



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**Análisis comparado de índices de valoración de la dominación
estocástica**

Autora: Paula María Izquierdo García

Tutor: Carlos Matrán Bea

Año 2025

Resumen

A lo largo de nuestras vidas realizamos comparaciones constantemente. Desde una edad temprana aprendemos a establecer comparaciones entre dos objetos o personas. A medida que vamos creciendo y entendiendo el entorno, vamos realizando comparaciones más complejas, comparando múltiples objetos, personas o ideas. Estas comparaciones son esenciales para evaluar opciones, tomar decisiones y comprender nuestro entorno.

En este trabajo, lo que se quiere hacer son comparaciones entre muestras que se han obtenido de poblaciones, y que nos permiten hacer comparaciones entre distribuciones de probabilidad a partir de las muestras que se han obtenido. En la literatura estadística, este tipo de comparaciones típicamente se ha hecho recurriendo a indicadores como la media, la mediana o la varianza, entre otros, aunque, como veremos, la dominancia estocástica es el método adecuado para realizarlas. Dado que en la práctica estadística resulta inservible la dominancia estocástica estricta, se han desarrollado diversas suavizaciones para medir el grado en que una distribución puede considerarse “mejor” o “mayor” que otra. Este trabajo recopila y analiza dichas propuestas, a partir de los enfoques desarrollados por Álvarez-Esteban et al. (2017), del Barrio et al. (2025) y del Barrio et al. (2018).

Palabras clave: Dominancia estocástica, funciones cuantil, familias de localización y escala, distribución de Laplace.

Abstract

Throughout our lives, we constantly make comparisons. From an early age, we learn to set up comparisons between two objects or people. As we grow up and get to know our surroundings, we start making more complex comparisons, sizing up multiple objects, people, or ideas. These comparisons are essential for weighing up options, making decisions, and figuring out our environment.

In this study, the aim is to make comparisons between samples that have been drawn from populations, which allow us to compare probability distributions based on the samples obtained. In the statistical literature, such comparisons have typically been carried out using indicators such as the mean, median, or variance, among others, however, as we will see, stochastic dominance is the appropriate method for making these comparisons. Since strict stochastic dominance is impractical in statistical practice, various relaxations have been developed to measure the extent to which one distribution can be considered “better” or “greater” than another. This work compiles and analyzes these proposals, drawing on the approaches developed by Álvarez-Esteban et al. (2017), del Barrio et al. (2025) and del Barrio et al. (2018).

Key words: Stochastic dominance, quantile functions, location-scale families, Laplace distribution.

Índice general

Introducción	4
1 Orden estocástico	5
2 Funciones cuantil	7
3 De la comparación de medias a la dominancia estocástica	9
3.1 La paradoja de San Petersburgo	10
4 Dominancia estocástica	12
5 Familia exponencial	14
6 Familias de localización y escala	17
6.1 Definición e interpretación de los parámetros	17
6.2 Representación cuantil y propiedades de transformación	18
6.3 Ejemplos de familias de localización y escala	19
6.4 Propiedades estadísticas del parámetro de localización	21
6.5 Conclusión	21
7 Distribución de Laplace	22
7.1 Propiedades de la distribución:	23
7.1.1 Ejemplo: distribución Laplace con varianza unitaria	24
7.2 Aplicaciones:	24
8 Dominancia estocástica en la distribución doble exponencial	26
8.1 Dominancia estocástica en modelos de localización y escala	28
9 Relajación de la dominancia estocástica basada en las funciones cuantil:	30
10 Relajación de la dominancia estocástica basada en el modelo de contaminación:	34
11 Comparación entre los índices $\gamma(F, G)$ y $\pi(F, G)$	38
12 Relajación de la dominancia estocástica basada en el transporte óptimo	40
Conclusiones	44

Introducción

Esta memoria tiene como objetivo analizar el orden estocástico desde una perspectiva estadística. Aunque el orden estocástico es una herramienta rigurosa para establecer cuándo una variable aleatoria o una población es menor que otra, en la práctica resulta difícil garantizar estadísticamente esta relación, lo que limita su aplicabilidad directa.

El trabajo realizado consiste en el estudio de algunos trabajos relacionados con el orden estocástico, en los que se planten modificaciones que permitan su utilización en la práctica estadística. Los trabajos en los que me he basado son: Álvarez-Esteban et al. (2017), del Barrio et al. (2025) y del Barrio et al. (2018). No se pretende desarrollar innovaciones teóricas, sino exponer de forma estructurada los conceptos y resultados más significativos de estos trabajos, asumibles con el nivel propio del Grado de Estadística. Para ello, se han incorporado algunos ejemplos nuevos y se han modificado los ejemplos de los trabajos originales.

La estructura del trabajo es la siguiente: en el capítulo 1 se da la definición del orden estocástico y vemos un ejemplo ilustrativo; en el capítulo 2 se tratan las funciones cuantil, que son una herramienta muy importante para obtener una mejor interpretación del concepto de orden estocástico; el capítulo 3 repasa la evolución histórica del problema de la comparación de muestras; en el capítulo 4 se proporciona la definición de dominancia estocástica basada en funciones cuantil. Los capítulos 5, 6 y 7 están dedicados al estudio de la familia exponencial, los modelos de localización y escala, y la distribución de Laplace, respectivamente. En el capítulo 8 se analiza la dominancia estocástica en el caso de la distribución de Laplace.

En los capítulos posteriores se plantean varias relajaciones de la dominancia estocástica, en el capítulo 9 la relajación con las funciones cuantil; en el capítulo 10 la relajación basada en el modelo de contaminación; el capítulo 11 compara los dos índices planteados hasta el momento; y por último el capítulo 12 nos ofrece la relajación basada en el transporte óptimo, que viene utilizándose mucho últimamente en cuestiones muy generales, como la comparación de lenguajes, entre otras.

Capítulo 1

Orden estocástico

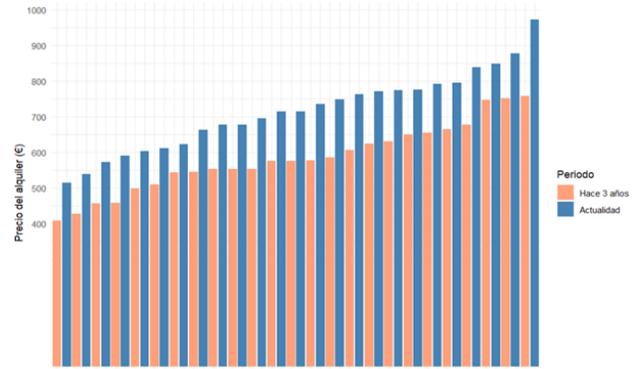
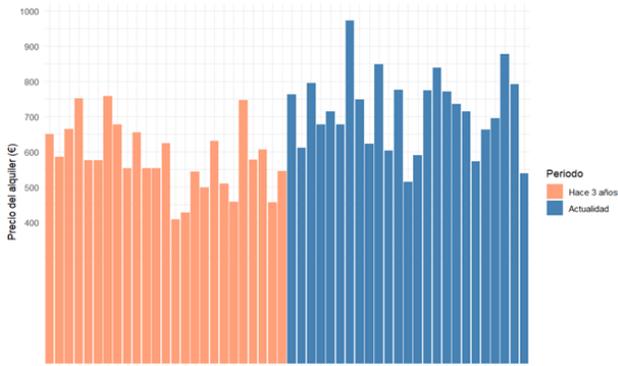
El concepto del orden estocástico fue planteado por primera vez por Francis Galton de forma indirecta; cuando su medio primo Charles Darwin le envió varias semillas, porque él estaba convencido que las semillas que habían sido cruzadas artificialmente producían plantas más vigorosas que las semillas autofecundadas, pero no sabía cómo demostrarlo cuantitativamente. Galton observó que había algunas plantas no cruzadas que eran más grandes que algunas cruzadas, por lo que no se podía comparar simplemente con máximos o promedios. Y se le ocurrió poner en una fila en orden las semillas cruzadas artificialmente y en otra fila ordenada las no cruzadas, e ir comparando la más pequeña con la más pequeña, la siguiente con la siguiente, y así con todas las posiciones. Por lo tanto, Galton descubrió lo que hoy llamamos una comparación de distribuciones por sus cuantiles; y anticipó el concepto de orden estocástico.

El objetivo del orden estocástico es establecer que la primera población es menor que la segunda en el sentido estocástico. Para ello, *afirmamos que F está estocásticamente dominada por G* , denotado como $F \leq_{st} G$.

La clásica definición establece que:

$$F \leq_{st} G \text{ si y solo si } F(x) \geq G(x) \forall x \in \mathbb{R}. \quad (1)$$

Para entender el significado del orden estocástico, vamos a plantearnos la afirmación: “El precio del alquiler de las viviendas ha aumentado respecto a hace 3 años”. Para ello vamos a tomar dos muestras de 25 viviendas en alquiler de toda España, una de las muestras de hace 3 años y la otra con datos de la actualidad.



En el primer gráfico podemos apreciar que las barras naranjas, que son las pertenecientes al precio de los alquileres de viviendas hace 3 años, son en principio más pequeñas que las de color azul, que pertenecen a los precios del alquiler de viviendas en la actualidad, aunque hay barras en color naranja que son más altas que algunas de color azul.

Para comprobar que los precios en la actualidad dominan estocásticamente a los precios de hace 3 años, vamos a ordenar los precios de hace 3 años de menor a mayor y los precios en la actualidad de menor a mayor, y comparar una a una las barras que corresponden a esa posición. Es decir, compararemos la barra menor del precio del alquiler hace 3 años con la barra menor de los alquileres en la actualidad, después la segunda menor del precio del alquiler hace 3 años con la segunda menor de los alquileres en la actualidad, y así con todos. Ahora en el segundo gráfico podemos ver que todas las barras azules son mayores que las barras naranjas.

Como veremos posteriormente, el segundo gráfico representa la comparación entre las funciones cuantil, y demostraría que la función cuantil asociada a los precios del alquiler de viviendas de hace 3 años es menor que la función cuantil correspondiente a los precios actuales. Por tanto, la situación ilustrada refleja el cumplimiento del orden estocástico.

Capítulo 2

Funciones cuantil

La función cuantil, F^{-1} , se define formalmente como:

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : t \leq F(x)\}, \quad t \in (0, 1) \quad (2)$$

Como la función de distribución es creciente y cumple $\lim_{x \rightarrow -\infty} F(x) = 0$, y $\lim_{x \rightarrow \infty} F(x) = 1$; entonces para todo $t \in (0, 1)$, existen $x_1, x_2 \in \mathbb{R}$ que cumplen $F(x_1) < t < F(x_2)$. Por tanto, el conjunto $\{x \in \mathbb{R} : t \leq F(x)\}$ no es vacío, ya que al menos contiene a x_2 , y además está acotado inferiormente por x_1 . Esta definición nos proporciona el valor mínimo que satisface la condición dada y garantiza la propiedad fundamental de la función cuantil, que actúa como una versión de la inversa de F . En particular, se cumple que:

$$F^{-1}(t) \leq x \quad \Leftrightarrow \quad t \leq F(x), \quad \forall t \in (0, 1).$$

Sabemos que la función cuantil es una función $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ y es creciente y continua por la izquierda.

Para demostrar que es creciente, podemos hacerlo viendo que si $t_1, t_2 \in (0, 1)$ con $t_1 < t_2$, entonces se tiene $\{x : t_2 \leq F(x)\} \subset \{x : t_1 \leq F(x)\}$.

La propiedad de ser continua por la izquierda puede razonarse: teniendo una sucesión $(t_n)_n \subset (0, 1)$ tal que $t_n \uparrow t_0$. La sucesión $(F^{-1}(t_n))_n$ es creciente y acotada por $F^{-1}(t_0)$, y si su límite $x_0 := \lim_n F^{-1}(t_n)$ verificara $x_0 < F^{-1}(t_0)$. Como sabemos $F^{-1}(t) \leq x \Leftrightarrow t \leq F(x)$, $\forall t \in (0, 1)$. Entonces tendríamos que $t_n \leq F(x_0) < t_0$, que es incompatible con $t_n \rightarrow t_0$. Por lo tanto, debe ser $F^{-1}(t_0) = \lim_{y \rightarrow t_0^-} F^{-1}(y)$.

Dado que la función cuantil es creciente, es una función $\beta(0, 1) \mid \beta$ -medible. Este hecho garantiza que, considerada como una variable aleatoria en el espacio probabilístico $((0, 1), \beta(0, 1), \ell)$, siendo ℓ la medida de Lebesgue, induzca una probabilidad en \mathbb{R} . Dicha probabilidad tendrá una función de distribución de la forma:

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

aunque no se haya visto en todo el grado de estadística la medida de Lebesgue, está incluida en el trabajo para poder entender mejor los conceptos que vamos a utilizar.

La función de distribución de esa variable aleatoria coincide con la función original que dio lugar a la función cuantil. Este resultado garantiza que cualquier función que satisfaga las propiedades de una función de distribución, corresponde efectivamente a la función de distribución de una probabilidad sobre la recta real.

La ventaja que tienen las funciones cuantil frente a las funciones de distribución es que cuando se tienen varias funciones cuantil, al estar todas definidas en el mismo espacio, se pueden hacer comparaciones entre ellas y resulta que las comparaciones del orden punto a punto coinciden con las que tendríamos con el orden estocástico.

Capítulo 3

De la comparación de medias a la dominancia estocástica

La media aritmética lleva usándose como medida de tendencia central desde los inicios de la estadística descriptiva. Sin embargo, las bases teóricas para ver cómo se distribuyen los datos en torno a la media no se establecen hasta el siglo XVIII con Pierre-Simon Laplace y Carl Friedrich Gauss, ambos establecieron las bases de la estadística moderna. Gauss introdujo la distribución normal, que describe cómo muchas variables tienden a agruparse en torno a un valor central, con una dispersión que disminuye a medida que nos alejamos del valor central.

Por su parte, una de las tantas cosas que podemos destacar de la investigación de Laplace es su trabajo sobre la convergencia en ley de sumas de variables aleatorias hacia la distribución normal. Aunque no lo formuló en términos modernos, es el precursor del Teorema Central del Límite, que nos indica que, bajo ciertas condiciones, la distribución de las medias muestrales tiende a una distribución normal independientemente de la forma de la distribución original de los datos. Este teorema supuso un avance crucial, ya que sienta las bases para la inferencia estadística.

Cuando queremos analizar la evaluación de tratamientos o mejoras en procesos productivos, uno de los enfoques más utilizados en estadística, es la comparación de medias de las distribuciones. Sin embargo, es fundamental actuar con cautela al inferir mejoras basadas en las medias u otras características de las distribuciones. Si llegamos a la conclusión de que $\mu_1 > \mu_2$; este resultado únicamente permitiría afirmar que, “en promedio”, el primer proceso genera valores más altos que el segundo.

A principios del siglo XX, William Sealy Gosset, bajo el seudónimo de “Student”, tuvo que enfrentarse al desafío de comparar medias de muestras pequeñas. El problema al que se enfrentaba era que no existían métodos adecuados basados en la distribución normal para comparar este tipo de datos. Ante este desafío, Gosset creó una nueva distribución que permitía realizar inferencias sobre las medias aun habiendo pocas observaciones, es la distribución t de Student. Esta distribución es la base de la prueba t de Student, una herramienta estadística muy utilizada para comparar medias. Poco después, en la década de 1920, Ronald Fisher introdujo el Análisis de Varianza (ANOVA), este método permitía comparar las medias de tres o más grupos de manera simultánea. El ANOVA

descompone la variabilidad de los datos en dos componentes: la variabilidad entre grupos y la variabilidad dentro de los grupos. Con el avance de la estadística, surge la necesidad de métodos que no dependan de la normalidad de los datos. Estos son los llamados métodos no paramétricos, los cuales son más robustos y flexibles.

Para tener una herramienta que compare las distribuciones completas, surge la dominancia estocástica. Su historia se remonta a la paradoja de San Petersburgo de Daniel Bernoulli, donde se introduce la idea de que las personas toman sus decisiones no solo basándose en el valor esperado de los resultados, sino también en su utilidad subjetiva. Fue el primer paso hacia la idea de que las distribuciones de probabilidad pueden compararse de manera más sofisticada que solo comparando sus medias. Es decir, la dominancia estocástica se origina de la teoría de la utilidad esperada en el siglo XVIII, aunque no llegó a formalizarse con el nombre de “dominación estocástica” hasta 1944 cuando John von Neumann y Oskar Morgenstern publicaron su obra “Theory of Games and Economic Behavior”.

3.1. La paradoja de San Petersburgo

La paradoja de San Petersburgo, tal como se expone en Levy (2016), páginas 24-26, fue un punto de referencia para la teoría de utilidad moderna. Consiste en que los inversores están dispuestos a pagar una cantidad muy pequeña por una inversión cuyo valor esperado es infinito. Para entender esto de una manera más fácil, reproduciremos el ejemplo clásico de la paradoja. El juego consiste en lanzar una moneda hasta que aparezca la primera cara; en cuanto aparece la primera cara, el juego se termina.

Sabemos que la probabilidad de que salga la primera cara en el lanzamiento número x es $\frac{1}{2^x}$, y la esperanza matemática del juego es $\sum_{x=1}^{\infty} \frac{1}{2^x} 2^{x-1} = \infty$.

Es decir, el valor esperado del juego es infinito. Según el Criterio del Máximo Retorno Esperado (MERC), el equivalente de certeza del juego es infinito, por lo que en teoría, un jugador racional estaría dispuesto a pagar cualquier cantidad, por grande que sea, por participar. Para que el juego sea justo la cantidad a pagar debería ser infinito. Entonces, ¿cuánto se está dispuesto a pagar? Experimentos con esta pregunta han revelado que la mayoría de los sujetos estarían dispuestos a pagar ante esta inversión arriesgada la cantidad de 2-3 \$. Ante esta contradicción, en el siglo XVIII, Nikolaus Bernoulli y Gabriel Cramer sugirieron que para los inversores lo que es importante es la utilidad derivada del dinero recibido y no el dinero en sí.

Pero hace falta una definición adecuada de utilidad y de utilidad esperada. La utilidad es un concepto fundamental en la teoría económica de la toma de decisiones bajo incertidumbre, y se utiliza para representar las preferencias de un individuo o un inversor frente a diferentes resultados inciertos. Estas preferencias pueden representarse matemáticamente a través de la función de utilidad, denominada $U(x)$, donde x representa los distintos resultados monetarios posibles.

La utilidad esperada, otro concepto en la teoría económica, permite representar y comparar las

preferencias de un individuo o inversor frente a diferentes resultados inciertos, se calcula como el valor esperado de la función de utilidad del resultado de esa inversión. Cuando se tienen dos distribuciones de probabilidad F y G para dos inversiones distintas, la utilidad esperada bajo cada una de ellas se calcula aplicando la función de utilidad del inversor a los posibles resultados de cada distribución, y luego tomando el valor esperado a esos resultados de utilidad.

El objetivo fundamental de un inversor es maximizar su utilidad esperada. Si se conoce por completo la función de utilidad, es posible realizar una ordenación completa de las distintas inversiones. Sin embargo, en la práctica, frecuentemente solo se dispone de información parcial sobre las preferencias del inversor. En estos casos, se recurre al concepto de orden estocástico, que permite establecer un ordenamiento parcial de las preferencias del inversor.

La definición de dominancia estocástica responde a la pregunta de si una distribución de probabilidad F domina estocásticamente a otra G . Si esto ocurre, implica que la utilidad esperada de F será mayor o igual que la utilidad esperada de G para todos los inversores. Por tanto, la dominancia estocástica es una herramienta para determinar qué inversión sería preferida por una amplia clase de inversores, constituyendo una alternativa más robusta al valor esperado para tomar decisiones.

Capítulo 4

Dominancia estocástica

En capítulos anteriores hemos visto que, en la práctica estadística, las comparaciones entre distribuciones se han realizado habitualmente a partir de sus medias. Sin embargo, este enfoque solo refleja los promedios. Para dar una definición más precisa, se recurre al concepto de orden estocástico, introducido formalmente en el capítulo 1 ecuación 1. Recordemos que, bajo esta definición clásica, una distribución G domina estocásticamente a otra distribución F cuando $F(x) \geq G(x) \forall x \in \mathbb{R}$. A continuación, veremos que esta relación 1 puede expresarse de forma más intuitiva utilizando las funciones cuantil, en cuyo caso se dice que:

$$F^{-1}(t) \leq G^{-1}(t), \quad \forall t \in (0, 1). \quad (3)$$

Demostración: A partir de la definición de función cuantil 2 y sabiendo la relación 1, tomemos un $t \in (0, 1)$. Esto implica que el conjunto $x : F(x) \geq t$ está contenido en $x : G(x) \geq t$.

Por lo tanto,

$$\inf\{x : F(x) \geq t\} \leq \inf\{x : G(x) \geq t\}$$

es decir:

$$F^{-1}(t) \leq G^{-1}(t)$$

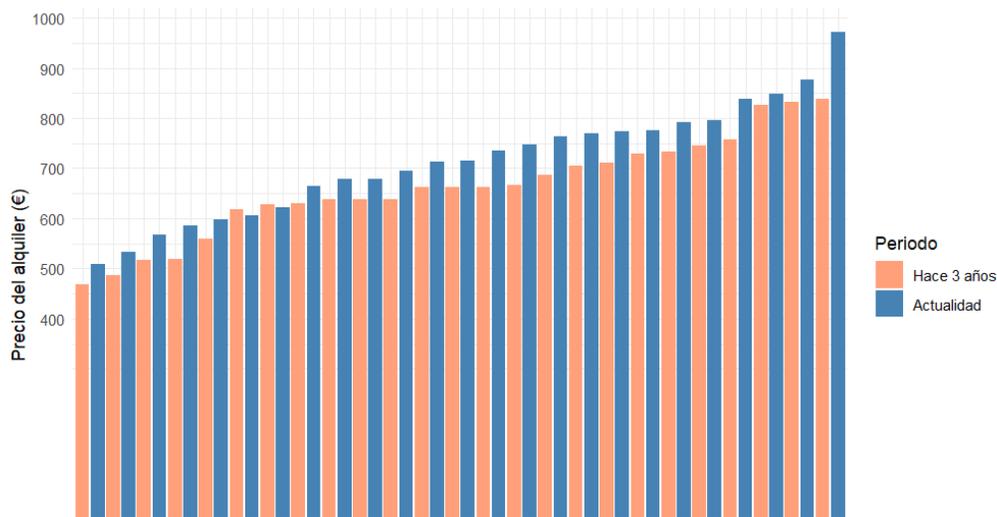
Esta formulación con funciones cuantil resulta particularmente útil en diversas aplicaciones estadísticas, al ofrecer una interpretación directa basada en los percentiles de las distribuciones.

Si tomamos de nuevo el ejemplo: “El precio del alquiler de las viviendas ha aumentado respecto a hace 3 años”, podemos interpretarlo en términos de funciones cuantil. Sean P y Q las distribuciones de los precios hace 3 años y en la actualidad, respectivamente, con funciones de distribución acumuladas F_P y F_Q , y funciones cuantil asociadas F_P^{-1} y F_Q^{-1} , definidas en $(0, 1)$. Entonces, al aplicar estas funciones a una muestra, se cumple que $L(F_P^{-1}(U)) = P$ y $L(F_Q^{-1}(U)) = Q$ si U es $\mathcal{U}(0, 1)$.

Por consiguiente, podemos explicar en términos generales $P \leq_{st} Q$ cuando existe un procedimiento de generación simultánea de datos de las distribuciones P y Q , esta simulación tiene la propiedad de que toda la muestra que salga de P va a ser uno a uno menor o igual a la correspondiente muestra de Q . Es decir, se cumple el orden elemento a elemento.

Si bien el interés por la dominancia estocástica es indiscutible, se ha señalado con frecuencia que esta relación puede ser demasiado estricta para garantizar su cumplimiento en la práctica. En particular, la dominancia podría darse en la mayor parte del dominio, pero fallar en una región acotada, o incluso ser desconocida o en ciertas partes del rango completo. Además, para concluir que la dominancia estocástica es válida, sería imprescindible recopilar evidencia estadística suficiente que permita rechazar la hipótesis nula en el marco de un problema de prueba adecuado:

$$H_0 : F \not\leq_{st} G, \quad H_1 : F \leq_{st} G.$$



Volviendo al ejemplo del precio de los alquileres, el gráfico anterior muestra una representación similar a la que vimos en el capítulo 1 del orden estocástico. La diferencia es que, en este caso, hay dos barras naranjas que están por encima de su correspondiente barra azul. Según la definición dada en 3 implica que no se cumple la dominancia estocástica, ya que el orden elemento a elemento falla para algunos pares. En nuestro caso, el orden se incumple en 2 de los 25 pares. Esto significa que para el 92 % de los valores correspondientes a los precios de las viviendas de hace 3 años (barras naranjas), ocuparían una posición menos favorable si se compararan dentro del contexto de los precios actuales de los alquileres (barras azules).

No obstante, la evaluación estadística de la dominancia entre distribuciones resulta inviable, dado que pequeñas variaciones en las colas de una distribución pueden evitar o facilitar el establecimiento de una relación de dominancia estocástica. Este resultado se formaliza en Berger (1988) en el contexto de una muestra única. Específicamente, aunque $F \leq_{st} G$, existen F^* y G^* tan cercanas como se desee a F y G , respectivamente, tales que $F^* \not\leq_{st} G^*$.

Capítulo 5

Familia exponencial

La familia exponencial ocupa un lugar destacado en la estadística clásica debido a que permite el uso de técnicas paramétricas. Una de sus principales ventajas es que, en este contexto, los procedimientos de estimación y de contraste de hipótesis pueden realizarse utilizando estadísticos suficientes. Estos permiten reducir la dimensión del problema a un número reducido de variables (una en el caso uniparamétrico, o k en el caso de familias k -paramétricas). Antes del desarrollo de herramientas computacionales potentes, era necesario contar con herramientas que fueran matemáticamente factibles para realizar las operaciones, y es lo que ofrecía la familia exponencial.

Como resultado, hubo determinadas distribuciones que adquirieron un gran protagonismo. Algunas de ellas son la binomial, adecuada para describir experimentos donde los resultados son éxitos o fracasos; la Poisson, utilizada para modelar sucesos raros; la normal, nos permite a través del teorema Central del Límite hacer aproximaciones a distribuciones generales; la doble exponencial, útil en contextos donde se requiere una distribución simétrica y en los que es común la aparición de eventos extremos o valores atípicos.

Formalmente, la familia exponencial se caracteriza porque la fórmula matemática de la distribución de densidad de probabilidad puede expresarse como:

$$f_X(x; \theta) = h(x) \exp(\eta(\theta)T(x) - A(\theta))$$

La familia exponencial cumple unas propiedades fundamentales que la convierten en una de las familias más importantes y útiles en el campo de la estadística. Una de las propiedades más destacadas es la suficiencia del estadístico, que se construye a partir de la función $T(x)$ y que permite realizar una reducción de los datos sin pérdida de información. Esta propiedad no depende de los valores concretos de los datos, sino de los resultados muestrales que tienen la misma imagen, es decir, de la que define el estadístico sobre el espacio muestral.

Para una muestra aleatoria independiente e idénticamente distribuida x_1, x_2, \dots, x_n , la función de densidad conjunta es:

$$f_X(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n h(x_i) \exp(\eta(\theta)T(x_i) - A(\theta)) =$$

$$\left[\prod_{i=1}^n h(x_i) \right] \exp \left(\eta(\theta) \sum_{i=1}^n T(x_i) - n * A(\theta) \right)$$

Lo que implica por el teorema de factorización de Fisher-Neyman, que $T(x) = \sum_{i=1}^n T(x_i)$ es el estadístico suficiente de θ , ya que toda la información sobre el parámetro contenida en los datos queda resumida en $T(x)$.

Otra propiedad esencial de las distribuciones pertenecientes a la familia exponencial es la Estimación por Máxima Verosimilitud. Una de las ventajas clave de estas familias es que la función de verosimilitud adopta una forma que permite una derivación sencilla de su estimador.

La función de verosimilitud para una muestra x_1, \dots, x_n es:

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n h(x_i) \exp(\eta(\theta)T(x_i) - A(\theta))$$

La log-verosimilitud se expresa como:

$$\begin{aligned} l(\theta; x_1, \dots, x_n) &= \sum_{i=1}^n \ln(h(x_i)) + \sum_{i=1}^n \eta(\theta) * T(x_i) - n * A(\theta) = \\ &= \sum_{i=1}^n \ln(h(x_i)) + \eta(\theta) \sum_{i=1}^n T(x_i) - n * A(\theta) \end{aligned}$$

El gradiente de la log-verosimilitud con respecto al vector de θ parámetros naturales es:

$$\nabla_{\theta} l(\theta; x_1, \dots, x_n) = \sum_{i=1}^n T(x_i) - n * \nabla_{\theta} A(\theta)$$

Por lo tanto, el estimador de máxima verosimilitud del parámetro natural θ_0 es el valor de θ que resuelve la ecuación:

$$\nabla_{\theta} A(\theta) = \frac{1}{n} \sum_{i=1}^n T(x_i)$$

Esta estructura permite derivar de forma más sencilla las ecuaciones de verosimilitud, y garantiza la existencia de un estimador de máxima verosimilitud único y globalmente óptimo.

Existen dos características especialmente relevantes del estimador de máxima verosimilitud. En primer lugar, el estimador depende únicamente de la media muestral del estadístico suficiente, es decir de: $\frac{1}{n} \sum_{i=1}^n T(x_i)$. En segundo lugar, como sabemos que $\nabla_{\theta} A(\theta) = E[T(X)]$, el estimador de máxima verosimilitud resuelve la ecuación

$$E_{\theta}[T(X)] = \frac{1}{n} \sum_{i=1}^n T(x_i).$$

Quiere decir, que el estimador de máxima verosimilitud se obtiene igualando la media muestral del estadístico suficiente con su media poblacional, proporcionando una interpretación intuitiva y práctica del procedimiento de estimación en el marco de las familias exponenciales.

La relación entre la familia exponencial y el orden estocástico proviene de la propiedad de razón de verosimilitud monótona (MLR) (Lehmann (1959)). Sabemos que una familia exponencial uniparamétrica tiene la propiedad MLR si el término $\eta(\theta)$ es una función no decreciente del parámetro θ .

Podemos demostrar que si $\eta(\theta)$ es monótono, se da MLR en las familias exponenciales. Si realizamos la razón de verosimilitud para dos valores del parámetro $\theta_1 < \theta_2$, con la forma de la distribución de densidad:

$$\frac{f(x; \theta_2)}{f(x; \theta_1)} = \frac{h(x) \exp(\eta(\theta_2) T(x) - A(\theta_2))}{h(x) \exp(\eta(\theta_1) T(x) - A(\theta_1))} = \exp\{(\eta(\theta_2) - \eta(\theta_1)) T(x) - (A(\theta_2) - A(\theta_1))\}$$

Para que la razón de verosimilitud sea una función no decreciente de $T(x)$, el coeficiente de $T(x)$ en el exponente debe ser positivo: $(\eta(\theta_2) - \eta(\theta_1)) \geq 0$; $\eta(\theta_2) \geq \eta(\theta_1)$. Lo cual se cumple solo si $\eta(\theta)$ es una función no decreciente de θ .

Si una familia de distribuciones $f_\theta(x)$ posee la propiedad MLR con respecto a un estadístico $T(x)$, entonces, en esta familia se va a dar la dominancia estocástica en x con respecto al parámetro θ . Específicamente, si $\theta_1 < \theta_2$, la distribución con parámetro θ_2 va a dominar estocásticamente a la distribución con el parámetro θ_1 , es decir: $F_{\theta_1} \leq_{st} F_{\theta_2}$.

En resumen, la familia exponencial permite reducir la dimensión del problema mediante la suficiencia del estadístico, facilita la obtención de estimadores de máxima verosimilitud de forma intuitiva, y establece una conexión directa entre la razón de verosimilitud monótona y la dominancia estadística. Estas características explican la importancia de esta familia tanto en la estadística clásica como en la moderna.

Capítulo 6

Familias de localización y escala

Acabamos de ver que la familia exponencial permite realizar cálculos eficientes y buscar optimizaciones, gracias a su estructura matemática particular. En un sentido semejante, las familias de localización y escala se construyen sobre una parametrización basada directamente en la localización y la forma de la distribución.

6.1. Definición e interpretación de los parámetros

Las familias de localización y escala combinan parámetros de localización y escala con el objetivo de formar un conjunto de distribuciones de probabilidad a partir de una distribución estándar. En la recta real estos modelos se definen a partir de una función de distribución F_0 , la cual consideramos como la distribución canónica del modelo. Asumimos que F_0 es una función de distribución no degenerada. Una distribución pertenece a esta familia si su función de distribución puede expresarse como:

$$F(x) = F_0\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}$$

para ciertos parámetros $\mu, \sigma \in \mathbb{R}$, con $\sigma > 0$.

El parámetro de localización determina la “ubicación” o el desplazamiento de la distribución, lo que significa que al cambiar el parámetro, la distribución entera se desplaza a lo largo del eje x sin alterar la forma de la distribución. Si $\mu < 0$ entonces el traslado es a la izquierda; y si $\mu > 0$ entonces es hacia la derecha.

El parámetro de escala determina la dispersión o variabilidad de la distribución, quiere decir que la modificación de este parámetro estira o comprime la distribución a lo largo del eje x . Si $\sigma > 1$ dilata la distribución, se vuelve más ancha; y si $0 < \sigma < 1$ la distribución se contrae, se vuelve más estrecha.

Si Z es una variable aleatoria con una distribución conocida, y se realiza:

$$X = \mu + \sigma Z, \quad \mu, \sigma \in \mathbb{R}, \quad \sigma > 0 \quad (4)$$

siendo μ el parámetro de localización y σ el parámetro de escala; entonces se establece un vínculo funcional entre X y Z , eso es una familia de localización y escala asociada a la distribución de probabilidad que tiene Z .

En particular, en el caso que Z sea una variable continua en todo el dominio y su función de densidad sea $f(x)$, podemos determinar la familia de localización y escala de $f(x)$ como:

$$\left\{ f(x | \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) : \mu \in \mathbb{R}, \sigma > 0 \right\}$$

en este caso μ es el parámetro de localización y σ el de escala.

6.2. Representación cuantil y propiedades de transformación

Como ya se mencionó, sea P_0 una distribución de probabilidad dada, con función de distribución F_0 , y función cuantil F_0^{-1} . Si buscamos una variable aleatoria X_0 que tenga como ley de probabilidad $P_0 : L(X_0) = P_0$, entonces su función de distribución es F_0 y su función cuantil es F_0^{-1} . En particular siempre podríamos tomar F_0^{-1} como una variable aleatoria, $F_0^{-1} : (0, 1) \rightarrow \mathbb{R}$ donde cumple que su función de distribución es F_0 . Es decir, manejada como variable aleatoria, la función cuantil es una representación inmediata de la distribución de probabilidad de la que es función cuantil.

Una de las principales ventajas de este enfoque es la equivarianza bajo transformaciones lineales crecientes. Si aplicamos una transformación afín del tipo $\mu + \sigma X_0$, con $\sigma > 0$, entonces $\mu + \sigma F_0^{-1}$ va a seguir siendo una función cuantil, donde se van a mantener las propiedades de ser una función creciente y continua por la izquierda de las funciones cuantil. Este resultado resulta especialmente útil en distribuciones que difieren en cambios de localización y escala.

Si tenemos la función de distribución de $\mu + \sigma F_0^{-1}$, su ley $L(\mu + \sigma F_0^{-1})$ es $F_0(\frac{x - \mu}{\sigma})$. Esta ley es lo que buscamos con los cambios de localización y escala. De hecho, siempre podemos calcular la función de distribución de una variable aleatoria $\mu + \sigma X_0$ a partir de la X_0 :

$$P(\mu + \sigma X_0 \leq x) = P(X_0 \leq \frac{x - \mu}{\sigma}) = F_0(\frac{x - \mu}{\sigma})$$

Pero además como X_0 es una variable aleatoria, y F_0^{-1} está igualmente distribuida que la X_0 :

$$F_0^{-1} \stackrel{d}{=} X_0 \Rightarrow \mu + \sigma F_0^{-1} \stackrel{d}{=} \mu + \sigma X_0$$

6.3. Ejemplos de familias de localización y escala

Existen muchas familias de distribuciones que pertenecen a la familia de localización y escala, algunas de ellas son: la distribución normal, la doble exponencial, la uniforme, o la distribución de Cauchy.

Entonces, si se toma como punto de partida una distribución de referencia con media cero y varianza uno, las familias de localización y escala basadas en la distribución de referencia van a permitir dar al parámetro de localización y al de escala el sentido tradicional de media y desviación típica. Podemos demostrarlo sabiendo 4 y que $E(Z) = 0$ y $Var(Z) = 1$, entonces se tiene:

$$E(X) = E(\mu + \sigma Z) = E(\mu) + E(\sigma Z) = \mu + \sigma * E(Z) = \mu + \sigma * 0 = \mu$$

$$Var(X) = Var(\mu + \sigma Z) = Var(\sigma Z) = \sigma^2 * Var(Z) = \sigma^2$$

A continuación se presentan dos ejemplos clásicos: la distribución normal estándar $N(0, 1)$ y la doble exponencial univariada clásica $L(0, 1)$.

Ejemplo 1: Sea $X \sim N(0, 1)$ una variable aleatoria con distribución normal de media 0 y varianza 1. Su función de densidad está dada por:

$$f_x(x) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})}, x \in \mathbb{R}$$

La esperanza matemática de X se define como:

$$E(X) = \int_{-\infty}^{\infty} x f_x(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})} dx$$

La función de densidad $f(x)$ es simétrica respecto al 0: $f(x) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})} = f(-x)$. Al estar dentro de la integral multiplicada por x , se convierte en una función impar, y podemos dividir la integral en dos:

$$E(X) = \int_{-\infty}^0 x f_x(x) dx + \int_0^{\infty} x f_x(x) dx$$

Como acabamos de ver que $x f(x)$ es una función impar: $\int_{-\infty}^0 x f_x(x) dx = - \int_0^{\infty} x f_x(x) dx$. Es decir, los dos términos se cancelan y se concluye que:

$$E(X) = 0$$

La varianza se define como: $Var(X) = E(X^2) - [E(X)]^2$, como $E(X) = 0 \Rightarrow$

$$Var(X) = E(X^2) = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{(-\frac{x^2}{2})} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{(-\frac{x^2}{2})} dx$$

Es conocido que: $\int_{-\infty}^{\infty} x^2 e^{-ax^2} dx = \frac{1}{2} \sqrt{\pi} a^{-3/2}$

En nuestra función $a = \frac{1}{2} \Rightarrow \int_{-\infty}^{\infty} x^2 e^{(-\frac{x^2}{2})} dx = \frac{1}{2} \sqrt{\pi} (\frac{1}{2})^{-3/2} = \frac{1}{2} \sqrt{\pi} 2^{3/2} = \sqrt{2\pi}$

Por tanto, volviendo a la varianza:

$$Var(X) = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1$$

Ejemplo 2: Ahora se va a realizar con una doble exponencial de parámetros $\mu = 0$ (localización) y $b = 1$ (escala). Su función de densidad es de la forma:

$$f_x(x) = \frac{1}{2} e^{-|x|}, x \in \mathbb{R}$$

Su esperanza matemática es:

$$E(X) = \int_{-\infty}^{\infty} x f_x(x) dx = \int_{-\infty}^{\infty} x \frac{1}{2} e^{-|x|} dx$$

La función $f_x(x)$ es simétrica respecto al origen (función par) al igual que en el caso de la normal estándar. Al estar multiplicada por x es una función impar.

$$E(X) = \int_{-\infty}^0 x f_x(x) dx + \int_0^{\infty} x f_x(x) dx = -A + A = 0$$

La varianza es: $Var(X) = E(X^2) - [E(X)]^2$, como hemos visto $E(X) = 0 \Rightarrow Var(X) = E(X^2)$

$$Var(X) = \int_{-\infty}^{\infty} x^2 f_x(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|} dx = \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|} dx$$

La función $x^2 e^{-|x|}$ es par, podemos reducir la integral al intervalo $[0, \infty)$ y multiplicar por 2:

$$Var(X) = \frac{1}{2} \left[\int_{-\infty}^0 x^2 e^x dx + \int_0^{\infty} x^2 e^{-x} dx \right] = \int_0^{\infty} x^2 e^{-x} dx$$

Es una integral de tipo gamma: $\int_0^{\infty} x^n e^{-x} dx = n!$ Por tanto:

$$Var(X) = 2$$

Como se ha demostrado, en el caso de la normal estándar el parámetro de escala coincide con la desviación típica, ya que la varianza es 1. En cambio, en la distribución de Laplace $L(0, 1)$, la varianza es 2, significando que el parámetro de escala $b = 1$ no representa directamente la desviación típica (que en este caso es $\sqrt{2}$). Este resultado evidencia que la interpretación del parámetro de escala como desviación típica depende de la varianza de la distribución de referencia. No siempre se puede asumir que dicho parámetro coincide con la desviación típica, salvo que la distribución base tenga varianza igual a 1.

6.4. Propiedades estadísticas del parámetro de localización

Si $f(x)$ es una función par, es decir, simétrica respecto de 0, entonces la función transformada $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ es simétrica respecto de μ , y por lo tanto μ es su mediana. Además, si el valor esperado de $f(x)$ existe, entonces μ también es el valor esperado de $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$.

El parámetro μ tiene la propiedad de que, entre todos los valores reales $a \in \mathbb{R}$, es el que minimiza la esperanza del error cuadrático $E[(X - a)^2]$. Es decir, $\mu = \arg \min_{a \in \mathbb{R}} E[(X - a)^2]$, significando que la esperanza matemática es el valor que mejor representa a la distribución en términos de proximidad cuadrática. A continuación se presenta la demostración formal de este resultado.

Sea X una v.a. con $E(X^2) < \infty$, queremos minimizar con respecto a $a \in \mathbb{R}$:

$$g(a) = E[(X - a)^2] = E[X^2 - 2aX + a^2] = E(X^2) - 2aE(X) + a^2$$

Derivando respecto a a e igualando la derivada a 0:

$$g'(a) = \frac{d}{da}[E(X^2) - 2aE(X) + a^2] = -2E(X) + 2a$$

$$g'(a) = 0 \Rightarrow -2E(X) + 2a = 0 \Rightarrow a = E(X)$$

Para verificar que es un mínimo realizamos la segunda derivada:

$$g''(a) = \frac{d^2}{da^2}g(a) = 2 > 0 \Rightarrow \text{Mínimo global}$$

Por tanto se confirma que, en cualquier distribución donde μ coincide con $E(X)$, el parámetro μ no tiene solo el sentido de localización sino también ser el punto que mejor representa a la función de probabilidad, por lograr minimizar la dispersión cuadrática respecto a la variable aleatoria X .

6.5. Conclusión

Estas familias desempeñan un papel fundamental en la teoría estadística y probabilística, ya que permiten modelar una gran variedad de fenómenos mediante simples transformaciones paramétricas. Su estructura facilita no solo la generación de nuevas distribuciones a partir de una distribución base, sino también la comparación entre distribuciones. En particular, la utilización de las funciones cuantil resulta especialmente útil en este contexto, al ser una herramienta eficaz para el análisis y la comparación de distribuciones dentro de estas familias.

Capítulo 7

Distribución de Laplace

En 1774 Pierre-Simon Laplace introdujo una distribución que hoy conocemos como distribución de Laplace, o doble exponencial; siendo esta la distribución para la cual la función de verosimilitud se maximiza al establecer el parámetro de localización igual a la mediana de los valores observados de un número impar de variables aleatorias independientes e idénticamente distribuidas. Este resultado se conoce como la primera Ley de Laplace, y se propuso como alternativa a la media para poder ajustar errores de medición simétricos. (Ver pág. 164 del libro Johnson et al. (1995))

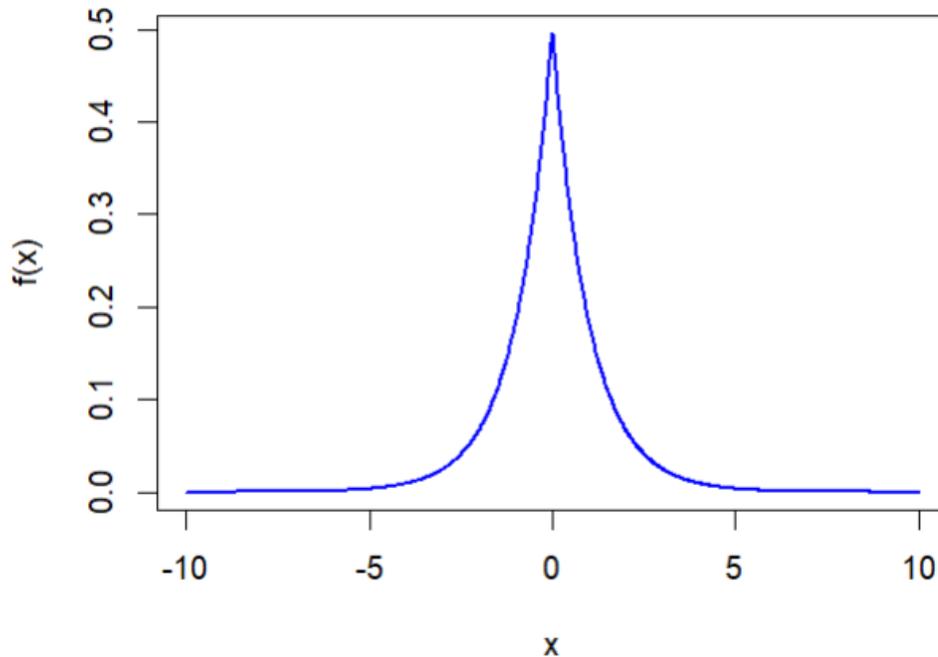
También se puede generar esta distribución como la diferencia de dos variables aleatorias exponenciales independientes e idénticamente distribuidas.

La función de densidad de la distribución de Laplace está definida como:

$$f(x) = \frac{1}{2b} e^{-\frac{|x - \mu|}{b}} \quad b > 0$$

donde μ representa el parámetro de localización y b el parámetro de escala.

Su representación es de la forma:



7.1. Propiedades de la distribución:

Algunas de las características estadísticas de la doble exponencial son que la media vale μ y su varianza es $2b^2$.

La función de distribución F vale:

$$F(x) = \begin{cases} \frac{1}{2}e^{\frac{x-\mu}{b}} & \text{si } x \in (-\infty, \mu) \\ 1 - \frac{1}{2}e^{-\frac{x-\mu}{b}} & \text{si } x \in [\mu, \infty) \end{cases}$$

La función cuantil F^{-1} , que es la inversa de la función de distribución F , se formula como:

$$F^{-1}(p) = \begin{cases} \mu + b * \ln(2p) & \text{si } 0 \leq p \leq \frac{1}{2} \\ \mu - b * \ln[2 * (1 - p)] & \text{si } \frac{1}{2} \leq p \leq 1 \end{cases}$$

El primer cuartil es $q_1 = \mu - b * \ln(2)$. El segundo cuartil es la mediana que tiene el valor $q_2 = \mu$. El tercer cuartil es $q_3 = \mu + b * \ln(2)$.

Al tener la función de densidad primitiva, podemos derivar de ella de forma explícita tanto la función de distribución como la función cuantil. Esto representa una ventaja significativa frente a distribuciones como la normal, cuya función cuantil existe, pero no puede expresarse con una fórmula exacta. En el caso de la Laplace, es posible calcular cuantiles, realizar simulaciones y ob-

tener resultados analíticos con facilidad, lo que la convierte en un excelente ejemplo ilustrativo.

7.1.1. Ejemplo: distribución Laplace con varianza unitaria

Obsérvese que en el ejemplo 2 del capítulo 6 (familias de localización y escala) se utilizaba la distribución $L(0, 1)$, obteniéndose $E(X) = 0$ y $Var(X) = 2$. Si en su lugar se hubiese utilizado la distribución $L(0, \frac{1}{\sqrt{2}})$, cuya función de densidad es:

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

entonces se obtendría una varianza unitaria, es decir, $E(X) = 0$ y $Var(X) = 1$.

Demostración: Como la distribución es simétrica respecto de 0, $E(X) = 0$ (ver ejemplo 2, capítulo 6). La varianza se calcula como:

$$Var(X) = E(X^2) - [E(X)]^2 = E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{\sqrt{2}}{2} e^{-\sqrt{2}|x|} dx = \sqrt{2} \int_0^{\infty} x^2 e^{-\sqrt{2}x} dx$$

Es una integral de tipo gamma: $\int_0^{\infty} x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$. En nuestro caso $n = 2$, $a = \sqrt{2}$.

$$Var(X) = \sqrt{2} \frac{2}{2\sqrt{2}} = 1$$

7.2. Aplicaciones:

La distribución de Laplace tiene muchas aplicaciones interesantes en diversas áreas, como los patrones de consumo y la estructura de preferencias de los consumidores; o en ingeniería para el análisis de errores, la resistencia de los materiales y la confiabilidad de estos.

En contextos más recientes, la distribución de Laplace ha cobrado importancia en el ámbito de Machine Learning, donde se utiliza la distribución de Laplace como distribución prioritaria en los modelos bayesianos en las regulaciones de tipo LASSO (ver Tibshirani (1996), capítulo 5).

La regulación de tipo L_1 , también llamada LASSO, se utiliza junto con la regulación L_2 o Ridge, para evitar el sobreajuste en modelos de Machine Learning. Estas técnicas penalizan los coeficientes grandes en los modelos de regresión. En la L_1 se añade un término a la función de coste que es proporcional a la magnitud de los pesos en valor absoluto.

Dado que los β_j son los coeficientes del modelo, la restricción del LASSO $\sum |\beta_j| \leq s$ es equivalente a añadir un término de penalización de la forma $\lambda \sum |\beta_j|$ a la suma de los cuadrados de los residuos. Esta penalización L_1 está estrechamente relacionada con la distribución de Laplace, porque es equivalente a asumir una distribución a priori doble exponencial sobre los coeficientes

β_j . La densidad de la distribución de Laplace es:

$$f(\beta_j) = \frac{1}{2\tau} \exp\left(-\frac{|\beta_j|}{\tau}\right)$$

donde $\tau = 1/\lambda$ es el parámetro de escala, relacionado con la intensidad de penalización λ . Al tomar logaritmos se obtiene:

$$\log f(\beta_j) = \log\left(\frac{1}{2\tau}\right) - \frac{|\beta_j|}{\tau}$$

Es decir, $|\beta_j|$ es proporcional a la (menos) log-densidad de la distribución de Laplace.

Dicho de otra manera, desde un punto de vista bayesiano, la utilización de la penalización L_1 es equivalente a asumir que los coeficientes β_j provienen de una distribución de Laplace centrada en cero. Esta distribución concentra más densidad cerca del 0 y de las colas, lo que implica que el LASSO tiende a producir estimaciones con muchos coeficientes que son exactamente igual a cero, o con valores grandes cuando son realmente necesarios.

Además, la distribución de Laplace se aplica en el llamado mecanismo de Laplace, uno de los mecanismos aleatorios más conocidos en privacidad diferencial; que consiste en ir añadiendo ruido a la respuesta siguiendo una distribución de Laplace. (Ver Near y Abuah (2025), capítulo 4).

La privacidad diferencial es un marco que busca proteger la información individual, asegurando que los resultados del análisis estadístico cambien lo menos posible ante la modificación de un único registro en la base de datos.

Formalmente, el mecanismo de Laplace para la función $f : D \rightarrow \mathbb{R}^M$ se define como:

$$M_L(D, f, \epsilon) = f(D) + (N_1, \dots, N_M)$$

donde $D \in \mathcal{D}$, y N_i son variables aleatorias independientes que se obtienen a partir de una distribución de Laplace $N_i \sim L\left(0, \frac{\Delta f}{\epsilon}\right)$.

Entonces, podemos destacar, que a medida que aumentamos ϵ , menor es el parámetro b , implicando que menor será el ruido añadido.

Estos dos ejemplos que acabamos de explicar aparecen en el documento a modo ilustrativo. Para profundizar en su desarrollo teórico o técnico, se recomienda consultar las referencias específicas.

Capítulo 8

Dominancia estocástica en la distribución doble exponencial

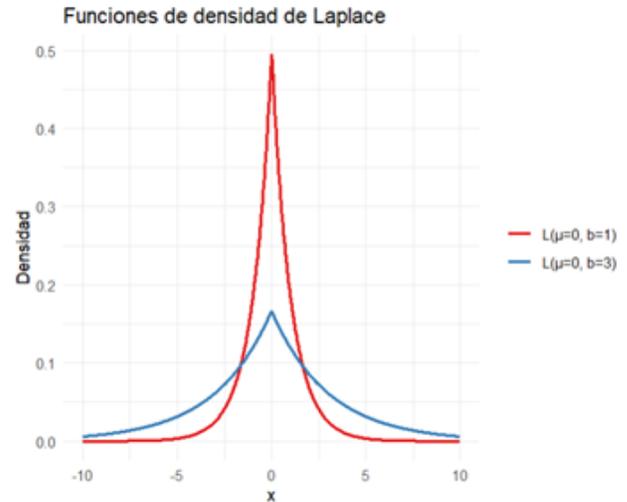
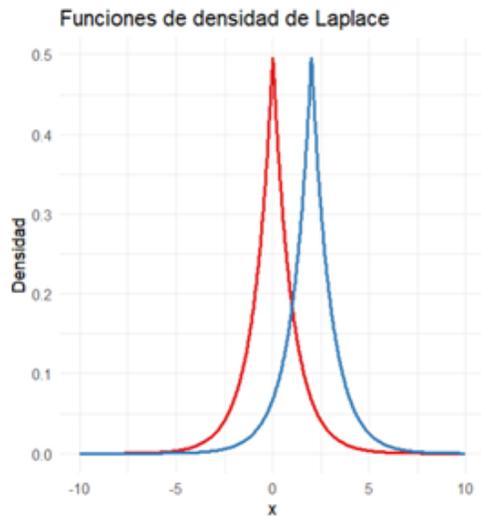
En el contexto del análisis de dominancia estocástica, de forma paralela a lo realizado en Álvarez-Esteban et al. (2017) vamos a manejar esta noción dentro del marco de la distribución doble exponencial.

Matemáticamente, la distribución de Laplace surge como la distribución de la diferencia entre dos variables aleatorias independientes con distribuciones exponenciales idénticas.

Una forma eficiente de simular una variable con distribución de Laplace estándar es utilizando dos variables auxiliares:

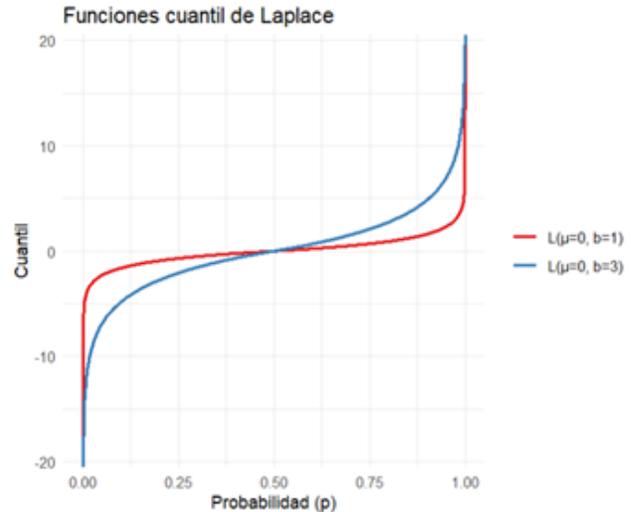
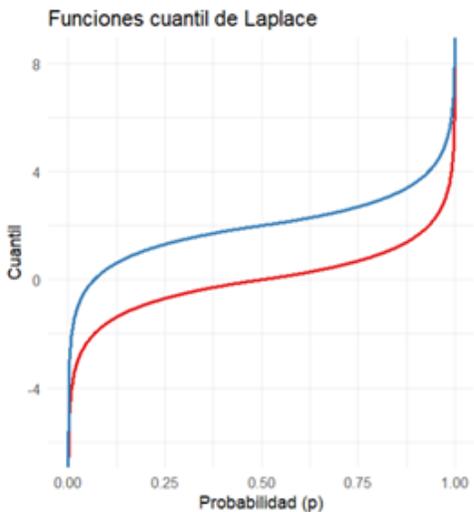
- $V \sim Exp(1)$, una variable aleatoria que sigue una distribución exponencial estándar.
- $I \sim Bernoulli(0,5)$, una variable aleatoria independiente que sigue una Bernoulli estándar.

Definiendo la variable $U = (2I - 1)V$, se obtiene $U \sim L(0, 1)$. El término $(2I - 1)$ actúa como una variable de signo: cuando $I = 1$, nos queda $(2I - 1) = 1$ y se asigna un signo positivo a V ; si $I = 0$, entonces $(2I - 1) = -1$ y V tiene signo negativo. Como I sigue una Bernoulli con probabilidad 0,5 ambos signos tienen la misma probabilidad. Para que se pueda llevar a cabo esta simulación es necesario que haya independencia entre I y V , para que se pueda mantener la simetría de la distribución de Laplace. Esta técnica de simulación se describe con más detalle en Siegrist (2025), capítulo 5.28.



En un primer análisis, se comparan dos distribuciones de Laplace con el objetivo de estudiar la influencia de la variación en el parámetro de ubicación μ , mientras el parámetro de escala b permanece constante. Para ello, se representan gráficamente las distribuciones $L(\mu = 0, b = 1)$ (en color rojo) y $L(\mu = 2, b = 1)$ (en color azul). Las gráficas permiten identificar las diferencias entre ambas distribuciones, derivadas del cambio en la ubicación del parámetro μ .

En un segundo gráfico, se estudia el efecto de modificar el parámetro de escala b , manteniendo constante el parámetro de ubicación μ . En este caso, se comparan las distribuciones $L(\mu = 0, b = 1)$ (en color rojo) y $L(\mu = 0, b = 3)$ (en color azul).



Además, se representan las funciones cuantil asociadas a estas distribuciones para proporcionar un análisis más detallado de la relación de dominancia estocástica entre ellas. Se observa que, en general, no se cumple la condición $F^{-1} \leq_{st} G^{-1}$ en todo el dominio, lo que implica que se puede rechazar la existencia de dominancia estocástica entre las distribuciones consideradas.

Por último, se pretende determinar hasta qué valores del dominio se verifica la dominancia estocástica, identificando los intervalos en los que la relación $F^{-1} \leq_{st} G^{-1}$ se mantiene válida. Este enfoque permite un estudio más preciso de las condiciones bajo las cuales una distribución domina estocásticamente a otra. En capítulos posteriores analizaremos varias relajaciones de la dominancia estocástica basadas en diferentes modelos.

8.1. Dominancia estocástica en modelos de localización y escala

Sean ahora F_1, F_2 dos funciones de distribución que pertenecen a la familia de localización y escala generada a partir de F_0 . Las correspondientes funciones cuantil adoptan la forma:

$$F_1(t) = F_0\left(\frac{t - \theta_1}{\lambda_1}\right) \implies F_1^{-1}(t) = \theta_1 + \lambda_1 F_0^{-1}(t), \quad t \in [0, 1],$$

$$F_2(t) = F_0\left(\frac{t - \theta_2}{\lambda_2}\right) \implies F_2^{-1}(t) = \theta_2 + \lambda_2 F_0^{-1}(t), \quad t \in [0, 1],$$

donde $\theta_1, \theta_2 \in \mathbb{R}$, y $\lambda_1, \lambda_2 \in \mathbb{R}^+$.

La relación de orden estocástico entre dos funciones de distribución $F_1 \leq_{st} F_2$ se describiría como $F_1(x) \geq F_2(x)$ para todo $x \in \mathbb{R}$. Como hemos visto anteriormente, podemos poner esta relación en función de las funciones cuantil, lo que facilita la interpretación:

$$F_1^{-1}(t) \leq F_2^{-1}(t) \quad \forall t \in [0, 1].$$

Analizaremos la relación entre funciones de distribución que pertenecen a una misma familia de localización y escala, asociada a una función de distribución no degenerada F_0 , a través de los parámetros involucrados:

$$\begin{aligned} F_0\left(\frac{t - \theta_1}{\lambda_1}\right) &\leq F_0\left(\frac{t - \theta_2}{\lambda_2}\right), \\ \theta_1 + \lambda_1 F_0^{-1}(t) &\leq \theta_2 + \lambda_2 F_0^{-1}(t), \\ (\lambda_1 - \lambda_2) F_0^{-1}(t) &\leq \theta_2 - \theta_1, \quad \forall t \in [0, 1]. \end{aligned}$$

Analizaremos ahora cómo se comporta la desigualdad anterior según distintas relaciones entre los parámetros de escala:

a) Si $\lambda_1 = \lambda_2$:

$$0 \cdot F_0^{-1}(t) \leq \theta_2 - \theta_1 \implies \theta_1 \leq \theta_2.$$

b) Si $\lambda_1 > \lambda_2$:

$$F_0^{-1}(t) \leq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}, \quad \forall t \in [0, 1].$$

Al ser $F_0^{-1}(t)$ una función creciente, el mayor valor del primer término de la desigualdad se alcanza en el extremo superior del intervalo, por lo tanto: $F_0^{-1}(1) \leq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$.

c) Si $\lambda_1 < \lambda_2$:

$$F_0^{-1}(t) \geq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}, \quad \forall t \in [0, 1].$$

Al ser $F_0^{-1}(t)$ una función creciente, el menor valor del primer término de la desigualdad se alcanza en el extremo inferior del intervalo, por lo tanto: $F_0^{-1}(0) \geq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$.

A partir de las relaciones establecidas previamente, se va a analizar casos en los que varíe el soporte. En la distribución doble exponencial, al igual que en el caso de la normal, el soporte no está acotado ni inferior ni superiormente, por tanto, el orden estocástico solo se va a dar en el caso a) cuando $\lambda_1 = \lambda_2$. Debido que en el caso b) el extremo superior del intervalo $F_0^{-1}(1)$ va a valer ∞ y nunca se va a dar la relación $F_0^{-1}(1) \leq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$. En el caso c), de una forma similar, el extremo inferior del intervalo $F_0^{-1}(0)$ vale $-\infty$ y es imposible que se de $F_0^{-1}(0) \geq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$.

Si ahora nos enfocamos en el caso en que el soporte de la distribución está acotado. Un ejemplo representativo de esta situación es la distribución uniforme. En este contexto, podemos garantizar que el orden estocástico se cumple únicamente si se verifica al menos una de las siguientes tres condiciones a), b) o c).

Consideremos ahora las distribuciones uniformes $U(0, 1)$, $U(a, b)$ y $U(c, d)$, a cuyas funciones de distribución asociadas las denotamos como F_0 , F_1 y F_2 , respectivamente. Podemos demostrar que, bajo cualquiera de las condiciones mencionadas, para que se cumpla la dominancia estocástica en la distribución uniforme es necesario que $a \leq c$ y $b \leq d$.

A continuación, se demuestra esta afirmación para cada uno de los casos:

a) $\lambda_1 = \lambda_2$ y $\theta_1 \leq \theta_2$:

Dado que $b - a = d - c$ y $a \leq c$ podemos demostrar que $b \leq d$.

b) $\lambda_1 > \lambda_2$ y $F_0^{-1}(1) \leq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$:

Como $b - a > d - c$ y $F_0^{-1}(1) \leq \frac{c-a}{(b-a)-(d-c)}$ sabemos que el denominador es positivo, por lo tanto, para que se cumpla la desigualdad se tiene que cumplir que $a \leq c$.

Podemos deducir que $b - a > d - c \implies b \leq d$

c) $\lambda_1 < \lambda_2$ y $F_0^{-1}(0) \geq \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$:

Al ser $b - a < d - c$ y $F_0^{-1}(0) \geq \frac{c-a}{(b-a)-(d-c)}$, sabemos que el denominador es negativo, por lo tanto, para que se cumpla la desigualdad se tiene que cumplir que $a \leq c$.

Podemos deducir que $b - a < d - c \implies b \leq d$

En resumen, se ha mostrado que la dominancia estocástica en modelos de localización y escala dependen únicamente del soporte de la distribución y la relación entre los parámetros. En particular, en la distribución de Laplace que no está acotada, la dominancia solo es posible bajo la igualdad de los parámetros de escala.

Capítulo 9

Relajación de la dominancia estocástica basada en las funciones cuantil:

Dada que las definiciones basadas en funciones de distribución acumulada y funciones cuantil son equivalentes, también resulta razonable fijar un intervalo dentro de $(0,1)$ y comprobar si la relación se mantiene dentro de dicho intervalo. En este sentido, la normalización proporcionada por la transformación cuantil ofrece una ventaja adicional. En lugar de enfrentar la elección arbitraria del intervalo, es posible analizar la longitud del conjunto donde la relación no se cumple, definida como:

$$\gamma(F, G) := \ell(t \in (0, 1) : F^{-1}(t) \leq G^{-1}(t)). \quad (5)$$

Este índice, introducido por Álvarez-Esteban et al. (2017), permite medir cómo de alejadas están F y G respecto al orden estocástico, siendo $\gamma(F, G) = 0$ indicativo de un ajuste perfecto. Por ejemplo, si tenemos un valor de $\gamma(F, G) = 0,05$ significa que el 95 % de los pacientes obtienen mejores resultados con el tratamiento asociado a G en comparación con el tratamiento asociado a F .

En el caso del modelo de localización y escala (LS), como ocurre con el modelo de Laplace, el cálculo y la interpretación de este índice son simples y altamente informativos.

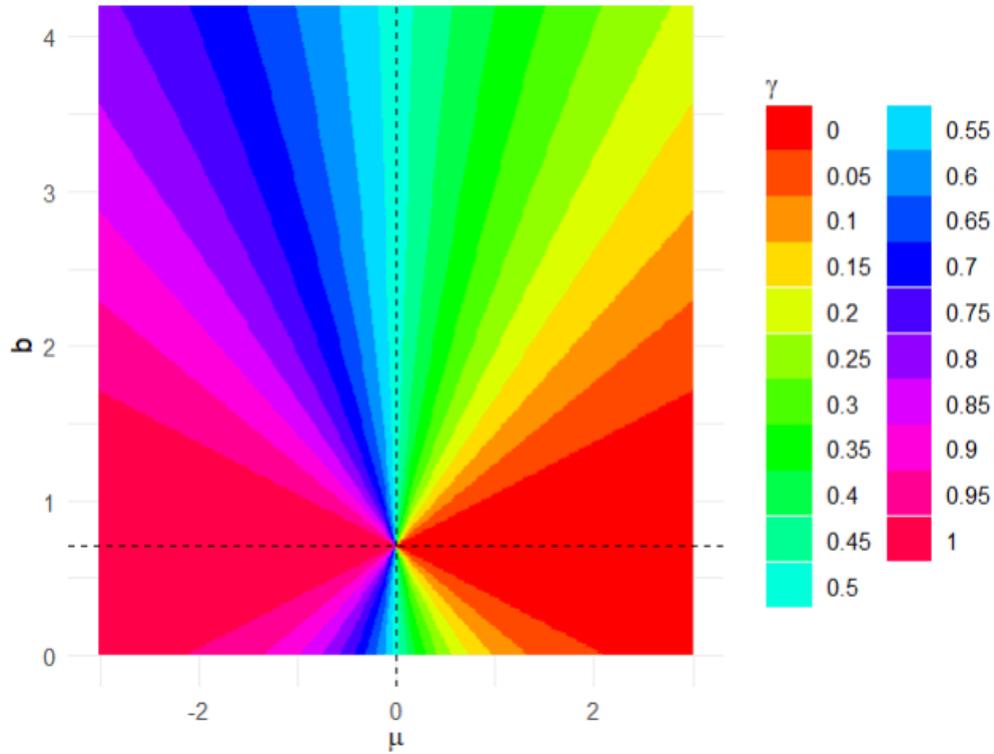


Figura 1: Gráfico de contorno de $\gamma(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$

Lo que estamos representando en el gráfico es el índice $\gamma(F, G)$ definido en 5, donde F es una doble exponencial $L(0, 1/\sqrt{2})$ y la G una doble exponencial de parámetros $(\mu, \sigma/\sqrt{2})$. En el eje x estamos representando el parámetro μ y en el eje y el parámetro b . El punto central del gráfico, correspondiente a $(0, 1/\sqrt{2})$, representa el caso en que los parámetros de F y G tienen los mismos valores.

El objetivo del gráfico es analizar para diferentes valores de los parámetros μ y b como afectan a los niveles del orden estocástico, sabiendo que el índice $\gamma(F, G) = 0$, como hemos dicho antes, es el ajuste perfecto y solo se obtiene cuando $\mu \geq 0$ y $b = 1/\sqrt{2}$.

El gráfico de contorno permite observar que niveles moderados e incluso altos de desacuerdo con el orden estocástico (hasta $\gamma(F, G) \sim 0,5$) pueden ser compatibles con un aumento en la media, incluso en el modelo de Laplace. Por ejemplo, si F representa la distribución de Laplace con varianza unitaria $L(0, 1/\sqrt{2})$ y G corresponde a la distribución $L(0,8, 2)$, entonces se obtiene $\gamma(F, G) = 0,25$. Como hemos visto antes, esto significa que se obtienen mejores resultados con el nuevo tratamiento, pero peores resultados para el 25 % de los pacientes.

Para poder realizar una comparación con la figura 3 de Álvarez-Esteban et al. (2017), donde se representa $\gamma(N(0, 1), N(\mu, \sigma^2))$, ha sido necesario normalizar la distribución de Laplace. Para ello, se transforma el parámetro de escala b a $b = \sigma/\sqrt{2}$, de modo que la varianza coincida con la de la normal. El resultado se muestra en la figura 1; obsérvese que el efecto visual es similar. Sin embargo, el área donde γ está entre 0 y 0,05 (y, respectivamente, donde γ está entre 0,95 y 1) es más amplia en el caso de la distribución normal. Esto se debe a la forma de sus colas: al ser más

delgadas en la normal, hay menor solapamiento entre distribuciones desplazadas, lo que facilita la dominancia estocástica casi segura en más regiones.

Dos funciones cuantil, en la familia de localización y escala, tienen un punto de corte, el punto t_0 , que existe si y solo si $\lambda_1 \neq \lambda_2$ y $F_0^{-1}(t_0) = \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$. En ese caso, si existe un punto de corte, es único. Se demuestra a continuación dicha afirmación, basándonos en la distribución doble exponencial.

Demostración: Supongamos que las dos funciones cuantil son de la forma:

$$F_1^{-1}(t) = \theta_1 + \lambda_1 F_0^{-1}(t), \quad F_2^{-1}(t) = \theta_2 + \lambda_2 F_0^{-1}(t) \quad t \in (0, 1)$$

Para buscar el punto de corte t_0 igualamos las dos expresiones

$$\theta_1 + \lambda_1 F_0^{-1}(t) = \theta_2 + \lambda_2 F_0^{-1}(t)$$

$$(\lambda_1 - \lambda_2) F_0^{-1}(t) = \theta_2 - \theta_1$$

$$F_0^{-1}(t) = \frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}$$

Para que exista el punto t_0 , tiene que suceder que $\lambda_1 \neq \lambda_2$. Si existe el punto, va a ser único, debido a que la función cuantil es continua y estrictamente creciente en $(0, 1)$, porque el soporte de la doble exponencial no está acotado ni inferior ni superiormente, por lo tanto, toma cada valor una sola vez.

También, si tenemos el conjunto $\{y : F_{\theta_1, \lambda_1}^{-1}(y) \leq F_{\theta_2, \lambda_2}^{-1}(y)\}$, puede expresarse como: $(0, F_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}))$ o $(F_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2}), 1)$. Entonces, para la longitud del intervalo $\ell(\{F_{\theta_1, \lambda_1}^{-1} > F_{\theta_2, \lambda_2}^{-1}\})$, dependiendo del signo de $\lambda_1 - \lambda_2$, puede calcularse como $1 - F_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})$ o como $F_0(\frac{\theta_2 - \theta_1}{\lambda_1 - \lambda_2})$.

Además, el índice puede expresarse en términos de la función de distribución de referencia F_0 :

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

lo que permite centrar el análisis en las comparaciones respecto a F_0 .

Ahora bien, dados dos funciones de distribución F y G dentro del modelo LS, si el objetivo es establecer un nivel de concordancia del 95 %, por ejemplo, con la dominancia estocástica de G sobre F según el índice $\gamma(F, G)$, sería suficiente identificar el punto de cruce de las funciones cuantil y comprobar si el intervalo correspondiente a $\{F^{-1} \leq G^{-1}\}$ presenta una longitud igual o superior a 0.95.

La longitud del intervalo en el modelo de Laplace que estamos analizando, se puede calcular como:

$$\gamma(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2})) = 1 - F_0\left(\frac{\mu}{1 - \sigma}\right)$$

Como podemos observar en la Figura 1:

$$\gamma(L(0, 1/\sqrt{2}), L(\mu, 1/\sqrt{2})) = \begin{cases} 0 & \text{si } \mu \geq 0, \\ 1 & \text{si } \mu < 0. \end{cases}$$

Otra de las apreciaciones que se pueden hacer de la Figura 1 es ver que $\gamma(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$ es constante a lo largo de todos los rayos $\{(\mu, b) : \mu = C|b - 1|, b > 0\}$, para algún $C > 0$, y se vuelve único en $\mu = 0$ y $\sigma = 1$. Al igual que cuando b es mayor que $1/\sqrt{2}$, es posible que $\mu > 0$ mientras que $\gamma(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$ es tan próximo a $\frac{1}{2}$ como se quiera. Esto confirma nuevamente que el nuevo tratamiento podría ser peor hasta para el 50% de la población.

Capítulo 10

Relajación de la dominancia estocástica basada en el modelo de contaminación:

Otra de las alternativas destacadas es la propuesta por Álvarez-Esteban et al. (2016), que está basada en el modelo de contaminación. Este enfoque plantea que siempre existe un $\alpha \in [0, 1]$, que permite una descomposición de mixtura de la forma:

$$\begin{cases} F = (1 - \alpha)\tilde{F} + \alpha R \\ G = (1 - \alpha)\tilde{G} + \alpha S \end{cases} \quad (6)$$

donde \tilde{F} , \tilde{G} , R, S son funciones de distribución, y $\tilde{F} \leq_{st} \tilde{G}$.

Para entender más fácilmente el modelo de contaminación, volvamos a la afirmación: “El precio del alquiler de las viviendas ha aumentado respecto a hace 3 años”.

Supongamos ahora que queremos comparar esta afirmación entre dos regiones distintas: el Principado de Asturias y Castilla y León. En particular, nos interesa analizar en cuál de las dos comunidades autónomas ha aumentado más el precio del alquiler.

Podemos modelar los precios de alquiler en Castilla y León como una mezcla:

$$F = (1 - \alpha)\tilde{F} + \alpha R$$

donde F representa a las viviendas alquiladas en Castilla y León; \tilde{F} son las viviendas en alquiler en las ciudades de Castilla y León; y R son las viviendas en alquiler en los pueblos de Castilla y León. De manera análoga para el Principado de Asturias:

$$G = (1 - \alpha)\tilde{G} + \alpha S$$

donde los términos son análogos: \tilde{G} representa las ciudades y S las zonas rurales.

En este contexto, se observa que en Castilla y León el precio del alquiler ha aumentado en las

ciudades, pero ha disminuido o se ha mantenido en los pueblos. Por tanto, cuando se incluye información rural en la muestra, esta “contamina” la señal de interés, que es la evolución del precio en áreas urbanas. En cambio, en el Principado de Asturias, el precio del alquiler ha aumentado tanto en las ciudades como en los pueblos, aunque más intensamente en las ciudades. Aun así, la presencia de los pueblos sigue introduciendo cierta mezcla.

El modelo de contaminación permite interpretar esta mezcla como una combinación de poblaciones, en la que una fracción α corresponde a una subpoblación secundaria que podría distorsionar la comparación directa entre regiones. Por eso, si queremos analizar el cambio real en los precios del alquiler entre las ciudades de ambas regiones, que son poblaciones homogéneas, deberíamos comparar directamente \tilde{F} y \tilde{G} . Si se verifica que $\tilde{F}(x) \geq \tilde{G}(x) \forall x$, entonces se concluye que las ciudades del Principado de Asturias dominan estocásticamente a las de Castilla y León en cuanto a precios de alquiler, es decir, los precios han aumentado más.

Esta descomposición, como hemos visto en el ejemplo anterior, puede interpretarse como un proceso de generación aleatoria de dos etapas, basadas en la distribución de Bernoulli. Al generarse valores con probabilidad $1 - \alpha$ a partir de F (respectivamente G), se elige la distribución \tilde{F} (respectivamente \tilde{G}) que cumple la relación de dominancia estocástica.

Sabemos que en la práctica los estadísticos procesan las muestras, eliminando los ruidos o “rarezas” que existan. Quiere decir que el enfoque que se está llevando a cabo resulta especialmente útil, debido a que la descomposición planteada (6), puede considerarse como una vecindad de contaminación en torno a algunas distribuciones latentes \tilde{F} y \tilde{G} .

Cuando α es suficientemente grande, la relación (6) siempre se va a cumplir, incluyendo el caso extremo $\alpha = 1$.

Por el contrario, si α es pequeño, se interpreta que la mayor parte de la distribución G domina a F . Por tanto, quiere decir que el menor α que permite las descomposiciones propuestas puede entenderse como la proporción de la población que no se ajusta al modelo de orden estocástico.

$$\alpha(F, G) := \inf \{ \alpha \in [0, 1] : \text{se verifica la descomposición de (6)} \}$$

La teoría desarrollada por Álvarez-Esteban et al. (2016) permite definir $\pi(F, G)$ como:

$$\pi(F, G) = \sup_{x \in \mathbb{R}} (G(x) - F(x)). \quad (7)$$

y demuestra que $\pi(F, G) = \alpha(F, G)$. Esta definición facilita la evaluación de $F \leq_{st}^{\pi} G$, es decir, permite cuantificar la proximidad a la dominancia estocástica siempre que $\pi \geq \alpha(F, G) = \pi(F, G)$.

Si $\pi(F, G) \in (0, 1)$ y F y G tienen densidades continuas f y g , respectivamente, entonces existe un punto x_0 tal que $\pi(F, G) = G(x_0) - F(x_0)$ y x_0 satisface $f(x_0) = g(x_0)$. Al igual que el índice $\gamma(F, G)$, $\pi(F, G)$ es invariante ante transformaciones estrictamente crecientes.

Para las funciones de distribución F y G , se verifica $F \leq_{st}^{\pi} G$ si y solo si

$$F^{-1}(y) \leq G^{-1}(\pi + y), \quad 0 < y < 1 - \pi.$$

Con esta definición podemos extraer algunas conclusiones como que $\pi(F, G) = 0$ si y solo si $F \leq_{st} G$. Aunque la situación más común sea la desigualdad estricta, también se mantiene $\pi(F, G) + \pi(G, F) \leq 1$, e incluso es posible que $\pi(F, G)$ y $\pi(G, F)$ sean pequeñas simultáneamente.

Este índice puede representarse de manera similar al índice definido mediante la transformación cuantil, ofreciendo una herramienta útil para el análisis de dominancia estocástica y sus aproximaciones.

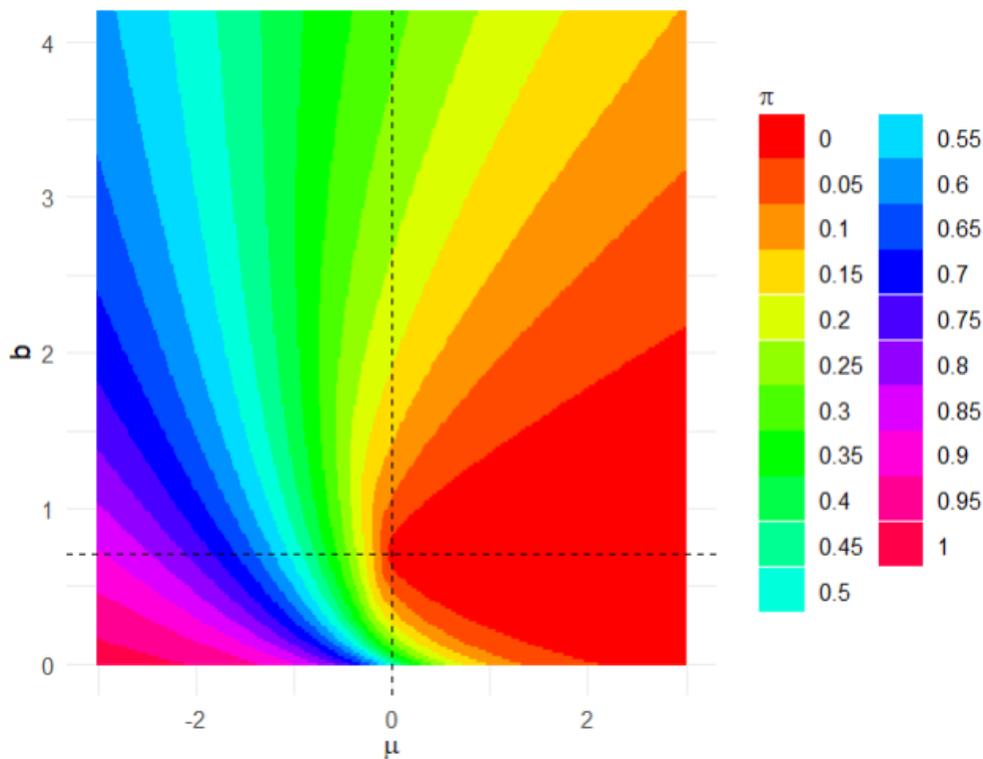


Figura 2: Gráfico de contorno de $\pi(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$

El gráfico de contorno de $\pi(F, G)$, permite observar el comportamiento del modelo de contaminación en una distribución de Laplace, es similar al realizado con el índice $\gamma(F, G)$, donde el eje x representa al parámetro de localización μ y el eje y al parámetro de escala b , que en términos de varianza es $\sigma/\sqrt{2}$. Las líneas de nivel tienen forma cóncava, el área de color rojo tiene una forma más simétrica, y a medida que se cambia de color y b aumenta, las líneas de nivel se estiran más en la parte superior.

Si analizamos, como en la figura 1, el ejemplo donde F es $L(0, 1/\sqrt{2})$ y G corresponde a $L(0, 8, 2)$, nos sale $\pi(F, G) = 0, 1$ indicando que el nuevo tratamiento tiene mejores resultados para el 90 % de los pacientes.

En comparación a la figura 5 de Álvarez-Esteban et al. (2017), el área roja es más pequeña en nuestro gráfico, pero a medida que cambiamos de nivel y aumentamos b , las líneas de nivel en nuestro gráfico se abren más en la parte superior.

Con base en esta interpretación, es posible plantear la hipótesis:

$$\begin{cases} H_0 : \pi(F_{L(\mu_1, b_1)}, F_{L(\mu_2, b_2)}) \geq 0,05 \\ H_1 : \pi(F_{L(\mu_1, b_1)}, F_{L(\mu_2, b_2)}) < 0,05 \end{cases}$$

Si rechazamos la hipótesis nula, podríamos concluir que el segundo tratamiento es mejor que el primero, si estamos dispuestos a excluir el 5% de las observaciones en cada extremo de las distribuciones.

Sin embargo, para el contraste $H_0 : \mu_2 \leq \mu_1$, $H_1 : \mu_2 > \mu_1$ no se obtendría una conclusión similar sobre la mejora asociada al nuevo tratamiento.

A continuación, se analiza el comportamiento del índice $\pi(F, G)$ bajo los modelos de localización y escala.

$$\sup_{x \in \mathbb{R}} (F_{\theta_2, \lambda_2}(x) - F_{\theta_1, \lambda_1}(x)) = \sup_{x \in \mathbb{R}} (F_{\frac{\theta_2 - \theta_1}{\lambda_1}, \frac{\lambda_2}{\lambda_1}}(x) - F_{0,1}(x)) \quad (8)$$

Por lo que tanto, considerando solo el caso $F = F_{0,1}$, se obtiene:

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

No existe una única expresión para todos los modelos de localización y escala para $\pi(F_{0,1}, F_{\theta, \lambda})$, ya que el mayor problema que se da es que π depende de la distribución base F_0 .

En el caso de la distribución de Laplace estándar tenemos $\pi(L(0, 1), L(\mu, b)) = L_0(\frac{\tilde{x} - \mu}{b\sqrt{2}}) - L_0(\tilde{x})$,

siendo $\tilde{x} = \frac{b * \ln(b) - \mu}{1 + b}$. En particular, si $b = 1$, vemos que el punto de corte es en $\tilde{x} = \frac{\mu}{2}$.

Si consideramos el caso donde L_0 represente a la distribución de Laplace con varianza unitaria: $\pi(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2})) = L_0(\frac{\tilde{x} - \mu}{\sigma}) - L_0(\tilde{x})$, siendo $\tilde{x} = \frac{\sigma * \ln(\sigma) - \mu}{1 + \sigma}$. En el caso en que $\sigma = 1$, el punto de corte se va a dar, igual que antes, en $\tilde{x} = \frac{\mu}{2}$.

Capítulo 11

Comparación entre los índices $\gamma(F, G)$ y $\pi(F, G)$

En esta sección, se analiza la relación entre los índices $\gamma(F, G)$ (definido en la ecuación (5)) y $\pi(F, G)$ (definido en la ecuación (7)).

Esta relación justifica la utilización de $P(X > Y)$, debido a que todos los índices pueden expresarse en función de esta probabilidad. Para cada par de variables aleatorias (X, Y) con funciones de distribución marginales F, G , se cumple que:

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

Si consideramos la representación de las funciones cuantil $(X, Y) = (F^{-1}, G^{-1})$ y también que $P(X > Y) = \ell(F^{-1} > G^{-1}) = \gamma(F, G)$. Como además sabemos $\pi \geq \pi(F, G)$ y la definición de $\pi(F, G)$. Se puede afirmar que, para cualquier par de funciones de distribución F y G :

$$\pi(F, G) \leq \gamma(F, G) \tag{9}$$

Ambos índices son relajaciones de la dominancia estocástica, que toman valores en el intervalo $[0, 1]$ y que cumplen que $\gamma(F, G) = 0$ si y solo si $\pi(F, G) = 0$. Las dos igualdades son equivalentes a $F \leq_{st} G$.

El índice $\pi(F, G)$ está relacionado con el concepto de recorte, en cambio, $\gamma(F, G)$ está relacionado con la winsorización. Ambas técnicas son muy populares en el análisis de datos, ya que se utilizan para evitar la influencia de las colas de las distribuciones, en especial cuando hay valores atípicos (outliers).

Para $\pi(F, G)$, tras eliminar, podar, una proporción δ_0 de la cola superior de F y la cola inferior de G , podemos recalcar que $\pi(F, G) \leq \delta_0$ si y solo si las funciones de distribución \tilde{F} y \tilde{G} satisfacen $\tilde{F} \leq_{st} \tilde{G}$.

En cambio, la winsorización consiste en remplazar las colas de las distribuciones por el valor percentil de cada extremo. Si consideramos que $F^{-1}(t) \leq G^{-1}(t)$ para t en algún intervalo de $(\gamma_1, 1 - \gamma_2)$, entonces $\gamma(F, G) \leq \delta_0$ si $\gamma_1 + \gamma_2 \leq \delta_0$.

Sin duda, $F^{-1} \leq G^{-1}$ en $(\gamma_1, 1 - \gamma_2)$ si y solo si las funciones de distribución \tilde{F} y \tilde{G} , obtenidas por winsorización de F y G cumplen $\tilde{F} \leq_{st} \tilde{G}$, donde los cuantiles para la winsorización son γ_1 y $1 - \gamma_2$.

En la Sección 3 de Álvarez-Esteban et al. (2017) se presentan métodos inferenciales para evaluar la consistencia de la hipótesis nula $H_0 : \pi(F, G) \geq \delta_0$ frente a la alternativa $H_1 : \pi(F, G) < \delta_0$, al igual que la evaluación de $H_0 : \gamma(F, G) \geq \delta_0$ contra $H_1 : \gamma(F, G) < \delta_0$. Estos procedimientos permiten obtener evidencia estadística para una dominancia estocástica aproximada.

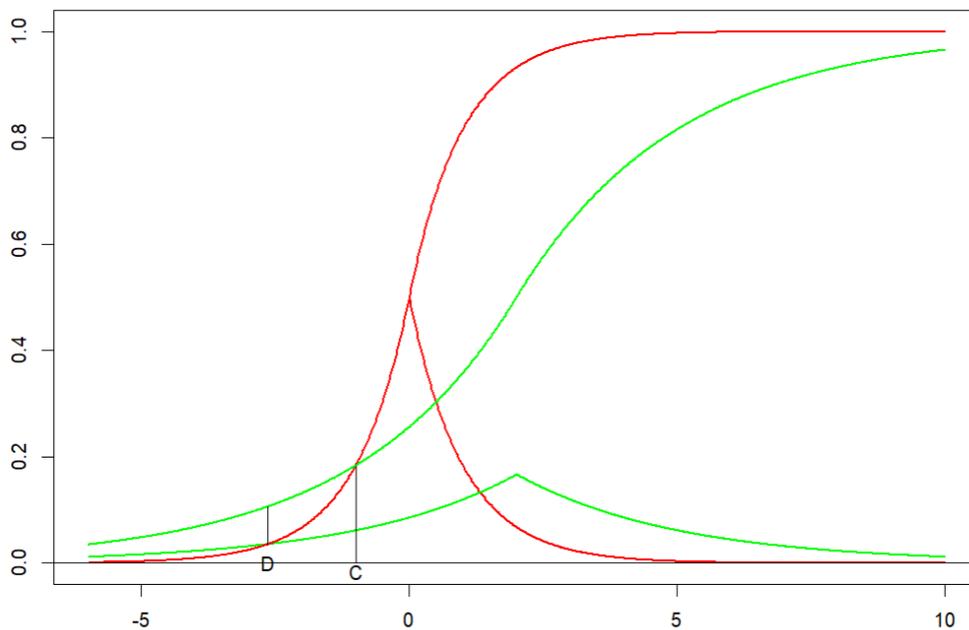


Figura 3: Densidad y distribución de Laplace de $L(0, 1)$ y $L(2, 3)$

Para comprender mejor el comportamiento de las funciones de $\pi(L(0, 1), L(\mu, b))$ y $\gamma(L(0, 1), L(\mu, b))$, vamos a utilizar la gráfica. En esta representación, las funciones de distribución y densidad de la distribución de Laplace $L(0, 1)$ aparecen representadas en color rojo, mientras que en color verde aparecen las de $L(2, 3)$.

El punto C es el punto donde se cortan las dos funciones de distribución. El valor $\gamma(L(0, 1), L(2, 3))$ es la longitud del punto C al eje de abscisas.

Por otro lado, el punto D representa el lugar de intersección entre las dos funciones de densidad. El valor de $\pi(L(0, 1), L(2, 3))$ se calcula como la longitud vertical entre las dos funciones de distribución acumulada en la posición definida por el punto D , y la función de distribución acumulada de $L(\mu, b)$ está por encima de la función de distribución acumulada $L(0, 1)$.

Con base en la interpretación gráfica, se puede concluir que $\pi(L(0, 1), L(\mu, b))$ tiene un comportamiento menor en comparación con $\gamma(L(0, 1), L(\mu, b))$.

Capítulo 12

Relajación de la dominancia estocástica basada en el transporte óptimo

Otro de los índices planteados para asegurar casi la dominancia estocástica es el denominado ε_{ω_2} , basado en el transporte óptimo. Pero más en particular basado en la versión L_2 de la distancia de Wasserstein.

Este índice, al igual que el modelo de contaminación, está basado en una descomposición de mixtura:

$$F = (1 - \pi) \tilde{F} + \pi H_F$$

$$G = (1 - \pi) \tilde{G} + \pi H_G$$

para alguna función de distribución $\tilde{F}, H_F, \tilde{G}, H_G$ con $\tilde{F} \leq_{st} \tilde{G}$.

Nos hemos basado en el trabajo del Barrio et al., 2018 en el que vienen hechas todas las demostraciones, por las que se llega a que el índice basado en el transporte óptimo es:

$$\varepsilon_{\mathcal{W}_2}(F, G) := \frac{\int (F^{-1}(t) - G^{-1}(t))_+^2 dt}{\mathcal{W}_2^2(F, G)} \quad (10)$$

Que es lo mismo que escribirlo de la forma:

$$\varepsilon_{\mathcal{W}_2}(F, G) := \frac{\int (F^{-1}(t) - G^{-1}(t))_+^2 dt}{\int |F^{-1}(t) - G^{-1}(t)|^2 dt}$$

Debido a que sabemos la que distancia de Wasserstein está definida como:

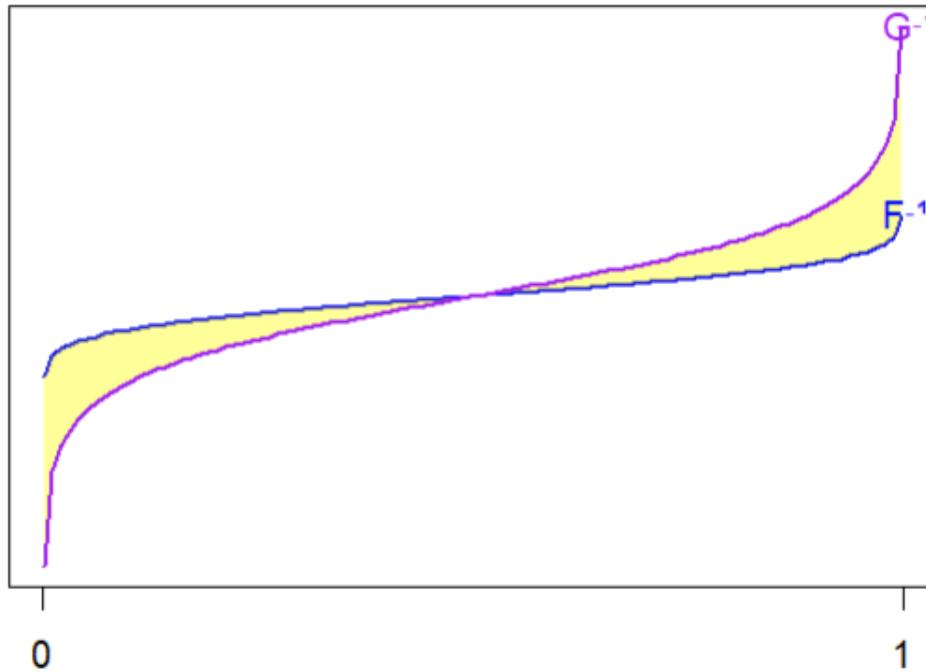
$$\omega_2(F, G) = \sqrt{\int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt}$$

y que $x_+ = \max(x, 0)$.

También podemos expresar otro índice con la versión L_1 de la distancia de Wasserstein, quedando de la forma:

$$\varepsilon_{W_1}(F, G) := \frac{\int (F^{-1}(t) - G^{-1}(t))_+ dt}{\int |F^{-1}(t) - G^{-1}(t)| dt}$$

Para entender la definición de ε_{ω_1} , se puede explicar con la siguiente gráfica:



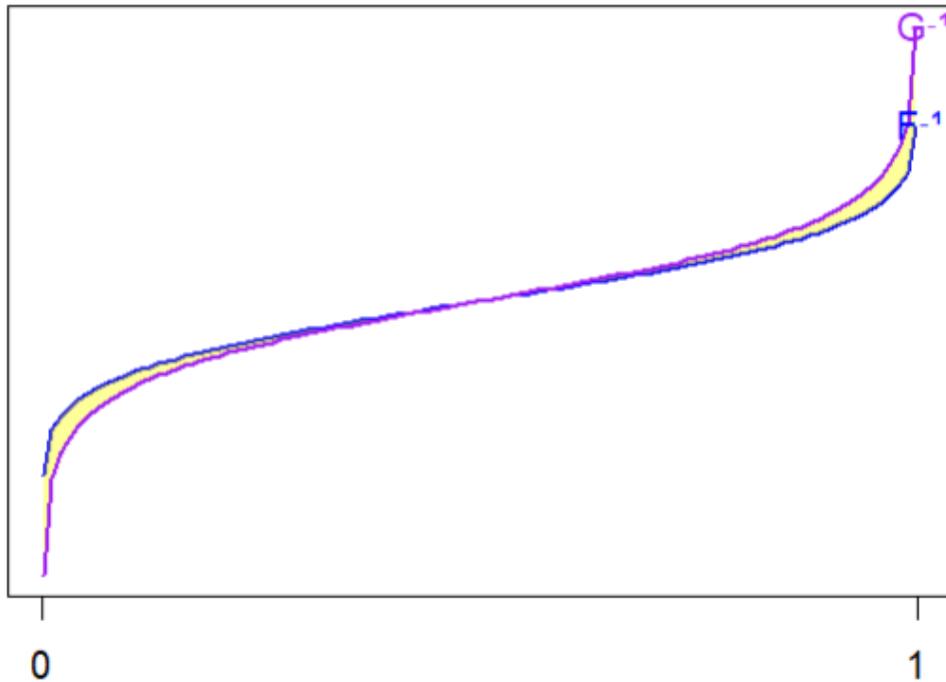
Vemos que tenemos representadas dos funciones cuantil F^{-1} y G^{-1} , donde la función F^{-1} está por encima hasta un punto (punto de corte) en el que cambia y empieza a estar por encima G^{-1} .

Lo que estamos representando en el numerador de ε_{ω_1} es la densidad que está entre las funciones F^{-1} y G^{-1} cuando F^{-1} está por encima de G^{-1} , es decir, el área amarilla del gráfico hasta el punto de corte de ambas funciones.

El denominador representa toda el área entre ambas funciones, que en el gráfico es toda el área coloreada en color amarillo.

Podemos utilizarlo también para entender la versión L_2 , sería lo mismo que L_1 pero elevando al cuadrado la diferencia del numerador y el valor absoluto del denominador.

El problema surge con la versión L_2 cuando el área entre las funciones es muy pequeña, debido a que F^{-1} y G^{-1} son muy parecidas en todo el dominio, como podemos ver en el gráfico a continuación.



Cuando estamos ante estos casos, el denominador de ε_{ω_2} se convierte en un número muy pequeño que genera inestabilidad numérica, causando sensibilidad extrema a pequeñas diferencias.

Al igual que en del Barrio et al., 2018, se va a representar un gráfico de contorno, pero en este caso en vez de realizar el índice con la distribución normal, vamos a verlo con la doble exponencial, al igual que hemos estado efectuando durante el documento.

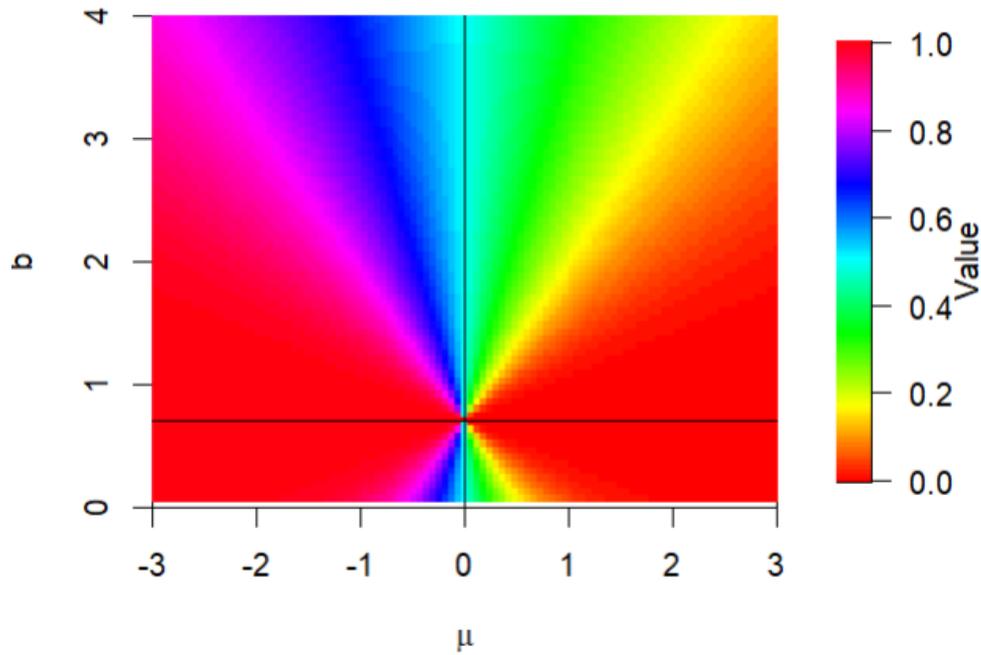


Figura 4: Gráfico de contorno de $\varepsilon_{\omega_2}(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$

Estamos analizando los valores de $\varepsilon_{\omega_2}(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2}))$, para $\mu \in \mathbb{R}$ y $\sigma > 0$. Se puede observar que el índice ε_{ω_2} permanece constante cuando $(\mu, \sigma/\sqrt{2})$ se mueve a lo largo de los rayos desde $(0, 1/\sqrt{2})$. Cuando $\mu > 0$ se da el caso en que $\varepsilon_{\omega_2}(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2})) < \frac{1}{2}$; mientras que, para $\mu < 0$, se cumple la situación contraria, donde $\varepsilon_{\omega_2}(L(0, 1/\sqrt{2}), L(\mu, \sigma/\sqrt{2})) > \frac{1}{2}$. Además, solo se cumple $\varepsilon_{\omega_2} = 0$ cuando $\sigma = 1$ y $\mu \geq 0$, reduciéndose la expresión a la forma $\varepsilon_{\omega_2}(L(0, 1/\sqrt{2}), L(\mu, 1/\sqrt{2}))$.

La principal diferencia que vemos respecto a la figura 1 de del Barrio et al., 2018 es que, en nuestro gráfico, la amplitud del área de color rojo es menor en comparación a la amplitud del área en color blanco del gráfico en del Barrio et al., 2018. Esta área representa el rango donde se cumple la dominancia entre un 95 % y un 100 %, lo cual permite afirmar, con una confianza muy alta, que se cumple la dominancia estocástica. Volviendo de nuevo a analizar $F \sim L(0, 1/\sqrt{2})$ y $G \sim L(0,8, 2)$, obtenemos que $\varepsilon_{\omega_2}(F, G) = 0,25$, coincidiendo con el resultado que obteníamos con la figura 1, donde un nuevo tratamiento nos da peores resultados para el 25 % de los pacientes.

Conclusiones

En este trabajo fin de grado hemos intentado recopilar la información presentada en Álvarez-Esteban et al. (2017), del Barrio et al. (2025) y del Barrio et al. (2018), con el fin de comprender el orden estocástico y algunas relajaciones.

Partiendo de la definición clásica del orden estocástico y el papel fundamental que desempeñan las funciones cuantil, se ha explicado la utilidad del concepto al igual que las limitaciones que tiene el orden estocástico en la práctica. Una de las partes más importantes del trabajo ha sido el estudio de la dominancia estocástica dentro de las familias de localización y escala. En particular, se ha prestado especial atención a la distribución doble exponencial como caso representativo. Se ha mostrado que, debido a su simetría y a su soporte no acotado, la dominancia estocástica solo se verifica en ciertos casos concretos.

Para adaptar el orden estocástico a contextos más realistas, donde los datos y las condiciones prácticas impiden obtener de forma exacta el orden estocástico, se ha profundizado en tres relajaciones cuantitativas: la dominancia basada en funciones cuantil, el modelo de contaminación y la formulación basada en el transporte óptimo. Cada uno de estos enfoques permite cuantificar, desde distintas perspectivas, el grado de cumplimiento del orden estocástico entre distribuciones.

Además, se han incluido ejemplos y gráficos que permiten una mejor interpretación de los conceptos y sus implicaciones. Estas herramientas, junto con las propiedades de las funciones cuantil, han permitido mostrar de forma clara la utilidad de los distintos índices y su comportamiento en situaciones concretas.

En conjunto, el trabajo proporciona una visión amplia y comprensible de cómo adaptar y medir la dominancia estocástica cuando no se cumple de forma estricta, y establece una buena base para su aplicación práctica en contextos reales.

Bibliografía

- Álvarez-Esteban, P. C., del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2016). A contamination model for the stochastic order. *TEST*, 25, 751-774.
- Álvarez-Esteban, P. C., del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2017). Models for the Assessment of Treatment Improvement: The Ideal and the Feasible. *Statistical Science*, 32(3), 469-485. <https://doi.org/10.1214/17-STS616>
- Berger, R. L. (1988). A nonparametric, intersection-union test for stochastic order [West Lafayette, Ind., 1986]. En *Statistical Decision Theory and Related Topics, IV* (pp. 253-264, Vol. 2). Springer.
- del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2025). *Invariant Measures of Disagreement with Stochastic Dominance* [Tentatively accepted].
- del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2018). An Optimal Transportation Approach for Assessing Almost Stochastic Order. En E. G. et al. (Ed.), *The Mathematics of the Uncertain* (pp. 33-44, Vol. 142). Springer International Publishing. https://doi.org/10.1007/978-3-319-73848-2_3
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Laplace (Double Exponential) Distributions. En *Continuous Univariate Distributions, Volume 2* (2nd, pp. 164-209). John Wiley & Sons.
- Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. John Wiley & Sons.
- Levy, H. (2016). *Stochastic Dominance: Investment Decision Making under Uncertainty* (3.^a ed.). Springer. https://doi.org/10.1007/978-3-319-21708-6_2
- Near, J. P., & Abuah, C. (2025). *Programming Differential Privacy*. <https://uvm-plaid.github.io/programming-dp/>
- Siegrist, K. (2025). *Probability, Mathematical Statistics, and Stochastic Processes*. LibreTexts. [https://espanol.libretexts.org/Estadisticas/Teoria_de_Probabilidad/Probabilidad%2C_estad%3%ADstica_matem%3%A1tica_y_procesos_estoc%3%A1sticos_\(Siegrist\)/05%3A_Distribuciones_especiales/5.28%3A_La_distribuci%3%B3n_de_Laplace](https://espanol.libretexts.org/Estadisticas/Teoria_de_Probabilidad/Probabilidad%2C_estad%3%ADstica_matem%3%A1tica_y_procesos_estoc%3%A1sticos_(Siegrist)/05%3A_Distribuciones_especiales/5.28%3A_La_distribuci%3%B3n_de_Laplace)
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>