

Fundamentos de Inferencia Estadística

Alex Sanchez Pla y Santiago Pérez Hoyos

Table of contents

Presentación	6
Objetivo	6
Prerequisitos y organización del material	6
Referencias	7
Agradecimiento y fuentes utilizadas	8
El proyecto Statmedia	8
Otros materiales utilizados	9
1 Probabilidad y Experimentos aleatorios	10
1.1 Introducción	10
1.1.1 Fenómenos deterministas y fenómenos aleatorios	10
1.1.2 Sucesos	10
1.2 Función de probabilidad	11
1.2.1 ¿Diferentes funciones de probabilidad para una misma experiencia aleatoria?	11
1.3 ¿Cómo se calculan las probabilidades?	12
1.4 Propiedades inmediatas de la probabilidad	12
1.4.1 Suceso nulo	13
1.4.2 Suceso implicado	13
1.4.3 Complementario de un suceso	13
1.4.4 Ocurrencia de algun suceso	13
1.4.5 Propiedad de que ocurra algun suceso	13
1.4.6 Probabilidad de que ocurran dos (o más) sucesos a la vez	13
1.4.7 Una formula que los relaciona	13
1.5 Probabilidad condicionada	14
1.5.1 Sucesos dependientes y sucesos independientes	15
1.5.2 Incompatibilidad e independencia	15
1.6 Dos Teoremas importantes	15
1.6.1 Teorema de las probabilidades totales	15
1.6.2 Teorema de Bayes	16
1.7 Introducción a los experimentos múltiples	16
1.8 Combinatoria	17
1.8.1 Permutaciones	17
1.8.2 Variaciones	17
1.8.3 Variaciones con repetición	18
1.8.4 Combinaciones	18
1.8.5 Permutaciones con repetición	19
1.9 Frecuencia relativa y probabilidad	19
2 Variables aleatorias y Distribuciones de probabilidad	21
2.1 El espacio muestral y sus elementos	21
2.2 Representación numérica de los sucesos elementales. Variables aleatorias	21

2.3	Caracterización de una variable aleatoria a través de la probabilidad. Función de distribución	22
2.4	Propiedades de la función de distribución	22
2.5	Clasificación de las variables aleatorias	24
2.5.1	Variables aleatorias discretas	24
2.5.2	Variables aleatorias continuas	24
2.6	Variable aleatoria discretas	25
2.6.1	Caracterización a través de la función de densidad o de probabilidad	26
2.6.2	Propiedades de la función de densidad discreta	27
2.6.3	Relaciones entre la función de distribución y la función de densidad discreta. Probabilidad de intervalos.	27
2.7	Variables aleatorias continuas	28
2.7.1	Ejemplos	28
2.7.2	Función de densidad continua	29
2.7.3	Relaciones entre la función de distribución y la función de densidad.	30
2.8	Independencia de variables aleatorias	32
2.8.1	Caracterización de la independencia	32
2.9	Caracterización de una variable aleatoria a través de parámetros	32
2.10	Esperanza de una variable aleatoria discreta	33
2.11	Esperanza de una variable aleatoria continua	33
2.12	Propiedades de la esperanza matemática	34
2.12.1	Linealidad de la esperanza matemática	34
2.12.2	Esperanza del producto	34
2.13	Varianza de una variable aleatoria	34
2.13.1	Propiedades de la varianza	35
2.14	Momentos (de orden k) de una variable aleatoria	35
2.15	Definición formal de variable aleatoria	36
3	Distribuciones Notables	38
3.1	Distribuciones discretas	38
3.1.1	La distribución de Bernoulli	38
3.1.2	La distribución Binomial	40
3.1.3	La distribución de Poisson	44
3.1.4	La distribución Multinomial	45
3.1.5	La distribución Hipergeométrica	48
3.1.6	La distribución Geométrica o de Pascal	49
3.1.7	La distribución Binomial negativa	50
3.1.8	Tabla resumen de las distribuciones discretas principales	50
3.2	DISTRIBUCIONES CONTINUAS	51
3.2.1	La distribución Uniforme	51
3.2.2	La distribución Exponencial	53
3.2.3	La distribución Normal	54
3.2.4	La distribución Gamma	56
3.2.5	La distribución de Cauchy	57
3.2.6	La distribución de Weibull	58
3.2.7	Tabla resumen de las principales distribuciones continuas	59
3.3	LA FAMILIA EXPONENCIAL DE DISTRIBUCIONES	59
4	Grandes muestras	60
4.1	Introducción: Aproximaciones asintóticas	60
4.1.1	Convergencia de variables aleatorias	60

4.2	Leyes de los grandes números	60
4.3	El teorema central del límite	60
4.3.1	Sumas de variables aleatorias	60
4.3.2	Definición de convergencia en ley	61
4.3.3	Enunciado del teorema central del límite	62
4.3.4	Aplicación del TCL a los ejemplos	62
4.3.5	Casos particulares más notables	63
4.3.6	Interpretación del teorema central del límite	63
4.3.7	Aproximaciones y errores numéricos	64
4.3.8	Acerca de las variables aproximadamente normales	64
5	Distribuciones de probabilidad multidimensionales	67
5.1	Variables aleatorias multidimensionales.	67
5.2	Distribuciones conjuntas, marginales y condicionales,.	67
5.3	Valores esperados, covariancia y correlación.	67
5.4	Independencia de variables aleatorias	67
5.5	Distribuciones multivariantes: multinomial y normal bivalente.	67
6	Introducción a la inferencia estadística	68
6.1	Los problemas de la inferencia estadística.	68
6.2	Muestreo y distribuciones en el muestreo.	68
6.3	La verosimilitud y su papel en la inferencia estadística	68
6.4	El problema de la estimación. Tipos de estimadores.	68
6.5	Métodos de obtención de estimadores. Estimadores máximo verosímiles y estimadores bayesianos.	68
6.6	Propiedades de los estimadores.	68
7	Estimación por intervalos	69
7.1	Preliminares: estimación del error estándar e Introducción al bootstrap	69
7.2	Estimadores por intervalo: intervalos de confianza	69
7.3	Intervalos de confianza para características de una población normal (media, varianza),	69
7.4	Intervalos de confianza bootstrap.	69
7.5	Intervalos de confianza para proporciones binomiales	69
7.6	Intervalos de confianza para parámetros en muestra grandes y para casos generales (tasas, OR, ...)	69
7.7	Aplicaciones: cálculo del tamaño muestral	69
8	Pruebas de hipótesis	70
8.1	Conceptos básicos: pruebas de hipótesis y de significación, pruebas unilaterales y bilaterales, tipos de error, valores críticos de test y p-valores	70
8.2	Potencia de un test. Cálculos de potencia y de tamaño de la muestra. Tamaño del efecto.	70
8.3	Métodos de construcción de tests.	70
8.4	Problemas asociados al uso de tests estadísticos. La crisis de la significación	70
9	Inferencia Aplicada	71
9.1	Pruebas de normalidad.Pruebas gráficas. El test de Shapiro-Wilks	71
9.2	Pruebas de hipótesis para contrastar variables cuantitativas: pruebas paramétricas t-test y Anova	71

9.3	Pruebas de hipótesis para contrastar variables cuantitativas: pruebas de hipótesis no paramétricas de Wilcoxon y Kruskal-Wallis	71
9.4	Contrastes para datos categóricos. Pruebas binomiales, ji cuadrado y test de Fisher.	71
9.5	Riesgo relativo y razón de «odds»	71
10	Computación Intensiva y <i>Multiple Testing</i>	72
10.1	Tests de permutaciones; ¿Qué?, ¿Cuándo?, ¿Cómo?	72
10.2	El bootstrap en contraste de hipótesis	72
10.3	El problema de las comparaciones múltiples	72
10.4	Métodos de control de error: FWER y FDR	72
	Referencias	73

Presentación

Objetivo

El objetivo de estas notas es presentar un material de soporte para la asignatura de “Inferencia Estadística” del [Máster interuniversitario de Bioestadística y Bioinformática](#) impartido conjuntamente por la [Universitat Oberta de Catalunya \(UOC\)](#) y la [Universidad de Barcelona \(UB\)](#).

Esta asignatura adolece de las características habituales de las asignaturas de posgrado, y especialmente de un posgrado de estadística (y bioinformática), que muestran algunas de las cosas que no debe de ser esta asignatura:

- No puede ser un primer curso de estadística, porque se supone que los estudiantes del máster ya lo han cursado en sus grados. Por no decir que, a quien viene a especializarse en estadística se le puede suponer una base mínima.
- Tampoco debe ser como los segundos cursos de estadística de algunos grados, que tratan temas como la regresión, el diseño de experimentos o el análisis multivariante, porque esto ya se trata en diversas asignaturas del máster.

¿Que debemos pues esperar que sea este curso?

- Puestos a pedir, este curso debería servir para repasar y consolidar los conceptos básicos que la mayoría de estudiantes traerán consigo.
- Además, y sobretodo, debe proporcionar una visión general, lo más completa posible dentro de las limitaciones de tiempo, del campo de la inferencia estadística
- Y, naturalmente, esto significa proporcionar aquellos conceptos sobre los que se apoyaran muchas de las restantes asignaturas como “Regresión modelos y métodos”, “Diseño de Experimentos”, “Análisis Multivariante”, “Análisis de la Supervivencia” o “Análisis de datos ómicos”.

Prerequisitos y organización del material

Uno de los problemas “eternos” en el estudio de la estadística ha sido siempre la falta de acuerdo, entre la comunidad de docentes, de cual debería ser el nivel matemático a que se impartan los cursos.

En los cursos de pre-grado ha habido un cierto consenso, y con los años el nivel de formalismo ha disminuido, incluso en estudios de tipo “STEM”, tendiendo a centrarse en la aplicación de los conceptos, por ejemplo usando R, más que en un tratamiento formal (“matemático”) de los mismos.

Aunque esto puede ser práctico para aquellos estudios en los que la estadística és una asignatura de un grado, es también obvio que dicha aproximación no permite profundizar en muchos de los puntos que se tratan.

Es por ello que en este curso seguiremos la indicación habitual en cursos similares de asumir que el estudiante:

- Se siente cómodo con el lenguaje algebraico, desarrollo de expresiones, sumatorios etc.
- Está familiarizado con el cálculo diferencial en una o varias variables, aunque esta familiaridad no será imprescindible para seguir la mayoría de los contenidos del curso.
- Conoce el lenguaje estadístico R, que en muchas ocasiones nos ofrecerá una solución directa a los problemas de cálculo.

Referencias

Los requisitos anteriores corresponden básicamente a las matemáticas del bachillerato. Algunas fuentes adicionales pueden ser:

- [Iniciación a las matemáticas para la ingeniería](#). M. Besalú y Joana Villalonga
 - [Colección de \(100\) videos de soporte a las matemáticas para la ingeniería](#)

Agradecimiento y fuentes utilizadas

Salvo que uno desee escribir un libro sobre algo muy extraño, siempre habrán otros libros o manuales similares al que se está planteando.

La respuesta a la pregunta, “Y entonces, ¿porque hacer un nuevo material?” suele ser más una excusa que una explicación sólida.

Una posible razón puede ser *para ajustarlo al máximo al perfil del curso para al que se destinan dichos materiales*, condición que otros textos, pensados para cursos y audiencias distintas, pueden no satisfacer. En este caso adoptaremos esta explicación y el tiempo decidirá si el objetivo se alcanza.

Dicho esto, debemos agradecer a las distintas fuentes utilizadas, el que hayan puesto a disposición sus materiales para poder reutilizarlos. Entre estos destacamos:

El proyecto Statmedia

Statmedia fue un proyecto para impulsar el aprendizaje de la estadística mediante gadgets interactivos que, sin duda fue pionero en este campo.

El programa contenía unas explicaciones de los contenidos de probabilidad y estadística para un sólido curso de introducción, complementado con: - Una serie de casos para motivar e ilustrar los conceptos introducidos. - Un conjunto de gadgets interactivos con los que interactuar y experimentar y - Ejercicios de respuesta múltiple para verificar los conceptos trabajados.

Este programa, sin embargo, como tantos otros, no sobrevivió al desarrollo tecnológico, y la evolución (o decadencia) del lenguaje Java lo llevo a dejar de ser funcional.

Para estos apuntes hemos recuperado, y en ocasiones adaptado o modificado, algunos de los contenidos de Statmedia, que habían estado escritos con gran pulcritud. Esto se ha hecho siguiendo las indicaciones de la licencia (CC-Share-alike) que permite adaptar contenidos atribuyendolo a sus autores y citando la fuente.

Los gadgets originales ya no son funcionales pero muchos de ellos han sido re-escritos en R como aplicaciones Shiny (disponibles en: [https://github.com/Statmedia](#)) y se enlazaran desde los puntos necesarios del texto.

Dejando aparte (además) de la licencia, vaya nuestro agradecimiento explícito al equipo de profesores del Departamento de Estadística de la Universidad de Barcelona, redactor de la versión inicial del proyecto, que es la que hemos utilizado: Antonio Arcas Pons, Miquel calvo Llorca, Antonio Miñarro Alonso, Sergi Civit Vives y Angel Vilarroya del Campo.

[Antoni Arcas, Antonio Miñarro and Miguel Calvo \(2008\) Statmedia projects in Statistical Education](#)

Otros materiales utilizados

- [Alex Sanchez y Francesc Carmona \(2002\). Apunts d'Estadística Matemàtica](#) Licencia CC0 1.0 Universal
- [Molina Peralta, I. and García-Portugués, E. \(2024\). *A First Course on Statistical Inference*. Version 2.4.1. ISBN 978-84-09-29680-4.](#) Licencia CC BY-NC-ND 4.0
- [Peter K. Dunn \(2024\) *The theory of distributions*.](#) Licencia CC BY-NC-ND 4.0

1 Probabilidad y Experimentos aleatorios

1.1 Introducción

1.1.1 Fenómenos deterministas y fenómenos aleatorios

Supongamos que disponemos de un dado regular con todas las caras pintadas de blanco y con un número, que irá de 1 a 6 sin repetir ninguno, en cada una de las seis caras.

Definamos los dos experimentos siguientes: Experimento 1: Tirar el dado y anotar el color de la cara resultante. Experimento 2: Tirar el dado y anotar el número de la cara resultante. ¿Qué diferencia fundamental observamos entre ambos experimentos? Muy simple! En el experimento 1, el resultado es obvio: saldrá una cara de color blanco. Es decir, es posible predecir el resultado. Se trata de un experimento o fenómeno determinista.

En cambio, en el experimento 2 no podemos predecir cuál será el valor resultante. El resultado puede ser : 1, 2, 3, 4, 5 o 6 . Se trata de un experimento o fenómeno aleatorio.

El conjunto de resultados se anotará con el símbolo: Ω . En este caso, $\Omega = \{1, 2, 3, 4, 5, 6\}$. En los fenómenos aleatorios, al hacer muchas veces la experiencia, la frecuencia relativa de cualquier elemento del conjunto de resultados debe aproximarse siempre hacia un mismo valor.

1.1.2 Sucesos

Supongamos que se ejecuta un experimento aleatorio. Se nos puede ocurrir emitir un enunciado que, una vez realizada la experiencia, pueda decirse si se ha verificado o no se ha verificado. A dichos enunciados los denominamos sucesos.

Por otro lado, los sucesos van asociados a subconjuntos del conjunto de resultados. Cada suceso se corresponderá exactamente con uno, y sólo con un, subconjunto del conjunto de resultados.

Veamos un ejemplo: Experimento: Tirar un dado regular. Conjunto de resultados : $\Omega = \{1, 2, 3, 4, 5, 6\}$ Enunciado: Obtener múltiplo de 3. Subconjunto al que se asocia el enunciado: $A = \{3, 6\}$ Nos referiremos habitualmente al suceso A.

1.1.2.1 Sucesos y conjuntos

Al conjunto de resultados Ω , se le denomina suceso seguro. Al conjunto \emptyset (conjunto sin elementos), se le denomina suceso imposible. Al complementario del conjunto A (A^c), se le denomina suceso contrario o complementario de A. A partir de dos sucesos A y B, podemos formar los sucesos siguientes:

- A intersección B, que anotaremos como:

$$A \cap B$$

- A unión B, que anotaremos como:

$$A \cup B$$

A intersección B, significa que se verifican a la vez A y B. A unión B, significa que se verifica A o B (se pueden verificar a la vez).

1.2 Función de probabilidad

Lógicamente, una vez tenemos un suceso, nos preocupa saber si hay muchas o pocas posibilidades de que al realizar la experiencia se haya verificado.

Por lo tanto, sería interesante el tener alguna función que midiera el grado de confianza a depositar en que se verifique el suceso.

A esta función la denominaremos función de probabilidad. La función de probabilidad será, pues, una aplicación entre el conjunto de resultados y el conjunto de números reales, que asignará a cada suceso la probabilidad de que se verifique.

La notación: $P(A)$ significará: probabilidad de que se verifique el suceso A . Pero claro, de funciones de probabilidad asociadas a priori a una experiencia aleatoria podrían haber muchas.

Lo que se hace para decir qué es y qué no es una función de probabilidad es construir una serie de propiedades (denominadas axiomas) que se exigirán a una función para poder ser catalogada como función de probabilidad.

Y, ¿cuáles son estos axiomas? Pues los siguientes: Sea S el conjunto de sucesos. Axioma 1: Para cualquier suceso A, la probabilidad debe ser mayor o igual que 0 . Axioma 2: $P(\Omega) = 1$ Axioma 3: Para sucesos A_i , de modo que cada par de sucesos no tengan ningún resultado común, se verifica que:

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

De este modo, pueden haber muchas funciones de probabilidad que se podrían asociar con la experiencia.

El problema pasa entonces al investigador para decidir cual o cuales son las funciones de probabilidad más razonables asociadas con la experiencia que está manejando.

1.2.1 ¿Diferentes funciones de probabilidad para una misma experiencia aleatoria?

Supongamos la experiencia de tirar un dado regular. A todo el mundo se le ocurriría pensar que la función de probabilidad se obtiene de contar el número de resultados que contiene el suceso dividido por 6 , que es el número total de resultados posibles. Así pues, la probabilidad de obtener un múltiplo de 3 sería igual a $2/6$, la probabilidad de obtener el número 2 sería $1/6$ i la probabilidad de obtener un número par sería $3/6$. Es decir, parece inmediato construir la función de probabilidad que, además, parece única. A nadie se le ocurre decir, por ejemplo, que la probabilidad de obtener un número par es $5/6$!

En este caso, todo ha sido muy fácil. Hemos visto que existe una única función de probabilidad que encaje de forma lógica con la experiencia y, además, ha sido muy sencillo encontrarla.

Pero esto, por desgracia, no siempre es así. En muchísimas ocasiones resulta muy complejo el decidir cuál es la función de probabilidad.

En el tema de variables aleatorias y de función de distribución se explica el problema de la modelización de muchas situaciones reales.

1.3 ¿Cómo se calculan las probabilidades?

No siempre es fácil conocer los valores de la función de probabilidad de todos los sucesos. Sin embargo, muchas veces se pueden conocer las probabilidades de algunos de estos sucesos. Con la ayuda de ciertas propiedades que se deducen de manera inmediata a partir de la axiomática es posible calcular las probabilidades de más sucesos.

Por otro lado, en caso de que el número de resultados sea finito y de que todos los resultados tengan las mismas posibilidades de verificarse, la probabilidad de un suceso cualquiera se puede calcular a partir de la regla de Laplace:

Si A es un suceso :

$$\text{Probabilidad } (A) = \frac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$$

donde: Número de casos favorables = Número de resultados contenidos en A (cardinal de A)
Número de casos posibles = Número total de resultados posibles (cardinal del conjunto total de resultados)

En este caso, el contar número de resultados, ya sean favorables o posibles, debe hacerse por medio de la combinatoria.

Veamos con unos ejemplos muy sencillos y visuales cómo se obtienen y qué representan los casos posibles y los casos favorables.

También es posible obtener de manera aproximada la probabilidad de un suceso si se puede repetir muchas veces la experiencia: la probabilidad del suceso sería el valor al que tendería la frecuencia relativa del suceso. Podéis consultar más detalles acerca de esta aproximación.

En este caso, la cuestión estriba en poder hacer muchas veces la experiencia en condiciones independientes.

1.4 Propiedades inmediatas de la probabilidad

Veremos a continuación una serie de propiedades que se deducen de manera inmediata de la axiomática de la probabilidad.

1.4.1 Suceso nulo

Probabilidad del suceso conjunto vacío (es decir del suceso que no contiene ningún resultado):

$$P(\emptyset) = 0$$

1.4.2 Suceso implicado

Si A es un suceso que está contenido en B (todos los resultados de A también pertenecen a B), entonces:

$$P(A) \leq P(B)$$

1.4.3 Complementario de un suceso

Sea A^c el suceso formado por todos los elementos de Ω que no pertenecen a A (Suceso complementario de A). La probabilidad de dicho suceso es igual a:

$$P(A^c) = 1 - P(A)$$

1.4.4 Ocurrencia de algun suceso

La probabilidad de la unión de dos sucesos A y B es igual a:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1.4.5 Propiedad de que ocurra algun suceso

Si tenemos una colección de k sucesos, la probabilidad de la unión de dichos sucesos será:

$$P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) + \dots + (-1)^{k+1} \cdot P(A_1 \cap \dots \cap A_k)$$

1.4.6 Probabilidad de que ocurran dos (o más) sucesos a la vez

1.4.7 Una formula que los relaciona

Se verifica que:

$$P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(\bar{A}_i)$$

1.5 Probabilidad condicionada

Imaginemos que en la experiencia de tirar un dado regular supiéramos de antemano que se ha obtenido un número par. Es decir, que se ha verificado el suceso: $B = \text{número par}$.

Pregunta: ¿Cuál es ahora la probabilidad de que se verifique el suceso mayor o igual a cuatro? Lógicamente, el resultado sería : $2/3$. Por lo tanto, la probabilidad del suceso $A = \text{mayor o igual a cuatro}$ se ha modificado. Evidentemente, ha pasado de ser $1/2$ (cuando no tenemos ninguna información previa) a ser $2/3$ (cuando sabemos que se ha verificado el suceso B). ¿Cómo podemos anotar esta última probabilidad ($2/3$) ? Muy sencillo. Anotaremos $P(A/B)$, que se lee como probabilidad de A condicionada a B . Así, en este ejemplo,

$$\begin{aligned}P(A/B) &= 2/3 \\ P(A) &= 1/2\end{aligned}$$

En términos generales, estamos en condiciones de poder definir la probabilidad condicionada, y lo hacemos como:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Podemos ahora visualizar de una manera práctica y divertida el ejemplo anterior. Siguiendo con la notación utilizada, el suceso A será lo que denominamos suceso de obtención, mientras que el suceso B será lo que denominamos suceso condicionado. La pantalla nos proporcionará los casos posibles para el condicionante elegido y los casos favorables, calculando mediante la regla de Laplace la probabilidad del suceso.

- 1) Elegid suceso a estudiar. Desplazad, si procede, las barras de puntos.
- 2) Elegir suceso condicionante. Desplazad, si procede, las barras de puntos.
- 3) Comprobad los sucesos posibles y los favorables.

La probabilidad condicionada se comporta, entonces, como una función de probabilidad. Es decir, verifica los tres axiomas siguientes:

Axioma 1:

$$P(A/B) \geq 0$$

Axioma 2:

$$P(\Omega/B) = 1$$

Axioma 3:

$$P\left(\bigcup_{i=1}^{\infty} A_i/B\right) = \sum_{i=1}^{\infty} P(A_i/B)$$

para sucesos A_i con intersección vacía dos a dos.

1.5.1 Sucesos dependientes y sucesos independientes

Sean A y B dos sucesos con probabilidad mayor que 0 . Evidentemente, si

$$P(A/B) = P(A)$$

B no ha modificado la probabilidad de que suceda A. En este caso diremos que son sucesos independientes.

En caso contrario diremos que son sucesos dependientes. En el ejemplo del apartado anterior, se observa que los sucesos son dependientes puesto que las probabilidades anteriores no coinciden.

Se verifica que independencia de los sucesos A y B es equivalente a decir que la probabilidad de la intersección es igual a producto de probabilidades de los dos sucesos.

Se verifica también que si A y B son independientes: a) El complementario del suceso A y el suceso B son independientes. b) El complementario del suceso A y el complementario del suceso B son independientes. c) El complementario del suceso B y el suceso A son independientes.

1.5.2 Incompatibilidad e independencia

Dos sucesos con intersección vacía se denominan sucesos incompatibles. Esto, ¿qué implica? Pues, que si se verifica uno seguro que no se verifica el otro, ya que no tienen resultados en común. Por lo tanto es el caso extremo de dependencia. Obtenemos en este caso que:

$$P(A/B) = 0$$

y, en consecuencia, si $P(A)$ y $P(B)$ son diferentes de cero, la probabilidad condicionada anterior es diferente de $P(A)$, y así se deduce la dependencia.

La única posibilidad de que se dé incompatibilidad e independencia a la vez, es que alguno de los dos sucesos tenga probabilidad igual a cero.

1.6 Dos Teoremas importantes

1.6.1 Teorema de las probabilidades totales

Sea Ω el conjunto total formado por una partición (colección de sucesos con intersección vacía dos a dos):

$$\Omega = H_1 \cup \dots \cup H_n$$

La probabilidad de cualquier otro suceso A , se puede obtener a partir de las probabilidades de los sucesos de la partición y de las probabilidades de A condicionado a los sucesos de la partición, de la manera siguiente:

$$P(A) = \sum_{i=1}^n P(A/H_i) \cdot P(H_i)$$

Esto es lo que se conoce como teorema de las probabilidades totales.

1.6.2 Teorema de Bayes

Es una consecuencia del teorema de las probabilidades totales. Sea Ω el conjunto total formado por una partición (colección de sucesos con intersección vacía dos a dos).

$$\Omega = H_1 \cup \dots \cup H_n$$

Ahora el interés se centrará en la obtención de la probabilidad de cualquier suceso de la partición condicionada a un suceso A cualquiera.

El resultado será:

$$P(H_i/A) = \frac{P(A/H_i) \cdot P(H_i)}{\sum_{i=1}^n P(A/H_i) \cdot P(H_i)}$$

Esto es conocido como teorema o regla de Bayes.

1.7 Introducción a los experimentos múltiples

Supongamos que tiramos a la vez un dado y una moneda. Tenemos una experiencia múltiple, puesto que la experiencia que se realiza es la composición de dos experiencias (experiencia 1 = tirar un dado regular; experiencia 2 = tirar una moneda regular). ¿Cuál es en este caso el conjunto de resultados? Si Ω_1 es el conjunto de resultados asociado con la experiencia tirar un dado y Ω_2 es el conjunto de resultados asociado con la experiencia tirar una moneda, el conjunto de resultados asociado a la experiencia múltiple será $\Omega_1 \times \Omega_2$.

Es decir, $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ $\Omega_2 = \{\text{cara}, \text{cruz}\}$ $\Omega_1 \times \Omega_2 = \{(1, \text{cara}), (2, \text{cara}), (3, \text{cara}), (4, \text{cara}), (5, \text{cara}), (6, \text{cara}), (1, \text{cruz}), (2, \text{cruz}), (3, \text{cruz}), (4, \text{cruz}), (5, \text{cruz}), (6, \text{cruz})\}$

Si P_1 y P_2 son, respectivamente, las funciones de probabilidad asociadas a las experiencias 1 y 2, ¿es posible calcular probabilidades de la experiencia múltiple a partir de P_1 y P_2 ?

Efectivamente! Pero hemos de distinguir dos situaciones:

- Experiencias independientes: cuando el resultado de una no influya en la otra.
- Experiencias dependientes: cuando el resultado de una influya en la otra.

En nuestro caso se trata de experiencias independientes, puesto que el resultado que se obtenga al tirar el dado no influye sobre el resultado que se obtenga al lanzar la moneda y al revés. ¿Como se calculan, pues, las probabilidades de la experiencia múltiple? Sea un suceso de la experiencia múltiple: $A \times B$.

- Caso de experiencias independientes:

$$P(A \times B) = P_1(A) \times P_2(B)$$

- Caso de experiencias dependientes:

$$P(A \times B) = P_1(A) \times P_2(B/A)$$

Entendemos que existe una P_2 para cada suceso A .

Esto que hemos explicado se puede, lógicamente, generalizar a una experiencia múltiple formada por n experiencias.

1.8 Combinatoria

Veamos algunas fórmulas simples que se utilizan en combinatoria y que nos pueden ayudar a calcular el número de casos posibles o el número de casos favorables.

1.8.1 Permutaciones

Sea un conjunto de n elementos. A las ordenaciones que se pueden hacer con estos n elementos sin repetir ningún elemento y utilizándolos todos se las denomina permutaciones. El número de permutaciones que se pueden realizar coincide con el factorial de n , y su cálculo es:

$$n! = n \cdot (n - 1) \cdot (n - 2) \dots \dots 2 \cdot 1$$

Ejemplo:

¿De cuántas maneras distintas podemos alinear a seis personas en una fila?

Respuesta

De $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$ maneras (permutaciones de 6 elementos).

1.8.2 Variaciones

Sea un conjunto de n elementos. Supongamos que deseamos ordenar r elementos de entre los n . A cada una de estas ordenaciones la denominamos variación. El número de variaciones que se pueden hacer de los n elementos tomados de r en r es:

$$V_n^r = n \cdot (n - 1) \dots \dots (n - r + 1)$$

Ejemplo

En una carrera de velocidad compiten diez atletas. ¿De cuántas maneras distintas podría estar formado el podio? (el podio lo forman el primer, el segundo y el tercer clasificado)

Respuesta

Cada podio posible es una variación de diez elementos tomado de tres en tres. Por tanto, el número posible de podios es:

$$V_{10}^3 = 10 \cdot 9 \cdot 8 = 720$$

1.8.3 Variaciones con repetición

Sea un conjunto de n elementos. Supongamos que se trata de ordenar r elementos que pueden estar repetidos. Cada ordenación es una variación con repetición. El número de variaciones con repetición para un conjunto de n tomados de r en r es :

$$RV_n^r = n^r$$

Ejemplo

En una urna tenemos cinco bolas numeradas del 1 al 5 . Se extraen tres bolas sucesivamente con reposición (devolviendo cada vez la bola a la urna). ¿Cuántos resultados distintos es posible obtener?

Respuesta: Se trata de variaciones con repetición de un conjunto de cinco bolas tomadas de tres en tres. En total tendremos:

$$RV_5^3 = 5^3 = 125$$

1.8.4 Combinaciones

Cuando se trata de contar el número de subconjuntos de x elementos en un conjunto de n elementos tenemos lo que se denomina combinaciones de x elementos en un conjunto de n . El cálculo del conteo se hace mediante el número combinatorio, de la manera siguiente:

$$C_n^x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Ejemplo

¿De cuántas maneras podemos elegir, en la urna anterior (recordemos que había cinco bolas), tres bolas en una única extracción?

Respuesta

Serán combinaciones de cinco elementos tomados de tres en tres, por tanto, tendremos:

$$C_5^3 = \binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

1.8.5 Permutaciones con repetición

Sea un conjunto de n elementos, de entre los cuales tenemos a elementos indistinguibles entre sí, b elementos indistinguibles entre sí, c elementos indistinguibles entre sí, etc. Cada ordenación de estos elementos se denominará permutación con repetición. El número de permutaciones con repetición es:

$$RP_n^{a,b,c,\dots} = \frac{n!}{a!b!c!\dots}$$

Ejemplo

¿Cuántas palabras con sentido o sin él pueden formarse con las letras PATATA?

Respuesta: Tenemos tres veces la letra A, dos veces la T y una vez la P. Por tanto, serán:

$$RP_6^{3,2,1} = \frac{6!}{3!2!1!} = 60$$

1.9 Frecuencia relativa y probabilidad

La definición moderna de probabilidad basada en la axiomática de Kolmogorov (presentada anteriormente) es relativamente reciente. Históricamente hubo otros intentos previos de definir el escurridizo concepto de probabilidad, descartados por diferentes razones. Sin embargo conviene destacar aquí algunas ideas que aparecen en la antigua definición basada en la frecuencia relativa, ya que permiten intuir algunas profundas propiedades de la probabilidad.

Recordemos antes que si en un experimento que se ha repetido n veces un determinado suceso A se ha observado en k de estas repeticiones, la frecuencia relativa f_r del suceso A es:

$$f_r = k/n$$

El interés por la frecuencia relativa y su relación con el concepto de probabilidad aparece a lo largo de los siglos XVIII a XX al observar el comportamiento de numerosas repeticiones de experimentos reales.

A título de ejemplo de un experimento de este tipo, supongamos que se dispone de una moneda ideal perfectamente equilibrada. Aplicando directamente la regla de Laplace resulta claro que el suceso A = obtener cara tiene probabilidad:

$$p(A) = 1/2 = 0,5$$

En el cuadro siguiente se simula por ordenador el comportamiento de la frecuencia relativa del suceso A = obtener cara. El cuadro inicia la simulación con el lanzamiento consecutivo de la moneda veinte veces, calculando la frecuencia relativa de cara y comparándolo con la $p(A) = 0.5$. Aunque no es imposible que coincidan, la mayoría de veces f_r será diferente.

El lector puede manipular el cuadro para observar qué ocurre con rachas entre $n = 1$ y $n = 1000$ lanzamientos. También puede empezarse una nueva racha de lanzamientos con el botón Reiniciar.

Al realizar este tipo de experimento ha de llamar la atención:

- el carácter errático del comportamiento de f_r entre los valores 0 y 1 , pero a pesar de ello se intuye que a mayor número de lanzamientos n , más improbable es que f_r se aleje mucho de $p(A)$.

La teoría moderna de la probabilidad enlaza formalmente estas ideas con el estudio de las leyes de los grandes números, que se discutirán brevemente en el capítulo dedicado a las “Grandes muestras” .

2 Variables aleatorias y Distribuciones de probabilidad

2.1 El espacio muestral y sus elementos

Cuando llevamos a cabo un experimento aleatorio, el conjunto Ω de resultados posibles forman el denominado espacio muestral. Sus elementos ω (resultados o sucesos elementales) deben ser conocidos por el investigador que realiza la experiencia, aun cuando no podamos determinar a priori el resultado particular de una realización concreta.

Supondremos que también conocemos la manera de asignar una probabilidad sobre el conjunto de enunciados o sucesos observables que se pueden construir a partir de Ω . Es decir, supondremos la existencia de un espacio de probabilidad construido a partir de los resultados de Ω .

Generalmente, la estructura del espacio muestral no permite, o por lo menos no facilita, su tratamiento matemático. Pensemos en la inmensa variedad en la naturaleza de resultados posibles de diferentes experimentos. Además es bastante frecuente que no nos interesen los resultados en sí, sino una característica que, de alguna manera, resuma el resultado del experimento.

2.2 Representación numérica de los sucesos elementales. Variables aleatorias

La forma de resumen que adoptaremos es la asignación a cada suceso elemental de un valor numérico, en particular, de un número real.

En la práctica la asignación de un valor numérico a cada elemento del espacio muestral se hace siguiendo una regla o enunciado, según el interés concreto del experimentador. Evidentemente, podemos construir diversas maneras de asignar valores numéricos a los mismos resultados de un experimento.

Hablando en términos coloquiales, podemos decir que cada regla de asignación corresponde a una determinada variable que se puede medir sobre los sucesos elementales.

Nótese que es posible construir múltiples variables sobre un mismo espacio de probabilidad. En términos algo más formales, las reglas de asignación se pueden interpretar como una aplicación de Ω en el conjunto de números reales.

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) \end{aligned}$$

X representa la variable o regla de asignación concreta. El conjunto de valores numéricos que puede tomar una variable, y que depende de la naturaleza de la misma variable, recibe el nombre de recorrido de la variable.

A partir de este momento, los sucesos elementales quedan substituidos por sus valores numéricos de acuerdo a una determinada variable y permiten un mayor tratamiento matemático en el marco de la teoría de la probabilidad.

El apelativo aleatoria que reciben las variables hace referencia al hecho de que los posibles valores que toman dependen de los resultados de un fenómeno aleatorio que se presentan con una determinada probabilidad.

Como un complemento al tema presentamos la definición formal de variable aleatoria, donde se introducen las restricciones a las reglas de asignación numérica que posibilitan el tratamiento matemático de las variables.

2.3 Caracterización de una variable aleatoria a través de la probabilidad. Función de distribución

Una vez que tenemos definida una variable aleatoria, ésta queda totalmente caracterizada en el momento en que somos capaces de determinar la probabilidad de que la variable tome valores en cualquier intervalo de la recta real. Dado que los posibles valores que puede tomar la variable, es decir, su recorrido, pueden ser muy grandes (infinitos de hecho), el problema de caracterizar una variable aleatoria se ha resuelto a través de la definición de una serie de funciones matemáticas.

La más general de dichas funciones es la función de distribución. Definimos la función de distribución de una variable aleatoria X como la aplicación

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\rightarrow F(x) = P(X \leq x) = P\{\omega \in \Omega \mid X(\omega) \leq x\} \end{aligned}$$

Por tanto, para cada punto de la recta real, el valor de la función de distribución es la probabilidad del suceso formado por los resultados del experimento que tienen asignado un valor de la variable aleatoria menor o igual a dicho punto. O también podemos decir que es la probabilidad inducida en el intervalo de la recta $(-\infty, x]$.

Hay que hacer notar que siempre será posible determinar dicha probabilidad gracias a los requerimientos exigidos en la definición formal de variable aleatoria. Por tanto, toda variable aleatoria tiene asociada una función de distribución. Nos referimos a esta función cuando decimos que conocemos la distribución de la variable aleatoria.

2.4 Propiedades de la función de distribución

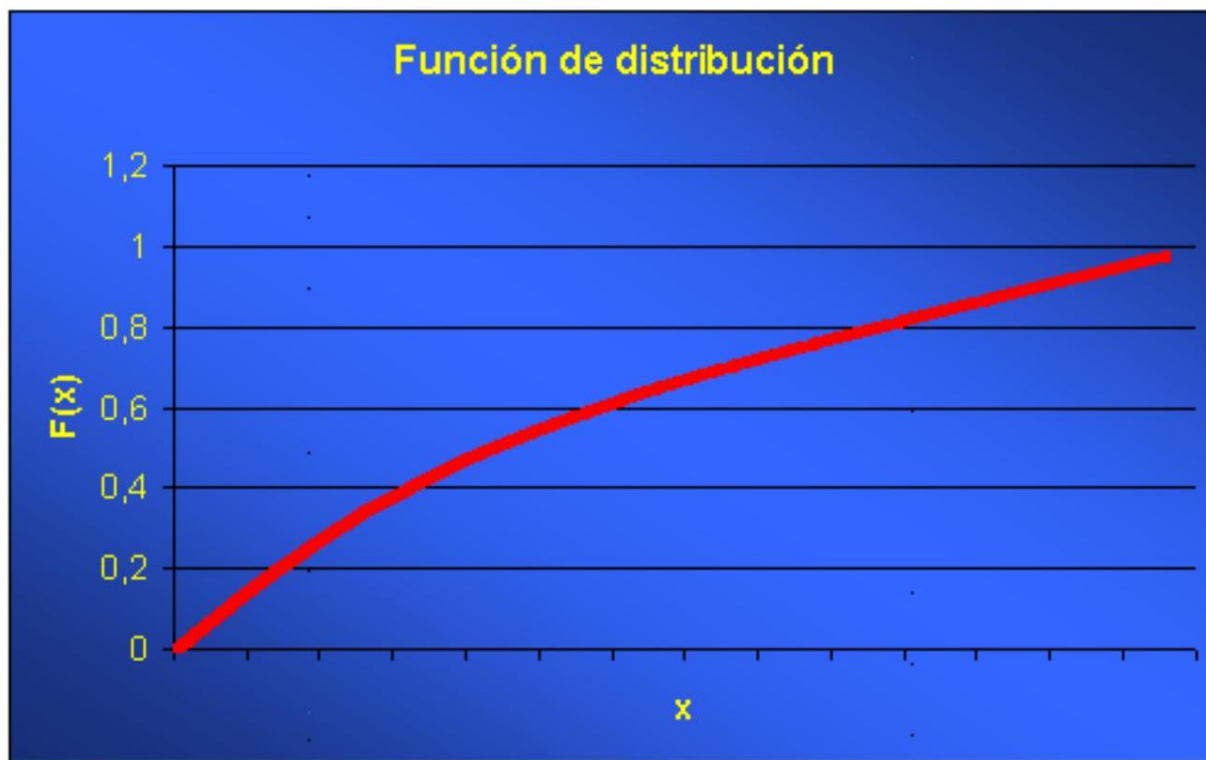
Resumimos a continuación las propiedades más importantes de la función de distribución:

1. $0 \leq F(x) \leq 1$
2. $\lim_{x \rightarrow +\infty} F(x) = 1$
3. $\lim_{x \rightarrow -\infty} F(x) = 0$
4. $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
5. $\lim_{x \rightarrow a^+} F(x) = F(a) \quad \forall a \in \mathbb{R}$ Función siempre continua por la derecha.

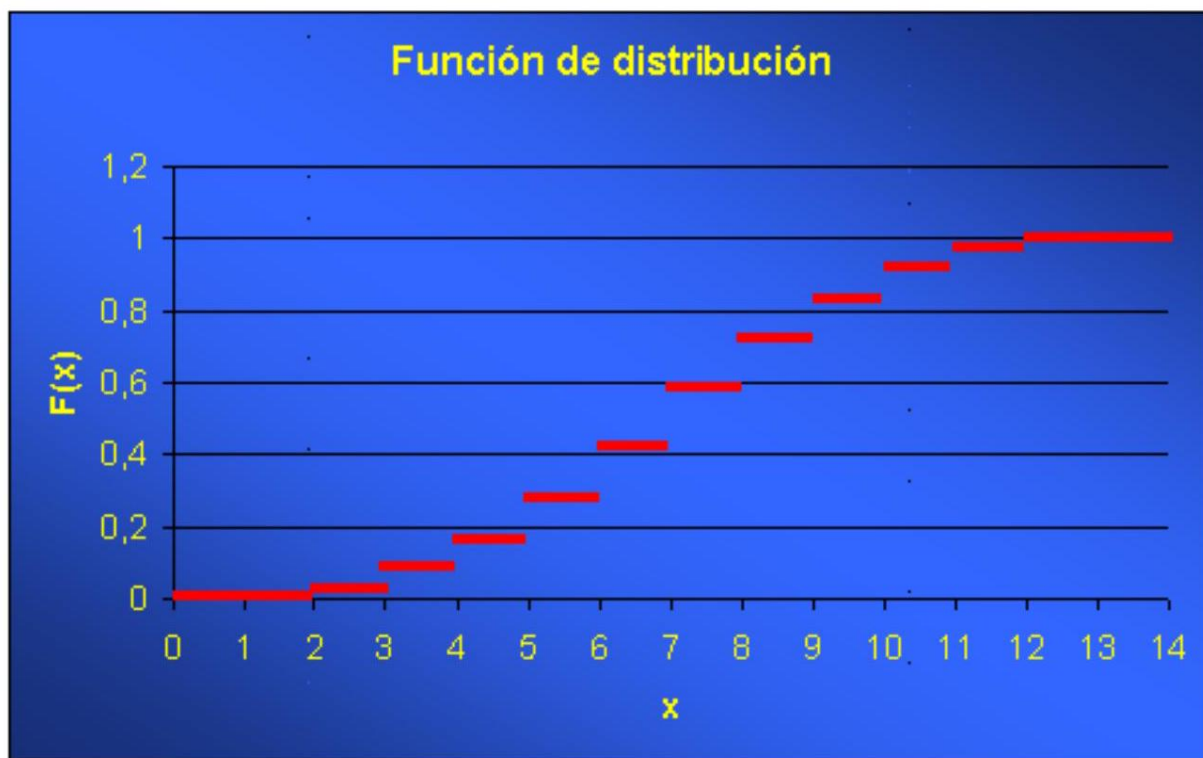
Toda función que verifique las propiedades anteriores es una función de distribución y toda función de distribución caracteriza una determinada variable aleatoria sobre algún espacio de probabilidad.

Las propiedades anteriores se traducen en un tipo de gráfica para la función de distribución del tipo de las que mostramos a continuación:

Primer tipo



Segundo tipo



Evidentemente, podrían aparecer distribuciones, y por tanto gráficas, que combinen las características de los dos modelos anteriores.

2.5 Clasificación de las variables aleatorias

Para su estudio, las variables aleatorias se clasifican en variables discretas o variables continuas.

2.5.1 Variables aleatorias discretas

Diremos que una variable aleatoria es discreta si su recorrido es finito o infinito numerable. Generalmente, este tipo de variables van asociadas a experimentos en los cuales se cuenta el número de veces que se ha presentado un suceso o donde el resultado es una puntuación concreta.

Los puntos del recorrido se corresponden con saltos en la gráfica de la función de distribución, que correspondería al segundo tipo de gráfica visto anteriormente.

2.5.2 Variables aleatorias continuas

Son aquellas en las que la función de distribución es una función continua. Se corresponde con el primer tipo de gráfica visto.

Generalmente, se corresponden con variables asociadas a experimentos en los cuales la variable medida puede tomar cualquier valor en un intervalo; mediciones biométricas, por ejemplo.

Un caso particular dentro de las variables aleatorias continuas y al cual pertenecen todos los ejemplos usualmente utilizados, son las denominadas variables aleatorias absolutamente continuas.

2.5.2.1 Variables aleatorias absolutamente continuas

Diremos que una variable aleatoria X continua tiene una distribución absolutamente continua si existe una función real f , positiva e integrable en el conjunto de números reales, tal que la función de distribución F de X se puede expresar como

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

Una variable aleatoria con distribución absolutamente continua, por extensión, se la clasifica como variable aleatoria absolutamente continua.

A la función f se la denomina función de densidad de probabilidad de la variable X .

Hay que hacer notar que no toda variable continua es absolutamente continua, pero los ejemplos son complicados, algunos utilizan para su construcción el conjunto de Cantor, y quedan fuera del nivel y del objetivo de este curso.

Igualmente indicaremos que los tipos de variables comentados anteriormente forman únicamente una parte de todos los posibles tipos de variables, sin embargo contienen prácticamente todas las variables aleatorias que encontramos usualmente.

Tal como se estudiará más adelante, existen algunas familias de funciones de distribución, tanto dentro del grupo de las discretas como de las continuas, que por su importancia reciben un nombre propio y se estudiarán en los capítulos siguientes.

En ocasiones encontramos variables de tipo mixto, es decir que se comportan como discretas o continuas para distintos grupos de valores.

2.6 Variable aleatoria discretas

Una variable aleatoria X diremos que es discreta si su recorrido es finito o infinito numerable, recorrido que denotaremos de la forma $\{x_1, x_2, \dots, x_k, \dots\}$.

El ejemplo más sencillo de variable aleatoria discreta lo constituyen las variables indicadoras. Sea A un suceso observable, se llama indicador de A a la variable aleatoria definida por

$$I_A : \Omega \rightarrow \mathbb{R} \\ \omega \rightarrow I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

2.6.0.1 Ejercicio:

Construir, a partir de las variables indicadoras de A y B , las siguientes variables indicadoras?

$$I_{A \cap B}; I_{A \cup B}; I_{A^c}; I_{\Omega}$$

2.6.0.1.1 Solución

$$\begin{aligned}I_{A \cap B} &= I_A \cdot I_B \\I_{A \cup B} &= I_A + I_B - I_{A \cap B} \\I_{A^c} &= 1 - I_A \\\Omega &= 1\end{aligned}$$

2.6.1 Caracterización a través de la función de densidad o de probabilidad

Las variables aleatorias discretas vienen caracterizadas a través de una función que asocia cada elemento del recorrido con su probabilidad. Dicha función recibe varios nombres según los autores: función o ley de probabilidad, o también función de densidad de la variable aleatoria discreta.

Podemos representarla de la manera siguiente:

$$\begin{aligned}f: \mathbb{R} &\rightarrow [0, 1] \\x &\rightarrow f(x) = P(X = x) = P\{\omega \in \Omega \mid X(\omega) = x\}\end{aligned}$$

La función definida anteriormente es nula en todo punto que no pertenezca al recorrido. Es evidente que, al ser una probabilidad, la función de densidad discreta está acotada $0 \leq f(x) \leq 1$. Toda función de densidad discreta puede expresarse de manera explícita a través de una tabla que asocie directamente puntos del recorrido con sus probabilidades.

Ejemplo: Consideremos la variable indicadora del suceso A :

$$\begin{aligned}I_A: \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}\end{aligned}$$

La función de densidad de esta variable sería la siguiente:

x	0	1
$f(x) = P(X = x)$	$1 - P(A) = P(A^c)$	$P(A)$

El recorrido está formado por dos valores: 1 y 0 , con las mismas probabilidades que las del suceso A y su complementario, respectivamente.

En otros casos la función de densidad se expresa a través de una fórmula matemática que define una regla de asignación de probabilidades para los valores del recorrido.

2.6.1.1 Ejemplo

$$P(X = x) = 0,2 \cdot 0,8^{x-1}, \quad x = 1, 2, \dots$$

es la función de densidad de una variable aleatoria discreta con recorrido numerable.

2.6.2 Propiedades de la función de densidad discreta

1.

$$0 \leq f(x) \leq 1$$

2. $\sum_{i=1}^n f(x_i) = 1$, si el recorrido es finito.

3. $\sum_{i=1}^{\infty} f(x_i) = 1$, si el recorrido es numerable.

2.6.3 Relaciones entre la función de distribución y la función de densidad discreta. Probabilidad de intervalos.

Existe una relación muy importante entre las funciones de distribución $F(x)$ y de densidad $f(x)$ de una variable aleatoria discreta. La función de distribución en un punto se obtiene acumulando el valor de la función de densidad para todos los valores del recorrido menores o iguales al punto en cuestión.

$$F(x) = \sum_{x_i \leq x} f(x_i) \quad \text{para todo } x_i \text{ perteneciente al recorrido de la variable.}$$

En efecto, supongamos que el recorrido de una variable discreta X es $\{x_1, x_2, \dots, x_k, \dots\}$ y que deseamos conocer el valor de la función de distribución en un punto x tal que $x_i \leq x < x_{i+1}$, entonces es inmediato que

$$F(x) = P(X \leq x) = P(X = x_1) + P(X = x_2) + \dots + P(X = x_i) = f(x_1) + f(x_2) + f(x_3) + \dots + f(x_i)$$

Por ejemplo, para una variable indicadora de un suceso A , tenemos la relación siguiente:

Valor de x	$f(x)$	$F(x)$
$(-\infty, 0)$		0
0	$P(A^c)$	$P(A^c)$
$(0, 1)$		$P(A^c)$
1	$P(A)$	$P(A^c) + P(A) = 1$
$(1, +\infty)$		1

A partir de las funciones de densidad y de distribución es posible expresar las probabilidades para cualquier posible intervalo de valores de la variable. Por ejemplo:

Intervalo
$P(X \leq a) = F(a)$
$P(X < a) = F(a) - f(a)$
$P(X > a) = 1 - F(a) = 1 - P(X \leq a)$
$P(X \geq a) = 1 - F(a) + f(a) = 1 - P(X > a)$
$P(a < X \leq b) = F(b) - F(a)$
$P(a < X < b) = F(b) - f(b) - F(a)$

Intervalo
$P(a \leq X \leq b) = F(b) - F(a) + f(a)$
$P(a \leq X < b) = F(b) - f(b) - F(a) + f(a)$

2.7 Variables aleatorias continuas

Una variable aleatoria X diremos que es continua si su función de distribución es una función continua. En la práctica, se corresponden con variables asociadas con experimentos en los cuales la variable medida puede tomar cualquier valor en un intervalo: mediciones biométricas, intervalos de tiempo, áreas, etc.

2.7.1 Ejemplos

- Resultado de un generador de números aleatorios entre 0 y 1. Es el ejemplo más sencillo que podemos considerar, es un caso particular de una familia de variables aleatorias que tienen una distribución uniforme en un intervalo $[a, b]$. Se corresponde con la elección al azar de cualquier valor entre a y b .
- Estatura de una persona elegida al azar en una población. El valor que se obtenga será una medición en cualquier unidad de longitud (m , cm , etc.) dentro de unos límites condicionados por la naturaleza de la variable. El resultado es impredecible con antelación, pero existen intervalos de valores más probables que otros debido a la distribución de alturas en la población. Más adelante veremos que, generalmente, variables biométricas como la altura se adaptan un modelo de distribución denominado distribución Normal y representado por una campana de Gauss.

Dentro de las variables aleatorias continuas tenemos las variables aleatorias absolutamente continuas.

Diremos que una variable aleatoria X continua tiene una distribución absolutamente continua si existe una función real f , positiva e integrable en el conjunto de números reales, tal que la función de distribución F de X se puede expresar como

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

Una variable aleatoria con distribución absolutamente continua, por extensión, se clasifica como variable aleatoria absolutamente continua.

En cuanto a nuestro manual, todas las variables aleatorias continuas con las que trabajemos pertenecen al grupo de las variables absolutamente continuas, en particular, los ejemplos y casos expuestos.

2.7.2 Función de densidad continua

La función que caracteriza las variables continuas es aquella función f positiva e integrable en los reales, tal que acumulada desde $-\infty$ hasta un punto x , nos proporciona el valor de la función de distribución en x , $F(x)$. Recibe el nombre de función de densidad de la variable aleatoria continua.

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

Las funciones de densidad discreta y continua tienen, por tanto, un significado análogo, ambas son las funciones que acumuladas (en forma de sumatorio en el caso discreto o en forma de integral en el caso continuo) dan como resultado la función de distribución.

La diferencia entre ambas, sin embargo, es notable.

- La función de densidad discreta toma valores positivos únicamente en los puntos del recorrido y se interpreta como la probabilidad de la que la variable tome ese valor $f(x) = P(X = x)$.
- La función de densidad continua toma valores en el conjunto de números reales y no se interpreta como una probabilidad. No está acotada por 1, puede tomar cualquier valor positivo. Es más, en una variable continua se cumple que probabilidades definidas sobre puntos concretos siempre son nulas.

$$P(X = x) = 0 \text{ para todo } x \text{ real.}$$

¿Cómo se interpreta, entonces, la función de densidad continua? Las probabilidades son las áreas bajo la función de densidad. El área bajo la función de densidad entre dos puntos a y b se interpreta como la probabilidad de que la variable aleatoria tome valores comprendidos entre a y b .

Por tanto, siempre se cumple lo siguiente:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

La función de densidad se expresa a través de una función matemática. La forma específica de la función matemática generalmente pasa por considerar a la variable aleatoria como miembro de una determinada familia de distribuciones, un determinado modelo de probabilidad. Estas familias generalmente dependen de uno o más parámetros y serán objeto de un estudio específico en un capítulo posterior. La atribución a una determinada familia depende de la naturaleza de la variable en cuestión.

Podemos ver, únicamente con ánimo ilustrativo, la expresión analítica y la gráfica para los ejemplos comentados con anterioridad:

- Resultado de un generador de números aleatorios entre a y b . Modelo Uniforme. $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$
- Estatura de una persona elegida al azar en una población. Modelo Normal.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-170)^2}{2}} - \infty < x < \infty$$

2.7.3 Relaciones entre la función de distribución y la función de densidad.

Para una variable continua, la relación entre las funciones de distribución y de densidad viene dada directamente a través de la definición. La función de distribución en un punto se obtiene integrando el valor de la función de densidad desde menos infinito hasta el punto en cuestión. Por ejemplo:

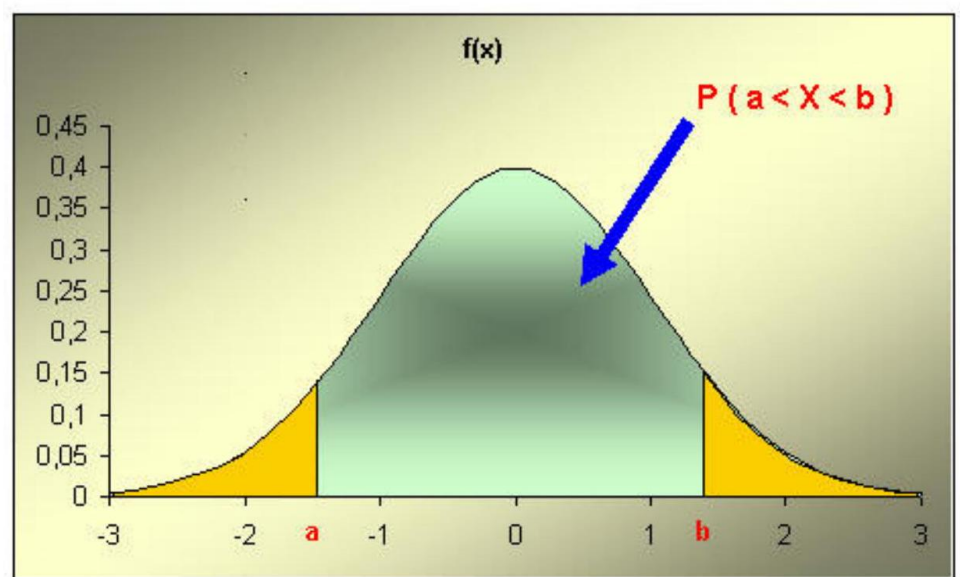
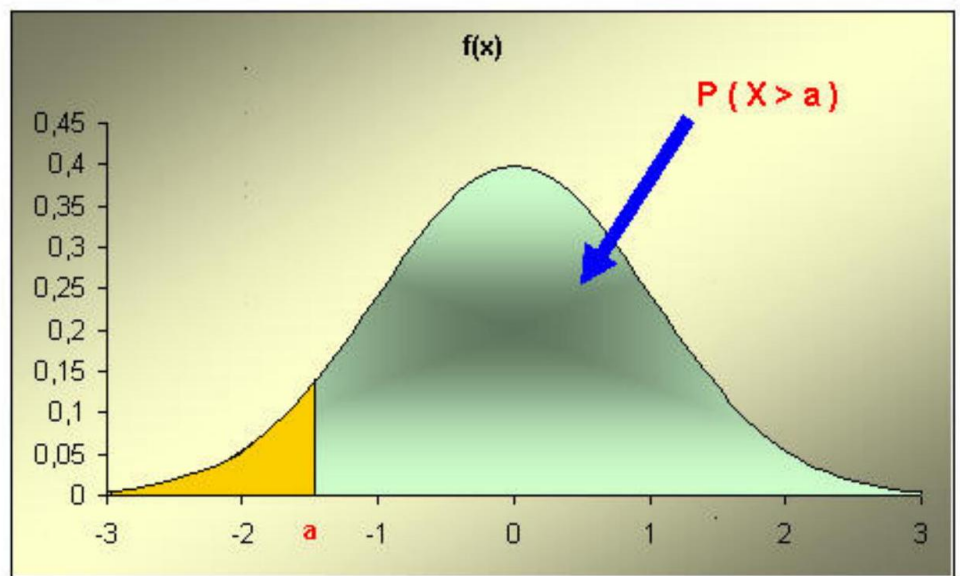
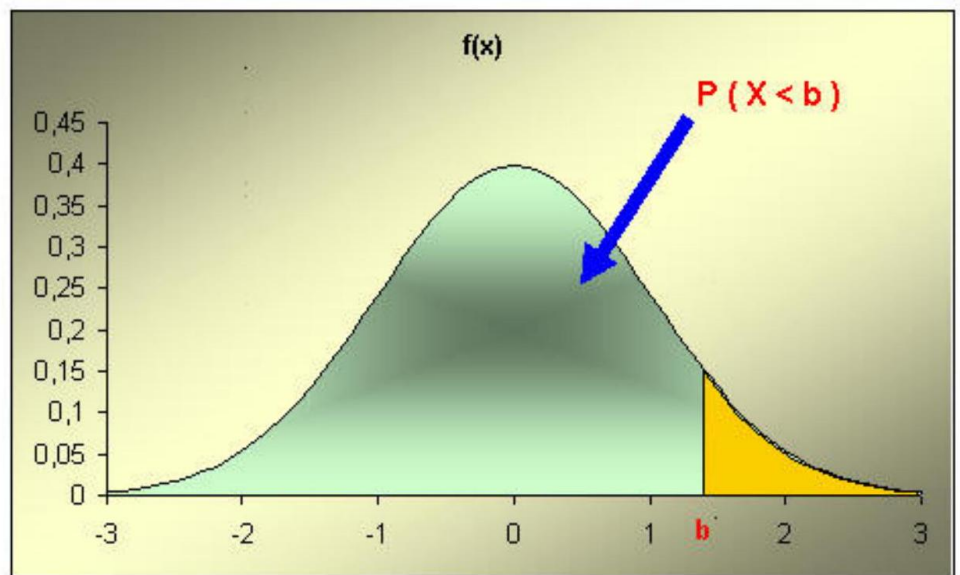
$$F(x) = \int_{-\infty}^x f(t)dt$$

2.7.3.1 Probabilidad de intervalos

A partir de las funciones de densidad y de distribución, y teniendo en cuenta que $P(X = x) = 0$ para todo x real, es posible expresar las probabilidades para cualquier posible intervalo de valores de la variable. Por ejemplo:

Intervalo
$P(X \leq a) = P(X < a) = F(a) = \int_{-\infty}^a f(x)dx$
$P(X \geq a) = P(X > a) = 1 - F(a) = \int_a^{+\infty} f(x)dx$
$P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b)$ $= F(b) - F(a) = \int^b f(x)dx$

Fijémonos que la probabilidad de los intervalos se corresponde con el área bajo la función de densi-



dad dentro del intervalo considerado.

2.8 Independencia de variables aleatorias

De manera intuitiva podemos decir que dos variables aleatorias son independientes si los valores que toma una de ellas no afectan a los de la otra ni a sus probabilidades.

En muchas ocasiones la independencia será evidente a partir del experimento, por ejemplo, es independiente el resultado del lanzamiento de un dado y el de una moneda, por tanto las variables Puntuación obtenida con el dado y Número de caras obtenidas al lanzar la moneda una vez serán variables independientes.

En otras ocasiones tenemos una dependencia clara, por ejemplo, al lanzar un dado consideremos las variables X = puntuación del dado Y = variable indicadora de puntuación par. Es evidente que existe una clara dependencia, si sabemos que $Y = 1$, la variable X sólo puede tomar los valores 2, 4 o 6; si sabemos que $X = 3$, entonces, $Y = 0$ forzosamente.

Algunas veces podemos suponer la existencia de una cierta relación entre variables, aunque sea en forma algo abstracta y sin concretar. Por ejemplo si realizamos unas mediciones sobre unos individuos, las variables altura en cm y peso en Kg probablemente estarán relacionadas, los valores de una influirán en los valores de la otra. Intentar determinar la naturaleza exacta de la relación entre ambas es lo que en estadística conocemos como un problema de regresión.

Si queremos una definición algo más formal, basta con que recordemos que dos sucesos son independientes si la probabilidad de la intersección es igual al producto de probabilidades, aplicando esta definición a sucesos del tipo $X \leq a$ tenemos la definición siguiente:

2.8.1 Caracterización de la independencia

Diremos que dos variables aleatorias X e Y son independientes si y sólo si

$$P(X \leq a \cap Y \leq b) = P(X \leq a) \cdot P(Y \leq b) = F_X(a) \cdot F_Y(b)$$

A la función $F(x, y) = P(X \leq a \cap Y \leq b)$ se la conoce como la función de distribución conjunta de X e Y . Como consecuencia inmediata de la independencia de X e Y , se cumple lo siguiente:

$$P(a < X \leq c \cap b < Y \leq d) = P(a < X \leq c) \cdot P(b < Y \leq d)$$

2.9 Caracterización de una variable aleatoria a través de parámetros

Hasta el momento hemos visto que toda variable aleatoria viene caracterizada a través de unas determinadas funciones matemáticas, las funciones de distribución y de densidad. Una vez caracterizada, y por tanto conocida, la distribución de una variable aleatoria, podemos obtener cualquier probabilidad asociada.

En ocasiones podemos acotar más el problema y reducir el estudio de una variable aleatoria a determinar una serie de características numéricas asociadas con la distribución de la variable. Dichas características tienen como propiedad fundamental el hecho de resumir gran parte de las propiedades de la variable aleatoria y juegan un papel muy destacado en las técnicas estadísticas que desarrollaremos a lo largo del curso.

Por ejemplo, supuesta la pertenencia de una variable aleatoria a una determinada familia de distribuciones de probabilidad, bien sea discreta o continua, los diferentes miembros de la familia diferirán en el valor de esas características numéricas. En este caso, denominaremos a tales características los parámetros de la distribución.

Existe un buen número de tales características, pero nos centraremos en las dos más importantes: la esperanza y la varianza. La primera nos informa sobre la localización de los valores de la variable y la segunda, sobre el grado de dispersión de estos valores.

2.10 Esperanza de una variable aleatoria discreta

La esperanza matemática de una variable aleatoria es una característica numérica que proporciona una idea de la localización de la variable aleatoria sobre la recta real. Decimos que es un parámetro de centralización o de localización.

Su interpretación intuitiva o significado se corresponde con el valor medio teórico de los posibles valores que pueda tomar la variable aleatoria, o también con el centro de gravedad de los valores de la variable supuesto que cada valor tuviera una masa proporcional a la función de densidad en ellos.

La definición matemática de la esperanza en el caso de las variables aleatorias discretas se corresponde directamente con las interpretaciones proporcionadas en el párrafo anterior. Efectivamente, supuesta una variable aleatoria discreta X con recorrido $\{x_1, x_2, \dots, x_k, \dots\}$ y con función de densidad $f(x)$, se define la esperanza matemática de X como el valor

$$E(X) = \sum_{x_i \in X(\Omega)} x_i f(x_i)$$

donde el sumatorio se efectúa para todo valor que pertenece al recorrido de X . En caso de que el recorrido sea infinito la esperanza existe si la serie resultante es absolutamente convergente, condición que no siempre se cumple.

La definición se corresponde con un promedio ponderado según su probabilidad de los valores del recorrido y, por tanto, se corresponde con la idea de un valor medio teórico.

2.11 Esperanza de una variable aleatoria continua

La idea intuitiva que más nos puede ayudar en la definición de la esperanza matemática de una variable aleatoria continua es la idea del centro de gravedad de los valores de la variable, donde cada valor tiene una masa proporcional a la función de densidad en ellos.

Dada una variable aleatoria absolutamente continua X con función de densidad $f(x)$, se define la esperanza matemática de X como el valor

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

suponiendo que la integral exista.

2.12 Propiedades de la esperanza matemática

1. Esperanza de una función de una variable aleatoria

- Variable discreta

$$E(h(X)) = \sum_{x_i \in X(\Omega)} h(x_i) f(x_i)$$

- Variable continua

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x) f(x) dx$$

2.12.1 Linealidad de la esperanza matemática

- $E(X + Y) = E(X) + E(Y)$
- $E(k \cdot X) = k \cdot E(X)$ para todo número real k .
- $E(k) = k$ para todo número real k .

$E(a \cdot X + b) = a \cdot E(X) + b$ para todo par de números reales a y b .

2.12.2 Esperanza del producto

- $E(X \cdot Y) = E(X) \cdot E(Y)$ únicamente en el caso de que X e Y sean variables aleatorias independientes.

2.13 Varianza de una variable aleatoria

La varianza de una variable aleatoria es una característica numérica que proporciona una idea de la dispersión de la variable aleatoria respecto de su esperanza. Decimos que es un parámetro de dispersión.

La definición es la siguiente:

$$\text{Var}(X) = E((X - E(X))^2)$$

Es, por tanto, el promedio teórico de las desviaciones cuadráticas de los diferentes valores que puede tomar la variable respecto de su valor medio teórico o esperanza.

En el caso de las variables discretas, la expresión se convierte en:

$$\text{Var}(X) = \sum_{x_i \in X(\Omega)} (x_i - E(X))^2 f(x_i)$$

mientras que para las variables continuas tenemos:

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$$

En ambos casos existe una expresión equivalente alternativa y generalmente de cálculo más fácil:

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Una de las características de la varianza es que viene expresada en unidades cuadráticas respecto de las unidades originales de la variable. Un parámetro de dispersión derivado de la varianza y que tiene las mismas unidades de la variable aleatoria es la desviación típica, que se define como la raíz cuadrada de la varianza.

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{E((X - E(X))^2)}$$

2.13.1 Propiedades de la varianza

1. $\text{Var}(X) \geq 0$
2. $\text{Var}(k \cdot X) = k^2 \cdot \text{Var}(X)$ para todo número real k .
3. $\text{Var}(k) = 0$ para todo número real k .
4. $\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$ para todo par de números reales a y b .
5. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ únicamente en el caso que X y Y sean independientes.

2.14 Momentos (de orden k) de una variable aleatoria

- Dada una variable aleatoria X , definimos el momento de orden k como:

$$m_k = E(X^k)$$

suponiendo que tal esperanza exista. Podemos ver que la esperanza es el momento de orden 1, $E(X) = m_1$.

- Definimos el momento central de orden k como:

$$\mu_k = E((X - E(X))^k)$$

Con la denominación anterior, la varianza es el momento central de orden 2, $\text{Var}(X) = \mu_2$.

- Es posible también definir momentos mixtos de dos variables aleatorias. Dadas dos variables aleatorias X y Y definimos el momento mixto de orden (r, k) como

$$m_{rk} = E(X^r \cdot Y^k)$$

y el momento mixto central de orden (r, k) como

$$\mu_{rk} = E(X - E(X))^r \cdot (Y - E(Y))^k$$

- El momento mixto central más importante es el μ_{11} , denominado la covarianza de X e Y , y con una interpretación en el sentido de cuantificar el grado de dependencia entre dos variables aleatorias, puesto que si X e Y son independientes se verifica que $\mu_{11} = 0$, mientras que si $\mu_{11} \neq 0$ entonces las variables son dependientes.

2.15 Definición formal de variable aleatoria

Tal como hemos comentado, la definición formal de variable aleatoria impone una restricción matemática en la formulación vista hasta el momento.

Definiremos una variable aleatoria como una aplicación de Ω en el conjunto de números reales

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) \end{aligned}$$

que verifique la propiedad siguiente

$$\forall x \in \mathbb{R} \quad \text{el conjunto } A = \{\omega \mid X(\omega) \leq x\} \text{ es un suceso observable}$$

es decir, para todo número real x , el conjunto de resultados elementales tales que la variable aleatoria toma sobre ellos valores inferiores o iguales a x ha de ser un suceso sobre el cual podamos definir una probabilidad.

Dicha propiedad recibe el nombre de medibilidad y por tanto podríamos decir que una variable aleatoria es una función medible de Ω en los reales.

Esta condición nos asegura que podremos calcular sin problemas, probabilidades sobre intervalos de la recta real a partir de las probabilidades de los sucesos correspondientes.

$$P(X \leq x) = P\{\omega \mid X(\omega) \leq x\}$$

La expresión anterior se leería de la manera siguiente: La probabilidad de que la variable aleatoria tome valores inferiores o iguales a x es igual a la probabilidad del suceso formado por el conjunto de resultados elementales sobre los que el valor de la variable es menor o igual que x .

La probabilidad obtenida de esta manera se denomina probabilidad inducida. Se puede comprobar que, a partir de la condición requerida, se pueden obtener probabilidades sobre cualquier tipo de intervalo de la recta real. Por ejemplo:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

La condición exigida para ser variable aleatoria discreta ahora puede ser expresada como:

$$\forall k = 1, 2, \dots \quad \text{el conjunto } A = \{\omega \mid X(\omega) = x_k\} = X^{-1}(\{x_k\}) \text{ es un suceso observable}$$

Toda variable aleatoria definida sobre un espacio de probabilidad finito es necesariamente discreta. La suma y el producto de variables aleatorias discretas, definido por:

$$(X + Y)(\omega) = X(\omega) + Y(\omega) \text{ y } (X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$$

es también una variable aleatoria discreta.

3 Distribuciones Notables

3.1 Distribuciones discretas

3.1.1 La distribución de Bernoulli

Es el modelo discreto más sencillo en que podamos pensar. Hace referencia a situaciones en las que el resultado de un experimento sólo puede ser: se ha dado el suceso A ó no se ha dado el suceso A . Por ejemplo, en el lanzamiento de una moneda sólo puede darse el suceso sale cara o su complementario no sale cara (sale cruz).

Por lo tanto, definimos la variable aleatoria X de la siguiente manera:

- $X = 1$ si se ha dado A .
- $X = 0$ si no se ha dado A , es decir, se ha dado el complementario A^c .

Si además, conocemos la probabilidad de que suceda A :

$$P[A] = p$$

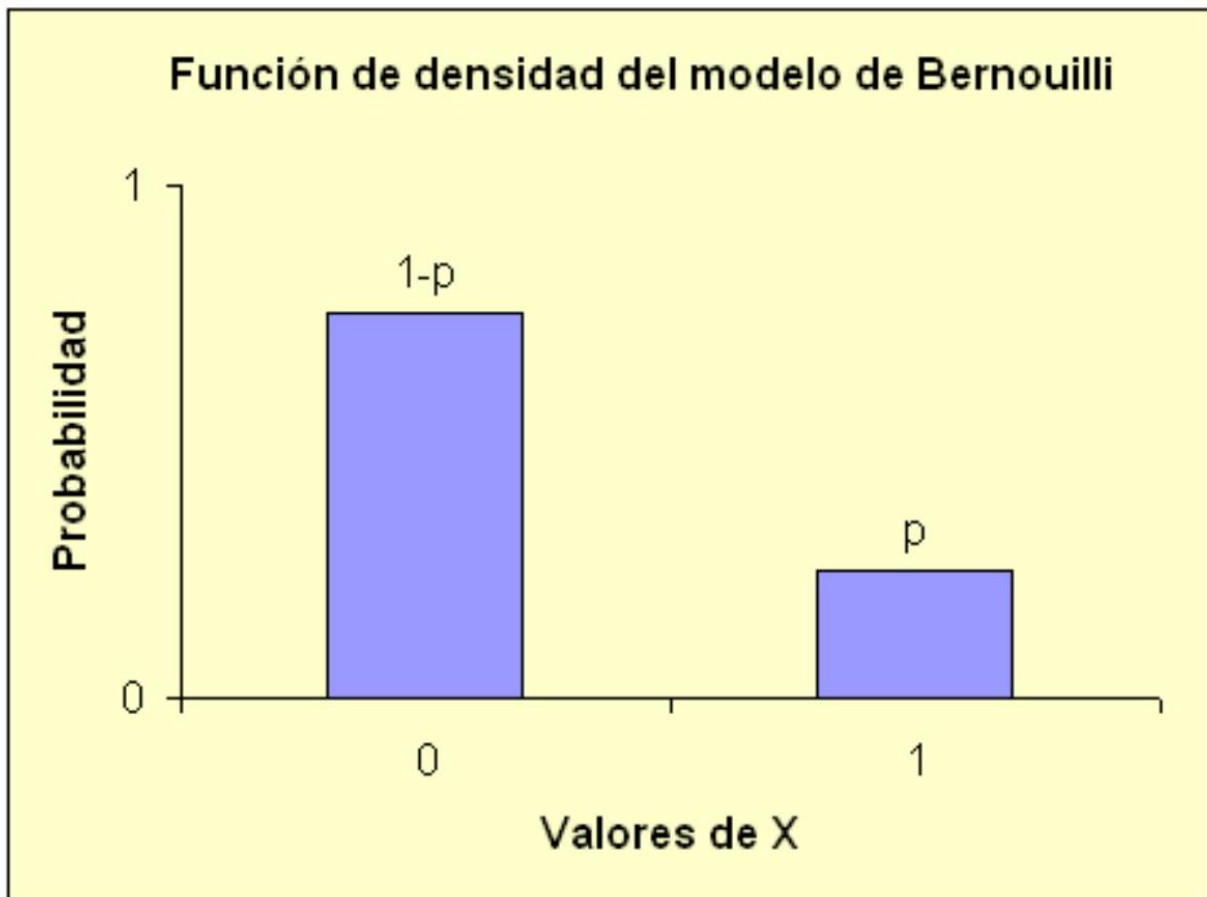
y, por tanto,

$$P[A^c] = 1 - p$$

ya podemos definir la distribución de la variable aleatoria X . En estas condiciones diremos que X sigue una distribución de Bernoulli de parámetro p , que abreviaremos así $X \sim \text{Bernoulli}(p)$, y su función de densidad se define así:

$$f(k) = P[X = k] = \begin{cases} p & \text{si } k = 1 \text{ (se ha dado } A \text{)} \\ 1 - p & \text{si } k = 0 \text{ (se ha dado } A^c \text{)} \end{cases}$$

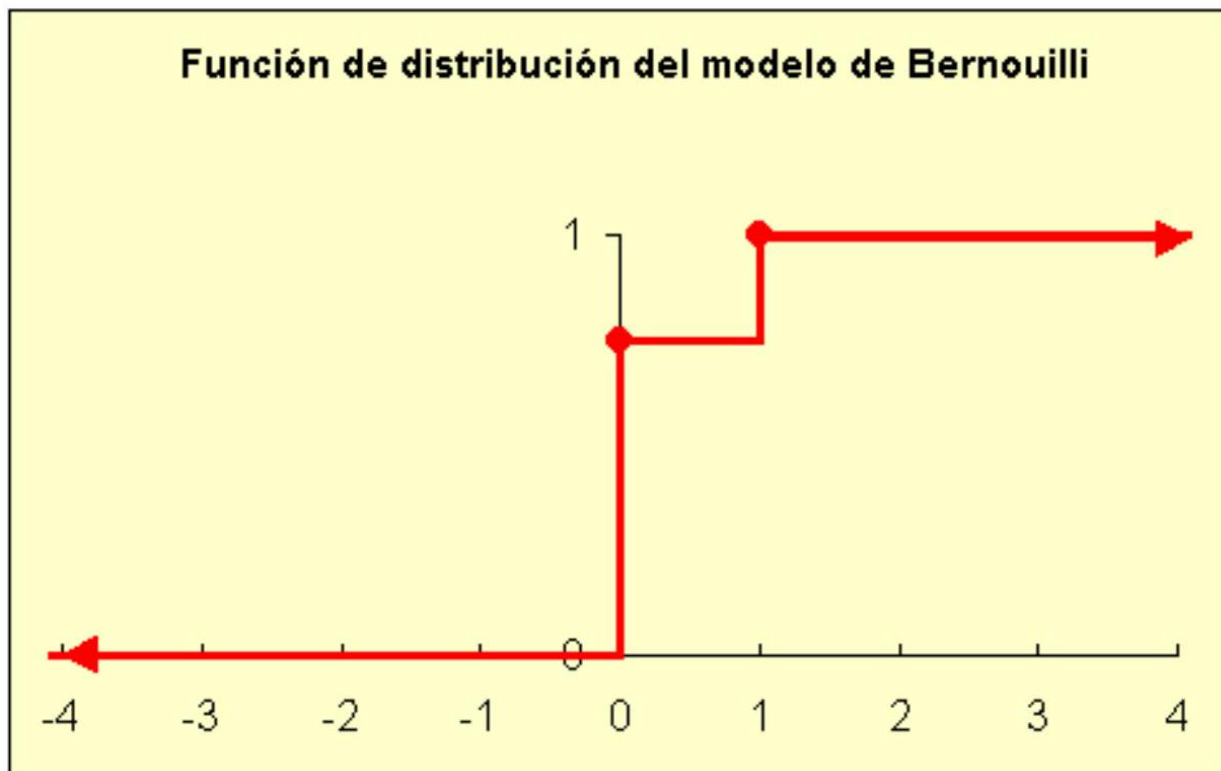
Gráficamente:



Mientras que la función de distribución será:

$$F(k) = P[X \leq k] = \begin{cases} 0 & \text{si } k < 0 \\ p & \text{si } 0 \leq k < 1 \\ 1 & \text{si } k \geq 1 \end{cases}$$

Gráficamente:



3.1.1.1 Propiedades del modelo de Bernoulli

- 1) La esperanza vale $E(X) = p$.
- 2) La varianza vale $V(X) = p(1 - p)$.

3.1.2 La distribución Binomial

Al igual que el modelo de Bernoulli, hace referencia a experiencias con resultados dicotómicos (el resultado sólo puede ser A o A^c). Sin embargo en este modelo estamos interesados en la repetición de n veces una experiencia de este tipo en condiciones independientes.

Tomemos el ejemplo del conteo del número de caras en el lanzamiento n veces de una moneda regular. Para concretar, vamos a suponer que disponemos de una moneda regular ($P[\text{cara}] = P[\text{cruz}] = 1/2$) que lanzamos cuatro veces. Es evidente que, en estas condiciones, la variable X : número de caras en cuatro lanzamientos independientes de una moneda regular es una variable aleatoria discreta que sólo puede tomar cinco posibles valores:

$$x = 0, 1, 2, 3, 4$$

Pasemos ahora a calcular la probabilidad de cada valor (en terminología estadística, vamos a calcular la función de densidad de la variable X).

Es evidente que la $P[X = 0]$ es igual a la probabilidad de salgan cuatro cruces seguidas:

$$P[X = 0] = P[\text{cruz}, \text{cruz}, \text{cruz}, \text{cruz}] = P[\text{cruz}]^4 = (1/2)^4 = 0,0625$$

ya que la moneda es regular y, por tanto, $P[\text{cara}] = P[\text{cruz}] = 1/2$. La $P[X = 3]$ corresponde al suceso de que salgan tres caras (c en adelante) y una cruz (+ en adelante). Sin embargo, en este caso tenemos hasta cuatro posibles maneras de obtener dicho resultado, según el orden en que aparezcan las tres caras y la cruz:

+ccc	c + cc	cc + c	ccc+
------	--------	--------	------

También debería resultar evidente que la probabilidad de cada uno de estos sucesos es la misma:

$$P[+ccc] = P[c + cc] = P[cc + c] = P[ccc+] = (1/2)^4 = (1/2)^4 = 0,0625$$

de manera que, finalmente, la probabilidad de que salgan tres caras y una cruz es la suma de las probabilidades de los 4 casos anteriores:

$$P[X = 3] = 4(1/2)^4 = 0,25$$

Y así podríamos ir calculando el resto de casos. Podemos ver que, en este ejemplo, todos los casos tienen la misma probabilidad (0,0625) y que el número total de casos posibles es 16 . En términos de combinatoria dicho número se obtendría como variaciones con repetición de dos valores (cara o cruz) tomados de cuatro en cuatro (el número de lanzamientos de la moneda):

$$VR_2^4 = 2^4 = 16$$

En la siguiente tabla se muestran los dieciséis posibles resultados:

$k = \text{número de caras}$	Casos
0	++++
1	+++c
	+ + c +
	+c + +
	c + ++
	++cc
	+c + c
	c + +c +
	c + c +
	cc++
	ccc+
	c + cc

Si hacemos uso de nuestros conocimientos de combinatoria, comprobamos que el número de casos para cada posible valor $k(k = 0, 1, 2, 3, 4)$ puede calcularse como permutaciones con repetición de cuatro elementos tomado de k y $4 - k$:

$$RP_4^{k,4-k} = \frac{4!}{k!(4-k)!} = \binom{4}{k}$$

y obtenemos finalmente el número combinatorio 4 sobre k . En efecto, para el caso $k = 3$, tendríamos:

$$\binom{4}{3} = \frac{4!}{3!1!} = 4$$

que son los cuatro posibles casos que nos dan tres caras y una cruz. Finalmente, recordando que todos los casos tienen la misma probabilidad, se construye la siguiente tabla:

$k = \text{número de caras}$	Número de casos	$P[X = k]$
0	1	0,0625
1	4	0,2500

2	6	0,3750
3	4	0,2500
4	1	0,0625
Total	16	1

3.1.2.1 Los parámetros de la distribución Binomial

La última tabla de la página anterior es, justamente, la función de densidad de nuestra variable X .

Función de densidad de X	
k	$P[X = k]$
0	0,0625
1	0,2500
2	0,3750
3	0,2500
4	0,0625
En otro caso	0

Como hemos visto, para obtener los resultados anteriores, hemos tenido que definir dos valores:

1. n : el número de lanzamientos (repeticiones de la experiencia aleatoria en condiciones independientes), en nuestro caso $n = 4$.
2. p : la probabilidad de que salga cara ($P[c]$), en nuestro caso $p = 1/2$.

Se dice, por tanto, que la distribución Binomial depende de dos parámetros: n y p . En nuestro ejemplo, diremos que X sigue una distribución Binomial de parámetros $n = 4$ i $p = 1/2$. De forma abreviada:

$$X \sim B(n = 4; p = 1/2)$$

En el ejemplo que hemos visto, suponíamos que la moneda era regular y, por tanto,

$$P[c] = P[+] = 1/2$$

Si tenemos una moneda trucada con las siguientes probabilidades:

$$P[c] = 2/3 \quad \text{y} \quad P[+] = 1/3$$

diremos que en este caso la variable X : número de caras en cuatro lanzamientos independientes de nuestra moneda trucada sigue una distribución Binomial de parámetros:

$$X \sim B(n = 4; p = 2/3)$$

El problema se nos complica levemente ya que ahora no todos los posibles resultados tienen la misma probabilidad. Veamos dos ejemplos:

- La probabilidad de obtener cuatro caras es:

$$P[cccc] = (2/3)^4 = 0,1975$$

- La probabilidad de que el primer lanzamiento sea cara y el resto sean cruces valdrá:

$$P[c^{+++}] = (2/3)'(1/3)^3 = 0,0247$$

Sin embargo sí se cumplirá que la probabilidad de que todos los casos que resulten en el mismo número de caras y cruces tendrán la misma probabilidad. Por ejemplo, para los cuatro casos en los que el número total de caras es 1 y el de cruces 3 :

$$P[c+++]=P[+c++]=P[++c+]=P[+++c]=(2/3)'(1/3)^3=0,0247$$

Y, por tanto, la probabilidad de obtener una sola cara en el lanzamiento de nuestra moneda trucada será:

$$P[X = 1] = 4'0,0247 = 0,0988$$

O, generalizando, si $P[A] = p$ y $P[A^c] = 1 - p$ tenemos que

$$P[X = k] = c(n, k)p^k(1 - p)^{n-k} \quad \text{si } k = 0, 1, \dots, n$$

donde $c(n, k)$ representa el número de posibles resultados en los que obtenemos k caras y $n-k$ cruces en n lanzamientos. Tal como hemos visto, dicho número se puede calcular como permutaciones con repetición de n unidades tomadas de k y $n - k$.

Todo lo anterior nos lleva a formular el modelo binomial a través de la siguiente función de densidad:

$$f(k) = P[X = k] = \begin{cases} \binom{n}{k} p^k (1 - p)^{n-k} & \text{si } k = 0, \dots, n \\ 0 & \text{en caso contrario} \end{cases}$$

con lo que la función de distribución se calcularía:

$$F(k) = P[X \leq k] = \begin{cases} 0 & \text{si } k < 0 \\ \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} & \text{si } 0 \leq k < n \\ 1 & \text{si } k \geq n \end{cases}$$

En el programa siguiente se muestra la forma de la función de densidad junto con los valores de la función de densidad y de la función de distribución para cualquier valor:

AVIS

Posar un exemple de com fer-ho amb R

3.1.2.2 Propiedades del modelo Binomial

1. La esperanza vale $E(X) = np$.
2. La varianza es $V(X) = np(1-p)$.
3. Es una generalización del modelo de Bernoulli. En efecto, la Binomial con $n = 1$ (una sola realización) coincide con la distribución de Bernoulli.
4. La suma de dos variables aleatorias binomiales independientes con igual parámetro p también sigue una distribución Binomial:

$$X_1 \sim B(n = n_1; p = p_0) \quad \text{y} \quad X_2 \sim B(n = n_2; p = p_0)$$

Si definimos $Z = X_1 + X_2$ entonces,

$$Z \sim B(n = n_1 + n_2; p = p_0)$$

3.1.3 La distribución de Poisson

Se trata de un modelo discreto, pero en el que el conjunto de valores con probabilidad no nula no es finito, sino numerable. Se dice que una variable aleatoria X sigue la distribución de Poisson si su función de densidad viene dada por:

$$f(k) = P[X = k] = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!} & \text{si } k = 0, 1, 2, \dots \\ 0 & \text{en caso contrario} \end{cases}$$

Como vemos, este modelo se caracteriza por un sólo parámetro λ , que debe ser positivo. Esta distribución suele utilizarse para contajes del tipo número de individuos por unidad de tiempo, de espacio, etc.

3.1.3.1 Propiedades del modelo de Poisson

1. Esperanza: $E(X) = \lambda$.
2. Varianza: $V(X) = \lambda$.

En esta distribución la esperanza y la varianza coinciden.

3. La suma de dos variables aleatorias independientes con distribución de Poisson resulta en una nueva variable aleatoria, también con distribución de Poisson, de parámetro igual a la suma de parámetros:

$$X_1 \sim P(\lambda = \lambda_1) \quad \text{y} \quad X_2 \sim P(\lambda = \lambda_2)$$

y definimos $Z = X_1 + X_2$, entonces,

$$Z \sim P(\lambda = \lambda_1 + \lambda_2)$$

Este resultado se extiende inmediatamente al caso de n variables aleatorias independientes con distribución de Poisson. En este caso, la variable suma de todas ellas sigue una distribución de Poisson de parámetro igual a la suma de los parámetros.

3.1.4 La distribución Multinomial

COMPTE:Distribució Multivariant!

Using callouts is an effective way to highlight content that your reader give special consideration or attention.

Este modelo se puede ver como una generalización del Binomial en el que, en lugar de tener dos posibles resultados, tenemos r resultados posibles.

Supongamos que el resultado de una determinada experiencia puede ser r valores distintos: A_1, A_2, \dots, A_r cada uno de ellos con probabilidad p_1, p_2, \dots, p_r , respectivamente.

$$P(A_1) = p_1; \quad P(A_2) = p_2; \quad \dots \quad P(A_r) = p_r; \quad \text{con} \quad \sum_{i=1}^r P(A_i) = 1$$

Si repetimos la experiencia n veces en condiciones independientes, podemos preguntarnos la probabilidad de que el suceso A_1 aparezca k_1 veces, el suceso A_2 , k_2 veces y así sucesivamente:

$$P[(A_1 = k_1) \cap (A_1 = k_2) \cap \dots \cap (A_r = k_r)]$$

Al modelo estadístico que nos da dicha probabilidad se le denomina Multinomial, y su función de densidad viene dada por:

$$f(k_1, k_2, \dots, k_r) = P[(A_1 = k_1) \cap (A_1 = k_2) \cap \dots \cap (A_r = k_r)] = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$
$$\text{con} \sum_{i=1}^r P(A_i) = 1 \quad \text{y} \quad \sum_{i=1}^r k_i = n$$

como se ve, el modelo Multinomial queda definido por los parámetros $(n, p_1, p_2, \dots, p_r)$. La fórmula anterior puede deducirse de forma análoga al caso Binomial. En realidad, si tomamos $r = 2$ tenemos exactamente el modelo Binomial.

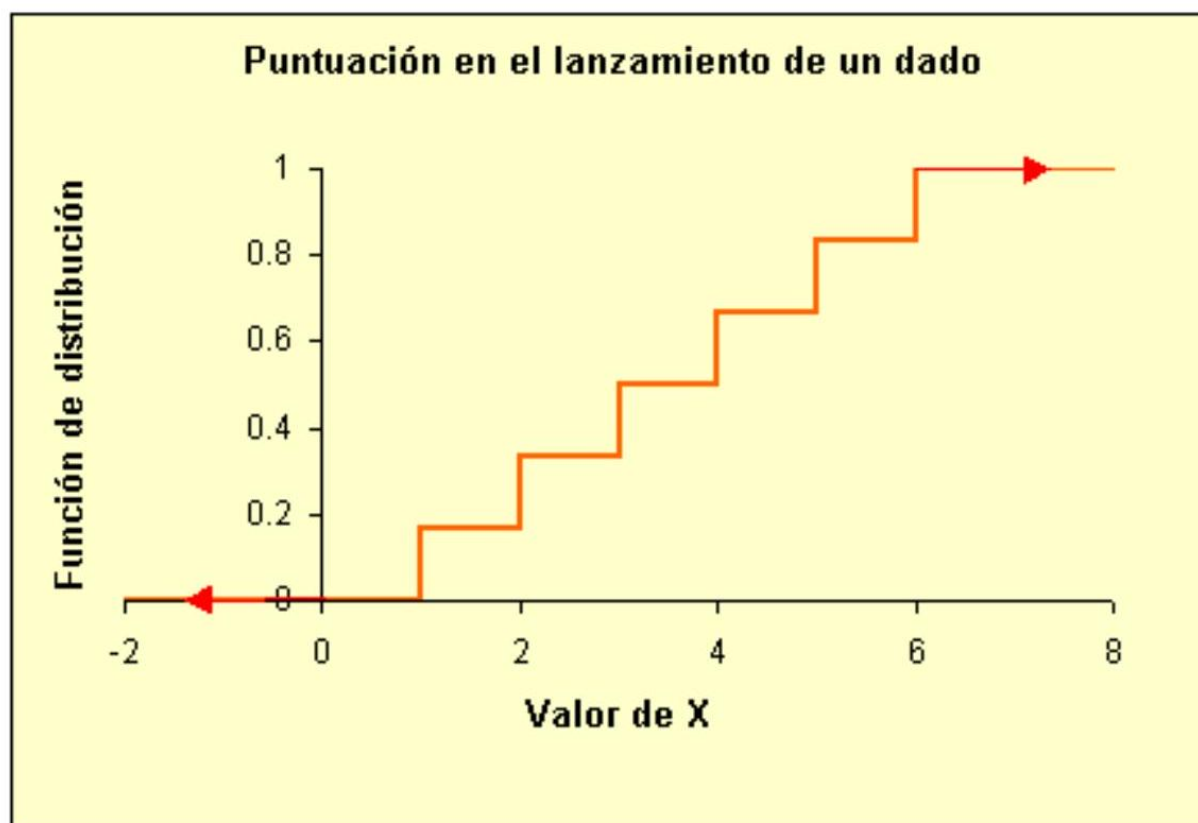
Se debe destacar que este modelo es un ejemplo de distribución multivariante, es decir, de distribución conjunta de varias (r) variables aleatorias. En efecto, si definimos la variable aleatoria X_1 como número de veces que se produce el suceso A_1 de un total de n experiencias, y así sucesivamente, tenemos un conjunto de r variables aleatorias discretas cuya función de densidad conjunta (valorada a la vez) viene definida por la anterior fórmula. Nótese que si consideramos cada una de estas variables $X_i (i = 1, 2, \dots, r)$ por separado, su distribución es la Binomial de parámetros n y p_i .

3.1.4.1 La distribución Uniforme discreta

Tenemos esta distribución cuando el resultado de una experiencia aleatoria puede ser un conjunto finito de n posibles resultados, todos ellos igualmente probables.

Un ejemplo puede ser la variable X , puntuación en el lanzamiento de un dado regular. Esta variable toma seis valores posibles, todos con la misma probabilidad $p = 1/6$. La función de densidad de esta variable será:

$$f(k) = P[X = k] = 1/6 \quad k = 1, 2, 3, 4, 5, 6$$



En general, si la variable X puede tomar $n(k = 1, 2, \dots, n)$ valores, todos con igual probabilidad, su función de densidad será:

$$f(k) = P[X = k] = 1/n \quad k = 1, 2, \dots, n$$

3.1.4.2 Propiedades del modelo Uniforme discreto

Sea n el número de valores equiprobables posibles:

3.1.4.3 Esperanza:

$$E(X) = \frac{n+1}{2}$$

3.1.4.4 Varianza:

$$V(X) = \frac{(n+1)[2(2n+1) - 3(n+1)]}{12}$$

3.1.5 La distribución Hipergeométrica

Este modelo presenta similitudes con el Binomial, pero sin la suposición de independencia de éste último. Veámoslo:

- Partimos de un conjunto formado por N individuos divididos en dos categorías mutuamente excluyentes: A y A^c ; de manera que N_1 individuos pertenecen a la categoría A y N_2 individuos, a la categoría A^c . Por tanto, se cumple que

$$N = N_1 + N_2$$

- Si del conjunto anterior extraemos n individuos sin reemplazamiento ($n \leq N$), la variable X que representa el número k de individuos que pertenecen a la categoría A (de los n extraídos) tiene por función de densidad:

$$f(k) = P[X = k] = \frac{\binom{N_1}{k} \binom{N_2}{n-k}}{\binom{N}{n}}$$

si $\max\{0, n - N_2\} \leq k \leq \min\{N_1, n\}$

La dependencia se debe al hecho de que N es finito y las extracciones se efectúan sin reemplazamiento. El caso de extracciones con reemplazamiento sería equivalente al de N infinito y se resolvería mediante el modelo Binomial.

El programa siguiente nos muestra la forma de la función de densidad de esta variable y el valor de la función de densidad y de la función de distribución en el punto que elijamos:

3.1.5.1 Propiedades del modelo hipergeométrico

1. Esperanza: $E(X) = nN_1/N_2$.
2. Varianza: $V(X) = (nN_1N_2(N - n)) / (N_2(N - 1))$

3.1.6 La distribución Geométrica o de Pascal

Definamos una experiencia aleatoria cuyo resultado sólo puede ser el suceso A o su complementario A^c , y que se repite secuencialmente hasta que aparece el suceso A por primera vez.

Definamos la variable aleatoria X como el número de veces que repetimos la experiencia en condiciones independientes hasta que se dé A por primera vez. Bajo estas condiciones, decimos que la variable X sigue una distribución geométrica o de Pascal de parámetro $p = P(A)$.

La función de densidad puede deducirse fácilmente de la definición:

$$f(k) = P[X = k] = (1 - p)^k p \quad k = 0, 1, 2, \dots$$

En el programa siguiente podéis ver su forma y obtener los valores de la función de densidad y de la de distribución:

Algunas puntualizaciones de la definición de X :

- Noté que, en esta definición, condiciones independientes significa que p , la probabilidad de A , y $1 - p$, la de su complementario A^c , no varían a lo largo de las sucesivas repeticiones de la experiencia.
- Tal y como la hemos definido, X se refiere al número de lanzamientos hasta que se produce A , pero sin contabilizar el último caso en que se da A . Por dicha razón X puede tomar los valores $k = 0, 1, 2, \dots$ con probabilidad no nula.

Un ejemplo de este modelo podría ser la experiencia consistente en lanzar sucesivamente un dado regular hasta que aparezca el número 6 . Si definimos la variable aleatoria X como el número de lanzamientos de un dado regular hasta que aparezca un 6 , queda claro que X sigue una distribución geométrica de parámetro $p = 1/6$.

3.1.6.1 Propiedades del modelo Geométrico o de Pascal

- 1) Esperanza: $E(X) = (1 - p)/p$
- 2) Varianza: $V(X) = (1 - p)/p^2$

3.1.6.2 Preguntas:

- ¿A que suceso nos referimos cuando decimos $X = 0$? Respuesta.
 - Cuando decimos que $X = 0$ nos referimos al caso en que el 6 aparece en el primer lanzamiento. La probabilidad de que esto suceda, suponiendo un dado regular, es de $1/6$:

$$P[X = 0] = 1/6$$

- ¿Cuál es la probabilidad de que el primer 6 aparezca en el cuarto lanzamiento? Respuesta.
 - La probabilidad de que el primer 6 aparezca en el cuarto lanzamiento corresponde a:

$$P[X = 3] = (5/6)^3 \cdot 1/6 = 0,0965$$

Fijémonos en que, si definimos A como el suceso sale un 6, la probabilidad anterior corresponde a la del suceso: $\{A^c A^c A^c A\}$ (en este orden).

3.1.7 La distribución Binomial negativa

Puede definirse como una generalización del modelo Geométrico o de Pascal. Así, dado un suceso A y su complementario A^c , cuando X representa el número de veces que se da A^c (ausencias, fallos, etc.) hasta que se produce r veces el suceso A , en una serie de repeticiones de la experiencia aleatoria en condiciones independientes, decimos que X sigue la distribución Binomial negativa. Nótese que, cuando $r = 1$, tenemos exactamente el modelo geométrico.

Este modelo queda definido por dos parámetros p (la probabilidad de $A : p = P(A)$) y r (el número de veces que debe producirse A para que detengamos la experiencia).

La función de densidad viene dada por:

$$f(k) = P[X = k] = \binom{k+r-1}{r-1} p^r q^k \quad k = 0, 1, 2, \dots$$

donde q representa el complementario de $p : q = 1 - p$.

3.1.7.1 Propiedades del modelo Binomial negativo

1. Esperanza: $E(X) = r'q/p$
2. Varianza: $V(X) = r'q/p^2$
3. Se cumplen las siguientes propiedades respecto la función de densidad:

$$f(0) = p^r \quad \text{y} \quad f(k+1) = \frac{(1-p)(k+r)}{k+1} f(k)$$

4. Este modelo se ajusta bien a contagios (números de individuos por unidad de superficie) cuando se produce una distribución contagiosa (los individuos tienden a agruparse).
5. La distribución Binomial negativa puede definirse con mayor generalidad si tomamos r como un número real positivo cualquiera (no necesariamente entero). Pero, en dicho caso, se pierde el carácter intuitivo del modelo y se complican ligeramente los cálculos. Por dichas razones, se ha excluido dicha posibilidad en esta presentación.

3.1.8 Tabla resumen de las distribuciones discretas principales

Distribución	Parámetros	Función de densidad	Esperanza	Varianza
Bernouilli	$0 \leq p \leq 1$	$p^k(1-p)^{1-k}$ $k = 0, 1$	p	$p(1-p)$
Binomial	$0 \leq p \leq 1$ $n = 1, 2, \dots$	$\binom{n}{k} p^k(1-p)^{n-k}$ $k = 0, 1, \dots, n$	np	$np(1-p)$
Poisson	$\lambda > 0$	$e^{-\lambda} \frac{\lambda^k}{k!}$ $k = 0, 1, 2, \dots$	λ	λ
Multinomial	$0 \leq p_1, \dots, p_r \leq 1$ $(p_1 + \dots + p_r = 1)$ $n = 1, 2, \dots$	$\frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$ $\sum_{i=1}^r k_i = n$	$\begin{pmatrix} np_1 \\ np_2 \\ \vdots \\ np_r \end{pmatrix}$	$\sigma_{ii} = np_i(1-p_i)$ $\sigma_{ij} = np_i p_j$ $i \neq j$
Uniforme discreta	$n = 1, 2, \dots$	$\frac{1}{n}$ $k = 1, 2, \dots, n$	$\frac{n+1}{2}$	$\frac{(n+1)[2(2n+1)-3(n+1)]}{12}$
Hipergeométrica	$\begin{cases} N = N_1 + N_2 \\ p = N_1/N \end{cases}$	$\frac{\binom{N_1}{k} \binom{N_2}{n-k}}{\binom{N}{n}}$ $\max\{0, n - N_2\} \leq k \leq \min\{N_1, n\}$	np	$np(1-p) \frac{N-n}{N-1}$
Pascal	$0 \leq p \leq 1$	$p(1-p)^k$ $k = 0, 1, 2, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Binomial negativa	$0 \leq p \leq 1$ $r > 0$		$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$

		$\binom{k+r-1}{r-1} p^r (1-p)^k$		
		$k = 0, 1, 2, \dots$		

3.2 DISTRIBUCIONES CONTINUAS

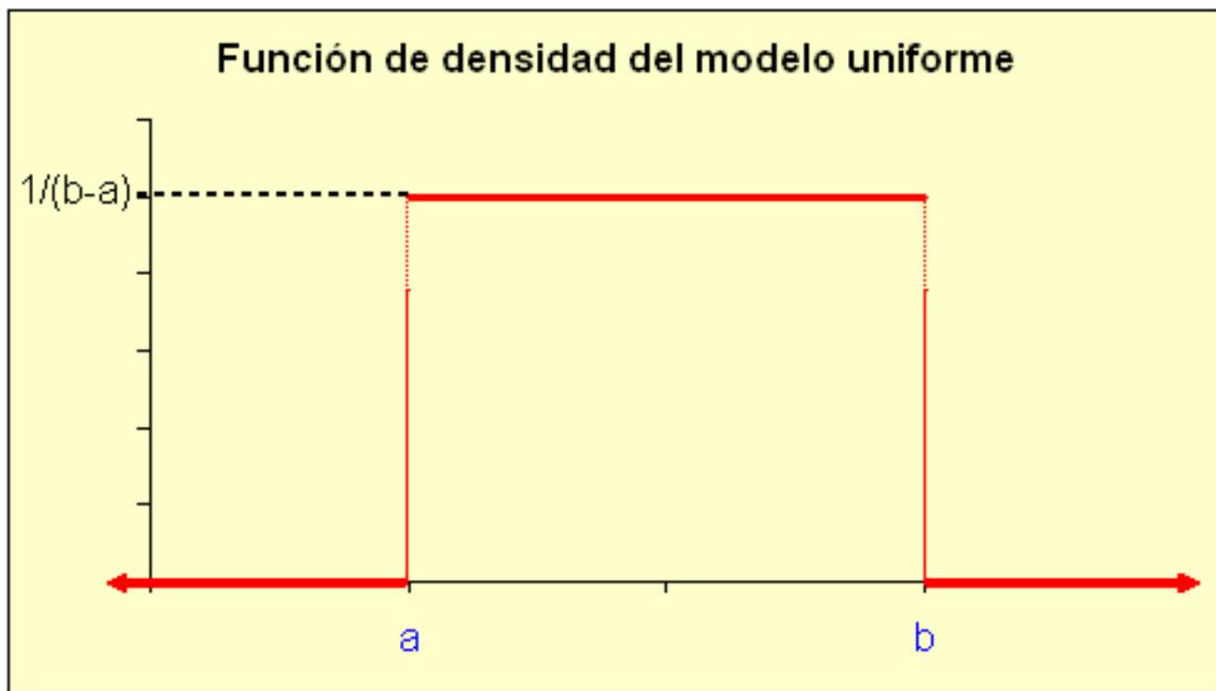
3.2.1 La distribución Uniforme

La distribución Uniforme es el modelo (absolutamente) continuo más simple. Corresponde al caso de una variable aleatoria que sólo puede tomar valores comprendidos entre dos extremos a y b , de manera que todos los intervalos de una misma longitud (dentro de (a, b)) tienen la misma probabilidad. También puede expresarse como el modelo probabilístico correspondiente a tomar un número al azar dentro de un intervalo (a, b) .

De la anterior definición se desprende que la función de densidad debe tomar el mismo valor para todos los puntos dentro del intervalo (a, b) (y cero fuera del intervalo). Es decir,

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b) \\ 0 & \text{si } x \notin (a, b) \end{cases}$$

Gráficamente:

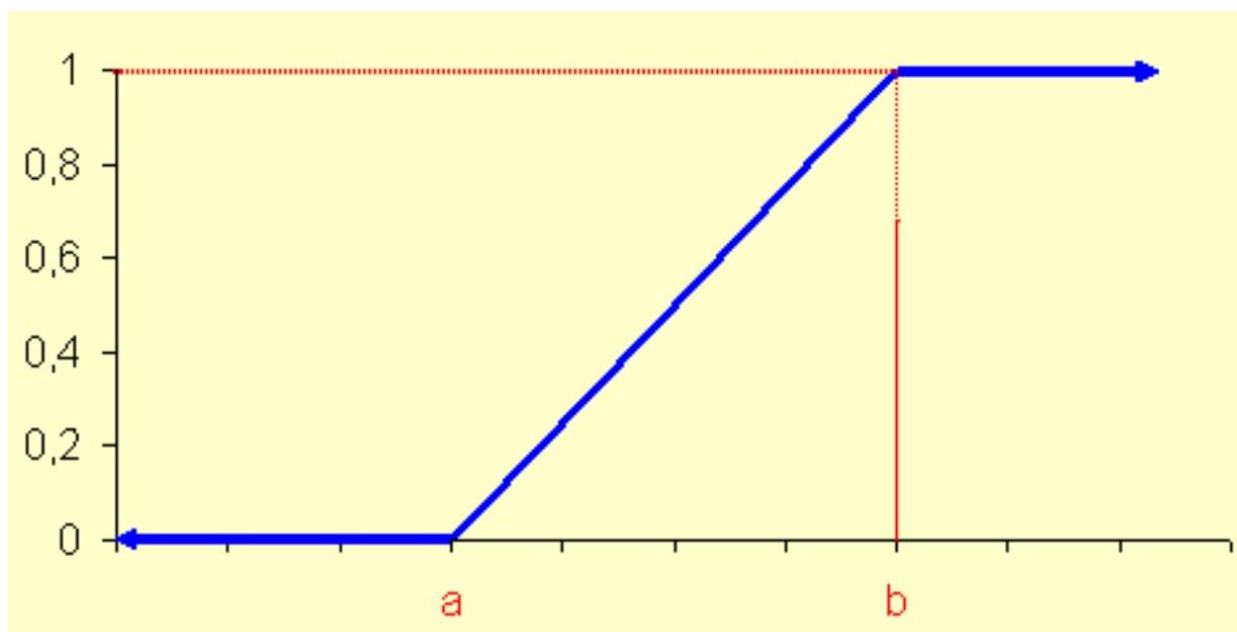


La función de distribución se obtiene integrando la función de densidad y viene dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in (a, b) \\ 1 & \text{si } x \geq b \end{cases}$$

Gráficamente:

Función de distribución del modelo uniforme



3.2.1.1 Propiedades del modelo Uniforme

1. Su esperanza vale $(b + a)/2$
2. Su varianza es $(b - a)^2/12$

3.2.1.2 Una aplicación del modelo Uniforme: el muestreo de Montecarlo

En ciertos casos es útil simular el muestreo de una variable aleatoria con una distribución dada. El muestreo de Montecarlo es un procedimiento general para obtener muestras aleatorias de cualquier tipo de variable (discreta o continua) si su función de distribución es conocida o se puede calcular. De hecho, todas las muestras artificiales de Statmedia han sido generadas a través del método de Montecarlo.

Supongamos que queremos generar una muestra procedente de una variable aleatoria X con función de distribución $F(x)$. El proceso comprende los siguientes pasos:

1. Obtener un valor aleatorio y entre cero y uno. Es decir, obtener una muestra de una distribución Uniforme entre cero y uno. La mayoría de lenguajes de programación incorporan un generador de este tipo.
2. Considerar el valor obtenido como el valor de la función de distribución a generar: $y = F(x)$.
3. El valor $x = F^{-1}(y)$ (la inversa de la función de distribución en el punto y) es un valor procedente de la distribución de la que deseábamos generar la muestra.
4. Si queremos obtener una muestra con n individuos debemos repetir los pasos anteriores n veces.

3.2.1.3 Generación de una muestra procedente de una distribución Binomial

Supongamos que queremos simular el experimento de contar el número de caras obtenidas en 5 lanzamientos de una moneda trucada con probabilidad de cara igual a 0,75 . Es decir, queremos obtener una muestra de una distribución Binomial con $n = 5$ y $p = 0,75$.

Siguiendo los pasos anteriores deberemos obtener un número al azar entre 0 y 1 (un valor procedente de una distribución Uniforme entre 0 y 1) y si este valor es menor o igual a 0,75 diremos que ha salido cara y, si es superior a 0,75 , cruz. Utiliza el siguiente programa para simular cinco lanzamientos con nuestra moneda trucada:

3.2.2 La distribución Exponencial

Este modelo suele utilizarse para variables que describen el tiempo hasta que se produce un determinado suceso.

Su función de densidad es de la forma:

$$f(x) = \begin{cases} \frac{1}{\alpha} \exp\left(-\frac{x}{\alpha}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Como vemos este modelo depende de un único parámetro α que debe ser positivo: $\alpha > 0$. A continuación se muestra un programa que nos permite ver cómo cambia la forma de la función de densidad según el parámetro α .

La función de distribución se obtiene integrando la de densidad y es de la forma:

$$F(x) = \begin{cases} 1 - \exp\left(-\frac{x}{\alpha}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Podemos utilizar el programa siguiente para calcular dicha función de distribución:

3.2.2.1 Propiedades del modelo Exponencial

1. Su esperanza es α .
2. Su varianza es α^2 .
3. Una propiedad importante es la denominada *carencia de memoria*, que podemos definir así: si la variable X mide el tiempo de vida y sigue una distribución Exponencial, significará que la probabilidad de que siga con vida dentro de 20 años es la misma para un individuo que a fecha de hoy tiene 25 años que para otro que tenga 60 años.
4. Cuando el número de sucesos por unidad de tiempo sigue una distribución de Poisson de parámetro λ (proceso de Poisson), el tiempo entre dos sucesos consecutivos sigue una distribución Exponencial de parámetro $\alpha = 1/\lambda$.

3.2.3 La distribución Normal

Se trata, sin duda, del modelo continuo más importante en estadística, tanto por su aplicación directa, veremos que muchas variables de interés general pueden describirse por dicho modelo, como por sus propiedades, que han permitido el desarrollo de numerosas técnicas de inferencia estadística. En realidad, el nombre de Normal proviene del hecho de que durante un tiempo se creyó, por parte de médicos y biólogos, que todas las variables naturales de interés seguían este modelo.

Su función de densidad viene dada por la fórmula:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad \text{donde } -\infty < x < +\infty$$

que, como vemos, depende de dos parámetros μ (que puede ser cualquier valor real) y σ (que ha de ser positiva). Por esta razón, a partir de ahora indicaremos de forma abreviada que una variable X sigue el modelo Normal así: $X \sim N(\mu, \sigma)$. Por ejemplo, si nos referimos a una distribución Normal con $\mu = 0$ y $\sigma = 1$ lo abreviaremos $N(0, 1)$.

A continuación vemos gráfica de esta función de densidad (podeis probar a cambiar los parámetros):

Como puedes ver, la función de densidad del modelo Normal tiene forma de campana, la que habitualmente se denomina campana de Gauss. De hecho, a este modelo, también se le conoce con el nombre de distribución gaussiana.

3.2.3.1 Propiedades del modelo Normal

1. Su esperanza es μ .
2. Su varianza es σ^2 y, por tanto, su desviación típica es σ .
3. Es simétrica respecto a su media μ , como puede apreciarse en la representación anterior.
4. Media, moda y mediana coinciden (μ).
5. Cualquier transformación lineal de una variable con distribución Normal seguirá también el modelo Normal. Si $X \sim N(\mu, \sigma)$ y definimos $Y = aX + b$ (con $a \neq 0$), entonces $Y \sim N(a\mu + b, |a|\sigma)$. Es decir, la esperanza de Y será $a\mu + b$ y su desviación típica, $|a|\sigma$.
6. Cualquier combinación lineal de variables normales independientes sigue también una distribución Normal. Es decir, dadas n variables aleatorias independientes con distribución $X_i \sim N(\mu_i, \sigma_i)$ para $i = 1, 2, \dots, n$ la combinación lineal: $Y = a_n X_n + a_{n-1} X_{n-1} + \dots + a_1 X_1 + a_0$ sigue también el modelo Normal:

$$Y \approx N\left(a_0 + \sum_{i=1}^n a_i \mu_i, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right)$$

La función de distribución del modelo Normal

La función de distribución del modelo Normal se debería calcular, como en el resto de distribuciones continuas, integrando la función de densidad:

$$F(x) = P[X \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$$

Pero nos encontramos con el problema de que no existe ninguna primitiva conocida para esta función, es decir, no sabemos resolver la anterior integral. Sin embargo, si somos incapaces de calcular la función distribución no podremos efectuar ningún cálculo con este modelo. ¿Cómo solucionamos el problema?

Una primera solución podría consistir en aproximar la integral a través de técnicas de cálculo numérico. Sin embargo, dado que el conjunto de valores que pueden tomar los parámetros μ y σ son infinitos, deberíamos repetir el proceso para cada valor diferente de algún parámetro. Afortunadamente, podemos ahorrarnos el esfuerzo aprovechando la propiedad de que cualquier transformación lineal de una variable Normal sigue también el modelo Normal. Por tanto, replantearemos cualquier problema en términos de una Normal concreta, que suele ser la $N(0, 1)$, de la siguiente manera:

Si $X \sim N(\mu, \sigma)$ y entonces definimos $Z = (X - \mu)/\sigma$ se cumplirá que $Z \sim N(0, 1)$

y, por tanto:

$$F_X(x) = P[X \leq x] = P\left[\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right] = P\left[Z \leq \frac{x - \mu}{\sigma}\right] = F_Z\left(\frac{x - \mu}{\sigma}\right)$$

A la distribución $N(0, 1)$, es decir, la que tiene por media cero y por desviación típica uno, se le denomina Normal reducida o tipificada. En cambio, al proceso de transformación del cálculo de la función de distribución de una Normal cualquiera a través de la Normal tipificada, se le denomina tipificación.

Debemos remarcar que el proceso de tipificación no resuelve el problema de la inexistencia de la función primitiva correspondiente. Sin embargo, sí es posible, mediante técnicas de cálculo numérico, obtener la integral numérica correspondiente y elaborar unas tablas que podemos consultar. Naturalmente, la tipificación permite que con una sola tabla, la de la $N(0, 1)$, tengamos suficiente.

Hoy en día, cada vez se utilizan menos tablas como la mencionada anteriormente, ya que los ordenadores, junto con los abundantes programas estadísticos existentes nos resuelven este problema. Sin embargo, la imposibilidad de integrar analíticamente la función de densidad persiste y, aunque nosotros no seamos conscientes, los programas informáticos realizan el proceso de tipificación para simplificar el problema.

A continuación se presenta un programa que permite comparar la función de densidad de una distribución Normal cualquiera con la de la Normal tipificada:

3.2.3.2 Cálculo de probabilidades del modelo Normal con Statmedia

AVIS

Posar un exemple de com fer-ho amb R O buscar si algun Shiny ho fa

El siguiente programa dibuja el área bajo de la función de densidad de una Normal cualquiera, a la izquierda de un valor x , es decir, el valor de la función de distribución en el punto x , cuyo valor también calcula el programa.

Para acabar, debemos recordar que Statmedia dispone de una calculadora estadística que también nos permite calcular la función de distribución de cualquier valor para una distribución Normal cualquiera. A continuación se muestra dicha calculadora una vez se ha escogido la opción de calculadora probabilística y, posteriormente, el modelo Normal:

Como podéis observar, calcula la función de densidad, la de distribución y la inversa de esta última. Además, incluye otras distribuciones ya vistas o que se verán posteriormente.

Para utilizar esta calculadora sólo tenéis que apretar el botón Calculadora de la barra de navegación.

3.2.4 La distribución Gamma

Este modelo es una generalización del modelo Exponencial ya que, en ocasiones, se utiliza para modelar variables que describen el tiempo hasta que se produce p veces un determinado suceso.

Su función de densidad es de la forma:

$$f(x) = \begin{cases} \frac{1}{\alpha^p \Gamma(p)} e^{-\frac{x}{\alpha}} x^{p-1} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Como vemos, este modelo depende de dos parámetros positivos: α y p . La función $\Gamma(p)$ es la denominada función Gamma de Euler que representa la siguiente integral:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$$

que verifica $\Gamma(p+1) = p\Gamma(p)$, con lo que, si p es un número entero positivo, $\Gamma(p+1) = p!$. El siguiente programa permite visualizar la forma de la función de densidad de este modelo (para simplificar, se ha restringido al caso en que p es un número entero).

3.2.4.1 Propiedades de la distribución Gamma

1. Su esperanza es $p\alpha$.
2. Su varianza es $p\alpha^2$.
3. La distribución Gamma ($\alpha, p = 1$) es una distribución Exponencial de parámetro α . Es decir, el modelo Exponencial es un caso particular de la Gamma con $p = 1$.
4. Dadas dos variables aleatorias con distribución Gamma y parámetro α común

$$X \sim G(\alpha, p_1) \text{ y } Y \sim G(\alpha, p_2)$$

se cumplirá que la suma también sigue una distribución Gamma

$$X + Y \sim G(\alpha, p_1 + p_2)$$

Una consecuencia inmediata de esta propiedad es que, si tenemos k variables aleatorias con distribución Exponencial de parámetro α (común) e independientes, la suma de todas ellas seguirá una distribución $G(\alpha, k)$.

3.2.5 La distribución de Cauchy

Se trata de un modelo continuo cuya función de densidad es:

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{para} \quad -\infty < x < \infty$$

Cuya integral nos proporciona la función de distribución:

$$F(x) = \int_{-\infty}^x \frac{1}{\pi(1+t^2)} dt = \frac{1}{\pi} [\arctan(t)]_{t=-\infty}^{t=x} = \frac{1}{2} + \frac{\arctan(x)}{\pi}$$

El siguiente programa permite visualizar la forma de la función de densidad de este modelo y el valor de la función de distribución:

3.2.5.1 Propiedades de la distribución de Cauchy

Se trata de un ejemplo de variable aleatoria que carece de esperanza (y, por tanto, también de varianza o cualquier otro momento), ya que la integral impropia correspondiente no es convergente:

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2x}{1+x^2} dx = \frac{1}{2\pi} \left[\lim_{x \rightarrow \infty} \ln(x^2) - \lim_{x \rightarrow -\infty} \ln(x^2) \right] = \frac{1}{2\pi} [\infty - \infty]$$

y nos queda una indeterminación. Por tanto, la esperanza de una distribución de Cauchy no existe. Cabe señalar que la función de densidad es simétrica respecto al valor cero (que sería la mediana y la moda), pero al no existir la integral anterior, la esperanza no existe.

3.2.6 La distribución de Weibull

Se trata de un modelo continuo asociado a variables del tipo tiempo de vida, tiempo hasta que un mecanismo falla, etc. La función de densidad de este modelo viene dada por:

$$f(x) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

que, como vemos, depende de dos parámetros: $\alpha > 0$ y $\beta > 0$, donde α es un parámetro de escala y β es un parámetro de forma (lo que proporciona una gran flexibilidad a este modelo).

La función de distribución se obtiene por la integración de la función de densidad y vale:

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$$

El siguiente programa permite visualizar la forma de la función de densidad de este modelo y el valor de la función de distribución:

3.2.6.1 Propiedades de la distribución Weibull

1. Si tomamos $\beta = 1$ tenemos una distribución Exponencial.
2. Su esperanza vale:

$$E(X) = \alpha \Gamma\left(\frac{1}{\beta} + 1\right)$$

3. Su varianza vale:

$$V(X) = \alpha^2 \left\{ \Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right) \right]^2 \right\}$$

donde $\Gamma(x)$ representa la función Gamma de Euler definida anteriormente.

3.2.7 Tabla resumen de las principales distribuciones continuas

Distribución	Parámetros	Función de densidad	Esperanza	Varianza
Uniforme	a, b	$\frac{1}{b-a} \quad a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial	$\alpha > 0$	$\frac{1}{\alpha} \exp\left(-\frac{x}{\alpha}\right) \quad x > 0$	α	α^2
Normal	$-\infty < \mu < \infty$ $\sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < +\infty$	μ	σ^2

Cauchy $\mid - \mid \frac{1}{\pi(1+x^2)} \quad -\infty < \mathbf{x} < \infty \mid - \mid - \mid$

Weibull $\mid \alpha > 0 \quad \beta > 0 \mid \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} \quad x \geq 0 \mid \alpha \Gamma\left(\frac{1}{\beta} + 1\right) \mid \alpha^2 \left\{ \Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right) \right]^2 \right\} \mid$

3.3 LA FAMILIA EXPONENCIAL DE DISTRIBUCIONES

4 Grandes muestras

4.1 Introducción: Aproximaciones asintóticas

4.1.1 Convergencia de variables aleatorias

4.2 Leyes de los grandes números

4.3 El teorema central del límite

El teorema central del límite (a partir de ahora, TCL) presenta un doble interés. Por un lado, proporciona a la estadística un resultado crucial para abordar el estudio de la distribución asintótica de muchos tipos de variables aleatorias. Como se verá en próximos capítulos, va a resultar básico en la construcción de contrastes de hipótesis y de intervalos de confianza, dos herramientas esenciales en estadística aplicada.

Además, el TCL proporciona una explicación teórica fundamentada a un fenómeno habitual en experimentos reales: las variables estudiadas presentan muchas veces una distribución empírica aproximadamente normal.

El TCL forma parte de un conjunto de propiedades relativas a las convergencias de variables aleatorias. En este tema se estudia sólo un tipo de convergencia, la convergencia en ley, ya que es necesaria para entender el enunciado del TCL. Se descarta, pues, en este documento el estudio de los otros tipos de convergencias (en probabilidad, casi segura, etc.) y el estudio de las leyes de los grandes números.

Posiblemente el lector con poca formación en análisis matemático hallará alguna dificultad en la primera lectura de la definición de convergencia en ley y en el enunciado del TCL. Si es este el caso, los ejemplos incluidos han de ayudar en su comprensión. Consideramos al TCL un resultado básico con el que hay que familiarizarse, ya que se aplicará repetidamente en los próximos temas.

4.3.1 Sumas de variables aleatorias

El TCL estudia el comportamiento de las sumas de variables aleatorias. En temas anteriores se han visto ya ejemplos de sumas de variables aleatorias.

Formalmente, la suma de dos variables aleatorias corresponde a la siguiente aplicación: si X_1 y X_2 son dos variables aleatorias definidas sobre Ω , la suma es:

$$\begin{aligned} X_1 + X_2 : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X_1(\omega) + X_2(\omega) \end{aligned}$$

La suma de dos variables puede extenderse sin dificultad a sumas de tres, cuatro,... y, en general, n variables aleatorias.

El TCL se ocupa de las sucesiones de variables aleatorias. En el contexto del TCL una sucesión corresponde a un conjunto donde el primer elemento es una variable aleatoria, el segundo elemento es la suma de dos variables aleatorias, el tercero es la suma de tres variables aleatorias, y así sucesivamente.

Una sucesión es un conjunto de elementos infinitos, que se designan simbólicamente mediante $\{X_n\}$. Cada uno de los elementos de la sucesión (que es una variable aleatoria) lleva asociada una determinada función de distribución:

$$X_n \rightarrow F_n$$

Así pues, la sucesión de variables aleatorias lleva asociada una secuencia paralela de funciones de distribución.

En los ejemplos se presentan sumas de variables aleatorias de diferentes tipos.

4.3.1.1 Presentación de los ejemplos

1. Ejemplo 1: sumas de variables binomiales.
2. Ejemplo 2: sumas de variables Poisson.
3. Ejemplo 3: sumas de n puntuaciones de dados.
4. Ejemplo 4: sumas de variables uniformes.
5. Ejemplo 5: sumas de variables exponenciales.

4.3.2 Definición de convergencia en ley

La siguiente definición se ocupa del comportamiento de las sucesiones. Sea $\{X_n\}$ una sucesión de variables aleatorias, y sea $\{F_n\}$ la correspondiente sucesión de funciones de distribución. Se dice que $\{X_n\}$ converge en ley a una variable aleatoria X de función de distribución F si:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{para todo } x \text{ donde } F \text{ es continua.}$$

Se indica que la sucesión converge en ley mediante el símbolo:

$$X_n \xrightarrow{L} X$$

El significado de la definición es que, al aumentar arbitrariamente n , las sucesivas funciones de distribución de la secuencia se aproximan a la distribución F de la variable X .

En los ejemplos se presentan gráficamente algunas situaciones donde diferentes sucesiones de variables aleatorias convergen en ley a una variable aleatoria normal.

4.3.2.1 Representación gráfica de la convergencia

1. Ejemplo 1: primeros elementos de una sucesión de sumas de variables binomiales.
2. Ejemplo 2: primeros elementos de una sucesión de sumas de variables Poisson.
3. Ejemplo 3: primeros elementos de una sucesión de sumas de variables discretas.
4. Ejemplo 4: primeros elementos de una sucesión de sumas de variables uniformes.
5. Ejemplo 5: primeros elementos de una sucesión de sumas de variables exponenciales.

4.3.3 Enunciado del teorema central del límite

A continuación se presenta el enunciado del TCL en la versión de Lindeberg y Lévy. Teorema: Sea X_1, X_2, \dots, X_n , un conjunto de variables aleatorias independientes idénticamente distribuidas, cada una de ellas con función de distribución F , y supongamos que $E(X_k) = \mu$ y $\text{var}(X_k) = \sigma^2$ para cualquier elemento del conjunto. Si designamos a la suma normalizada de n términos con el símbolo:

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

entonces la sucesión de sumas normalizadas converge en ley a la variable aleatoria normal tipificada $Z \sim N(0, 1)$, es decir:

$$S_n^* \xrightarrow{L}$$

El teorema anterior tiene dos importantes corolarios:

1. Si consideramos la suma ordinaria de las n variables aleatorias, es decir, $S_n = X_1 + X_2 + \dots + X_n$, entonces la sucesión de sumas ordinarias converge en ley a una normal de media $n\mu$ y varianza $n\sigma^2$.
2. Si consideramos el promedio de las n variables aleatorias, es decir, $n^{-1}S_n$, entonces la sucesión de promedios converge en ley a una normal de media μ y varianza $n^{-1}\sigma^2$.

4.3.3.1 Comentarios al teorema:

1. La convergencia a la normal tipificada se produce con cualquier tipo de variable que cumpla las condiciones del teorema, sea discreta o absolutamente continua.
2. Un sinónimo para indicar que una sucesión converge en ley a una normal es señalar que es asintóticamente normal.
3. El TCL presenta el comportamiento de sumas infinitas de variables aleatorias. Veremos posteriormente como interpretar el resultado para valores finitos.
4. Existen otras versiones del TCL donde se relajan las condiciones de la versión de Lindeberg y Lévy, que, como se ha visto, obliga a las variables aleatorias a tener idénticas medias y varianzas. Dichas versiones del TCL necesitan el conocimiento de conceptos matemáticos que exceden el nivel al que se orienta Statmedia, y por esta razón se omite su enunciado.

4.3.4 Aplicación del TCL a los ejemplos

- Ejemplo 1: normalidad asintótica de la Binomial.
- Ejemplo 2: normalidad asintótica de la Poisson.
- Ejemplo 3: normalidad asintótica de la suma de puntuaciones de un dado.
- Ejemplo 4: normalidad asintótica de la suma de uniformes.
- Ejemplo 5: normalidad asintótica de la suma de exponenciales.

4.3.5 Casos particulares más notables

Aunque el TCL tiene multitud de casos particulares interesantes, son especialmente relevantes para el desarrollo de los próximos temas los siguientes casos:

4.3.5.1 Promedio de n variables aleatorias

Al considerar n variables independientes, todas con la misma distribución, cada una de ellas con esperanza igual a μ y varianza igual a σ^2 , el promedio es asintóticamente normal con media μ y varianza $n^{-1}\sigma^2$. Este resultado proporciona una distribución asintótica a la media de n observaciones en el muestreo aleatorio simple que se estudiará en el próximo tema.

4.3.5.2 Binomial de parámetros n y p

Es asintóticamente normal con media np y varianza $np(1-p)$. Históricamente (de Moivre, 1733), es el primer resultado demostrado de convergencia a una normal.

4.3.5.3 Poisson de parámetro $n\lambda$

Es asintóticamente normal con media $n\lambda$ y varianza $n\lambda$.

Los dos últimos puntos se han detallado en los ejemplos [1 \(binomial\)](#) y [2 \(Poisson\)](#).

4.3.6 Interpretación del teorema central del límite

El TCL hace referencia a sucesiones infinitas, por tanto, la igualdad de las distribuciones se alcanza sólo en el límite, y hace mención a una distribución final teórica o de referencia.

Sin embargo, puede utilizarse esta distribución final de referencia para aproximar distribuciones correspondientes a sumas finitas. Algunos casos particulares importantes (binomial, Poisson, etc.) alcanzan grados de aproximación suficientes para sumas con no demasiados términos.

Los resultados que se indican a continuación son, por tanto, aproximaciones que se consideran usualmente suficientes, pero conllevan errores numéricos de aproximación.

1. Binomial: aproximar si $n \geq 30$ y $0.1 \leq p \leq 0.9$ a una normal de media np , varianza $np(1-p)$. Ver aquí más detalles.
2. Poisson: aproximar si $\lambda \geq 10$ a una normal de media λ y varianza λ . Ver aquí más detalles.

Para evaluar aproximadamente el error cometido en las aproximaciones, puede consultarse los cuadros gráficos de los ejemplos de este tema.

El TCL permite aproximar funciones de distribución, independientemente del carácter (continuo o discreto) de las variables sumadas. No sirve, por tanto, para aproximar la funciones de densidad discretas por una normal. En el caso continuo sí puede establecerse también una convergencia de las densidades asociadas.

Finalmente, es conveniente mencionar que existen resultados teóricos que permiten estudiar la velocidad de convergencia de una suma de variables aleatorias a la normal, sin embargo la dificultad

técnica que conllevan trasciende el nivel marcado para el conjunto de documentos marcado para Statmedia.

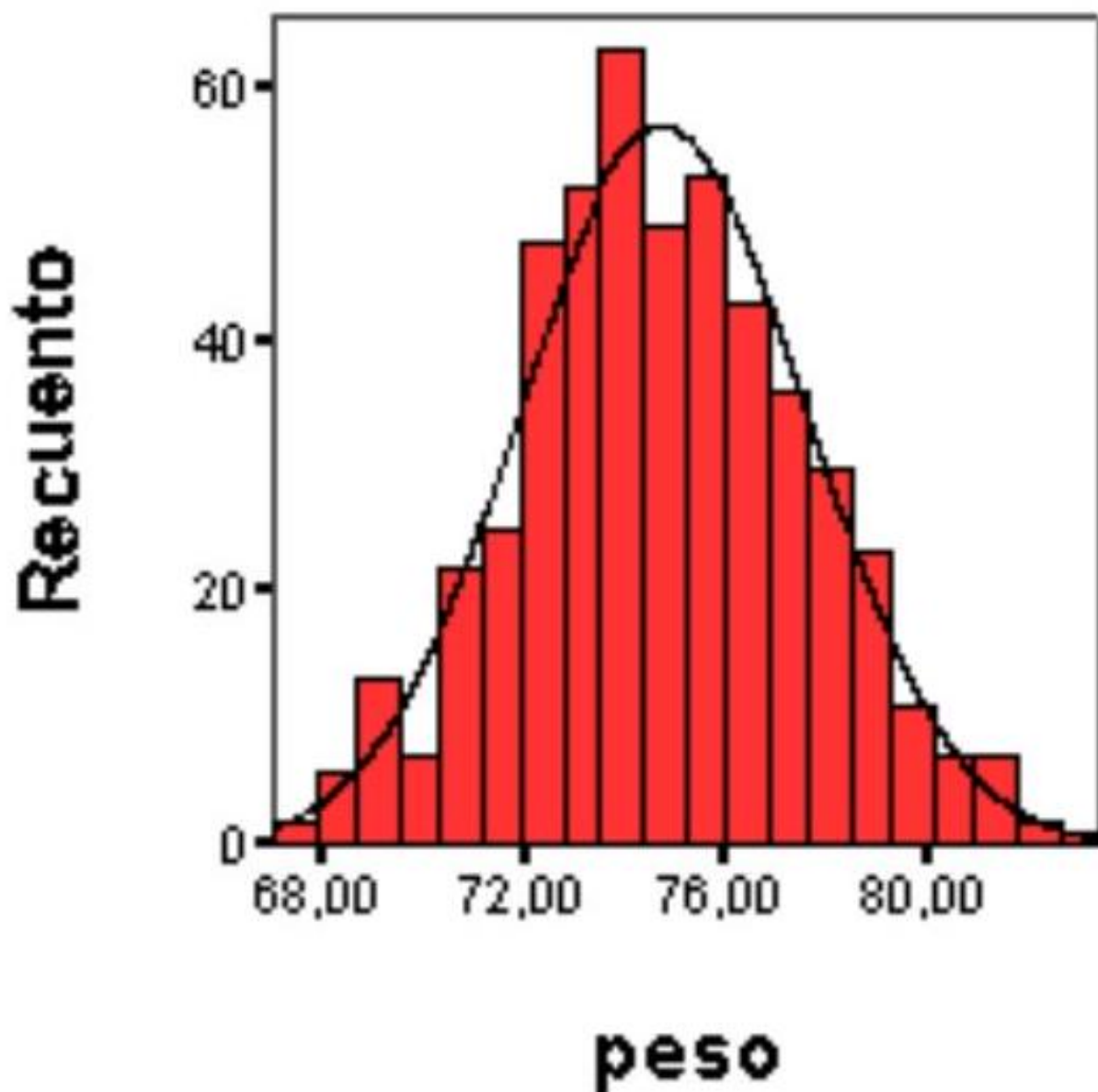
4.3.7 Aproximaciones y errores numéricos

- Ejemplo 1: error en la aproximación de la binomial.
- Ejemplo 2: error en la aproximación de la Poisson.
- Ejemplo 3: error en la aproximación de la suma de puntuaciones de un dado.
- Ejemplo 4: error en la aproximación de la suma de uniformes.
- Ejemplo 5: error en la aproximación de la suma de exponenciales.

4.3.8 Acerca de las variables aproximadamente normales

En general, cuando se estudia en experimentos reales una determinada variable no se conoce su distribución teórica. Sin embargo, puede establecerse su distribución empírica a partir de una muestra más o menos amplia.

Una forma habitual de presentar la distribución empírica es construir el histograma de clases de dicha variable. Es un hecho conocido desde el siglo XIX que esta distribución empírica presenta muchas veces una forma que es aproximadamente normal. Por ejemplo, al realizar un estudio sobre el peso de adultos varones de dieciocho años en Catalunya, se observó la distribución siguiente en la muestra:



El TCL permite dar una explicación a este fenómeno. La variable peso de un adulto viene determinada en cada individuo por la conjunción de multitud de diferentes factores. Algunos de estos factores son ambientales (dietas, ejercicio, enfermedades, etc.) y otros son congénitos. Con el nivel actual de conocimiento no se pueden desglosar completamente todos los factores que intervienen, pero puede aceptarse en cambio que la variable peso es el resultante de la suma de diferentes variables primarias, congénitas o ambientales, y que posiblemente no todas tienen el mismo grado de influencia. Seguramente, estas variables primarias tampoco tienen la misma media, varianza o, incluso, la misma distribución.

La versión del TCL que se ha presentado aquí exige estas condiciones para la convergencia a la normal, pero, como ya se ha comentado antes otras versiones más elaboradas del TCL permiten modelar la suma de variables de forma menos restringida. En este contexto, al considerar la variable peso como una suma más o menos extensa (pero finita) de diferentes variables primarias, es esperable que ocurra que la variable resultante, el peso, siga una distribución aproximadamente normal.

De forma similar es explicable la normalidad aproximada que se observa en muchas variables

biométricas (pesos, alturas, longitudes, concentraciones de metabolitos, distribuciones de edad, etc.) así cómo en muchos otros contextos (distribución de rentas, errores de medición, etc.). A pesar de esta ubicuidad de la distribución normal, el lector no debe inferir que es forzosamente, ni mucho menos, la distribución de referencia en todo estudio aplicado.

5 Distribuciones de probabilidad multidimensionales

En este capítulo se extiende el concepto de variable aleatoria a un conjunto de variables que pueden interpretarse asociadas a un conjunto de medidas distintas y que pueden estar, o no relacionadas.

Tras introducir los conceptos de distribuciones multidimensionales, condicionales y marginales, se pasa a considerar el caso más habitual en inferencia estadística en el que las componentes de los vectores son independientes entre ellas.

Este es, de hecho, el punto de partida de muchos modelos y métodos en estadística.

5.1 Variables aleatorias multidimensionales.

5.2 Distribuciones conjuntas, marginales y condicionales,.

5.3 Valores esperados, covariancia y correlación.

5.4 Independencia de variables aleatorias

5.5 Distribuciones multivariantes: multinomial y normal bivalente.

6 Introducción a la inferencia estadística

Se plantean los problemas que trata la inferencia. Se relaciona con el capítulo anterior a través de la idea del muestreo aleatorio simple y las distribuciones en el muestreo.

Se plantea el problema de la estimación como una forma de aproximación a las características de las distribuciones a partir de muestras aleatorias simples.

Se abordan las distintas formas de construcción de estimadores.

6.1 Los problemas de la inferencia estadística.

6.2 Muestreo y distribuciones en el muestreo.

6.3 La verosimilitud y su papel en la inferencia estadística

6.4 El problema de la estimación. Tipos de estimadores.

6.5 Métodos de obtención de estimadores. Estimadores máximo verosímiles y estimadores bayesianos.

6.6 Propiedades de los estimadores.

7 Estimación por intervalos

Se plantea el problema de la estimación como una forma de aproximación a las características de las distribuciones a partir de muestras aleatorias simples.

Se abordan las distintas filosofías para la construcción de estimadores.

7.1 Preliminares: estimación del error estándar e Introducción al bootstrap

7.2 Estimadores por intervalo: intervalos de confianza

7.3 Intervalos de confianza para características de una población normal (media, varianza),

7.4 Intervalos de confianza bootstrap.

7.5 Intervalos de confianza para proporciones binomiales

7.6 Intervalos de confianza para parámetros en muestra grandes y para casos generales (tasas, OR, ...)

7.7 Aplicaciones: cálculo del tamaño muestral

8 Pruebas de hipótesis

Se plantea el problema de las pruebas de hipótesis. Se discuten las aproximaciones y los conceptos asociados. Se trata el problema de la crisis de la significación.

8.1 Conceptos básicos: pruebas de hipótesis y de significación, pruebas unilaterales y bilaterales, tipos de error, valores críticos de test y p-valores

8.2 Potencia de un test. Cálculos de potencia y de tamaño de la muestra. Tamaño del efecto.

8.3 Métodos de construcción de tests.

8.4 Problemas asociados al uso de tests estadísticos. La crisis de la significación

9 Inferencia Aplicada

Se muestra como deducir y aplicar algunos de los tests mas populares.

9.1 Pruebas de normalidad. Pruebas gráficas. El test de Shapiro-Wilks

**9.2 Pruebas de hipótesis para contrastar variables cuantitativas:
pruebas paramétricas t-test y Anova**

**9.3 Pruebas de hipótesis para contrastar variables cuantitativas:
pruebas de hipótesis no paramétricas de Wilcoxon y Kruskal-Wallis**

**9.4 Contrastes para datos categóricos. Pruebas binomiales, ji
cuadrado y test de Fisher.**

9.5 Riesgo relativo y razón de «odds»

10 Computación Intensiva y *Multiple Testing*

Se introducen distintos métodos cuyo nexo común es la computación intensiva.

10.1 Tests de permutaciones; ¿Qué?, ¿Cuándo?, ¿Cómo?

10.2 El bootstrap en contraste de hipótesis

10.3 El problema de las comparaciones múltiples

10.4 Métodos de control de error: FWER y FDR

Referencias