la evolución del índice que marca la desviación respecto al orden estocástico a lo largo del tiempo.

En los periodos 1991-1995 y 2020-2030 tenemos que  $\gamma(F_n, G_m) = 0$ , esto es, se produce orden estocástico. Esto quiere decir que al final del periodo se ha producido un envejecimiento general en toda la población respecto a la situación que había al inicio. Observemos que el valor más alto para este índice es de 0,091 y se produce en el periodo 2005-2010. Este valor, que mide el nivel de

Periodo	$\gamma(F_n, G_m)$
<b>1991-1995</b>	0
<b>1995-2000</b>	0.0006043986
<b>2000-2005</b>	0.0530607515
<b>2005-2010</b>	0.0911119579
<b>2010-2015</b>	0.0370305439
<b>2015-2020</b>	0.0030000969
<b>2020-2030</b>	0
<b>2030-2040</b>	0.0310456531
<b>2040-2050</b>	0.0599448091

Tabla 2.3: Cálculo de  $\gamma(F_n, G_m)$  en cada periodo de tiempo considerado

desacuerdo con el orden estocástico, nos muestra que durante este periodo la disconformidad con la idea de envejecimiento en los datos es mayor, lo que quiere decir que el envejecimiento ha decelerado en la población española.

Si observamos detenidamente los periodos de tiempo elegidos y los valores de  $\gamma(F_n, G_m)$  en el gráfico y lo relacionamos con los acontecimientos que han tenido lugar en España durante ese tiempo, hay un hecho fundamental que los determina: **la crisis económica que comenzó en el año 2008**.

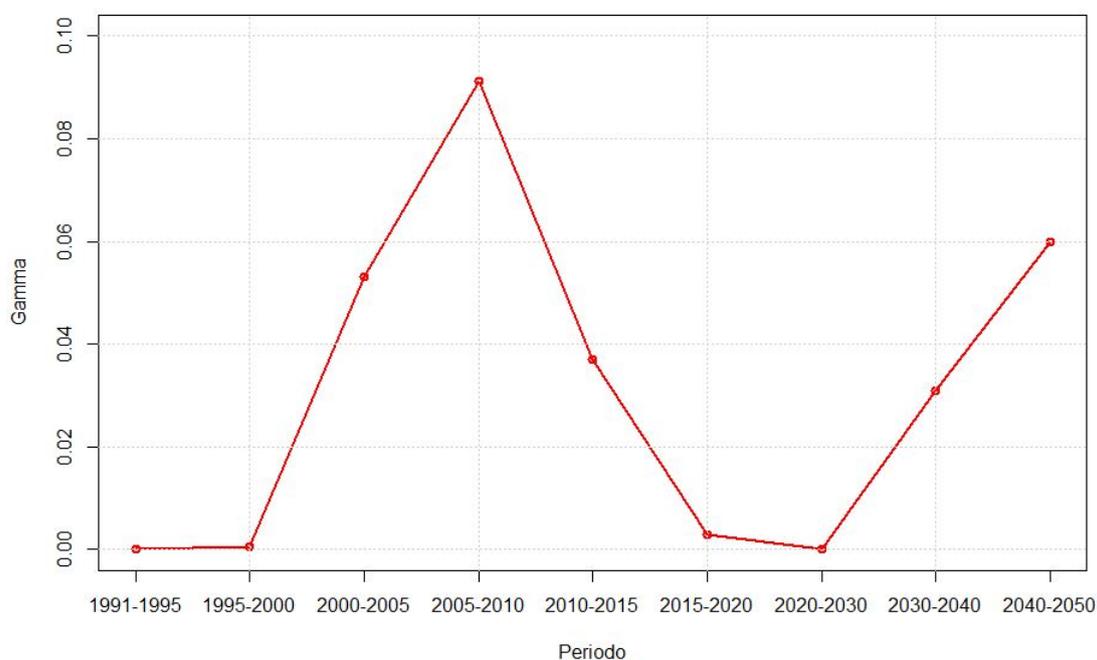


Figura 2.13: Evolución del índice  $\gamma(F_n, G_m)$  en cada periodo

Hasta el periodo de 2005-2010 se produce un aumento del índice  $\gamma(F_n, G_m)$  (cuanto mayor sea este valor, mayor desacuerdo con la idea de envejecimiento de la población), lo cual muestra que

hasta entonces el envejecimiento se ha ralentizado motivado, en cierta parte, por un aumento en la natalidad y en la inmigración (como puede verse al observar y comparar las pirámides de población de la figura 2.14 en los años 1995 y 2005, ya que la población entre 0-10 años ha tenido una tendencia creciente y la joroba intermedia ha crecido de tamaño). Sin embargo, a partir del periodo 2005-2010 en el que se incluye el inicio de la crisis económica de 2008 este índice comienza a decrecer (puede observarse por ejemplo en la pirámide de 2015 de la figura 2.14 que la tendencia de nacimientos es decreciente), hasta el punto de que en la actualidad estamos volviendo a tener un envejecimiento generalizado de la totalidad de la población. Hasta el año 2030 la situación no revertirá (en la pirámide de 2030 de la figura 2.14 se aprecia una estabilización de la natalidad); sin embargo, las previsiones que se han realizado para los periodos 2030-2040 y 2040-2050 nos muestran que el envejecimiento de la población comenzará a descender (por ejemplo, en la pirámide de 2050 en 2.14 se observa un ligero ascenso en la población joven).

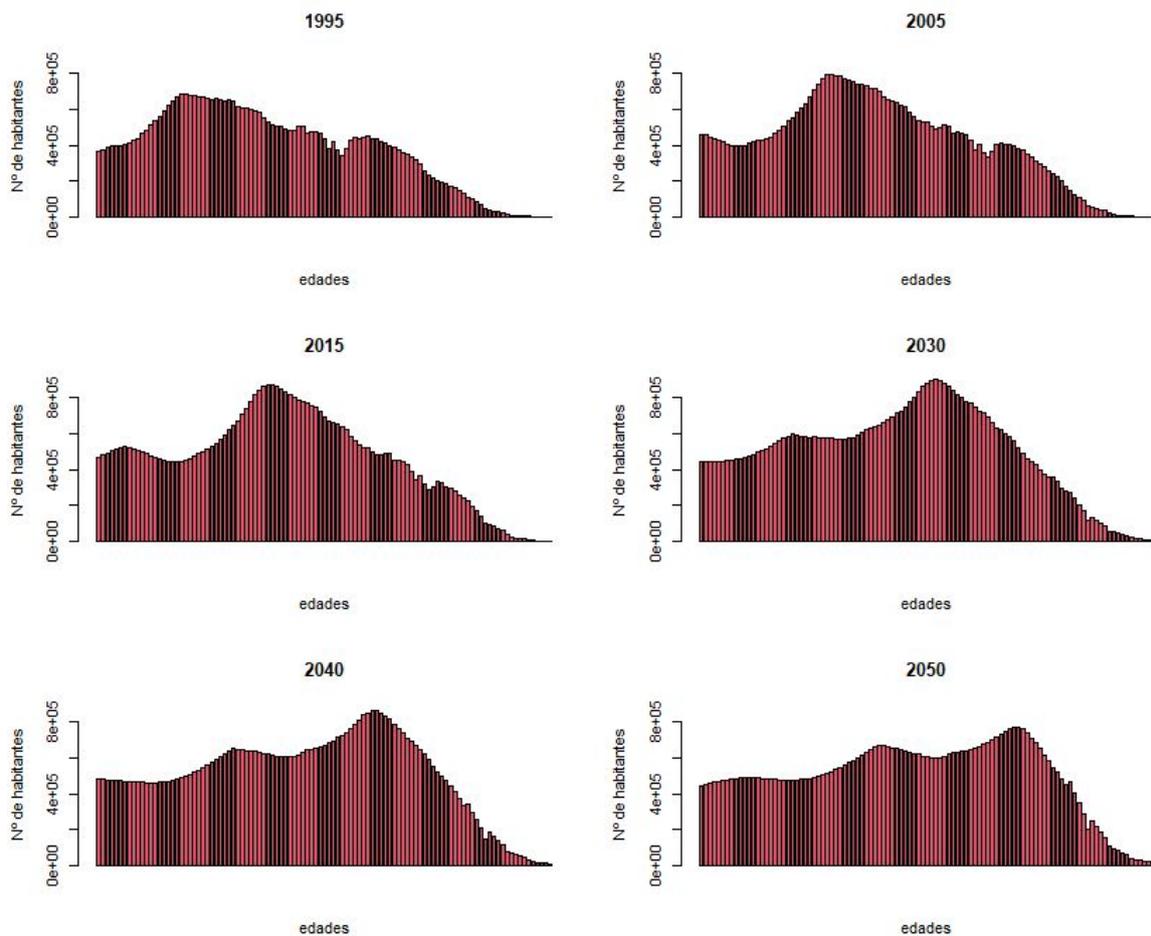


Figura 2.14: Pirámides de población en España en distintos años

Para finalizar conviene añadir que estas previsiones se hicieron a inicios de 2020, sin tener en cuenta los efectos de la pandemia mundial de COVID-19 que se desató a partir de marzo de 2020. Dejamos por tanto una cuestión en el aire: ¿las previsiones seguirán como aquí hemos visto o tendrá impacto la crisis socioeconómica provocada por la pandemia, al igual que pasó con la crisis del 2008?

## Capítulo 3

# Órdenes estocásticos convexos

El orden estocástico usual sobre el que hemos hablado en los capítulos anteriores se centra en comparar dos poblaciones (más en concreto, sus distribuciones) con el objetivo de determinar si una cierta característica puede considerarse mayor en una población que en la otra. Otro factor distinto al que podemos atender es a la variabilidad entre distribuciones, es decir, a ver cómo de dispersos están los valores que toman las variables correspondientes. Por ejemplo, imaginemos que  $X$  e  $Y$  son dos variables aleatorias que describen los beneficios ante dos grandes inversiones, sobre las que sabemos que el retorno esperado (es decir, la media) es el mismo. Un inversor con aversión al riesgo elegirá aquella que tenga menor variabilidad: aunque opte a menos ganancias, también tiene menor probabilidad de perder dinero.

La variabilidad entre variables aleatorias tiene una relación directa con los órdenes estocásticos basados en **funciones convexas**. Recordemos que una función  $f$  a valores reales se dice que es *convexa* si

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y),$$

para todos  $x$  e  $y$  y para todo  $\alpha \in [0, 1]$ . Se dice que  $f$  es *cóncava* si  $-f$  es convexa. El concepto de convexidad tiene una interpretación geométrica en funciones de variable real: si  $f$  es convexa, al tomar dos puntos cualesquiera  $x$  e  $y$  entonces la gráfica de  $f$  entre  $f(x)$  y  $f(y)$  queda por debajo de la recta que une los puntos  $f(x)$  y  $f(y)$ . Las funciones convexas y cóncavas están muy presentes en las matemáticas y en las ciencias en las que se aplican. Por ejemplo, en economía son la base de la optimización así como de la teoría de la utilidad, que estudia el grado de satisfacción que proporcionan los bienes que utiliza un consumidor. De hecho, muchos manuales sobre orden estocástico (como el de Levy [23]) realizan parte del estudio a partir de funciones de utilidad, que son un caso particular de funciones cóncavas (según nuestra definición).

**Definición 3.0.1.** Sean  $X$  e  $Y$  variables aleatorias con esperanza finita. Se dice que:

1.  $X$  es menor que  $Y$  respecto al orden estocástico convexo (escrito  $X \leq_{cx} Y$ ) si

$$Ef(X) \leq Ef(Y)$$

para toda función real  $f$  **convexa** siempre y cuando las esperanzas anteriores sean finitas;

2.  $X$  es menor que  $Y$  respecto al orden estocástico convexo creciente (escrito  $X \leq_{icx} Y$ ) si

$$Ef(X) \leq Ef(Y)$$

para toda función real  $f$  **convexa** y **creciente** siempre y cuando las esperanzas anteriores sean finitas;

3.  $X$  es menor que  $Y$  respecto al orden estocástico cóncavo creciente (escrito  $X \leq_{icv} Y$ ) si

$$Ef(X) \leq Ef(Y)$$

para toda función real  $f$  **cóncava** y **creciente** siempre y cuando las esperanzas anteriores sean finitas.

**Observación 3.0.2.** Es muy sencillo comprobar que las tres son relaciones de orden parcial utilizando el teorema de anulación para la integral. Al orden  $X \leq_{icv} Y$  se le llama en muchos manuales *segundo orden estocástico*, especialmente en aquellos en los que se utilizan funciones de utilidad. Es conveniente decir que en algunos libros (como en el de Shaked y Shanthikumar, [30]) el estudio de los órdenes  $\leq_{icx}$  y  $\leq_{icv}$  se efectúa a la inversa, dado que no existe un criterio único para establecer lo que es una función convexa y una función cóncava. Por tanto es importante revisar de qué manera se define una función convexa en cada manual, razón por la cual algunos resultados han sido adaptados a nuestra definición anterior a partir de estos.

A lo largo de todo el capítulo **asumiremos que todas las variables aleatorias tratadas tienen esperanza finita**. Si la esperanza de una variable aleatoria  $X$  no es finita y  $f$  es cualquier función convexa o cóncava, entonces  $Ef(X)$  solo puede ser finita si  $f$  es constante, en caso contrario  $Ef(X)$  es o infinito o no existe. Si esto sucede, las definiciones anteriores no tienen sentido, por lo que haremos esta suposición.

Del hecho de que  $f$  es convexa si y solo si  $-f$  es cóncava se deduce que  $X \leq_{icx} Y$  si y solo si  $-X \leq_{icv} -Y$ , por lo que solo estudiaremos aquí los resultados correspondientes al orden  $\leq_{icx}$ , los relativos al orden  $\leq_{icv}$  son análogos (solamente es cambiar el signo). Además el estudio de los órdenes convexos nos obliga a tener que recordar algunas propiedades de las funciones convexas:

- Si  $I$  es un intervalo y  $f : I \rightarrow \mathbb{R}$  es una función convexa, entonces para todo  $a \in I$  los cocientes incrementales definidos como  $f_a(x) = \frac{f(x)-f(a)}{x-a}$  para todo  $x \in I \setminus \{a\}$  son funciones crecientes;
- $f$  es derivable por la izquierda y por la derecha, y por lo tanto es continua en todo punto  $a \in I^\circ$  (esto es, en el interior del intervalo). Además para las derivadas laterales se cumple que

$$f'(a-) = \sup\{f_a(x) : x \in I, x < a\}$$

y

$$f'(a+) = \inf\{f_a(x) : x \in I, a < x\}$$

para todo  $a \in I^\circ$ .

- Por ser  $f$  convexa y la derivada creciente, se cumple que  $f'(x-) \leq f'(x+)$  para cada  $x \in I$ .

Los órdenes convexo y convexo creciente también están relacionados, como establece el siguiente teorema recogido en el libro de Müller y Stoyan [27]:

**Teorema 3.0.3.** Dadas variables aleatorias  $X$  e  $Y$ , las siguientes afirmaciones son equivalentes:

1.  $X \leq_{cx} Y$ ,
2.  $X \leq_{icx} Y$  y  $EX = EY$ .

*Demostración.* Supongamos que  $X \leq_{cx} Y$ . Es obvio que  $X \leq_{icx} Y$ , pues si  $Ef(X) \leq Ef(Y)$  para toda función real  $f$  convexa, también es cierto en particular para funciones convexas crecientes. Para ver que  $EX = EY$ , tenemos que considerar las funciones  $f(x) = x$  y  $g(x) = -x$ , ambas convexas. Por un lado,  $X \leq_{icx} Y$  implica que  $EX \leq EY$  (utilizando  $f$ ), y por otro lado,  $-EX \leq -EY$  (utilizando  $g$ ), luego al juntar ambas obtenemos la igualdad.

Recíprocamente, sea  $f$  una función convexa arbitraria. Supongamos que existe  $\alpha < \infty$  de modo que la función  $g(x) = f(x) + \alpha x$  es creciente (y además así construida convexa). Entonces  $Eg(X) = Ef(X) + \alpha EX \leq Ef(Y) + \alpha EY = Eg(Y)$  (pues se verifica  $X \leq_{icx} Y$ ). Como además  $EX = EY$ , obtenemos  $Ef(X) \leq Ef(Y)$ , luego  $X \leq_{cx} Y$ . Si no existe tal  $\alpha$ , entonces  $f$  se puede aproximar de forma monótona por

$$f_n(x) = \begin{cases} f(x) & \text{si } x \geq -n, \\ f(-n) + f'_+(-n)(x+n) & \text{en otro caso.} \end{cases}$$

donde  $f'_+$  denota la derivada por la derecha, la cual siempre existe por ser  $f$  convexa. La sucesión  $(f_n)$  está formada por funciones convexas y todas ellas son de la misma forma que  $g$  tomando  $\alpha = -f'_+(-n)$ . Utilizando el teorema de la convergencia monótona para las esperanzas, completamos la demostración.  $\square$

**Observación 3.0.4.** El teorema anterior evidencia que en el caso de que las medias sean iguales los órdenes  $\leq_{cx}$  y  $\leq_{icx}$  son equivalentes, luego los resultados que veamos para uno serán ciertos también para el otro si hay igualdad de medias. También es importante resaltar una de las implicaciones del teorema anterior: **si dos variables aleatorias no tienen la misma esperanza, entonces no son comparables en el sentido del orden  $\leq_{cx}$** . El siguiente teorema nos da la relación que existe entre el orden  $\leq_{cx}$  y la varianza de las distribuciones implicadas:

**Corolario 3.0.5.**

1. Si  $X \leq_{cx} Y$  entonces

$$EX^n \leq EY^n \text{ y } E(X - EX)^n \leq E(Y - EY)^n \text{ para } n = 2, 4, 6, \dots$$

2. Si  $X \leq_{cx} Y$  entonces  $Var(X) \leq Var(Y)$ .

3. Además, si  $X$  e  $Y$  son variables aleatorias no negativas tales que  $X \leq_{cx} Y$ , entonces

$$EX^n \leq EY^n \text{ para todo } n \in \mathbb{N}.$$

*Demostración.* Basta estudiar el carácter de las funciones  $f(x) = x^n$ . Para  $n$  par,  $f$  es convexa para todo  $x \in \mathbb{R}$ , puesto que si elegimos dos puntos de la curva la recta que los une siempre queda por encima, luego 1 es cierto (pues además se supone  $EX < \infty$ ). 2 simplemente se basa en observar que la varianza es el segundo momento de  $X$  respecto a la media. Para  $n$  impar, que  $f$  sea convexa solo sucede si  $x \geq 0$ , en cuyo caso se tiene 3.  $\square$

Es importante apreciar que tanto el orden estocástico usual como los órdenes convexo y convexo creciente tienen una caracterización en términos de sus esperanzas, dando lugar al estudio de *órdenes estocásticos integrales*. Estos se pueden describir mediante una clase  $\mathfrak{F}$  de funciones reales de la siguiente forma:

$$X \preceq Y \text{ si y solo si } Ef(X) \leq Ef(Y) \text{ para todo } f \in \mathfrak{F}.$$

En el caso del orden estocástico usual el conjunto  $\mathfrak{F}$  es el de todas las funciones reales crecientes, para el orden convexo es el de todas las funciones reales convexas y para el orden convexo creciente es el de todas las funciones reales convexas y crecientes. A partir de esto, es evidente el siguiente resultado:

**Proposición 3.0.6.** Sean  $X$  e  $Y$  variables aleatorias reales. Si  $X \leq_{st} Y$ , entonces  $X \leq_{icx} Y$ .

*Demostración.* Las funciones convexas y crecientes representan un subconjunto dentro de las funciones crecientes.  $\square$

### 3.1. Caracterizaciones de los órdenes estocásticos convexas

Al igual que hicimos con el orden estocástico usual, a continuación vamos a dar algunas caracterizaciones equivalentes de los órdenes estocásticos convexas. Dado que la definición de estos órdenes es más compleja que la del orden usual, así serán también sus caracterizaciones.

En primer lugar veamos una caracterización de los órdenes convexas en términos de las **funciones cuantil**. Para la demostración son necesarios algunos conocimientos sobre análisis funcional y reordenamientos decrecientes que recogemos en el apéndice, y que pueden verse completos en el artículo de M. A. Sordo y H. M. Ramos [31]. Conviene tener bastante en cuenta la reflexión realizada en la observación B.1.1 y en la proposición B.1.2, que afirma que si  $X$  es una v.a. con función de distribución  $F$ , entonces  $\bar{F}^{-1}(t) := F^{-1}(1 - t)$  es su reordenamiento decreciente, y se verifica que  $X \stackrel{d}{=} \bar{F}^{-1}$ .

**Teorema 3.1.1.** Sean  $X$  e  $Y$  dos variables aleatorias con funciones de distribución  $F$  y  $G$  respectivamente. Entonces  $X \leq_{icx} Y$  si y solo si

$$\int_p^1 F^{-1}(t)dt \leq \int_p^1 G^{-1}(t)dt \text{ para todo } p \in [0, 1]. \quad (3.1)$$

*Demostración.* Vamos a reescribir 3.1 en términos de  $\bar{F}^{-1}$ . Como  $\bar{F}^{-1}(t) = F^{-1}(1 - t)$ , realizando el cambio de variable  $t = 1 - s$  en 3.1 vemos que equivale a

$$\int_0^p \bar{F}^{-1}(s) ds \leq \int_0^p \bar{G}^{-1}(s) ds \text{ para todo } p \in [0, 1]. \quad (3.2)$$

Ahora sean  $(\Omega_X, \sigma_X, P_X)$  y  $(\Omega_Y, \sigma_Y, P_Y)$  los espacios de probabilidad sobre los que se definen  $X$  e  $Y$ . Tomamos  $f(\omega) = X(\omega)$  para todo  $\omega \in \Omega_X$  y  $g(\omega) = Y(\omega)$  para todo  $\omega \in \Omega_Y$ . Los reordenamientos decrecientes de  $f$  y  $g$  vienen dados por  $f^*(s) = \bar{F}^{-1}(s)$  y  $g^*(s) = \bar{G}^{-1}(s)$  para todo  $s \in [0, 1]$ . Por el teorema B.1.3 del apéndice, 3.2 equivale a

$$\int_{\Omega_X} \Psi(X) dP_X \leq \int_{\Omega_Y} \Psi(Y) dP_Y,$$

para toda función  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  convexa creciente, lo que equivale a  $X \leq_{icx} Y$ .  $\square$

**Las condiciones que definen los órdenes estocásticos convexos resultan imposibles de comprobar en la práctica**, pues hay que verificar la desigualdad  $Ef(X) \leq Ef(Y)$  para *toda* función  $f$  convexa o *toda* función  $f$  convexa y creciente. Necesitamos pues caracterizaciones equivalentes que sean más fáciles de comprobar, y en lo que sigue nos dirigiremos a probar un resultado de importancia en la práctica. Para llegar a él utilizaremos un subconjunto adecuado del conjunto de todas las funciones convexas, formado por las llamadas *funciones rampa*  $\phi_t$ , las cuales fijado  $t \in \mathbb{R}$  se definen como

$$\phi_t(x) = (x - t)_+ = \max\{x - t, 0\}.$$

La *función rampa* es una función que vale 0 si  $x$  es más pequeño que  $t$  y vale  $x - t > 0$  si  $x$  es más grande que  $t$ .

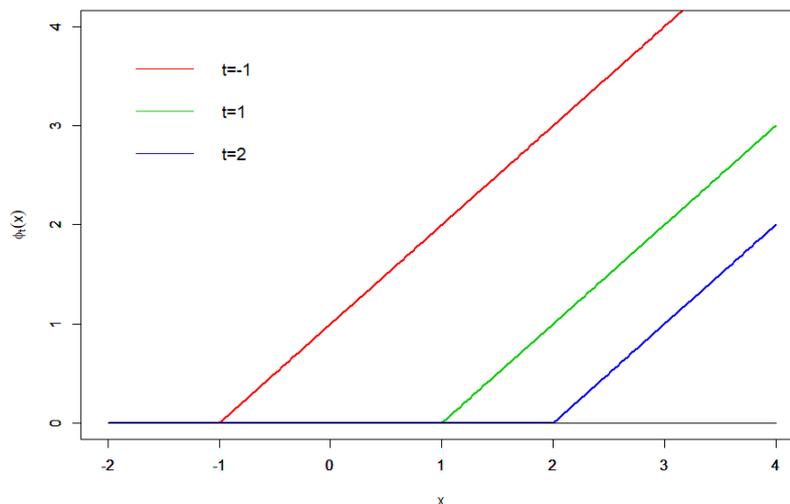


Figura 3.1: Funciones rampa  $\phi_t$

En la figura 3.1 podemos apreciar varios ejemplos de funciones rampa: son funciones convexas, crecientes y no negativas.

Para los resultados que siguen a continuación es importante tener en cuenta el siguiente teorema, que trata sobre cómo podemos aproximar funciones convexas por funciones lineales afines.

**Teorema 3.1.2.** Sea  $f : [a, b] \rightarrow \mathbb{R}$  una función convexa. Existe una colección numerable de funciones lineales  $d_i + c_i w$  para las cuales la función convexa  $\psi(w) = \sup_{i \in \mathbb{N}} (d_i + c_i w)$  verifica siempre  $\psi(w) \leq f(w)$  dándose la igualdad en todos los puntos excepto quizá en los extremos  $w = a$  o  $w = b$ , donde  $\psi(a) = f(a+)$  y  $\psi(b) = f(b-)$ .

El teorema anterior puede generalizarse para funciones  $f$  convexas definidas en todo  $\mathbb{R}$  pues solo se aplicaría la primera parte del argumento, sin preocuparnos sobre qué pasa en los extremos.

**Observación 3.1.3.** En el caso de que la función  $f$  sea convexa, positiva y creciente (imaginémos el caso de la función  $e^x$ ) podemos aproximarla mediante las *funciones rampa*  $\phi_t(x) = (x - t)_+$ . Son funciones no negativas, convexas y crecientes, se corresponderían con las rectas  $l_i(x) = c_i + d_i x$  pero al ser  $f$  positiva solo sería necesaria su parte positiva, correspondiéndose con las funciones  $\phi_t$ .

Todos los resultados necesarios de análisis convexo anteriores pueden encontrarse en el libro de Rockafellar [28]. Tomando esto como punto de partida, es de interés el siguiente teorema, que aparece en el libro de Müller y Stoyan [27]:

**Teorema 3.1.4.** Sean  $X$  e  $Y$  variables aleatorias. Son equivalentes:

1.  $X \leq_{icx} Y$ ;
2.  $E(X - t)_+ \leq E(Y - t)_+$  para todo  $t \in \mathbb{R}$ .

*Demostración.* (1)  $\Rightarrow$  (2) es trivial, puesto que las funciones  $\phi_t$  son crecientes y convexas para cualquier valor  $t$  fijado.

Para demostrar (2)  $\Rightarrow$  (1), vamos a ver que toda función creciente y convexa  $f$  se puede escribir como límite de una combinación lineal de funciones positivas  $\phi_t$ . Supongamos que  $E(X - t)_+ \leq E(Y - t)_+$  para todo  $t \in \mathbb{R}$  y sea  $f$  una función creciente y convexa. Consideraremos tres casos:

1. Si  $\lim_{t \rightarrow -\infty} f(t) = 0$ , por ser  $f$  creciente y convexa entonces también es no negativa (imagínemos  $e^x$ ). Por el teorema 3.1.2 se sabe que  $f$  es el máximo de un conjunto numerable  $\{l_1, l_2, \dots\}$  de funciones lineales y crecientes (por serlo así  $f$ ). Se define

$$f_n(t) = \max\{0, l_1(t), \dots, l_n(t)\}.$$

El teorema 3.1.2 nos asegura que  $f_n$  converge creciendo hacia  $f$ , donde cada  $f_n$  es una función lineal a trozos. Entonces  $f_n$  se puede escribir como

$$f_n(x) = \sum_{i=1}^n c_{in} (x - d_{in})_+$$

para constantes  $c_{in} \geq 0$  y  $d_{in} \in \mathbb{R}$ . Tomando esperanzas y utilizando la hipótesis

$$E f_n(X) = \sum_{i=1}^n c_{in} E(X - d_{in})_+ \leq \sum_{i=1}^n c_{in} E(Y - d_{in})_+ = E f_n(Y).$$

Basta aplicar el teorema de la convergencia monótona para concluir que  $Ef(X) \leq Ef(Y)$ .

2. Si  $\lim_{t \rightarrow -\infty} f(t) = \alpha \in \mathbb{R}$ , entonces se puede aplicar 1 a la función  $f - \alpha$ .
3. Si  $\lim_{t \rightarrow -\infty} f(t) = -\infty$ , se pueden considerar las funciones

$$f_n(x) = \text{máx}\{f(x), -n\}$$

que verifican 2 para todo  $n$ , y  $f_n$  converge a  $f$  de forma monótona. Basta pues aplicar el teorema de la convergencia monótona para concluir. □

El anterior teorema nos muestra que el orden convexo creciente se puede caracterizar a través de la función  $\pi_X(t) = E(X - t)_+$ , con  $t \in \mathbb{R}$ . Podemos calcular esta esperanza utilizando el teorema del Transfer (A.3.6 en el apéndice), a partir del cual integrando por partes obtenemos que

$$\pi_X(t) = E(X - t)_+ = \int_t^\infty (x - t)dF(x) = \int_t^\infty (1 - F(x))dx = \int_t^\infty \bar{F}(x)dx \quad (3.3)$$

A la función  $\pi_X$  la llamaremos *integral de función de supervivencia*.

Ahora estamos en condiciones de demostrar otra caracterización de los órdenes  $\leq_{st}$ ,  $\leq_{icx}$  y  $\leq_{cx}$  que nos permitirá comprobarlos fácilmente al utilizar distribuciones:

**Teorema 3.1.5.**

1.  $X \leq_{st} Y$  si y solo si la función  $t \mapsto \pi_Y(t) - \pi_X(t)$  es decreciente.
2.  $X \leq_{icx} Y$  si y solo si  $\pi_X(t) \leq \pi_Y(t)$  para todo  $t \in \mathbb{R}$ , esto es

$$\int_t^\infty [\bar{F}_Y(x) - \bar{F}_X(x)]dx \geq 0 \text{ para todo } t \in \mathbb{R}, \quad (3.4)$$

o en términos de la función de distribución,

$$\int_t^\infty [F_X(x) - F_Y(x)]dx \geq 0 \text{ para todo } t \in \mathbb{R}. \quad (3.5)$$

3.  $X \leq_{cx} Y$  si y solo si  $\pi_X(t) \leq \pi_Y(t)$  para todo  $t \in \mathbb{R}$  y  $EX = EY$ .

*Demostración.*

1. Puesto que

$$\pi_Y(t) - \pi_X(t) = \int_t^\infty [\bar{F}_Y(x) - \bar{F}_X(x)]dx \quad (3.6)$$

se tiene el orden estocástico usual  $X \leq_{st} Y$  si y solo si el integrando es positivo para todo  $t$ . Como  $t$  es el extremo inferior de la integral, al hacerlo crecer el valor de la integral solo puede disminuir, por ello la función  $t \mapsto \pi_Y(t) - \pi_X(t)$  es decreciente.

2. Este es el contenido del teorema 3.1.4 utilizando 3.3.
3. El teorema 3.0.3 nos muestra que el orden  $\leq_{cx}$  es equivalente al orden  $\leq_{icx}$  suponiendo que las esperanzas de las dos variables son iguales.

□

**Observación 3.1.6.** Como ya advertimos, debemos tener en cuenta qué definición estamos tomando como función convexa. En nuestro caso estamos siguiendo el libro de Müller y Stoyan [27], pero en otros manuales como en el de Shaked y Shanthikumar [30] o en el de Levy [23] las definiciones de función convexa y cóncava se toman a la inversa. Por tanto debemos tener precaución con los resultados que ahí aparecen formulados, y en todo caso adaptarlos a nuestra definición.

Para los órdenes convexos, lo que establece el teorema anterior es que el área acumulada debajo de  $F_Y$  no puede ser menor que el área acumulada debajo de  $F_X$ , para cualquier de  $t$  valor real hasta el cual medimos. A diferencia del orden usual, en esta ocasión **ambas funciones de distribución pueden cortarse** muchas veces con tal de que el balance entre las áreas resulte positivo para  $Y$ . En el caso de funciones de distribución que se corten una única vez tenemos un criterio mucho más potente para determinar la dominación, como veremos a continuación.

**Definición 3.1.7.** Sean  $X$  e  $Y$  variables aleatorias con funciones de distribución  $F$  y  $G$  respectivamente. Se dice que  $X$  es menos peligroso que  $Y$  (escrito  $X \leq_D Y$ ) si  $EX \leq EY$  y si existe algún valor  $t_0 \in \mathbb{R}$  tal que  $F(t) \leq G(t)$  para todo  $t < t_0$  y  $F(t) \geq G(t)$  para todo  $t \geq t_0$ .

Esta relación no es de orden pues no se cumple la propiedad transitiva, ya que si  $F$  y  $G$  se cortan una vez,  $G$  y  $H$  se cortan una vez, entonces  $F$  y  $H$  puede que se crucen dos veces. Sin embargo, es una condición suficiente para que exista orden  $\leq_{icx}$ , como establece el siguiente teorema:

**Teorema 3.1.8 (Criterio de corte).**  $X \leq_D Y$  implica  $X \leq_{icx} Y$ , esto es, si  $X$  e  $Y$  son variables aleatorias con funciones de distribución  $F$  y  $G$  respectivamente, cuando  $F$  y  $G$  se corten una sola vez y si  $F$  corta a  $G$  desde abajo,  $EX \leq EY$  es condición suficiente para que  $X \leq_{icx} Y$ , siempre que exista al menos un punto  $t_0$  tal que  $F(t) \leq G(t)$  para todo  $t \in (-\infty, t_0]$  y  $F(t) \geq G(t)$  para todo  $t \in [t_0, +\infty)$ .

*Demostración.* Observemos que

$$EY - EX = \int_{-\infty}^{\infty} [F(t) - G(t)]dt = \int_{-\infty}^{t_0} [F(t) - G(t)]dt + \int_{t_0}^{\infty} [F(t) - G(t)]dt \geq 0. \quad (3.7)$$

Para probar que  $X \leq_{icx} Y$  vamos a utilizar la caracterización del teorema 3.1.5 para este orden, luego tenemos que comprobar que  $\int_t^{\infty} [F(x) - G(x)]dx \geq 0$  para todo  $t \in \mathbb{R}$ .

Si  $t_0 \leq t$  entonces  $F(t) \geq G(t)$ , por lo tanto se tiene que  $\int_t^{\infty} [F(x) - G(x)]dx \geq 0$ .

Si  $t \leq t_0$ , entonces podemos escribir

$$\begin{aligned} \int_t^\infty [F(x) - G(x)]dx &= \int_{-\infty}^\infty [F(x) - G(x)]dx - \int_{-\infty}^t [F(x) - G(x)]dx = \\ &= \int_{-\infty}^\infty [F(x) - G(x)]dx + \int_{-\infty}^t [G(x) - F(x)]dx \geq 0, \end{aligned}$$

donde el primer sumando es positivo por 3.7 y el segundo también lo es, pues si  $t \leq t_0$  entonces  $F(t) \leq G(t)$ .  $\square$

**Observación 3.1.9.** El teorema anterior establece la dominación en el caso de que el área entre  $F$  y  $G$  a la derecha de la intersección (donde  $F > G$ ) sobrepase al área a la izquierda (donde  $F < G$ ). También se puede ver con las funciones de supervivencia, para las que tendría que suceder  $\bar{F}(x) \geq \bar{G}(x)$  para todo  $x \leq x_0$  y  $\bar{F}(x) \leq \bar{G}(x)$  para todo  $x \geq x_0$ . Este criterio se puede incluso generalizar al caso en que se produzca un número finito de cortes, como se recoge en el libro de Levy [23].

**Observación 3.1.10.** Dado que si  $EX = EY$  entonces los órdenes  $\leq_{cx}$  y  $\leq_{icx}$  son equivalentes, el criterio de corte se puede extender al orden  $\leq_{cx}$ , de modo que si  $EX = EY$  y se produce la dominación de las funciones como se dice en el teorema entonces  $X \leq_{cx} Y$ . Si rehacemos la demostración anterior, vemos que esto se traduce en que ha de ser

$$\int_{-\infty}^{t_0} [G(t) - F(t)]dt = \int_{t_0}^\infty [F(t) - G(t)]dt,$$

esto es, el área entre  $F$  y  $G$  a un lado de  $t_0$  ha de ser igual al área en el otro lado.

Este criterio de corte es muy útil para comprobar la existencia de orden  $\leq_{icx}$  entre distribuciones de probabilidad que se corten una vez ya que solamente debemos hacer una comparación de medias.

**Ejemplo 3.1.11.** Supongamos que tenemos variables  $X \sim \varepsilon(\lambda)$  e  $Y \sim N(\mu, 1)$ . Queremos ver qué condiciones hay que imponer a los parámetros  $\lambda > 0$  y  $\mu$  para determinar si existe orden  $\leq_{icx}$  o no. Para ello podemos aplicar el criterio de corte (pues ambas funciones son inyectivas y se cortan como mucho una única vez) y hacer una comparación de medias, las cuales se calculan como  $EX = \frac{1}{\lambda}$  y  $EY = \mu$ . Entonces para que  $X \leq_{icx} Y$  es necesario que

$$EX \leq EY \iff \frac{1}{\lambda} \leq \mu \iff \lambda\mu \geq 1.$$

Por ejemplo, si  $X \sim \varepsilon(2)$  e  $Y \sim N(1, 1)$  entonces como  $2 \cdot 1 \geq 1$  se tiene que  $EX \leq EY$ , por tanto por el criterio de corte (pues si  $t < 0$  ambas funciones no se cortan y si  $t \geq 0$  ambas funciones son inyectivas, luego se cortan una vez) se deduce que  $X \leq_{icx} Y$ .

En la figura 3.2 podemos interpretar la condición del teorema 3.1.5 por medio de las áreas entre ambas funciones. Si llamamos  $t_0$  al punto de corte de ambas, el área a la derecha de  $t_0$  (donde  $F_X > F_Y$ ) es mayor que el área a la izquierda de  $t_0$  (donde  $F_X < F_Y$ ) por lo tanto al restar ambas el valor de  $\pi_Y - \pi_X$  es positivo. Con este ejemplo podemos resaltar que aunque existe orden  $\leq_{icx}$ , no

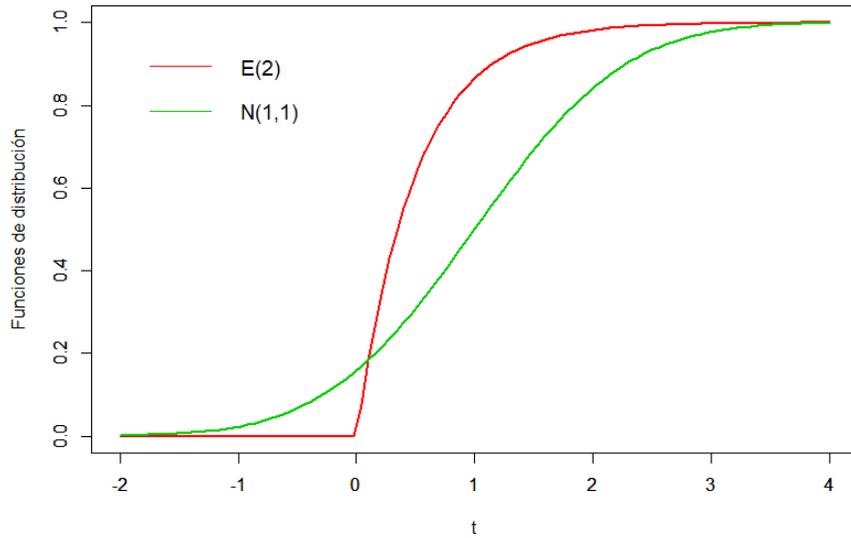


Figura 3.2: Comparación de distribución exponencial y normal para el orden convexo creciente

se produce orden  $\leq_{st}$ , ya que no se produce la dominación entre las funciones de distribución para todo  $x \in \mathbb{R}$ . Así vemos que  $\leq_{icx} \not\Rightarrow \leq_{st}$ .

### 3.2. El teorema de Strassen

El orden estocástico convexo admite también una caracterización en el contexto de las **martingalas** a través del teorema de Strassen (1965), y que nos permite trabajar con el orden estocástico convexo entre variables que están definidas sobre un mismo espacio de probabilidad. Con esta subsección se pretende además mostrar la complejidad que pueden alcanzar los resultados sobre órdenes estocásticos convexos a pesar de la simpleza en su formulación, pero que son sencillos de utilizar en la práctica.

**Teorema 3.2.1 (de Strassen).** Sean  $X$  e  $Y$  dos variables aleatorias con esperanza finita. Son equivalentes:

1.  $X \leq_{cx} Y$ ;
2. existen variables aleatorias  $\hat{X}$  e  $\hat{Y}$  definidas sobre el mismo espacio de probabilidad tales que  $X \stackrel{d}{=} \hat{X}$  e  $Y \stackrel{d}{=} \hat{Y}$  de modo que

$$E[\hat{Y}|\hat{X}] = \hat{X} \text{ c.s..} \tag{3.8}$$

Strassen demostró este resultado en un contexto mucho más general, utilizando *espacios polacos* (espacios topológicos metrizables, separables y completos) y algunos pilares fundamentales del análisis funcional como el teorema de Hahn-Banach o el teorema de representación de Riesz. Este resultado general queda fuera de los objetivos de este trabajo, pero sí se puede hacer una demostración constructiva de esta versión probabilística del teorema.

La implicación (2)  $\Rightarrow$  (1) es directa aplicando la desigualdad de Jensen para la esperanza (proposición A.2.1 del apéndice A). Dada una función  $\phi$  convexa,

$$E[\phi(X)] = E[\phi(\hat{X})] = E\phi(E[\hat{Y}|\hat{X}]) \leq E(E[\phi(\hat{Y})|\hat{X}]) = E[\phi(\hat{Y})] = E[\phi(Y)],$$

donde hemos hecho uso de la propiedad de la esperanza condicionada  $EA = E(E(A|B))$ .

El recíproco es bastante más complicado y exige el desarrollo de una artillería que puede consultarse detalladamente en el libro de R. Szekli [32], que requiere del concepto de núcleo de Markov.

**Definición 3.2.2.** Un *núcleo de Markov* es una función  $Q(x, E)$ ,  $x \in \mathbb{R}$ ,  $E \in \beta$  de modo que si se fija  $x \in \mathbb{R}$  entonces  $Q(x, \cdot)$  es una probabilidad sobre  $(\mathbb{R}, \beta)$ , y si se fija  $E \in \beta$  entonces  $Q(\cdot, E)$  es una función tal que  $\{x : Q(x, E) \leq a\} \in \beta$ , para todo  $a \in [0, 1]$  (esto es,  $Q(\cdot, E)$  es una variable aleatoria - función medible sobre  $(\mathbb{R}, \beta)$ ).

Se dice que  $Q$  *conserva la media* si la esperanza de la medida de probabilidad  $Q(x, \cdot)$  es igual a  $x$  para todo  $x \in \mathbb{R}$ , esto es, si para la correspondiente función de distribución

$$Q(x, y) := Q(x, (-\infty, y]) \text{ si } y \in \mathbb{R}$$

se tiene que

$$\int_{-\infty}^{\infty} y dQ(x, y) = \int_{-\infty}^{\infty} y Q(x, dy) = x \text{ para todo } x \in \mathbb{R}.$$

Utilizando un núcleo  $Q$  adecuado se puede obtener cualquier función de distribución  $G$  a partir de una dada  $F$  mediante

$$G(y) = \int_{-\infty}^{\infty} Q(x, (-\infty, y]) dF(x),$$

lo cual denotaremos como  $G = QF$ .

La clave de la demostración del teorema de Strassen se basa en que existe una relación entre lo anterior y el orden  $\leq_{cx}$  como muestra el siguiente teorema, el cual no demostraremos aquí pero que puede encontrarse en el artículo de Blackwell [14].

**Teorema 3.2.3.** Supongamos que  $Q$  es un núcleo de Markov que conserva la media. Si  $X$  e  $Y$  son variables aleatorias con funciones de distribución  $F$  y  $G$ , entonces  $X \leq_{cx} Y$  si y solo si  $G = QF$ .

La demostración de lo anterior se basa en que se pueden construir variables  $\hat{X}$  e  $\hat{Y}$  sobre el mismo espacio de probabilidad y una medida  $\mathbb{P}$  tal que  $\mathbb{P}(\hat{Y} \in B | \hat{X} = x) = Q(x, B)$  y  $\mathbb{P}(\hat{X} \leq x) = F(x)$  para un núcleo de Markov  $Q$  y elementos  $B \in \beta$  y  $x \in \mathbb{R}$ .

El teorema de Strassen puede extenderse también a los órdenes  $\leq_{icx}$  y  $\leq_{icv}$ , como podemos ver en el siguiente corolario que aparece en el libro de Müller y Stoyan [27].

**Corolario 3.2.4.** Dadas dos variables aleatorias  $X$  e  $Y$  con esperanza finita, son equivalentes:

1.  $X \leq_{icx}$  ( $\leq_{icv}$  resp.)  $Y$ ;

2. existen variables aleatorias  $\widehat{X}$  e  $\widehat{Y}$  definidas sobre el mismo espacio de probabilidad tales que  $X \stackrel{d}{=} \widehat{X}$  e  $Y \stackrel{d}{=} \widehat{Y}$  de modo que

$$E[\widehat{Y}|\widehat{X}] \geq (\leq \text{ resp.}) \widehat{X} \text{ c.s.} \quad (3.9)$$

**Observación 3.2.5.** Resulta bastante curiosa (a la par que interesante) la demostración que se hace en el artículo de Armbruster [4] para el teorema de Strassen y el corolario anterior. En este, primero se hace una demostración utilizando variables discretas y se trata como un problema de programación lineal, sin utilizar argumentos de teoría de la medida. Posteriormente se realiza una extensión a distribuciones generales mediante el estudio de algunos funcionales.

El teorema de Strassen y el corolario anterior se pueden utilizar para dar condiciones suficientes acerca de los órdenes estocásticos convexos en familias de localización y escala. Esto lo veremos en la extensión que haremos de los órdenes estocásticos al caso multivariante, en la sección 3.6. También lo podemos aplicar al caso de que tengamos muestras aleatorias para obtener una condición complementaria sobre la ley fuerte de los grandes números:

**Teorema 3.2.6.** Sean  $X_1, \dots, X_n$  v.a. independientes e idénticamente distribuidas. Si  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  es la media de  $X_1, \dots, X_n$ , entonces

$$\bar{X}_n \leq_{cx} \bar{X}_{n-1}.$$

*Demostración.* Como las v.as.  $X_1, \dots, X_n$  son independientes e idénticamente distribuidas, se tiene que

$$E\left(X_1 \mid \bar{X}_n\right) = \dots = E\left(X_n \mid \bar{X}_n\right).$$

Por lo tanto,

$$\begin{aligned} \bar{X}_n &= E\left(\bar{X}_n \mid \bar{X}_n\right) = E\left(\frac{1}{n} \sum_{i=1}^n X_i \mid \bar{X}_n\right) = \\ &= \frac{1}{n} \sum_{i=1}^n E\left(X_i \mid \bar{X}_n\right) = E\left(X_i \mid \bar{X}_n\right), \end{aligned}$$

para cada  $i = 1, \dots, n$ . De modo análogo a lo anterior podemos comprobar que  $E\left(\bar{X}_{n-1} \mid \bar{X}_n\right) = E\left(X_i \mid \bar{X}_n\right)$ , por lo tanto se tiene que

$$E\left(\bar{X}_{n-1} \mid \bar{X}_n\right) = \bar{X}_n.$$

Por el teorema de Strassen, esto equivale a que  $\bar{X}_n \leq_{cx} \bar{X}_{n-1}$ . □

En el artículo de Denuit y Vermandele [18] puede verse otra demostración del resultado anterior utilizando el concepto de v.as. intercambiables (una generalización de v.as. i.i.d.).

La *ley fuerte de los grandes números* (LFGN) establece que si tenemos variables aleatorias i.i.d. con esperanza finita entonces la media muestral converge hacia  $EX_1$  c.s.. Por las propiedades del orden estocástico convexo, de  $\bar{X}_n \leq_{cx} \bar{X}_{n-1}$  se sigue que  $Var(\bar{X}_n) \leq Var(\bar{X}_{n-1})$ , de modo que  $\bar{X}_n$

es un estimador para la media más *eficiente* que  $\bar{X}_{n-1}$ . El teorema 3.2.6 nos muestra además que **la convergencia de la LFGN es monótona**, en el sentido de que la variabilidad de la media muestral decrece (en términos del orden estocástico convexo) de forma monótona según aumenta el número de observaciones.

### 3.3. Órdenes estocásticos convexos en familias de localización y escala

Al igual que hicimos para el orden estocástico usual en 1.2, podemos estudiar los órdenes convexos entre distribuciones pertenecientes a la misma familia de localización y escala. Solo tiene sentido para familias cuya esperanza es finita, puesto que sino los órdenes convexos no tienen sentido, y podemos suponer que el elemento de referencia  $\mathcal{F}_0$  tiene esperanza igual a 0. La clave reside en el siguiente teorema:

**Teorema 3.3.1.** Sea  $X$  una variable aleatoria con  $EX = 0$ . Si  $0 \leq a \leq c$  y  $b \leq d$ , entonces  $aX + b \leq_{icx} cX + d$ .

*Demostración.* En primer lugar, se tiene que  $F_{aX+b}(t) = F_X(\frac{t-b}{a})$  y  $F_{cX+d}(t) = F_X(\frac{t-d}{c})$ . Siguiendo el mismo argumento que en 1.6 sobre la existencia de un punto de corte  $t_0$  se tiene que  $F_{aX+b}(t) \leq F_{cX+d}(t)$  si y solo si  $t \leq t_0 = \frac{bc-ad}{c-a}$  y  $F_{aX+b}(t) \geq F_{cX+d}(t)$  si y solo si  $t \geq t_0$ . En segundo lugar, como  $b \leq d$  y  $EX = 0$ , por la linealidad de la esperanza tenemos que  $E(aX + b) = b < d = E(cX + d)$ . Por tanto se cumple que  $aX + b \leq_D cX + d$  y por el criterio de corte concluimos que  $aX + b \leq_{icx} cX + d$ .  $\square$

El teorema anterior nos muestra las condiciones que tenemos que poner en una familia de localización y escala para que exista orden  $\leq_{icx}$ . Si  $X$  e  $Y$  son v.a. con funciones de distribución  $F_{\theta_1, \lambda_1}$  y  $F_{\theta_2, \lambda_2}$  respectivamente pertenecientes a la misma familia de localización y escala basada en un elemento  $\mathcal{F}_0$  con esperanza finita e igual a 0, entonces si  $Z$  tiene como función de distribución  $\mathcal{F}_0$  sucede que  $X = \lambda_1 Z + \theta_1$  y que  $Y = \lambda_2 Z + \theta_2$ . Tenemos el siguiente resultado:

**Proposición 3.3.2.** Sean  $X$  e  $Y$  variables aleatorias con esperanza finita pertenecientes a la misma familia de localización y escala con parámetros de localización  $\theta_1, \theta_2$  y parámetros de escala  $\lambda_1, \lambda_2$  respectivamente. Si  $\theta_1 \leq \theta_2$  y  $\lambda_1 \leq \lambda_2$  entonces  $X \leq_{icx} Y$ .

Además, como los órdenes estocásticos convexos solo tienen sentido si las variables tienen esperanza finita, se puede tomar la media como parámetro de localización. En la proposición 3.0.3 vimos que los órdenes  $\leq_{icx}$  y  $\leq_{cx}$  son equivalentes en el caso de que las esperanzas de las variables sean iguales. Esto sirve de justificación para lo siguiente:

**Proposición 3.3.3.** Sean  $X$  e  $Y$  variables aleatorias con esperanza finita pertenecientes a la misma familia de localización y escala con parámetros de localización  $\theta_1 = EX, \theta_2 = EY$  y parámetros de escala  $\lambda_1, \lambda_2$  respectivamente. Si  $\theta_1 = \theta_2$  y  $\lambda_1 \leq \lambda_2$  entonces  $X \leq_{cx} Y$ .

En el caso de familias cuyos dos primeros momentos sean finitos, se pueden tomar estos como los parámetros de localización y escala. Dadas variables  $X$  e  $Y$  con medias  $\mu_X, \mu_Y$  finitas y desviaciones típicas  $\sigma_X, \sigma_Y$  finitas, entonces **las condiciones suficientes para que  $X \leq_{icx} Y$  son  $\mu_X \leq \mu_Y$**

y  $\sigma_X \leq \sigma_Y$ . Por lo anterior tenemos que el orden  $\leq_{cx}$  es equivalente al orden  $\leq_{icx}$  si y solo si  $EX = EY$ , luego **para que se produzca orden  $\leq_{cx}$  se ha de pedir que  $\mu_X = \mu_Y$  y que  $\sigma_X \leq \sigma_Y$ .**

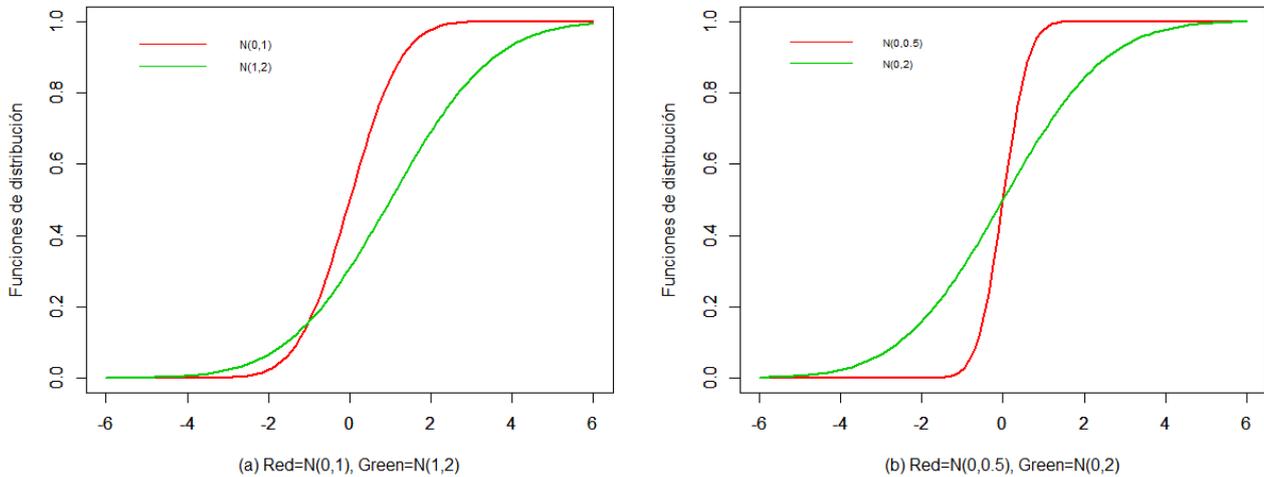


Figura 3.3: Comparación de variables normales según los órdenes convexo y convexo creciente

**Ejemplo 3.3.4.** Los teoremas 3.3.2 y 3.3.3 se pueden aplicar a la familia de localización y escala formada por las distribuciones normales. En la figura 3.3 podemos ver dos ejemplos donde sí existe orden  $\leq_{icx}$  entre distribuciones. En (a), si  $X \sim N(0, 1)$  e  $Y \sim N(1, 2)$  se tiene que  $\mu_X = 0 < 1 = \mu_Y$  y  $\sigma_X = 1 < 2 = \sigma_Y$ , luego  $X \leq_{icx} Y$ . En (b), si  $X \sim N(0, \frac{1}{2})$  e  $Y \sim N(0, 2)$  se tiene que  $\mu_X = 0 = \mu_Y$  y  $\sigma_X = \frac{1}{2} < 2 = \sigma_Y$ , luego  $X \leq_{icx} Y$ , y como  $EX = EY$  en este caso se tiene además que  $X \leq_{cx} Y$ . Podemos apreciar además todas las condiciones del criterio de corte sobre cómo se produce la dominación a ambos lados de la intersección, y cómo el área a la derecha es mayor que el área a la izquierda (en el segundo caso iguales).

**Observación 3.3.5.** Ya comentamos que el orden estocástico convexo estaba muy relacionado con la varianza de las distribuciones, y lo podemos ver en el anterior ejemplo, puesto que si  $\sigma_X \leq \sigma_Y$  entonces  $\sigma_X^2 = Var(X) \leq Var(Y) = \sigma_Y^2$  y por el teorema anterior si  $EX = EY$  se tiene que  $X \leq_{cx} Y$ . Así si las medias son iguales, el orden estocástico convexo hace menor aquella variable que tiene menor varianza, resultando de gran aplicación en áreas como las finanzas o la economía.

### 3.4. Consecuencias de los órdenes estocásticos convexos

A diferencia del orden estocástico usual, en general el orden convexo no es cerrado para la convergencia en distribución. Debemos añadir hipótesis adicionales, como recoge el siguiente teorema tomado del libro de Müller y Stoyan [27]:

**Proposición 3.4.1.** Sean  $(X_n)$  e  $(Y_n)$  sucesiones de variables aleatorias con  $X_n \leq_{icx} Y_n$  para cada  $n \in \mathbb{N}$ . Si  $X_n \xrightarrow{d} X$  e  $Y_n \xrightarrow{d} Y$  y si además  $E(X_n)_+ \rightarrow EX_+$  y  $E(Y_n)_+ \rightarrow EY_+$  entonces  $X \leq_{icx} Y$ .

*Demostración.* Por el Teorema 3.1.4 es suficiente probar que  $E(X - t)_+ \leq E(Y - t)_+$  para todo  $t$  real. Fijemos  $t$ . Por hipótesis  $X_n \leq_{icx} Y_n$  equivale a  $E(X_n - t)_+ \leq E(Y_n - t)_+$ . Como  $(x - t)_+ = \max\{x, t\} - t$  y  $-\max\{x, t\} = \min\{-x, -t\}$ , se tiene que  $E(X_n - t)_+ = E(X_n)_+ - Eg_t(X_n)$ , donde  $g_t(x) = \min\{x, t\}$  es una función continua y acotada. La convergencia en distribución implica que  $Eg_t(X_n) \rightarrow Eg_t(X)$  por ser  $g$  continua. Como hemos supuesto que  $E(X_n)_+ \rightarrow EX_+$  se concluye que  $E(X - t)_+ \leq E(Y - t)_+$ .  $\square$

Como hemos visto el orden estocástico convexo está muy relacionado con la varianza de las distribuciones. Según el corolario 3.0.5, se debe interpretar en el sentido de que la variable dominante tiene mayor varianza que la dominada. También vimos que si las esperanzas de dos variables no eran iguales, entonces no eran comparables según el orden convexo  $\leq_{cx}$ . Veremos ahora qué ocurre si las varianzas son iguales:

**Proposición 3.4.2.** Sean  $X$  e  $Y$  variables aleatorias tales que  $X \leq_{cx} Y$  y  $Var(X) = Var(Y)$ . Entonces  $X$  e  $Y$  tienen la misma distribución.

*Demostración.* Podemos calcular  $Var(X) = EX^2 - (EX)^2$ . Haciendo integración por partes en  $EX^2$  obtenemos que

$$\begin{aligned} EX^2 &= \int_{-\infty}^{\infty} t^2 dF(t) = \int_{-\infty}^0 2t(-F_X(t))dt + \int_0^{\infty} 2t(1 - F_X(t))dt = \\ &= \int_{-\infty}^0 2(\pi_X(t) + t - EX)dt + \int_0^{\infty} 2\pi_X(t)dt. \end{aligned}$$

Suponiendo que  $EX = EY$  llegamos a que

$$EY^2 - EX^2 = 2 \int_{-\infty}^{\infty} (\pi_Y(t) - \pi_X(t))dt.$$

Si  $X \leq_{cx} Y$  entonces es cierto que  $EX = EY$ , y como además  $Var(X) = Var(Y)$  se tiene

$$2 \int_{-\infty}^{\infty} (\pi_Y(t) - \pi_X(t))dt = EY^2 - EX^2 = Var(Y) - Var(X) = 0.$$

pues  $Var(X) = E(X^2) - (EX)^2$  y como las esperanzas son iguales se restan. El orden  $X \leq_{cx} Y$  implica que  $\pi_Y - \pi_X > 0$  por el teorema 3.1.5. Tenemos pues que la integral de una función positiva es igual a 0, por lo tanto la función ha de ser igual a 0 c.s. Finalmente, por la expresión en 3.3 las dos funciones de supervivencia han de ser iguales, por lo que  $X =^d Y$ .  $\square$

El resultado anterior admite una formulación mucho más fuerte, como muestra la siguiente proposición tomada del artículo de Denuit et al. [17]:

**Proposición 3.4.3.** Si  $X$  e  $Y$  son dos variables aleatorias tales que  $X \leq_{cx} Y$  y  $E\phi(X) = E\phi(Y)$  para alguna función  $\phi$  estrictamente convexa de modo que las esperanzas anteriores son finitas, entonces  $X$  e  $Y$  tienen la misma distribución. El resultado sigue siendo cierto si cambiamos  $\leq_{cx}$  y “convexa” por  $\leq_{icx}$  y “creciente y convexa”.

Es también importante destacar que el orden convexo se conserva si sumamos variables aleatorias independientes, como muestran los siguientes resultados que aparecen en el libro de Shaked y Shanthikumar [30]:

**Proposición 3.4.4.** Si  $X \leq_{cx}$  (resp.  $\leq_{icx}$ )  $Y$  y  $Z$  es una variable aleatoria independiente de  $X$  e  $Y$ , entonces

$$X + Z \leq_{cx} \text{ (resp. } \leq_{icx} \text{)} Y + Z.$$

*Demostración.* Sea  $f$  una función convexa. Definimos  $g(z) = Ef(X + z)$  y  $h(z) = Ef(Y + z)$ . La aplicación  $x \mapsto f(x + z)$  es convexa para todo  $z$  real, luego si  $X \leq_{cx} Y$  entonces  $g(z) \leq h(z)$  para todo  $z \in \mathbb{R}$  (consiste en hacer una traslación). La esperanza es monótona, luego

$$Ef(X + Z) = Eg(Z) \leq Eh(Z) = Ef(Y + Z).$$

La segunda parte se prueba de forma análoga tomando  $f$  convexa y creciente. □

**Corolario 3.4.5.** Sean  $X_1, \dots, X_m$  un conjunto de variables aleatorias independientes e  $Y_1, \dots, Y_m$  otro conjunto de variables aleatorias independientes. Si  $X_i \leq_{cx}$  (resp.  $\leq_{icx}$ )  $Y_i$  para cada  $i = 1, \dots, m$  entonces

$$\sum_{j=1}^m X_j \leq_{cx} \text{ (resp. } \leq_{icx} \text{)} \sum_{j=1}^m Y_j$$

*Demostración.* Consiste en aplicar reiteradamente el argumento del teorema anterior. □

### 3.5. Contrastando el orden estocástico convexo

En el capítulo anterior vimos que la relación de orden estocástico usual  $X \leq_{st} Y$  nos aporta mucha más información sobre las distribuciones que realizar una comparación de las medias  $\mu_X$  y  $\mu_Y$  de dos poblaciones cuando se quiere comprobar que una cierta característica es mayor en una que en la otra. A lo largo de este capítulo ya hemos comentado que el orden estocástico convexo está relacionado con la varianza de las distribuciones. De modo análogo, podemos también deducir que la relación  $X \leq_{cx} Y$  que hemos estudiado aquí nos proporciona más información que el mero hecho de comparar las varianzas y comprobar que  $Var(X) \leq Var(Y)$ .

Utilizando las caracterizaciones que hemos visto en la sección 3.1, podemos estudiar algunos índices de desviación que nos permitirán ver cómo de lejos estamos de la relación de orden estocástico convexo. Supongamos que  $X$  e  $Y$  son dos variables aleatorias integrables con funciones de distribución  $F$  y  $G$  respectivamente. A partir de la caracterización dada en el teorema 3.1.1 por medio de las funciones cuantil  $F^{-1}$  y  $G^{-1}$ , en el artículo de Berrendero y Cárcamo [9] se define el índice  $\Delta_\omega(F, G)$  como

$$\Delta_\omega(F, G) := \int_0^1 \left( G^{-1}(t) - F^{-1}(t) \right) \omega(t) dt, \tag{3.10}$$

siempre que la integral anterior tenga sentido y donde  $\omega$  es una función a valores reales definida en el intervalo  $[0, 1]$ . En el terreno de las métricas probabilísticas, el índice anterior puede considerarse

como una *métrica de Wasserstein con signo y con peso*.

El teorema que sigue a continuación, tomado del artículo de Berrendero y Cárcamo [9] (2012), muestra que el orden convexo  $F \leq_{cx} G$  se caracteriza por el hecho de que el índice  $\Delta_\omega(F, G)$  es no negativo para cualquier función peso  $\omega$  no decreciente. Además, si  $X$  e  $Y$  son variables con distribuciones distintas para las que se produce la dominación,  $\Delta_\omega(F, G)$  ha de ser necesariamente positivo para cualquier función peso  $\omega$  estrictamente creciente.

**Teorema 3.5.1.** Sea  $\mathcal{I} := \{\omega : [0, 1] \rightarrow \mathbb{R} \text{ no decrecientes}\}$  y sea  $\mathcal{I}_0 := \{\omega \in \mathcal{I} : \omega(0) \geq 0\}$ . Además sea  $\mathcal{I}^* := \{\omega \in \mathcal{I} \text{ estrictamente crecientes}\}$  e  $\mathcal{I}_0^* := \mathcal{I}_0 \cup \mathcal{I}^*$ . Entonces:

1.  $X \leq_{cx} Y$  si y solo si  $\Delta_\omega(F, G) \geq 0$ , para toda  $\omega \in \mathcal{I}$ . La equivalencia sigue siendo cierta si “ $\leq_{cx}$ ” e “ $\mathcal{I}$ ” se cambian por “ $\leq_{icx}$ ” e “ $\mathcal{I}_0$ ”, respectivamente.
2. Si  $X \leq_{cx} Y$  y  $\Delta_\omega(F, G) = 0$  para alguna  $\omega \in \mathcal{I}^*$ , entonces  $X$  e  $Y$  tienen la misma distribución. El resultado sigue siendo cierto si “ $\leq_{cx}$ ” e “ $\mathcal{I}^*$ ” se cambian por “ $\leq_{icx}$ ” e “ $\mathcal{I}_0^*$ ”, respectivamente.

Cuando tratamos con muestras, procediendo análogamente a como vimos en el capítulo 2, el test para probar la existencia de orden estocástico convexo debería plantearse para contrastar

$$H_0 : X =^d Y \text{ contra } H_a : X \leq_{cx} Y. \quad (3.11)$$

Bajo la hipótesis nula  $H_0$ ,  $X$  e  $Y$  están igualmente distribuidas y  $\Delta_\omega(F, G) = 0$ , mientras que bajo la hipótesis alternativa  $H_a$ ,  $X \leq_{cx} Y$  y  $\Delta_\omega(F, G) > 0$ , para cada  $\omega \in \mathcal{I}$ . De este modo, podemos observar que para cada función  $\omega$  no decreciente el índice  $\Delta_\omega(F, G)$  es una medida de discrepancia para el test. Una vez que la función  $\omega$  no decreciente está fijada, el test se determina calculando el índice de desviación  $\Delta_\omega(F, G)$  entre las dos distribuciones y se rechaza  $H_0$  cuando  $\Delta_\omega(F, G)$  es suficientemente grande. Un test análogo se podría considerar para el orden  $\leq_{icx}$ , utilizando una función  $\omega$  estrictamente creciente.

Para trabajar con muestras es necesario utilizar el concepto de **mayorización** que trataremos en el capítulo 4 de este trabajo y unas funciones peso  $\omega$  específicas obtenidas a partir de los datos. La información detallada acerca de cómo se calculan puede consultarse en el artículo de Berrendero y Cárcamo [10] (2011).

En otro artículo de Berrendero y Cárcamo [9] (2012) se propone también otro punto de vista distinto para tratar el contraste de hipótesis planteado en 3.11. Utilizando el teorema 3.4.3 que vimos en la sección anterior, los autores plantean también el índice

$$\Lambda_\phi(X, Y) := E\phi(Y) - E\phi(X), \quad (3.12)$$

como medida de discrepancia para el test, donde  $\phi$  es una función estrictamente convexa. En el caso de que  $X$  e  $Y$  sean v.a. con distribuciones distintas,  $X \leq_{cx} Y$  solo puede ocurrir si  $E\phi(X) \neq E\phi(Y)$  para alguna función estrictamente convexa. Por lo tanto, una vez fijada  $\phi$  estrictamente convexa, el

test se determina calculando el índice  $\Lambda_\phi(X, Y)$  entre las dos distribuciones y se rechaza  $H_0$  cuando sea distinto de cero. Lo mismo se puede plantear para el orden  $\leq_{icx}$  pero utilizando una función  $\phi$  estrictamente creciente y estrictamente convexa (piénsese en la función  $e^x$ ).

### 3.6. Órdenes convexos multivariantes

En este caso no surge ningún problema para extender los órdenes convexos univariantes al caso multivariante, y siguen estando relacionados con la comparación de la variabilidad entre vectores aleatorios.

**Definición 3.6.1.** Sean  $\mathbf{X}$  e  $\mathbf{Y}$  vectores aleatorios con esperanza finita. Se dice que:

1.  $\mathbf{X}$  es menor que  $\mathbf{Y}$  respecto al orden estocástico convexo (escrito  $\mathbf{X} \leq_{cx} \mathbf{Y}$ ) si

$$Ef(\mathbf{X}) \leq Ef(\mathbf{Y})$$

para toda función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  **convexa** siempre y cuando las esperanzas anteriores sean finitas;

2.  $\mathbf{X}$  es menor que  $\mathbf{Y}$  respecto al orden estocástico convexo creciente (escrito  $\mathbf{X} \leq_{icx} \mathbf{Y}$ ) si

$$Ef(\mathbf{X}) \leq Ef(\mathbf{Y})$$

para toda función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  **convexa** y **creciente** siempre y cuando las esperanzas anteriores sean finitas;

3.  $\mathbf{X}$  es menor que  $\mathbf{Y}$  respecto al orden estocástico cóncavo creciente (escrito  $\mathbf{X} \leq_{icv} \mathbf{Y}$ ) si

$$Ef(\mathbf{X}) \leq Ef(\mathbf{Y})$$

para toda función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  **cóncava** y **creciente** siempre y cuando las esperanzas anteriores sean finitas.

Se puede demostrar el siguiente resultado que es análogo al Teorema de Strassen univariante:

**Teorema 3.6.2.**

1.  $\mathbf{X} \leq_{cx} \mathbf{Y}$  si y solo si existen vectores aleatorios definidos sobre un mismo espacio de probabilidad  $\widehat{\mathbf{X}} =^d \mathbf{X}$  e  $\widehat{\mathbf{Y}} =^d \mathbf{Y}$  tales que  $E[\widehat{\mathbf{Y}}|\widehat{\mathbf{X}}] = \widehat{\mathbf{X}}$  c.s..
2.  $\mathbf{X} \leq_{icx} \mathbf{Y}$  si y solo si existen vectores aleatorios definidos sobre un mismo espacio de probabilidad  $\widehat{\mathbf{X}} =^d \mathbf{X}$  e  $\widehat{\mathbf{Y}} =^d \mathbf{Y}$  tales que  $E[\widehat{\mathbf{Y}}|\widehat{\mathbf{X}}] \geq \widehat{\mathbf{X}}$  c.s..
3.  $\mathbf{X} \leq_{icv} \mathbf{Y}$  si y solo si existen vectores aleatorios definidos sobre un mismo espacio de probabilidad  $\widehat{\mathbf{X}} =^d \mathbf{X}$  e  $\widehat{\mathbf{Y}} =^d \mathbf{Y}$  tales que  $E[\widehat{\mathbf{Y}}|\widehat{\mathbf{X}}] \leq \widehat{\mathbf{X}}$  c.s..

Utilizando el teorema de Strassen se pueden demostrar resultados como los siguientes, que pueden encontrarse en el artículo de Müller [26]:

**Proposición 3.6.3.** Sean  $\mathbf{X}$  y  $\mathbf{X}'$  vectores aleatorios n-dimensionales con distribuciones normales  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$  respectivamente. Las siguientes condiciones son equivalentes:

1.  $\mathbf{X} \leq_{cx} \mathbf{X}'$ ;
2.  $\boldsymbol{\mu} = \boldsymbol{\mu}'$  y  $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}'$  es una matriz definida positiva.

*Demostración.* Si  $\boldsymbol{\mu} = \boldsymbol{\mu}'$  y  $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}'$  es una matriz definida positiva, entonces  $\mathbf{X}'$  tiene la misma distribución que  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$  donde  $\mathbf{X}$  y  $\mathbf{Z}$  son independientes y  $\mathbf{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}')$ . Se cumple que  $E[\mathbf{Y}|\mathbf{X}] = \mathbf{X}$ , y como  $\mathbf{X}' \stackrel{d}{=} \mathbf{Y}$  por lo tanto  $\mathbf{X} \leq_{cx} \mathbf{X}'$ .

Recíprocamente, supongamos que  $\mathbf{X} \leq_{cx} \mathbf{X}'$ . Como las funciones  $f(\mathbf{x}) = x_i$  y  $g(\mathbf{x}) = -x_i$  son convexas para  $i = 1, \dots, n$ , por la definición de orden  $\leq_{cx}$  utilizando las funciones  $f$  y  $g$  obtenemos que  $\boldsymbol{\mu} = \boldsymbol{\mu}'$ . Supongamos ahora que  $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}'$  no fuera definida positiva, esto es, existe un vector  $\mathbf{a} \in \mathbb{R}^n$  tal que  $\mathbf{a}^T(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}')\mathbf{a} < 0$ . Entonces

$$E(\mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu}))^2 = \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a} > \mathbf{a}^T \boldsymbol{\Sigma}' \mathbf{a} = E(\mathbf{a}^T(\mathbf{X}' - \boldsymbol{\mu}))^2.$$

Esto va en contra de que  $\mathbf{X} \leq_{cx} \mathbf{X}'$ , puesto que la función  $f(\mathbf{x}) = (\mathbf{a}^T(\mathbf{X} - \boldsymbol{\mu}))^2$  es convexa.  $\square$

De forma similar se puede probar la siguiente proposición:

**Proposición 3.6.4.** Sean  $\mathbf{X}$  y  $\mathbf{X}'$  vectores aleatorios n-dimensionales con distribuciones normales  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$  respectivamente. Entonces

1. Si  $\boldsymbol{\mu} \leq \boldsymbol{\mu}'$  y  $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}'$  es definida positiva, entonces  $\mathbf{X} \leq_{icx} \mathbf{X}'$ .
2. Si  $\mathbf{X} \leq_{icx} \mathbf{X}'$ , entonces  $\boldsymbol{\mu} < \boldsymbol{\mu}'$  y  $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}'$  es una matriz semidefinida positiva para vectores positivos, esto es, se tiene que  $\mathbf{a}^T(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}')\mathbf{a} \geq 0$  para todo  $\mathbf{a} \geq \mathbf{0}$ .

*Demostración.* La primera parte es análoga al teorema anterior. Respecto a la segunda parte, si  $\mathbf{X} \leq_{icx} \mathbf{X}'$  entonces  $\boldsymbol{\mu} = \boldsymbol{\mu}'$  pues las funciones  $f(\mathbf{x}) = x_i$  son convexas y crecientes para todo  $i = 1, \dots, n$ . Sea  $\mathbf{a} \geq \mathbf{0}$ . Entonces la función  $f_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}^T \mathbf{x})$  es convexa y creciente para cualquier función  $f : \mathbb{R} \rightarrow \mathbb{R}$ . En particular, para la identidad,  $\mathbf{X} \leq_{icx} \mathbf{X}'$  implica que  $\mathbf{a}^T \mathbf{X} \leq_{icx} \mathbf{a}^T \mathbf{X}'$  y necesariamente  $Var(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a} \leq \mathbf{a}^T \boldsymbol{\Sigma}' \mathbf{a} = Var(\mathbf{a}^T \mathbf{X}')$ , obteniendo el resultado.  $\square$



## Capítulo 4

# Mayorización

En esta sección veremos cómo podemos comparar vectores aleatorios para determinar cómo de *similares* son analizando cómo de dispersas están sus componentes. La bibliografía básica que se ha seguido aquí ha sido el libro de Marshall y Olkin [25]. En los resultados que estudiaremos a continuación podremos observar que se hace uso de técnicas propias del álgebra, del análisis y de la probabilidad, mostrando que todas ellas comparten puntos en común.

Dado un vector  $\mathbf{x}^T = (x_1, \dots, x_n)^T$  denotaremos por

$$x_{[1]} \geq \dots \geq x_{[n]}$$

a las componentes de  $\mathbf{x}$  ordenadas de manera decreciente, y denotaremos por

$$x_{(1)} \leq \dots \leq x_{(n)}$$

a las componentes de  $\mathbf{x}$  ordenadas de manera creciente. Además, sean

$$\mathbf{x}_{\downarrow}^T = (x_{[1]}, \dots, x_{[n]})^T \text{ y } \mathbf{x}_{\uparrow}^T = (x_{(1)}, \dots, x_{(n)})^T$$

el reordenamiento decreciente y el reordenamiento creciente del vector  $\mathbf{x}$ , respectivamente.

El concepto de *mayorización* surgió a partir de la idea de medir la desigualdad de los salarios o de las riquezas entre dos poblaciones. Para ello, consideremos una población de  $n$  individuos, y sea  $x_i$  la riqueza del individuo  $i$ -ésimo,  $i = 1, \dots, n$ . Si los ordenamos del más pobre al más rico obtenemos la ordenación  $x_{(1)} \leq \dots \leq x_{(n)}$ . Podemos sumar las riquezas de los  $k$  individuos más pobres, de modo que  $S_0 = 0$  y  $S_k = \sum_{i=1}^k x_{(i)}$ ,  $k = 1, \dots, n$ . Ahora vamos a representar los puntos  $(\frac{k}{n}, \frac{S_k}{S_n})$ ,  $k = 0, 1, \dots, n$  y vamos a unir los que sean consecutivos mediante segmentos hasta obtener una curva que conecta el origen con el punto  $(1, 1)$ . A esta curva se le llama *curva de Lorenz*.

Si la riqueza de la población se distribuye de manera uniforme entre la población, entonces la curva de Lorenz se correspondería con una recta, como podemos ver en la figura 4.1, curva A. Si esto no sucede, entonces la curva es convexa y queda por debajo de la recta anterior. En efecto,

$S_k - S_{k-1} = x_{(k)}$ , y cada uno de los estadísticos  $x_{(k)}$  es mayor que los anteriores  $x_{(k-1)}$ ,  $k = 1, \dots, n$ . Si suponemos  $S_n = 1$  y consideramos los cocientes incrementales

$$\frac{S_k - S_{k-1}}{\frac{k}{n} - \frac{k-1}{n}} = \frac{x_{(k)}}{\frac{1}{n}} = nx_{(k)}$$

podemos observar que estos son crecientes. Como la curva es lineal a trozos, la derivada es constante en cada subintervalo y es creciente, por lo tanto ha de ser convexa. Cuanto más abajo quede el codo de la curva, más concentrada está la riqueza en menor número de personas. Por ejemplo, de nuevo en la figura 4.1 la curva C queda por debajo de la curva B, luego la concentración de la riqueza es más desigual en C que en B. Esto quiere decir que en la población B la cantidad de personas pobres es menor que en C.

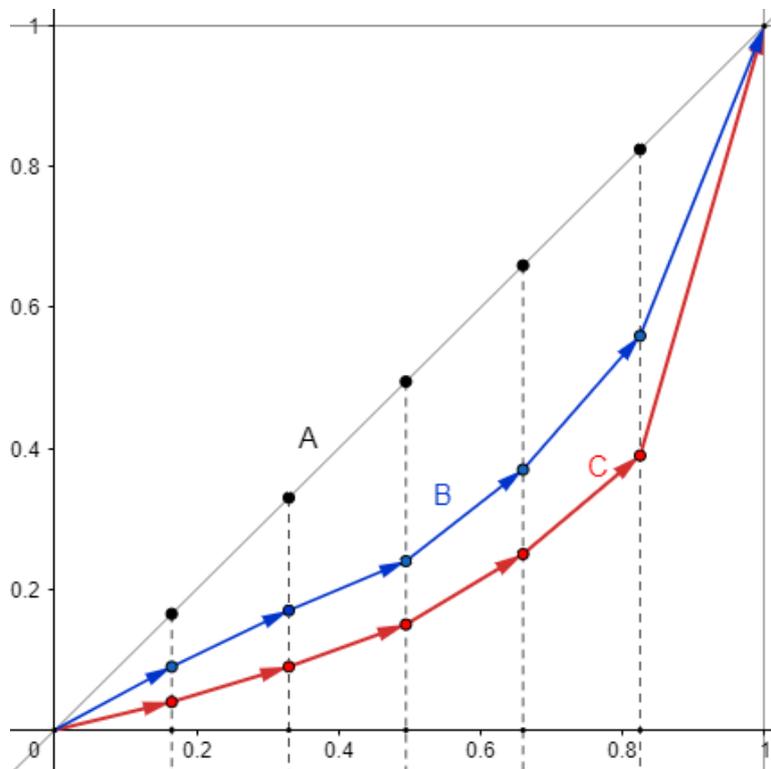


Figura 4.1: Curvas de Lorenz de la riqueza de varias poblaciones

Si  $\mathbf{x}^T = (x_1, \dots, x_n)^T$  representa la distribución de la riqueza total  $T$  de una población B y si  $\mathbf{y}^T = (y_1, \dots, y_n)^T$  representa la distribución de la misma riqueza total  $T$  de otra población C, entonces según la idea de Lorenz la distribución de  $\mathbf{x}$  es más uniforme que la distribución de  $\mathbf{y}$  si y solo si

$$\sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)} \text{ para } k = 1, \dots, n-1,$$

y obviamente,

$$\sum_{i=1}^n x_{(i)} = \sum_{i=1}^n y_{(i)} = T.$$

Esta es la idea que se corresponde con el concepto de mayorización.

**Definición 4.0.1.** Dados vectores aleatorios  $\mathbf{x}^T = (x_1, \dots, x_n)^T$  e  $\mathbf{y}^T = (y_1, \dots, y_n)^T$  de  $\mathbb{R}^n$  se dice que  $\mathbf{x}$  está mayorizado por  $\mathbf{y}$  (escrito  $\mathbf{x} \leq_M \mathbf{y}$ ) si y solo si

$$\begin{cases} \sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)} & \text{para } k = 1, \dots, n-1; \\ \sum_{i=1}^n x_{(i)} = \sum_{i=1}^n y_{(i)}, \end{cases}$$

o, equivalentemente

$$\begin{cases} \sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]} & \text{para } k = 1, \dots, n-1; \\ \sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}. \end{cases}$$

**Observación 4.0.2.** La segunda parte correspondería a sumar los valores de los  $k$  individuos más ricos, lo cual es equivalente a lo explicado anteriormente salvo que en este caso la curva de Lorentz sería cóncava. Para ver que son equivalentes ambas definiciones basta observar que  $x_{(j)} = x_{[n-j+1]}$  para cada  $j = 1, \dots, n$  y hacer un cambio de índices.

La relación  $\mathbf{x} \leq_M \mathbf{y}$  es solamente un *preorden* entre vectores aleatorios. No se cumple la propiedad antisimétrica, ya que si  $\mathbf{x} \leq_M \mathbf{y}$  e  $\mathbf{y} \leq_M \mathbf{x}$  entonces ambos tienen las mismas componentes pero no tienen por qué estar colocadas en el mismo orden, esto es,  $\mathbf{x} = P\mathbf{y}$  para alguna matriz de permutación  $P$ . Si decimos que  $\mathbf{x} \sim \mathbf{y}$  siempre que  $\mathbf{x} = P\mathbf{y}$  para alguna matriz de permutación  $P$ , entonces  $\sim$  define una relación de equivalencia sobre  $\mathbb{R}^n$ . De este modo, si consideramos el espacio cociente  $\mathbb{R}^n / \sim$  resultante entonces  $\leq_M$  define un orden parcial sobre este espacio. Esta relación también es un orden parcial si nos restringimos al subconjunto  $\{\mathbf{x} \in \mathbb{R}^n : x_1 \geq \dots \geq x_n\}$ .

## 4.1. Principio de transferencia

Un aspecto muy importante en la mayorización es el *principio de transferencia*, que fue estudiado en el contexto de intercambios de riqueza por Pigou en 1912 en su libro *Wealth and Welfare* y descrito por Dalton en 1920 en su artículo *The measurement of the inequality of incomes*. Si la población consta de solo dos individuos y se produce un intercambio de dinero de la persona más rica a la más pobre, entonces de este modo se disminuye la desigualdad. Hay una condición evidente: el tamaño de la transferencia no debe ser tan grande como para que, al recibirlo, la persona más pobre supere en riqueza al rico, ni tampoco como para crear igualdad entre ambos. Esto mismo se puede generalizar a una población con  $n$  individuos: una persona rica  $j$  puede dar dinero a una persona más pobre  $k$  de modo que no haya cambios en el orden de la cantidad de riqueza. En términos matemáticos, si estos individuos son  $j$  y  $k$  con  $1 \leq j, k \leq n$ ,  $\delta > 0$  es la cantidad de riqueza transferida de modo que

$$x_i = \begin{cases} y_i - \delta & \text{para } i = j; \\ y_i + \delta & \text{para } i = k; \\ y_i & \text{en otro caso,} \end{cases} \quad (4.1)$$

y siempre que

$$x_j \geq x_k. \quad (4.2)$$

**Definición 4.1.1.** Si se cumple 4.1 y 4.2, diremos que  $\mathbf{x}$  se ha obtenido a partir de  $\mathbf{y}$  por medio de una *T-transformación*.

Parece intuitivo que a través de estos intercambios de riqueza, es decir, si  $\mathbf{x}$  se obtiene a partir de  $\mathbf{y}$  por medio de una T-transformación, entonces  $\mathbf{x} \leq_M \mathbf{y}$ . Pero el recíproco también se cumple, como muestra el siguiente teorema que hemos adaptado de Marshall y Olkin [25]:

**Teorema 4.1.2.** Las siguientes condiciones son equivalentes:

1.  $\mathbf{x} \leq_M \mathbf{y}$ ;
2.  $\mathbf{x}$  se puede obtener a partir de  $\mathbf{y}$  por medio de un número finito de T-transformaciones.

*Demostración.* Supongamos que  $\mathbf{x} \leq_M \mathbf{y}$ . Vamos a ver que toda T-transformación se puede describir mediante una aplicación lineal cuya matriz tiene la forma

$$T = \lambda I + (1 - \lambda)Q \quad (4.3)$$

donde  $0 \leq \lambda \leq 1$ ,  $I$  es la matriz identidad y  $Q$  es una matriz de permutación que intercambia dos coordenadas (supongamos que intercambia las coordenadas  $j$  y  $k$ ).

Al aplicarlo a un vector cualquiera  $\mathbf{z}^T = (z_1, \dots, z_n)^T$  se tiene que

$$T\mathbf{z} = (z_1, \dots, z_{j-1}, \lambda z_j + (1 - \lambda)z_k, z_{j+1}, \dots, z_{k-1}, \lambda z_k + (1 - \lambda)z_j, z_{k+1}, \dots, z_n)^T. \quad (4.4)$$

Podemos suponer que  $\mathbf{x}$  no se obtiene a partir de  $\mathbf{y}$  intercambiando coordenadas (pues la demostración sería trivial con una matriz de permutación) y que  $x_1 \geq \dots \geq x_n$  e  $y_1 \geq \dots \geq y_n$  sin pérdida de generalidad.

Sea  $j$  el índice más grande de modo que  $x_j < y_j$  y sea  $k > j$  el índice más pequeño (pero mayor que  $j$ ) de modo que  $x_k > y_k$ . (Es decir, un lugar donde se produce un desequilibrio en las riquezas). Tal par  $j, k$  debe existir, pues el índice  $i$  más grande para el cual  $x_i \neq y_i$  ha de verificar  $x_i > y_i$  por la condición de mayorización. Por la elección de  $j$  y  $k$  se cumple que

$$y_j > x_j \geq x_k > y_k. \quad (4.5)$$

Tomemos  $\delta = \min(y_j - x_j, x_k - y_k)$  y definimos  $1 - \lambda = \frac{\delta}{y_j - y_k}$ , y sea

$$\mathbf{y}^{*T} = (y_1, \dots, y_{j-1}, y_j - \delta, y_{j+1}, \dots, y_{k-1}, y_k + \delta, y_{k+1}, \dots, y_n)^T. \quad (4.6)$$

Por 4.5 se cumple que  $0 < \lambda < 1$  y además

$$\mathbf{y}^{*T} = \lambda \mathbf{y}^T + (1 - \lambda)(y_1, \dots, y_{j-1}, y_k, \dots, y_{k-1}, y_j, y_{k+1}, \dots, y_n)^T.$$

Luego  $\mathbf{y}^* = T\mathbf{y}$  donde  $T = \lambda I + (1 - \lambda)Q$  y  $Q$  intercambia las coordenadas  $j$  y  $k$ . Por la expresión en 4.6,  $\mathbf{y}^* \leq_M \mathbf{y}$ . También es cierto que  $\mathbf{x} \leq_M \mathbf{y}^*$ . Para ello notemos que, por construcción,

$$\begin{aligned} \sum_{i=1}^v y_i^* &= \sum_{i=1}^v y_i \geq \sum_{i=1}^v x_i \quad \text{si } v = 1, \dots, j-1; \\ y_j^* &\geq x_j, y_i^* = y_i \quad \text{si } i = j+1, \dots, k-1; \\ \sum_{i=1}^v y_i^* &= \sum_{i=1}^v y_i \geq \sum_{i=1}^v x_i \quad \text{si } v = k+1, \dots, n; \\ \sum_{i=1}^n y_i^* &= \sum_{i=1}^n y_i = \sum_{i=1}^n x_i \end{aligned}$$

Dados dos vectores  $\mathbf{u}, \mathbf{v}$ , denotemos por  $d(\mathbf{u}, \mathbf{v})$  al número de veces que  $u_i - v_i \neq 0$ . Puesto que  $y_j^* = x_j$  si  $\delta = y_j - x_j$  e  $y_k^* = x_k$  si  $\delta = y_k - x_k$  estamos consiguiendo entonces que dos coordenadas sean iguales, por lo tanto  $d(\mathbf{x}, \mathbf{y}^*) \leq d(\mathbf{x}, \mathbf{y}) - 1$ . Por lo tanto, podemos obtener  $\mathbf{x}$  a partir de  $\mathbf{y}$  en un número finito de T-transformaciones (pasando entre medias por  $n - 1$  vectores como  $\mathbf{y}^*$ ).  $\square$

**Observación 4.1.3.** De hecho, si  $\mathbf{x} \leq_M \mathbf{y}$  entonces  $\mathbf{x}$  se puede obtener a partir de  $\mathbf{y}$  aplicando a lo sumo  $n - 1$  T-transformaciones, puesto que  $d(\mathbf{u}, \mathbf{v}) \leq n$  y  $d(\mathbf{u}, \mathbf{v}) \neq 1$  (pues no puede suceder que  $\sum u_i \neq \sum v_i$ ). Partiendo de  $\mathbf{y}$  iríamos construyendo aproximaciones sucesivas  $\mathbf{y}_1^*, \dots, \mathbf{y}_{n-1}^*$  hasta llegar al vector  $\mathbf{x}$ . Por lo tanto, para los resultados que siguen, basta demostrarlo para  $n = 2$ , puesto que para valores de  $n$  mayores se obtiene como composición de  $n - 1$  T-transformaciones.

**Definición 4.1.4.** Una matriz  $n \times n$  con entradas no negativas  $\mathbb{P} = (p_{i,j})$  se dice que es *doblemente estocástica* si verifica que la suma de cada una sus filas y la suma de cada una sus columnas es 1, esto es

$$\sum_{i=1}^n p_{i,j} = 1 \quad \text{para } 1 \leq j \leq n$$

y

$$\sum_{j=1}^n p_{i,j} = 1 \quad \text{para } 1 \leq i \leq n.$$

Las matrices doblemente estocásticas juegan un papel muy importante en las cadenas de Markov, ya que si un proceso tiene una matriz de transición  $\mathbb{P}$  con tal característica, entonces la probabilidad de salir de cualquier estado es la misma que la de volver a entrar en él. Es sencillo comprobar que una matriz es doblemente estocástica: ha de cumplirse que  $\mathbb{P}\mathbf{1} = \mathbf{1}$  y que  $\mathbf{1}^T\mathbb{P} = \mathbf{1}^T$  donde  $\mathbf{1} = (1, \dots, 1)$  (es simplemente la condición de que sumen 1 filas y columnas). Estas matrices también están relacionadas con la mayorización, como muestra el siguiente teorema cuya demostración ha sido ampliada del libro de R. Bhatia [11]:

**Teorema 4.1.5.** Las siguientes condiciones son equivalentes:

1.  $\mathbf{x} \leq_M \mathbf{y}$ ;
2. Existe una matriz doblemente estocástica  $\mathbb{P}$  de modo que  $\mathbf{x} = \mathbb{P}\mathbf{y}$ .

*Demostración.* (1)  $\Rightarrow$  (2): se basa en el teorema 4.1.2. Las T-transformaciones son aplicaciones lineales cuya matriz es doblemente estocástica (puede verse fácilmente en la expresión 4.4 al comprobar que  $\mathbb{P}\mathbf{1} = \mathbf{1}$  y que  $\mathbf{1}^T\mathbb{P} = \mathbf{1}^T$ ). Como el producto de matrices doblemente estocásticas vuelve a ser una matriz doblemente estocástica, para aplicar varias T-transformaciones lo que tenemos que hacer es multiplicar sus matrices. Por lo tanto  $\mathbb{P}$  es el producto de las matrices que aparecen en 4.1.2, y es una matriz doblemente estocástica que verifica  $\mathbf{x} = \mathbb{P}\mathbf{y}$ .

(2)  $\Rightarrow$  (1): Supongamos que  $\mathbb{P}$  es una matriz doblemente estocástica que verifica  $\mathbf{x} = \mathbb{P}\mathbf{y}$ . Para probar que  $\mathbf{x} \leq_M \mathbf{y}$  podemos suponer sin pérdida de generalidad que las coordenadas de  $\mathbf{x}$  e  $\mathbf{y}$  están ordenadas de manera decreciente. Notemos que para todo  $k$  con  $1 \leq k \leq n$  tenemos que

$$\sum_{j=1}^k x_j = \sum_{j=1}^k \sum_{i=1}^n p_{ij} y_i.$$

Si tomamos  $t_i = \sum_{j=1}^k p_{ij}$ , entonces  $0 \leq t_i \leq 1$  y

$$\sum_{i=1}^n t_i = \sum_{i=1}^n \sum_{j=1}^k p_{ij} = \sum_{j=1}^k \left( \sum_{i=1}^n p_{ij} \right) = \sum_{j=1}^k 1 = k.$$

Por otro lado,

$$\sum_{j=1}^k x_j = \sum_{i=1}^n \left( \sum_{j=1}^k p_{ij} \right) y_i = \sum_{i=1}^n t_i y_i.$$

Utilizando lo anterior tenemos que

$$\begin{aligned} \sum_{j=1}^k x_j - \sum_{j=1}^k y_k &= \sum_{i=1}^n t_i y_i - \sum_{j=1}^k y_k = \\ &= \sum_{i=1}^n t_i y_i - \sum_{j=1}^k y_k + \left( k - \sum_{i=1}^n t_i \right) y_k = \\ &= \sum_{j=1}^k (t_i - 1)(y_i - y_k) + \sum_{i=k+1}^n t_i (y_i - y_k) \leq 0 \end{aligned}$$

Recordemos que hemos elegido las coordenadas de  $\mathbf{y}$  en orden decreciente, por tanto el primer sumatorio es negativo ( $0 \leq t_i \leq 1$ , luego  $t_i - 1 \leq 0$  y si  $i \leq k$  entonces  $y_i \geq y_k$ ) y el segundo también lo es (en esta ocasión, como  $i > k$ , se tiene que  $y_i \leq y_k$ ). Además en el caso que sea  $k = n$  la diferencia entre ambos es 0, pues  $\mathbb{P}$  es una matriz doblemente estocástica,  $t_i = 1$  y en el segundo sumatorio no hay términos.  $\square$

## 4.2. Relación entre la mayorización y los órdenes estocásticos convexos

Existe otra caracterización en términos de funciones convexas encontrada en el libro de Belzunce, Martínez-Riquelme y Mulero [7]:

**Teorema 4.2.1.** Son equivalentes:

1.  $\mathbf{x} \leq_M \mathbf{y}$ ;
2.  $\sum_{i=1}^n g(x_i) \leq \sum_{i=1}^n g(y_i)$  para toda función  $g$  convexa.

Este teorema nos muestra que **la mayorización se podría interpretar como una versión discreta de los órdenes estocásticos convexos**, ahora en el caso de probabilidades con soporte finito al sustituir la integral por la suma. Es posible extender el concepto de mayorización a variables continuas, es decir, que tienen una función de densidad. Para ello es necesario utilizar los contenidos del apéndice B.1 sobre análisis real y sobre reordenamientos decrecientes, en concreto el teorema B.1.3. Si denotamos por  $f^*$  al reordenamiento decreciente de una función de densidad  $f$  (recordemos que  $\int f(x)dx = 1$ ), entonces se dice que  $f$  está mayorizada por  $g$  (escrito  $f \leq_M g$ ) si

$$\int_0^t f^*(x) dx \leq \int_0^t g^*(x) dx \text{ para todo } t > 0.$$

Como podemos ver, la condición sobre las sumas parciales al ordenar las componentes de forma decreciente que define a la mayorización en el caso discreto se corresponde con la idea de los reordenamientos decrecientes en el caso continuo. Si restringimos a variables aleatorias, esta dualidad se puede apreciar también en los resultados que ya hemos visto (como la caracterización del orden  $\leq_{cx}$  en términos de las funciones cuantil, que involucraba los reordenamientos decrecientes). Por ejemplo, el teorema anterior que se establece en términos de la suma equivale a la definición del orden estocástico convexo en el que se trataba con la integral. El siguiente teorema, que se puede consultar en el artículo de J. V. Ryff [29], nos muestra otras afirmaciones que nos hacen ver esa dualidad:

**Teorema 4.2.2.**

1. Dadas densidades  $f$  y  $g$  integrables Lebesgue en un determinado espacio, son equivalentes:

- a)  $f \leq_M g$ ;
- b)  $\int u(f(x))dx \geq \int u(g(x))dx$  para cualquier función  $u$  continua y convexa;
- c)  $\int (f(x) - a)_+ dx \geq \int (g(x) - a)_+ dx$  para todo  $a > 0$ .

2. Si  $f(x) = \int k(x, y)g(y)dy$  para algún núcleo  $k$  no negativo con

$$\int k(x, y)dx = \int k(x, y)dy = 1 \text{ para todos } x \text{ e } y,$$

entonces  $f \leq_M g$ .

Si sustituimos tales funciones  $f$  y  $g$  por variables aleatorias con esperanza finita, entonces es evidente que  $\leq_M$  en el caso continuo es lo mismo que  $\leq_{cx}$ . Nótese que la condición de la existencia de tal núcleo (propia del teorema de Strassen, como ya vimos) equivale a la matriz doblemente estocástica de la que hablábamos anteriormente.

### 4.3. Consecuencias de la mayorización

Schur introdujo en 1923 el concepto de que una función sea monótona con respecto a la mayorización:

**Definición 4.3.1.** Una función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  se dice que es *Schur-convexa* si

$$\mathbf{x} \leq_M \mathbf{y} \implies f(\mathbf{x}) \leq f(\mathbf{y}).$$

Como ya vimos en la observación 4.1.3, para las demostraciones solo es necesario considerarlo en el caso  $n = 2$ .  $\mathbf{x} \leq_M \mathbf{y}$  equivale a que  $\mathbf{x} = \mathbb{P}\mathbf{y}$  donde  $\mathbb{P}$  es una matriz doblemente estocástica, luego ha de ser

$$\mathbb{P} = \begin{pmatrix} \alpha & \bar{\alpha} \\ \bar{\alpha} & \alpha \end{pmatrix} \quad (4.7)$$

para algún valor  $\alpha \in [0, 1]$  de modo que  $\bar{\alpha} = 1 - \alpha$ . Entonces  $\mathbf{x} = (x_1, x_2)$  se puede escribir como  $x_1 = \alpha y_1 + \bar{\alpha} y_2$  y  $x_2 = \bar{\alpha} y_1 + \alpha y_2$ . Por lo tanto, para  $n = 2$  la condición de que  $f$  sea Schur-convexa equivale a que existe  $\alpha \in [0, 1]$  tal que

$$f(\alpha y_1 + (1 - \alpha)y_2, (1 - \alpha)y_1 + \alpha y_2) \leq f(y_1, y_2).$$

En el libro de Marshall y Olkin [25] puede encontrarse el siguiente teorema que nos será de utilidad:

**Teorema 4.3.2.** Sea  $g$  una función real cualquiera. Definimos

$$f(\mathbf{x}) = \sum_{i=1}^n g(x_i) \text{ para } \mathbf{x} \in \mathbb{R}^n.$$

Entonces  $f$  es Schur-convexa si y solo si  $g$  es convexa.

*Demostración.* Basta probarlo para  $n = 2$ . Por hipótesis  $g$  es convexa, luego

$$\begin{aligned} f(\mathbf{x}) &= g(x_1) + g(x_2) = g(\alpha y_1 + \bar{\alpha} y_2) + g(\bar{\alpha} y_1 + \alpha y_2) \leq \\ &\leq [\alpha g(y_1) + \bar{\alpha} g(y_2)] + [\bar{\alpha} g(y_1) + \alpha g(y_2)] = g(y_1) + g(y_2) = f(\mathbf{y}), \end{aligned}$$

luego  $f$  es Schur-convexa. El recíproco es inmediato.  $\square$

Con él podemos ver otra relación entre la mayorización y el orden convexo entre variables aleatorias: cada vector  $n$ -dimensional  $\mathbf{x} = (x_1, \dots, x_n)$  puede identificarse con una v.a.  $X$  que asigne probabilidad  $1/n$  a cada  $x_i$ , si todas ellas son diferentes. Cuando  $k$  componentes sean iguales, entonces se asigna probabilidad  $k/n$  a ese valor.

Formalmente, sea  $\Omega = \{1, \dots, n\}$  y consideremos como  $\sigma$ -álgebra al conjunto de partes de  $\Omega$ ,  $\mathcal{P}(\Omega)$ . Tomamos  $P$  como la distribución uniforme sobre  $\Omega$ , que asigna probabilidad  $1/n$  a cada  $\omega$ .

Entonces, dado un vector  $\mathbf{x} = (x_1, \dots, x_n)$  la variable aleatoria  $X$  se define sobre  $(\Omega, \mathcal{P}(\Omega), P)$  como

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ i &\longmapsto x_i. \end{aligned}$$

Si  $\mathbf{x}$  es un vector que contiene las riquezas de la población, entonces  $X$  se puede considerar como la riqueza de un individuo tomado aleatoriamente.

Usando esta construcción y aplicando el teorema 4.3.2 obtenemos el siguiente corolario, tomado del libro de Müller y Stoyan [27]:

**Corolario 4.3.3.** Dados vectores aleatorios  $\mathbf{x}$  e  $\mathbf{y}$ , sean  $X$  e  $Y$  las variables aleatorias construidas arriba. Entonces  $\mathbf{x} \leq_M \mathbf{y}$  si y solo si  $X \leq_{cx} Y$ .

Aparte de en economía, la mayorización se emplea en muchos otros ámbitos como en la **física**. Por ejemplo, consideremos  $n$  cuerpos cada uno en equilibrio térmico y todos ellos con la misma capacidad calorífica. Sea  $x_i$  la cantidad de calor del cuerpo  $i$ . Se tiene una situación muy similar a la transición de riquezas entre individuos. En este caso, es posible conectar algunos de los cuerpos durante un periodo de tiempo para que se produzca una transferencia de calor entre ellos. Si no hay pérdidas de calor, esto llevará a una distribución más uniforme de la distribución del calor. Si todos los cuerpos están conectados durante un largo tiempo, entonces finalmente todos ellos tienen la misma cantidad de calor.

También tiene aplicaciones en otras áreas de las matemáticas, como por ejemplo en **variable compleja** para estudiar la geometría de los polinomios. Un ejemplo es la generalización del teorema de Gauss-Lucas: si  $p(z) = \prod_{k=1}^n (z - z_k)$ ,  $\bar{z} = \frac{1}{n} \sum_{k=1}^n z_k$  y  $w_1, \dots, w_{n-1}$  son los ceros de  $p'(z)$ , entonces  $(w_1, \dots, w_{n-1}, \bar{z}) \leq_M (z_1, \dots, z_n)$ . De esto se deduce que existe una matriz  $\mathbb{P}$  doblemente estocástica que relaciona los ceros de un polinomio  $P$  y los de su polinomio derivado  $P'$  sobre  $\mathbb{C}$ .

### 4.3.1. Midiendo la aleatoriedad

Volviendo a nuestro ámbito, según podemos ver en el artículo de Hickey (1982) [19] la mayorización se puede usar también para comparar la heterogeneidad de distribuciones de probabilidad sobre un conjunto de  $n$  puntos, que tomaremos sin pérdida de generalidad como  $\Omega = \{1, \dots, n\}$ . Sea

$$\Sigma = \{\mathbf{p} = (p_1, \dots, p_n) : \sum_{i=1}^n p_i = 1, p_i \geq 0\},$$

el conjunto formado por todas las funciones de masa de probabilidad sobre  $\Omega$ . Notemos que también se podría hacer una extensión al caso de probabilidades discretas donde el soporte es infinito, trabajando de este modo con sucesiones (un subespacio de  $\ell^1$ ). La mayorización sobre  $\Sigma$  es un orden que compara la disimilaridad entre funciones de masa de probabilidad, esto es, cómo de diferentes son unas de otras respecto a la aleatoriedad. Esto da sentido a la siguiente definición:

**Definición 4.3.4.** Sean  $X$  e  $Y$  variables aleatorias con soporte finito y sean  $\mathbf{p} = (p_1, \dots, p_n)$  y  $\mathbf{q} = (q_1, \dots, q_n)$  sus funciones de masa de probabilidad (definidas sobre la unión de ambos soportes). Se dice que  $X$  es menos aleatoria que  $Y$  (escrito  $X \leq_{rand} Y$ ) si  $\mathbf{q} \leq_M \mathbf{p}$ .

La relación  $\leq_{rand}$  es reflexiva y transitiva, pero no es antisimétrica (salvo un cambio de coordenadas, que podríamos arreglarlo pidiendo que estén ordenadas de forma creciente). A diferencia de todos los órdenes que hemos estudiado anteriormente, el orden  $\leq_{rand}$  tiene una peculiaridad. Para los anteriores era importante saber *dónde* se localizaba la masa de probabilidad, y por lo tanto conocer la estructura topológica y de orden del espacio sobre el que se construyen (que, hasta ahora, ha sido la recta real). Ahora para el orden  $\leq_{rand}$  lo que nos importa es saber *cuanta* masa de probabilidad hay en cada punto, sin importarnos las propiedades del espacio. Para este orden basado en la mayorización es importante el siguiente resultado, que se puede encontrar en el artículo de Hickey (1983) [19] y en el libro de Marshall y Olkin [25]:

**Proposición 4.3.5.** En el conjunto de todas las distribuciones con soporte finito fijado, la distribución uniforme es el único máximo y las distribuciones puntuales son los elementos mínimos con respecto al orden  $\leq_{rand}$ .

*Demostración.* Para la mayorización siempre se cumple que

$$\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \leq_M (p_1, \dots, p_n) \leq_M (1, \dots, 0)$$

donde  $p_i \geq 0$  y  $\sum p_i = 1$  (el elemento a la derecha puede cambiarse por cualquiera de los elementos canónicos  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$  con un 1 en la posición  $i$ ). Por lo tanto la distribución uniforme es el máximo para el orden  $\leq_{rand}$  (el cual es único pues no hay otra forma de repartir la probabilidad para que se haga mínima en el orden  $\leq_M$ ), y las distribuciones unipuntuales son el mínimo para el orden  $\leq_{rand}$ .  $\square$

**Observación 4.3.6.** Lo que nos viene a decir la proposición anterior es que las distribuciones degeneradas, en las que la probabilidad está totalmente concentrada en uno de los elementos, es menos aleatoria que cualquier distribución no degenerada, y que la distribución uniforme  $\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$  es la más aleatoria de entre todas aquellas que pertenecen a  $\Sigma$ .

La misma proposición nos justifica también el hecho de que los  $n$  cuerpos de los que hablábamos anteriormente lleguen al equilibrio térmico cuando se los deja en contacto un largo periodo de tiempo (adquiriendo así una distribución uniforme), y que si se mantienen separados y no hay pérdidas de calor se mantienen con el mismo calor (lo que son las distribuciones unipuntuales).

### 4.3.2. Dualidad con el concepto de entropía

Según se muestra en el artículo de Hickey (1983) [20], para medir la diferencia entre distribuciones de probabilidad discretas  $P = (p_1, \dots, p_n)$  se suele utilizar la llamada *función de entropía*

$$\phi_E(P) = - \sum_{i=1}^n p_i \log_n p_i$$

que está bien definida si decimos que  $0 \log_n 0 = 0$ . La función  $f(t) = -t \log_n t$  es convexa para todo  $t \in [0, 1]$  (según la definición que hemos considerado de convexidad), por lo tanto  $\phi_E$  es una función Schur-convexa por el teorema 4.3.2. De este modo podemos observar que aparece una conexión entre el orden  $\leq_{rand}$  y el concepto de entropía, habitual en la *teoría de la información* y que ha sido estudiado en la asignatura Criptografía durante el grado.

Por ejemplo, para el lanzamiento de una moneda en el que apostamos a cara o cruz, estamos considerando el espacio  $\Sigma = \{(p, 1 - p) : p \in [0, 1]\}$ . Suponiendo que en la moneda la probabilidad de que salga cara es igual a la probabilidad de que salga cruz, entonces estaríamos considerando el elemento  $P_u = \left(\frac{1}{2}, \frac{1}{2}\right) \in \Sigma$ , esto es, la distribución uniforme. Para ella se tiene que

$$\phi_E(P_u) = - \sum_{i=1}^2 p_i \log_2 p_i = -2 \left( \frac{1}{2} \log_2 \frac{1}{2} \right) = -\log_2 \frac{1}{2} = 1,$$

mientras que para los casos en los que solo puede salir cara o solo puede salir cruz ( $P_C = (1, 0)$  y  $P_X = (0, 1)$  respectivamente) se tiene que

$$\phi_E(P_C) = \phi_E(P_X) = -\log_2 1 = 0,$$

de modo que los valores extremos para el orden  $\leq_{rand}$  se corresponden con los de la función  $\phi_E$ . De hecho, tomando  $\Sigma = \{(p, 1 - p) : p \in [0, 1]\}$  podemos representar la función  $\phi_E(p) = -p \log_2(p) - (1 - p) \log_2(1 - p)$  para  $p \in [0, 1]$ :

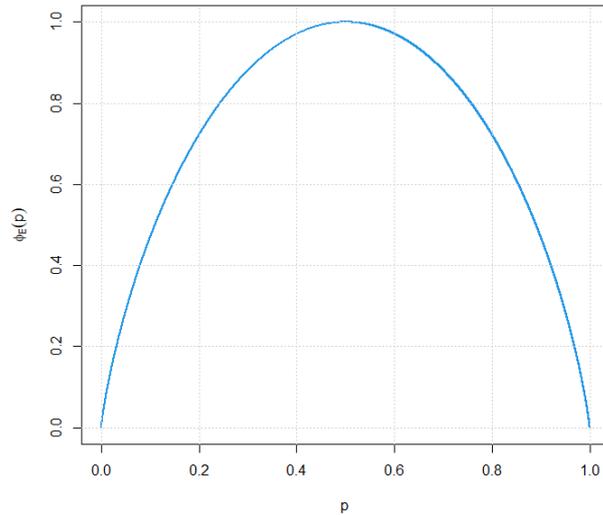


Figura 4.2: Función de entropía  $\phi_E$  para el caso  $n = 2$

Como podemos ver en la figura 4.2, el valor máximo de  $\phi_E$  se alcanza para la distribución uniforme (la más aleatoria de todas), mientras que los valores mínimos se alcanzan para las distribuciones concentradas (las menos aleatorias). Recordemos que  $\phi_E$  es una función Schur-convexa, luego es monótona para la mayorización, por lo tanto todos los valores intermedios corresponden a distribu-

ciones intermedias para el orden  $\leq_{rand}$ . Observamos de este modo la dualidad que existe para medir la aleatoriedad entre distribuciones.

# Apéndice A

## Resultados de teoría de la probabilidad

En este apéndice recogemos algunas cuestiones básicas de teoría de la probabilidad, necesarias para algunas de las partes contempladas en los capítulos de este trabajo.

Consideraremos un espacio probabilístico  $(\Omega, \sigma, P)$  formado por un conjunto  $\Omega$  al que denominaremos *espacio muestral*, una  $\sigma$ -álgebra  $\sigma$  definida sobre  $\Omega$  formada por elementos a los que llamaremos *sucesos* y una función de probabilidad  $P : \sigma \rightarrow [0, 1]$  que asigna a cada suceso una probabilidad. Reservaremos la letra  $\beta$  para el caso en el que utilicemos la  $\sigma$ -álgebra de Borel definida sobre  $\mathbb{R}$ .

### A.1. Función cuantil

Sea  $X : \Omega \rightarrow \mathbb{R}$  una *variable aleatoria* real con *función de distribución*  $F$ . A partir de  $F$  podemos considerar su *función cuantil*, que denotaremos por  $F^{-1}$ , y que se define como

$$F^{-1}(t) := \min\{x : t \leq F(x)\}, \quad t \in (0, 1).$$

Es sencilla de interpretar:  $F^{-1}(t)$  es el valor más pequeño a partir del cual la función de distribución supera el valor  $t \in [0, 1]$ . La función cuantil tiene algunas propiedades importantes:

#### Proposición A.1.1. (Propiedades de la función cuantil)

1.  $F^{-1}(p) \leq x$  si y solo si  $p \leq F(x)$ .
2.  $\{x \in \mathbb{R} \mid t \leq F(x)\} = [F^{-1}(t), +\infty)$ .
3. Si  $F$  es una función de distribución continua y estrictamente creciente, entonces  $F^{-1}$  es la inversa analítica de  $F$ .
4.  $p \leq F(F^{-1}(p))$ .
5.  $F^{-1}(F(x)) \leq x$ .

6. Dada una v.a.  $U$  uniformemente distribuida en el intervalo  $(0, 1)$ , si  $F$  es una función de distribución entonces  $X = F^{-1}(U)$  es una variable aleatoria con función de distribución  $F$ .

*Demostración.*

1.  $F^{-1}(p) \leq x$  quiere decir que  $x \in \{z : F(z) \geq p\}$  y por lo tanto  $p \leq F(x)$ . Por el otro lado, si  $p \leq F(x)$ , entonces  $x \in \{z : F(z) \geq p\}$  y por tanto  $F^{-1}(p) \leq x$  puesto que  $F^{-1}(p)$  es el ínfimo y  $F$  es continua por la derecha.
2. Inmediato a partir de 1.
3. Resulta trivial pues con tales propiedades  $F$  es una biyección y está correctamente definida su inversa, que es  $F^{-1}$ .
4. Si llamamos  $x = F^{-1}(p)$ , entonces se tiene que  $p \leq F(x)$ . Por la propiedad 1. esto equivale a  $F^{-1}(p) \leq x$ , y obviamente  $F^{-1}(p) \leq F^{-1}(p)$ .
5.  $F^{-1}(F(x)) = \inf\{z : F(z) \geq F(x)\}$  y  $x \in \{z : F(z) \geq F(x)\}$  por lo que  $x$  ha de ser mayor que el ínfimo.
6. Supongamos que  $U$  es una variable aleatoria distribuida uniformemente en el intervalo  $(0, 1)$ . Veamos que  $X = F^{-1}(U)$  tiene función de distribución  $F$ :

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

□

## A.2. Momentos de una variable aleatoria

**Proposición A.2.1 (Desigualdad de Jensen).** Sea  $X$  una variable aleatoria. Para toda función  $\varphi$  convexa se verifica que

$$\varphi(EX) \leq E\varphi(X).$$

**Proposición A.2.2.** Si  $X$  es una variable aleatoria no negativa con esperanza finita, entonces

$$EX = \int_0^{+\infty} \bar{F}(t) dt.$$

El teorema de Fubini es bien conocido en el contexto del análisis matemático. En él se utiliza la medida de Lebesgue, aunque admite una extensión a medidas más generales siempre que sean  $\sigma$ -finitas (véase la definición B.2.2 en el apéndice B). Puede consultarse esta versión más general en el libro de Billingsley [13] de la que haremos uso en la siguiente proposición:

**Proposición A.2.3.** Sea  $X$  una variable aleatoria. Entonces

$$EX = \int_0^{\infty} [1 - F_X(t)] dt - \int_{-\infty}^0 F_X(t) dt.$$

*Demostración.* Podemos descomponer  $X = X^+ - X^-$  donde  $X^+$  y  $X^-$  son las partes positiva y negativa de  $X$ . Dado  $\omega \in \Omega$  arbitrario,

$$X^-(\omega) = \int_{-X^-(\omega)}^0 dx = \int_{-\infty}^0 I_{\{x \mid X^-(\omega) \geq -x\}} dx = \int_{-\infty}^0 I_{\{(\omega, x) \mid X(\omega) \leq x\}} dx.$$

La última igualdad es cierta puesto que si  $X^-(\omega) \geq -x$  con  $x \leq 0$  entonces  $X^+(\omega) = 0$  y por tanto  $X(\omega) \leq x$ . Recíprocamente, si  $X(\omega) \leq x$  con  $x \leq 0$  entonces  $X^+(\omega) = 0$  y  $X^-(\omega) \geq -x$ .

El integrando de la expresión anterior es no negativo, luego aplicando el teorema de Fubini podemos intercambiar el orden de integración al tomar esperanzas, esto es,

$$\begin{aligned} E(X^-) &= \int_{\Omega} \left( \int_{-\infty}^0 I_{\{(\omega, x) \mid X(\omega) \leq x\}} dx \right) dP = \int_{-\infty}^0 \left( \int_{\Omega} I_{\{\omega \mid X(\omega) \leq x\}} dP \right) dx = \\ &= \int_{-\infty}^0 P(X \leq x) dx = \int_{-\infty}^0 F(x) dx. \end{aligned}$$

Del mismo modo se puede probar la proposición A.2.2 anterior, entonces, como la esperanza es lineal, se tiene que

$$EX = EX^+ - EX^- = \int_0^{\infty} [1 - F_X(t)] dt - \int_{-\infty}^0 F_X(t) dt.$$

□

### A.3. Convergencias

En primer lugar daremos los conceptos de convergencia necesarios para este trabajo. Sus caracterizaciones equivalentes pueden encontrarse en el libro de P. Billingsley [13].

**Definición A.3.1.** Se dice que una sucesión de variables aleatorias  $(X_n)_{n \geq 1}$  converge *casi seguro* a una variable aleatoria límite  $X$  definidas todas ellas sobre un mismo espacio  $\Omega$  si

$$P\left(\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1.$$

Se representa como  $X_n \xrightarrow{c.s.} X$ .

**Definición A.3.2.** Supongamos que las variables  $X_n$  y  $X$  tienen funciones de distribución  $F_n$  y  $F$  respectivamente. Se dice que una sucesión de variables aleatorias reales  $(X_n)_{n \geq 1}$  converge *en distribución* a una variable aleatoria real  $X$  definidas todas ellas sobre un mismo espacio  $\Omega$  si

$$F_n(x) \longrightarrow F(x)$$

para cada punto  $x$  de continuidad de la función  $F$ . Se representa como  $X_n \xrightarrow{d} X$ .

**Observación A.3.3.** A la convergencia en distribución también se le llama *convergencia en ley* y se representa como  $\mathcal{L}(X_n) \longrightarrow \mathcal{L}(X)$ .

A continuación veremos algunos resultados de convergencias bajo el signo integral, análogos a los propios del análisis de variables reales.

**Teorema A.3.4 (de la convergencia monótona).** Sea  $\{X_n : n \geq 1\}$  una sucesión no decreciente de variables aleatorias no negativas con con límite medible  $X$  c.s.. Entonces

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n(\omega) dP(\omega) = \int_{\Omega} X(\omega) dP(\omega).$$

**Teorema A.3.5 (de la convergencia dominada).** Sea  $\{X_n : n \geq 1\}$  una sucesión de variables aleatorias tal que  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$  c.s., y  $|X_n| \leq Y$  para todo  $n \geq 1$  con  $\int_{\Omega} Y(\omega) dP(\omega) < \infty$ . Entonces

$$\int_{\Omega} |X(\omega)| dP(\omega) < \infty$$

y

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n(\omega) dP(\omega) = \int_{\Omega} X(\omega) dP(\omega).$$

**Teorema A.3.6 (Teorema del Transfer).** Sea  $X$  una variable aleatoria con llegada a un espacio medible  $(A, \mathcal{A})$  y  $g : A \rightarrow \mathbb{R}$  una función  $\mathcal{A}|\mathcal{R}$ -medible. Entonces  $g \circ X$  es  $P$ -integrable si y solo si  $g$  es  $P_X$ -integrable, en cuyo caso se tiene que

$$E[g(X)] = \int_A g(u) dP_X(u) = \int_{\Omega} g(X(\omega)) dP(\omega).$$

# Apéndice B

## Algunos resultados de análisis

En este apéndice se darán algunos resultados de análisis que son necesarios para poder seguir correctamente los contenidos del trabajo. En teoría de la probabilidad es muchas veces necesario analizar algunos *funcionales* y algunas *medidas* para obtener ciertos resultados de relevancia, como en este caso sucede para la comparación entre variables aleatorias. Por ello se incluyen contenidos de análisis funcional y análisis real que se utilizan para la demostración de algunos teoremas.

En un contexto general, un *espacio de medida* es una terna  $(\Omega, \Lambda, \mu)$  formada por un conjunto  $\Omega$  para el que se ha definido una  $\sigma$ -álgebra  $\Lambda$  de *conjuntos medibles* y una *función medida*  $\mu$  que asigna un valor real o medida a cada elemento de la  $\sigma$ -álgebra.

Algunos ejemplos notables de espacios de este tipo son los espacios probabilísticos  $(\Omega, \sigma, P)$  donde todo conjunto medible tiene una medida finita y  $P(\Omega) = 1$ , o la terna  $(\mathbb{R}, \beta, m)$  formada por el conjunto de los números reales, la  $\sigma$ -álgebra de Borel  $\beta$  y la medida de Lebesgue  $m$ , basada en la longitud de los intervalos.

### B.1. Reordenamientos decrecientes

Denotemos por  $M(\Omega, \mu)$  al conjunto de todas las funciones medibles con valores en  $\overline{\mathbb{R}} = [-\infty, \infty]$  definidas sobre un espacio de medida  $(\Omega, \Lambda, \mu)$ . Dada una función  $f \in M(\Omega, \mu)$  se define su *reordenamiento decreciente*  $f^*$  como

$$f^*(t) = \inf\{s \in \overline{\mathbb{R}} : D_f(s) \leq t\}, t \in [0, \mu(\Omega)],$$

donde

$$D_f(s) = \mu(\{x : f(x) > s\}),$$

para cada  $s \in [-\infty, \infty]$ . Esta nueva función conserva la medida de los conjuntos de nivel de  $f$ .

**Observación B.1.1.** En el caso particular de los espacios probabilísticos,  $M(\Omega, P)$  se corresponde con el conjunto de todas las funciones  $\sigma|\overline{\beta}$ -medibles, esto es, el conjunto de todas las variables

aleatorias con valores en  $\overline{\mathbb{R}}$ . Dada una variable aleatoria  $X$  con función de distribución  $F$  (y función de supervivencia  $\overline{F}$ ) se tiene que

$$D_X(s) = P(\{\omega : X(\omega) > s\}) = 1 - F(s) = \overline{F}(s).$$

Las funciones de distribución  $F$  son funciones crecientes y continuas por la derecha. Al construir su función cuantil  $F^{-1}$  obtenemos una función que es creciente y continua por la izquierda, y que además tiene la propiedad de que si  $X$  es una variable aleatoria con función de distribución  $F$ , entonces  $X \stackrel{d}{=} F^{-1}$ . Sin embargo, nada nos impide considerar la función  $\overline{F}^{-1}(t) := F^{-1}(1 - t)$ , que por construcción es decreciente y continua por la derecha. Dada una variable aleatoria  $U$  uniforme en  $(0, 1)$ , también se verifica que  $Y := \overline{F}^{-1}(U)$  tiene como función de distribución  $F$ :

$$P(Y \leq t) = P(\overline{F}^{-1}(U) \leq t) = P(F^{-1}(1 - U) \leq t) = P(1 - U \leq F(t)) = P(V \leq F(t)) = F(t),$$

donde hemos utilizado la propiedad A.1.1.1 de la función cuantil y el hecho de que  $V := 1 - U$  también tiene distribución uniforme sobre  $(0, 1)$ .

**Proposición B.1.2.**  $\overline{F}^{-1}$  es el reordenamiento decreciente de una variable aleatoria  $X$  con función de distribución  $F$ .

*Demostración.* Para ello basta observar que, por definición, para todo  $t \in [0, 1]$ ,

$$\begin{aligned} X^*(t) &= \inf\{s \in \overline{\mathbb{R}} : D_X(s) \leq t\} = \inf\{s \in \overline{\mathbb{R}} : \overline{F}(s) \leq t\} = \\ &= \inf\{s \in \overline{\mathbb{R}} : F(s) \geq 1 - t\} = F^{-1}(1 - t) = \overline{F}^{-1}(t). \end{aligned}$$

□

Recordemos que  $L^1(\Omega, \mu)$  se define como el espacio de funciones medibles cuyo valor absoluto es integrable respecto a la medida  $\mu$  y en el cual se identifican todas las funciones que son iguales c.s., esto es, está formado por funciones  $f \in M(\Omega, \mu)$  para las cuales  $\int_{\Omega} |f| d\mu < \infty$ .

Se tiene el siguiente teorema, que nos permitirá resolver algunas cuestiones relacionadas con los órdenes convexos:

**Teorema B.1.3.** Supongamos que  $(\Omega, \Lambda, \mu)$  y  $(\Omega', \Lambda', \mu')$  son espacios de medida tales que  $\mu(\Omega) = \mu'(\Omega') = a < \infty$  y denotemos por  $m$  a la medida de Lebesgue en  $\mathbb{R}$ . Si  $f \in L^1(\Omega, \mu)$ ,  $g \in L^1(\Omega', \mu')$  y  $f^*$ ,  $g^*$  son sus reordenamientos decrecientes, entonces

$$\int_0^t f^* dm \leq \int_0^t g^* dm \text{ para todo } t \in [0, a],$$

si y solo si

$$\int_{\Omega} \Psi(f) d\mu \leq \int_{\Omega'} \Psi(g) d\mu'$$

para toda función  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  convexa creciente.

Se puede encontrar información adicional sobre el resultado anterior en el artículo de M. A. Sordo y H. M. Ramos [31] o en los de Kong-Ming Chong [15] y [16].

## B.2. El teorema de Radon-Nikodym

En esta subsección se hace un pequeño resumen de cómo se pueden caracterizar las medidas con densidad. Si  $f$  es una función no negativa definida sobre un espacio de medida  $(\Omega, \Lambda, \mu)$ , entonces dado  $A \in \Lambda$  se tiene que

$$\nu(A) = \int_A f d\mu$$

define una nueva medida sobre  $\Lambda$ . Se dice que  $\nu$  es la *medida con densidad  $f$  respecto de  $\mu$* . El propósito de esta subsección es ver qué condiciones han de cumplir dos medidas  $\mu$  y  $\nu$  para poder asegurar que existe una densidad que las relacione. Para ello será necesario introducir los conceptos que siguen a continuación.

**Definición B.2.1.** Sea  $(\Omega, \Lambda, \mu)$  un espacio de medida y  $\nu$  otra medida definida en el mismo espacio medible. Se dice que  $\nu$  es *absolutamente continua* respecto de  $\mu$  (escrito  $\nu \ll \mu$ ) si se verifica

$$\nu(M) = 0 \text{ para todo } M \in \Lambda \text{ tal que } \mu(M) = 0.$$

**Definición B.2.2.**

1. Un conjunto  $A \in \Lambda$  se dice que es *de medida finita* si  $\mu(A) < \infty$ .
2. Un espacio de medida  $(\Omega, \Lambda, \mu)$  se dice que es *finito* (o que la medida  $\mu$  es *finita*) si  $\mu(\Omega) < \infty$ .
3. Un espacio de medida  $(\Omega, \Lambda, \mu)$  se dice que es  *$\sigma$ -finito* (o que la medida  $\mu$  es  *$\sigma$ -finita*) si  $\Omega$  se puede escribir como unión numerable de conjuntos de  $\Lambda$  de medida finita.

Por ejemplo, la medida de Lebesgue  $m$  sobre  $(\mathbb{R}, \beta)$  no es finita pero sí  $\sigma$ -finita, puesto que  $\mathbb{R}$  se puede escribir como la unión numerable de los intervalos  $[k, k+1)$  para todo  $k \in \mathbb{Z}$ , cada uno de los cuales tiene medida finita e igual a 1.

Si  $\nu$  es la *medida con densidad  $f$  respecto de  $\mu$* , para cada  $A \in \Lambda$  se tiene que si  $\mu(A) = 0$ , entonces  $\nu(A) = 0$ , esto es,  $\nu \ll \mu$ . A continuación veremos recíprocamente que si esta condición se satisface y además  $\nu$  y  $\mu$  son medidas  $\sigma$ -finitas sobre  $\Lambda$ , entonces  $\nu$  es una medida con densidad respecto de  $\mu$ , para una determinada densidad  $f$ . Esto es lo que afirma el conocido teorema de Radon-Nikodym.

**Teorema B.2.3 (de Radon-Nikodym).** Si  $\mu$  y  $\nu$  son dos medidas  $\sigma$ -finitas definidas sobre un espacio medible  $(\Omega, \Lambda)$  tales que  $\nu$  es absolutamente continua respecto de  $\mu$ , entonces existe una densidad  $f$  no negativa que verifica

$$\nu(A) = \int_A f d\mu, \text{ para todo } A \in \Lambda.$$

Además, para dos densidades  $f$  y  $g$  que cumplen lo anterior se tiene que  $\mu(\{f \neq g\}) = 0$ .

En el contexto de la teoría de la medida, se dice que  $f$  es la *derivada de Radon-Nikodym de  $\nu$  respecto de  $\mu$* . Según lo anterior, cualquier otra derivada de  $\nu$  respecto de  $\mu$  es igual a esta en casi todo  $\Omega$ , y se denota como  $\frac{d\nu}{d\mu}$ .

La demostración de este resultado, la cual es bastante laboriosa, puede encontrarse en el libro de P. Billingsley [13], así como más información sobre cómo se aplica al caso de los espacios probabilísticos.

# Apéndice C

## Programas en R

En este apéndice se incluye el código que se ha elaborado para calcular los índices de desviación explicados en el capítulo 2  $\gamma(F, G)$  y  $\alpha(F, G)$  al tratar con muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  tomadas a partir de dos poblaciones con funciones de distribución  $F$  y  $G$  respectivamente. Se recomiendan ejecutar en el entorno RStudio.

### C.1. `gamma()`

Los argumentos de entrada son dos vectores `c1` y `c2` que contienen los datos. Sólo es válido en el caso de que  $n = m$ , es decir, que las dos muestras tengan el mismo número de datos. La base del programa consiste en ordenar ambos vectores `c1` y `c2` y contar el número de veces que el valor  $i$ -ésimo de `c1_ordenado` es mayor que el valor  $i$ -ésimo de `c2_ordenado`,  $i$  in  $1:n$ .

```
1 #gamma
2 gamma<-function(c1,c2){
3   if(length(c1)!=length(c2)){
4     stop("Los vectores han de tener la misma longitud")
5   }else{
6     contador<-0; gama<-0;
7     x<-sort(c1); y<-sort(c2);
8     n<-length(c1);
9     for(i in 1:n){
10      if(x[i]>y[i]){ #contamos el numero de veces que x[i]>y[i]
11        contador<-contador+1;
12      }
13    }
14    gama<-contador/n;
15    gama;
16  }
17 }
```

## C.2. alpha()

Los argumentos de entrada son dos vectores  $c1$  y  $c2$  que contienen los datos. Funciona para longitudes  $n$  y  $m$  cualesquiera, y no tiene en cuenta ningún tipo de información adicional sobre las muestras que se consideran. La base del programa (conviene fijarse en la gráfica 2.3 para entenderlo) es la siguiente: dadas dos muestras  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$  ya ordenadas y tomadas a partir de dos distribuciones  $F$  y  $G$ , para calcular el índice  $\alpha(F_n, G_m)$  podemos utilizar el carácter especial de las funciones de distribución empíricas, que son escalonadas, además de observar que dicho índice se define a través de dos integrales. Por lo tanto para calcularlo basta con hallar el área que se forma entre las dos funciones cuando se cruzan, que se reduce a sumar áreas de rectángulos (para el numerador de  $\alpha$  solo sumaremos cuando  $G > F$ ). La clave para resolver el problema se basa en ordenar todos los valores  $x_1, \dots, x_n, y_1, \dots, y_m$  obteniendo una partición  $z_1, \dots, z_{n+m}$  de modo que se van comparando  $z_1$  con  $z_2$ ,  $z_2$  con  $z_3, \dots$  de la siguiente forma: si  $z_1$  y  $z_2$  son valores de  $F$ , entonces  $F_n$  ha crecido  $1/n$  [ $G_m$  ahí vale 0] y el área se calcula como  $(1/n) * [z_2 - z_1]$ . Si ahora por ejemplo sucede que  $z_3$  es un valor de  $G$ , entonces la diferencia entre  $G_m$  y  $F_n$  vale  $d = 1/m - 1/n$ : si este valor  $d$  es positivo,  $G_m$  está por encima; si es negativo, lo está  $F_n$  (esto se puede observar en las diferencias  $d_i$  que aparecen marcadas en la figura 2.3). Podemos calcular este área como  $d * (z_3 - z_2)$ . El procedimiento se repite hasta llegar a algún valor  $z_i$  que sea igual a  $x_n$  o a  $y_m$ . A partir de ahí  $F_n$  o  $G_m$  ya valen 1, luego se ha de tratar el problema aparte, como se muestra en el programa.

```
1 #alpha
2 alpha<-function(c1,c2){
3   diferencia<-0; #Calcula el valor G(x) - F(x)
4   I1<-0; #Numerador de alpha, suma area cuando G > F
5   I2<-0; #Denominador de alpha, suma las areas de los rectangulos
6   x<-sort(c1); y<-sort(c2);
7   n<-length(c1); m<-length(c2);
8   i<-1;j<-1;
9
10  while((i<n)&(j<m)){ #Vamos comparando 1 a 1 los valores ordenados
11    if(y[j]<=x[i]){ #Primero valor de G
12      diferencia<-diferencia+(1/m);
13      if(y[j+1]<=x[i]){ #y[j] e y[j+1] seguidos
14        if(diferencia>0){ #Por encima G
15          I1<-I1+diferencia*(y[j+1]-y[j]);
16          I2<-I2+diferencia*(y[j+1]-y[j]);
17        }else{ #Por encima F
18          I2<-I2-diferencia*(y[j+1]-y[j]);
19        }
20      }else{ #y[j] y x[i] seguidos
21        if(diferencia>0){ #Por encima G
22          I1<-I1+diferencia*(x[i]-y[j]);
23          I2<-I2+diferencia*(x[i]-y[j]);
24        }else{ #Por encima F
```

```

25         I2<-I2-diferencia*(x[i]-y[j]);
26     }
27 }
28 j<-j+1; #Avanzamos un dato de F
29 }else{ #Primero valor de F
30     diferencia<-diferencia-(1/n);
31     if(x[i+1]<=y[j]){ #x[i] y x[i+1] seguidos
32         if(diferencia>0){ #Por encima G
33             I1<-I1+diferencia*(x[i+1]-x[i]);
34             I2<-I2+diferencia*(x[i+1]-x[i]);
35         }else{ #Por encima F
36             I2<-I2-diferencia*(x[i+1]-x[i]);
37         }
38     }else{ #x[i] e y[j] seguidos
39         if(diferencia>0){ #Por encima G
40             I1<-I1+diferencia*(y[j]-x[i]);
41             I2<-I2+diferencia*(y[j]-x[i]);
42         }else{ #Por encima F
43             I2<-I2-diferencia*(y[j]-x[i]);
44         }
45     }
46     i<-i+1; #Avanzamos un dato de G
47 }
48 }
49
50 while((j==m)&(i<n)){ #Queda UN SOLO dato de G
51     if(x[i]<=y[j]){
52         diferencia<-diferencia-(1/n);
53         if(x[i+1]<=y[j]){ #x[i] y x[i+1] seguidos
54             if(diferencia>0){ #Por encima G
55                 I1<-I1+diferencia*(x[i+1]-x[i]);
56                 I2<-I2+diferencia*(x[i+1]-x[i]);
57             }else{ #Por encima F
58                 I2<-I2-diferencia*(x[i+1]-x[i]);
59             }
60         }else{ #x[i] e y[m] seguidos
61             if(diferencia>0){ #Por encima G
62                 I1<-I1+diferencia*(y[j]-x[i]);
63                 I2<-I2+diferencia*(y[j]-x[i]);
64             }else{ #Por encima F
65                 I2<-I2-diferencia*(y[j]-x[i]);
66             }
67         }
68         i<-i+1; #Avanzamos dato de F
69     }
70     else{ #Se acaban los datos de G
71         diferencia<-diferencia+(1/m);

```

```

72     I1<-I1+diferencia*(x[i]-y[j]);
73     I2<-I2+diferencia*(x[i]-y[j]);
74     while(i<n){
75         diferencia<-diferencia-(1/n);
76         I1<-I1+diferencia*(x[i+1]-x[i]);
77         I2<-I2+diferencia*(x[i+1]-x[i]);
78         i<-i+1;
79     }
80     i<-i+3; #CONDICION DE PARADA DEL BUCLE
81 }
82 }
83
84 while((i==n)&(j<=m)){ #Queda UN SOLO dato de F
85     if(j<m){
86         if(y[j]<=x[i]){
87             diferencia<-diferencia+(1/m);
88             if(y[j+1]<=x[i]){ #y[j] e y[j+1] seguidos
89                 if(diferencia>0){ #Por encima G
90                     I1<-I1+diferencia*(y[j+1]-y[j]);
91                     I2<-I2+diferencia*(y[j+1]-y[j]);
92                 }else{ #Por encima F
93                     I2<-I2-diferencia*(y[j+1]-y[j]);
94                 }
95             }else{ #y[j] y x[n] seguidos
96                 if(diferencia>0){ #Por encima G
97                     I1<-I1+diferencia*(x[i]-y[j]);
98                     I2<-I2+diferencia*(x[i]-y[j]);
99                 }else{ #Por encima F
100                    I2<-I2-diferencia*(x[i]-y[j]);
101                }
102            }
103            j<-j+1; #Avanzamos un dato de G
104        }else{ #Se acaban los datos de F
105            diferencia<-diferencia-(1/n);
106            I2<-I2-diferencia*(y[j]-x[i]);
107            while(j<m){
108                diferencia<-diferencia+(1/m);
109                I2<-I2-diferencia*(y[j+1]-y[j]);
110                j<-j+1;
111            }
112            j<-j+3; #CONDICION DE PARADA DEL BUCLE
113        }
114    }
115    if(j==m){
116        if(y[j]<=x[i]){
117            diferencia<-diferencia+(1/m);
118            I1<-I1+diferencia*(x[i]-y[j]);

```

```
119     I2<-I2+diferencia*(x[i]-y[j]);
120   }else{ #x[n] < y[m]
121     diferencia<-diferencia-(1/n);
122     I2<-I2-diferencia*(y[j]-x[i]);
123   }
124   j<-j+1;
125 }
126 }
127
128 alfa<-I1/I2;
129 alfa;
130 }
```



# Bibliografía

- [1] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. y MATRÁN, C. (2014). A contamination model for approximate stochastic order: extended version. Disponible en [arxiv.org/abs/1412.1920](https://arxiv.org/abs/1412.1920).
- [2] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. y MATRÁN, C. (2017). Models for the Assessment of Treatment Improvement: The Ideal and the Feasible, *Statistical Science* **32**, No. 3, 469–485. [MR3696006](https://doi.org/10.1214/15-STATSCIENCE)
- [3] ARCONES, M. A., KVAM, P. H. y SAMANIEGO, F. J. (2002). Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint, *Journal of the American Statistical Association* **97**, No. 457, 170–182. [MR1947278](https://doi.org/10.1198/016214502000000000)
- [4] ARMBRUSTER, B. (2013). *A Short Proof of Strassen's Theorem Using Convex Analysis*, Northwestern University. Disponible en [arxiv.org/abs/1603.00137](https://arxiv.org/abs/1603.00137).
- [5] DEL BARRIO, E., CUESTA-ALBERTOS, J. A. y MATRÁN, C. (2016). Profiles of pyramid ages in American countries: a trimmed k-barycenters approach. Technical Report.
- [6] DEL BARRIO, E., CUESTA-ALBERTOS, J. A. y MATRÁN, C. (2018). Some indices to measure departures from stochastic order. Disponible en [arxiv.org/abs/1804.02905](https://arxiv.org/abs/1804.02905).
- [7] BELZUNCE, F., MARTÍNEZ-RIQUELME, C. y MULERO, J. (2016). *An introduction to Stochastic Orders*, Academic Press. [MR2751296](https://doi.org/10.1007/978-1-4939-9712-9)
- [8] BERGER, R. L. (1988). A nonparametric, intersection-union test for stochastic order, *Statistical Decision Theory and Related Topics IV*, Vol. 2, 253-264. [MR0927137](https://doi.org/10.1007/978-1-4613-1000-0_10)
- [9] BERRENDERO, J. R. y CÁRCAMO, J. (2011). Tests for the Second Order Stochastic Dominance Based on L-Statistics, *J. Bus. Econom. Statist.* **29**, No. 2, 260–270. [MR2807880](https://doi.org/10.1177/0898010110382800)
- [10] BERRENDERO, J. R. y CÁRCAMO, J. (2012). Tests for stochastic orders and mean order statistics, *Comm. Statist. Theory Methods* **41**, No. 8, 1497–1509. [MR2903001](https://doi.org/10.1080/03605310.2012.683001)
- [11] BHATIA, R. (1997). *Matrix Analysis*, New York: Springer-Verlag. [MR1477662](https://doi.org/10.1007/978-1-4612-0763-4)
- [12] BHATTACHARJEE, M.C. y SETHURAMAN, J. (1990). Families of life distributions characterized by two moments, *Journal of Applied Probability* **27**, 720-725. [MR1067037](https://doi.org/10.2307/3214037)
- [13] BILLINGSLEY, P. (1995). *Probability and Measure*, 3<sup>a</sup> ed. New York: Wiley. [MR1324786](https://doi.org/10.1002/9781119197580)

- [14] BLACKWELL, D. (1953). Equivalent comparisons of Experiments, *Annals of Mathematical Statistics* **24**, No. 2, 265-272. [MR0056251](#)
- [15] CHONG, K.-M. (1974). An induction principle for spectral and rearrangement inequalities, *Transactions of the American Mathematical Society* **196**, 371-383. [MR0344396](#)
- [16] CHONG, K.-M. (1974). Some extensions of a theorem of Hardy, Littlewood and Pólya and their applications, *Canadian Journal of Maths* **26**, 1321-1340. [MR0352377](#)
- [17] DENUIT, M., LEFÈVRE, C. y SHAKED, M. (2000). On the theory of high convexity stochastic orders, *Statist. Probab. Lett.* **47**, No. 3, 287–293. [MR1747489](#)
- [18] DENUIT, M. y VERMANDELE, C. (1998). Optimal reinsurance and stop-loss order, *Insurance: Mathematics and Economics* **22**, 229-233. [MR1650288](#)
- [19] HICKEY, R.J. (1982). A note on the measurement of randomness, *J. Appl. Probab.* **19**, 229-232. [MR0644436](#)
- [20] HICKEY, R.J. (1983). Majorisation, randomness and some discrete distributions, *J. Appl. Probab.* **20**, 897-902. [MR0720482](#)
- [21] LEHMANN, E. (1955). Ordered Families of Distributions, *Ann. Math. Statist.* **26**, No. 3, 399-419. [MR0071684](#)
- [22] LESHNO, M. y LEVY, H. (2002). Preferred by “All” and Preferred by “Most” Decision Makers: Almost Stochastic Dominance, *Management Science* **48**, No. 8, 955-1101. [MNSC.48.8.1074.169](#)
- [23] LEVY, H. (2006). *Stochastic Dominance*, 3<sup>a</sup> ed. Springer. [MR3525602](#)
- [24] MANN, H. B. y WHITNEY, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, *Annals of Mathematical Statistics* **18**, No. 1, 50-60. [MR0022058](#)
- [25] MARSHALL, A. W. y OLKIN, I. (1979). *Inequalities: Theory of Majorization and its Applications*, Academic Press. [MR0552278](#)
- [26] MÜLLER, A. (2001). Stochastic ordering of Multivariate Normal Distributions, *Annals of the Institute of Statistical Mathematics* **53**, No. 3, 567-575. [MR1868892](#)
- [27] MÜLLER, A. y STOYAN, D. (2002). *Comparison Methods for Stochastic Models and Risks*, Wiley. [MR1889865](#)
- [28] ROCKAFELLAR, R. T. (1970). *Convex Analysis*, 2<sup>a</sup> ed. Princeton University Press. [MR0274683](#)
- [29] RYFF, J. V. (1963). On the representation of doubly stochastic operators, *Pacific Journal of Mathematics* **13**, No. 4, 1379-1386. [MR0163171](#)
- [30] SHAKED, M. y SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*, Springer. [MR2265633](#)

- [31] SORDO, M.A. y RAMOS, H.M. (2007). Characterization of stochastic orders by L-functionals, *Statistical Papers* **48**, 249-263. [MR2295814](#)
- [32] SZEKLI, R. (1995). *Stochastic ordering and dependence in applied probability*, New York: Springer-Verlag. [MR1329324](#)