Modelos de Probabilidad e Inferencia Estadística

Julia de Haro García
José Luis Iranzo Acosta

PIRÁMIDE

Modelos de Probabilidad e Inferencia Estadística

JULIA DE HARO GARCÍA
PROFESORA TITULAR DE UNIVERSIDAD. UNIVERSIDAD DE MÁLAGA
JOSÉ LUIS IRANZO ACOSTA
PROFESOR ASOCIADO. UNIVERSIDAD DE MÁLAGA

Modelos de Probabilidad e Inferencia Estadística

EDICIONES PIRÁMIDE

COLECCIÓN «ECONOMÍA Y EMPRESA»

Director:
Miguel Santesmases Mestre
Catedrático de la Universidad de Alcalá

Edición en versión digital

Para cualquier información pueden dirigirse a piramide_legal@anaya.es
Juan Ignacio Luca de Tena, 15. 28027 Madrid
Teléfono: 91 393 89 89
www.edicionespiramide.es
ISBN digital: 978-84-368-4531-0

Índice

Prólogo

El manual que a continuación presentamos trata básicamente dos aspectos de los cuales se ha de nutrir la Estadística llamada Teórica o Inferencial: por un lado, los modelos de probabilidad que recogen cómo determinados fenómenos pueden ser analizados basándose en el comportamiento de los mismos dando lugar a las distribuciones o modelos de probabilidad —sirva de ejemplo, la tan aplicada y socorrida distribución normal— y, por otro lado, la parte inferencial propiamente dicha, la cual extrapola lo ocurrido en una muestra al resto de la población.

El profesorado que ha realizado este libro, tras numerosos años impartiendo asignaturas relacionadas con el contenido de este manual, ha querido plasmar todo el material que a lo largo de estos años ha servido para formar a las distintas promociones en titulaciones relacionadas con el ámbito económico y financiero de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Málaga. Pero, además, sin perder en ningún momento el afán simplificador y pretendiendo ir un poco más allá de los contenidos habituales de los programas oficiales universitarios, hemos considerado oportuno incorporar «otros contenidos» que vienen a satisfacer las dudas y el interés de aquel alumnado que desea profundizar un poco más en este campo científico. Prueba de ello han sido las innumerables conversaciones que la autora y el autor han tenido con estudiantes en este sentido. Igualmente, es un manual que puede ser empleado por docentes, investigadores e incluso por todo aquel que quiera iniciarse en la Inferencia Estadística.

Dado que estas materias son de difícil comprensión, se han desarrollado los diferentes conceptos en un lenguaje fácil de entender, evitando en lo posible demostraciones que aparten al lector de las ideas básicas que se desarrollan en cada una de las unidades. Siguiendo con esta finalidad didáctica, aparecen al final de cada tema ejercicios resueltos que permiten una mejor comprensión de las definiciones, procedimientos y técnicas estadísticas analizadas.

Como ya se ha comentado, este manual se puede estructurar en dos grandes bloques. Los capítulos 1 y 2 se enmarcan dentro del Cálculo de Probabilidades. En ellos se analizan el concepto de variable aleatoria, las funciones de densidad y distribución y las principales características de una variable aleatoria (unidimensional y bidimensional), así como se muestran los principales modelos de probabilidad que son objeto de interés en el campo de la Economía, Empresa y Finanzas. El resto de los capítulos se destinan a proporcionar unas nociones de Inferencia Estadística: estudio de las distribuciones muestrales (capítulo 3), propiedades de los estimadores y métodos de estimación por puntos (capítulo 4), estimación por intervalos de confianza (capítulo 5) y la verificación de hipótesis basada en poblaciones normales (capítulo 6). Hemos de señalar que el enfoque adoptado en los capítulos de Inferencia Estadística es de tipo paramétrico y clásico.

Por último, damos las gracias a nuestros familiares que han soportado estoicamente la elaboración de este libro durante la situación de pandemia y posterior proceso de desescalada, así como, a los compañeros y las compañeras del Departamento de Economía Aplicada (Estadística y Econometría) de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Málaga, por las sugerencias y comentarios realizados. Igualmente, agradecer a Ediciones Pirámide la oportunidad que nos ha brindado, haciendo una especial mención a nuestra editora Inmaculada por su paciencia y gran profesionalidad que hacemos extensible al resto del equipo.

Málaga, marzo de 2022.

JULIA DE HARO GARCÍA
JOSÉ LUIS IRANZO ACOSTA

1 Variables aleatorias

1.1. INTRODUCCIÓN

La estadística que se estudia en la actualidad es el fruto de la unión de dos disciplinas que evolucionaron de forma independiente hasta el siglo XIX: por un lado, el Cálculo de Probabilidades, que surge en el siglo XVII a partir de los juegos de azar, y por otro lado, la Estadística o ciencia del Estado (del latín *status*). Desde esta óptica se puede entender la Estadística «*como la exteriorización*[1] *cuantitativa de las cosas del Estado*». La unión de estas dos líneas da lugar a una ciencia que se ocupa de cómo obtener conclusiones de la investigación empírica por medio de modelos matemáticos.

Sin embargo, se puede afirmar que gracias a la vinculación con el Cálculo de Probabilidades, la Estadística ha pasado de ser una ciencia que no solamente describe la realidad, sino que también la modeliza. A partir de lo anterior se puede afirmar que esta disciplina está formada por tres partes que son objeto de estudio en los diversos grados[2] que se imparten en las facultades de Ciencias Económicas y Empresariales de Europa:

- **Estadística Descriptiva.** Se ocupa de la obtención, clasificación, representación y análisis de los datos relativos a las características de interés observados en todos los individuos de un colectivo con el objeto exclusivo de describir las regularidades o el comportamiento de ese colectivo, sin pretender extraer conclusiones que trasciendan del conjunto

[1] Los ejemplos del origen histórico de esta vertiente de la estadística se pueden encontrar en el Pentateuco, dentro del libro denominado de los Números, donde ya se habla de un censo de personas. Además, existen evidencias históricas que proporcionan informaciones sobre censos en los imperios de la antigüedad (China, Egipto, Mesopotamia, Grecia o Roma).

[2] Los grados que incluyen asignaturas en los que se desarrollan contenidos de tipo estadístico son, entre otros: Administración y Dirección de Empresas, Economía, Estadística y Empresa, Finanzas y Contabilidad, *Marketing* e Investigación de Mercados y Gestión y Administración Pública.

analizado. Este tipo de estudio es de naturaleza deductiva (no se tratarán en este libro al ser contenidos propios de cursos introductorios).

- **Teoría de la Probabilidad.** Estudia los fenómenos aleatorios. Se trata de una teoría de base axiomática cuyos teoremas se obtienen mediante un razonamiento lógico-deductivo.
- **Inferencia Estadística**. Esta rama se nutre de las dos anteriores y su objetivo consiste en obtener conclusiones de la población a través de la información muestral. Para ello se servirá de los instrumentos que le proporciona el cálculo de probabilidades mediante la observación y análisis de una muestra representativa de la población. Se realiza, por tanto, un proceso inductivo (Sarrión, Benítez, Iranzo e Isla, 2012).

1.2. CONCEPTO Y TIPOS DE VARIABLES ALEATORIAS

Una descripción completa de la probabilidad exige conocer tres elementos básicos:

- El espacio muestral, E, que representa todos los resultados posibles del experimento aleatorio (e.a.).
- El álgebra, $\mathcal{A}$, formada por todos o parte de los subconjuntos (sucesos) de E.
- La función de conjunto, P, para la medida de probabilidad.

La terna $(E, \mathcal{A}, P)$ se denomina espacio probabilístico, cuya obtención depende de la información disponible sobre el experimento aleatorio y de la naturaleza de los problemas que pretendamos resolver. Centrándonos en el tercer componente, P, se define como una función de conjunto cuyo dominio es la familia $\mathcal{A}$ y cuyo recorrido es un subconjunto de $\mathbb{R}$ (conjunto de números reales), concretamente, es el subconjunto formado por todos los puntos del intervalo cerrado $[0, 1]$ siempre que cumpla con los axiomas básicos de no negatividad, certeza y aditividad.

En este punto vamos a definir qué es una variable aleatoria (v.a.): concepto clave en el cálculo de probabilidades y que nos permitirá modelizar los resultados de un e.a. en el conjunto de los números reales.

Con frecuencia, los resultados que generan los espacios muestrales pueden ser difíciles de manejar debido a que tales elementos no son números, sino resultados cualitativos. Trabajar con resultados numéricos permite hacer uso de desarrollos matemáticos para aplicarlos adecuadamente, haciendo necesario pasar del espacio muestral original, E, cuyos resultados puede que no sean numéricos, a un nuevo espacio muestral inducido, *E'*, cuyos resultados van a ser números. Esto se consigue mediante la función denominada variable aleatoria[3], *X*, función que asigna un número real, y solo uno, $X(e) = x$, a cada resultado (e) del espacio *E*; el dominio de dicha función es *E* y su recorrido es el conjunto $\mathbb{R}$ o un subconjunto de $\mathbb{R}$:

$$X: E \to \mathbb{R}$$
$$e \to X(e)$$

Para aclarar este concepto consideremos el e.a. consistente en lanzar tres monedas de forma que c = *cara* y k = *cruz*, siendo los resultados los que se muestran en la tabla 1.1:

TABLA 1.1

Descripción de los sucesos al lanzar tres monedas al aire

Sucesos	Composición	Sucesos	Composición
e_1	(c, c, c)	e_5	(k, k, c)
e_2	(c, c, k)	e_6	(k, c, k)
e_3	(c, k, c)	e_7	(c, k, k)
e_4	(k, c, c)	e_8	(k, k, k)

Si ahora contamos el número de caras, en cada lanzamiento, tenemos un nuevo espacio muestral, *E'*, representado por resultados numéricos (véase figura 1.1).

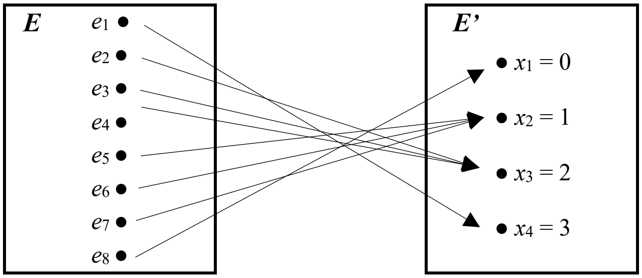

Figura 1.1. Espacio muestral original, *E*, e inducido, *E'*.

[3] El término v. a. puede ser algo confuso tal y como se ha definido, porque *X* no es propiamente dicha una variable, sino una función y, tampoco es aleatoria, porque en su definición no se hace referencia a las probabilidades de los eventos. Para Barbancho (1992b, p. 54) «*estas contradicciones no deben causar extrañeza pues proceden de la evolución del pensamiento científico que se ha servido de viejos términos, acuñados por la historia, para nuevas concepciones en el modo de explicar y estudiar los fenómenos aleatorios*».

El cómputo del número de caras, X, proporciona cuatro resultados numéricos diferentes, los cuales definen el espacio muestral inducido $E' = \{x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3\}$. En la figura anterior se observa la relación entre los resultados de E y E': a cada elemento e_i de E le corresponde uno y solo un x_j de E' (el recíproco no es cierto).

En general el número de elementos que conforman los dos espacios muestrales son distintos, solamente en el caso de que E conste de resultados numéricos, y son estos resultados los que nos interesan, se verifica la coincidencia entre E y E'.

Una variable aleatoria X se puede clasificar en tres tipos:

- **Discreta.** Variable que solo toma un número finito de valores reales o un número infinito, pero numerable. La designaremos por la letra X y sus valores concretos por x, cuando se trata del caso finito de k valores distintos, es decir, $x_1, x_2, ..., x_k$; o bien, por x_i $\forall$ $i = 1, 2, \ldots$ para un número infinito pero numerable de valores diferentes.
- **Continua.** Es aquella que toma un número infinito no numerable de valores distintos pudiendo ser toda la recta real $\mathbb{R}$ o un subconjunto o intervalo de $\mathbb{R}$.
- **Mixta.** Es una variable combinación o mixtura de una v.a. discreta y otra continua.

TABLA 1.2

Algunos ejemplos de variables aleatorias

Tipo de variables aleatorias		**Ejemplos**
Discreta	Número finito de valores	Número de asignaturas suspensas de un alumno que se ha matriculado en 5, número de opositores que superan una oposición,…
	Número infinito de valores pero numerable	Número de aviones que despegan en un aeropuerto al año, número de coches que pasan por un peaje a cierta hora,…
Continua	Toma todos los posibles valores reales, $\mathbb{R}$, o un subconjunto del mismo	Errores de medida de una báscula, resultados obtenidos en un año determinado (beneficios o pérdidas) de las pymes de la Unión Europea, tiempo de espera para ser atendido en una oficina bancaria en horario comercial,…

Es importante conocer el tipo de variable aleatoria que representa el fenómeno aleatorio ya que tendrá implicaciones en su modelización matemática.

1.3. FUNCIÓN DE DENSIDAD

Hemos visto que una v.a. X supone la transformación del espacio muestral E a E', pues bien, cuando relacionamos cada uno de los eventos elementales de E' con sus respectivas probabilidades obtenemos la función de densidad o de probabilidad de X. Esta función se definirá tanto para v.a. discreta como para v.a. continua.

1.3.1. Función de densidad o de probabilidad para el caso discreto

Dada una v.a. X discreta (v.a.d.), se define su función de densidad o cuantía, $f(x)$, como aquella función que a cada valor de x le asigna su probabilidad de ocurrencia

$$P(X = x_i) = f(x_i) = p_i, \qquad \text{para } i = 1,\ 2, \dots$$

Esta función de probabilidad o de densidad cumple las siguientes propiedades:

- Se encuentra acotada entre $0 \leq f(x) \leq 1, \forall\, x \in \mathbb{R}$.
- La suma de probabilidades asociadas a todos los valores de la v.a. X es la unidad

$$\sum_{x_i} f(x_i) = 1$$

Esta función se puede presentar mediante una tabla con dos columnas que recogen los pares de valores $(x_i, f(x_i))$, $\forall\, i = 1, \dots, k$ o proporcionando la función matemática que los genera (tanto para el caso finito como infinito numerable). La representación gráfica de $f(x)$ se hace mediante un diagrama de barras.

1.3.2. Función de densidad para el caso continuo

En el caso de que la variable aleatoria X sea continua (v.a.c.) se define su función de densidad, $f(x)$, como aquella función tal que para cualquier par de valores reales a y b cumple

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \ \ a \text{ y } b \in \mathbb{R} \text{ tal que } a \leq b$$

La función de densidad de X, siendo continua, cumple las siguientes propiedades:

- No negatividad, $f(x) \geq 0, \forall\, x \in \mathbb{R}$ (solamente está acotada inferiormente).
- $\int_{-\infty}^{\infty} f(x)dx = 1$.
- Para cualquier subintervalo $B \in \mathbb{R}$, se cumple $P(X \in B) = \int_{X \in B} f(x)dx$.

Si la v.a. X es discreta se comprueba que la función de densidad, $f(x)$, proporciona la probabilidad correspondiente a cada valor de X, x_i. Sin embargo, **en el caso continuo, se reparte una cantidad finita de masa (la unidad) entre una infinidad no numerable de valores, y ocurre que, para todo número real x del intervalo de definición de la variable se verifica $P(X = x) = 0$.** Luego para calcular[4] probabilidades distintas de 0 hay que referirse necesariamente a intervalos y como X es continua la expresión gráfica $f(x)$ será generalmente una curva que encierra entre la misma y el eje de abscisas un área de valor 1. La probabilidad de que X esté en el intervalo (a, b) será el área bajo la curva correspondiente a dicho intervalo y se calcula con una integral.

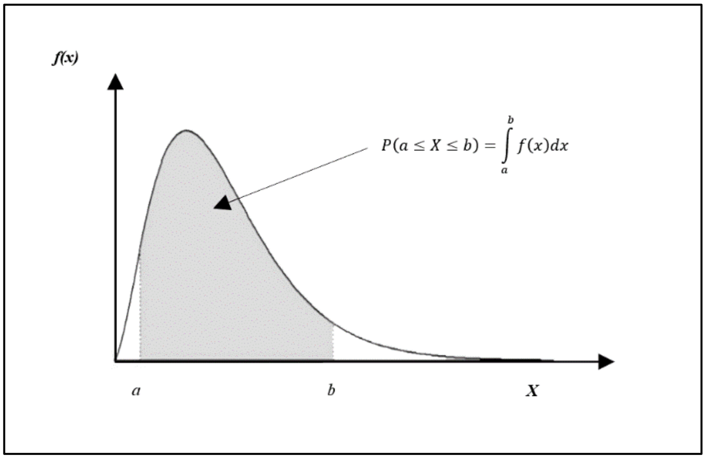

Figura 1.2. Representación gráfica de $f(x)$ y cálculo de probabilidades.

[4] En el caso continuo, cada ordenada de $f(x)$ no da la probabilidad de que X tome el valor x, sino la densidad de masa en el punto x (la probabilidad en un punto x siempre es nula).

Por tanto, cuando la v.a. X es continua y disponemos de su función de densidad, $f(x)$, el cálculo de probabilidades se efectúa mediante la integral de la función de densidad correspondiente en el intervalo deseado, verificándose que:

$$P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x)dx$$

1.4. FUNCIÓN DE DISTRIBUCIÓN

La función de distribución, $F(x)$, proporciona la probabilidad de que un valor observado de la v.a. X, sea menor o igual que el número real x.

1.4.1. Función de distribución para el caso discreto

En el caso de que la v.a. X sea discreta la función de distribución se obtiene como suma acumulativa empleando la función de densidad:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Las propiedades que cumple $F(x)$ son las siguientes:

- Función monótona no decreciente: si $a \leq b \ \ \forall\, a, b \in \mathbb{R}$ entonces $F(a) \leq F(b)$.
- La función, al ser una probabilidad, cumple $0 \leq F(x) \leq 1, \forall\, x \in \mathbb{R}$.
- Los valores extremos son:

$$\lim_{x \to -\infty} F(x) = 0 \text{ y } \lim_{x \to \infty} F(x) = 1$$

- La función $f(x)$ se puede obtener por medio de $F(x)$ de la siguiente forma:

$$f(x_i) = P(X = x_i) = F(x_i) - F(x_{i-1})$$

- El cálculo de probabilidades en el intervalo definido por a y b, siendo $a \leq b$, dependerá de la definición de los extremos del mismo (tabla 1.3):

TABLA 1.3

Cálculo de probabilidades empleando la función de distribución

Definición del intervalo		Cálculo de probabilidades
Extremos cerrados	$[a,b] \to a \leq X \leq b$	$P(a \leq X \leq b) = F(b) - F(a) + f(a)$
Extremos semiabiertos	$(a,b] \to a < X \leq b$	$P(a < X \leq b) = F(b) - F(a)$
	$[a,b) \to a \leq X < b$	$P(a \leq X < b) = F(b) - f(b) - F(a) + f(a)$
	$(a,b) \to a < X < b$	$P(a < X < b) = F(b) - f(b) - F(a)$

- Es una función continua por la derecha y discontinua por la izquierda de forma que $F(x)$ experimenta en el punto x un salto de cuantía $P(X=x)$. La representación gráfica de $F(x)$ es un diagrama escalonado.

1.4.2. Función de distribución para el caso continuo

La función de distribución $F(x)$ siendo X continua se define como

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt\,;\ \text{siendo } t \text{ una variable real auxiliar}$$

La función cumple básicamente las mismas propiedades que si la v.a. X fuera discreta, aunque hay que tener en cuenta que:

- Se trata de una función continua tanto por la derecha como por la izquierda.
- La función de densidad se obtiene derivando en $F(x)$ con respecto a x (punto en donde la función es continua)

$$f(x) = \frac{dF(x)}{dx}$$

1.5. CARACTERÍSTICAS DE LA VARIABLE ALEATORIA

La descripción de una v.a. X se hace fijándonos en sus características más importantes. Existen muchas características que pueden considerarse, aunque la práctica estadística ha demostrado que las que pueden aportar mayor información son las siguientes:

TÍTULOS RELACIONADOS

Análisis cuantitativo de la actividad turística, *J. Alegre Martín, M. Cladera Munar, C. N. Juaneda Sampol.*

Análisis multivariante de datos. Cómo buscar patrones de comportamiento en BIG DATA, *G. Mateos-Aparicio Morales y A. Hernández Estrada.*

Árboles de decisión y ELECTRA I, *M. Ruiz Rodríguez, S. Martínez Fierro, J. M.ª Biedma Ferrer y A. Martín Navarro.*

Curso elemental de Estadística Descriptiva, *A. Hernández Bastida.*

Econometría. Ejercicios resueltos, *R. M.ª García Fernández, J. M. Herrerías Velasco y F. Palacios González.*

Econometría. *M. Díaz Fernández y M.ª del M. Llorente Marrón.*

Ejercicios de econometría I, *F. Palacios González (coord.), R. M.ª García Fernández y J. M. Herrerías Velasco.*

Ejercicios de estadística descriptiva y probabilidad para economía y administración de empresas. *J. M. Casas Sánchez, C. García Pérez, L. F. Rivera Galicia y A. I. Zamora Sanz.*

Ejercicios de inferencia estadística y muestreo para economía y administración de empresas. *J. M. Casas Sánchez, C. García Pérez, L. F. Rivera Galicia y A. I. Zamora Sanz.*

Estadística. Problemas resueltos, *M.ª J. Peralta Astudillo, A. Rúa Vieytes, R. Redondo Palomo y C. del Campo Campos.*

Introducción a la econometría. *F. J. Trívez Bielsa.*

Introducción a la inferencia estadística, *J. L. Iranzo Acosta y J. de Haro García.*

Introducción a la optimización de decisiones, *J. Niño Mora.*

Introducción a las matemáticas financieras (Manual + Guía del alumno), *S. Cruz Rambaud y M.ª del C. Valls Martínez.*

Introducción a las matemáticas financieras. Problemas resueltos, *M.ª del C. Valls Martínez y S. Cruz Rambaud.*

Introducción a las técnicas de muestreo, *J. Boza Chirino, J. V. Pérez-Rodríguez y J. de León Ledesma.*

Matemáticas de las operaciones financieras, *E. Navarro Arribas.*

Matemática de los seguros de vida, *R. Moreno Ruiz, O. Gómez Pérez-Cacho, E. Trigo Martínez.*

Matemáticas para la economía y la empresa, *S. Calderón Montero y M. L. Rey Borrego.*

Matemáticas para el éxito empresarial, *E. M. Fedriani Martel y M.ª del C. Melgar Hiraldo.*

Microeconometría y decisión, *B. Cabrer Borrás, A. Sancho Pérez y G. Serrano Domingo.*

Modelos de Probabilidad e Inferencia Estadística, *J. de Haro García y J. L. Iranzo Acosta.*

Prácticas de estadística con R, *J. M.ª Sarabia, F. Prieto y V. Jordá.*

Predicción y simulación aplicada a la economía y gestión de empresas, *A. Pulido San Román y A. M.ª López García.*

Problemas resueltos de estadística, *S. Zubelzu y A. Ercoreca.*

Problemas resueltos de estadística para las ciencias sociales, *J. M.ª Sarabia, C. Trueba, L. Remuzgo, V. Jordá y F. Prieto.*

Sistemas y tecnologías de la información en las organizaciones. *M. Fugini, P. Maggiolini, D. Pagani y R. Salvador Vallés.*