Permanentes de matrices y aplicaciones en probabilidad y en estadística

Pablo Arnedo Magallón
Trabajo de fin de grado en Matemáticas
Universidad de Zaragoza

Directora del trabajo: Carmen Sangüesa Lafuente
12 de septiembre de 2019

Prólogo

En 1812, Cauchy desarrolló la teoría de los determinantes como un tipo de funciones alternantes simétricas (esto es, funciones definidas mediante la suma de términos positivos y negativos, que se van alternando), distinguiéndolas de las funciones simétricas ordinarias, a las que llamó *fonctions symétriques permanentes*. Asimismo, también introdujo un tipo especial de estas funciones, que más tarde fueron denominadas *permanentes* por Muir, nombre que se conserva hoy en día. Por otro lado, y de manera independiente, también Binet en el mismo año introdujo los permanentes, dando un método para calcular los permanentes de matrices $m \times n$, incluso si éstas no son cuadradas, cuando $m \leq 4$, que aparece en [9]. En cualquier caso, en este trabajo nos centraremos únicamente en permanentes de matrices cuadradas.

En el siglo que siguió a las publicaciones de Binet y Cauchy, se publicaron unos 20 textos sobre permanentes, la mayoría de los cuales trabajaba con identidades que incluían determinantes y permanentes. Los resultados más interesantes publicados en estos artículos son las identidades de Borchardt, Cayley y Muir, que dan diferentes fórmulas para el producto del determinante y el permanente de una matriz. Las tres identidades aparecen en [9], así como fórmulas más sencillas para algunos grupos particulares de matrices.

Todos estos resultados son más o menos sencillos, aunque algunas demostraciones son complicadas debido a la complejidad inherente a las funciones multilineales. Sin embargo, el verdadero punto de inflexión en la historia de los permanentes llegó a principios del siglo XX con el teorema de Muirhead, que fue posteriormente extendido por Hardy, Littlewood y Pólya. Tanto el teorema como su generalización aparecen en [9] y tienen aplicaciones en muchas áreas de las matemáticas. Poco después, llegaron el problema de Pólya, que buscaba la transformación más sencilla posible T tal que $Per\,T(A) = Det\,A$; la introducción en 1918 por Schur del concepto de *funciones de matriz generalizadas* (también conocidas como *funciones de Schur*); y la conjetura de van der Waerden. Todo esto aparece en [9].

El estudio de los permanentes tuvo una pausa entre 1926 y 1959. Si bien la mayoría de los resultados combinatorios obtenidos en este período se podían expresar en términos de permanentes, sus autores los probaron mediante métodos combinatorios y no hicieron mención alguna a los permanentes. Sin embargo, en 1959 se produjo un reavivamiento del interés en los permanentes cuando, de manera casi simultánea, Brenner demostró que los permanentes de matrices diagonal-dominantes no se cancelan, Caianiello los usó para expresar las expansiones de perturbación en teoría de campos cuánticos de una forma algebraica compacta, Levine extendió la identidad de Cayley a matrices 4×4 y Marcus y Newman publicaron su artículo sobre la conjetura de van der Waerden. Se considera que este artículo es el principal responsable de dicho reavivamiento, que ha llevado a la publicación de más de 200 artículos de más de 100 autores hasta la publicación de [9].

Pasando a un ámbito más estadístico, es obvio que siempre se ha buscado la comparación entre fenómenos aleatorios. En este sentido, la forma más simple de comparar distribuciones de probabilidad es comparando alguna medida de dichas distribuciones. Sin embargo, este camino tiene varios inconvenientes: es imposible que la medida elegida pueda abarcar toda la información sobre la distribución de referencia, puede formar parte de un conjunto más amplio de medidas de esa característica y puede no existir para esa distribución en concreto. Para solventar estos problemas, entonces, necesitamos una comparación más general, que son los órdenes estocásticos. Su estudio comenzó con el libro sobre desigualdades de Hardy, Littlewood y Pólya (1934, reeditado en 1952). [6] A partir de entonces, el número de órdenes estocásticos no ha dejado de crecer, teniendo además muchas y muy diversas aplicaciones:

estadística aplicada, biología, ingenería y fiabilidad de sistemas, economía (estudio de la desigualdad), etc. [11]

Además, estos órdenes estocásticos también sirven si trabajamos con dos muestras diferentes al mismo tiempo. De hecho, utilizaremos tanto los permanentes como los órdenes estocásticos en el último capítulo del trabajo, en el que trataremos de comparar muestras de estadísticos ordenados en orden de razón de verosimilitudes (el orden estocástico más fuerte de los que trataremos en el trabajo).

Resumen

The objective of this project is to study the main properties of permanents of matrices (a concept very similar to the determinant, as we will see at the beginning of chapter 1) and order statistics, as well as the main stochastic orders; and then to use all this to see how to deal with order statistics when we have two different samples. Before that (and also before this summary), we have a preface where we explain a bit of history of the study of this subject and what led to studying some of the points of the project.

For that, we will see all the different points in the following order:

1. In the first chapter, to start, we will see what is a permanent (the concept is the same as the one of a determinant but, in this case, all the terms of the sum are positive) and the main properties of the permanents with proofs (which are very similar to the ones of the determinants). Then, we will see how to differentiate the permanent of a functions matrix, result that we will use in chapter 4. To finish, we will see the Alexandroff inequality, which is the main result for permanents, using a previous result to prove it. This inequality states that if $C_1,...,C_{n-1},C$ are positive vectors (i.e., all their components are > 0) and $(C_1|...|C_n)$ denotes the matrix whose columns are $C_1,...,C_n$, then

$$[Per\,(C_1|...|C_{n-1}|C)]^2 \geq Per\,(C_1|...|C_{n-2}|C_{n-1}|C_{n-1})Per\,(C_1|...|C_{n-2}|C|C).$$

2. In the second chapter, we will start seeing the definition of order statistics (given a random sample, order statistics are the same variables, ordered increasingly) and then some general situations in life which justify the importance of order statistics, with special attention to k-out-of-n systems, which will also be mentioned in chapter 4. After that, we will see the distribution and density functions of order statistics, with some examples, both discrete and absolutely continuous, for random variables: first, that are independent and identically distributed; and then, independent but not identically distributed. We will also remind some concepts that are useful for these results, e.g., binomial and multinomial distributions, random sample.

3. In the third chapter, we will see the main stochastic orders, seen as another way of ordering random variables, based on their distribution and density functions. For each of them (usual, hazard rate, reversed hazard and likelihood ratio), we will see the definition and the main properties. We will say that two random variables are ordered in the usual stochastic order if the distribution function of the first variable is less or equal than the one of the second variable, and the rest of orders are stronger than the usual one. We will also remind what is weak convergence, as this is central in the main result of each type, and talk (less) about almost everywhere order, seen as a natural candidate for an order that compares the values that a pair of random variables can take. We will finish the chapter seeing what are log-concave and log-convex densities, some examples of both types (normal, gamma and Weibull distributions) and an important result about them that will be used in chapter 4.

4. In the fourth and last chapter, we will use some of the things seen in the previous chapters to compare order statistics from two different samples in likelihood ratio order. The objective of this chapter is to prove that, given $X_1,...,X_p$ with distrbution function F and $X_{p+1},...,X_n$ with

distribution function G, with $X_1,\dots,X_n$ an independent random sample and $n = p+q$, for any $k = 1,\dots,n$,

$$G \leq_{lr} F \Longrightarrow X_{(k)}(p,q) \leq_{lr} X_{(k)}(p+1,q-1),$$

where $\leq_{lr}$ denotes the likelihood ratio order among two distributions, the first number in the parenthesis denotes the number of variables with distribution F and the second number in the parenthesis denotes the number of variables with distribution G.

Índice general

Capítulo 1

Permanentes

El concepto de permanente fue introducido por primera vez en 1812 por Binet y Cauchy, casi a la vez que el concepto de determinante. La conjetura de van der Waerden, que trataba de encontrar el mínimo permanente en el conjunto de las matrices doblemente estocásticas, fue la que más atención de los matemáticos atrajo hacia la teoría de permanentes. Dicha conjetura fue demostrada independientemente por Egorychev y Falikman en torno a 1980, lo cual aumentó la actividad en el área de los permanentes. En este trabajo veremos algunas aplicaciones en estadística, con especial atención a los estadísticos ordenados, en particular a los sistemas k de n. [2]

1.1. Definición y primeras propiedades

Definición. Sea $A = ((a_{i,j}))$ una matriz cuadrada de dimensión $n \times n$. Definimos el **permanente** de A de la siguiente manera:

$$\mathrm{Per}\, A = \sum_P \prod_{j=1}^{n} a_{j,i_j},$$

donde $\sum_P$ denota la suma a lo largo de todas las posibles permutaciones $(i_1, i_2, ..., i_n)$ de $(1, 2, ..., n)$. Esta definición es similar a la del determinante de A, con la única diferencia de que no hay alternancia de signos (dependiendo de si la permutación es de orden par o impar). [2]

En el siguiente resultado, vamos a ver cómo los permanentes tienen propiedades similares a los determinantes. Además, las demostraciones se realizan de manera análoga a la de los determinantes, que podemos ver en el texto de González-Meneses. [5]

Proposición 1. *La definición anterior da lugar a las siguientes propiedades:*

1. *$\mathrm{Per}\, A$ no cambia si intercambiamos dos filas o dos columnas de A entre sí.*

2. *Podemos desarrollar el permanente de una matriz por cualquier fila o columna, es decir, si $A(i, j)$ denota la matriz resultante de eliminar en A la fila i y la columna j (de dimensión $n-1 \times n-1$, entonces tenemos que*

$$\mathrm{Per}\, A = \sum_{i=1}^{n} a_{i,j} \mathrm{Per}\, A(i, j) = \sum_{j=1}^{n} a_{i,j} \mathrm{Per}\, A(i, j),$$

donde la primera fórmula la podemos aplicar para cualquier $j = 1, ..., n$ y la segunda fórmula la podemos aplicar para cualquier $i = 1, ..., n$

3. *Si A^* denota la matriz resultante de multiplicar los elementos de la fila i de A por c, entonces*

$$\mathrm{Per}\, A^* = c\mathrm{Per}\, A$$

4. *Si descomponemos la fila i de A como suma de dos vectores $\mathbf{v}$ y $\mathbf{w}$ y denotamos A^* y A^{**} a las matrices resultantes de sustituir en A la fila i por $\mathbf{v}$ y $\mathbf{w}$ respectivamente, entonces*

$$Per\,A = Per\,A^* + Per\,A^{**}$$

Demostración. Vamos a ver la demostración de algunas de estas propiedades (en todos los casos, las vamos a hacer por inducción sobre n):

1. Para $n = 1$, la propiedad no tiene sentido. Para $n = 2$ la propiedad se puede obtener teniendo en cuenta que
$$Per\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad + bc,$$
que evidentemente no cambia al permutar filas o columnas. Ahora, supongamos que se cumple para $n-1$ y vamos a probarla para $n > 2$. En primer lugar, supongamos que las columnas que se intercambian son consecutivas (sean j y $j+1$ las columnas intercambiadas para $j < n$, y notemos que el proceso sería el mismo si intercambiáramos dos filas consecutivas). Sea A' la matriz resultante de intercambiar estas columnas. Entonces, $A(1,k)$ (con $k \neq j, j+1$) se transforma en $A'(1,k)$. Por lo tanto, por hipótesis de inducción, $Per\,A(1,k) = Per\,A'(1,k)$.

   Por otro lado, $A(1,j)$ resulta de eliminar la fila 1 y la columna j de A, pero esto es lo mismo que eliminar la fila 1 y la columna $j+1$ de A'. Así, $A(1,j) = A'(1,j+1)$. Análogamente, $A(1,j+1) = A'(1,j)$, luego $Per\,A(1,j) = Per\,A'(1,j+1)$ y $Per\,A'(1,j) = Per\,A(1,j+1)$. Como además $a_{1,j} = a'_{1,j+1}$ y $a'_{1,j} = a_{1,j+1}$, aplicando la segunda propiedad con $i = 1$, se tiene que

$$\begin{aligned} Per\,A &= \left(\sum_{k\neq j,j+1} a_{1,k} Per\,A(1,k)\right) + a_{1,j} Per\,A(1,j) + a_{1,j+1} Per\,A(1,j+1) \\ &= \left(\sum_{k\neq j,j+1} a'_{1,k} Per\,A'(1,k)\right) + a'_{1,j+1} Per\,A'(1,j+1) + a'_{1,j} Per\,A'(1,j) \\ &= Per\,A'. \end{aligned}$$

   Por último, para intercambiar columnas no consecutivas, notemos que basta con realizar un determinado número (finito) de intercambios de columnas consecutivas en los que el permanente de A no cambia.

2. Este método para calcular un permanente lo obtenemos sacando como factores comunes en la definición de permanente los términos de la fila o columna por la que desarrollamos. Notemos que no necesitamos las otras tres propiedades para ello (de hecho, utilizamos esta propiedad en las demostraciones de las otras propiedades). La prueba de este método la podemos encontrar (para determinantes) en el texto de González-Meneses. [5]

3. Para $n = 1$ es obvio. Supongamos que se cumple para $n-1$ y vamos a probarlo para $n > 1$. Si A^* es la matriz resultante de multiplicar los elementos de la fila i de A por c, entonces $a^*_{i,1} = ca_{i,1}$ y además $A(i,1) = A^*(i,1)$. Ahora, si $k \neq i$, entonces $a^*_{k,1} = a_{k,1}$ y $A^*(k,1)$ se obtiene multiplicando una de las filas de $A(k,1)$ por c. Por hipótesis de inducción, $Per\,A^*(k,1) = c Per\,A(k,1)$. Por lo tanto, desarrollando por la primera columna:
$$Per\,A^* = a^*_{i,1} Per\,A^*(i,1) + \sum_{k\neq i} a^*_{k,1} Per\,A^*(k,1) = ca_{i,1} Per\,A(i,1) + \sum_{k\neq i} ca_{k,1} Per\,A(k,1) = c Per\,A$$

4. Sean $\mathbf{v} = (v_1,...,v_n)$ y $\mathbf{w} = (w_1,...,w_n)$. Para $n = 1$ es evidente. Supongamos que es cierto para $n-1$ y vamos a probarlo para $n > 1$. En primer lugar, notemos que al descomponer la fila i tenemos que $a_{i,1} = v_1 + w_1 = a^*_{i,1} + a^{**}_{i,1}$. Ahora, para $k \neq i$, $a_{k,1} = a^*_{k,1} = a^{**}_{k,1}$ y además $A^*(k,1)$

y $A^{**}(k,1)$ se obtienen descomponiendo en dos sumandos una fila de $A(k,1)$. Por hipótesis de inducción, $Per\,A(k,1) = Per\,A^*(k,1) + Per\,A^{**}(k,1)$. Por lo tanto, desarrollando por la primera columna:

$$\begin{aligned}
Per\,A &= a_{i,1} Per\,A(i,1) + \sum_{k\neq i} a_{k,1} Per\,A(k,1) \\
&= (a^*_{i,1} + a^{**}_{i,1}) Per\,A(i,1) + \sum_{k\neq i} a_{k,1}(Per\,A^*(k,1) + Per\,A^{**}(k,1)) \\
&= \left(a^*_{i,1} Per\,A^*(i,1) + \sum_{k\neq i} a^*_{k,1} Per\,A^*(k,1)\right) + \left(a^{**}_{i,1} Per\,A^{**}(i,1) + \sum_{k\neq i} a^{**}_{k,1} Per\,A^{**}(k,1)\right) \\
&= Per\,A^* + Per\,A^{**}
\end{aligned}$$

□

El siguiente resultado nos indica cómo debemos derivar el permanente de una matriz de funciones y lo utilizaremos en el capítulo 4.

Teorema 1. *Sea $A(x) = (a_{i,j}(x))$, $i = 1,\ldots,n$, $j = 1,\ldots,n$ una matriz de funciones $a_{i,j} : I \longrightarrow \mathbb{R}$, donde I es un intervalo de $\mathbb{R}$. Supongamos que todas estas funciones son derivables. Entonces*

$$(Per\,A)'(x) = \sum_{i=1}^{n} Per\,A_i(x),$$

donde $A_i(x)$ se obtiene sustituyendo la fila i de $A(x)$ por las derivadas de cada función ($a_{i,j}(x)$ se sustituye por $a'_{i,j}(x)$ para cada $j = 1,\ldots,n$). [2]

Demostración. Vamos a probarlo por inducción sobre n. Para $n = 0$ es trivial. Para $n > 0$, desarrollando por la primera fila, tenemos que:

$$Per\,A(x) = \sum_{j=1}^{n} a_{1,j}(x) Per\,A(1,j)(x).$$

Derivando esta expresión, tenemos que:

$$(Per\,A)'(x) = \sum_{j=1}^{n} a'_{1,j}(x) Per\,A(1,j)(x) + \sum_{j=1}^{n} a_{1,j}(x)(Per\,A(1,j))'(x).$$

Notemos que el primer sumando no es más que $Per\,A_1(x)$. Para el segundo, podemos aplicar la hipótesis de inducción, de manera que:

$$(Per\,A(1,j))'(x) = \sum_{k=1}^{n-1} Per\,A_k(1,j)(x).$$

Si además tenemos en cuenta que

$$\sum_{j=1}^{n} a_{1,j}(x) Per\,A_k(1,j)(x) = Per\,A_{k+1}(x),$$

entonces:

$$\sum_{j=1}^{n} a_{1,j}(x)(Per\,A(1,j))'(x) = \sum_{j=1}^{n}\sum_{k=1}^{n-1} a_{1,j}(x) Per\,A_k(1,j)(x) = \sum_{k=1}^{n-1} Per\,A_{k+1}(x),$$

de manera que el segundo sumando de la expresión anterior es igual a $Per\,A_2(x) + \ldots + Per\,A_n(x)$, lo que demuestra el teorema. □

1.2. Desigualdad de Alexandroff

La desigualdad de Alexandroff fue inicialmente enunciada y demostrada usando una función llamada *discriminante mixto* y sin hacer mención alguna a los permanentes. Sin embargo, la versión que vamos a ver es suficiente para el objetivo del trabajo. Probaremos primero el siguiente resultado:

Lema 1. *Sean $C_1,\ldots,C_{n-1}$ vectores positivos en $\mathbb{R}^n$, $n\geq 2$ (esto es, vectores de $\mathbb{R}^n$ tales que todas sus componentes son >0). Sea C_n otro vector positivo en $\mathbb{R}^n$ y supongamos que C_{n-1} y C_n son linealmente independientes. Si $(C_1|\ldots|C_n)$ denota la matriz cuyas columnas son $C_1,\ldots,C_n$, entonces la ecuación*

$$Per\,(C_1|\ldots|C_{n-2}|C_n+yC_{n-1}|C_n+yC_{n-1})=0$$

tiene dos raíces reales distintas, $\alpha_1<\alpha_2$. Además, para cualquier vector positivo $Z\in\mathbb{R}^n$, la raíz de la ecuación lineal

$$Per\,(C_1|\ldots|C_{n-2}|C_n+yC_{n-1}|Z)=0$$

está en el intervalo (α_1,α_2). [3]

Demostración. Vamos a proceder por inducción sobre n. En todos los casos, vamos a denotar a_{ji} a la i-ésima componente del vector C_j. Para $n=2$, sea $A=(C_1|C_2)$. Entonces se cumple que

$$Per\,(C_2+yC_1|C_2+yC_1)=Per\begin{bmatrix} a_{21}+ya_{11} & a_{21}+ya_{11}\\ a_{22}+ya_{12} & a_{22}+ya_{12}\end{bmatrix}=0$$

si y solo si $2(a_{21}+ya_{11})(a_{22}+ya_{12})=0$, es decir, si y solo si $y=-\frac{a_{21}}{a_{11}}$ o bien $y=-\frac{a_{22}}{a_{12}}$, luego éstas son las dos raíces de la ecuación, que son reales y distintas ya que C_1 y C_2 son linealmente independientes.

Ahora, para cualquier vector positivo (z_1,z_2), la función lineal

$$\phi(y)=Per\begin{bmatrix} a_{21}+ya_{11} & z_1\\ a_{22}+ya_{12} & z_2\end{bmatrix}=z_2(a_{21}+ya_{11})+z_1(a_{22}+ya_{12})$$

toma los valores $z_1(a_{22}-\frac{a_{21}}{a_{11}}a_{12})$ en $y=-\frac{a_{21}}{a_{11}}$ y $z_2(a_{21}-\frac{a_{22}}{a_{12}}a_{11})$ en $y=-\frac{a_{22}}{a_{12}}$. Es claro que uno de los valores anteriores es positivo y el otro negativo. Por lo tanto, aplicando el teorema de Bolzano, la raíz de $\phi(y)=0$ debe estar entre $-\frac{a_{21}}{a_{11}}$ y $-\frac{a_{22}}{a_{12}}$.

Sea $n\geq 3$ y supongamos que el lema es cierto para $n-1$. Sea

$$C_y=(C_1|\ldots|C_{n-2}|C_n+yC_{n-1}|C_n+yC_{n-1}).$$

Entonces, desarrollando por la primera fila,

$$Per\,C_y=\left(\sum_{j=1}^{n-2}a_{j1}Per\,C_y(1,j)\right)+2(a_{n1}+ya_{n-1,1})Per\,C_y(1,n). \tag{1.1}$$

Sea y_0 la raíz de $Per\,C_y(1,n)=0$. Por hipótesis de inducción, y_0 se encuentra entre las dos raíces de $Per\,C_y(1,j)=0$, $j=1,2,\ldots,n-2$. Para aplicar esta hipótesis de inducción, tomamos como Z la columna j de $C_y(1,n)$ y hacemos una trasposición de columnas para colocar la columna con el vector Z al final de la matriz.

Dado que $Per\,C_y(1,j)\to\infty$ cuando $y\to\pm\infty$, $j=1,2,\ldots,n-2$ (porque $Per\,C_y$ es una forma cuadrática en y con coeficientes positivos), concluimos que

$$Per\,C_{y_0}(1,j)<0,\ \ j=1,2,\ldots,n-2. \tag{1.2}$$

Por lo tanto, por (1.1), $Per\,C_{y_0}<0$. Entonces $Per\,C_y=0$ debe tener dos raíces reales distintas, $\alpha_1<\alpha_2$. De esta manera queda probada la primera parte. Para la segunda parte, 1.1 y 1.2 nos aseguran que $Per\,C_{y_0}<0$ y por lo tanto, $y_0\in(\alpha_1,\alpha_2)$. Análogamente podemos probar que la raíz de $Per\,C_y(i,n)=0$

está en (α_1, α_2), $i = 1, 2, ..., n$. En particular, dado que $C_y(1,n)$ es una forma lineal en y con coeficientes positivos y por tanto es creciente en y, se sigue que

$$Per\, C_{\alpha_1}(i,n) < 0,\ Per\, C_{\alpha_2}(i,n) > 0,\ i = 1, 2, ..., n. \tag{1.3}$$

Ahora, sea Z un vector positivo en $\mathbb{R}^n$. Entonces, desarrollando por la última columna:

$$Per\,(C_1|...|C_{n-2}|C_n + yC_{n-1}|Z) = \sum_{i=1}^{n} z_i Per\, C_y(i,n). \tag{1.4}$$

A partir de (1.3) y (1.4) se sigue que

$$Per\,(C_1|...|C_{n-2}|C_n + \alpha_1 C_{n-1}|Z) < 0,\ Per\,(C_1|...|C_{n-2}|C_n + \alpha_2 C_{n-1}|Z) > 0,$$

y por lo tanto, aplicando el teorema de Bolzano, la raíz de

$$Per\,(C_1|...|C_{n-2}|C_n + yC_{n-1}|Z) = 0$$

está en (α_1, α_2). □

Teorema 2 (Desigualdad de Alexandroff)**.** *Sean $C_1, ..., C_{n-1}$ vectores positivos en $\mathbb{R}^n$, $n \geq 2$. Entonces, para todo C vector positivo en $\mathbb{R}^n$, se cumple:*

$$[Per\,(C_1|...|C_{n-1}|C)]^2 \geq Per\,(C_1|...|C_{n-2}|C_{n-1}|C_{n-1}) Per\,(C_1|...|C_{n-2}|C|C).$$

Además, la igualdad se da si y solo si C_{n-1} y C son linealmente dependientes. [3]

Demostración. Si C_{n-1} y C son linealmente dependientes, entonces podemos poner $C_{n-1} = \lambda C$, siendo λ un número real distinto de 0. Sustituyendo en ambos lados de la igualdad en la expresión del enunciado y aplicando la tercera propiedad de la proposición 1, tenemos la igualdad. Supongamos entonces que son linealmente independientes. Aplicando las dos últimas propiedades a las dos últimas columnas de

$$Per\,(C_1|...|C_{n-2}|C + yC_{n-1}|C + yC_{n-1}), \tag{1.5}$$

obtenemos la siguiente forma cuadrática en y:

$$y^2 Per\,(C_1|...|C_{n-2}|C_{n-1}|C_{n-1}) + 2y Per\,(C_1|...|C_{n-1}|C) + Per\,(C_1|...|C_{n-2}|C|C),$$

cuyo discriminante es

$$4[Per\,(C_1|...|C_{n-1}|C)]^2 - 4 Per\,(C_1|...|C_{n-2}|C_{n-1}|C_{n-1}) Per\,(C_1|...|C_{n-2}|C|C).$$

Por el lema 1, la forma cuadrática en (1.5) tiene dos raíces reales distintas y por lo tanto su discriminante debe ser positivo. Esto completa la demostración. □

Capítulo 2

Estadísticos ordenados

2.1. Introducción

Dado $(\Omega, \mathscr{F}, P)$ un espacio de probabilidad, supongamos que $(X_1, ..., X_n) : \Omega \longrightarrow \mathbb{R}^n$ es un vector aleatorio n-dimensional. Entonces los estadísticos ordenados son cada una de las X_k anteriores ordenadas de menor a mayor. Las denotamos de la siguiente manera: $X_{(1)}$ la más pequeña, $X_{(2)}$ la segunda más pequeña, y así sucesivamente hasta $X_{(n)}$, que es la más grande, esto es, para cada $k = 1, 2, ..., n$:

$$X_{(k)} = \{k\text{-ésimo valor más pequeño de } X_1, ..., X_n\}.$$

Por tanto:

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}.$$

Notemos que en esta definición no necesitamos que las X_k sean independientes ni idénticamente distribuidas. Trabajaremos primero con variables independientes e idénticamente distribuidas (i.i.d.) y extenderemos los resultados a variables que no sean idénticamente distribuidas (seguirán siendo independientes) utilizando lo visto en el capítulo anterior.

Veamos ahora algunas situaciones donde se usan los estadísticos ordenados: [1]

1. **Muestreo con censura**: 50 máquinas caras son iniciadas en un experimento para estimar la esperanza de vida de una máquina de esas características. Si, como cabría esperar, son lo suficientemente fiables como para ser puestas a la venta, llevaría mucho tiempo esperar al primer fallo de cada máquina. En lugar de ello, podemos ahorrar tiempo y costes de fabricación si consideramos únicamente los primeros fallos (que serían los primeros tiempos en la muestra de estadísticos ordenados). Asimismo, podemos eliminar del experimento las máquinas que no han fallado para su venta y/o modificación. Aquellas que hayan fallado se destruyen.

2. **Análisis de fiabilidad**: Consideremos un cable con n filamentos. Puede ocurrir que la rotura del primer filamento cause la rotura del cable, pero es más probable que el cable funcione siempre que k de los filamentos no se rompan (con $1 \leq k \leq n$). Un sistema así se llama k de n y funciona siempre que al menos k de la n componentes funcionen. Con respecto a un fallo de neumáticos, un coche es un sistema 4 de 5 (recordemos que es obligatorio llevar un neumático de repuesto). En términos de sistemas eléctricos, un sistema n de n se llama sistema en serie. El fallo de cualquier componente es catastrófico. En el otro extremo, un sistema 1 de n se llama sistema en paralelo (funciona siempre que cualquiera de los componentes funcione). En general, si el vector $(X_1, ..., X_n)$ representa el tiempo de vida de cada componente, el tiempo de vida de un sistema k de n es $X_{(n-k+1)}$, es decir, el tiempo hasta que menos de k componentes funcionan.

3. **Récords olímpicos**: La marca de Bob Beamon en la prueba de salto de longitud correspondiente a los Juegos Olímpicos de 1968 sigue siendo el récord vigente en dicha prueba. Sin embargo, en muy raras ocasiones un récord dura tanto tiempo. Si modelamos cada una de las mejores

marcas de cada edición como variables aleatorias i.i.d., entonces los récords serían cada vez más difíciles de batir. Sin embargo, esta modelización no es adecuada, porque las poblaciones se van incrementando en tamaño y mejorando sus medios a medida que pasa el tiempo, lo que hace que estas variables aleatorias no estén idénticamente distribuidas. En cualquier caso, nos estamos centrado en un máximo (o mínimo si miramos las pruebas de tiempo), es decir, en un estadístico ordenado.

4. **Distribución de premios en metálico en torneos**: En una prueba de atletismo, por ejemplo, los atletas se ordenan según el tiempo que necesitan para recorrer una determinada distancia (cuanto menor sea el tiempo, se llevan un premio mayor). Se puede hacer algo análogo con la distancia alcanzada en un salto o en el lanzamiento de un objeto o bien con la altura alcanzada en un salto (en este caso, cuanto mayor sea la distancia o altura alcanzada, mayor será el premio). Evidentemente, todo esto se hace a partir de estadísticos ordenados. Lo más previsible es que el participante con mejor nivel de habilidad resulte campeón del torneo, aunque determinados eventos aleatorios pueden hacer que sea otro participante quien gane el torneo.

2.2. Resultados para v.a.i.i.d.

En esta sección, vamos a enunciar y demostrar los principales resultados sobre estadísticos ordenados, esto es, las funciones de distribución y densidad del estadístico ordenado $X_{(k)}$ para cualquier k. Asimismo, vamos a ver algunos ejemplos concretos del uso de estos resultados, incluidos los casos particulares del mínimo y del máximo. Pero antes, vamos a ver algunas definiciones que nos resultarán útiles para el primer resultado, tanto en el enunciado como en la demostración, así como en el resto del trabajo.

Definición. Decimos que una variable aleatoria X tiene distribución binomial con parámetros (n, p) (con $n \in \mathbb{N}$, $p \in [0,1]$) si su función de masa de probabilidad es:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Recordemos que la distribución binomial cuenta el número de éxitos en n ensayos tales que en cada uno de ellos la probabilidad de éxito es p.

Definición. Una muestra aleatoria simple (abreviadamente m.a.s.) de tamaño n es un vector de v.a. $(X_1, \ldots, X_n)$ i.i.d. En estadística representa una selección con reemplazamiento de los elementos de una población, donde todos ellos tienen la misma probabilidad de ser incluidos en la muestra.

Definición. Sea $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria. La función de distribución de X es una función $F : \mathbb{R} \longrightarrow [0,1]$ que a cada $x \in \mathbb{R}$ le asocia $F(x) = P(X \leq x)$.

Teorema 3. *Sea $X_1, \ldots, X_n$ una m.a.s. de una población que tiene función de distribución $F(x)$ y sea $X_{(1)}, \ldots, X_{(n)}$ la muestra ordenada. Entonces la función de distribución $F_{(i)}(x)$ del estadístico ordenado de orden i (para cada $i = 1, \ldots, n$) se puede expresar como: [1]*

$$F_{(i)}(x) = \sum_{j=i}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} = \int_0^{F(x)} \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt, \forall x \in \mathbb{R} \qquad (2.1)$$

Demostración. Como el número de elementos de la muestra que son $\leq x$ es una binomial (para cualquier

$x \in \mathbb{R}$), tenemos que:

$$\begin{aligned} F_{(i)}(x) &= P(X_{(i)} \leq x) \\ &= P(\textit{Al menos } i \textit{ de los } X_1, \ldots, X_n \textit{ son } \leq x) \\ &= \sum_{j=i}^{n} P(\textit{Exactamente } j \textit{ de los } X_1, \ldots, X_n \textit{ son } \leq x) \\ &= \sum_{j=i}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} \end{aligned}$$

Entonces, la función de distribución de $X_{(i)}$ es simplemente la probabilidad de cola (empezando en i) de una variable aleatoria binomial con parámetros n y $F(x)$. Además si usamos que

$$\sum_{j=i}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \int_0^p \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt, \forall p \in (0,1),$$

que se puede probar integrando por partes, entonces podemos escribir la función de distribución $F_{(i)}(x)$ como en el enunciado. □

Nota. Notemos que como estamos utilizando variables aleatorias genéricas, este resultado es válido tanto para variables aleatorias discretas como para variables aleatorias absolutamente continuas.

Nota. En particular, la distribución del mínimo es

$$F_{(1)}(x) = \sum_{j=1}^{n} \binom{n}{j} F(x)^j (1-F(x))^{n-j} = 1-(1-F(x))^n$$

y la distribución del máximo es

$$F_{(n)}(x) = (F(x))^n.$$

[1]

Ejemplo. Sea $U \equiv U(0,1)$ (esto es, U tiene una densidad uniforme en el intervalo $(0,1)$). Sabemos, para cada $u \in (0,1)$, que la densidad de U es $f(u) = 1$ y su función de distribución es $F(u) = u$. Denotemos por $U_{(i)}$ al estadístico ordenado de orden i. Entonces, por 2.1,

$$F_{(i)}(u) = \int_0^u \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt,$$

y derivando,

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i}, \ \forall x \in (0,1),$$

que es una densidad beta.

Ejemplo. Ahora vamos a ver cómo expresar la función de probabilidad de los estadísticos ordenados para variables aleatorias discretas. Para ello, sea $X_1, \ldots, X_n$ una m.a.s. de una variable aleatoria que toma valores en los enteros no negativos (Poisson, Binomial,...). Sea $F(x)$ su función de distribución. Aplicando 2.1, tenemos que la función de probabilidad de $X_{(i)}$, esto es,

$$f_{(i)}(x) := P(X_{(i)} = x), \ x = 0,1,2,\ldots$$

se escribe como:

$$\begin{aligned} f_{(i)}(x) &= F_{(i)}(x) - F_{(i)}(x-1) \\ &= \int_0^{F(x)} \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt - \int_0^{F(x-1)} \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt \\ &= \int_{F(x-1)}^{F(x)} \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i} dt \end{aligned}$$

Cuando la muestra es de v.a. absolutamente continuas, la función de densidad de $X_{(i)}$ se puede obtener derivando la segunda expresión de 2.1. Vamos a ver una forma alternativa que, a partir de su demostración, nos permitirá extender el resultado a variables aleatorias independientes pero no idénticamente distribuidas. Para ello, utilizaremos probabilidades de una variable aleatoria multinomial, por lo que vamos a recordar lo que es antes de ver esta forma alternativa.

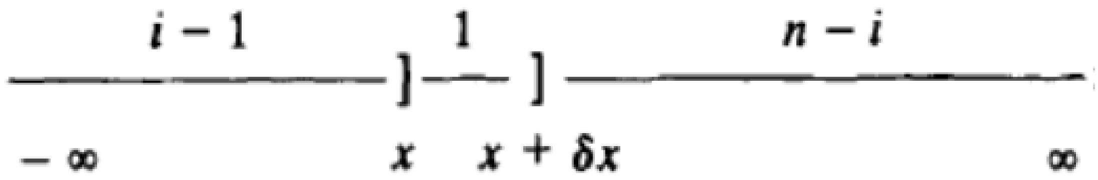
Definición. Decimos que un vector aleatorio $(M_1,...,M_k)$ tiene distribución multinomial con parámetros $(n, p_1,..., p_k)$ (con $n \in \mathbb{N}$, $p_1,..., p_k \in [0,1]$ de manera que $\sum_{i=1}^{k} p_i = 1$) si su función de masa de probabilidad es:

$$P(M_1 = x_1,...,M_k = x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} p_i^{x_i}, \tag{2.2}$$

donde $\sum_{i=1}^{k} x_i = n$. Notemos que esta distribución es una generalización de la distribución binomial, para el caso en el que hay más de dos resultados posibles para cada ensayo, de manera que cada M_i cuenta el número de resultados de cada tipo.

Teorema 4. *Sea $X_1,...,X_n$ una muestra aleatoria de una población absolutamente continua con densidad $f(x)$ y función de distribución $F(x)$ y sea $X_{(1)},...,X_{(n)}$ la muestra ordenada. Entonces la función de densidad $f_{(i)}(x)$ del estadístico ordenado de orden i (para cada $i = 1,...,n$) se puede expresar como: [1]*

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1} (1-F(x))^{n-i} f(x), \forall x \in \mathbb{R} \tag{2.3}$$



Demostración. Sea $\delta > 0$. Entonces el suceso $x < X_{(i)} \leq x+\delta x$ se puede expresar como: $X_j \leq x$ para $i-1$ de los X_j, $x < X_j \leq x+\delta x$ para un único X_j y $X_j > x+\delta x$ para los restantes $n-i$ X_j, tal como se puede ver en la imagen anterior. Además, también puede ocurrir que 2 o más X_i sean observados en $(x, x+\delta x]$.

Como el número de observaciones en cada uno de los intervalos anteriores es la realización de un vector aleatorio (M_1, M_2, M_3) trinomial con n ensayos y probabilidades $F(x)$, $F(x+\delta x) - F(x)$, $1-F(x+\delta x)$; podemos escribir:

$$P(M_1 = i-1, M_2 = 1, M_3 = n-i) = \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1} (1-F(x+\delta x))^{n-i} (F(x+\delta x) - F(x)), \tag{2.4}$$

donde M_i es el número de ensayos que se observan en cada uno de los intervalos (en orden creciente). Por lo tanto:

$$P(x < X_{(i)} \leq x+\delta x) = P(M_1 = i-1, M_2 = 1, M_3 = n-i) + P(x < X_{(i)} \leq x+\delta x, M_2 \geq 2). \tag{2.5}$$

Con respecto al primer término, sabiendo que

$$f(x) = \lim_{\delta \to 0} \left(\frac{F(x+\delta x) - F(x)}{\delta x} \right), \tag{2.6}$$

por ser f la derivada de F, se tiene que

$$\begin{aligned} f_{(i)}(x) &= \lim_{\delta \to 0} \left(\frac{P(x < X_{(i)} \leq x+\delta x)}{\delta x} \right) \\ &= \lim_{\delta \to 0} \left(\frac{P(M_1 = i-1, M_2 = 1, M_3 = n-i)}{\delta x} \right) + \lim_{\delta \to 0} \left(\frac{P(x < X_{(i)} \leq x+\delta x, M_2 \geq 2)}{\delta x} \right). \end{aligned}$$

Tenemos fácilmente por 2.4 y 2.6 que

$$\lim_{\delta\to 0} \frac{P(M_1 = i-1, M_2 = 1, M_3 = n-i)}{\delta x} = \frac{n!}{(i-1)!(n-i)!} F(x)^{i-1}(1-F(x))^{n-i} f(x).$$

Luego falta probar que el segundo término de la expresión anterior tiende a 0 cuando $\delta \to 0$.

Para ello, basta tener en cuenta que

$$P(x < X_{(i)} \leq x + \delta x, M_2 \geq 2) \leq P(M_2 \geq 2) = \sum_{k=2}^{\infty} \binom{n}{k} (F(x+\delta x) - F(x))^k$$

y que

$$\lim_{\delta\to 0} \frac{(F(x+\delta x) - F(x))^k}{\delta x} = \lim_{\delta\to 0} \frac{F(x+\delta x) - F(x)}{\delta x} (F(x+\delta x) - F(x))^{k-1} = 0$$

para todo $k \geq 2$. □

Nota. En particular, la densidad del mínimo es

$$f_{(1)}(x) = n(1-F(x))^{n-1} f(x)$$

y la densidad del máximo es

$$f_{(n)}(x) = n(F(x))^{n-1} f(x).$$

[1]

Ejemplo. Vamos a aplicar estas últimas fórmulas al caso en el que $X_1, \ldots, X_n$ es el resultado de una m.a.s. de una población $Exp(\theta)$. Sabemos que $F(x) = 1 - e^{-\theta x}$ si $x \geq 0$ y 0 en otro caso y que $f(x) = \theta e^{-\theta x}$ si $x \geq 0$ y 0 en otro caso. Por lo tanto,

$$f_{(1)}(x) = ne^{-(n-1)\theta x} \theta e^{-\theta x} = n\theta e^{-(n-1)\theta x} e^{-\theta x} = n\theta e^{-n\theta x},\ \forall x \geq 0,$$

por lo que $f_{(1)}(x) \equiv Exp(n\theta)$, y

$$f_{(n)}(x) = n(1-e^{-\theta x})^{n-1} \theta e^{-\theta x} = n\theta (1-e^{-\theta x})^{n-1} e^{-\theta x},\ \forall x \geq 0.$$

Nota. La expresión del teorema también se puede obtener directamente a partir de la función de densidad conjunta de los n estadísticos ordenados. Para verlo, en primer lugar tenemos que notar que, si $x_{(1)} \leq \ldots \leq x_{(n)}$ son las observaciones de los n estadísticos ordenados, entonces las variables X_i originales únicamente pueden tomar alguno de los $x_{(i)}$ anteriores (y además todos ellos deben ser tomados por alguna de las X_i), lo cual asigna (por simetría) la misma densidad a cada una de las permutaciones de $(1, \ldots, n)$. A partir de aquí, la función de densidad conjunta de los n estadísticos ordenados es:

$$f_{(1),(2),\ldots,(n)}(x_1, x_2, \ldots, x_n) = n! \prod_{j=1}^{n} f(x_j),$$

para $-\infty < x_1 \leq x_2 \leq \ldots \leq x_n < \infty$.

A partir de aquí, si integramos las variables $(X_{(1)}, \ldots, X_{(i-1)})$ y $(X_{(i+1)}, \ldots, X_{(n)})$, obtenemos que la densidad de $X_{(i)}$ es:

$$\begin{aligned} f_{(i)}(x) = n! f(x) &\left[\int_{-\infty}^{x} \cdots \int_{-\infty}^{x_2} f(x_1) \cdots f(x_{i-1}) dx_1 \cdots dx_{i-1} \right] \\ &\times \left[\int_{x}^{\infty} \cdots \int_{x}^{x_{i+2}} f(x_{i+1}) \cdots f(x_n) dx_{i+1} \cdots dx_n \right] \end{aligned} \tag{2.7}$$

Ahora, mediante una integración directa, podemos obtener que:

$$\int_{-\infty}^{x} \cdots \int_{-\infty}^{x_2} f(x_1)\cdots f(x_{i-1})dx_1\cdots dx_{i-1} = \frac{F(x)^{i-1}}{(i-1)!} \tag{2.8}$$

y

$$\int_{x}^{\infty} \cdots \int_{x}^{x_{i+2}} f(x_{i+1})\cdots f(x_n)dx_{i+1}\cdots dx_n = \frac{(1-F(x))^{n-i}}{(n-i)!}. \tag{2.9}$$

Si sustituimos los dos conjuntos de integrales de 2.7 por los valores obtenidos en 2.8 y 2.9, obtenemos la misma expresión para $f_{(i)}(x)$ que la obtenida en el teorema 4. [1]

2.3. Resultados para v.a. cualesquiera

La distribución de los estadísticos ordenados que hemos visto, aunque sea aparentemente simple, se complica si eliminamos la hipótesis de idéntica distribución. En este sentido, un modelo conocido es el de un dato atípico, donde $X_1,...,X_n$ son variables aleatorias independientes, $X_1,...,X_{n-1}$ pertenecen a una población con distribución $F(x)$ y densidad $f(x)$ y X_n pertenece a otra población con distribución $G(x)$ y densidad $g(x)$. Como antes, denotamos $X_{(1)} \leq ... \leq X_{(n)}$ a los estadísticos ordenados resultantes de estas variables aleatorias.

Ahora, tomamos $X_1,...,X_n$ de esta manera, consideramos (M_1,M_2,M_3) la multinomial que nos cuenta cuántas de las $n-1$ primeras variables están en $(-\infty,x]$, $(x,x+\delta x]$ y $(x+\delta x,\infty)$, aplicamos la ley de probabilidad total para X_n, y tenemos: [2]

$$\begin{aligned} P(x < X_{(i)} \leq x+\delta x) &= P(M_1 = i-2, M_2 = 1,\ X_n \in (-\infty,x]) \\ &\quad + P(M_1 = i-1, M_2 = 0,\ X_n \in (x,x+\delta x]) \\ &\quad + P(M_1 = i-1, M_2 = 1,\ X_n \in (x+\delta x,\infty)) \\ &\quad + P(x < X_{(i)} \leq x+\delta x,\ 2 \text{ o más observaciones en } (x,x+\delta x]). \end{aligned}$$

Aplicando 2.2 y siguiendo un razonamiento análogo al del teorema 4, llegamos a la siguiente expresión:

$$\begin{aligned} f_{(i)}(x) &= \frac{(n-1)!}{(i-2)!(n-i)!}(F(x))^{i-2}G(x)f(x)(1-F(x))^{n-i} \\ &\quad + \frac{(n-1)!}{(i-1)!(n-i)!}(F(x))^{i-1}g(x)(1-F(x))^{n-i} \\ &\quad + \frac{(n-1)!}{(i-1)!(n-i-1)!}(F(x))^{i-1}f(x)(1-F(x))^{n-i-1}(1-G(x)) \end{aligned}$$

Antes de pasar al caso más general, vamos a introducir la siguiente notación que nos resultará útil para abordarlo:

$$\begin{array}{r} i_1\{ \\ i_2\{ \\ \end{array} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & \cdots & . \end{pmatrix}$$

denota la matriz en la que la fila $(a_{11},a_{12},...,a_{1n})$ está repetida i_1 veces, la fila $(a_{21},a_{22},...,a_{2n})$ está repetida i_2 veces, y así sucesivamente (entendiendo que si en una fila no hay índice lateral esa fila está repetida una vez). De esta forma, tenemos los siguientes resultados, que generalizan los teoremas 3 y 4, respectivamente, a variables aleatorias independientes pero no necesariamente idénticamente distribuidas.

Teorema 5. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes tales que, para cada $j = 1, \ldots, n$, X_j tiene función de distribución F_j. Sean $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ los estadísticos ordenados obtenidos a partir de estas variables aleatorias. Entonces la función de distribución del estadístico ordenado $X_{(j)}$ (para cada $j = 1, \ldots, n$) se puede escribir como: [2]*

$$F_{(j)}(x) = \sum_{i=j}^{n} \frac{1}{i!(n-i)!} \mathrm{Per}\, B_1,$$

donde

$$B_1 = \begin{array}{r} i\{ \\ n-i\{ \end{array} \begin{pmatrix} F_1(x) & F_2(x) & \cdots & F_n(x) \\ 1-F_1(x) & 1-F_2(x) & \cdots & 1-F_n(x) \end{pmatrix}$$

Demostración. Siguiendo un razonamiento análogo al usado para demostrar el teorema 3, para cualquier $x \in \mathbb{R}$, tenemos que:

$$\begin{aligned} F_{(j)}(x) &= P(X_{(j)} \leq x) \\ &= \sum_{i=j}^{n} P(\text{Exactamente } i \text{ de los } X_1, \ldots, X_n \text{ son } \leq x) \\ &= \sum_{i=j}^{n} \frac{1}{i!(n-i)!} \sum_{P} F_{j_1}(x) \ldots F_{j_i}(x)(1 - F_{j_{i+1}}(x)) \ldots (1 - F_{j_n}(x)), \end{aligned}$$

donde $\sum_P$ denota la suma en todas las permutaciones $(j_1, j_2, \ldots, j_n)$ de $(1, 2, \ldots, n)$ y la división entre los factoriales se debe al número de veces que estamos repitiendo la misma probabilidad al sumar todas las permutaciones. A partir de aquí, y utilizando la definición de permanente, podemos escribir la distribución de $X_{(j)}$ como en el enunciado. □

Teorema 6. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes tales que, para cada $j = 1, \ldots, n$, X_j tiene distribución F_j y densidad f_j. Sean $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ los estadísticos ordenados obtenidos a partir de estas variables aleatorias. Entonces la función de densidad del estadístico ordenado $X_{(j)}$ (para cada $j = 1, \ldots, n$) se puede escribir como: [2]*

$$f_{(j)}(x) = \frac{1}{(j-1)!(n-j)!} \mathrm{Per}\, B_2,$$

donde

$$B_2 = \begin{array}{r} j-1\{ \\ \\ n-j\{ \end{array} \begin{pmatrix} F_1(x) & F_2(x) & \cdots & F_n(x) \\ f_1(x) & f_2(x) & \cdots & f_n(x) \\ 1-F_1(x) & 1-F_2(x) & \cdots & 1-F_n(x) \end{pmatrix}$$

Demostración. En primer lugar, teniendo en cuenta que $x < X_{(j)} \leq x + \delta x$ si y solo si $j-1$ de las variables están por debajo de x, otra está en $(x, x+\delta x]$ y el resto por encima de $x + \delta x$, más el resto de situaciones (que quedan recogidas en una o pequeña), tenemos que:

$$\begin{aligned} P(x < X_{(j)} \leq x + \delta x) \quad = \quad & \frac{1}{(j-1)!(n-j)!} \sum_{P} F_{i_1}(x) \ldots F_{i_{j-1}}(x)(F_{i_j}(x - \delta x) - F_{i_j}(x)) \\ & \times (1 - F_{i_{j+1}}(x + \delta x)) \ldots (1 - F_{i_n}(x + \delta x)) + o(\delta), \end{aligned}$$

donde la división por los factoriales se debe a la misma razón de antes.

Si dividimos ambos lados de la igualdad por δx y hacemos que δ decrezca hacia 0, entonces tenemos que la función de densidad de $X_{(j)}$ es:

$$f_{(j)}(x) = \frac{1}{(j-1)!(n-j)!} \sum_{P} F_{i_1}(x) \ldots F_{i_{j-1}}(x) f_{i_j}(x)(1 - F_{i_{j+1}}(x)) \ldots (1 - F_{i_n}(x))$$

A partir de aquí, y aplicando la definición de permanente, es inmediato ver que la función de densidad de $X_{(j)}$ se puede escribir como en el enunciado. □

Ejemplo. Sean $X_i \equiv Exp(\theta_i)$ independientes para $i = 1,2,3$ y supongamos que los θ_i son distintos. Entonces, aplicando el teorema 6, tenemos que:

$$\begin{aligned}
f_{(1)}(x) &= \frac{1}{2}Per\begin{pmatrix} f_1(x) & f_2(x) & f_3(x) \\ 1-F_1(x) & 1-F_2(x) & 1-F_3(x) \\ 1-F_1(x) & 1-F_2(x) & 1-F_3(x) \end{pmatrix} = \frac{1}{2}f_1(x)Per\begin{pmatrix} 1-F_2(x) & 1-F_3(x) \\ 1-F_2(x) & 1-F_3(x) \end{pmatrix} \\
&+\frac{1}{2}f_2(x)Per\begin{pmatrix} 1-F_1(x) & 1-F_3(x) \\ 1-F_1(x) & 1-F_3(x) \end{pmatrix} + \frac{1}{2}f_3(x)Per\begin{pmatrix} 1-F_1(x) & 1-F_2(x) \\ 1-F_1(x) & 1-F_2(x) \end{pmatrix} \\
&= f_1(x)(1-F_2(x))(1-F_3(x)) + f_2(x)(1-F_1(x))(1-F_3(x)) + f_3(x)(1-F_1(x))(1-F_2(x)) \\
&= \theta_1 e^{-\theta_1 x}e^{-\theta_2 x}e^{-\theta_3 x} + \theta_2 e^{-\theta_2 x}e^{-\theta_1 x}e^{-\theta_3 x} + \theta_3 e^{-\theta_3 x}e^{-\theta_1 x}e^{-\theta_2 x} \\
&= (\theta_1+\theta_2+\theta_3)e^{-(\theta_1+\theta_2+\theta_3)x},
\end{aligned}$$

por lo que la densidad de $X_{(1)}$ es la de una variable aleatoria exponencial de parámetro $\theta_1+\theta_2+\theta_3$, como sabemos que ocurre con el mínimo de variables aleatorias exponenciales independientes e idénticamente distribuidas.

Aplicando nuevamente el teorema 6, la densidad de $X_{(2)}$ es:

$$\begin{aligned}
f_{(2)}(x) &= Per\begin{pmatrix} F_1(x) & F_2(x) & F_3(x) \\ f_1(x) & f_2(x) & f_3(x) \\ 1-F_1(x) & 1-F_2(x) & 1-F_3(x) \end{pmatrix} = F_1(x)Per\begin{pmatrix} f_2(x) & f_3(x) \\ 1-F_2(x) & 1-F_3(x) \end{pmatrix} \\
&+F_2(x)Per\begin{pmatrix} f_1(x) & f_3(x) \\ 1-F_1(x) & 1-F_3(x) \end{pmatrix} + F_3(x)Per\begin{pmatrix} f_1(x) & f_2(x) \\ 1-F_1(x) & 1-F_2(x) \end{pmatrix} \\
&= F_1(x)[f_2(x)(1-F_3(x)) + f_3(x)(1-F_2(x))] + F_2(x)[f_1(x)(1-F_3(x)) + f_3(x)(1-F_1(x))] \\
&+F_3(x)[f_1(x)(1-F_2(x)) + f_2(x)(1-F_1(x))] \\
&= (1-e^{-\theta_1 x})(\theta_2 e^{-\theta_2 x}e^{-\theta_3 x} + \theta_3 e^{-\theta_3 x}e^{-\theta_2 x}) + (1-e^{-\theta_2 x})(\theta_1 e^{-\theta_1 x}e^{-\theta_3 x} + \theta_3 e^{-\theta_3 x}e^{-\theta_1 x}) \\
&+(1-e^{-\theta_3 x})(\theta_1 e^{-\theta_1 x}e^{-\theta_2 x} + \theta_2 e^{-\theta_2 x}e^{-\theta_1 x})
\end{aligned}$$

Volviendo a aplicar el teorema 6, tenemos que la densidad del máximo (X_3) es:

$$\begin{aligned}
f_{(3)}(x) &= \frac{1}{2}Per\begin{pmatrix} F_1(x) & F_2(x) & F_3(x) \\ F_1(x) & F_2(x) & F_3(x) \\ f_1(x) & f_2(x) & f_3(x) \end{pmatrix} = \frac{1}{2}f_1(x)Per\begin{pmatrix} F_2(x) & F_3(x) \\ F_2(x) & F_3(x) \end{pmatrix} \\
&+\frac{1}{2}f_2(x)Per\begin{pmatrix} F_1(x) & F_3(x) \\ F_1(x) & F_3(x) \end{pmatrix} + \frac{1}{2}f_3(x)Per\begin{pmatrix} F_1(x) & F_2(x) \\ F_1(x) & F_2(x) \end{pmatrix} \\
&= f_1(x)F_2(x)F_3(x) + f_2(x)F_1(x)F_3(x) + f_3(x)F_1(x)F_2(x) \\
&= \theta_1 e^{-\theta_1 x}(1-e^{-\theta_2 x})(1-e^{-\theta_3 x}) + \theta_2 e^{-\theta_2 x}(1-e^{-\theta_1 x})(1-e^{-\theta_3 x}) \\
&+\theta_3 e^{-\theta_3 x}(1-e^{-\theta_1 x})(1-e^{-\theta_2 x})
\end{aligned}$$

Como vemos, las funciones de densidad obtenidas en este caso son mucho más complicadas que las obtenidas para el mínimo y el máximo en el caso $\theta_1 = \theta_2 = \theta_3 = \theta$, especialmente en el caso del máximo.

Capítulo 3

Órdenes estocásticos univariantes

Hasta ahora, hemos visto cómo se distribuyen los estadísticos ordenados de un conjunto de variables aleatorias, sean éstas i.i.d. o no (de hecho, también lo hemos visto en el caso particular de variables aleatorias i.i.d.). En este capítulo, vamos a introducir otras formas de ordenar variables aleatorias, basadas en la distribución y en la densidad de las variables aleatorias implicadas. Asimismo, para cada una de estas formas, vamos a ver una serie de resultados, algunos de los cuales serán de interés en el capítulo 4, donde ordenaremos estadísticos ordenados pertenecientes a diferentes muestras.

3.1. Introducción

Las relaciones de orden estocástico son un caso particular de las relaciones de orden parcial. Recordemos que éstas últimas son por definición reflexivas, antisimétricas y transitivas. Si consideramos el conjunto S de las funciones de distribución de todas las variables aleatorias que toman valores reales, esto es, dado un espacio de probabilidad $(\Omega, \mathscr{F}, P)$, todas las posibles funciones de distribución asociadas a variables aleatorias $X : \Omega \to \mathbb{R}$; entonces una relación de orden en el conjunto S se dice **orden estocástico**. Usaremos el símbolo $\preceq$ para referirnos a una relación genérica de este tipo.

Ejemplo. Sea $U \equiv U(0,1)$. Entonces $1-U$ tiene la misma distribución y además se tiene que $U \preceq 1-U$ en la mayoría de los órdenes que vamos a considerar (excepto el llamado orden casi seguro, como veremos más adelante).

Por convenio, si las variables aleatorias X e Y tienen distribuciones F_X y F_Y respectivamente de manera que $F_X \preceq F_Y$, entonces puede resultar conveniente escribir $X \preceq Y$, entendiendo entonces que la antisimetría de la relación se da entre distribuciones pero no entre variables aleatorias, ya que puede haber varias variables aleatorias diferentes que tengan la misma distribución. [10]

3.2. Orden estocástico usual

El ejemplo más claro de orden estocástico es la comparación puntual entre funciones de distribución. En este sentido, si $F_X(t) \geq F_Y(t)$, $\forall t \in \mathbb{R}$, entonces X tiene mayor probabilidad que Y de tomar valores pequeños, mientras que tiene menor probabilidad que Y de tomar valores grandes. Esto lleva a la siguiente definición.

Definición. Sean X e Y dos variables aleatorias. Se dice que X es menor que Y con respecto al orden estocástico usual (denotado $X \leq_{st} Y$) si

$$F_X(t) \geq F_Y(t), \ \forall t \in \mathbb{R},$$

es decir, si

$$S_X(t) \leq S_Y(t), \ \forall t \in \mathbb{R},$$

donde $S_X(t)$ denota la función de supervivencia de X dada por

$$S_X(t) = P(X > t) = 1 - F_X(t).$$

[10]

Nota. Esta relación de orden es especialmente conocida como *orden estocástico*. Para distinguirla de otros órdenes estocásticos, la llamaremos *orden estocástico usual*.

Nota. A primera vista puede parecer extraño decir que $F_X \leq_{st} F_Y$ si $F_X(t) \geq F_Y(t)$, $\forall t \in \mathbb{R}$. Pero téngase en cuenta que queremos que Y sea estocásticamente mayor que X si toma valores grandes con mayor probabilidad. Como la función de distribución describe la probabilidad de tomar valores pequeños, hay que hacer un cambio de sentido en la desigualdad. Además, $\leq_{st}$ puede considerarse una generalización de $\leq$ en la recta real, dado que $\forall a, b \in \mathbb{R}$, se cumple que

$$a \leq b \Longrightarrow a \leq_{st} b$$

si consideramos a y b como variables aleatorias degeneradas. En otras palabras:

$$a \leq b \Longrightarrow \delta_a \leq_{st} \delta_b,$$

donde δ_ρ es la delta de Dirac que corresponde a la función de distribución asociada a una variable aleatoria degenerada en el punto ρ. [10]

Nota. Un candidato natural para una relación de orden que compare el tamaño de variables aleatorias es la relación $X \preceq_{c.s.} Y$, que se da si y solo si $X(\omega) \leq Y(\omega)$ para casi todo ω (es decir, para todo ω salvo los que formen parte de un conjunto de probabilidad nula). Sin embargo, esta relación no depende únicamente de las distribuciones. Esto se puede ver si tenemos en cuenta que siempre $X \preceq_{c.s.} X$, pero $X \preceq_{c.s.} Y$ no se cumple si X e Y son v.a.i.i.d. con la misma distribución (no degenerada). Para ello, podemos pensar en el ejemplo anterior, en el que teníamos las variables U y $1-U$ con la misma distribución ($U(0,1)$). Sin embargo, $U \npreceq_{c.s.} 1-U$, ni a la inversa. Esto hace que la definición de $\leq_{st}$ sea mucho más útil. [10]

Ahora, vamos a enunciar (sin demostración) algunas de las propiedades más relevantes del orden estocástico usual.

Teorema 7. *$X \leq_{st} Y$ si y solo si $Ef(X) \leq Ef(Y)$, para toda función f creciente para la que existen ambas esperanzas. Además, si una desigualdad de este tipo se cumple para todas X e Y con $X \leq_{st} Y$, entonces f es creciente. [10]*

Definición. Sea $(X_n)_{n\geq 1}$ una sucesión de variables aleatorias. Se dice que X_n converge en distribución a la variable aleatoria X (denotado $X_n \longrightarrow X$ (D)) si $\forall x$ donde F (la función de distribución de X) es continua se cumple que:

$$\lim_{n\to\infty} F_n(x) = F(x),$$

donde F_n es la función de distribución de X_n para cada $n \in \mathbb{N}$. Este tipo de convergencia también es conocido como **convergencia débil**.

Teorema 8. *La relación $\leq_{st}$ es cerrada con respecto a la convergencia débil, es decir, si $F_n \leq_{st} G_n$, $\forall n$, y las sucesiones $(F_n)_{n\geq 1}$ y $(G_n)_{n\geq 1}$ convergen débilmente a F y G respectivamente, entonces $F \leq_{st} G$. [10]*

Definición. Recordemos que una función creciente de $\mathbb{R}$ en $\mathbb{R}$ es una función f tal que si $x \leq y$, entonces $f(x) \leq f(y)$. Esta definición se puede extender a funciones de $\mathbb{R}^n$ en $\mathbb{R}$, teniendo en cuenta que dados los vectores $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n)$ de $\mathbb{R}^n$, decimos que $\mathbf{x} \leq \mathbf{y}$ si $x_i \leq y_i$ para cada $i = 1, \ldots, n$.

Teorema 9. *Sean $X_1, \dots, X_n, Y_1, \dots, Y_n$ variables aleatorias independientes tales que $X_i \leq_{st} Y_i$ para todo $i = 1, \dots, n$ y $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}$ una función creciente. Entonces*

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

[10]

Corolario 1. *Manteniendo las hipótesis del teorema anterior, sean $X_{(1)}, \dots, X_{(n)}$ los estadísticos ordenados correspondientes a $X_1, \dots, X_n$ e $Y_{(1)}, \dots, Y_{(n)}$ los estadísticos ordenados correspondientes a $Y_1, \dots, Y_n$. Entonces, para todo $i = 1, \dots, n$ se cumple que*

$$X_{(i)} \leq_{st} Y_{(i)}.$$

[10]

Demostración. La función $\varphi_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ definida como $\varphi_i(x_1, \dots, x_n) =$ *i-ésimo valor más pequeño* es una función creciente. A partir de aquí, podemos aplicar el teorema anterior y así tenemos el resultado. □

3.3. Orden de tasa de fallo

Hay muchas situaciones en las que necesitamos o nos puede resultar útil un orden más fuerte que el orden estocástico usual. Por ejemplo, si alguien quiere comprar un coche y duda entre dos modelos con tiempos de vida X e Y con $X \leq_{st} Y$ que valen lo mismo, se decantará por el segundo modelo. Sin embargo, si esta persona quiere comprar un coche de segunda mano con un año de vida, de manera que los tiempos de vida restantes son X' e Y' con $P(X' > t) = P(X > 1 + t | X > 1)$ y análogamente para Y', ¿se sigue cumpliendo que $X' \leq_{st} Y'$? El siguiente ejemplo demuestra que no necesariamente. [10]

Ejemplo. Sea $X \equiv U(0,3)$ e Y con la siguiente densidad:

$$f(x) = \begin{cases} \frac{1}{6} & \text{si } 0 < x \leq 1 \\ \frac{1}{2} & \text{si } 1 < x \leq 2 \\ \frac{1}{3} & \text{si } 2 < x < 3 \end{cases}$$

Entonces es fácil ver que $X \leq_{st} Y$. Para comparar X' e Y', vamos a ver, en primer lugar, que $X' \equiv U(0,2)$:

Es obvio que el rango de distribución de X' es el intervalo $(0,2)$. A partir de aquí, por definición, para cada $x \in (0,2)$, $P(X' > x) = P(X > 1 + x | X > 1)$, por lo que aplicando la definición de probabilidad condicional tenemos que:

$$P(X' \leq x) = P(X \leq 1 + x | X > 1) = \frac{P(1 < X \leq 1 + x)}{P(X > 1)} = \frac{x/3}{2/3} = \frac{x}{2},$$

luego la densidad de X' es la derivada de esta función, que es $\frac{1}{2}$, es decir, la densidad de la distribución $U(0,2)$.

Análogamente podemos comprobar que Y' tiene densidad

$$g(x) = \begin{cases} \frac{3}{5} & \text{si } 0 < x \leq 1 \\ \frac{2}{5} & \text{si } 1 < x < 2 \end{cases}$$

por lo que es fácil ver que $X' \geq_{st} Y'$

Ahora nos podemos preguntar: ¿qué hipótesis más fuertes debemos tomar para asegurar que el orden estocástico usual se cumple para todos los coches de segunda mano de edad t?, es decir, ¿cuándo se cumple $[X|X>t] \leq_{st} [Y|Y>t]$ para todo t? Aplicando la definición de $\leq_{st}$, podemos poner esta desigualdad de la siguiente manera (siempre que los condicionantes tengan sentido):

$$P(X > s+t|X > t) \leq P(Y > s+t|Y > t),\ \forall s \geq 0.$$

Y esto es equivalente (aplicando la definición de probabilidad condicional como antes) a que

$$\frac{S_Y(t)}{S_X(t)} \leq \frac{S_Y(s+t)}{S_X(s+t)},$$

siempre que los denominadores no se anulen.

Entenderemos a partir de ahora la desigualdad $\frac{a}{b} \leq \frac{c}{d}$ como $ad \leq bc$. Notemos que con esta convención la desigualdad anterior se cumple para todo $s > 0$ y $t \in \mathbb{R}$. Además, a partir de ahora, creciente significará no decreciente (análogamente con decreciente). Todo esto nos lleva a la siguiente definición:

Definición. Sean X e Y dos variables aleatorias. Decimos que X es menor que Y con respecto al orden de tasa de fallo (denotado $X \leq_{hr} Y$), si la función

$$t \longmapsto \frac{S_Y(t)}{S_X(t)}$$

es creciente. [10]

Como en la sección anterior, veamos algunas propiedades útiles de este orden: [10]

Teorema 10. *El orden de tasa de fallo es cerrado con respecto a la convergencia débil, es decir, si $F_n \leq_{hr} G_n$, $\forall n$, y las sucesiones $(F_n)_{n\geq 1}$ y $(G_n)_{n\geq 1}$ convergen débilmente a F y G respectivamente, entonces $F \leq_{hr} G$.*

Teorema 11. *Si $X \leq_{hr} Y$, entonces $X \leq_{st} Y$.*

Demostración. Como $\lim_{t\to-\infty} \frac{S_Y(t)}{S_X(t)} = 1$ (trivial ya que $\lim_{t\to-\infty} S(t) = 1$ para cualquier función de supervivencia S), si $X \leq_{hr} Y$, entonces $\frac{S_Y(t)}{S_X(t)} \geq 1$, luego $F_X(t) \geq F_Y(t)$, $\forall t \in \mathbb{R}$. □

A veces es útil considerar, en la definición de orden de tasa de fallo, la función resultante de sustituir la función de supervivencia por la función de distribución. Esto nos da el orden de tasa de fallo invertido.

Definición. Sean X e Y dos variables aleatorias. Decimos que X es menor que Y con respecto al orden de tasa de fallo invertido (denotado $X \leq_{rh} Y$), si la función

$$t \longmapsto \frac{F_Y(t)}{F_X(t)}$$

es creciente. [10]

El orden de tasa de fallo invertido comparte muchas de las propiedades del orden de tasa de fallo. De hecho, hay una gran dualidad entre $\leq_{hr}$ y $\leq_{rh}$. Esta dualidad se debe al siguiente resultado.

Teorema 12. *Sea g una función continua y estrictamente decreciente. Entonces $X \leq_{hr} Y$ si y solo si $g(X) \geq_{rh} g(Y)$. [10]*

El siguiente resultado son dos consecuencias de este teorema. [10]

Proposición 2. *Sean X e Y dos variables aleatorias. Entonces:*

1. *El orden de tasa de fallo invertido es cerrado con respecto a la convergencia débil, es decir, si $F_n \leq_{rh} G_n$, $\forall n$, y las sucesiones $(F_n)_{n\geq 1}$ y $(G_n)_{n\geq 1}$ convergen débilmente a F y G respectivamente, entonces $F \leq_{rh} G$.*

2. *Si $X \leq_{rh} Y$, entonces $X \leq_{st} Y$.*

3.4. Orden de razón de verosimilitudes

Una característica interesante del orden de tasa de fallo es que $X \leq_{hr} Y$ si y solo si para todo $t \in \mathbb{R}$ se cumple que $[X|X > t] \leq_{st} [Y|Y > t]$. Esto es importante para el análisis de las distribuciones de tiempos de vida. Sin embargo, en algunas situaciones nos puede interesar tener $[X|X \in A] \leq_{st} [Y|Y \in A]$ para todos los posibles eventos A. Esto nos lleva al orden de razón de verosimilitudes.

Definición. Sean X e Y variables aleatorias. Decimos que X es menor que Y en orden de razón de verosimilitudes (denotado $X \leq_{lr} Y$) si X e Y tienen densidades f_X y f_Y respecto a alguna medida dominante μ tales que:

$$f_X(t)f_Y(s) \leq f_X(s)f_Y(t), \ \forall s \leq t. \tag{3.1}$$

Esto es, $P(X \in A) = \int_A f_X(x)d\mu(x)$ para todo A boreliano de $\mathbb{R}$ y análogamente con Y, es decir, $P(Y \in A) = \int_A f_Y(x)d\mu(x)$. [10]

Nota. Notemos lo siguiente: [10]

a) Hemos puesto una definición con una medida dominante arbitraria μ para incluir el caso de densidades de distribuciones continuas (donde μ es la medida de Lebesgue), y el caso de distribuciones discretas (donde μ es la medida de contar), así como situaciones más generales como mixturas de ambas.

b) La desigualdad 3.1 simplemente dice que la tasa $\frac{f_Y}{f_X}$ es creciente, pero, como hemos indicado, escrito de una forma que elude la dificultad de decir por separado qué hacer si el numerador, el denominador, o ambos se cancelan.

c) Esta relación también puede definirse como: $X \leq_{lr} Y$ si para toda medida dominante μ y todo par de densidades f_X y f_Y, se satisface 3.1 μ-c.s. en la unión de los soportes de X e Y (esto es, el conjunto donde X o Y pueden tomar valores), ya que en conjuntos de medida nula, las densidades se pueden definir arbitrariamente de modo que se verifique la anterior desigualdad.

El siguiente resultado es de gran interés, ya que muestra que el orden de razón de verosimilitudes es más exigente que los tres anteriores. [10]

Teorema 13. *Si $X \leq_{lr} Y$, entonces $X \leq_{hr} Y$ y $X \leq_{rh} Y$.*

Este último resultado nos dice que el orden de razón de verosimilitudes es el más fuerte de los vistos en el trabajo, como hemos indicado antes. A pesar de ello, a veces es el más fácil de verificar. Como ejemplo de ello, vamos a considerar las densidades $\Gamma(\alpha,\lambda)$ y $\Gamma(\beta,\mu)$ con $\alpha \leq \beta$ y $\lambda \geq \mu$. Recordemos que la función de densidad de una variable aleatoria Γ con parámetro de forma α y parámetro de escala λ es

$$f_{\Gamma(\alpha,\lambda)} = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & , \text{si } x \geq 0 \\ 0 & , \text{si } x < 0 \end{cases} \tag{3.2}$$

En primer lugar, vamos a comprobar que $\leq_{lr}$ se verifica cuando tienen el mismo parámetro de escala y diferente parámetro de forma. Es decir, vamos a comprobar que $\Gamma(\alpha,\lambda) \leq_{lr} \Gamma(\beta,\lambda)$. Para ello, sean s y t con $s \leq t$. Si $s \leq 0$ y/o $t \leq 0$, entonces ambos lados de la desigualdad de la definición se anulan, por lo que ésta se verifica (al no ser estricta). En otro caso, como $\left(\frac{t}{s}\right)^{\alpha-1} \leq \left(\frac{t}{s}\right)^{\beta-1}$, tenemos que:

$$\frac{\lambda^\alpha t^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)} \frac{\lambda^\beta s^{\beta-1}e^{-\lambda s}}{\Gamma(\beta)} \leq \frac{\lambda^\alpha s^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)} \frac{\lambda^\beta t^{\beta-1}e^{-\lambda s}}{\Gamma(\beta)} = \frac{\lambda^\alpha s^{\alpha-1}e^{-\lambda s}}{\Gamma(\alpha)} \frac{\lambda^\beta t^{\beta-1}e^{-\lambda t}}{\Gamma(\beta)}.$$

Ahora, vamos a comprobar que $\leq_{lr}$ se verifica cuando tienen el mismo parámetro de forma y diferente parámetro de escala. Es decir, vamos a comprobar que $\Gamma(\beta,\lambda) \leq_{lr} \Gamma(\beta,\mu)$. Para ello, sean s y t con

$s \leq t$. Si $s \leq 0$ y/o $t \leq 0$, entonces ambos lados de la desigualdad de la definición se anulan, por lo que ésta se verifica (al no ser estricta). En otro caso, como $e^{-\lambda(t-s)} \leq e^{-\mu(t-s)}$, tenemos que:

$$\frac{\lambda^\beta t^{\beta-1}e^{-\lambda t}}{\Gamma(\beta)}\frac{\mu^\beta s^{\beta-1}e^{-\mu s}}{\Gamma(\beta)} \leq \frac{\lambda^\beta t^{\beta-1}e^{-\lambda s}}{\Gamma(\beta)}\frac{\mu^\beta s^{\beta-1}e^{-\mu t}}{\Gamma(\beta)} = \frac{\lambda^\beta s^{\beta-1}e^{-\lambda s}}{\Gamma(\beta)}\frac{\mu^\beta t^{\beta-1}e^{-\mu t}}{\Gamma(\beta)}.$$

Para terminar la comprobación, basta aplicar transitividad para obtener que $\Gamma(\alpha,\lambda) \leq_{lr} \Gamma(\beta,\mu)$.

Para este orden, también se verifica la preservación del orden bajo convergencia débil:

Teorema 14. *El orden de razón de verosimilitudes es cerrado con respecto a la convergencia débil, es decir, si $F_n \leq_{lr} G_n$, $\forall n$, y las sucesiones $(F_n)_{n\geq 1}$ y $(G_n)_{n\geq 1}$ convergen débilmente a F y G respectivamente, entonces $F \leq_{lr} G$. [10]*

Utilizaremos estos dos últimos resultados en el siguiente capítulo.

3.5. Densidades log-cóncavas y log-convexas

Las densidades log-cóncavas y log-convexas son importantes porque aparecen muy frecuentemente y también porque tienen propiedades interesantes. Además, la log-concavidad de la densidad implica otras propiedades con claro significado físico y es más fácil de comprobar que otras propiedades más débiles pero a veces más interesantes. Dicha importancia fue reconocida quizá por primera vez por Karlin y Rubin (1956) [7]. Más recientemente, ha jugado un papel en teoría económica. En este apartado, vamos a incluir un resultado acerca del orden de razón de verosimilitudes donde se utiliza dicha propiedad. [8]

Definición. Sea F la función de distribución de una variable aleatoria absolutamente continua y sea f la función de densidad de esta misma variable aleatoria. [8]

- Decimos que F tiene densidad **log-cóncava** si $\log f$ es cóncavo (en el intervalo donde $f > 0$).
- Decimos que F tiene densidad **log-convexa** si $f(x) = 0$ para $x < 0$ y $\log f$ es convexo en $[0,\infty)$ (en este caso, se puede comprobar que f no se anula en $[0,\infty)$).

Proposición 3. *Algunos ejemplos de densidades log-cóncavas y log-convexas: [8]*

1. *Las densidades normales con media μ y desviación típica σ (esto es, las que son de la forma $\frac{1}{\sqrt{2}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ para todo $x \in \mathbb{R}$) son log-cóncavas.*
2. *Las densidades gamma con parámetro de forma α y parámetro de escala λ, que hemos recordado en 3.2, son log-cóncavas si $\alpha \geq 1$ y log-convexas si $0 < \alpha \leq 1$. En particular, las densidades exponenciales (que son densidades gamma con $\alpha = 1$) son log-cóncavas y log-convexas.*
3. *Las densidades de Weibull con parámetro de forma α y parámetro de escala λ (esto es, las que son de la forma $\alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^\alpha}$ si $x \geq 0$, 0 en otro caso [12]) son log-cóncavas si $\alpha \geq 1$ y log-convexas si $0 < \alpha \leq 1$.*

Demostración. En todos los casos, estas propiedades se pueden comprobar verificando la monotonía de la derivada del logaritmo de f. Por ejemplo, para el caso gamma,

$$\log f = \log\left(\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}\right) = \alpha\log\lambda + (\alpha-1)\log x - \lambda x - \Gamma(\alpha).$$

A partir de aquí, las dos primeras derivadas salen de la siguiente manera:

$$(\log f)' = \frac{\alpha-1}{x} - \lambda$$

$$(\log f)'' = \frac{1-\alpha}{x^2}$$

Como $x > 0$ (necesario para que tenga sentido $\log x$), $(\log f)''$ tiene el mismo signo que $1-\alpha$, es decir, es positivo para $0 < \alpha \leq 1$ y negativo para $\alpha \geq 1$. De aquí deducimos que $(\log f)'$ es creciente para $0 < \alpha \leq 1$ y decreciente para $\alpha \geq 1$, por lo que $\log f$ es convexo para $0 < \alpha \leq 1$ y cóncavo para $\alpha \geq 1$, como queríamos probar. □

El siguiente resultado, que será utilizado en el próximo capítulo, nos habla de la preservación del orden lr bajo convoluciones de distribuciones log-cóncavas. Su prueba se puede encontrar en el libro de Shaked y Shanthikuman (2007), página 48. [12]

Teorema 15. *Sean X e Y dos v.a. independientes y sea Z otra v.a., independiente de las dos anteriores y con densidad log-cóncava. Si $X \leq_{lr} Y$, se tiene que*

$$X + Z \leq_{lr} Y + Z.$$

[12]

Ejemplo. Vamos a considerar la convolución de variables aleatorias cualesquiera (X e Y) con Z normal con media 0 y varianza ε^2, que utilizaremos en el siguiente capítulo. Sabemos que Z tiene densidad log-cóncava por ser normal. Supongamos que $X \leq_{lr} Y$. Entonces, si se satisface la independencia de X, Y y Z (dos a dos), aplicando el teorema 15, $X + Z \leq_{lr} Y + Z$.

Capítulo 4

Aplicación: Comparaciones en orden lr de estadísticos ordenados

4.1. Introducción

Como vimos en el capítulo 2, los estadísticos ordenados se utilizan en muchas áreas de ámbitos muy diferentes. Además, en los últimos años se ha publicado una gran cantidad de textos sobre comparaciones estocásticas de estadísticos ordenados. Un ejemplo de ello es el siguiente resultado:

Proposición 4. *Sean $X_1,...,X_m,Y_1,...,Y_m$ variables aleatorias independientes. Si $X_j \leq_{lr} Y_i$ para cualesquiera i y j, entonces $X_{(k)} \leq_{lr} Y_{(k)}$, para $k = 1,...,m$.*

Esta proposición es un caso particular de un teorema cuya prueba aparece en el libro de Shaked y Shantikumar. [12] Notemos que la condición de esta proposición es muy exigente (requiere que todas las X_i sean menores o iguales en el orden lr que cualquiera de las Y_i), por lo que la vamos a tratar de debilitar, pero sólo en un caso concreto.

En este capítulo, vamos a considerar $X_1,...,X_p$ una muestra aleatoria de tamaño p de una distribución F y $X_{p+1},...,X_n$ otra muestra aleatoria independiente de tamaño q de una distribución G, donde $n = p+q$ y $0 \leq p < n$. Vamos a denotar

$$X_{(1)}(p,q) \leq X_{(2)}(p,q) \leq ... \leq X_{(n)}(p,q)$$

a los estadísticos ordenados de la muestra conjunta, ya que ahora también es importante el tamaño de cada muestra. En el artículo de Ding *et al.* [4] se demostró que para todo $1 \leq k \leq n$ y $0 \leq p < n$:

$$G \leq_{hr} F \Longrightarrow X_{(k)}(p,q) \leq_{hr} X_{(k)}(p+1,q-1)$$

$$G \leq_{rh} F \Longrightarrow X_{(k)}(p,q) \leq_{rh} X_{(k)}(p+1,q-1).$$

Como ejemplo, recordemos la relación entre los estadísticos ordenados y los sistemas k de n vista en la introducción al capítulo 2. Dividimos el sistema en dos secciones B_1 y B_2, en cada una de las cuales todos los componentes tienen la misma distribución. Entonces, si consideramos un sistema k de n en el que tomamos p componentes de B_1 y $q = n-p$ componentes de B_2, y suponemos que todos los componentes funcionan de manera independiente, entonces el tiempo de vida del sistema es $X_{(n-k+1)}(p,q)$ y las desigualdades anteriores nos permiten comparar el tiempo de vida del sistema en función del número de componentes más resistentes.

El objetivo de este capítulo es probar la implicación análoga para el orden de razón de verosimilitudes (recordemos que este orden es el más fuerte de los que hemos visto en el capítulo anterior), es decir:

$$G \leq_{lr} F \Longrightarrow X_{(k)}(p,q) \leq_{lr} X_{(k)}(p+1,q-1).$$

Para ello, nos vamos a basar en la demostración que aparece en el artículo de Yang y Zhuang. [13]

4.2. Resultado principal

Teorema 16. *Sea $X_1,\dots,X_p$ una muestra aleatoria de tamaño p de una distribución F y $X_{p+1},\dots,X_n$ otra muestra aleatoria independiente de tamaño q de una distribución G, de manera que ambas distribuciones son absolutamente continuas con densidades respectivas f y g, donde $n=p+q$ y $0\le p<n$. Entonces, para cada $k=1,\dots,n$:*

$$G\le_{lr} F \Longrightarrow X_{(k)}(p,q)\le_{lr} X_{(k)}(p+1,q-1).$$

Demostración. Vamos a considerar dos casos:

Caso 1: f y g son derivables y estrictamente positivas en todo $\mathbb{R}$. Vamos a utlizar la siguiente notación:

$$[i,j,l]_{p,q} = Per \begin{array}{l} i\{ \\ j\{ \\ l\{ \end{array}\begin{pmatrix} F(x) & \cdots & F(x) & G(x) & \cdots & G(x) \\ f(x) & \cdots & f(x) & g(x) & \cdots & g(x) \\ 1-F(x) & \cdots & 1-F(x) & 1-G(x) & \cdots & 1-G(x) \end{pmatrix} \tag{4.1}$$

$$[i,\overline{j},l]_{p,q} = Per \begin{array}{l} i\{ \\ j\{ \\ l\{ \end{array}\begin{pmatrix} F(x) & \cdots & F(x) & G(x) & \cdots & G(x) \\ f'(x) & \cdots & f'(x) & g'(x) & \cdots & g'(x) \\ 1-F(x) & \cdots & 1-F(x) & 1-G(x) & \cdots & 1-G(x) \end{pmatrix} \tag{4.2}$$

En ambos casos, la primera columna se repite p veces, la última columna se repite q veces y además $i+j+l=p+q=n$.

Aplicando el teorema 6 para $j=n-k+1$ y sabiendo que $F_1(x)=\dots=F_p(x)=F(x)$ y $F_{p+1}(x)=\dots=F_n(x)=G(x)$, tenemos que la densidad de X_{n-k+1}, esto es, $f_{(n-k+1)}(x,p)=[n-k,1,k-1]_{p,q}$. Por lo tanto, aplicando la definición de orden de razón de verosimilitudes, para probar el resultado, tenemos que probar que la función

$$\varphi_p(x)=\frac{[n-k,1,k-1]_{p+1,q-1}}{[n-k,1,k-1]_{p,q}},\ k=1,2,\dots,n$$

es creciente. Notemos, por el teorema 1, que la derivada del denominador en la expresión anterior es

$$(n-k)[n-k-1,2,k-1]_{p,q}+[n-k,\overline{1},k-1]_{p,q}-(k-1)[n-k,2,k-2]_{p,q}$$

y la derivada del numerador es

$$(n-k)[n-k-1,2,k-1]_{p+1,q-1}+[n-k,\overline{1},k-1]_{p+1,q-1}-(k-1)[n-k,2,k-2]_{p+1,q-1}.$$

De esta manera, tomando la derivada de φ_p, considerando solo su numerador (ya que el denominador es siempre positivo) y denotando $\equiv$ a la igualdad de signo tenemos que:

$$\begin{aligned}\varphi_p'(x) \quad &\equiv \quad [n-k,1,k-1]_{p,q}\Big[(n-k)[n-k-1,2,k-1]_{p+1,q-1}+[n-k,\overline{1},k-1]_{p+1,q-1}\\ &\quad -(k-1)[n-k,2,k-2]_{p+1,q-1}\Big]-[n-k,1,k-1]_{p+1,q-1}\Big[(n-k)[n-k-1,2,k-1]_{p,q}\\ &\quad +[n-k,\overline{1},k-1]_{p,q}-(k-1)[n-k,2,k-2]_{p,q}\Big]\\ &= \quad (n-k)J_1+J_2+(k-1)J_3,\end{aligned}$$

donde

$$J_1=[n-k-1,2,k-1]_{p+1,q-1}[n-k,1,k-1]_{p,q}-[n-k-1,2,k-1]_{p,q}[n-k,1,k-1]_{p+1,q-1} \tag{4.3}$$

$$J_2=[n-k,\overline{1},k-1]_{p+1,q-1}[n-k,1,k-1]_{p,q}-[n-k,\overline{1},k-1]_{p,q}[n-k,1,k-1]_{p+1,q-1} \tag{4.4}$$

$$J_3 = [n-k,1,k-1]_{p+1,q-1}[n-k,2,k-2]_{p,q} - [n-k,1,k-1]_{p,q}[n-k,2,k-2]_{p+1,q-1} \quad (4.5)$$

Ahora, vamos a definir las siguientes matrices (de dimensión $n-1 \times n$ y con una interpretación similar a la de 4.1):

$A_1 = (n-k-1,1,k-1)_{p+1,q-1}$ $B_1 = (n-k-1,1,k-1)_{p,q}$
$A_2 = (n-k,0,k-1)_{p+1,q-1}$ $B_2 = (n-k,0,k-1)_{p,q}$
$A_3 = (n-k,1,k-2)_{p+1,q-1}$ $B_3 = (n-k,1,k-2)_{p,q}$

Asimismo, vamos a denotar $[A_j^m]$ y $[B_j^m]$ a los permanentes de las matrices resultantes de eliminar la columna m a A_j y B_j, respectivamente. Notemos, además, que si $A_j = (a_j,b_j,c_j)_{p+1,q-1}$, entonces $[A_j^1] = [a_j,b_j,c_j]_{p,q-1}$ y $[A_j^n] = [a_j,b_j,c_j]_{p+1,q-2}$. Análogamente, si $B_j = (a_j,b_j,c_j)_{p,q}$, entonces $[B_j^1] = [a_j,b_j,c_j]_{p-1,q}$. Como además el permanente de una matriz es invariante bajo permutación de columnas, podemos aplicar la desigualdad de Alexandroff a las columnas p y $p+1$ de $[A_j^1]$, obteniendo:

$$[A_j^1]^2 = [a_j,b_j,c_j]_{p,q-1}^2 \geq [a_j,b_j,c_j]_{p+1,q-2}[a_j,b_j,c_j]_{p-1,q} = [A_j^n][B_j^1].$$

Todo lo anterior es válido para $j = 1,2,3$. Además, es inmediato que $[A_j^1] = [B_j^n]$ para $j = 1,2,3$. A partir de aquí podemos definir

$$\alpha_j := (p+1)q[A_j^1][B_j^n] - p(q-1)[A_j^n][B_j^1] = (p+1)q[A_j^1]^2 - p(q-1)[A_j^n][B_j^1] \geq 0, \quad (4.6)$$

para $j = 1,2,3$ aplicando la desigualdad anterior.

Recordemos que si $G \leq_{lr} F$, entonces $G \leq_{rh} F$ y $G \leq_{hr} F$. Además, si $G \leq_{lr} F$, entonces $\frac{f}{g}$ es creciente, por lo que

$$f'(x)g(x) - f(x)g'(x) \geq 0, \quad (4.7)$$

ya que esta expresión es el numerador de la derivada de $\frac{f}{g}$, y el denominador de esta misma derivada es siempre positivo. Análogamente, si $G \leq_{hr} F$, entonces

$$f(x)G(x) - F(x)g(x) \geq 0, \quad (4.8)$$

y si $G \leq_{rh} F$, entonces

$$(1-F(x))g(x) - f(x)(1-G(x)) \geq 0. \quad (4.9)$$

Desarrollando los primeros factores de J_1 en 4.3 por la fila intermedia (la que contiene a las f) y los segundos factores por la primera fila (la que contiene a las F), tenemos que:

$$\begin{aligned}
J_1 &= \Big((p+1)f(x)[A_1^1] + (q-1)g(x)[A_1^n]\Big)\Big(pF(x)[B_1^1] + qG(x)[B_1^n]\Big) \\
&\quad - \Big(pf(x)[B_1^1] + qg(x)[B_1^n]\Big)\Big((p+1)F(x)[A_1^1] + (q-1)G(x)[A_1^n]\Big) \\
&= \Big((p+1)q[A_1^1][B_1^n] - p(q-1)[B_1^1][A_1^n]\Big)(f(x)G(x) - F(x)g(x)) \\
&= \alpha_1(f(x)G(x) - F(x)g(x)) \geq 0
\end{aligned}$$

por 4.6 y 4.8. Siguiendo un proceso análogo y desarrollando los primeros factores de J_2 en 4.4 por la fila que contiene a las f' y los segundos factores por la fila que contiene a las f:

$$\begin{aligned}
J_2 &= \Big((p+1)f'(x)[A_2^1] + (q-1)g'(x)[A_2^n]\Big)\Big(pf(x)[B_2^1] + qg(x)[B_2^n]\Big) \\
&\quad - \Big(pf'(x)[B_2^1] + qg'(x)[B_2^n]\Big)\Big((p+1)f(x)[A_2^1] + (q-1)g(x)[A_2^n]\Big) \\
&= \alpha_2(f'(x)g(x) - f(x)g'(x)) \geq 0
\end{aligned}$$

por 4.6 y 4.7. Siguiendo un proceso análogo y desarrollando los primeros factores de J_3 en 4.5 por la fila que contiene a las $1-F$ y los segundos factores por la fila que contiene a las f:

$$\begin{aligned}
J_3 &= \Big((p+1)(1-F(x))[A_3^1] + (q-1)(1-G(x))[A_3^n]\Big)\Big(pf(x)[B_3^1] + qg(x)[B_3^2]\Big) \\
&\quad - \Big(p(1-F(x))[B_3^1] + q(1-G(x))[B_3^2]\Big)\Big((p+1)f(x)[A_3^1] + (q-1)g(x)[A_3^2]\Big) \\
&= \alpha_3((1-F(x))g(x) - f(x)(1-G(x))) \geq 0.
\end{aligned}$$

por 4.6 y 4.9.

Por lo tanto, $\varphi_p'(x) \geq 0,\ \forall x \in \mathbb{R}$, luego $\varphi_p(x)$ es creciente en $\mathbb{R}$. Esto demuestra el caso 1.

Caso 2: f y g no verifican las hipótesis del caso 1. Sean $Z_1,\dots,Z_n$ v.a.i.i.d. con distribución $N(0,\varepsilon^2)$ para un $\varepsilon > 0$. Vamos a denotar H_ε a esta distribución y vamos a definir

$$\overline{X}_i = X_i + Z_i,\ i = 1,\dots,n.$$

Denotemos $\overline{X}_{(1)}(p,q) \leq \overline{X}_{(2)}(p,q) \leq \dots \leq \overline{X}_{(n)}(p,q)$ a los estadísticos ordenados de estas nuevas variables aleatorias. Como $X_1,\dots,X_p$ tienen distribución F, $\overline{X}_1,\dots,\overline{X}_p$ tienen distribución $F * H_\varepsilon$, que viene dada por:

$$F * H_\varepsilon(x) = P(X+Z \leq x) = \int_{\mathbb{R}} P(X \leq x-u) f_\varepsilon(u)du = \frac{1}{\varepsilon\sqrt{2\pi}} \int_{\mathbb{R}} F(x-u) e^{-\frac{u^2}{2\varepsilon^2}},\ \forall x \in \mathbb{R}.$$

Si hacemos el cambio de variable $x-u=v$, obtenemos

$$F * H_\varepsilon(x) = \frac{1}{\varepsilon\sqrt{2\pi}} \int_{\mathbb{R}} F(v) e^{-\frac{(x-v)^2}{2\varepsilon^2}}. \tag{4.10}$$

Notemos que esta expresión es infinitamente derivable en x, por serlo la densidad normal, aplicando el teorema de derivación bajo el signo integral.

Análogamente como $X_{p+1},\dots,X_n$ tienen distribución G, $\overline{X}_{p+1},\dots,\overline{X}_n$ tienen distribución $G*H_\varepsilon$, cuya distribución es la misma de la expresión anterior cambiando, en el último término, F por G, por lo que haciendo el mismo cambio de variable, vemos que también es infinitamente derivable. Como las Z_i tienen densidades log-cóncavas, aplicando el teorema 15, como $G \leq_{lr} F$ por hipótesis, se tiene que

$$G * H_\varepsilon \leq_{lr} F * H_\varepsilon.$$

Estas dos últimas distribuciones tienen densidades derivables. Lo podemos ver partiendo de 4.10, que hemos visto que es infinitamente derivable. Por lo tanto, podemos aplicar el caso anterior, por el que $\overline{X}_{(k)}(p,q) \leq_{lr} \overline{X}_{(k)}(p+1,q-1)$ para cualesquiera $k = 1,\dots,n$ y $p = 0,\dots,n-1$.

Veamos que, si $\varepsilon \to 0$, $\overline{X}_{(k)}(p,q)$ converge débilmente a $X_{(k)}(p,q)$ para cualesquiera k y p: Sabemos que la función característica de la convolución es el producto de las funciones características de las distribuciones implicadas. Utilizando esto y la función característica de Z, que es $e^{-\frac{\varepsilon^2 t^2}{2}}$ y converge a 1 cuando $\varepsilon \to 0$, tenemos que la función característica de las $\overline{X}_i$, que es

$$\varphi_{\overline{X}_i}(t) = Ee^{it(X_i+Z_i)} = \varphi_{X_i}(t) Ee^{itZ_i} = \varphi_{X_i}(t) e^{-\frac{\varepsilon^2 t^2}{2}},$$

converge a la de las X_i, por lo que las $\overline{X}_i$ convergen en distribución a las X_i. Ahora, utilizando el teorema 5, tenemos que los estadísticos ordenados de las $\overline{X}_i$ convergen en distribución a los estadísticos ordenados de las X_i.

Como el orden de razón de verosimilitudes es cerrado con respecto a la convergencia débil (como vimos en el capítulo 3), tenemos que

$$X_{(k)}(p,q) \leq_{lr} X_{(k)}(p+1,q-1),$$

lo que demuestra el caso 2 y el teorema. □

Bibliografía

[1] ARNOLD, B. C., BALAKRISHNAN, N. Y NAGARAJA, H. N., *A First Course in Order Statistics* (1992), editorial Wiley, Nueva York.

[2] BALAKRISHNAN, N., *Permanents, Order Statistics, Outliers, and Robustness*, *Revista Matemática Complutense*, volumen 20 (2007), páginas 7-107, artículo completo en `https://dmle.icmat.es/pdf/REVISTAMATEMATICACOMPLUTENSE_2007_20_01_01.pdf`.

[3] BAPAT, R. B. Y RAGHAVAN, T. E. S., *Nonnegative Matrices and Applications* (1997), editorial Cambridge University Press, Nueva York.

[4] DING, W., ZHANG, Y. Y ZHAO, P., *Comparisons of k-out-of-n systems with heterogeneous components*, revista *Statistics and Probability Letters*, volumen 83 (2013), páginas 493-502.

[5] GONZÁLEZ-MENESES, J., *Apuntes de Álgebra lineal*, Universidad de Sevilla, curso 2008/09, disponibles en `http://matematicas.unex.es/~navarro/algebralineal/meneses.pdf`.

[6] HARDY, G. H., LITTLEWOOD, J. E. Y PÓLYA, G., *Inequalities* (1952), editorial Cambridge University Press, Cambridge.

[7] KARLIN, S. Y RUBIN, H., *The theory of decision procedures for distributions with monotone likelihood ratio*, revista *The Annals of Mathematical Statistics*, volumen 27 (1956), páginas 272-299.

[8] MARSHALL, A. W. Y OLKIN, I., *Life distributions* (2007), editorial Springer, Nueva York.

[9] MINC, H., *Permanents* (1978), editorial Addison-Wesley, Massachusetts.

[10] MÜLLER, A. Y STOYAN, D., *Comparison Methods for Stochastic Models and Risks* (2002), editorial Wiley, Chichester.

[11] Presentación del grupo de trabajo *Ordenaciones estocásticas y sus aplicaciones*, disponible en `http://oea.seio.es/`.

[12] SHAKED, M. Y SHANTHIKUMAR, J. G., *Stochastic Orders* (2007), editorial Springer, Nueva York.

[13] YANG, J. Y ZHUANG, W., *A note on likelihood ratio ordering of order statistics from two samples*, revista *Statistics and Probability Letters*, volumen 84 (2014), páginas 135-139.