



Métodos Estadísticos de la Ingeniería

Mathieu Kessler

- Los métodos estadísticos persiguen extraer de la manera más eficiente posible la información presente en conjuntos de datos. Su uso se ha generalizado en todos los campos de la ingeniería y son varias las asignaturas aplicadas en las titulaciones de Ingeniero Industrial o Ingeniero Técnico Industrial que presuponen por parte del alumno el manejo básico de técnicas estadísticas sencillas. Este manual recorre el camino desde la exploración previa de un conjunto de datos, la formulación de un modelo aleatorio para el mecanismo de generación de éstos, hasta la introducción a las técnicas de inferencia que formalizan el carácter significativo o no de las conclusiones que se puede extraer de los datos resultantes de un experimento. Se ha optado por una presentación intuitiva de los conceptos intentando en la medida de lo posible relacionarlos con la experiencia práctica o el sentido común...
- Mathieu Kessler es Catedrático de Universidad en el área de Estadística e Investigación Operativa en el Departamento de Matemática Aplicada y Estadística de la Universidad Politécnica de Cartagena. Es doctor en Estadística por la Universidad de Paris VI y autor de numerosas publicaciones tanto sobre estadística teórica como sobre aplicaciones de la estadística en revistas internacionales.

Autor: Mathieu Kessler
ISBN: 978-84-96997-07-3
D. Legal: MU-1949-2008

*Si el año 2013
es el Año Internacional
de la Estadística*



*¿significa que aumentan
tus probabilidades
de aprobar?*

Prólogo

Este manual se ha desarrollado a partir de los apuntes que el autor usa como soporte para sus clases en la asignatura de “Métodos Estadísticos de la Ingeniería” que ha impartido en numerosas ocasiones, tanto en la titulación de Ingeniero Industrial como en la de Ingeniero Técnico Industrial de varias especialidades.

Se pueden destacar tres objetivos para esta asignatura:

- capacitar al alumno para extraer, resumir y comunicar información a partir de conjuntos de datos experimentales,
- proporcionarle las herramientas para la construcción de modelos para estos datos a través de variables aleatorias,
- finalmente, introducir al alumno en los conceptos de la inferencia estadística, permitiéndole llegar a conclusiones significativas a partir de una muestra.

El temario de la asignatura recorre estos objetivos, escogiendo deliberadamente una presentación intuitiva de los conceptos e intentando en la medida de lo posible relacionarlos con la experiencia práctica o el sentido común de los alumnos. En la primera parte, se pone especial énfasis en el uso intensivo de gráficas para la exploración de datos.

Quiero dar las gracias aquí en primer lugar, a todos los alumnos que he tenido y que, por sus preguntas y dudas me han obligado a precisar y pulir mis explicaciones, e incluso mi propia comprensión de los conceptos. Muchos de ellos me han regalado su atención, su ilusión, su interés, y por ello, me han hecho disfrutar con mi trabajo.

También estoy agradecido a mis compañeros del área de Estadística e I.O y del Departamento de Matemática Aplicada y Estadística, por contribuir a crear un ambiente de trabajo agradable y estimulante, así como a la Escuela Técnica Superior de Ingenieros Industriales de la UPCT por su apoyo en una primera edición de este manual.

Finalmente dedico este libro a Graci, Quique y David por ser la alegría de mis días, por su admiración ciega y en absoluto fundamentada en sus conocimientos estadísticos, y por estar siempre allí cuando vuelvo a casa...

Prólogo a la segunda edición (2013)

En esta segunda edición, realizada por la celebración en 2013 del Año Internacional de la Estadística, se han corregido erratas, actualizado algunos conjuntos de datos, e incluido como cabecera de cada capítulo un curiosidad destinada a ilustrar el papel omnipresente de la estadística en el funcionamiento de nuestras sociedades.

Índice general

I	Exploración de datos	1
I.1	Introducción	1
I.2	Unos cuantos términos	1
I.3	Tabulación y representaciones gráficas	2
I.3.1	Gráficas para variable cualitativa	2
I.3.2	Gráficas para una variable cuantitativa	3
I.4	Medidas numéricas	8
I.4.1	Medidas de centro	9
I.4.2	Medidas de dispersión	10
I.4.3	Un resumen gráfico: el diagrama de caja-bigotes	11
I.5	Ajuste por mínimos cuadrados	12
I.5.1	Planteamiento	12
I.5.2	Criterio de mínimos cuadrados	13
I.5.3	Casos concretos	16
II	Fundamentos de la teoría de la probabilidad.	27
II.1	Conceptos básicos relacionados con un experimento	27
II.1.1	Experimento aleatorio	27
II.1.2	Suceso elemental	28
II.1.3	Espacio muestral	28
II.1.4	Suceso	28
II.1.5	Diagrama de Venn	29
II.1.6	Leyes de Morgan	29
II.2	Concepto de Probabilidad	30
II.2.1	Definición informal de la probabilidad - propiedades.	30
II.2.2	El caso de un espacio muestral finito y la definición de Laplace.	31
II.3	La probabilidad condicionada.	33
II.3.1	Definición	33
II.3.2	Regla del producto.	34
II.3.3	Propiedad	34
II.4	Sucesos independientes	34
II.4.1	Definición para dos sucesos	34
II.4.2	Definición para n sucesos	35
II.5	Ejemplos de probabilidades condicionadas en la vida diaria	35
II.5.1	Eslogan publicitario para la lotería	35
II.5.2	Tabaquismo y cáncer de pulmón	35
II.5.3	Tabaquismo y esperanza de vida	36
II.6	Fórmula de la probabilidad total y teorema de Bayes	36

II.6.1	Condiciones de aplicación	36
II.6.2	Los resultados	36
II.6.3	Ejemplo	37
III	Variable aleatoria I	41
III.1	Concepto de variable aleatoria	41
III.1.1	Definición	41
III.1.2	Distribución de una variable aleatoria	42
III.2	Función de distribución de una v.a	43
III.2.1	Definición	43
III.2.2	Cálculo para el ejemplo de las tres monedas	43
III.2.3	Propiedades	43
III.3	Variable aleatoria discreta	44
III.3.1	Definición	44
III.3.2	Función puntual de probabilidad	44
III.3.3	Características de una variable discreta	45
III.3.4	Modelos más usados de v.a. discretas	47
III.4	Variable continua	51
III.4.1	Definición	51
III.4.2	Función de densidad	51
III.4.3	Medidas numéricas asociadas a una v.a continua	54
III.4.4	Modelos más comunes de v.a continua	56
III.5	Algunas propiedades útiles de la esperanza y la varianza	63
IV	Variable Aleatoria II	67
IV.1	Introducción	67
IV.2	Variable bidimensional discreta	67
IV.2.1	Función puntual de probabilidad conjunta	68
IV.2.2	Esperanza	69
IV.3	Variable bidimensional continua	69
IV.3.1	Función de densidad conjunta	69
IV.3.2	Esperanza	71
IV.4	Distribuciones condicionadas	71
IV.4.1	V.a bidimensional discreta	71
IV.4.2	Para una v.a bidimensional continua	72
IV.4.3	Esperanza condicionada	73
IV.5	Variables independientes	73
IV.5.1	Definición	73
IV.5.2	Consecuencias prácticas	74
IV.6	Medidas numéricas para una v.a bidimensional	74
IV.6.1	Definiciones	74
IV.6.2	Propiedades	76
IV.7	Algunos modelos de v.a. multidimensional	76
IV.7.1	Modelo multinomial	76
IV.7.2	El modelo Normal multidimensional	77
V	Muestreo y distribuciones muestrales	83
V.1	Introducción	83

V.2	Muestra	87
V.3	La media muestral	87
	V.3.1 Esperanza y varianza de \bar{X}	87
	V.3.2 Distribución de la media muestral	89
V.4	La varianza muestral	90
V.5	Distribución t de Student	90
V.6	La proporción muestral	92
	V.6.1 Cálculos exactos para la distribución de \hat{p}	93
	V.6.2 Distribución aproximada de \hat{p}	93
V.7	Introducción a las gráficas de control	94
	V.7.1 Gráfica de control \bar{X}	94
	V.7.2 Gráfica de control \hat{p}	95
	V.7.3 Otra señal de alarma	96
VI	Introducción a la teoría de la estimación	99
VI.1	Introducción	99
VI.2	Estimación puntual	99
	VI.2.1 Definición	99
	VI.2.2 Propiedades deseables para un estimador	100
	VI.2.3 Métodos de construcción de estimadores	101
VI.3	Estimación por intervalos	105
	VI.3.1 Idea básica	105
	VI.3.2 Intervalo de confianza para la media μ de una distribución Normal con varianza conocida	105
	VI.3.3 Comentarios importantes	108
	VI.3.4 Determinación del tamaño muestral	109
VII	Introducción a los contrastes de hipótesis	113
VII.1	Introducción	113
VII.2	Planteamiento general	114
	VII.2.1 Hipótesis estadística	114
	VII.2.2 Regla de decisión	114
	VII.2.3 Evaluación del error	114
	VII.2.4 Procedimiento	115
VII.3	Contraste de hipótesis para la media μ de una distribución Normal con varianza conocida.	116
	VII.3.1 Hipótesis bilateral	116
	VII.3.2 Hipótesis unilateral	117
	VII.3.3 Ejemplos	118
VII.4	Concepto de p-valor	119
VII.5	Potencia del test	120
	VII.5.1 Definición	120
	VII.5.2 Cálculo de la potencia	121
	VII.5.3 Ejemplo de cálculo de la potencia	122
	VII.5.4 Factores que influyen la potencia	123
VII.6	Inferencia para la media	123
	VII.6.1 Contraste de hipótesis para la media μ de una distribución Normal con varianza desconocida	123

VII.7 Inferencia para dos medias	125
VII.7.1 Estadísticos muestrales	126
VII.7.2 Intervalos y contrastes	127



2013: Año Internacional de la Estadística. ¿Sabías qué...?

La creación de la Comisión de Estadística del Reino el 3 de noviembre de 1856 marca el comienzo de la estadística oficial en España, siendo su primer trabajo el Censo de Población que se realizó en mayo de 1857. En diciembre de 1945 se crea el Instituto Nacional de Estadística que tiene como misión la elaboración de las estadísticas demográficas, económicas y sociales del país.

Fuente: Instituto Nacional de Estadística, <http://www.ine.es> (Breve reseña histórica)

statistics

Exploración de datos

I.1. Introducción

La estadística utiliza datos para conseguir comprensión sobre un fenómeno. Básicamente, esta comprensión es una consecuencia de la combinación entre conocimientos previos sobre el fenómeno y nuestra capacidad para utilizar gráficos y cálculos para extraer información de los datos.

En contextos industriales se recogen a menudo grandes conjuntos de datos correspondientes a un gran número de variables. Un efecto contradictorio aparece: por una parte, cuanto más datos, más información podemos extraer sobre las variables de interés, pero a la vez es más difícil su extracción.

En este contexto aparece una primera etapa fundamental frente a un conjunto de datos: la *exploración*, que se realiza a través de representaciones gráficas y del cálculo de unas cuantas medidas numéricas bien escogidas.

Para tener las ideas claras, unos cuantos gráficos pueden proporcionarnos información más valiosa que procedimientos sofisticados que no dominamos. En esta asignatura, veremos en temas posteriores métodos más sofisticados de análisis pero dedicamos ahora un capítulo a recordar las técnicas elementales con el objetivo de fomentar reacciones sanas frente a un conjunto de datos.

Aun cuando el conjunto de datos presenta varias variables, se debe empezar por el estudio individual de cada una.

I.2. Unos cuantos términos

- Un conjunto de datos describe **individuos**, que pueden ser personas pero también objetos. Por ejemplo, asociados a esta clase, podemos considerar que los individuos son los alumnos.
- Consideramos variables asociadas a este conjunto de datos, distinguiremos entre **variable cuantitativa**, que asocia un número a cada individuo, o **variable cualitativa**, que coloca cada individuo en una categoría. Ejemplos de

variables cuantitativas asociadas a la clase: peso, altura o edad. El sexo o el grupo sanguíneo son en cambio variables cualitativas.

- Un concepto fundamental que utilizaremos con frecuencia corresponde a la **distribución** de una variable X asociada a un conjunto de datos. Describir la distribución de X corresponde a establecer la lista de los valores que toma X junto con la frecuencia con la que toma cada valor. Hablaremos de **frecuencia absoluta** de un valor para denotar el número de veces que aparece este valor en el conjunto de datos, mientras que la **frecuencia relativa** corresponde a la proporción (o el porcentaje) de veces que aparece este valor.

En particular, una de las características interesantes de un conjunto de datos consiste en determinar si presenta mucha o poca **variabilidad**.

Ejemplo I.2.1 Consideremos por ejemplo la distribución del grupo sanguíneo en una clase presentada en la tabla siguiente:

Grupo	Frec. absoluta	Frec. relativa
A	51	$51/145=0.35$
B	19	0.13
O	5	0.03
AB	70	0.49

¿Qué representa la suma de la segunda columna (Frec. absoluta)? ¿Cuanto vale la suma de la tercera columna?

I.3. Tabulación y representaciones gráficas

Las representaciones gráficas son una herramienta fundamental para extraer información de forma visual de un conjunto de datos. Pueden ser mucho más útiles que procedimientos sofisticados que uno no domina...

I.3.1. Gráficas para variable cualitativa

Para un conjunto de datos descritos por una variable cualitativa, podemos realizar dos tipos de gráficas:

I.3.1.1. Diagrama de barras

Para cada valor que toma la variable en el conjunto y que indicamos en el eje horizontal, representamos en el eje vertical su frecuencia absoluta o relativa, en forma de una barra. En el caso del ejemplo I.2.1, obtenemos el diagrama de barra de la figura I.1. Cabe destacar que se suelen ordenar los valores de la variable por orden decreciente de frecuencias.

I.3.1.2. Diagrama de sectores

Si el conjunto no presenta demasiados valores distintos, también podemos utilizar el diagrama de sectores, donde cada valor ocupa un sector circular cuya área es proporcional a su frecuencia.

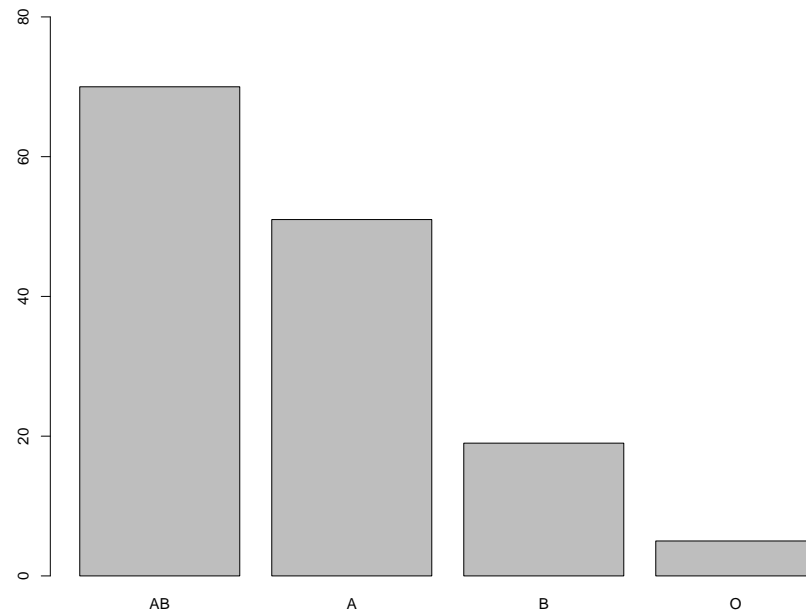


Figura I.1: Diagrama de barras, frecuencias absolutas, para el ejemplo I.2.1 del grupo sanguíneo,

Para el ejemplo I.2.1, calculemos el ángulo que ocupará el sector para cada uno de los valores AB, A, B, O. Por una regla de tres, deducimos que si el círculo entero (360 grados) representará el número total de datos en el conjunto, es decir 145 individuos, el valor AB con una frecuencia de 70 individuos deberá ocupar un sector de $70/145 \times 360 = 174^\circ$. Asimismo, el valor A ocupará 126° , el valor B 48° , mientras que el valor O ocupará solamente 12° . El diagrama de sectores correspondiente se representa en la figura I.2.

I.3.2. Gráficas para una variable cuantitativa

Nos centramos ahora en variables cuantitativas. Los conjuntos que examinaremos se presentarán o bien en forma bruta: un fichero con una columna para cada variable, donde cada fila representa un individuo, o bien en forma ya tabulada, es decir donde los datos están agrupados.

Para datos agrupados, consideremos mediciones del contenido en nitrato de una muestra de agua:

Valor	Frecuencia	Valor	Frecuencia
0.45	1	0.49	8
0.46	2	0.50	10
0.47	4	0.51	5
0.48	8	0.51	8

También se puede representar gráficamente mediante un diagrama de barras esta distribución de frecuencias, indicando en el eje Ox los valores que puede tomar la

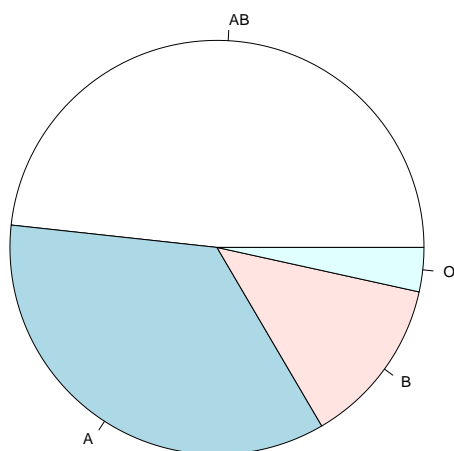


Figura I.2: Diagrama de sectores para el ejemplo I.2.1 del grupo sanguíneo, variable y en el eje Oy sus frecuencias. Obtenemos así un diagrama de barras en el ejemplo de las mediciones de la concentración en nitrato, ver figura I.3.

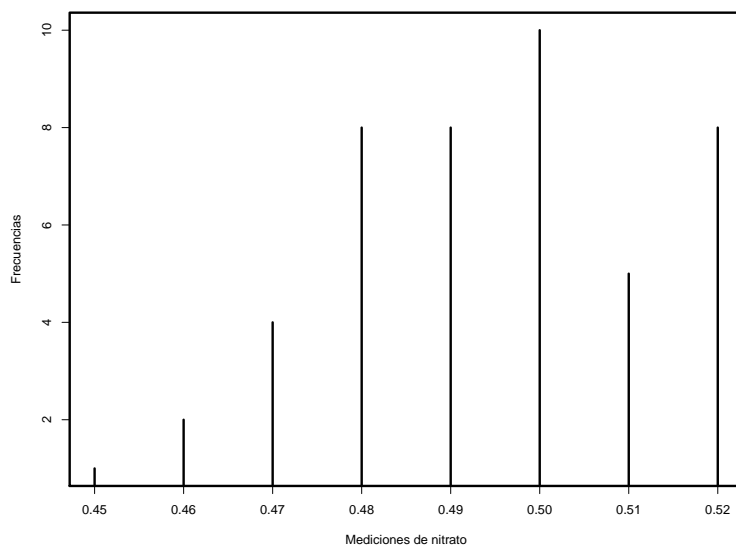


Figura I.3: Diagrama de barras para las concentraciones de nitrato

En el caso en que el conjunto presente muchos valores próximos pero distintos,

agrupamos los datos por clases, tal como lo veremos en los apartados siguientes.

I.3.2.1. Ejemplo: mediciones de la velocidad de la luz

Consideramos para ilustrar los conceptos que introduciremos en el resto del tema el conjunto de datos de Newcomb (<http://www.dmae.upct.es/~mathieu>). Newcomb fue el primero en conseguir ¡en 1882! una estimación bastante precisa de la velocidad de la luz. Las mediciones recogidas a continuación corresponden a los tiempos codificados que tardó un rayo de luz en recorrer el camino de ida y vuelta desde el laboratorio de Simon Newcomb situado en el Río Potomac hasta un espejo situado en la base del “Washington Monument”, en total una distancia de 7400m. Para obtener los tiempos en nano segundos ($10^{-9}s$) no codificados, hay que añadir 24800 a cada dato.¹

Tiempos codificados: 28, 26, 33, 24, 34, -44, 27, 16, 40, -2, 29, 22, 24, 21, 25, 30, 23, 29, 31, 19, 24, 20, 36, 32, 36, 28, 25, 21, 28, 29, 37, 25, 28, 26, 30, 32, 36, 26, 30, 22, 36, 23, 27, 27, 28, 27, 31, 27, 26, 33, 26, 32, 32, 24, 39, 28, 24, 25, 32, 25, 29, 27, 28, 29, 16, 23

Al observar estos datos, podemos realizar dos comentarios:

1. ¿Por qué Newcomb repitió tantas veces las mediciones, y no se limitó a realizar el experimento una vez? Porque los datos resultados del experimento presentan una cierta variabilidad: por mucho que haya intentado controlar las condiciones experimentales para mantenerlas constantes, el resultado es imprevisible. La medición está siempre perturbada por un “ruido” incontrolable...
2. ¿Qué hacer con estos datos? A la vista de estos datos, ¿cuál es el valor que podríamos tomar como la velocidad de la luz? Debemos encontrar un valor que sea representativo de las 66 mediciones realizadas. Se suele escoger la media, pero para asegurarnos de que ésta es representativa del conjunto, es útil establecer la tabla de frecuencias y visualizar el conjunto a través de un histograma, tal como lo vemos en la sección siguiente...

I.3.2.2. Tabla de frecuencias y histograma

En el caso en que el conjunto presente muchos valores próximos pero distintos, empezamos por agrupar los datos por clases: ordenamos los datos por orden creciente, dividimos el rango de los valores en clases de igual amplitud, y colocamos cada dato en la clase que le toca. A continuación podemos realizar el recuento de las frecuencias de cada clase.

¿Cuántas clases escoger? *La elección del número de clases es un problema que no admite una solución perfecta que sirva para todos los conjuntos de datos. Una regla aproximada llamada regla de Sturges consiste en escoger $1 + \log_2(n)$ clases para un conjunto con n datos.*

Para le ejemplo de las mediciones de Newcomb, los datos ordenados se presentan como:

¹Fuente: Moore, David S. and McCabe, George P. (1989). Introduction to the Practice of Statistics, W. H. Freeman and Company: New York, NY, pp 3-16.

Pos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dato	-44	-2	16	16	19	20	21	21	22	22	23	23	23	24	24
Pos	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Dato	24	24	24	25	25	25	25	25	26	26	26	26	26	27	27
Pos	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Dato	27	27	27	27	28	28	28	28	28	28	28	29	29	29	29
Pos	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Dato	29	30	30	30	31	31	32	32	32	32	32	33	33	34	36
Pos	61	62	63	64	65	66									
Dato	36	36	36	37	39	40									

Utilizamos por ejemplo clases de amplitud 5 empezando en -45 y acabando en 40, y realizamos el recuento de las frecuencias de cada clase:

Clase	Frecuencia	Clase	Frecuencia	Clase	Frecuencia
] - 45, -40]	1] - 15, -10]	0]15, 20]	4
] - 40, -35]	0] - 10, -5]	0]20, 25]	17
] - 35, -30]	0] - 5, 0]	1]25, 30]	26
] - 30, -25]	0]0, 5]	0]30, 35]	10
] - 25, -20]	0]5, 10]	0]35, 40]	7
] - 20, -15]	0]10, 15]	0		

Cuando establecemos la tabla de frecuencias de una variable cuantitativa, indicamos también las **frecuencias acumuladas** de cada clase: la frecuencia absoluta (relativa) acumulada de una clase es el número (proporción) de datos que pertenecen a esta clase o a alguna clase anterior.

La tabla completa de frecuencias tal como nos la suele presentar un programa de estadística incluye las frecuencias absolutas y relativas así como las frecuencias acumuladas absolutas y relativas. Para el ejemplo de las mediciones de Newcomb, la tabla completa se puede ver en la Tabla I.1 más abajo.

Por otra parte, los datos tabulados se examinan con más comodidad a través de representaciones gráficas. En el eje Ox aparecen las clases y en el eje Oy las frecuencias, el diagrama resultante se llama histograma. En la figura I.4, aparece el histograma para las mediciones de Newcomb. Se pueden representar histogramas de frecuencias absolutas, relativas, absolutas acumuladas o relativas acumuladas.

I.3.2.3. Cómo interpretar un histograma

Las representaciones gráficas describen la distribución de la variable en el conjunto. Al examinarlos hay que intentar contestar a las siguientes preguntas, para resumir las características de la distribución.

1. ¿ Es el histograma simétrico? Es decir, ¿aparece un punto central, respecto al cual, los valores se van repartiendo de manera aproximadamente simétrica? Esta es la situación clásica para un conjunto de mediciones: el valor central sería lo más representativo de lo que intentamos medir, y las mediciones van sobrevalorando e infravalorando de manera simétrica este valor. Si no consideramos los valores -44 y -2 en el conjunto de Newcomb, por ser muy diferentes

Clase	Frecuencias		Frec. Acumuladas	
	Absolutas	Relativas(%)	Absolutas	Relativas(%)
] - 45, -40]	1	1.5	1	1.5
] - 40, -35]	0	0.0	1	1.5
] - 35, -30]	0	0.0	1	1.5
] - 30, -25]	0	0.0	1	1.5
] - 25, -20]	0	0.0	1	1.5
] - 20, -15]	0	0.0	1	1.5
] - 15, -10]	0	0.0	1	1.5
] - 10, -5]	0	0.0	1	1.5
] - 5, 0]	1	1.5	2	3.0
]0, 5]	0	0.0	2	3.0
]5, 10]	0	0.0	2	3.0
]10, 15]	0	0.0	2	3.0
]15, 20]	4	6	6	9
]20, 25]	17	25.7	23	34.7
]25, 30]	26	39.3	49	74
]30, 35]	10	15.3	59	89.3
]35, 40]	7	10.7	66	100
TOTAL	66	100.0		

Tabla I.1: Tabla de frecuencias, mediciones de Newcomb.

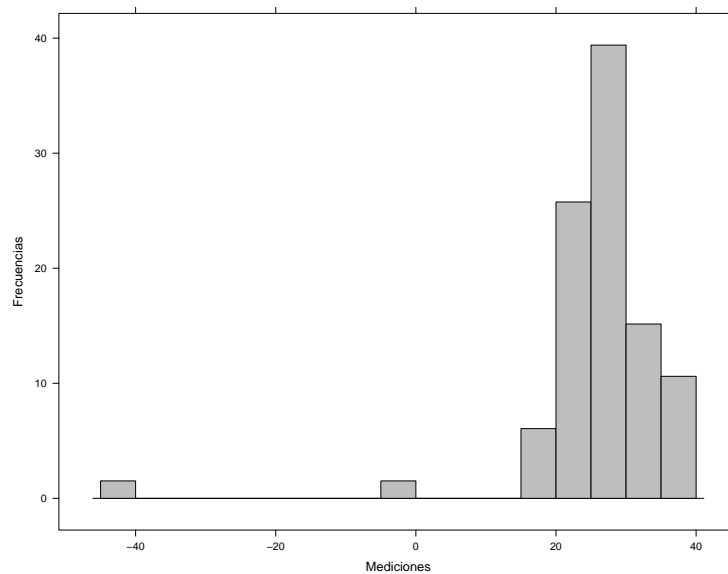


Figura I.4: Histograma para las mediciones de Newcomb

del resto del conjunto, podemos decir que la distribución de las mediciones es aproximadamente simétrica.

2. ¿Posee la distribución colas largas?

3. ¿Posee el histograma un máximo claro único? En este caso hablamos de histograma unimodal.
4. ¿Aparecen datos atípicos?, es decir datos que se alejan del patrón global de los datos. Para el conjunto de Newcomb, dos datos aparecen claramente atípicos: -44 y -2, mientras que las 64 mediciones restantes están entre 15 y 40. Al detectar datos atípicos, debemos comprobar que no se deban a errores tipográficos, y buscar si están asociados a unas circunstancias experimentales especiales. Podremos entonces decidir corregirlos u omitirlos del estudio.
5. ¿Donde localizamos el centro aproximado de los datos?
6. ¿Presentan los datos mucha dispersión?, lo que se traduce en la forma puntiaguda o chata del histograma. En el caso de mediciones, el hecho de que los datos estén concentrados revela que se consiguió una buena regularidad en el proceso de medición...

En la figura I.5, presentamos varios patrones de histogramas.

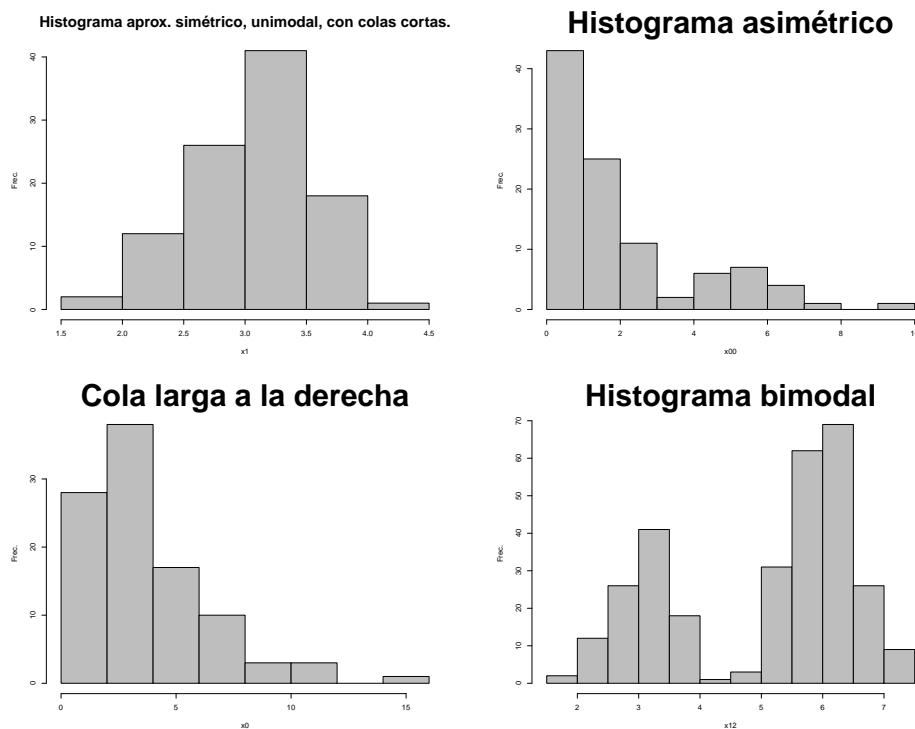


Figura I.5: Distintos patrones de histogramas.

I.4. Medidas numéricas

Para variables cuantitativas, se suele acompañar las representaciones gráficas de las distribuciones con medidas numéricas que proporcionen un resumen de sus características principales. Existen medidas numéricas para contestar a cada pregunta

(y alguna más...) planteadas en el apartado anterior a la hora de examinar el histograma. Nos limitaremos a las medidas de centro y de dispersión, es decir las que proporcionen una respuesta a las preguntas 5 y 6.

I.4.1. Medidas de centro

Buscamos ahora medidas numéricas que sean representativas del centro del conjunto de dato.

I.4.1.1. La media:

Si x_1, \dots, x_n son los datos, sabemos todos que la media es

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}.$$

En el caso en que los datos ya están tabulados y tenemos los valores distintos x_1, \dots, x_m junto con sus frecuencias n_1, \dots, n_m , deberemos tener en cuenta estas frecuencias para el cálculo de la media:

$$\bar{x} = \frac{n_1x_1 + \dots + n_mx_m}{(n_1 + \dots + n_m)}.$$

En este caso, ¿cuántos individuos tenemos en el conjunto?

Nota: Representa el centro de gravedad de los datos, es decir que si a cada dato le damos un peso unidad, la media representa el punto en el que el conjunto está en equilibrio.

En particular, deducimos que la media es muy sensible a datos atípicos en el conjunto de datos: si añadido un dato (peso) alejado del centro de gravedad, el punto de equilibrio debe desplazarse mucho hacia éste para que se mantenga el equilibrio.

Para paliar estos inconvenientes, se considera también la mediana:

I.4.1.2. La mediana:

La mediana es el punto que deja el 50% de los datos a su izquierda y el otro 50% a su derecha. Es una medida de centralización más adecuada que la media en el caso en que la distribución de los datos es asimétrica (lo que se ve en el histograma) o si hay datos atípicos. Si la distribución es simétrica, la media y la mediana coinciden.

Para calcular la mediana de un conjunto de n datos, x_1, x_2, \dots, x_n , empiezo por ordenar los datos por orden creciente. La mediana es el dato ordenado $n^\circ (n+1)/2$.

Ejemplo: 125, 129, 134, 185, 200. La mediana es el dato ordenado $n^\circ 3$, y es igual a 134.

11, 15, 20, 23: la mediana es el dato ordenado $n^\circ 2.5$, que tomamos por convención igual al punto medio entre el dato $n^\circ 2$ y el dato $n^\circ 3$. En este caso, la mediana es igual a 17.5.

La mediana no es sensible a datos atípicos, para convencerse de ello, se puede considerar el ejemplo anterior donde se sustituye el valor 23 por 1000... La mediana no cambia... Por lo tanto, la mediana es más representativa del centro del conjunto si hay algún dato atípico o si la distribución es algo asimétrica...

I.4.2. Medidas de dispersión

I.4.2.1. La desviación típica

Mide lo lejos que están situados los datos respecto de su centro de gravedad, la media. Empezamos por definir **la varianza**:

$$s^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}, \quad (\text{I.1})$$

que representa aproximadamente el promedio de las distancias al cuadrado entre los datos y su media. La desviación típica s es la raíz cuadrada de s^2 .

Para calcularla en la práctica se suele preferir la fórmula siguiente

$$s^2 = \frac{n}{n - 1}(\overline{x^2} - (\bar{x})^2), \quad (\text{I.2})$$

donde $\overline{x^2}$ representa la media de los datos que hemos previamente elevado al cuadrado, mientras que $(\bar{x})^2$ representa el cuadrado del valor de la media. Como ejemplo, supongamos que quiero calcular la varianza de los datos siguientes 4, 5,5, 6,5, 8.

Necesito por una parte \bar{x} , que calculo como $\bar{x} = (4 + 5,5 + 6,5 + 8)/4 = 6$, y por otra parte $\overline{x^2}$ que calculo como $\overline{x^2} = (4^2 + 5,5^2 + 6,5^2 + 8^2)/4 = 38,125$. Por lo tanto, deduzco

$$s^2 = \frac{4}{3}[38,125 - (6)^2] = 2,8333.$$

Naturalmente, la desviación típica es representativa de la dispersión del conjunto de datos solo si la media es representativa de su centro.

Es bueno ser consciente de que la desviación típica, al igual que la media, se expresa en las mismas unidades que los datos, mientras que la varianza en *(unidades)²*.

Una medida alternativa de dispersión que puede ser más representativa en el caso en que la distribución es asimétrica o en presencia de datos atípicos, es el **rango intercuartílico**.

I.4.2.2. El rango intercuartílico (RIC)

Hemos definido la mediana como el punto que separa el conjunto en dos partes de mismo tamaño. Definimos de la misma manera los cuartiles como los puntos que separan el conjunto en cuatro partes de mismo tamaño. El primer cuartil Q_1 deja el 25% de los datos ordenados a su izquierda, y el otro 75% a su derecha, mientras que el tercer cuartil Q_3 deja el 75% de los datos ordenados a su izquierda, y el otro 25% a su derecha. Por lo tanto el par (Q_1, Q_3) nos proporciona información sobre la dispersión presente en los datos: cuanto más alejados estén los cuartiles, más dispersos están los datos. Por ello, calculamos el rango intercuartílico RIC como la diferencia entre Q_3 y Q_1 .

Para calcular los cuartiles, empezamos por calcular la mediana Me de los datos. El primer cuartil es la mediana del grupo de datos que queda a la izquierda de Me (Me excluida), mientras que el tercer cuartil se calcula como la mediana del grupo que queda a su derecha (Me excluida).

El RIC también se utiliza para detectar datos atípicos:

Regla: Se consideran como atípicos los datos que son menores de $Q_1 - 1,5 \times RIC$, o mayores de $Q_3 + 1,5 \times RIC$.

I.4.3. Un resumen gráfico: el diagrama de caja-bigotes

El diagrama de caja-bigotes es un resumen gráfico que permite visualizar, para un conjunto de datos, la tendencia central, la dispersión y la presencia posible de datos atípicos. Para realizarlo se necesita calcular la mediana, el primer cuartil, y el tercer cuartil de los datos:

El diagrama de caja-bigotes presenta de manera gráfica estas informaciones, tal como está recogida en la figura I.6.

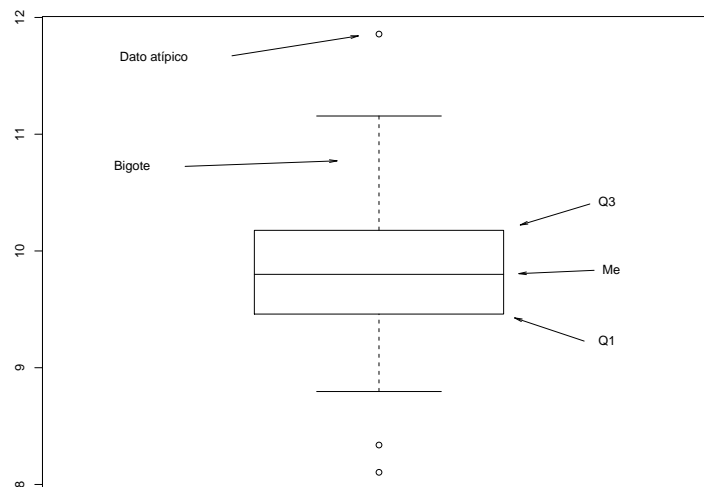


Figura I.6: Componentes del diagrama caja-bigotes

Los segmentos 1.5 RIC (llamados bigotes) se recortan hasta : el dato del conjunto inmediatamente superior a $Q_1 - 1,5 \times \text{RIC}$ para el bigote inferior, y el dato inmediatamente inferior a $Q_3 + 1,5 \times \text{RIC}$, para el bigote superior.

La mayor utilidad de los diagramas caja-bigotes es para comparar dos o más conjuntos de datos.

Ejemplo

La puntuación de los equipos de la liga española al final de las temporadas 01/02 y 02/03 en primera división se pueden comparar con un diagrama caja-bigotes, como aparece en la figura I.7

Comentarios: No hay datos atípicos, es decir que no hay equipo que se haya destacado por arriba o por abajo del resto de los equipos. Hay más diferencia de puntos entre el primer y el último clasificado para la liga 02/03 que en la liga anterior. Los equipos del tercer cuarto de la clasificación están muy apretados en la liga 02/03.

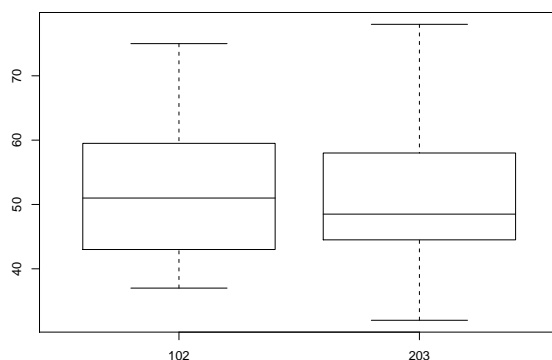


Figura I.7: Comparación puntuación final, temporadas 01/02 y 02/03

I.5. Ajuste por mínimos cuadrados

I.5.1. Planteamiento

Es muy normal considerar más de una variable asociada a un experimento. En este caso, más que la distribución de cada variable por separado, nos puede interesar en particular las relaciones que existan entre ellas. Nos centraremos aquí en el caso en que distinguimos una variable llamada “respuesta”, cuya amplitud depende de los valores de otras variables llamadas “explicativas”, y aprenderemos cómo deducir un modelo para la evolución de la primera en función de estas últimas.

Hay dos utilidades principales al disponer de un modelo: podemos primero explicar la manera en la que cambios en los valores de una variable explicativa induce cambios en el valor de la variable respuesta. Por ejemplo, si pienso que la temperatura media Y en agosto en San Javier evoluciona en función del año según el modelo:

$$\text{Temperatura} = -582,5 + 0,31\text{año},$$

deduciré que en promedio, la temperatura media en agosto aumenta de 0.3 grados cada año.

Por otra parte, si dispongo de un modelo para la evolución de la variable respuesta, me permite también realizar predicciones del valor que tomará para valores de las explicativas que no hemos observado.

Acabamos esta sección de presentación con cuatro ejemplos con datos reales tomados de campos diferentes. Las nubes de puntos correspondientes están presentadas en la figura I.8

- Estudio de la resistencia del cemento en función del tiempo de fraguado en días. Fuente: Hald, A. (1952) *Statistical theory for engineering applications*, Wiley & Sons New-York, pág 541. ¿Cómo evoluciona la resistencia de piezas de cemento en función del tiempo de fraguado? ¿Cuánto tiempo hay que esperar para conseguir el 90 % de la resistencia máxima? Este es el tipo de preguntas a las que podemos contestar con el estudio de este conjunto de datos.

- Todos los años Venecia se inunda durante las “acqua alta”. Sin embargo, parece que el nivel máximo al que llega el mar está cada año más alto, haciendo temer por la conservación de la ciudad y de sus monumentos. Es por lo tanto de interés estudiar la evolución del nivel máximo del mar en función del año. Fuente: Smith, R.L (1986) “Extreme value theory based on the r largest annual events”, *Journal of Hydrology*, **86**.
- Evolución de la producción mundial de petróleo desde 1880. Fuente: Data and Stories Library <http://lib.stat.cmu.edu/DASL/>.
- En 1929, Edwin Hubble investigó la relación entre la distancia de una galaxia a la tierra y la velocidad con la que está alejándose. En efecto se piensa que las galaxias se alejan como consecuencia del “Big Bang”. Hubble pensó que disponiendo de un modelo que relacionara la velocidad de recesión con la distancia a la tierra proporcionaría información sobre la formación del universo y sobre lo que podría pasar en el futuro. Los datos recogidos incluyen distancias en megaparsecs (1 megaparsec= 3.26 años luz) y velocidad de recesión en km/s. Fuente: Data and Stories Library, http://lib.stat.cmu.edu/DASL.

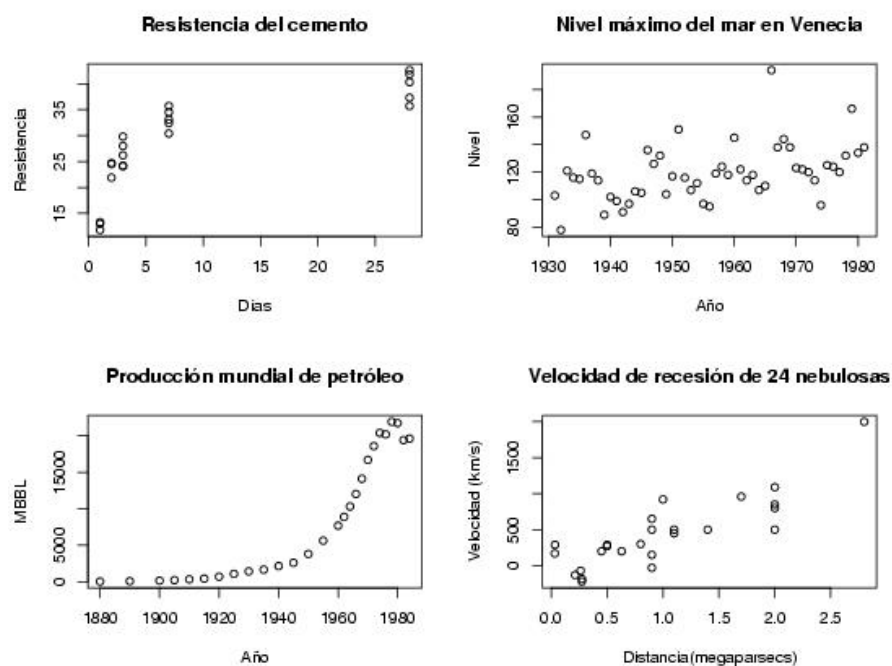


Figura I.8: Cuatro ejemplos de conjuntos de datos

I.5.2. Criterio de mínimos cuadrados

Para ilustrar las nociones nos limitamos primero al caso de una variable respuesta que llamaremos Y y una variable explicativa que llamaremos X .

Los datos se presenta en forma de pares:

X	x_1	x_2	\cdots	x_n
Y	y_1	y_2	\cdots	y_n

es decir que, para varios valores X observamos los valores correspondientes de Y . Para visualizar el conjunto recurrimos a la nube de puntos, también llamada diagrama de dispersión, en el que representamos los pares (x_i, y_i) , $i = 1, \dots, n$, en unos ejes Ox , Oy , ver figura I.9

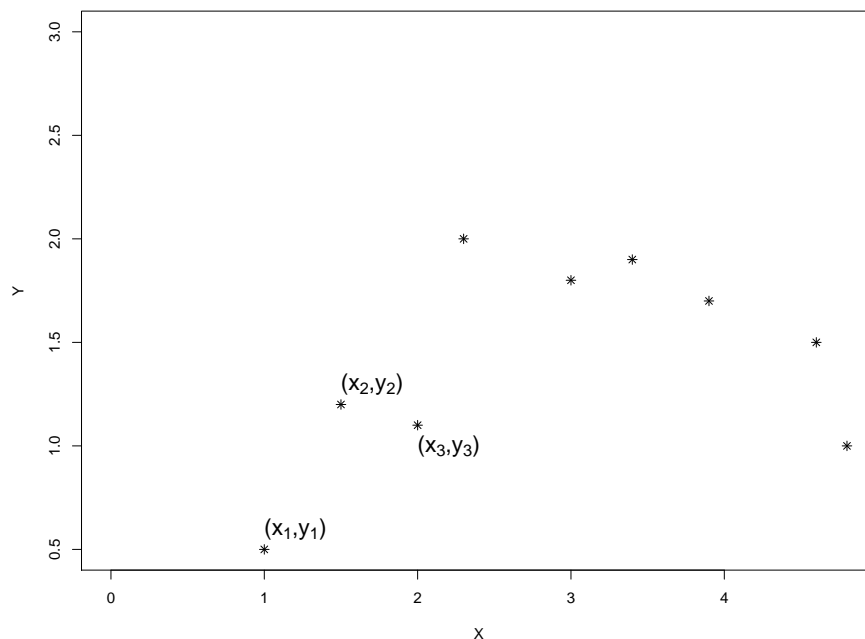


Figura I.9: Ejemplo de nube de puntos

Por conocimientos previos sobre el fenómeno que estudiamos o por la propia nube de puntos, decidimos ajustar a ésta una curva de una determinada forma funcional: podría ser por ejemplo una recta, de ecuación $Y = aX + b$, o una parábola $Y = a_0 + a_1X + a_2X^2$. La forma de la curva está fijada pero intervienen en la ecuación constantes, también llamadas parámetros, cuyo valor tenemos que ajustar para obtener el “mejor” ajuste posible: en el caso de la recta, debemos encontrar los valores de la pendiente b y de la ordenada en el origen a .

En una formulación general, escogemos una familia paramétrica de funciones

$$x \mapsto f(\theta, x) \quad \theta = (\theta_1, \dots, \theta_k), \quad (\text{I.3})$$

donde θ es el vector de parámetros. Buscar la función de la familia que mejor se ajusta a la nube de puntos es equivalente a encontrar el valor $\hat{\theta}$ de θ , que corresponde a esta función.

Debemos ahora dar sentido a la noción de “mejor”; debemos fijarnos un criterio que nos permita decidir que una función de la familia se ajusta mejor a la nube de puntos que otra. El criterio que seguimos en este tema es el de **mínimos cuadrados**.

Definimos la suma de cuadrados asociada a una función de la familia como la suma de los cuadrados de las distancias verticales entre la curva correspondiente y los datos observados de la nube de puntos. Tal como viene reflejado en la figura I.10, la distancia vertical entre por ejemplo el punto (x_3, y_3) y la curva es $y_3 - f(\theta, x_3)$, por lo tanto la suma de cuadrados se escribe

$$SC(\theta) = (y_1 - f(\theta, x_1))^2 + (y_2 - f(\theta, x_2))^2 + \cdots + (y_n - f(\theta, x_n))^2. \quad (\text{I.4})$$

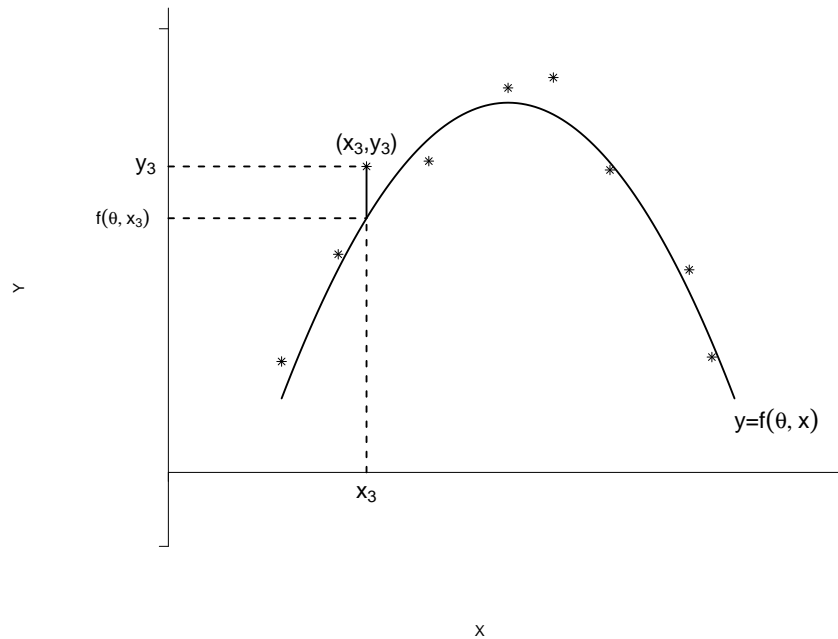


Figura I.10: Ajuste de una curva a la nube de puntos.

Buscamos el valor $\hat{\theta}$ de θ que minimiza la cantidad $\theta \mapsto SC(\theta)$, en muchos casos, es imposible encontrar este mínimo explícitamente y tenemos que recurrir a algoritmos numéricos. Nos centraremos en este tema en el caso en que la forma paramétrica de f es particularmente simple y permite el cálculo explícito de $\hat{\theta}$.

Supongamos que hemos ajustado la curva, es decir que hemos encontrado el valor $\hat{\theta}$ de θ que minimiza la suma de cuadrados, introduzcamos unos cuantos términos:

- La curva de ecuación $y = f(\hat{\theta}, x)$ se llama la **curva ajustada**.
- Las ordenadas de la curva ajustada correspondientes a los datos observados, es decir los valores $\hat{y}_1 = f(\hat{\theta}, x_1), \dots, \hat{y}_n = f(\hat{\theta}, x_n)$ se llaman los **valores ajustados**.
- Las distancias verticales entre los puntos observados y la curva ajustada se llaman los **residuos** e_1, \dots, e_n . Tenemos

$$e_i = y_i - \hat{y}_i, \quad i = 1, \dots, n.$$

- La suma de cuadrados

$$SC(\hat{\theta}) = \sum_{i=1}^n e_i^2$$

se llama **suma de cuadrados residuales**.

- Calcularemos en algunas ocasiones la varianza de los residuos, también llamada varianza residual

$$s_e^2 = \frac{1}{n-1} \sum_{i=1}^n (e_i - \bar{e})^2.$$

I.5.3. Casos concretos

Describimos ahora con más detalle unos pocos casos concretos en los que es posible obtener de manera explícita la expresión de $\hat{\theta}$, que minimiza la suma de cuadrados residuales. Estos casos corresponden todos a la llamada regresión lineal: son casos para los cuales los parámetros $(\theta_1, \dots, \theta_k)$ intervienen de manera lineal en la ecuación (I.3).

I.5.3.1. Recta $y = ax + b$

El caso más utilizado de ajuste por mínimo por mínimos cuadrados al ajuste por una recta, es decir cuando consideramos una variable explicativa X y buscamos ajustar un modelo de la forma

$$Y = aX + b.$$

Corresponde al caso en que θ consta de dos parámetros a y b , y la función f descrita en la sección I.5.2 es $f(\theta, x) = ax + b$. En este caso, decimos que el ajuste corresponde a la regresión lineal simple.

En el caso en que la pendiente a es positiva, hablamos de asociación positiva entre X e Y : cuando crece X , crece Y , cuando decrece X , decrece Y , y viceversa. En cambio, si la pendiente a es negativa, hablamos de asociación negativa entre X e Y (cuando crece una variable, decrece la otra).

a). **Obtención de la recta ajustada** La suma de cuadrados se escribe

$$SC(\theta) = SC(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2,$$

Los candidatos a alcanzar el mínimo de esta función satisfacen

$$\begin{aligned} \frac{\partial}{\partial a} SC(a, b) &= 0 \\ \frac{\partial}{\partial b} SC(a, b) &= 0. \end{aligned}$$

Deducimos de unas cuantas manipulaciones algebraicas que las soluciones a este sistema de ecuaciones son

$$\begin{aligned} \hat{a} &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \\ \hat{b} &= \bar{y} - \hat{a}\bar{x}. \end{aligned}$$

Introducimos la cantidad

$$s_{xy} = \frac{n}{n-1}(\overline{xy} - \bar{x}\bar{y}), \quad (\text{I.5})$$

que llamamos la **covarianza** de X e Y . El coeficiente \hat{a} se puede por lo tanto escribir como

$$\hat{a} = \frac{s_{xy}}{s_x^2},$$

donde s_x^2 es la varianza de X que introdujimos en la sección I.4.2.1. Con estas notaciones, se puede escribir la ecuación de la recta ajustada en una forma compacta:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}).$$

Nota La covarianza es una cantidad que puede ser positiva o negativa. De hecho tiene el mismo signo que la pendiente de la recta ajustada. Por lo tanto, si la covarianza es positiva, Y y X presentan una asociación positiva mientras que, si la covarianza es negativa Y y X presentan una asociación negativa.

b). Bondad del ajuste Para la regresión lineal simple, los residuos son

$$\begin{aligned} e_1 &= y_1 - f(\hat{\theta}, x_1) = y_1 - \hat{a}x_1 - \hat{b} \\ &\vdots \\ e_n &= y_n - f(\hat{\theta}, x_n) = y_n - \hat{a}x_n - \hat{b}, \end{aligned}$$

y tienen las siguientes propiedades

Propiedades de los residuos

- La media de los residuos es nula.

Demostración:

$$\begin{aligned} \bar{e} = \frac{e_1 + \dots + e_n}{n} &= \frac{1}{n}[(y_1 + \dots + y_n) - \hat{a}(x_1 + \dots + x_n) - n\hat{b}] \\ &= \bar{y} - \hat{a}\bar{x} - \hat{b} = 0 \end{aligned}$$

- Se puede demostrar sin dificultad que la varianza residual se escribe como

$$s_e^2 = s_y^2 \left(1 - \frac{(s_{xy})^2}{s_x^2 s_y^2} \right). \quad (\text{I.6})$$

De esta ecuación deducimos que la cantidad $\frac{(s_{xy})^2}{s_x^2 s_y^2}$ puede medir la calidad del ajuste. De hecho le damos un nombre especial:

Definición I.5.1 La cantidad $r = \frac{s_{xy}}{s_x s_y}$ se llama *coeficiente de correlación (de Pearson)* de X e Y .

La cantidad $R^2 = \frac{(s_{xy})^2}{s_x^2 s_y^2}$ se llama *coeficiente de determinación*.

Propiedades de r y R^2 De la fórmula $s_e^2 = s_y^2(1 - R^2)$, ver (I.6), deducimos

- R^2 está siempre comprendido entre 0 y 1, y cuanto más cercano esté de 1, mejor es el ajuste, puesto que corresponderá a una varianza residual menor. En particular, deducimos que si $R^2 = 1$, la varianza residual s_e^2 es nula, lo que quiere decir que la dispersión de los residuos es nula: todos los residuos son iguales, y por lo tanto iguales a su media, que vale 0, todos los puntos de la nube están situados en la recta, el ajuste es perfecto. Se suele considerar un valor de R^2 mayor que 0.8 como correspondiente a un ajuste bueno, mientras que un valor mayor que 0.9 corresponde a un ajuste muy bueno.
- Puesto que $R^2 = r^2$ y $0 \leq R^2 \leq 1$, deducimos que el coeficiente de correlación r está siempre comprendido entre -1 y 1 . Si $r = \pm 1$, el ajuste de los puntos observados por una recta es perfecto. El coeficiente de correlación se interpreta en general como una cantidad que cuantifica la asociación lineal que existe entre dos variables: cuanto más cerca de ± 1 , más se aproxima la nube de puntos a una recta.

Además por la definición de r , sabemos que r es del mismo signo de la covarianza. Por lo tanto, si r es positivo y cercano a 1, los datos apoyan la existencia de una asociación lineal positiva entre las dos variables, mientras que si es negativo y cercano a -1 , presentan una asociación lineal negativa.

Sin embargo, es necesario tener precaución a la hora de interpretar valores del coeficiente de correlación: sólo es un resumen, fiable en el caso en que está próximo a ± 1 para indicar que existe una fuerte asociación lineal entre las variables pero mucho menos fiable si toma un valor alejado de ± 1 . Anscombe (1973), "Graphs in statistical analysis", *American Statistician*, **27**, pp 17-21, construyó cuatro conjuntos de datos artificiales que dan lugar al mismo coeficiente de correlación y a las mismas rectas de regresión, pero cuyos aspectos son completamente diferentes. Los datos se presentan en el apéndice, y se deja su estudio en ejercicio.

c). Un ejemplo Para ilustrar el procedimiento que se sigue para calcular los valores de \hat{a} y \hat{b} , consideremos el ejemplo muy sencillo a continuación:

Queremos estudiar la relación entre el peso y la altura en un grupo de individuos. Los datos son

Peso(kg)	54	70	65	78	68	85	Y
Altura(cm)	160	170	172	185	160	175	X

Se deja en ejercicio al lector la representación de este conjunto a través de una nube de puntos... Buscamos ajustar una recta a la nube y pasamos a calcular la ecuación de la recta de regresión que en su forma compacta se escribe

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}).$$

Para calcular s_{xy} y s_x^2 utilizaremos las fórmulas (I.2) y (I.5), necesitamos por lo tanto \bar{x} , \bar{x}^2 , \bar{y} , \bar{y}^2 y \bar{xy} . Tenemos

$$\begin{aligned}\bar{x} &= \frac{160+170+\dots+175}{6} = 170,33, & \bar{y} &= \frac{54+70+\dots+85}{6} = 70, \\ \overline{x^2} &= \frac{160^2+170^2+\dots+175^2}{6} = 29089, & \overline{y^2} &= \frac{54^2+70^2+\dots+85^2}{6} = 4995,7, \\ \overline{xy} &= \frac{160 \times 54 + 170 \times 70 + \dots + 175 \times 85}{6} = 11984,2\end{aligned}$$

Deducimos que

$$\begin{aligned}s_x^2 &= \frac{n}{n-1}(\overline{x^2} - (\bar{x})^2) = \frac{6}{5}[29089 - (170,33)^2] \simeq 90,7, \\ s_y^2 &= \frac{n}{n-1}(\overline{y^2} - (\bar{y})^2) = \frac{6}{5}[4995,7 - (70)^2] \simeq 144,8, \\ s_{xy} &= \frac{n}{n-1}(\overline{xy} - (\bar{x})(\bar{y})) = \frac{6}{5}[11984,2 - 170,33 \times 70] \simeq 73.\end{aligned}$$

La ecuación de la recta es por lo tanto $y - 70 = \frac{73}{90,7}(x - 170,33)$, es decir

$$y = 0,80x - 67,1.$$

El modelo teórico propuesto para relacionar el peso y la altura es $Peso \simeq 0,8Altura - 67,1$.

En cuanto a la bondad del ajuste, tenemos que

$$R = \frac{s_{xy}}{s_x s_y} = \frac{73}{\sqrt{90,7}\sqrt{144,8}} \simeq 0,715,$$

lo que implica que $R^2 \simeq 0,51$, un ajuste malo.

d). Predicción Tal como lo mencionamos en la introducción del tema, si disponemos del modelo ajustado podemos utilizarlo para predecir el valor de la respuesta para valores no observados de X :

Si x_0 es un valor no observado, nuestra predicción del valor de Y será

$$y_{x_0} = \hat{a}x_0 + \hat{b}.$$

Si consideramos el ejemplo de la relación entre peso y altura del apartado anterior, podemos contestar a la pregunta ¿a qué peso correspondería una altura de 180cm? Sustituimos x por 180 en la ecuación de la recta ajustada, y encontramos que el peso asociado sería $0,80 \times 180 - 67,1 \simeq 76,9kg$.

Sin embargo, debemos tener mucho cuidado al extrapolar nuestro modelo fuera del rango de valores de X que hemos observado, al no disponer de valores fuera de este rango, tampoco sabemos si el modelo deducido seguirá valido. Para el ejemplo de los pesos, si queremos utilizar el modelo ajustado para saber a qué peso correspondería la altura de un niño de 80cm por ejemplo, obtenemos $0,80 \times 80 - 67,1 \simeq -3,1kg$, ¡lo que no tiene sentido!

Nota. El motivo por el cual, muy a menudo una recta suele ajustarse bastante bien a una nube de puntos, corresponde a que la fórmula de Taylor nos dice que localmente, cualquier función derivable se puede aproximar por una recta: aunque la relación entre Y y X no sea lineal sino de la forma $Y = f(\theta, X)$, f general, si f es derivable y observamos valores de X no muy dispersos alrededor, f se comporta aproximadamente como la tangente en un X central.

I.5.3.2. Recta forzada por el origen

Hay situaciones en las que pensamos recurrir a un ajuste lineal, pero sabemos por motivos físicos que un valor de X nulo corresponde necesariamente a un valor de Y nulo también. En este caso, no tenemos por que considerar todas las rectas, sino podemos restringirnos a las rectas que pasan por el origen $(0, 0)$. La ecuación de una recta forzada por el origen es

$$y = ax.$$

Dos ejemplos de situaciones en las que un valor nulo de X implica un valor nulo de Y :

- Medimos la evolución en función del tiempo (X) de la concentración (Y) de un producto que se va creando en una reacción química. Cuando empezamos la reacción $X = 0$, todavía no puede haber producto, por lo tanto $Y = 0$.
- Queremos medir el tiempo t que tarda un objeto que soltamos desde una altura h , en alcanzar el suelo. La relación física proporcionada por la teoría es $h = gt^2$, donde g es la constante de la gravedad. Si queremos comprobar que los datos empíricos confirman esta relación, buscaremos si es cierto que

$$t = \frac{1}{\sqrt{g}}\sqrt{h}.$$

Consideraremos $X = \sqrt{h}$, $Y = t$, y buscaremos ajustar una recta $y = ax$.

Las fórmulas que vimos para el caso de una recta general ya no son válidas. Calculemos la ecuación de la recta forzada por el origen: disponemos de n pares de datos $(x_1, y_1), \dots, (x_n, y_n)$, puesto que la función que buscamos ajustar es $f(\theta, x) = ax$, $\theta = a$ y la suma de cuadrados de la fórmula (I.4) se escribe

$$SC(\theta) = SC(a) = \sum_{i=1}^n (y_i - ax_i)^2.$$

El candidato a minimizar $SC(a)$ satisface la ecuación $\frac{dSC(a)}{da} = 0$. Calculamos

$$\frac{dSC(a)}{da} = \sum_{i=1}^n -x_i 2(y_i - ax_i) = 2[-\sum_{i=1}^n x_i y_i + a \sum_{i=1}^n x_i^2].$$

Por lo tanto, la solución a la ecuación $\frac{dSC(a)}{da} = 0$ es

$$\hat{a} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\overline{xy}}{\overline{x^2}}.$$

Puesto que la derivada segunda de $SC(a)$ es positiva, se trata efectivamente de un mínimo.

I.5.3.3. Algunas transformaciones útiles

Sólo hemos descrito cómo calcular la curva ajustada para dos familias específicas de funciones $y = ax$ e $y = ax + b$. Para una especificación más general de la función f que queremos ajustar, se recurre a algoritmos numéricos para encontrar el valor de los parámetros que minimicen la suma de cuadrados $SC(\theta)$.

Sin embargo, hay algunos tipos de modelos no lineales que se pueden abordar con los resultados del caso lineal después de realizar unas transformaciones convenientes.

a). **Modelo exponencial** Supongamos que queremos ajustar un modelo exponencial a una nube de puntos. La ecuación de las funciones que consideramos son $y = be^{ax}$, con $b > 0$. En el caso en que a es positivo, modelizamos un crecimiento exponencial, mientras que, si a es negativa, modelizamos un decrecimiento exponencial.

La relación entre Y y X es altamente no lineal, sin embargo una simple transformación puede llevarlo a un modelo lineal:

Modelo teórico original		Modelo transformado
$y = be^{ax}$	$\xrightarrow{\text{cojo ln}}$	$\ln(y) = \ln(b) + ax$ $y' = b' + a'x'$

Si introducimos las variables transformadas $Y' = \ln(Y)$, y $X' = X$, éstas satisfacen una relación lineal: $Y' = a'X' + b'$.

Nuestro procedimiento para ajustar un modelo exponencial consistirá por lo tanto en

1. Calculamos los datos transformados, es decir pasar de

$$\begin{array}{c|cccc} X & x_1 & x_2 & \dots & x_n \\ \hline Y & y_1 & y_2 & \dots & y_n \end{array} \quad y = be^{ax}$$

a

$$\begin{array}{c|cccc} X' & x_1 & x_2 & \dots & x_n \\ \hline Y' & \ln(y_1) & \ln(y_2) & \dots & \ln(y_n) \end{array} \quad y' = a'x' + b'$$

2. Ajustamos una recta a las variables transformadas, encontramos $y' = \hat{a}'x' + \hat{b}'$.
3. Volvemos al modelo original, haciendo la transformación inversa (en este caso exponencial)

$$y' = \hat{a}'x' + \hat{b}' \xrightarrow{\text{cojo exp}} y = e^{\hat{a}'x' + \hat{b}'} = e^{\hat{b}'} e^{\hat{a}'x'}$$

Ejemplo 1. Queremos ajustar un modelo exponencial a los siguientes datos

$$\begin{array}{c|cccc} X & 2.3 & 5 & 7.1 & 8 \\ \hline Y & 2.92 & 3.69 & 6.19 & 6.36 \end{array}$$

Transformamos los datos:

$$\begin{array}{c|cccc} X' & 2.3 & 5 & 7.1 & 8 \\ \hline Y' = \ln(Y) & 1.07 & 1.31 & 1.82 & 1.85 \end{array}$$

Ajustamos una recta a los datos transformados, calculando \bar{x}' , $\overline{x'^2}$, \bar{y}' , $\overline{y'^2}$ y $\overline{x'y'}$, para obtener \hat{a}' y \hat{b}' : $y' = 0,148x' + 0,682$, es decir que $\ln(y) = 0,148x + 0,682$, lo que implica que

$$y = e^{0,148x} e^{0,682} = 1,18e^{0,148x}$$

Ejemplo 2. El Ministerio de Fomento publica los datos de precios del metro cuadrado de las viviendas en España. En la gráfica siguiente, figura I.11, se ha representado la evolución del precio del metro cuadrado de la vivienda libre en la Región de Murcia

por cuatrimestre entre 1995 y 2012. En la primera fase de la evolución (1995-2006), aparece una tendencia exponencial, podemos representar también el logaritmo del precio en este periodo para ver si la evolución es entonces lineal. La gráfica correspondiente aparece en la figura I.12. Notar que entre 1996 y 2002, la curva del logaritmo sigue presentando una curvatura, lo que implica que ¡la subida fue más que exponencial!

Si nos preguntamos si este patrón de evolución se repite para el conjunto de las provincias españolas, podemos hacer gráficas conjuntas aprovechando los datos proporcionados por el Ministerio de Fomento. A título de ejemplo, se representa en la figura I.13 la evolución de los precios en la Región de Murcia, junto con las dos provincias con patrones más extremos: la de Málaga donde el incremento fue aun más pronunciado, y la de Badajoz, donde lo fue mucho menos.

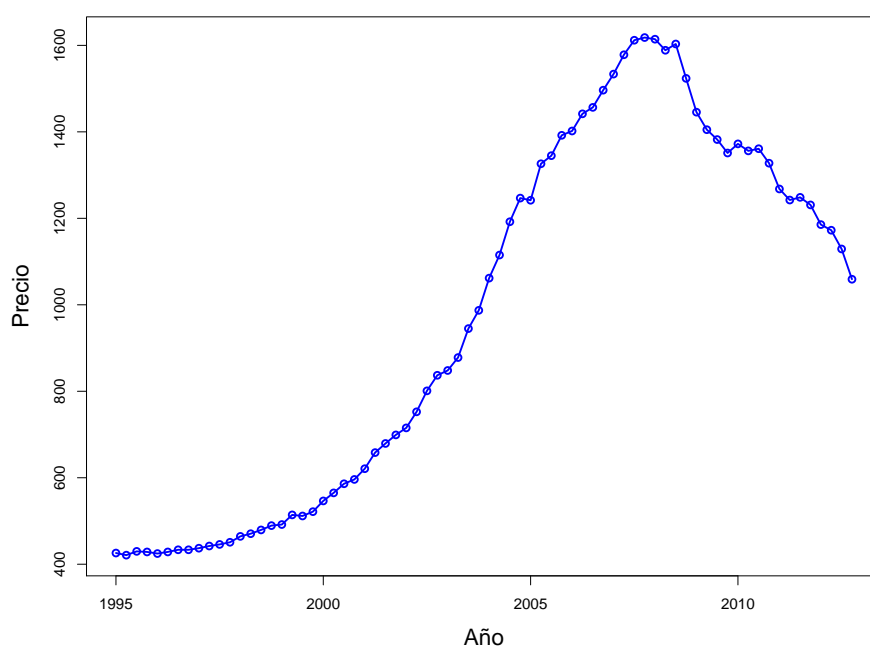


Figura I.11: Evolución del precio en euros del metro cuadrado de la vivienda libre en la Región de Murcia, datos cuatrimestrales, 1995-2012. Fuente: Ministerio de Fomento.

b). Modelo potencial El modelo potencial es de la forma $y = bX^a$. La forma de la nube de puntos correspondiente depende del valor de a . La transformación que utilizamos es la misma que para el modelo exponencial: aplicamos los logaritmos.

$$\begin{array}{ccc}
 \text{Modelo teórico original} & & \text{Modelo transformado} \\
 y = bx^a & \xrightarrow{\text{cojo } \ln} & \ln(y) = \ln(b) + a \ln(x) \\
 & & y' = b' + a'x'
 \end{array}$$

Introducimos las variables transformadas $Y' = \ln(Y)$, y $X' = \ln(X)$, éstas satisfacen una relación lineal: $Y' = a'X' + b'$. Seguimos los mismos pasos que en el apartado

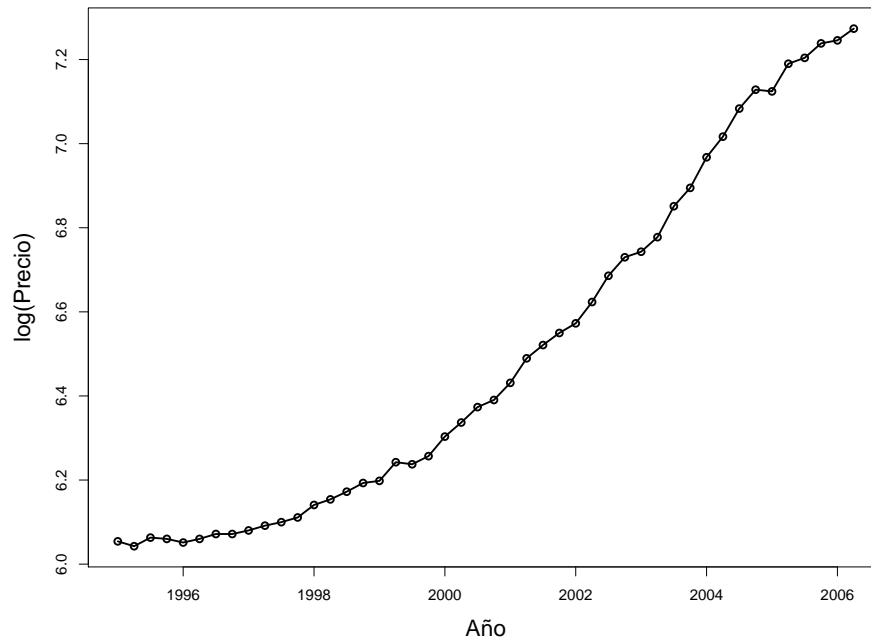


Figura I.12: Evolución del logaritmo del precio en euros del metro cuadrado de la vivienda libre en la Región de Murcia, datos cuatrimestrales, 1995-2006.

anterior con los datos transformados.

Ejemplo. Queremos ajustar un modelo potencial a los siguientes datos

X	3	7.34	20.1	54.6
Y	10.3	13.5	18.2	24.5

Transformamos los datos:

$X' = \ln(X)$	1.1	2	3	4
$Y' = \ln(Y)$	2.3	2.6	2.9	3.2

Ajustamos una recta a los datos transformados, calculando \bar{x}' , $\overline{x'^2}$, \bar{y}' , $\overline{y'^2}$ y $\overline{x'y'}$, para obtener \hat{a}' y \hat{b}' : $y' = 0,298x' + 2,006$, es decir que $\ln(y) = 0,298 \ln(x) + 2,006$, lo que implica que

$$y = e^{0,298 \ln(x)} e^{2,006} = 7,433x^{0,298}.$$

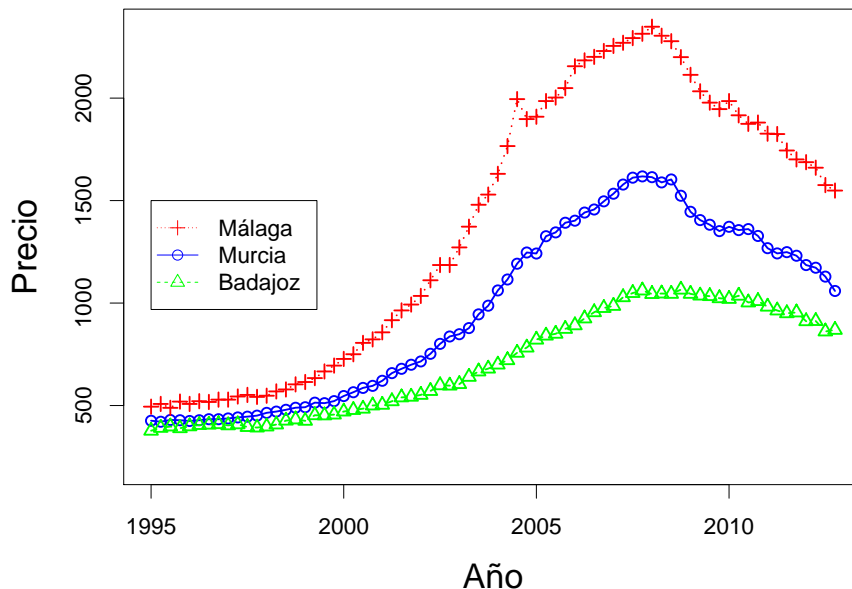


Figura I.13: Evolución del logaritmo del precio en euros del metro cuadrado de la vivienda libre en la Región de Murcia y las provincias de Málaga y Badajoz, datos cuatrimestrales, 1995-2012.

Apéndice

A continuación se presentan los datos de Anscombe (1973), "Graphs in statistical analysis", *American Statistician*, **27**, pp 17-21, se recomienda calcular las medias de X_1 , X_2 , X_3 , y X_4 así como de Y_1 , Y_2 , Y_3 y Y_4 y a continuación calcular las rectas de regresión de Y_i sobre X_i para $i=1, 2, 3, 4$. Finalmente, realizar las cuatro gráficas de Y_i en función de X_i para $i=1, 2, 3, 4$.

X1	Y1	X2	Y2	X3	Y3	X4	Y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6



2013: Año Internacional de la Estadística. ¿Sabías qué...?

Los precios de las cuotas y primas de todo tipo de seguro se basan en los cálculos de probabilidad de que ocurra un siniestro y usan de manera esencial la estadística. La ciencia asociada se llama ciencia actuarial. Un hito en la historia de los seguros se remonta a Edmond Halley, el famoso astrónomo inglés, quien estimó en 1693 a partir de datos estadísticos la probabilidad de fallecimiento de los hombres a cada edad, y ¡lo aplicó al precio de productos financieros de la época!

Fuente: A short history of Mathematical Population Dynamics, N Bacaër, 2011, Springer Verlag.

statistics

Fundamentos de la teoría de la probabilidad.

En el tema anterior, hemos estudiado algunos conjuntos de datos reales que presentan variabilidad aun cuando son generados por un experimento realizado en condiciones que nos esforzamos por mantener constantes. Es más, si consideramos el ejemplo de una reacción química de primer orden visto en la sección sobre ajuste de curvas, disponemos de una teoría química que describe la evolución de la concentración de los reactivos en función del tiempo como solución de un sistema de ecuaciones diferenciales y sin embargo, los datos provenientes de un experimento nunca se ajustan perfectamente a la curva teórica esperada. ¿Qué tipo de afirmaciones sobre el resultado de tal experimento podríamos entonces realizar? Estas afirmaciones tendrán que tener en cuenta la incertidumbre ligada al experimento. La teoría de la probabilidad es una teoría matemática que permite modelizar experimentos aleatorios, es decir experimentos cuyo resultado es imposible predecir de manera exacta. Son los cimientos sobre los que está construida toda la estadística.

II.1. Conceptos básicos relacionados con un experimento

Empecemos por introducir unos términos y conceptos relacionados con un experimento

II.1.1. Experimento aleatorio

Un experimento aleatorio es aquel que, aun realizado en las mismas condiciones, produce resultados posiblemente distintos.

Se opone a la noción de experimento determinístico, en el que conociendo las condiciones iniciales, se puede prever de manera exacta el resultado. En la práctica, aunque muchos experimentos son verdaderamente aleatorios, en muchos casos se puede tener por despreciable la variabilidad en los resultados y el considerar el experimento como determinístico proporciona conclusiones satisfactorias. Sin embargo,

hay muchas situaciones en las que es sólo al tener en cuenta el carácter aleatorio de un fenómeno que se llega a conclusiones válidas.

Un ejemplo sencillo de experimento aleatorio consiste en tirar un dado.

II.1.2. Suceso elemental

Un resultado posible del experimento se llama un suceso elemental.

II.1.3. Espacio muestral

El conjunto S de todos los resultados posibles de un experimento aleatorio se llama el espacio muestral de este experimento.

Si consideramos el experimento que consiste en tirar el dado, el espacio muestral es $\{1, 2, 3, 4, 5, 6\}$.

II.1.4. Suceso

Cualquiera colección de resultados posibles, es decir cualquier subconjunto de S , se llama un suceso posible asociado al experimento considerado.

Un suceso siempre se podrá describir de dos formas: utilizando palabras, o de forma matemática, utilizando el formalismo de los conjuntos:

Ejemplo. Asociado al experimento que consiste en tirar un dado, podemos considerar el suceso A : "Sacar un número par". A se puede también describir como $A = \{2, 4, 6\}$.

Consideremos un suceso A , y llevamos a cabo una realización del experimento, se dice que "Ocurre A " si el resultado del experimento pertenece a A . En el ejemplo anterior, donde A es "sacar un número par", si llevamos a cabo el experimento y sale un 4, diremos que ha ocurrido A .

Podemos para describir sucesos de interés, utilizar el formalismo de la teoría de conjuntos :

II.1.4.1. Operaciones elementales con sucesos

- Unión de dos sucesos A y B : la unión de A y B es el suceso formado por todos los elementos de A y todos los elementos de B .

$$A \cup B = \{x \in S : x \in A \text{ o } x \in B\},$$

Notar que "Ocurre $A \cup B$ " si y sólo si "Ocurre A " o "Ocurre B ". Por ejemplo, si B es el suceso "Sale un número mayor o igual a 5", es decir $B = \{5, 6\}$, $A \cup B = \{2, 4, 5, 6\}$.

- Intersección de dos sucesos A y B : la intersección de A y B está formada por los elementos comunes a A y a B .

$$A \cap B = \{x \in S : x \in A \text{ y } x \in B\}$$

"Ocurre $A \cap B$ " si y sólo si "Ocurre A " y "Ocurre B ". En el ejemplo anterior, $A \cap B = \{6\}$

Disponemos también de las propiedades siguientes de las operaciones con sucesos:

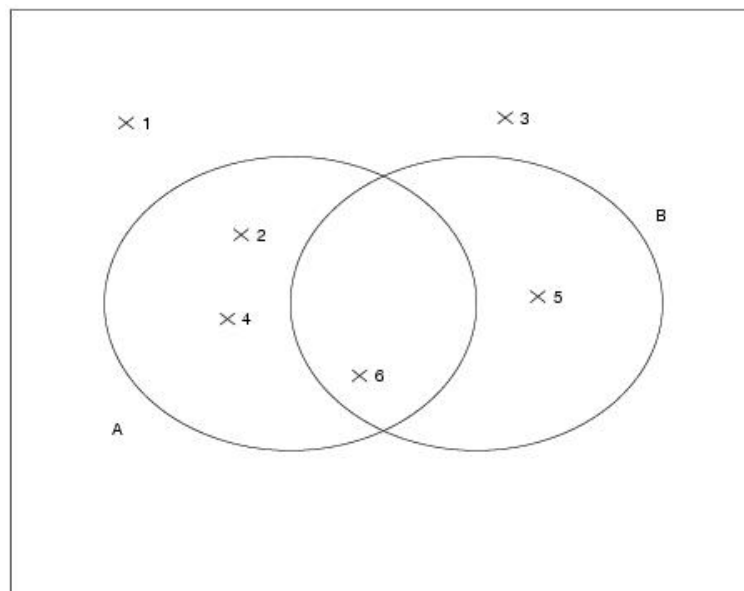
- Comutatividad $A \cup B = B \cup A$
 $A \cap B = B \cap A$
- Asociatividad $A \cup (B \cup C) = (A \cup B) \cup C$
 $A \cap (B \cap C) = (A \cap B) \cap C$
- Distributividad $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

II.1.4.2. Algunos términos más.

- El suceso seguro es S , el espacio muestral entero.
- El suceso imposible es el conjunto vacío
- Diremos que dos sucesos A y B son incompatibles, si no pueden ocurrir a la vez, es decir $A \cap B = \emptyset$ y diremos que los sucesos A_1, A_2, A_3, \dots son incompatibles dos a dos, si para todos $i \neq j$, $A_i \cap A_j = \emptyset$.
- Suceso complementario de A ($A^c = \{x \in S : x \notin A\}$). Notar que “Ocurre A^c ” si y sólo si “No Ocurre A ”. En nuestro ejemplo, $A^c = \{1, 3, 5\}$.

II.1.5. Diagrama de Venn

Es una manera gráfica de representar los sucesos: un rectángulo representa S el espacio muestral entero, y vamos agrupando los sucesos elementales en sucesos. Por ejemplo, volviendo al ejemplo del dado:



II.1.6. Leyes de Morgan

Para dos sucesos A y B ,

$$(A \cap B)^c = A^c \cup B^c,$$

es decir que, afirmar que “no ha ocurrido (A y B)” es lo mismo que decir “o bien no ha ocurrido A o bien no ha ocurrido B ”.

$$(A \cup B)^c = A^c \cap B^c,$$

es decir que, afirmar que “no ha ocurrido (A o B)” es lo mismo que decir “no ha ocurrido A y tampoco ha ocurrido B ”.

II.2. Concepto de Probabilidad

Al llevar a cabo una realización de un experimento aleatorio, somos conscientes de que no podemos predecir el resultado, sin embargo tenemos a menudo información sobre las “posibilidades” que tiene un determinado suceso de ocurrir. Queremos cuantificar de alguna manera esta información que llamaríamos la probabilidad del suceso.

II.2.1. Definición informal de la probabilidad - propiedades.

Más que formalizar una definición, preferimos indicar qué propiedades tendrá que tener la cantidad escogida para que refleje la creencia que tenemos de que un determinado suceso ocurra.

Dados todos los sucesos asociados a un experimento aleatorio, asignaremos a cada suceso A , una cantidad que denotaremos por $\mathbb{P}(A)$ y que llamaremos la “probabilidad del suceso A .” Pero al realizar esta asignación deberemos tener en cuenta que se deberá cumplir:

1. La probabilidad de un suceso es un número entre 0 y 1:

$$0 \leq \mathbb{P}(A) \leq 1,$$

2. considerando que la probabilidad asociada al suceso imposible es nula:

$$\mathbb{P}(\emptyset) = 0,$$

mientras que la probabilidad asociada al suceso seguro es 1 :

$$\mathbb{P}(S) = 1.$$

3. La probabilidad de que un suceso no ocurra es 1– la probabilidad de que ocurra:

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^C).$$

4. Si un suceso tiene más resultados posibles que otro, su probabilidad será mayor, es decir,

Si A y B son dos sucesos tales que $A \subset B$, entonces

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

5. Reglas de adición:

- a) Si A y B son dos sucesos incompatibles, es decir que no pueden ocurrir a la vez, la probabilidad de que ocurra uno o el otro es la suma de las probabilidades de cada suceso:

$$\text{Si } A \cap B = \emptyset, \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Esta fórmula seguirá válida si se trata de la unión de tres o más sucesos.

- b) En cambio si A y B son dos sucesos cualesquiera (en particular, podrían ocurrir a la vez), un diagrama de Venn nos convence de que la fórmula correcta es

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$

puesto que, al sumar $\mathbb{P}(A)$ y $\mathbb{P}(B)$, hemos contabilizado dos veces la probabilidad de la intersección $\mathbb{P}(A \cap B)$, y debemos restarla una vez para obtener $\mathbb{P}(A \cup B)$.

- c) Esta última fórmula se puede generalizar a más de dos sucesos, nos limitaremos a enunciar el caso de tres:

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ &\quad + \mathbb{P}(A \cap B \cap C). \end{aligned}$$

En todo lo que sigue, entenderemos como probabilidad la asignación de un número a cada suceso posible asociado a un experimento aleatorio, que cumpla con las cinco propiedades que acabamos de enumerar.

II.2.2. El caso de un espacio muestral finito y la definición de Laplace.

II.2.2.1. Espacio muestral finito

En el caso en que hay un número finito de resultados posibles del experimento, es decir el caso de un espacio muestral finito, la definición de una probabilidad asociada al experimento pasará por la asignación de la probabilidad de cada suceso elemental. En particular, diremos que los sucesos elementales son *equiprobables* si todos tienen la misma probabilidad de ocurrir. Para cumplir con las propiedades anteriores, está claro que si hay n sucesos elementales que son además equiprobables, la probabilidad de cada uno de ellos es $1/n$. En este caso, la probabilidad de un suceso A se podrá siempre calcular como (Regla de Laplace)

$$\mathbb{P}(A) = \frac{\text{n}^\circ \text{ de elementos en } A}{\text{n}^\circ \text{ de elementos totales}} = \frac{\text{n}^\circ \text{ de casos favorables}}{\text{n}^\circ \text{ de casos posibles}}.$$

Para ilustrar esta manera de calcular la probabilidad de un suceso, nos preguntamos ¿qué es más fácil acertar, el gordo de la lotería, la primitiva, o una quiniela de catorce resultados?

El gordo de la lotería, el experimento consiste en escoger al azar un número entre 100000 posibles, si A designa el suceso "acierto", la probabilidad de acertar es de una entre 100000,

$$\mathbb{P}(A) = \frac{1}{100000}.$$

La *primitiva*, el experimento consiste en extraer 6 números sin reposición entre 49. El número de casos posibles se calcula como las combinaciones sin repetición de 49 elementos tomados de 6 en 6 (en el anexo podéis encontrar un breve recordatorio sobre reglas de conteo), es $\binom{49}{6} = 13984000$. La probabilidad de acertar es una entre casi 14 millones:

$$\mathbb{P}(A) = \frac{1}{13984000} \simeq 7.15 \times 10^{-8}.$$

La *quiniela*, el experimento consiste en escoger quince veces uno de los tres resultados posibles 1, X, 2. El número de casos posibles es $3 \times 3 \times \dots \times 3 = 3^{15} = 14348907$. La probabilidad de acertar es de una entre 14 millones.

$$\mathbb{P}(A) = \frac{1}{14348907} \simeq 6.97 \times 10^{-8}.$$

Por supuesto, aquí no hemos tenido en cuenta que no se suele rellenar las quinielas escogiendo al azar una de las tres posibilidades 1, X, 2...

Euro-millón, el juego consiste en acertar 5 números de una tabla de 50 (Del nº 1 al nº 50) y además acertar 2 números (estrellas) de una tabla de 9 (Del nº 1 al nº 9). Es decir, para tener derecho al primer premio hay que acertar 7 números (5 + 2). Para calcular el número de casos posibles, tenemos en cuenta que para escoger los 5 primeros números de la tabla de 50, tenemos $\binom{50}{5}$ posibilidades, y para cada una de estas combinaciones, tenemos $\binom{9}{2}$ posibilidades para escoger las dos estrellas. En total tenemos por lo tanto

$$\binom{50}{5} \times \binom{9}{2} = 76275360$$

casos posibles, es decir que la probabilidad de acertar es de una entre más de 76 millones.

En cuanto a premios, en 2006, un acertante del euro-millón podría haber ganado hasta 180 millones de euros! El mayor premio que se ganó con la primitiva fue de casi 25 millones de euros, y fue en el 2005, mientras que en la quiniela, el mayor premio fue de 9 millones de euros (fue en la temporada 05/06)

Por último, hay un participante que siempre gana: el estado. En 2005 por ejemplo, Loterías y Apuestas del Estado, la sociedad que gestiona los juegos estatales, ingresó al Tesoro Público casi 2300 millones de euros (fuente: Memoria de Loterías y Apuestas del Estado 2005). Hay que decir que los españoles se gastaron en juegos en 2005, sumando los de gestión privada (casino, bingo, máquinas), la ONCE, y los de gestión pública, ¡una cantidad de 28 000 millones de euros!

II.2.2.2. Interpretación “frecuentista” de la probabilidad

En cuanto dejamos el marco de los sucesos elementales equiprobables, la asignación de una probabilidad a cada suceso es mucho más complicada. Muchas interpretaciones de resultados ligados a probabilidades están relacionadas con la definición de Laplace, llamada la “interpretación frecuentista” de la probabilidad:

Para un suceso A , realizamos un gran número de veces n el experimento, y consideramos que

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \frac{\text{n}^\circ \text{ de veces que ha ocurrido } A \text{ entre las } n \text{ realizaciones}}{n}.$$

II.3. La probabilidad condicionada.

Corresponde a una re-asignación o actualización de la probabilidad de un suceso cuando tenemos información sobre el resultado del experimento.

II.3.1. Definición

Si A y B son dos sucesos tales que $\mathbb{P}(B) > 0$, la probabilidad de A condicionada a B se denota por $\mathbb{P}(A|B)$ y es igual a

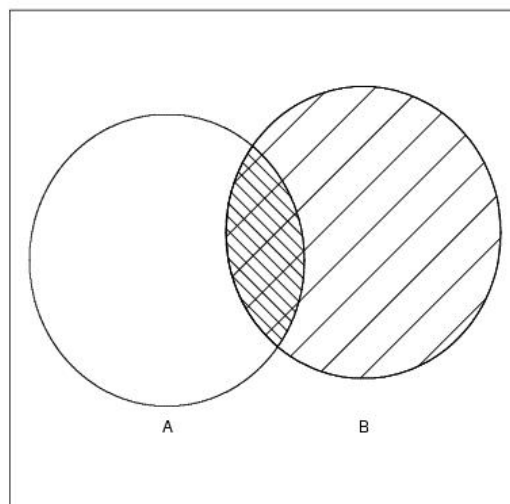
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Su interpretación es: realizamos el experimento, y sabemos que ha ocurrido B , ¿cuál es la probabilidad de que haya ocurrido A también? Por ejemplo, en una tirada de un dado he apostado por el "6". Tiran el dado sin que yo pueda ver el resultado, pero me dicen que ha salido un número par. Teniendo esta información, ¿cuál es la probabilidad de que haya ganado la apuesta? Es intuitivamente claro que es de un tercio (un caso favorable, el "6" entre tres posibles, el "2", el "4" y el "6".) Si introduzco los sucesos $A =$ "sale un 6", y $B =$ "sale un número par", quiero calcular $\mathbb{P}(A|B)$, utilizo la definición para encontrar:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{6\})}{\mathbb{P}\{2, 4, 6\}} = \frac{1/6}{3/6} = 1/3,$$

lo que coincide con nuestra intuición.

Al considerar el siguiente diagrama de Venn,



nos convencemos de que la definición $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ es intuitivamente razonable: realizamos el experimento y sabemos que el resultado pertenece a B , y nos preguntamos cuál es la probabilidad de que el resultado pertenezca también a A : B es nuestro nuevo espacio muestral puesto que sabemos que los resultados posibles pertenecen a B , y la probabilidad de que pertenezca a A es el cociente $\mathbb{P}(A \cap B)/\mathbb{P}(B)$.

II.3.2. Regla del producto.

(i) Si A y B son dos sucesos con $\mathbb{P}(B) > 0$,

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A|B)\mathbb{P}(B) \\ &= \mathbb{P}(B|A)\mathbb{P}(A)\end{aligned}$$

(ii) En el caso de tres sucesos, A , B y C , tenemos

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C)\mathbb{P}(C),$$

siempre que las probabilidades que aparecen estén bien definidas, es decir $\mathbb{P}(B \cap C) > 0$ y $\mathbb{P}(C) > 0$.

II.3.3. Propiedad

Para un suceso B fijado, la probabilidad condicionada a B , $\mathbb{P}(\cdot|B)$ satisface todas las propiedades de una probabilidad. En particular cumple por ejemplo, para cualquier suceso A , $0 \leq \mathbb{P}(A|B) \leq 1$, $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$; y para dos sucesos A y C , $\mathbb{P}(A \cup C|B) = \mathbb{P}(A|B) + \mathbb{P}(C|B) - \mathbb{P}(A \cap C|B)$.

II.4. Sucesos independientes

Una de las situaciones más importantes en probabilidad aparece cuando, considerando dos sucesos, el hecho de que uno ocurra no influye la probabilidad de que el otro ocurra. Este concepto se llama independencia de dos sucesos y pasamos a definirlo.

II.4.1. Definición para dos sucesos

A y B son dos sucesos independientes si se cumple

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Notar que esta definición es equivalente, por la definición de la probabilidad condicionada, a que $\mathbb{P}(A|B) = \mathbb{P}(A)$ y $\mathbb{P}(B|A) = \mathbb{P}(B)$. Es decir A y B son independientes si el hecho de saber que, por ejemplo, B ha ocurrido, no cambia la probabilidad que asignamos a A , y vice versa.

Una buena ilustración de este concepto: “¿Sabéis por qué un matemático siempre se lleva una bomba cuando viaja en avión? - Porque es mucho menos probable que haya dos bombas en un mismo avión que sólo una...”

¿Qué os parece este argumento?

II.4.2. Definición para n sucesos

Los n sucesos A_1, A_2, \dots, A_n son independientes si para cualquier subfamilia $A_{i_1}, A_{i_2}, \dots, A_{i_k}$, se cumple

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

En particular se cumple que $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ para todo i y j entre 1 y n .

II.5. Ejemplos de probabilidades condicionadas en la vida diaria

II.5.1. Eslogan publicitario para la lotería

En Francia, hubo hace unos años, un eslogan publicitario para un juego de lotería que rezaba:

El 100 % de los acertantes probaron suerte...

Los creadores de este eslogan jugaron con el efecto causado al usar una probabilidad condicionada: si P denota el suceso “probar suerte” y A el suceso “acertar”, el eslogan está diciendo $\mathbb{P}(P|A) = 1$, pero la gente piensa en $\mathbb{P}(A|P)$ que es muchísima más baja por supuesto...

II.5.2. Tabaquismo y cáncer de pulmón

Del informe “La situación del cáncer en España, 2005”, elaborado por el Centro nacional de Epidemiología, se deduce en particular los datos siguientes: el cáncer es la primera causa de muerte en términos absolutos (p9), y en particular, el cáncer de pulmón es el tumor de mayor incidencia y de mayor mortalidad entre los hombres. Por otra parte, en la información publicada por la Asociación Española contra el Cáncer (AECC) en su página web, se indica que el 90 % de los pacientes con cáncer de pulmón son fumadores.

¿Se puede deducir de esta última estadística de que el tabaquismo es un factor de riesgo para el cáncer de pulmón? En principio, parece que sí, pero en realidad ¡depende de la tasa de fumadores en la población!

Traduzcamos estos datos con sucesos: consideramos el experimento “escoger una persona al azar en la población española”. Introducimos los sucesos T = “tiene un tumor asociado a un cáncer de pulmón”, F = “es fumador”. Nos dicen que

$$P(F|T) = 0,9$$

pero en realidad, para saber si el hecho de ser fumador incrementa el riesgo de desarrollar un cáncer de pulmón, queremos saber si $\mathbb{P}(T|F)$ es mayor que $\mathbb{P}(T)$.

Para relacionar $\mathbb{P}(T|F)$ y $\mathbb{P}(T)$, podemos utilizar la definición de la probabilidad condicionada:

$$\mathbb{P}(T|F) = \frac{\mathbb{P}(T \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F|T)\mathbb{P}(T)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F|T)}{\mathbb{P}(F)} \times \mathbb{P}(T).$$

Por lo tanto, el concluir si el tabaquismo incrementa el riesgo de desarrollar un cáncer de pulmón dependerá del cociente $\mathbb{P}(F|T)/\mathbb{P}(F)$.

Según la Encuesta Nacional de Salud 2003, que se puede obtener del Instituto Nacional de Estadística, aproximadamente el 30% de la población española son fumadores diarios. El cociente $\mathbb{P}(F|T)/\mathbb{P}(F)$ es por lo tanto igual aproximadamente a $0.9/0.3=3$. Deducimos que **el hecho de ser un fumador diario multiplica por tres el riesgo de padecer un cáncer de pulmón.**

Pero, se puede enfatizar que la única afirmación “El 90% de los pacientes con cáncer de pulmón son fumadores” no implica de por sí que el tabaquismo sea un factor de riesgo para el cáncer de pulmón.

II.5.3. Tabaquismo y esperanza de vida

Un dato clásico en epidemiología es muy sorprendente en primera lectura:

Si nos limitamos a la gente mayor de 70 años, ¡la esperanza de vida de los fumadores es mayor que la de los no fumadores!

¿Cómo puede ser esto cierto? En realidad este dato no es tan sorprendente si uno se lo piensa un poco: muy poca gente llega a los 70 años fumando... De hecho, según la AECC, la edad media de fallecimiento por cáncer de pulmón es de 68 años para los hombres y 66 años para las mujeres. La gente que llega a los 70 años y son fumadores tienen un sistema inmunológico muy resistente y un mecanismo de control de células tumorosas muy eficiente, lo que implica que, de todas maneras, tendrían una vida muy larga...

II.6. Fórmula de la probabilidad total y teorema de Bayes

II.6.1. Condiciones de aplicación

- Tenemos n sucesos A_1, A_2, \dots, A_n que forman una partición del espacio muestral S , es decir que son mutuamente incompatibles ($A_i \cap A_j = \emptyset$, para $1 \leq i, j \leq n$), y cuya unión es el espacio muestral entero, i.e. $A_1 \cup A_2 \cup \dots \cup A_n = S$. Además conocemos la probabilidad de cada uno de ellos, es decir $\mathbb{P}(A_1), \mathbb{P}(A_2), \dots, \mathbb{P}(A_n)$.

Nota: A menudo nos encontraremos con la situación en la que sólo son dos sucesos, i.e. $n = 2$, en este caso tenemos $A_1 = A$ y $A_2 = A^c$.

- Tenemos otro suceso B y, para cada $i = 1, \dots, n$, conocemos, en el caso de que ocurra A_i , la probabilidad de B , es decir conocemos $\mathbb{P}(B|A_1), \dots, \mathbb{P}(B|A_n)$.

II.6.2. Los resultados

Si se cumplen las condiciones de aplicación del apartado anterior,

- **Fórmula de la probabilidad total** Se puede calcular $\mathbb{P}(B)$ descomponiendo B sobre cada uno de los sucesos de la partición:

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \dots + \mathbb{P}(B|A_n)\mathbb{P}(A_n).$$

- **Teorema de Bayes** Para cualquier $i = 1, \dots, n$, tenemos

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1) + \dots + \mathbb{P}(B|A_n)\mathbb{P}(A_n)}.$$

Demostración. Utilizamos, al formar A_1, A_2, \dots, A_n una partición del espacio muestral, la descomposición del suceso B

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n).$$

Los sucesos $(B \cap A_1), \dots, (B \cap A_n)$ son incompatibles dos a dos, y aplicamos la regla de la adición

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \dots + \mathbb{P}(B \cap A_n).$$

Utilizamos ahora la regla del producto $\mathbb{P}(B \cap A_i) = \mathbb{P}(B|A_i)\mathbb{P}(A_i)$ para todo $i = 1, \dots, n$ para la fórmula de la probabilidad total

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \dots + \mathbb{P}(B|A_n)\mathbb{P}(A_n).$$

Por otra parte por la definición de la probabilidad condicionada $\mathbb{P}(A_i|B) = \mathbb{P}(A_i \cap B)/\mathbb{P}(B)$, para todo $1 \leq i \leq n$. Si sustituimos en el numerador $\mathbb{P}(A_i \cap B)$ por $\mathbb{P}(B|A_i)\mathbb{P}(A_i)$ y en el denominador $\mathbb{P}(B)$ por la fórmula de la probabilidad total, obtenemos el teorema de Bayes. \square

II.6.3. Ejemplo

En un la transmisión de un mensaje por correo electrónico, la calidad de la recepción de un fichero depende de la sobrecarga de la red. Si la red está sobrecargada, la proporción de ficheros dañados es de 1 %, mientras que si no lo está, esta proporción sólo es del 0.01 %. Estimamos que la probabilidad de que la red esté sobrecargada es igual a 0.02. ¿Cuál es la proporción total de ficheros dañados en la transmisión? Suponiendo que recibo un fichero dañado, ¿cuál es la probabilidad de que la red estuviera sobrecargada durante la transmisión?

Empecemos por introducir los sucesos convenientes para traducir los datos que se nos proporciona. Sea RS el suceso “La red está sobrecargada”, y D el suceso “El archivo está dañado”. Se nos pide calcular $\mathbb{P}(D)$ y $\mathbb{P}(RS|D)$. Nos damos cuenta de que si $A_1 = RS$ y $A_2 = RS^C$, los sucesos A_1 y A_2 son incompatibles y su reunión es el espacio muestral entero, por lo tanto forman una partición del espacio muestral. Además conocemos sus probabilidades: $\mathbb{P}(A_1) = 0,02$ y $\mathbb{P}(A_2) = 0,98$. Por otra parte conocemos $\mathbb{P}(D|A_1)$, y $\mathbb{P}(D|A_2)$, estamos por lo tanto en las condiciones de aplicación de la fórmula de la probabilidad total y del teorema de Bayes. Deducimos

$$\begin{aligned} \mathbb{P}(D) &= \mathbb{P}(D|RS)\mathbb{P}(RS) + \mathbb{P}(D|RS^C)\mathbb{P}(RS^C) \\ &= 0,01 \cdot 0,02 + 0,0001 \cdot 0,98 = 0,000298 \simeq 0,0003, \end{aligned}$$

es decir que la proporción total de ficheros dañados es de 3 por 10000. Por otra parte, por el teorema de Bayes,

$$\begin{aligned} \mathbb{P}(RS|D) &= \frac{\mathbb{P}(D|RS)\mathbb{P}(RS)}{\mathbb{P}(D|RS)\mathbb{P}(RS) + \mathbb{P}(D|RS^C)\mathbb{P}(RS^C)} \\ &= \frac{0,01 \cdot 0,02}{0,000298} \simeq 0,67, \end{aligned}$$

por lo tanto, sabiendo que recibo un fichero dañado, la probabilidad de que la red estuviera sobrecargada es igual a 0.67.



2013: Año Internacional de la Estadística. ¿Sabías qué...?

Entre los algoritmos anti-spam más eficientes que usan los proveedores de correo como Gmail o Hotmail, están los basados en estadística Bayesiana, que estiman la probabilidad de que un determinado mensaje sea spam, al comparar sus contenidos con una gran base de datos de mensajes indeseados....

Fuente: J. A. Zdziarski “Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification”, (2005), No Starch Press.

statistics

TEMA III

Variable aleatoria I

Las nociones teóricas que hemos introducido responden a la necesidad de construir modelos matemáticos que den cuenta del carácter aleatorio de los fenómenos que nos interesan. Hemos puesto en el tema anterior las primeras piedras en este sentido describiendo experimento aleatorio, sucesos y probabilidad asociada a un suceso, pero nos falta la noción fundamental de variable aleatoria: en problemas concretos, estamos interesados en funciones definidas sobre el espacio de los resultados posibles del experimento aleatorio, y los sucesos que queremos estudiar se expresan a través de estas funciones. Puesto que nos es imposible predecir de manera exacta el valor de una variable aleatoria al realizar el experimento, nuestro modelo consistirá en describir las probabilidades asociadas a cualquier suceso relacionado con esta variable, descripción que conseguiremos gracias a la función de distribución.

III.1. Concepto de variable aleatoria

Consideramos un experimento aleatorio y su espacio muestral asociado.

III.1.1. Definición

Una variable aleatoria- de ahora en adelante v.a.- asocia un número o más generalmente una característica a todo resultado posible del experimento.

Por ejemplo, si consideramos el experimento que consiste en realizar una medición de la concentración de un producto en una solución, nos interesa la v.a X = “valor medido de la concentración.” Otro ejemplo de variable aleatoria se asocia, en un proceso de fabricación, al experimento de escoger un dispositivo producido, y considerar la v.a. X = “duración hasta el fallo”.

Finalmente ilustraremos algunos conceptos de este tema con un ejemplo sencillo: el experimento consiste en lanzar tres veces una moneda no trucada. Si denotamos por $+$ el resultado “cruz” y por c el resultado “cara” al lanzar una moneda, el espacio

muestral se describe como

$$S = \{ccc, cc+, c+c, c++, +cc, +c+, ++c, +++\}.$$

Consideraremos la v.a. X = “número de veces que ha salido cruz en los tres lanzamientos”. Puede tomar cualquiera de los valores 0, 1, 2 y 3.

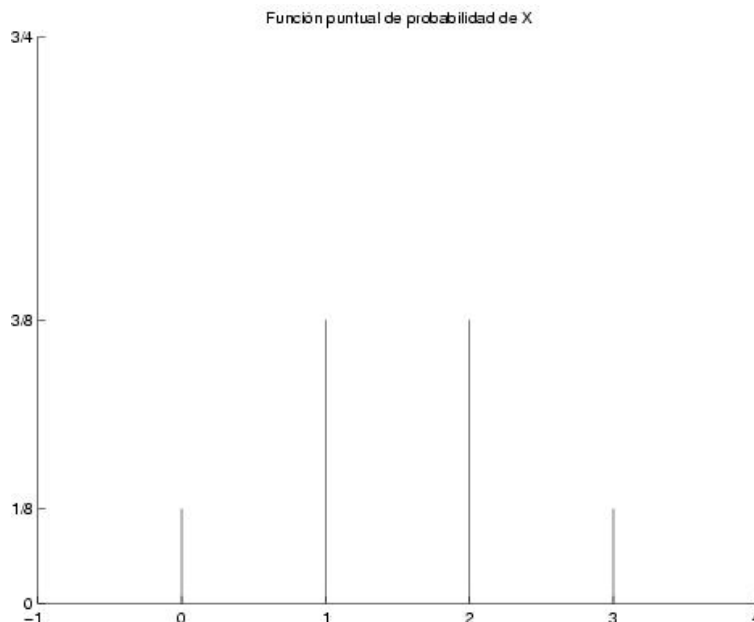
III.1.2. Distribución de una variable aleatoria

Conocer la distribución de los valores de una v.a. X consiste en saber asignar a cualquier suceso relacionado con X una probabilidad. Decidir de una distribución para una v.a. de interés en un problema concreto es por lo tanto escoger un modelo para describir el comportamiento de esta variable.

Para el ejemplo de los tres lanzamientos de una moneda, la distribución de X = “número de veces que ha salido cruz en los tres lanzamientos” está completamente determinada por la lista de los valores posibles junto con la probabilidad con la que X toma cada valor. Al ser la moneda no trucada, escogemos un modelo en el que los sucesos elementales de S son equiprobables, calculamos $\mathbb{P}(X = i)$ para $i = 0, 1, 2, 3$ con la regla casos favorables / casos posibles y obtenemos

Valor	Probabilidad
0	1/8
1	3/8
2	3/8
3	1/8

Se puede representar de manera gráfica la distribución de X :



Podremos fijarnos en las características principales de esta distribución (simetría, máximo, colas...)

III.2. Función de distribución de una v.a

Se trata de una manera de describir la distribución de una variable X .

III.2.1. Definición

La función de distribución de una v.a. X es la función F_X que asocia a cualquier número real t la probabilidad de que X sea menor o igual a t , i.e.

$$F_X(t) = \mathbb{P}(X \leq t).$$

III.2.2. Cálculo para el ejemplo de las tres monedas

Para calcular $F_X(t) = \mathbb{P}(X \leq t)$, debemos considerar los intervalos definidos por los valores posibles de X es decir 0, 1, 2 y 3 que inducen los cinco intervalos para t : $t < 0$, $0 \leq t < 1$, $1 \leq t < 2$, $2 \leq t < 3$ y $t > 3$.

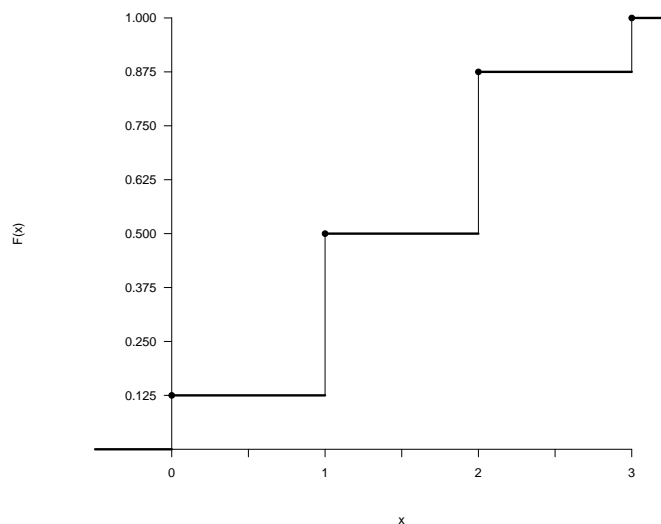
Si $t < 0$, el suceso $(X \leq t)$ es el suceso imposible puesto que todos los valores que puede tomar X son mayores o igual que 0. Por lo tanto, $F_X(t) = \mathbb{P}(X \leq t) = 0$. Si $0 \leq t < 1$, el suceso $(X \leq t)$ se cumple si y solamente si X toma el valor 0. Deducimos $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X = 0) = 1/8$.

Si $1 \leq t < 2$, el suceso $(X \leq t)$ se cumple si y solamente si X toma el valor 0 ó 1, es decir $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}[(X = 0) \cup (X = 1)] = \mathbb{P}[X = 0] + \mathbb{P}[X = 1] = 1/8 + 3/8 = 1/2$.

Si $2 \leq t < 3$, el suceso $(X \leq t)$ se cumple si y solamente si X toma el valor 0, 1 ó 2, es decir $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \mathbb{P}[X = 2] = 1/2 + 3/8 = 7/8$.

Finalmente, si $t > 3$, el suceso $(X \leq t)$ es el suceso seguro puesto que todos los valores que puede tomar X son menores o igual que 3. Por lo tanto $F_X(t) = \mathbb{P}(X \leq t) = 1$.

La gráfica de F_X en este ejemplo es



III.2.3. Propiedades

La función de distribución de una v.a. X cumple las propiedades siguientes:

- $0 \leq F_X(t) \leq 1$, para todo $t \in \mathbb{R}$.
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$ mientras que $\lim_{t \rightarrow +\infty} F_X(t) = 1$.
- F_X es una función creciente, puesto que si $a \leq b$, tenemos $(X \leq a) \subset (X \leq b)$ lo que implica que $\mathbb{P}(X \leq a) \leq \mathbb{P}(X \leq b)$.
- F_X es una función continua por la derecha.
- Finalmente la propiedad más importante que utilizaremos muy a menudo: para todos números reales $a \leq b$,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

La demostración de esta propiedad es inmediata si utilizamos la descomposición $(X \leq b) = (X \leq a) \cup (a < X \leq b)$ junto con la regla de la adición.

III.3. Variable aleatoria discreta

III.3.1. Definición

En el caso en que la v.a. X puede tomar un número finito o infinito numerable¹ de valores. En el ejemplo de los tres lanzamientos de una moneda, la v.a. $X =$ “Número de veces que ha salido cruz” es una v.a discreta puesto que sólo puede tomar cuatro valores.

III.3.2. Función puntual de probabilidad

III.3.2.1. Definición

Si X es una v.a. discreta, y $x_1, x_2, \dots, x_n, \dots$ representan sus valores posibles, la función puntual de probabilidad de X es la función f_X que asocia a cada x_i la probabilidad $\mathbb{P}(X = x_i)$, para $i = 1, \dots, n \dots$

$$f_X : x_i \mapsto f_X(x_i) = \mathbb{P}(X = x_i).$$

Ejemplo. En el experimento del lanzamiento de las tres monedas, hemos calculado la distribución de X , el número de veces que ha salido cruz en el apartado 1.2. Los valores posibles de X son 0, 1, 2 y 3; por lo tanto

Valor	f_X
0	1/8
1	3/8
2	3/8
3	1/8

¹Un conjunto infinito numerable es un conjunto del que se puede enumerar todos los elementos. \mathbb{N} , \mathbb{Z} y \mathbb{Q} son ejemplos de conjuntos infinitos numerables. En cambio un conjunto infinito no numerable es un conjunto que no se puede poner en biyección con \mathbb{N} , es decir para el cual es imposible enumerar los elementos. El intervalo de números reales $[0, 1]$ es infinito no numerable por ejemplo.

III.3.2.2. Propiedades

- La función puntual de probabilidad de una v.a. discreta permite calcular la función de distribución: si notamos que

$$(X \leq t) = \cup_{x_i \leq t} (X = x_i),$$

obtenemos que

$$\mathbb{P}(X \leq t) = \sum_{x_i \leq t} \mathbb{P}(X = x_i) = \sum_{x_i \leq t} f_X(x_i).$$

- Además, si consideremos dada una función f definida en un conjunto discreto de valores $\{x_1, \dots, x_n, \dots\}$, se puede demostrar que f es una función puntual de probabilidad de una v.a. X si y solamente si cumple

- $0 \leq f(x)$ para $x = x_1, \dots, x_n, \dots$
- $\sum_{x_i} f_X(x_i) = 1.$

III.3.3. Características de una variable discreta

Al igual que en el tema 1 para un conjunto de datos, queremos disponer de herramientas para describir la distribución de valores de una v.a. De hecho, todas las medidas descriptivas de un conjunto de datos tienen su contra-parte para la distribución de una v.a. Nos limitaremos por razones de tiempo a una medida de centralización y otra de dispersión: la esperanza y la varianza.

III.3.3.1. Esperanza

Si queremos considerar el valor medio de la distribución de valores de una v.a., es natural calcular la suma de estos valores ponderados por la probabilidad que se le asigna.

Definición III.3.1 *La media, o esperanza, o valor esperado, o promedio, de una v.a. discreta X se define como*

$$\mathbb{E}[X] = \sum_{x_i} x_i \mathbb{P}(X = x_i).$$

Representa una medida de centralización de la distribución de valores de X pero con la misma puntualización que en el tema 1: es representativa del centro de la distribución si ésta es aproximadamente simétrica pero puede ser una mala medida de centralización si la distribución es asimétrica y/o presenta colas pronunciadas.

Por supuesto, la esperanza de una v.a. X se expresa en las mismas unidades que X .

Será útil para una distribución de valores ser capaz de calcular el valor medio no solamente de X sino también de una función de X ; está claro por ejemplo que el valor medio de la distancia al cuadrado de X a su media será una medida de dispersión de la distribución de valores de X . Por ello, definimos la esperanza de una función cualquiera $f(X)$ de X .

Definición III.3.2 Sea X una v.a. discreta y f una función de \mathbb{R} en \mathbb{R} . La esperanza de $f(X)$ es la suma de los valores de $f(X)$ ponderados por la probabilidad de que X tome cada valor, es decir,

$$\mathbb{E}[f(X)] = \sum_{x_i} f(x_i)\mathbb{P}(X = x_i).$$

III.3.3.2. Varianza

Para disponer de una medida numérica de la dispersión de valores de una v.a. X , calcularemos el valor promedio de la distancia al cuadrado de X a su media. Al igual que en el tema 1, llamamos esta cantidad la varianza de X .

Definición III.3.3 La varianza de una v.a. discreta X , designada por $\text{var } X$ o σ_X^2 , está definida por

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Por la definición III.3.2 deducimos que $\text{var}(X)$ se puede calcular como

$$\text{var}(X) = \sum_{x_i} (x_i - \mathbb{E}[X])^2 \mathbb{P}(X = x_i).$$

Por otra parte, se suele calcular la varianza utilizando la fórmula equivalente siguiente:

Fórmula equivalente para el cálculo de la varianza. Tenemos

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Demostración:

$$\begin{aligned} \text{var}(X) &= \sum_{x_i} (x_i - \mathbb{E}[X])^2 \mathbb{P}(X = x_i) \\ &= \sum_{x_i} (x_i^2 - 2x_i\mathbb{E}[X] + \mathbb{E}[X]^2) \mathbb{P}(X = x_i) \\ &= \sum_{x_i} x_i^2 \mathbb{P}(X = x_i) - \sum_{x_i} 2x_i\mathbb{E}[X] \mathbb{P}(X = x_i) + \sum_{x_i} \mathbb{E}[X]^2 \mathbb{P}(X = x_i) \\ &= \sum_{x_i} x_i^2 \mathbb{P}(X = x_i) - 2\mathbb{E}[X] \sum_{x_i} x_i \mathbb{P}(X = x_i) + \mathbb{E}[X]^2 \sum_{x_i} \mathbb{P}(X = x_i) \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}$$

□

Finalmente, la desviación típica se define como la raíz cuadrada de la varianza

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Será la medida que calcularemos para dar cuenta de la dispersión de la distribución: cuanto más pequeña sea la desviación típica, más concentrada estará la distribución alrededor de su media. En particular, si la desviación típica de X es nula, deducimos

por la primera fórmula para el cálculo de la varianza, que todos los valores de X son iguales: X sólo puede tomar un valor, y lo toma con probabilidad 1.

Por otra parte, es bueno resaltar que la desviación típica se expresa en las mismas unidades que la variable X .

Nota III.3.1 *En la fórmula equivalente para la varianza aparecen las cantidades $\mathbb{E}[X^2]$ y $\mathbb{E}[X]$. En general para un entero k , llamamos a $\mathbb{E}[X^k]$ el momento de orden k . Así la media es el momento de orden 1. También hablamos de momento centrado de orden k para la cantidad $\mathbb{E}[(X - \mathbb{E}[X])^k]$. La varianza es por lo tanto el momento centrado de orden 2.*

III.3.3.3. Ejemplo

Calculemos para el ejemplo del lanzamiento de tres monedas la esperanza y la varianza de la v.a X "número de cruces".

Por una parte,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x_i} x_i \mathbb{P}(X = x_i) = 0 \cdot 1/8 + 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8 \\ &= 3/2\end{aligned}$$

y por otra parte

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sum_{x_i} x_i^2 \mathbb{P}(X = x_i) - (3/2)^2 \\ &= 0^2 \cdot 1/8 + 1^2 \cdot 3/8 + 2^2 \cdot 3/8 + 3^2 \cdot 1/8 - (3/2)^2 \\ &= 3/4\end{aligned}$$

La desviación típica es por lo tanto

$$\sigma_X = \sqrt{3}/2.$$

III.3.4. Modelos más usados de v.a. discretas

No debemos olvidar que nuestro objetivo es modelizar un fenómeno. Proponer un modelo no consiste en proporcionar una descripción de la realidad, sino disponer de una aproximación que dé cuenta de los resultados observados del experimento para unas condiciones experimentales dadas. Ningún modelo se ajusta perfectamente al fenómeno observado, así que considerarlo adecuado o válido es equivalente a considerar que el grado de precisión conseguido es satisfactorio para el uso que queremos hacer del modelo.

En este contexto, hay situaciones típicas de modelización que presentan las mismas características y para las cuales se han propuesto modelos de distribuciones bien estudiados y conocidos.

III.3.4.1. Variable de Bernoulli

Se trata de una variable que sólo puede tomar dos valores, 0 ó 1. Llamamos p la probabilidad de que tome el valor 1. Varios valores de p , (comprendidos entre 0 y 1,

puesto que p es una probabilidad) dan varias distribuciones de Bernoulli. Para un valor p concreto, hablamos de la distribución de Bernoulli de *parámetro* p .

Propiedades

- Valores posibles: $\{0, 1\}$,
 $\mathbb{P}(X = 0) = 1 - p$ $\mathbb{P}(X = 1) = p$.

- Esperanza:

$$\mathbb{E}[X] = \sum x_i \mathbb{P}(X = x_i) = 0 \times (1 - p) + 1 \times p = p$$

- Varianza:

Tenemos: $\mathbb{E}[X^2] = \sum x_i^2 \mathbb{P}(X = x_i) = 0^2 \times (1 - p) + 1^2 \times p = p$, por lo tanto

$$\text{var}(X) = p - p^2 = p(1 - p).$$

Ejemplo. Transmito un fichero por la red, en promedio 3 de cada 10000 ficheros transmitidos resultan dañados. Al experimento aleatorio: “transmitir un fichero por la red”, asocio la variable X que toma el valor 1 si el fichero se transmite correctamente y 0 si resulta dañado. La variable X sigue una distribución de Bernoulli de parámetro 0,9997.

III.3.4.2. Distribución binomial

a). Definición La distribución binomial aparece cuando se dan las condiciones siguientes:

- Tenemos un primer experimento aleatorio simple, con una situación dicotómica, es decir una situación con dos sucesos posibles A y A^c (o ocurre A o no ocurre A).
- Repetimos este experimento simple n veces de manera independiente.
- Consideramos la variable X = “Número de veces que ha ocurrido A en las n realizaciones del experimento simple.

En esta situación, la variable X sigue una distribución Binomial, de parámetros n (el número de veces que repetimos el experimento simple) y p (la probabilidad de que, en una realización del experimento simple, ocurra A). Lo denotamos por

$$X \sim \mathcal{B}(n, p),$$

donde el símbolo \sim se utiliza para “sigue una distribución”...

b). Ejemplo Una empresa produce piezas con 1 % de defectuosas. Las piezas se empaquetan en cajas de 10 unidades. Si consideramos el experimento aleatorio que consiste en escoger al azar una caja entre la producción, ¿cuál es la distribución de la variable X = “número de piezas defectuosas en la caja”.

Para completar una caja, se ha repetido 10 veces el experimento aleatorio simple “escojo una pieza en la producción” al que va asociado una situación dicotómica: o bien ocurre A = “la pieza escogida es defectuosa”, o bien ocurre A^c = “la pieza

escogida es correcta”. Contar el número de piezas defectuosas en la caja es por lo tanto equivalente a contar el número de veces que ha ocurrido A entre las 10 realizaciones del experimento simple. Deducimos que la distribución de X es una distribución Binomial con parámetros $n = 10$, y $p = \mathbb{P}(A)$, la probabilidad de que ocurra A en el experimento simple. Concluimos

$$X \sim \mathcal{B}(10, 0,01).$$

c). Propiedades

- Valores posibles: $0, 1, 2, \dots, n$.
- Distribución - Función puntual de probabilidad. $i = 0, 1, \dots, n$ $f_X(i) = \mathbb{P}(X = i)$. Para calcular estas probabilidades, introduzcamos los sucesos:
 - $A_1 =$ “ha ocurrido A en la primera realización del exp. simple”
 - $A_2 =$ “ha ocurrido A en la segunda realización del exp. simple”
 - \vdots
 - $A_n =$ “ha ocurrido A en la n -ésima realización del exp. simple”

Estos sucesos son independientes.

Empecemos por calcular $\mathbb{P}(X = 0)$:

El suceso $X = 0$ se puede escribir $A_1^c \cap A_2^c \cap \dots \cap A_n^c$, por lo tanto

$$\mathbb{P}(X = 0) = \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_n^c) = \mathbb{P}(A_1^c) \dots \mathbb{P}(A_n^c) = (1 - p)^n,$$

por la regla del producto para sucesos independientes.

De manera similar, calculamos $\mathbb{P}(X = 1)$:

El suceso $(X = 1)$ se escribe como

$$(X = 1) = (A_1 \cap A_2^c \cap \dots \cap A_n^c) \cup (A_1^c \cap A_2 \cap \dots \cap A_n^c) \cup \dots \cup (A_1^c \cap A_2^c \cap \dots \cap A_n)$$

Aplicando la regla de la adición para sucesos incompatibles y a continuación la regla del producto para sucesos independientes, obtenemos

$$\begin{aligned} \mathbb{P}(X = 1) &= \mathbb{P}(A_1 \cap A_2^c \cap \dots \cap A_n^c) + \mathbb{P}(A_1^c \cap A_2 \cap \dots \cap A_n^c) + \dots \\ &\quad + \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_n) \\ &= p(1 - p)^{n-1} + p(1 - p)^{n-1} + \dots + p(1 - p)^{n-1} = np(1 - p)^{n-1} \end{aligned}$$

De la misma manera, podemos demostrar que, para un i cualquiera entre 0 y n , la probabilidad $\mathbb{P}(X = i)$ se descompone como la suma de términos todos iguales, siendo el primero de ellos $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_i \cap A_{i+1}^c \cap \dots \cap A_n^c)$, que es igual a $p^i(1 - P)^{n-i}$. Sólo nos queda determinar el número de términos en esta suma, corresponde al número de maneras de escoger i sucesos diferentes entre n : es una cantidad básica en combinatoria, se llama el número de combinaciones de n elementos tomados de i en i , y se denota por $\binom{n}{i}$. En resumen, para $i = 0, 1, \dots, n$,

$$f_X(i) = \mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i},$$

donde

$$\binom{n}{i} = \frac{n!}{i! \cdot (n-i)!},$$

y se utiliza la convención $0! = 1$.

Nota: ¿se cumple que $\sum_{i=1}^n \binom{n}{i} p^i (1-p)^{n-i} = 1$? La respuesta es sí, por el

binomio de Newton: $(a+b)^n = \sum_{i=1}^n \binom{n}{i} a^i b^{n-i}$, y por lo tanto

$$\sum_{i=1}^n \binom{n}{i} p^i (1-p)^{n-i} = (p+1-p)^n = 1.$$

- Esperanza y varianza:
Es posible demostrar que, si $X \sim \mathcal{B}(n, p)$,

$$\mathbb{E}[X] = n \cdot p, \quad \text{var}(X) = n \cdot p \cdot (1-p).$$

III.3.4.3. Distribución Geométrica

a). Definición Es el modelo más sencillo para un tiempo de espera discreto: consideramos, al igual que para una distribución binomial, un experimento simple con una situación dicotómica, ocurre A o A^C con probabilidades p y $1-p$ respectivamente. Estamos dispuestos a realizar este experimento simple un cierto número de veces hasta que ocurra A . Introducimos la variable X : "Número de veces que debemos realizar el experimento simple hasta que ocurra A por primera vez".

La variable X sigue una distribución geométrica de parámetro p . Escribimos

$$X \sim \mathcal{Geo}(p)$$

b). Propiedades

- X puede tomar los valores $1, 2, \dots$
- Función puntual de probabilidad de X : queremos calcular $\mathbb{P}(X = i)$ para $i \in \mathbb{N}^*$.
Introducimos los sucesos: A_1 ="ocurre A en la primera realización del experimento simple", A_2 ="ocurre A en la segunda realización del experimento simple", etc....

Está claro que

$$\mathbb{P}(X = i) = \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_{i-1}^c \cap A_i),$$

y, por la regla del producto para sucesos independientes, deducimos

$$\mathbb{P}(X = i) = (1-p)^{i-1} p.$$

- Esperanza y varianza de $X \sim \mathcal{Geo}(p)$.
Utilizando resultados clásicos sobre suma de series geométricas, obtenemos

$$\begin{aligned} \mathbb{E}[X] &= 1/p, \\ \text{Var}(X) &= \frac{1-p}{p^2}. \end{aligned}$$

III.3.4.4. Distribución de Poisson

a). **Definición** La distribución de Poisson aparece en situaciones en las que se cuenta el número de apariciones de un determinado suceso o bien en un intervalo de tiempo dado (como el número de partículas emitidas en un segundo por un material radioactivo, o el número de clientes que llegan a una cola en un intervalo de tiempo dado) o bien en un recinto físico (como el número de fallos en un metro de alambre de hierro producido).

Si λ es el número medio de apariciones del suceso de interés por intervalo de tiempo, la variable $X =$ “número de veces que ha aparecido el suceso en un intervalo de tiempo escogido al azar”, sigue una distribución de Poisson de parámetro λ . Escribimos

$$X \sim \mathcal{P}(\lambda).$$

b). Propiedades

- Valores posibles: $0, 1, \dots, n, \dots$, es decir todos los números enteros...
- Función puntual de probabilidad: para $i = 0, 1, \dots$,

$$f_X(i) = \mathbb{P}(X = i) = \frac{\lambda^i e^{-\lambda}}{i!}.$$

Podemos comprobar que $\sum_{i=0}^{+\infty} \frac{\lambda^i e^{-\lambda}}{i!} = 1$, si utilizamos el hecho de que la suma de la serie de potencias $\sum_{i=0}^{+\infty} \frac{x^i}{i!} = e^x$.

- Esperanza y varianza.

Es fácil comprobar repitiendo cálculos similares a los del punto anterior, que la esperanza de una distribución de Poisson de parámetro λ , es, tal como se anunció en la definición, λ . Por otra parte, se puede demostrar que su varianza es λ también: si $X \sim \mathcal{P}(\lambda)$

$$\mathbb{E}[X] = \lambda, \quad \text{var}(X) = \lambda.$$

III.4. Variable continua

III.4.1. Definición

Si una v.a X puede tomar un número infinito no numerable de valores, se le llama v.a continua.

III.4.2. Función de densidad

III.4.2.1. Presentación

Queremos disponer de una manera de describir la distribución de una v.a continua, es decir que nos permita calcular la probabilidad asignada a cualquier suceso relacionado con X . Para una v.a discreta, hemos visto que utilizamos la función puntual de probabilidad que asocia a cada valor posible la probabilidad de que X tome este valor: el cálculo de la probabilidad de un suceso involucra entonces una suma de valores de la función puntual de probabilidad. Puesto que una v.a continua

puede tomar un número infinito no numerable de valores, no asignaremos una probabilidad a cada valor posible, sino que definiremos una “densidad” de probabilidad, que indique en qué zonas del espacio de los valores posibles de X es más probable que se encuentre X .

III.4.2.2. Definición

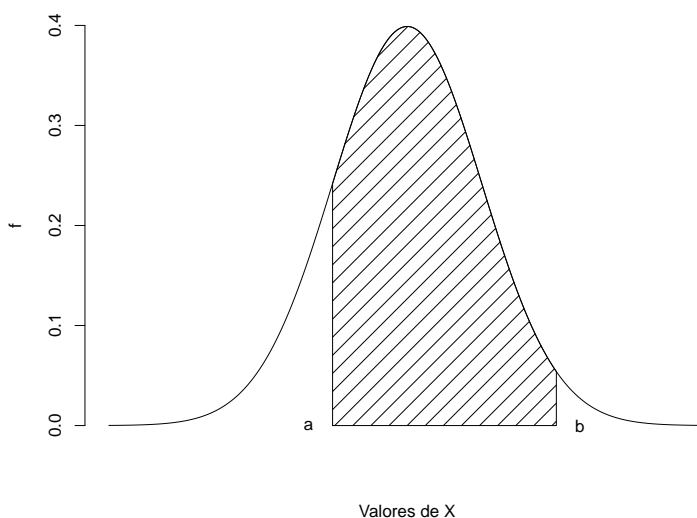
Para una v.a continua X existe una función f_X positiva, tal que, para todos a y b , $a \leq b$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

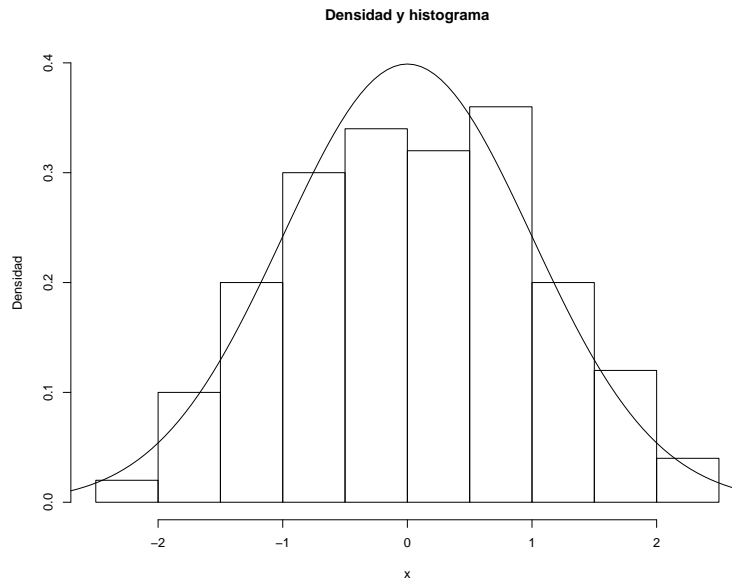
La función f_X se llama la función de densidad de la v.a X . Notar que se trata de una terminología coherente con la analogía mencionada anteriormente entre probabilidad y peso: para un cuerpo no homogéneo, el peso de una parte de este cuerpo se calcula integrando la densidad en el volumen correspondiente.

Nota:

- Al ser f_X una función positiva, y $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$, la probabilidad de que X esté entre a y b corresponde al área debajo de la curva de f_X comprendida entre a y b , tal como está ilustrado en la figura siguiente:



- Si disponemos de un conjunto de datos con una variable X , generados a partir de realizaciones de un experimento, y si nuestra descripción del mecanismo de generación de los datos a través de un modelo para X , es adecuada, la función de densidad de X tiene mucha relación con el histograma. En efecto, la probabilidad de que X pertenezca a una clase debe explicar la frecuencia de datos que aparecen en esta clase, y por lo tanto la forma del histograma debe corresponder a la forma de la densidad, tal como viene reflejado en la figura:



- El área total debajo de la curva de f_X debe corresponder a la probabilidad de que X tome un valor real, y es igual a 1:

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

- Si X es una v.a continua, la probabilidad de que tome un valor dado a es nula, puesto que la integral de f_X entre a y a es cero: la distribución de una v.a continua sólo asigna probabilidades positivas a intervalos de valores y no a puntos individuales. En particular deducimos por la regla de la adición que, si X es una v.a continua,

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b).$$

¡Por supuesto este tipo de igualdades no es válida en general para una v.a discreta!

III.4.2.3. Propiedades

a). Relaciones entre f_X y F_X . La función de distribución acumulada de X es, ver sección III.2 calcula para todo real t la probabilidad de que X tome un valor menor o igual que t : $F_X(t) = \mathbb{P}(X \leq t)$. Por la definición de la función de densidad f_X deducimos que

$$F_X(t) = \int_{-\infty}^t f_X(x)dx.$$

Por lo tanto, F_X es una primitiva de f_X , o equivalentemente, f_X se puede calcular como la derivada, en los puntos donde existe, de la función de distribución acumulada $t \mapsto F_X(t)$.

b). **Condiciones para que una función f sea la función de densidad de una v.a continua X .** Está claro que, para que una función f sea la función de densidad de una v.a continua X , es necesario que se cumplan las dos condiciones:

1. $f(x) \geq 0$, para todo $x \in \mathbb{R}$,
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

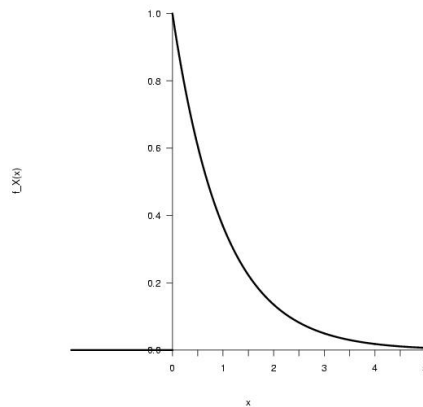
Se puede demostrar que son también condiciones suficientes para que exista una v.a X con función de densidad igual a f .

III.4.2.4. Ejemplo

El tiempo de vida expresado en miles de horas de un dispositivo electrónico escogido al azar en la producción de una fábrica es una v.a X . Después de un estudio, se opta por modelizar esta v.a como una v.a continua con una función de densidad dada por

$$f_X(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

La representación gráfica de f_X es



Notar que por la gráfica de esta función de densidad, comprobamos que la probabilidad de que X pertenezca a un intervalo de números negativos, por ejemplo $[-2, -3]$ es nula (la densidad de probabilidad es nula en \mathbb{R}^-), o que es mucho menos probable que un dispositivo dure entre 4000 y 5000 horas que dure entre 1000 y 2000h.

Si nos preguntamos precisamente cuál es la proporción de dispositivos en la producción que duran entre 1000 y 2000h, debemos calcular

$$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 f_X(x)dx = \int_1^2 e^{-x} dx = [-e^{-x}]_1^2 \simeq 0,235.$$

Según nuestro modelo, alrededor del 23% de la producción tendrá una duración entre 1000 y 2000 horas.

III.4.3. Medidas numéricas asociadas a una v.a continua

De la misma manera que para distribuciones de variables en un conjunto de datos, se pueden resumir algunas características de las distribuciones de variables asociadas a experimentos aleatorios.

III.4.3.1. Esperanza

Sea X una variable con densidad f , definimos la media de X , también llamada esperanza o valor esperado, como

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f(x) dx.$$

Es una medida de centro de la distribución si ésta es relativamente simétrica, se interpreta como el centro de gravedad de la distribución, ver figura III.1. Otra vez es coherente con la analogía entre el peso y la probabilidad.

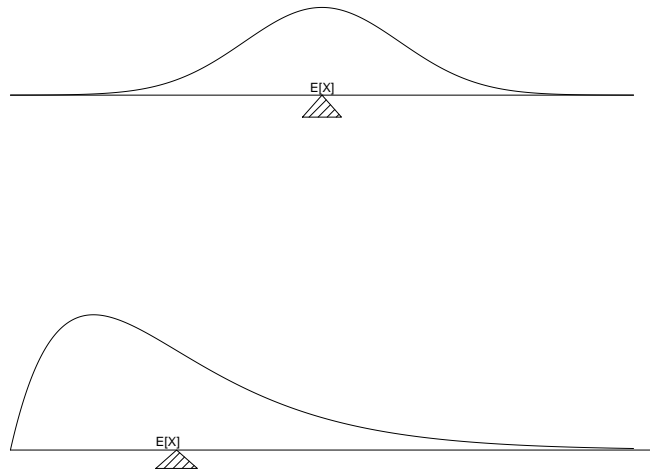


Figura III.1: La esperanza es el centro de gravedad

Tal como lo hicimos para una v.a discreta, es conveniente definir para una función g de X la esperanza de $g(X)$:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

III.4.3.2. Varianza - Desviación típica

La varianza se define como el promedio de la distancia al cuadrado entre X y su media:

$$\sigma_X^2 = \text{var}(X) = \mathbb{E}[(X - \mu_X)^2] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx.$$

Al desarrollar la integral, es fácil obtener la fórmula alternativa, más práctica para el cálculo:

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x) dx - (\mathbb{E}[X])^2.$$

y la desviación típica es $\sigma_X = \sqrt{\sigma_X^2}$.

La desviación típica mide la dispersión de la distribución de los valores de X respecto a su media.

III.4.3.3. Un ejemplo

Calculemos la duración media y la desviación típica en el ejemplo de la duración de los dispositivos electrónicos de la sección III.4.2.4. Tenemos que

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_{-\infty}^0 x \cdot f_X(x) dx + \int_0^{+\infty} x \cdot f_X(x) dx \\ &= 0 + \int_0^{+\infty} x \cdot e^{-x} dx \\ &= 1,\end{aligned}$$

hemos descompuesto la integral inicial según los intervalos de definición de f_X , sustituido la expresión de f_X en las integrales resultantes, y calculado por partes la última integral que aparece. La duración media de los dispositivos es por lo tanto de 1000h.

De la misma manera, calculamos la varianza de X :

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 0 + \int_0^{+\infty} x^2 \cdot e^{-x} dx - 1 = 1.$$

III.4.4. Modelos más comunes de v.a continua

Algunas situaciones de modelización presentan rasgos comunes y se han establecido modelos “estándar” que resultan adecuados para distintos contextos.

III.4.4.1. Variable aleatoria uniforme

El modelo de v.a. continua más sencillo corresponde a la situación en la que X puede tomar cualquier valor entre dos números a y b , sin que favorezca ninguna zona del intervalo $[a, b]$. La probabilidad de que X esté entre a y b será igual a 1, mientras que la probabilidad de que esté en un subintervalo de $[a, b]$ será sencillamente proporcional a su longitud. Intuitivamente, queremos que la función de densidad de X sea nula fuera de $[a, b]$, y constante en el intervalo $[a, b]$. Para que el área total debajo de la curva de densidad sea igual a 1, esta constante deberá ser igual a $1/(b-a)$. La función de densidad será por lo tanto dada por:

$$f_X(x) = \begin{cases} \frac{1}{(b-a)} & \text{si } a \leq x \leq b, \\ 0 & \text{en otro caso.} \end{cases}$$

La representación gráfica de f_X se encuentra en la figura III.2. Una v.a X que tenga esta función de densidad se llama una v.a uniforme entre a y b . Lo denotaremos por

$$X \sim \mathcal{U}([a, b]).$$

El comando “RANDOM” de varios lenguajes de programación, que también aparece en casi todas las calculadoras científicas, simula una variable uniforme entre 0 y 1. ¿Puede ser realmente una v.a uniforme?

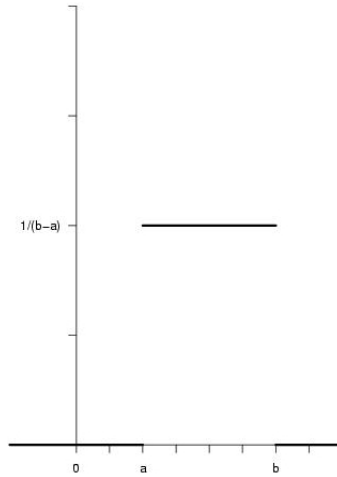


Figura III.2: Densidad de una v.a uniforme

Por otra parte calculemos la esperanza y la varianza de una v.a $X \sim \mathcal{U}([a, b])$. Antes de llevar a cabo los cálculos, y examinando la gráfica de la densidad de X , ¿cuánto piensa que vale $\mathbb{E}[X]$?

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = 0 + \int_a^b x \cdot \frac{1}{b-a} dx + 0 \\ &= \frac{b^2 - a^2}{2} \cdot \frac{1}{b-a} = \frac{a+b}{2} \end{aligned}$$

¿Corresponde con su intuición?. Se deja en ejercicio al lector comprobar que la varianza de una v.a $X \sim \mathcal{U}([a, b])$ es

$$\text{var}(X) = \frac{(b-a)^2}{12},$$

es decir que la desviación típica es sencillamente proporcional a $(b-a)$, otro resultado natural, ¿no?

III.4.4.2. Modelo exponencial

a). Definición En el mismo contexto que para una v.a de Poisson (ocurrencias de sucesos aleatorios en el tiempo), denotando por λ el número medio de ocurrencias por intervalo de tiempo, consideramos la v.a X que mide el tiempo entre dos ocurrencias consecutivas del suceso, la distribución de la v.a X se llama distribución exponencial de parámetro λ y se denota por

$$X \sim \text{Exp}(\lambda).$$

Dos ejemplos corresponden al tiempo entre dos emisiones consecutivas de una partícula por un material radioactivo, o entre dos llegadas de clientes en una cola.

b). Propiedades

- La función de densidad de una v.a $X \sim \text{Exp}(\lambda)$ es

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Su gráfica es parecida a la del ejemplo de la sección III.4.2.4. De hecho, resulta que la densidad de este ejemplo es la densidad de una distribución exponencial de parámetro λ .

- *Función de distribución acumulada.* Para todo t ,

$$F_X(t) = \int_{-\infty}^t f_X(x) dx.$$

Deducimos que, si $t < 0$, $F_X(t)$ es nula, mientras que, si $t \geq 0$,

$$F_X(t) = 0 + \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}.$$

En particular, tenemos que $\mathbb{P}(X > t) = e^{-\lambda t}$.

- *Esperanza y varianza.* Demostramos de la misma manera que para el ejemplo de la sección III.4.2.4, utilizando la integración por partes que

$$\mathbb{E}[X] = 1/\lambda, \quad \text{var}(X) = 1/\lambda^2.$$

- *Propiedad de falta de memoria de la distribución exponencial.* La distribución exponencial tiene una propiedad particular: “olvida su pasado”... Más concretamente, supongamos que $X \sim \text{Exp}(\lambda)$ y modeliza el tiempo entre dos llegadas sucesivas de clientes en una cola. Llega un cliente, y espero hasta que llegue el siguiente cliente... Han pasado tres minutos y no ha llegado, la probabilidad de que tenga que esperar por lo menos otro minuto más (es decir que el tiempo transcurrido entre las dos llegadas sea mayor que cuatro minutos) es la misma que la probabilidad de que X sea mayor que 1 minuto: ¡el hecho de saber que ya he esperado 3 minutos no cambia la probabilidad de que todavía tenga que esperar otro minuto más! Es decir, para todos $t_1 > 0$, $t_2 > 0$,

$$\mathbb{P}(X > t_1 + t_2 | X > t_1) = \mathbb{P}(X > t_2).$$

Demostración: Por la definición de la probabilidad condicionada,

$$\mathbb{P}(X > t_1 + t_2 | X > t_1) = \frac{\mathbb{P}((X > t_1 + t_2) \cap (X > t_1))}{\mathbb{P}(X > t_1)}.$$

Por otra parte, puesto que el suceso $(X > t_1 + t_2)$ está incluido en el suceso $(X > t_1)$, el denominador es sencillamente $\mathbb{P}(X > t_1 + t_2)$. Pero al calcular un poco más arriba la función de distribución acumulada de una distribución exponencial, hemos notado que $\mathbb{P}(X > t) = e^{-\lambda t}$. Por lo tanto

$$\mathbb{P}(X > t_1 + t_2 | X > t_1) = \frac{e^{-\lambda(t_1+t_2)}}{e^{-\lambda t_1}} = e^{-\lambda t_2} = \mathbb{P}(X > t_2).$$

□

III.4.4.3. La distribución Normal

a). **Definición** Sea μ un número real y σ^2 un número real positivo, la v.a X sigue una distribución Normal de parámetros μ y σ^2 si su densidad es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

cuya representación gráfica es la famosa “campana de Gauss”, ver Figura III.3.

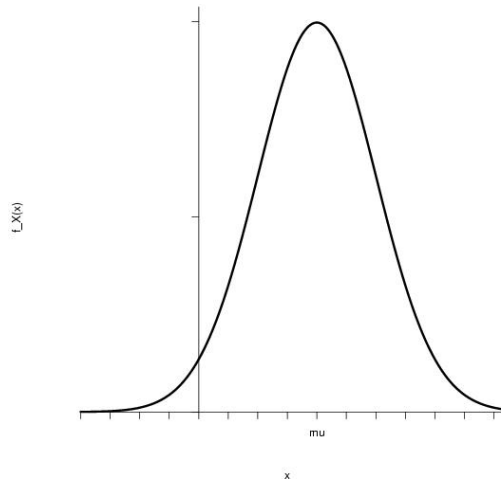


Figura III.3: Densidad Normal

Si X sigue una distribución Normal de parámetros μ y σ^2 , escribiremos $X \sim \mathcal{N}(\mu, \sigma^2)$.

La distribución Normal es, sin dudas, la distribución más utilizada en situaciones prácticas: aparece en la inmensa mayoría de los procedimientos estadísticos que se llevan a cabo de manera rutinaria (control de calidad, mediciones, etc...) En particular, está típicamente presente cuando se modeliza los valores proporcionados por un aparato de medición. De hecho, si consideramos los datos de las mediciones de la luz por S. Newcomb que estudiamos en el Tema 1, ver sección I.3.2.1, podemos comprobar que las frecuencias de aparición de los datos experimentales se ajustan bastante bien a un modelo Normal. En la figura III.4, se ha ajustado una curva Normal al histograma de los datos recogidos por Newcomb, después de omitir los dos datos atípicos -44 y -2 . Para ello, hemos fijado el valor de μ y σ^2 basándonos en el centro y la dispersión de la distribución de los datos experimentales.

b). Propiedades

- La curva de la densidad Normal es simétrica respecto al eje vertical $x = \mu$. En particular deducimos que $\mathbb{P}(X \geq \mu) = \mathbb{P}(X \leq \mu) = 1/2$.
- La curva de la densidad Normal nunca se cruza con el eje Ox .

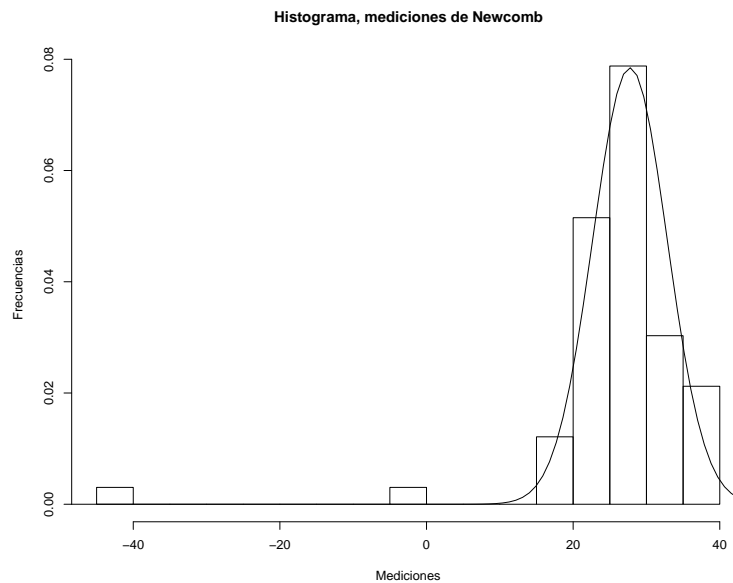


Figura III.4: Ajuste de una densidad Normal al histograma de Newcomb

- *Esperanza y varianza:* Es posible comprobar que, si $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[X] = \mu, \quad \text{var}(X) = \sigma^2.$$

- *Función de distribución acumulada.* La función f_X no admite primitiva en una forma cerrada, y por lo tanto no hay expresión simple para calcular la probabilidad de que una variable Normal pertenezca a un intervalo dado, o en general para su función de distribución. Se debe por lo tanto recurrir por lo tanto a aproximaciones numéricas de la integral

$$\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

para obtener $\mathbb{P}(a < X \leq b)$. Los programas informáticos de análisis de datos como R disponen de algoritmos que permitan calcular para cualquier t la probabilidad $\mathbb{P}(X \leq t)$. También existen calculadoras estadísticas.

A pesar de que no exista una expresión simple para las probabilidades asociadas a una distribución Normal, es muy útil conocer la regla siguiente: si X es una Normal $\mathcal{N}(\mu, \sigma^2)$, tenemos

$$\begin{aligned} \mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) &\simeq 0,68 \\ \mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\simeq 0,95 \\ \mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\simeq 0,997, \end{aligned}$$

lo que queda reflejado en la figura III.5: el 68 % del área debajo de la curva Normal está comprendida entre $\mu - \sigma$ y $\mu + \sigma$, el 95 % entre $\mu - 2\sigma$ y $\mu + 2\sigma$, y el 99.7 % entre $\mu - 3\sigma$ y $\mu + 3\sigma$.

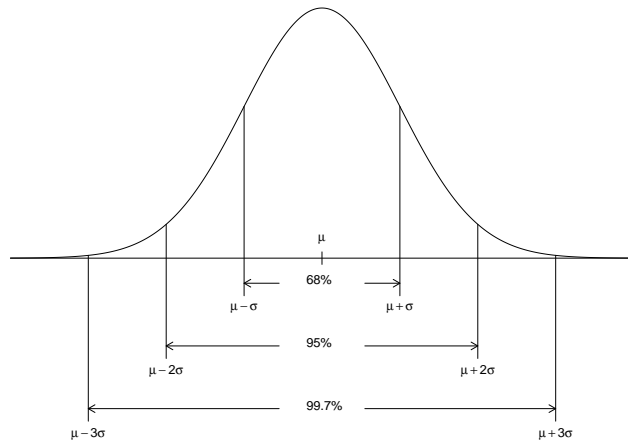


Figura III.5: Regla del 68 % - 95 % - 99.7 %

c). ¿Cómo calcular probabilidades asociadas a una distribución Normal

(i) Para una distribución $Z \sim \mathcal{N}(0, 1)$.

La distribución Normal con parámetros $\mu = 0$ y $\sigma^2 = 1$ se llama distribución Normal estándar. Su función de distribución acumulada se denota por ϕ y los valores de ϕ están tabulados. La tabla para valores de ϕ está incluida en el apéndice de este tema.

Notar que en la tabla sólo aparece valores de $\phi(t)$ para valores positivos de t . Para deducir $\phi(t)$ para valores negativos de t , utilizamos la simetría de la distribución normal que implica que, para todo t ,

$$\phi(-t) = 1 - \phi(t).$$

Comprobar con la tabla que sabéis calcular las probabilidades siguientes:

$$\begin{aligned} \mathbb{P}(Z \leq 2,68) \simeq 0,9963 & \quad \mathbb{P}(Z \leq 1,12) \simeq 0,8686 & \quad \mathbb{P}(Z \leq -0,9) \simeq 0,1841 \\ \mathbb{P}(1,1 \leq Z \leq 1,3) \simeq 0,04 & \quad \mathbb{P}(-0,9 \leq Z \leq -0,5) \simeq 0,13 & \quad \mathbb{P}(-1 \leq Z \leq 1) \simeq 0,68 \end{aligned}$$

(ii) Para una distribución $X \sim \mathcal{N}(\mu, \sigma^2)$.

El cálculo de probabilidades para una distribución Normal con parámetros μ y σ^2 se basa en la siguiente propiedad que no demostraremos:

Propiedad: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, la variable

$$Z = \frac{X - \mu}{\sigma}$$

sigue una distribución Normal con media 0 y varianza 1.

Pasar de $X \sim \mathcal{N}(\mu, \sigma^2)$ a $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ se llama tipificar la variable X , y la variable Z se llama la v.a X tipificada.

Para calcular una probabilidad relacionada con X , reescribiremos el suceso de interés, tipificando la v.a.

Supongamos por ejemplo que $X \sim \mathcal{N}(\mu = 1, \sigma^2 = 0,25)$. Tenemos

$$\mathbb{P}(X \leq 1,25) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{1,25 - \mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{1,25 - 1}{0,5}\right) = \mathbb{P}(Z \leq 0,5) \simeq 0,69.$$

y

$$\begin{aligned} \mathbb{P}(0,5 \leq X \leq 1,5) &= \mathbb{P}\left(\frac{0,5 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{1,5 - \mu}{\sigma}\right) = \mathbb{P}\left(\frac{0,5 - 1}{0,5} \leq Z \leq \frac{1,5 - 1}{0,5}\right) \\ &= \mathbb{P}(-1 \leq Z \leq 1) \simeq 0,68. \end{aligned}$$

III.4.4.4. Aproximación de una distribución Binomial por una distribución Normal

En el caso en que sólo disponemos de una calculadora sencilla, el cálculo de probabilidades asociadas a una distribución Binomial X puede resultar laborioso si éstas requieren evaluar la función puntual de X en muchos valores. Por ejemplo, supongamos que $X \sim \mathcal{B}(100, 0,1)$, el cálculo de $\mathbb{P}(X \geq 15)$ implica que calculemos 86 probabilidades individuales ($\mathbb{P}(X = 16)$, $\mathbb{P}(X = 17)$, \dots , $\mathbb{P}(X = 100)$) o pasando al suceso complementario 15 probabilidades, que siguen siendo muchos cálculos...

Para algunas combinaciones de valores de n y p , resulta que la distribución Binomial se puede aproximar de manera satisfactoria por una distribución normal, es decir que para calcular la probabilidad de un suceso relacionado con una v.a Binomial $X \sim \mathcal{B}(n, p)$, podremos hacer como si X tuviera una distribución normal.

Propiedad. Consideramos una v.a $X \sim \mathcal{B}(n, p)$. Si $n \cdot p \geq 5$ y $n(1 - p) \geq 5$, se puede aproximar de manera satisfactoria la distribución de X por la distribución de $W \sim \mathcal{N}(\mu, \sigma)$, con $\mu = n \cdot p$ y $\sigma = \sqrt{n \cdot p(1 - p)}$, con la fórmula

$$\text{para todo } x, \quad \mathbb{P}(X \leq x) \simeq \mathbb{P}(W \leq x + 0,5).$$

El término “+0.5” que aparece en el término de la derecha de la fórmula corresponde a la llamada “corrección por continuidad”: aproximamos la distribución de una v.a discreta, X , que sólo puede tomar valores enteros por una v.a continua W que puede tomar cualquier valor real. Para conseguir una equivalencia, podemos considerar que un valor entero x para la v.a. Binomial X corresponde al intervalo $[x - 0,5, x + 0,5]$ para la v.a Normal W , tal como está ilustrado en la Figura III.6, para unos pocos valores de X .

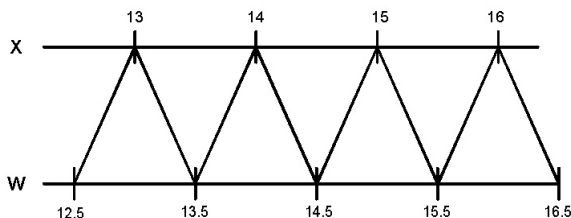


Figura III.6: Aproximación de una distribución Binomial por una distribución Normal

En particular deducimos de esta figura que aproximaremos las probabilidades relacionadas con X de la manera siguiente:

$$\begin{aligned}\mathbb{P}(X = 15) &\simeq \mathbb{P}(14,5 < W \leq 15,5) \\ \mathbb{P}(X > 15) &\simeq \mathbb{P}(W \geq 15,5) \\ \mathbb{P}(X \geq 15) &\simeq \mathbb{P}(W \geq 14,5) \\ \mathbb{P}(X \leq 16) &\simeq \mathbb{P}(W \leq 16,5) \\ \mathbb{P}(X < 16) &\simeq \mathbb{P}(W \leq 15,5) \\ \mathbb{P}(13 \leq X < 15) &\simeq \mathbb{P}(12,5 \leq W \leq 14,5)\end{aligned}$$

III.5. Algunas propiedades útiles de la esperanza y la varianza

Acabamos el capítulo con una sección “cajón de sastre” en la que mencionamos algunos resultados sobre esperanza y varianza.

Sean a y b dos números reales, y X una variable aleatoria. No es difícil demostrar, utilizando las definiciones de esperanza y varianza tanto para v.a discreta como para v.a continua que se cumplen las siguientes propiedades:

$$\begin{aligned}\mathbb{E}[aX + b] &= a\mathbb{E}[X] + b \\ \text{var}(aX + b) &= a^2 \text{var}(X) \\ \sigma_{aX+b} &= |a|\sigma_X\end{aligned}$$

Intuitivamente son resultados naturales: si multiplico todos los valores de una v.a por a y traslado el resultado de b unidades, el centro de gravedad de los datos (la esperanza) se multiplica por a y se traslada de b unidades, mientras que la dispersión (la desviación típica) sólo se multiplica por $|a|$, puesto que la traslación de los datos no cambia su dispersión.

Finalizamos con un último resultado asociado a la varianza de una variable: la desigualdad de Chebichev:

Propiedad: Sea cual sea la distribución de X , si conocemos el valor de la varianza de X , tenemos la siguiente cota para la probabilidad de que X esté en un intervalo centrado en su media μ_X :

$$\text{Para cualquier } a > 0, \mathbb{P}(|X - \mu_X| \leq a) \geq 1 - \frac{\text{Var}(X)}{a^2}.$$

Deducimos también una cota para el suceso complementario:

$$\text{Para cualquier } a > 0, \mathbb{P}(|X - \mu_X| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

La primera desigualdad se interpreta de la manera siguiente: sabemos que una proporción de los datos de al menos $\text{Var}(X)/a^2$ se encuentra en el intervalo $\mu_X \pm a$, mientras que la segunda desigualdad se lee: una proporción de los datos de como mucho $\text{Var}(X)/a^2$ se encuentra fuera del intervalo $\mu_X \pm a$.

Distribución Normal:

$$\mathbb{P}(Z \leq t) = \phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

t	$\mathbb{P}(Z \leq t)$	t	$\mathbb{P}(Z \leq t)$	t	$\mathbb{P}(Z \leq t)$	t	$\mathbb{P}(Z \leq t)$
0.00	0.5000	0.80	0.7881	1.60	0.9452	2.40	0.9918
0.02	0.5080	0.82	0.7939	1.62	0.9474	2.42	0.9922
0.04	0.5160	0.84	0.7995	1.64	0.9495	2.44	0.9927
0.06	0.5239	0.86	0.8051	1.66	0.9515	2.46	0.9931
0.08	0.5319	0.88	0.8106	1.68	0.9535	2.48	0.9934
0.10	0.5398	0.90	0.8159	1.70	0.9554	2.50	0.9938
0.12	0.5478	0.92	0.8212	1.72	0.9573	2.52	0.9941
0.14	0.5557	0.94	0.8264	1.74	0.9591	2.54	0.9945
0.16	0.5636	0.96	0.8315	1.76	0.9608	2.56	0.9948
0.18	0.5714	0.98	0.8365	1.78	0.9625	2.58	0.9951
0.20	0.5793	1.00	0.8413	1.80	0.9641	2.60	0.9953
0.22	0.5871	1.02	0.8461	1.82	0.9656	2.62	0.9956
0.24	0.5948	1.04	0.8508	1.84	0.9671	2.64	0.9959
0.26	0.6026	1.06	0.8554	1.86	0.9686	2.66	0.9961
0.28	0.6103	1.08	0.8599	1.88	0.9699	2.68	0.9963
0.30	0.6179	1.10	0.8643	1.90	0.9713	2.70	0.9965
0.32	0.6255	1.12	0.8686	1.92	0.9726	2.72	0.9967
0.34	0.6331	1.14	0.8729	1.94	0.9738	2.74	0.9969
0.36	0.6406	1.16	0.8770	1.96	0.9750	2.76	0.9971
0.38	0.6480	1.18	0.8810	1.98	0.9761	2.78	0.9973
0.40	0.6554	1.20	0.8849	2.00	0.9772	2.80	0.9974
0.42	0.6628	1.22	0.8888	2.02	0.9783	2.82	0.9976
0.44	0.6700	1.24	0.8925	2.04	0.9793	2.84	0.9977
0.46	0.6772	1.26	0.8962	2.06	0.9803	2.86	0.9979
0.48	0.6844	1.28	0.8997	2.08	0.9812	2.88	0.9980
0.50	0.6915	1.30	0.9032	2.10	0.9821	2.90	0.9981
0.52	0.6985	1.32	0.9066	2.12	0.9830	2.92	0.9982
0.54	0.7054	1.34	0.9099	2.14	0.9838	2.94	0.9984
0.56	0.7123	1.36	0.9131	2.16	0.9846	2.96	0.9985
0.58	0.7190	1.38	0.9162	2.18	0.9854	2.98	0.9986
0.60	0.7257	1.40	0.9192	2.20	0.9861	3.00	0.9987
0.62	0.7324	1.42	0.9222	2.22	0.9868	3.10	0.9990
0.64	0.7389	1.44	0.9251	2.24	0.9875	3.20	0.9993
0.66	0.7454	1.46	0.9279	2.26	0.9881	3.30	0.9995
0.68	0.7517	1.48	0.9306	2.28	0.9887	3.40	0.9997
0.70	0.7580	1.50	0.9332	2.30	0.9893	3.50	0.9998
0.72	0.7642	1.52	0.9357	2.32	0.9898	3.60	0.9998
0.74	0.7704	1.54	0.9382	2.34	0.9904	3.80	0.9999
0.76	0.7764	1.56	0.9406	2.36	0.9909	4.00	1.0000
0.78	0.7823	1.58	0.9429	2.38	0.9913	4.50	1.0000



2013: Año Internacional de la Estadística. ¿Sabías qué...?

La estilometría es el análisis estadístico del estilo de obras literarias, y busca por ejemplo, determinar la autoría de un texto, basándose en características cuantificables propias del autor y no del género o época. Una de estas características es la longitud de palabra y fue usada para discriminar entre obras de Shakespeare y Bacon por ejemplo.

Fuente: Girón, F.J, Ginebra, J & Riba, A. "Literatura y estadística: el problema de la autoría de Tirant lo Blanc", BEIO (2005) 22, 6-10.

statistics

TEMA IV

Variable Aleatoria II

IV.1. Introducción

Es frecuente que haya más de una variable aleatoria de interés asociada a un experimento aleatorio. Supongamos por ejemplo que consideramos n variables X_1, X_2, \dots, X_n , formaremos el vector aleatorio $\mathbb{X} = (X_1, X_2, \dots, X_n)$. Diremos que \mathbb{X} es una variable aleatoria multidimensional. Para el caso particular en que $n = 2$, hablaremos de variable aleatoria bidimensional.

Describir la distribución de una v.a. multidimensional consiste en asignar una probabilidad a sucesos conjuntos, es decir sucesos que involucren X_1, X_2, \dots, X_n . En este caso hablamos de **distribución conjunta** de (X, Y) , mientras que si consideramos las distribuciones de X e Y por separadas, hablamos de **distribuciones marginales** de X y de Y respectivamente.

Un ejemplo de suceso asociado a la distribución conjunta de X e Y es $(X+Y > 3)$ o $(X = 1 \cap Y > 2)$ mientras que el suceso $(X > 5)$ y el suceso $(Y = 4)$ hacen referencia a las distribuciones marginales de X y de Y respectivamente.

En este tema nos centraremos sobre todo en el caso de una variable bidimensional.

IV.2. Variable bidimensional discreta

Si tanto X como Y son variables discretas, basta con describir la probabilidad de los sucesos $(X = x) \cap (Y = y)$. Lo realizaremos a través de la función puntual de probabilidad conjunta de X e Y :

IV.2.1. Función puntual de probabilidad conjunta

IV.2.1.1. Definición

La función puntual de probabilidad conjunta de (X, Y) asocia a cualquier par de valores (x, y) la probabilidad del suceso $((X = x) \cap (Y = y))$. La denotamos

$$f_{XY}(x, y) = \mathbb{P}((X = x) \cap (Y = y)).$$

Los valores que toma una función puntual de probabilidad conjunta se pueden presentar en una tabla:

X	Y			
	120	130	140	150
0	0.03	0.1	0.15	0.2
1	0.05	0.06	0.1	0.1
2	0.21	0	0	0

Deducimos en particular de esta tabla que la probabilidad que X tome el valor 0 y a la vez Y tome el valor 140 es igual a 0.15.

IV.2.1.2. Propiedad

Para que una función $f : (x, y) \mapsto f(x, y)$ sea la función puntual de probabilidad conjunta de una variable bidimensional discreta (X, Y) es necesario y suficiente que cumpla

1. $f_{XY}(x_i, y_j) \geq 0, \forall x_i, y_j$.
2. $\sum_{x_i} \sum_{y_j} f_{XY}(x_i, y_j) = 1$.

IV.2.1.3. Relación entre funciones puntuales de probabilidad conjunta y marginales

Si conocemos la distribución conjunta de (X, Y) a través de una tabla como la descrita en el apartado IV.2.1.1, podemos calcular la distribución de X o de Y por separado: éstas se llaman las distribuciones marginales. En efecto, para calcular $\mathbb{P}(X = 0)$ por ejemplo, basta con utilizar

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(X = 0 \cap Y = 120) + \mathbb{P}(X = 0 \cap Y = 130) \\ &\quad + \mathbb{P}(X = 0 \cap Y = 140) + \mathbb{P}(X = 0 \cap Y = 150) = 0,48. \end{aligned}$$

Tenemos por lo tanto las relaciones siguientes:

$$\begin{aligned} \forall x_i, \quad f_X(x_i) &= \sum_{y_j} f_{XY}(x_i, y_j), \\ \forall y_j, \quad f_Y(y_j) &= \sum_{x_i} f_{XY}(x_i, y_j). \end{aligned}$$

Se suele representar en la misma tabla de la f.p.p. conjunta de la manera siguiente:

X	Y				f_X
	120	130	140	150	
0	0.03	0.1	0.15	0.2	0.48
1	0.05	0.06	0.1	0.1	0.31
2	0.21	0	0	0	0.21
f_Y	0.29	0.16	0.25	0.3	

IV.2.2. Esperanza

Sea $g : (x, y) \mapsto g(x, y)$ una función de dos variables que toma sus valores en \mathbb{R} . Definimos la esperanza (o media, o valor esperado, o valor promedio) de $g(X, Y)$ como

$$\begin{aligned}\mathbb{E}[g(X, Y)] &= \sum_{x_i} \sum_{y_j} g(x_i, y_j) \mathbb{P}(X = x_i \cap Y = y_j) \\ &= \sum_{x_i} \sum_{y_j} g(x_i, y_j) f_{XY}(x_i, y_j).\end{aligned}$$

IV.3. Variable bidimensional continua

Consideramos ahora el par (X, Y) donde X e Y son ambas v.a continuas. Para describir la distribución conjunta de (X, Y) , introducimos la función de densidad conjunta.

IV.3.1. Función de densidad conjunta

IV.3.1.1. Definición.

La función de densidad conjunta de (X, Y) es una función f_{XY} que permite calcular la probabilidad de cualquier suceso de la forma $(a \leq X \leq b) \cap (c \leq Y \leq d)$ a través de la fórmula:

$$\mathbb{P}((a \leq X \leq b) \cap (c \leq Y \leq d)) = \int_{x \in [a, b]} \int_{y \in [c, d]} f_{XY}(x, y) dx dy.$$

IV.3.1.2. Ejemplo

Consideremos un experimento que consista en producir dos componentes de dos tipos, y denotamos por X e Y el tiempo de vida en miles de horas del primer y segundo componente respectivamente. Modelizamos su distribución conjunta a través de la función de densidad siguiente

$$f_{XY}(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{si } x > 0 \text{ y } y > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Para calcular la probabilidad de que ambos componentes duren menos de 1000 horas, por ejemplo,

$$\begin{aligned}\mathbb{P}((X < 1) \cap (Y \leq 1)) &= \int_{-\infty}^1 \int_{-\infty}^1 f_{XY}(x, y) dx dy \\ &= \int_0^1 \int_0^1 2e^{-x} e^{-2y} dx dy = (1 - e^{-1})(1 - e^{-2}) \simeq 0,54.\end{aligned}$$

IV.3.1.3. Propiedades

Para que una función $f : (x, y) \mapsto f(x, y)$ con valores en \mathbb{R} sea la función de densidad conjunta de una v.a bidimensional continua, es necesario y suficiente que cumpla

1. $f(x, y) \geq 0, \quad \forall x, y,$

- 2.

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

IV.3.1.4. Relación entre funciones de densidad conjunta y marginales

Al igual que para una v.a discreta, se puede obtener de la función de densidad conjunta las funciones marginales, pero ahora en lugar de sumar, debemos integrar respecto de la otra variable.

Tenemos por lo tanto las relaciones siguientes:

$$\begin{aligned}\forall x, \quad f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dy, \\ \forall y, \quad f_Y(y) &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dx.\end{aligned}$$

Calculemos para ilustrar estas fórmulas las densidades marginales de X y de Y para el ejemplo del apartado IV.3.1.2. La función de densidad conjunta es

$$f_{XY}(x, y) = \begin{cases} 2e^{-x} e^{-2y} & \text{si } x > 0 \text{ y } y > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Deducimos la densidad marginal de X :

$$\forall x, \quad f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy.$$

Si $x \leq 0$, $f_{XY}(x, y) = 0$ para todo y , y por lo tanto $f_X(x) = 0$ también.

Si $x > 0$,

$$\begin{aligned}f_X(x) &= \int_0^{+\infty} 2e^{-x} e^{-2y} dy = e^{-x} [-e^{-2y}]_0^{+\infty} \\ &= e^{-x}.\end{aligned}$$

IV.3.2. Esperanza

Al disponer de una función de densidad conjunta f_{XY} para la v.a. bidimensional (X, Y) , podemos calcular el valor esperado de una función de las dos variables X e Y : **Definición.** Sea una función $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, la esperanza de $g(X, Y)$ se define como

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dx dy.$$

En particular podemos calcular por ejemplo la esperanza de la suma de dos variables:

$$\begin{aligned} \mathbb{E}[X + Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{XY}(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f_{XY}(x, y) dy \right) dx + \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f_{XY}(x, y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy = \mathbb{E}[X] + \mathbb{E}[Y], \end{aligned}$$

donde hemos utilizado para el último paso la relación entre funciones de densidades marginales y conjunta del apartado IV.3.1.4. Hemos por lo tanto demostrado una relación por otra parte muy intuitiva: la media de la suma de dos variables aleatorias es la suma de las dos medias...

IV.4. Distribuciones condicionadas

Consideremos un experimento al que va asociada una v.a. bidimensional (X, Y) . Por algún motivo, al realizar el experimento, sólo observamos el valor de Y y no el de X . ¿Qué información puedo deducir, basándome en el valor de Y , sobre la distribución de los posibles valores de X ?

Un contexto típico en ingeniería en la que se da esta situación es el siguiente: me interesa un señal X_1, X_2, \dots, X_n , pero no puedo observar directamente los valores de X sino a través de un aparato de medición que induce una perturbación aleatoria, que denotaremos por ε . Como resultado observo

$$\begin{aligned} Y_1 &= X_1 + \varepsilon_1, \\ &\vdots \\ Y_n &= X_n + \varepsilon_n. \end{aligned}$$

Disponiendo de los valores de Y_1, \dots, Y_n , persigo deducir la distribución de X_1, \dots, X_n condicionada a Y_1, \dots, Y_n . Obtener esta distribución condicionada se llama realizar el filtrado de la señal Y_1, \dots, Y_n . De los filtros basados en modelos probabilísticos, el más usado en práctica se llama el filtro de Kalman.

IV.4.1. V.a. bidimensional discreta

Sea (X, Y) una v.a. bidimensional discreta.

IV.4.1.1. Definición de la función puntual de probabilidad condicionada

Sea y un valor de Y tal que $\mathbb{P}(Y = y) > 0$, la función puntual de probabilidad de X condicionada a $Y = y$ asocia a cada valor posible x de X la probabilidad del suceso $X = x$ condicionada a $(X = x)$.

$$f_{X|Y=y}(x) = \mathbb{P}(X = x|Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Para ilustrar este concepto, calculemos para el ejemplo de v.a bidimensional introducido anteriormente la función puntual de probabilidad de X condicionada a $Y = 130$. Recordemos que la tabla de las f.p.p conjunta y marginales de (X, Y) era

X	Y				f_X
	120	130	140	150	
0	0.03	0.1	0.15	0.2	0.48
1	0.05	0.06	0.1	0.1	0.31
2	0.21	0	0	0	0.21
f_Y	0.29	0.16	0.25	0.3	

Por lo tanto $f_{X|Y=130}$ toma los valores:

Valores posibles de X	0	1	2
$f_{X Y=130}$	$0,1/0,16 = 0,625$	$0,06/0,16 = 0,375$	$0/0,16 = 0$

IV.4.2. Para una v.a bidimensional continua

Consideramos ahora una v.a. bidimensional continua (X, Y) .

IV.4.2.1. Definición

Sea (X, Y) una v.a continua con densidad conjunta f_{XY} . Consideramos un valor y para el cual $f_Y(y) > 0$. La función de densidad de X condicionada a $Y = y$ está definida por

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Nota: la densidad de Y condicionada a X se obtiene intercambiando los papeles de X e Y en la fórmula anterior.

IV.4.2.2. Ejemplo

Consideremos el ejemplo de la subsección IV.3.1.2. Calculemos, para un valor $y > 0$ genérico, la función de densidad de X condicionada a $Y = y$. Obtuvimos que la densidad marginal de Y , si $y > 0$ es $f_Y(y)2e^{-2y}$. Deducimos que la densidad que buscamos es

$$f_{X|Y=y}(x) = \begin{cases} \frac{2e^{-x}e^{-2y}}{2e^{-2y}} = e^{-x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Observamos que, en este caso, coincide con la densidad marginal de X .

IV.4.3. Esperanza condicionada

Es fácil comprobar que, para un valor y tal que $f_Y(y) > 0$, $x \mapsto f_{X|Y=y}(x)$ cumple con los dos requisitos (ver secciones III.3.2.2 y b)) que permiten deducir que se trata de una función de densidad (caso continuo) o puntual de probabilidad (caso discreto). Por ello, hablamos de distribución de X condicionada a $Y = y$, aunque sólo podemos interpretar las probabilidades asociadas como probabilidades condicionadas en el caso de una v.a discreta.

También podemos por lo tanto definir la esperanza condicionada de una función $g(X)$ dado $Y = y$.

Definición IV.4.1 *Sea una función $g : \mathbb{R} \rightarrow \mathbb{R}$, la esperanza condicionada de $g(X)$ dado $Y = y$ se define como*

- Si (X, Y) es una v.a. discreta

$$\mathbb{E}[g(X)|Y = y] = \sum_x g(x)f_{X|Y=y}(x).$$

- Si (X, Y) es una v.a continua

$$\mathbb{E}[g(X)|Y = y] = \int_{-\infty}^{+\infty} g(x)f_{X|Y=y}(x)dx.$$

La noción de esperanza condicionada permite en particular obtener resúmenes de las características principales de la distribución condicionada de X dado $Y = y$. Si consideramos el problema de predecir el valor de X dado que hemos observado el valor y para Y , se puede demostrar que la esperanza condicionada de X dado $Y = y$ es el mejor predictor posible en el sentido siguiente:

Llamamos predictor a cualquier función de Y , $h(Y)$ diseñada para aproximar el valor de X que no hemos observado. Denotamos, para todo y , por $h^*(y)$ la esperanza condicionada $\mathbb{E}[X|Y = y]$. Consideramos la función de Y , $h^*(Y)$, se trata de un predictor de X . Se puede probar que para cualquier predictor $h(Y)$ de X se cumple

$$\mathbb{E}[(X - h(Y))^2] \geq \mathbb{E}[(X - h^*(Y))^2],$$

es decir que el error cuadrático medio que se comete al predecir X por $h^*(Y)$ es el menor de los errores posibles.

IV.5. Variables independientes

En el tema 2 hemos definido el concepto de sucesos independientes. Introducimos ahora el concepto de variables aleatorias independientes:

IV.5.1. Definición

Definición IV.5.1 *Dos variables X e Y son independientes si se cumple*

$$\text{para todo } x \text{ e } y, \quad f_{XY}(x, y) = f_X(x)f_Y(y).$$

Las funciones f_{XY} , f_X y f_Y se refieren a funciones de densidad o funciones puntuales de probabilidad según si la v.a. (X, Y) es continua o discreta respectivamente.

Deducimos en particular que, si X e Y son independientes, la distribución condicionada de X (resp. Y) no depende del valor de Y (resp. X): el hecho de conocer el valor de una de las variables no proporciona información sobre la distribución de valores de la otra. En particular, deducimos que si X e Y son independientes, podemos describir completamente su distribución conjunta si conocemos sus dos distribuciones marginales.

En el ejemplo de la v.a discreta de la sección IV.2.1.1, notamos que $f_{XY}(0, 120) = 0,03 \neq f_X(0)f_Y(120)$. Por lo tanto X e Y no son independientes. En cambio, es fácil comprobar para el ejemplo de v.a continua de la sección IV.3.1.2, que se cumple que, para todo x e y , $f_{XY}(x, y) = f_X(x)f_Y(y)$: en este caso, las variables X e Y sí son independientes.

IV.5.2. Consecuencias prácticas

- Si X e Y son independientes, es fácil comprobar que cualquier suceso asociado con X es independiente de cualquier suceso asociado con Y . Es decir que

$$\mathbb{P}(a \leq X \leq b) \cap (c \leq Y \leq d) = \mathbb{P}(a \leq X \leq b)\mathbb{P}(c \leq Y \leq d).$$

- Si X e Y son independientes, se puede calcular de manera sencilla la esperanza de una función de X y de una función de Y :

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].$$

La noción de variables independientes se generaliza a más de dos variables de manera natural: X_1, X_2, \dots, X_n son v.a independientes si los sucesos asociados son independientes.

IV.6. Medidas numéricas para una v.a bidimensional

Al disponer de un modelo para la distribución conjunta de X e Y , es útil poder recurrir a alguna medida numérica que nos permita por ejemplo cuantificar el grado de asociación entre las dos variables.

IV.6.1. Definiciones

IV.6.1.1. Covarianza

La covarianza de X e Y se define como

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Utilizando la definición de la esperanza de una función de X e Y en el caso discreto y en el caso continuo, obtenemos la fórmula equivalente para la covarianza

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Notar que el cálculo de $\text{cov}(X, Y)$ se realiza por lo tanto de la manera siguiente

- (X, Y) v.a discreta:

$$cov(X, Y) = \sum_x \sum_y xyf_{XY}(x, y) - \mathbb{E}[X]\mathbb{E}[Y],$$

donde los sumatorios se realizan sobre los valores posibles de X e Y .

- (X, Y) es una v.a. continua:

$$cov(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_{XY}(x, y)dx dy - \mathbb{E}[X]\mathbb{E}[Y].$$

Notar también que la covarianza de una variable X consigo mismo es igual a la varianza de X : $cov(X, X) = \sigma_X^2$.

IV.6.1.2. Correlación

La correlación de X e Y se define como

$$\rho_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}.$$

La correlación de X e Y corresponde por lo tanto a la covarianza de las versiones tipificadas de X e Y . En particular la correlación de una variable X consigo mismo es igual a 1.

IV.6.1.3. Ejemplo para una v.a. (X, Y) discreta

Volvamos al ejemplo de la sección IV.2.1.1, su función puntual de probabilidad es

X	Y				f_X
	120	130	140	150	
0	0.03	0.1	0.15	0.2	0.48
1	0.05	0.06	0.1	0.1	0.31
2	0.21	0	0	0	0.21
f_Y	0.29	0.16	0.25	0.3	

Para calcular la covarianza de X e Y necesitamos por una parte $\mathbb{E}[X]$ y $\mathbb{E}[Y]$ y por otra parte $\mathbb{E}[XY]$. Obtenemos utilizando las distribuciones marginales de X e Y :

$$\mathbb{E}[X] = 0 \cdot 0,48 + 1 \cdot 0,31 + 2 \cdot 0,21 = 0,73$$

$$\mathbb{E}[Y] = 120 \cdot 0,29 + 130 \cdot 0,16 + 140 \cdot 0,25 + 150 \cdot 0,3 = 135,6$$

Nos queda calcular $\mathbb{E}[XY]$.

$$\begin{aligned} \mathbb{E}[XY] &= 0 \cdot 120 \cdot 0,03 + 0 \cdot 130 \cdot 0,1 + 0 \cdot 140 \cdot 0,15 + 0 \cdot 150 \cdot 0,2 \\ &+ 1 \cdot 120 \cdot 0,05 + 1 \cdot 130 \cdot 0,06 + 1 \cdot 140 \cdot 0,1 + 1 \cdot 150 \cdot 0,1 \\ &+ 2 \cdot 120 \cdot 0,21 + 2 \cdot 130 \cdot 0 + 2 \cdot 140 \cdot 0 + 2 \cdot 150 \cdot 0 \\ &= 93,2 \end{aligned}$$

Deducimos que $cov(X, Y) = 93,2 - 0,73 \cdot 135,6 = -5,78$. Para calcular la correlación de X e Y nos hacen falta además las desviaciones típicas de X e Y . Se comprueba que $\sigma_X^2 = 0,617$ mientras que $\sigma_Y^2 = 142,64$. Por lo tanto

$$\rho_{XY} = \frac{-5,78}{\sqrt{0,617} \sqrt{142,64}} = -0,62.$$

IV.6.1.4. Matriz de covarianzas y matriz de correlación

En el caso en que consideramos varias variables aleatorias X_1, X_2, \dots, X_n , podemos calcular las covarianzas y las correlaciones de cada par posible de variables, se suele presentar los resultados en forma de una matriz: la matriz de covarianzas de X_1, \dots, X_n es la matriz $n \times n, \Sigma$ cuyo elemento Σ_{ij} es igual a la covarianza de X_i y X_j , mientras que la matriz de correlaciones de X_1, \dots, X_n es la matriz $n \times n, Corr$ cuyo elemento $Corr_{ij}$ es igual a la correlación de X_i y X_j .

IV.6.2. Propiedades

1. Se puede demostrar (ver problema número 14 de la hoja de problemas de este tema) que

$$|cov(X, Y)| \leq \sigma_X \sigma_Y,$$

es decir que, para dos variables cualesquiera X e Y ,

$$-1 \leq \rho_{XY} \leq 1.$$

2. Si X e Y son independientes,

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])][\mathbb{E}[(Y - \mathbb{E}[Y])] = 0.$$

También implica que $\rho_{XY} = 0$. En cambio si $\rho_{XY} = \pm 1$, se puede demostrar que existe dos constantes a y b tal que $Y = ax + b$: existe una relación lineal determinista entre X e Y . De ahí que la correlación es una medida del grado de asociación lineal entre dos variables.

3. Usando la propiedad de linealidad de la esperanza es fácil obtener que

$$Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y).$$

En el caso particular en el que X e Y son independientes, esta relación se simplifica, dando lugar a la fórmula de propagación de los errores:

$$Var(X + Y) = Var(X) + Var(Y),$$

puesto que $cov(X, Y) = 0$.

IV.7. Algunos modelos de v.a. multidimensional

IV.7.1. Modelo multinomial

El modelo multinomial aparece como una generalización del modelo binomial: consideremos

- Tenemos un primer experimento aleatorio simple, con un k sucesos posibles A_1, \dots, A_k , que forman una partición del espacio muestral. Denotamos por $p_1 = \mathbb{P}(A_1), \dots, p_k = \mathbb{P}(A_k)$.
- Repetimos este experimento simple n veces de manera independiente.

- Consideramos la variable X_1 ="Número de veces que ha ocurrido A_1 en las n realizaciones del experimento simple, X_2 ="Número de veces que ha ocurrido A_2 en las n realizaciones del experimento simple, etc hasta X_k ="Número de veces que ha ocurrido A_k en las n realizaciones del experimento simple.

Proposición IV.7.1 *Se cumple que, para todos n_1, \dots, n_k enteros positivos o nulos tal que $n_1 + n_2 + \dots + n_k = n$,*

$$\mathbb{P}(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}.$$

Se dice que (X_1, \dots, X_k) sigue una distribución multinomial de parámetros p_1, \dots, p_k y n .

Es fácil comprobar que todos las distribuciones marginales de una multinomial son binomiales, ¿con qué parámetros?

IV.7.2. El modelo Normal multidimensional

IV.7.2.1. Caso bidimensional

Definición IV.7.1 *Consideremos un par de números reales $\vec{\mu} = (\mu_1, \mu_2) \in \mathbb{R}^2$ y una matriz Σ 2×2 simétrica y definida positiva (es decir que, para todo x en \mathbb{R}^2 , $x^T \Sigma x \geq 0$). La variable (X_1, X_2) sigue una distribución Normal bidimensional con parámetros (μ_1, μ_2) y Σ si su densidad es*

$$x = (x_1, x_2) \mapsto \frac{1}{2\pi|\Sigma|} e^{-\frac{1}{2}(x-\vec{\mu})^T \Sigma^{-1}(x-\vec{\mu})}.$$

En este caso escribimos $(X_1, X_2) \sim \mathcal{N}(\vec{\mu}, \Sigma)$.

Se puede comprobar que, si $(X_1, X_2) \sim \mathcal{N}(\vec{\mu}, \Sigma)$,

$$\mathbb{E}[X_1] = \mu_1, \quad \mathbb{E}[X_2] = \mu_2, \quad \Sigma \text{ es la matriz de covarianzas de } (X_1, X_2).$$

De la forma de la densidad Normal bidimensional, deducimos en particular la siguiente propiedad:

Propiedad: Si (X_1, X_2) sigue una distribución normal bidimensional, se cumple que X_1 y X_2 son independientes, si y solamente si su covarianza es nula.

Las curvas de nivel de la densidad bidimensional Normal son muy ilustrativas a la hora de visualizar las campanas de Gauss asociadas (estas campanas son en tres dimensiones). En la figura IV.1, las dos componentes X_1 y X_2 son independientes y además sus varianzas son iguales, más concretamente $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1$, $\Sigma_{22} = 1$ y $\Sigma_{12} = 0$.

En la figura IV.2, las dos componentes X_1 y X_2 siguen siendo independientes pero ahora sus varianzas son distintas, más concretamente $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1$, $\Sigma_{22} = 0,25$ y $\Sigma_{12} = 0$. Las curvas de nivel aparecen como elipses, cuyos ejes coinciden con los ejes del sistema de coordenadas.

Finalmente, si las dos componentes no son independientes, las curvas de nivel siguen formando elipses pero sus ejes presenten un ángulo respecto a los ejes del sistema de coordenada. En la figura IV.3, se representan las curvas de nivel para la densidad Normal bidimensional si $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1,125$, $\Sigma_{22} = 0,5$ y $\Sigma_{12} = 0,375$. Esto implica en particular que su correlación es $\rho_{X_1 X_2} = 0,5$.

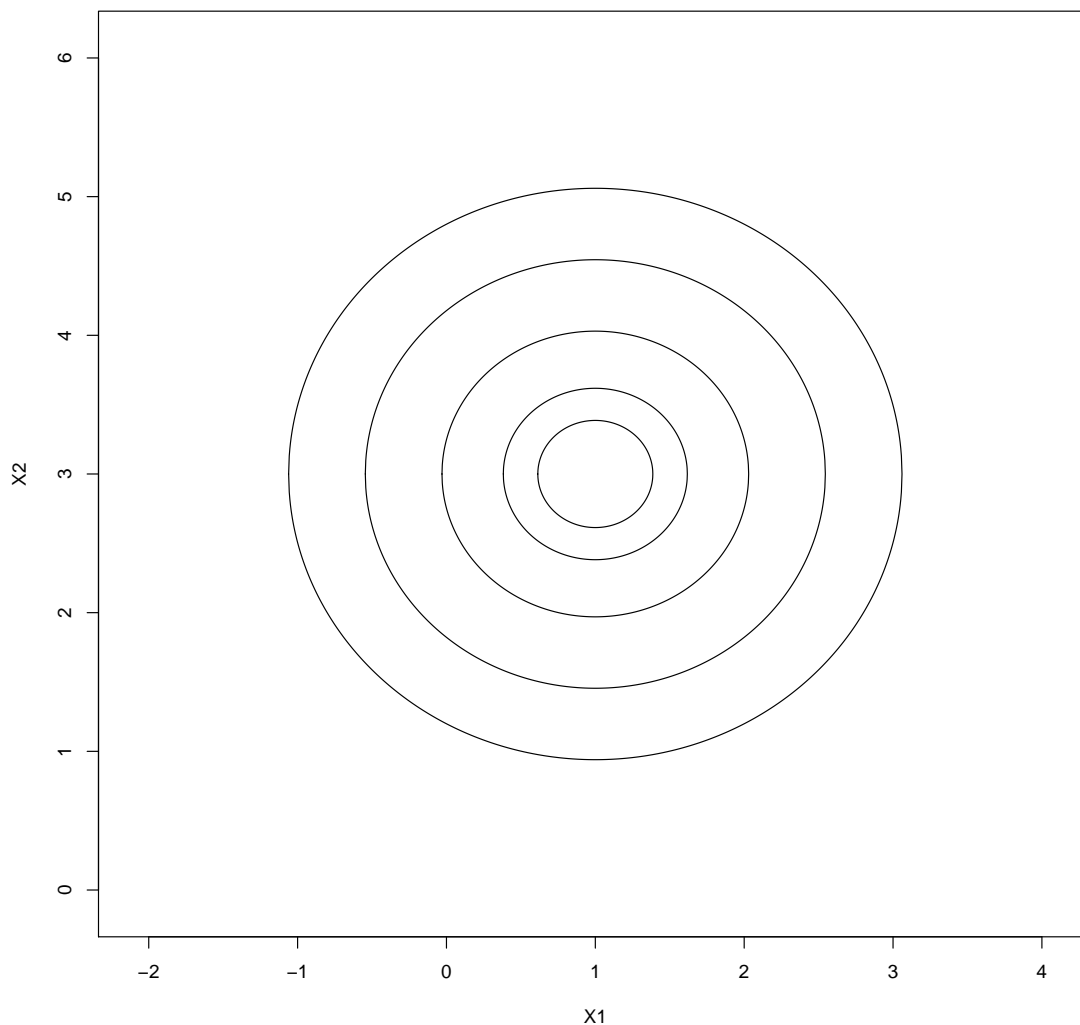


Figura IV.1: Curvas de nivel de la densidad Normal bidimensional si los dos componentes son independientes con varianzas iguales, $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1$, $\Sigma_{22} = 1$ y $\Sigma_{12} = 0$.

IV.7.2.2. Caso n -dimensional

Definición IV.7.2 Consideremos $\vec{\mu} = (\mu_1, \dots, \mu_n)$ en \mathbb{R}^n y una matriz Σ $n \times n$ simétrica y definida positiva.

La variable n -dimensional $X = (X_1, \dots, X_n)$ sigue una distribución Normal n -dimensional con parámetros $\vec{\mu}$ y Σ si su densidad es

$$x \in \mathbb{R}^n \mapsto \frac{1}{(2\pi|\Sigma|)^{n/2}} e^{-\frac{1}{2}(x-\vec{\mu})^T \Sigma^{-1}(x-\vec{\mu})}.$$

Se puede comprobar que la media de cada X_i es μ_i y que Σ es la matriz de covarianza de X .

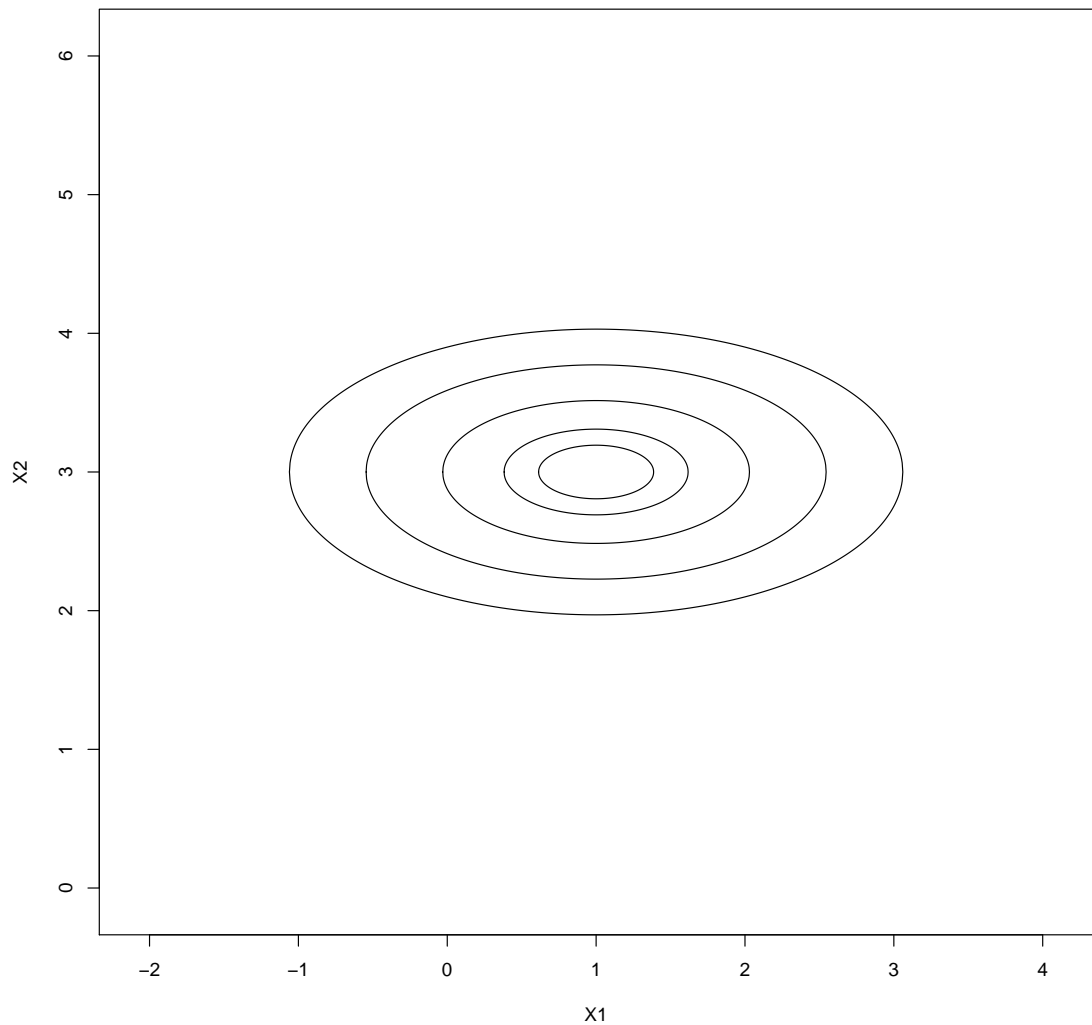


Figura IV.2: Curvas de nivel de la densidad Normal bidimensional si los dos componentes son independientes, pero sus varianzas son distintas, $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1$, $\Sigma_{22} = 0,25$ y $\Sigma_{12} = 0$.

Acabamos el tema con una propiedad fundamental de la distribución Normal n -dimensional, llamada propiedad de reproductividad de la distribución Normal.

Proposición IV.7.2 Si $X = (X_1, \dots, X_n) \sim \mathcal{N}(\vec{\mu}, \Sigma)$, para todos números reales a_1, \dots, a_n , se cumple que

$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n \text{ sigue una distribución Normal.}$$

¿Podrías caracterizar su media y su varianza?

Se deduce en particular de la proposición que las distribuciones marginales de una variable Normal n -dimensional son todas normales.

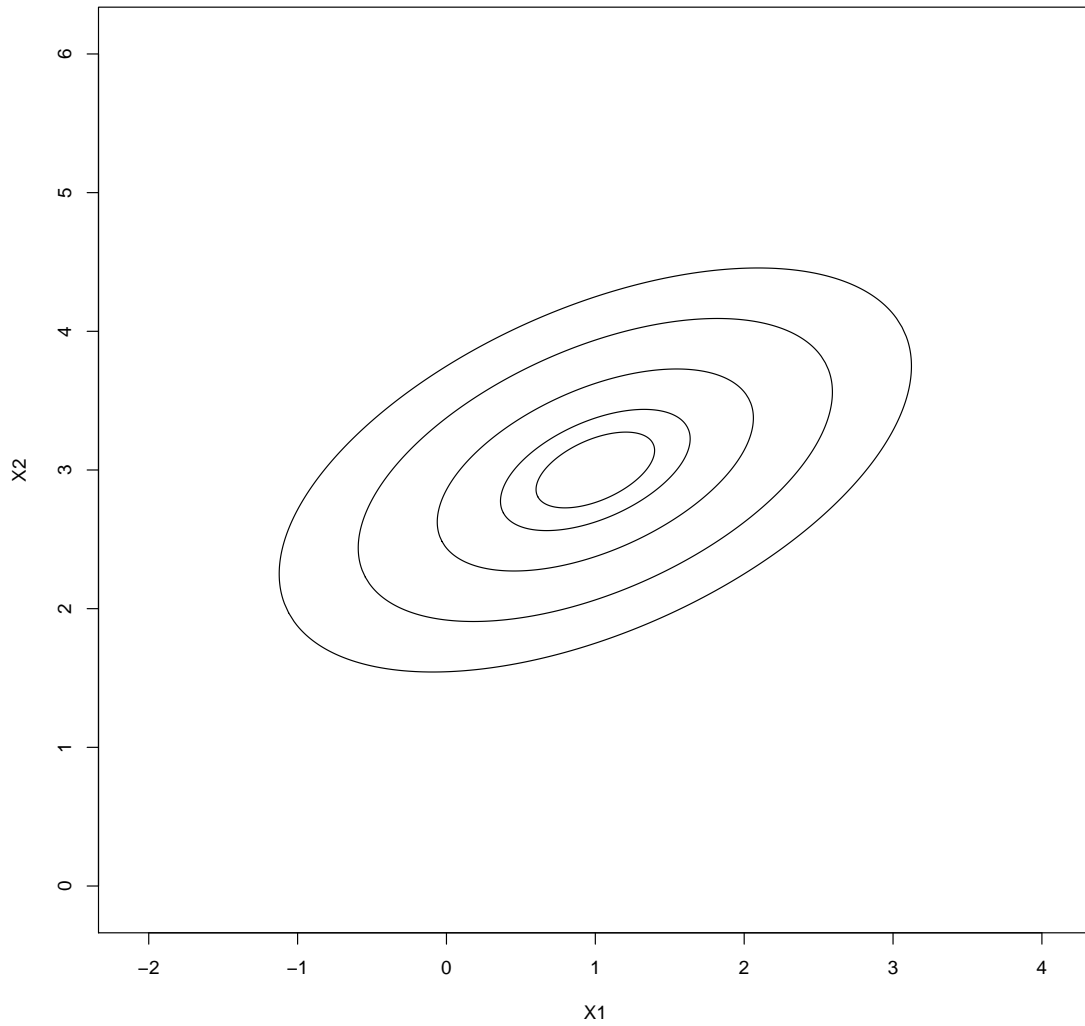


Figura IV.3: Curvas de nivel de la densidad Normal bidimensional si los dos componentes no son independientes, $\mu_1 = 1$, $\mu_2 = 3$, $\Sigma_{11} = 1,125$, $\Sigma_{22} = 0,5$ y $\Sigma_{12} = 0,375$, lo que implica $\rho_{X_1 X_2} = 0,5$.



2013: Año Internacional de la Estadística. ¿Sabías qué...?

Se puede utilizar las estadísticas de búsquedas que los usuarios de Google realizan cada día acerca de consejos y ayudas sobre la gripe, para monitorizar y anticipar día a día la actividad de esta enfermedad en nuestros países.. (ver <http://www.google.org/flutrends/>)

Fuente: J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski & L. Brilliant "Detecting influenza epidemics using search engine query data", Nature (2009), 457, 1012-1014

statistics

Muestreo y distribuciones muestrales

V.1. Introducción

Consideramos un experimento aleatorio para el cual estamos dispuestos a escoger un modelo, posiblemente con uno o varios parámetros que tendremos que ajustar.

Ejemplos

- Me interesa una moneda para tirar a cara o cruz. El experimento es “Tirar la moneda” y la variable X corresponde al resultado, su distribución se describe como: X puede tomar dos valores c (Cara) o $+$ (Cruz) con las probabilidades: $\mathbb{P}[X = c] = p$ y $\mathbb{P}[X = +] = 1 - p$. p es por lo tanto la probabilidad de que salga cara, y es un parámetro de nuestro modelo. En el caso en que confiamos en que la moneda no está trucada, nuestro modelo considerará que $p = 1/2$. Para sacar información sobre p y comprobar en particular que la moneda no está trucada, repetiremos un cierto número de veces el experimento.
- Para las próximas elecciones generales, queremos determinar la proporción de gente que tiene intención de ir a votar, es decir queremos estimar la tasa de participación. El censo electoral para España tiene unos 32 millones de personas. Es claramente imposible entrevistar a todas las personas del censo. En cambio realizaremos una encuesta, escogiendo al azar una muestra de unas 3000 personas entre el censo y preguntándoles si tienen intención de ir a votar.
- El índice de audiencias manda en la programación de televisión. Pero ¿cómo saben cuántos espectadores vieron un partido dado o un programa determinado? A mí nunca me han preguntado... En realidad, una encuesta se realiza de manera automática y continua: una empresa especializada llamada SOFRES (<http://www.sofresam.com>) ha escogido al azar unos 3300 hogares que representan unas 10000 personas de entre un total de aproximadamente 39 500 000 espectadores potenciales. En cada uno de estos hogares, instala un aparato llamado “audímetro” que graba cuál es el programa que se está viendo en cada momento.

- Quiero conocer la concentración de un determinado producto en una solución. Pienso que es razonable que la distribución de los valores proporcionados por mi aparato de medición sea una normal con media μ y desviación típica σ desconocidas. El centro de esta distribución, es decir μ , será por lo tanto lo más representativo de la concentración que intento determinar. Para estimar μ , repetiré la medición varias veces.

Pero surge una pregunta evidente:

Pregunta: ¿Cómo sabemos que nuestra estimación es fiable? ¿Por qué limitándose a unas 3000 personas, se puede extrapolar el resultado con confianza a una población de 30 millones? Además está claro que el resultado que obtengo depende de la muestra particular que haya escogido, si escojo otra muestra me sale otro resultado. Este hecho se llama la variabilidad muestral.

Intento de respuesta: Consideremos el caso del sondeo en el que se busca estimar la tasa de participación antes de unas elecciones. Para intentar convencer al lector de que el riesgo que corro al extrapolar el resultado de una muestra de 3000 personas a la población de 32 millones no es excesivo, llevo a cabo un estudio de simulación:

- Construyo en mi ordenador un fichero con 32 millones de ceros y unos, que representará el censo electoral. Los unos representarán a las personas que sí tienen la intención de ir a votar, mientras que los ceros a los que no piensan ir a votar. En el fichero que construyo, el 70 % de los 32 millones de datos son unos, mientras que el 30 % son ceros. (70 % es una tasa razonable de participación en unas elecciones)
- Extraigo al azar una muestra de 3000 datos del fichero completo, hago el recuento de los unos, y encuentro que la proporción de unos en esta muestra es de 0.71. Por lo tanto, en este caso, mi estimación es muy buena: estimo la tasa de participación en 71 % mientras que la auténtica, es decir, la de la población (el fichero) es de 70 %. ¿Os he convencido? Seguro que algún lector desconfiado dirá: “ *no demuestra nada, ha tenido suerte de que en la muestra que ha escogido, la proporción de unos sea próxima a la proporción poblacional, pero con otra muestra podría salir otro resultado peor.*” De acuerdo, el argumento es válido... Pero para convencerle, voy a coger otra muestra al azar de 3000 datos, y encuentro que la proporción muestral de unos es 0.72. Sigue estando muy bien, ¿no? ¿Sigue sin convencerle? Bueno, puedo repetir la extracción de muestras hasta 10 000 veces por ejemplo, y guardo los valores que encuentro para la proporción de 1 en cada una de estas 10000 muestras en una variable llamada \hat{p} .
- Realizo un histograma de los 10000 valores de \hat{p} , el resultado aparece en la figura V.1. Una primera conclusión se impone: la gran mayoría de las muestras han proporcionado un valor de \hat{p} entre 0.68 y 0.72, lo que corresponde a una muy buena estimación del valor de la proporción poblacional. Por lo tanto este estudio simulado demuestra que al escoger una muestra de 3000 personas, es muy probable que el valor de la proporción de 1 en la muestra esté bastante próxima (menos de dos puntos) de la proporción de 1 en la población, aunque ésta sea muchísimo más grande que la muestra.

Podemos dar un paso más en la utilización de este estudio simulado: si considero ahora el experimento “extraer una muestra de tamaño 3000 en la población”, \hat{p} es la

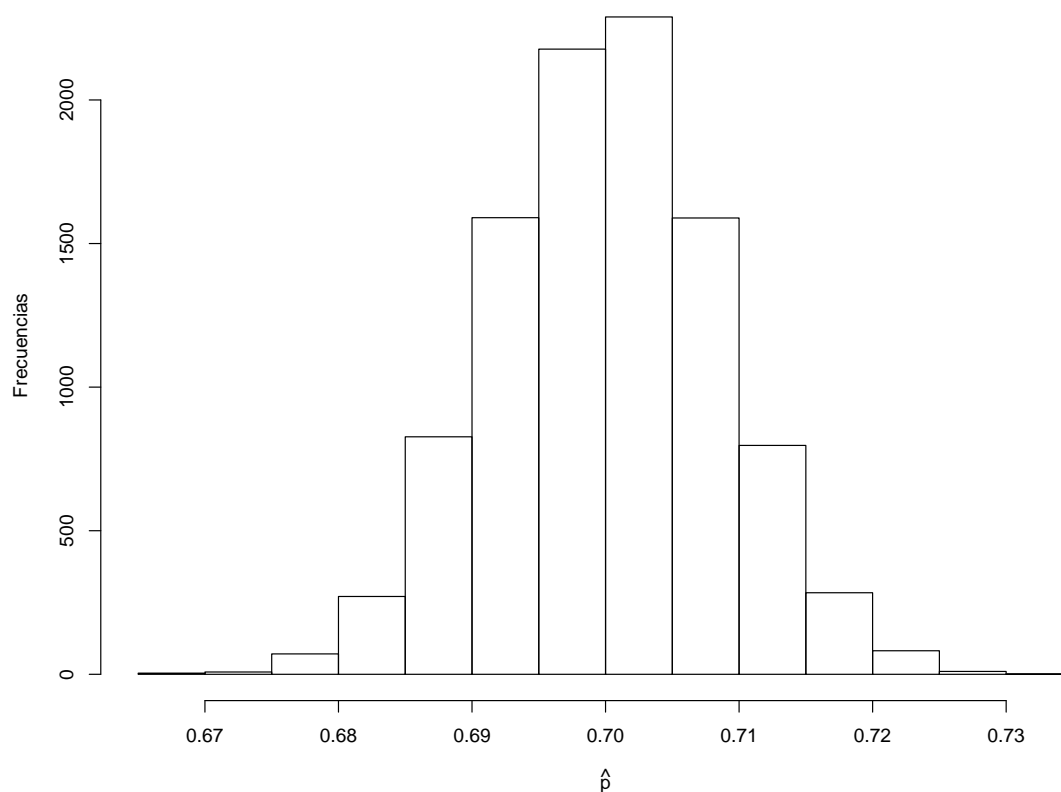


Figura V.1: Histograma de los valores de \hat{p} para 10000 muestras extraídas

variable “proporción de 1 en la muestra extraída”. Quiero formular un modelo para su distribución. El histograma en la figura V.1 me sugiere que puedo escoger una distribución normal para \hat{p} . De hecho en la figura V.2, se aprecia que el ajuste por una normal con media $\mu = 0,70$ y desviación típica $\sigma = 0,008$ es muy bueno. Utilizando entonces la regla de 68 % - 95 % - 99.7 %, deduzco en particular que al escoger al azar en la población una muestra de tamaño 3000, la probabilidad de que la proporción muestral \hat{p} se encuentre entre $0,7 - 2 \times 0,008 = 0,694$ y $0,7 + 2 \times 0,008 = 0,716$ es del 95 %.

Nota. Puesto que escoger una muestra de 3000 personas da tan buen resultado, podríamos preguntarnos si podríamos ahorrarnos algo y extraer una muestra más pequeña. Repitamos por ejemplo el estudio simulado con muestras de sólo 100 personas. El histograma que obtenemos aparece en la figura V.3. Observamos que en este caso el histograma es muchísimo más chato, y que la dispersión de los valores de \hat{p} es mucho mayor: es más probable, al escoger una muestra de 100, que la proporción muestral esté bastante alejado del objetivo 0.7.

Toda la teoría desarrollada acerca de los sondeos utiliza de manera crucial el hecho de que antes de extraer la muestra, se dispone de un modelo para la distribución de \hat{p} por ejemplo, tal como lo hemos ilustrado con nuestro ejemplo simulado. Este

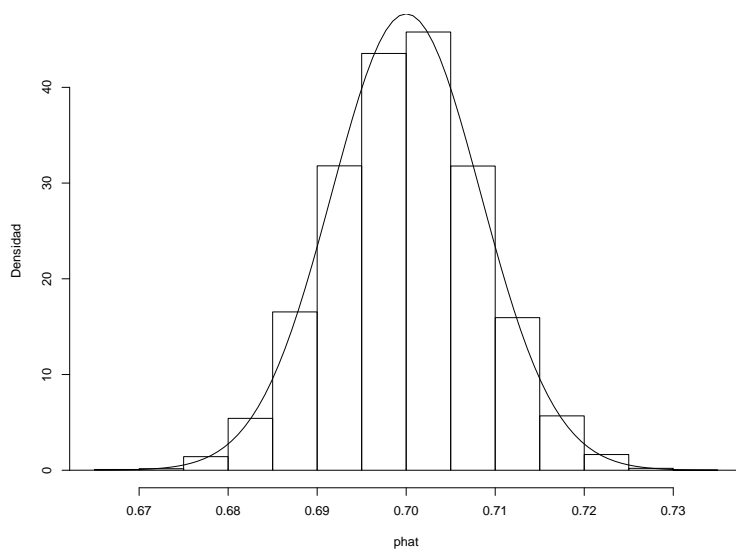


Figura V.2: Ajuste de una normal al histograma de los valores de \hat{p}

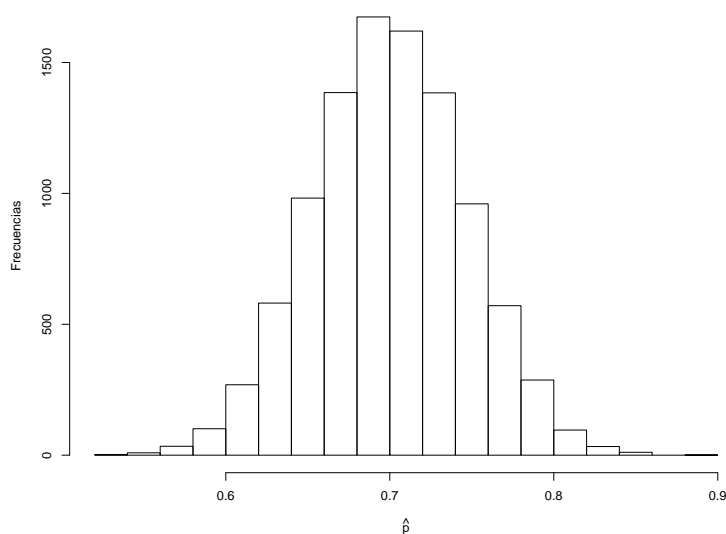


Figura V.3: Histograma de los valores de \hat{p} para 10000 muestras de tamaño 100 extraídas

modelo permite en particular decidir si, fijado el error máximo que se está dispuesto a cometer respecto a la proporción poblacional, el tamaño de la muestra es suficiente como para que el riesgo de cometer un error mayor es lo suficientemente pequeño.

Introducimos dos términos fundamentales en estadística:

Definición. Cualquier cantidad calculada a partir de las observaciones de una muestra se llama **estadístico**. La distribución de los valores que puede tomar un estadístico respecto a todas las muestras de tamaño n que se podría extraer se llama **distribución muestral** de este estadístico.

V.2. Muestra

Formalizamos el contexto y introducimos el concepto de muestra:

Consideramos un experimento aleatorio y una v.a X .¹ Al querer obtener información sobre algún parámetro del modelo que hemos escogido para la distribución de los valores de X , vamos a repetir el experimento n veces de manera independiente y consideramos las variables X_1 “valor de X obtenido en la primera realización del experimento”, \dots , X_n “valor de X obtenido en la n -ésima realización del experimento”. Las variables X_1, X_2, \dots, X_n son independientes y claramente la distribución de cada variable X_i coincide con la distribución de X . En este caso decimos que (X_1, X_2, \dots, X_n) constituye una *muestra aleatoria simple* de la distribución de X .

V.3. La media muestral

Supongamos que nos interesamos por el valor μ , la media de la v.a X . Escogemos una muestra, y calcularemos la media de esta muestra, llamada media muestral. Para controlar lo próximo que estará su valor de μ , consideramos el experimento que consiste en extraer una muestra aleatoria simple de la distribución de X , la media muestral es la variable aleatoria (su valor depende de la muestra escogida)

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}.$$

¿Qué podemos decir de la distribución de los valores que puede tomar \bar{X} ? Empezaremos por estudiar cuál será el centro y la dispersión de esta distribución.

V.3.1. Esperanza y varianza de \bar{X}

V.3.1.1. Esperanza

Tenemos que

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n}\mathbb{E}[X_1 + \dots + X_n] = \frac{1}{n}(\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]).$$

Puesto que la distribución de cada X_i es la misma que la distribución de X , deducimos que $\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n] = \mu$, y

$$\mathbb{E}[\bar{X}] = \frac{1}{n}(n \cdot \mu) = \mu,$$

es decir que el centro de la distribución de la media muestral coincide con el centro de la distribución de X .

¹En algunos casos, este experimento aleatorio consistirá en escoger al azar un individuo de una población muy grande, y X será el valor de la variable de interés para este individuo concreto. Llamaremos entonces media de X la media *poblacional* y su varianza, la varianza *poblacional*

V.3.1.2. Varianza

Utilizando la fórmula de propagación de los errores, ver Tema 4, obtenemos que

$$\text{var}[\bar{X}] = \text{var}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n^2} \text{var}[X_1 + \dots + X_n] = \frac{1}{n^2} (\text{var}[X_1] + \dots + \text{var}[X_n]),$$

lo que implica que

$$\text{var}(\bar{X}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

o de forma equivalente

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

¡La dispersión que presentan los valores de \bar{X} es \sqrt{n} más pequeña que la dispersión de X !

V.3.1.3. Consecuencia práctica

Quiero realizar una medición con un aparato. El experimento aleatorio es “llevar a cabo una medición”, mientras que la variable X es “valor proporcionado por el aparato”.

Los valores de X variarán pero lo deseable es que su centro μ coincida con el valor exacto de la cantidad que busco determinar: si $\mathbb{E}[X] = \text{valor exacto}$, decimos que el aparato es exacto.

Por otra parte, queremos que los valores proporcionen presenten la menor dispersión posible: si $\sigma = \sigma_X$ es pequeña, decimos que el aparato es preciso. Tenemos entonces varios casos posibles, tal como está ilustrado en la Figura V.4, con la analogía de la medición con un disparo en una diana: el centro de la diana representa el valor exacto de lo que buscamos determinar...

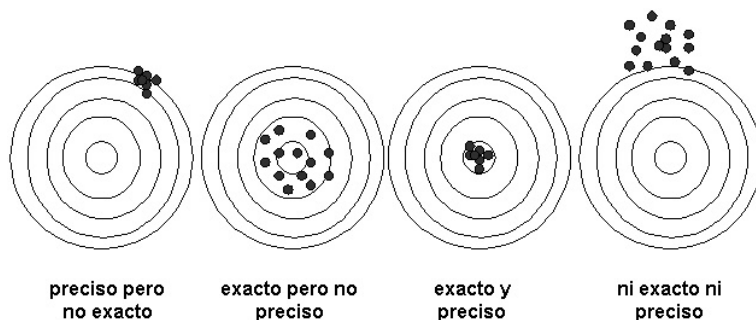


Figura V.4: Analogía de la medición con un disparo en una diana

Si nuestro aparato de medición no es exacto, podemos intentar calibrarlo para corregir la desviación sistemática que presenta. En cambio, si no es preciso, tiene difícil arreglo. Sin embargo exista una manera de mejorar la precisión de un aparato de medición: basta con repetir un número suficiente de veces la medición y proporcionar la media de los valores obtenidos: la desviación típica de los valores que proporcionaría con este método es \sqrt{n} veces más pequeña que la de los valores proporcionados si me limito a una medición.

V.3.2. Distribución de la media muestral

En la subsección anterior, hemos caracterizado la media y la desviación típica de la distribución de los valores de la media muestral \bar{X} . Hay que enfatizar el hecho de que estos resultados se obtienen sin hipótesis sobre la forma de la distribución de X . ¿Podemos decir algo más sobre la distribución de los valores de \bar{X} , ahora que sabemos cuáles son su centro y su dispersión?

V.3.2.1. Si la distribución de X es Normal

Si hemos modelizado la v.a X por una distribución Normal $\mathcal{N}(\mu, \sigma^2)$ y consideramos una muestra aleatoria simple de X , sabemos por la reproductividad de la distribución Normal que $X_1 + X_2 + \dots + X_n$ sigue también una distribución normal. Se cumple por lo tanto

Proposición V.3.1 *Si $X \sim \mathcal{N}(\mu, \sigma^2)$, y si \bar{X} es la media muestral basada en una muestra aleatoria simple de la distribución de X ,*

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

o, de manera equivalente,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Como ejemplo, consideremos un aparato de medición que proporciona valores que se distribuyen según una Normal, con una media de 120 y una desviación típica de 12. Por la propiedad de la distribución Normal, el 95 % de los valores están entre $\mu - 2\sigma$ y $\mu + 2\sigma$, es decir entre 96 y 144. En cambio, si repito 9 veces la medición y proporciono la media de estas nueve mediciones, el 95 % de los valores que obtendría con este procedimiento se encontrarían entre $\mu - 2\sigma/\sqrt{n}$ y $\mu + 2\sigma/\sqrt{n}$, es decir entre 112 y 128, lo que implica una precisión mucho mayor.

V.3.2.2. Si la distribución de X es desconocida o no es normal

Si la distribución de X es desconocida, no podemos hacer milagros: no podemos decir nada exacto sobre la distribución de \bar{X} , excepto sobre su media y su desviación típica, ver sección V.3.1. Sin embargo, si el tamaño muestral n es grande, se sabe que esta distribución se puede aproximar por una distribución Normal.

Teorema V.3.1 *Teorema Central del Límite Consideremos (X_1, \dots, X_n) una muestra aleatoria simple de la distribución de X con media μ y varianza σ^2 . Si n es “suficientemente” grande, se puede aproximar la distribución de \bar{X} por una Normal con media μ y varianza σ^2/n :*

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ aproximadamente.}$$

¿Cuándo se considera que n es “suficientemente” grande? No hay por desgracia ninguna respuesta universal, depende de la forma de la distribución de X : si ésta no es muy diferente de una distribución Normal, no hace falta un n muy grande para que la aproximación de la distribución de la media muestral por una Normal sea

satisfactoria. En cambio, si es muy distinta de una distribución Normal, será necesario una muestra grande. Se suele considerar como indicación que n mayor de 30 es suficiente en la mayoría de los casos (pero no es más que una indicación...)

Por otra parte, este teorema, fundamental en estadística, explica la importancia de la distribución Normal: aparece de manera natural, asociada a cualquier distribución, si consideramos la distribución de la media muestral, o de la suma de realizaciones independientes. En particular, si un error de medición se puede considerar como la suma de muchas pequeñas perturbaciones independientes, el Teorema Central del Límite implica que la distribución de sus valores es aproximadamente Normal.

V.4. La varianza muestral

Consideremos ahora un experimento al que asociamos una v.a X cuya distribución de valores modelizamos por una Normal con media μ y varianza σ^2 . Repetimos n veces el experimento y obtenemos una m.a.s (X_1, X_2, \dots, X_n) de la distribución de X . ¿Qué podemos decir de la distribución de la varianza muestral

$$s^2 = \frac{n}{n-1}(\overline{X^2} - (\bar{X})^2)?$$

Es posible demostrar la proposición siguiente

Proposición V.4.1 1. Las v.a \bar{X} y s^2 son independientes.

2. La densidad de $(n-1)s^2/\sigma^2$ es proporcional a

$$x^{(n-1)/2}e^{-x/2}, \quad \text{si } x > 0.$$

La distribución correspondiente se llama χ^2 (ji-cuadrado) con $(n-1)$ grados de libertad. Escribimos

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

En general, una v.a. X sigue una distribución χ^2 con $k \in \mathbb{N}$ grados de libertad si su densidad es proporcional a

$$x \mapsto x^{k/2}e^{-x/2}, \quad \text{si } x > 0.$$

En la figura V.5, se representa la densidad de una distribución χ^2 con distintos grados de libertad.

V.5. Distribución t de Student

En la sección 3, hemos utilizado el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \tag{V.1}$$

que sigue una distribución Normal estándar si \bar{X} es la media de una muestra aleatoria simple de una distribución Normal $\mathcal{N}(\mu, \sigma^2)$.

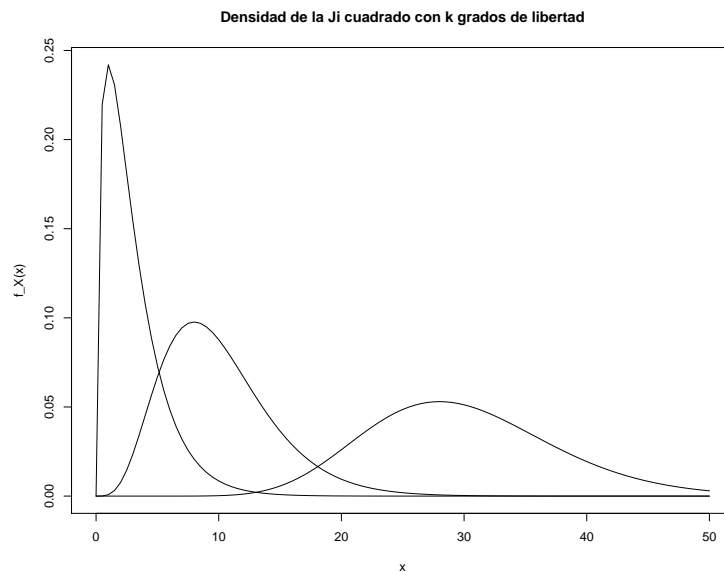


Figura V.5: Densidad de la distribución χ^2 con $k = 3, 10$ y 30 grados de libertad (respectivamente de izquierda a derecha)

Si desconocemos el valor de σ , lo estimaremos por S la desviación típica muestral

$$S = \sqrt{\frac{n}{n-1}(\overline{X^2} - (\bar{X})^2)}.$$

El estadístico que resulta de sustituir en (V.1) σ por S es

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Definición V.5.1 Consideramos (X_1, \dots, X_n) una muestra aleatoria simple de una distribución $\mathcal{N}(\mu, \sigma^2)$, sea \bar{X} la media muestral, la distribución de los valores de

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

se llama *distribución t de Student con $n-1$ grados de libertad*. Escribimos $T \sim t_{n-1}$.

La distribución de T depende por lo tanto del tamaño n de la muestra, a través de los llamados “grados de libertad”. Se puede demostrar que la densidad f_{t_k} de la distribución t de Student con k grados de libertad admite la siguiente expresión:

$$f_{t_k}(t) \propto \frac{1}{(1+t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty,$$

donde el símbolo \propto significa “es proporcional a”, es decir que existe una constante K tal que $f_{t_k}(t) = K \frac{1}{(1+t^2/p)^{(p+1)/2}}$. Por las propiedades de una función de densidad se puede deducir que la constante es

$$K = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{\sqrt{p\pi}},$$

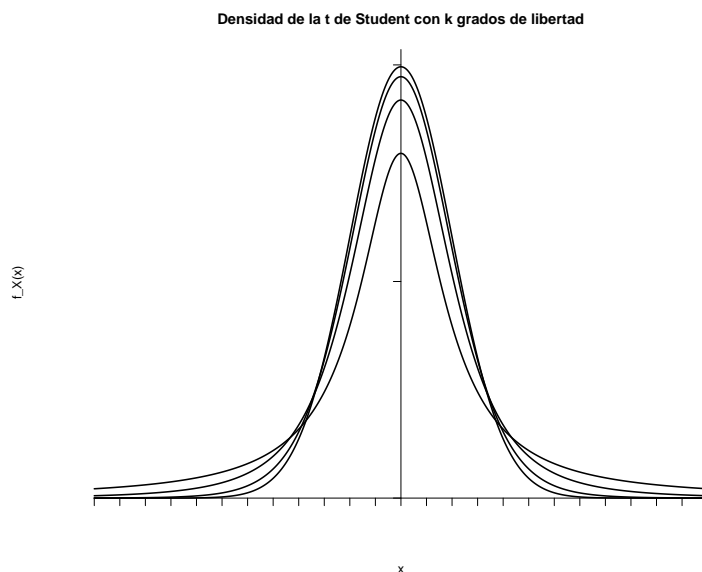


Figura V.6: Densidad de la distribución t de Student con 1, 3, 10 y 150 grados de libertad respectivamente (de la densidad más chata a la más puntiaguda)

donde Γ denota la función Gamma².

La distribución t tiene colas más pesadas que la distribución Normal, lo que es intuitivamente natural puesto que, al obtenerse T sustituyendo σ por S , el denominador de T presenta ahora también variabilidad. Esta variabilidad en el denominador resulta en que T puede tomar con más probabilidad valores más extremos. Sin embargo, si los grados de libertad aumentan, la variabilidad de S disminuye, y la distribución t de Student asociada se parece más a una Normal.

En la figura V.6, se representa la densidad de la distribución T de Student para varios valores de los grados de libertad.

V.6. La proporción muestral

Hay situaciones en las que la v.a X de interés tan sólo puede tomar el valor 0 ó 1, éste último con la probabilidad p , pensamos por ejemplo, en el experimento que consiste en producir una pieza con una máquina que produce una proporción p de defectuosos, X toma el valor 1 si la pieza es defectuosa, y 0 si la pieza es correcta, o en el ejemplo del sondeo para estimar la tasa de participación antes de unas elecciones. Para sacar información sobre p , repetiremos el experimento n veces de manera independiente, contaremos el número N de veces que la v.a X ha tomado el valor 1, es decir que fabricamos n piezas con la máquina y contamos el número N de defectuosas, o preguntaremos a n personas si tienen intención de ir a votar, para los dos ejemplos concretos que hemos mencionado. La proporción de “Unos” en la muestra se llama la proporción muestral y la denotamos por \hat{p} . Está claro que

²La función Gamma tiene la expresión siguiente: para cualquier real $\alpha > 0$, $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$.

tenemos

$$\hat{p} = \frac{N}{n}.$$

V.6.1. Cálculos exactos para la distribución de \hat{p}

El número de “Unos” en la muestra es el número de veces que ha salido “1” en n realizaciones independientes del experimento, su distribución es por lo tanto Binomial de parámetros n y p , la probabilidad de que salga “1” en una realización del experimento:

$$N \sim B(n, p).$$

Cálculos exactos para la distribución de \hat{p} se podrán realizar utilizando que $\hat{p} = N/n$ y el hecho que $N \sim B(n, p)$, tal como viene ilustrado en el ejemplo siguiente:

Ejemplo V.6.1 *Cuando está bien ajustada, una máquina produce piezas con sólo 1 % de defectuosos. Para realizar un control de la calidad de la producción, se extrae diariamente una muestra de 100 piezas, y se calcula la proporción muestral de defectuosos. Si la máquina está bien ajustada, ¿cuál es la probabilidad de que, en una de estas muestras, haya más de 2 % de defectuosos?*

Queremos calcular

$$\mathbb{P}(\hat{p} > 0,02) = \mathbb{P}\left(\frac{N}{100} > 0,02\right) = \mathbb{P}(N > 2),$$

siendo $N \sim B(100, 0,01)$ si la máquina está bien ajustada. Tenemos

$$\begin{aligned} \mathbb{P}(N > 2) &= 1 - \mathbb{P}(N \leq 2) = 1 - [\mathbb{P}(N = 0) + \mathbb{P}(N = 1) + \mathbb{P}(N = 2)] \\ &= 1 - \left[\binom{100}{0} 0,01^0 0,99^{100} + \binom{100}{1} 0,01^1 0,99^{99} + \binom{100}{2} 0,01^2 0,99^{98} \right] \simeq 0,08 \end{aligned}$$

Por lo tanto, si la máquina está bien ajustada, sólo hay una probabilidad de 0.08 de observar 3 o más piezas defectuosas en una muestra de 100.

En particular, si un día observo 3 piezas defectuosas en la muestra que he extraído, hay dos posibilidades: a) la máquina está bien ajustada pero he tenido mala suerte (sólo había 8 posibilidades entre 100 de que esto ocurriera), b) en realidad es un síntoma de que la máquina está mal ajustada... Este simple ejemplo ilustra la idea básica del control estadístico de calidad.

V.6.2. Distribución aproximada de \hat{p}

Los cálculos exactos que hemos descrito en el apartado anterior se pueden volver muy laboriosos si se necesita evaluar un gran número de probabilidades individuales. En el caso en que se cumplen las condiciones de aproximación de la distribución Binomial, la distribución de N se puede aproximar por una Normal $\mathcal{N}(np, np(1-p))$, y por lo tanto \hat{p} sigue aproximadamente una distribución Normal con media $np/n = p$ y varianza $np(1-p)/n^2 = p(1-p)/n$:

$$\text{Si } np > 5, n(1-p) > 5 \quad \hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right), \quad \text{aproximadamente}$$

Esta propiedad de aproximación justifica en particular las formas de campanas de Gauss que aparecen para los histogramas de \hat{p} en la introducción, ver Figuras V.2 y V.3.

Notar por otra parte que para el ejemplo del apartado anterior no se cumplen las condiciones de aproximación...

V.7. Introducción a las gráficas de control

Conocer las distribuciones muestrales de algunos estadísticos destacados como la media muestral, la varianza muestral o la proporción muestral ha propiciado que se propongan procedimientos de control estadístico de calidad en contextos industriales. Veremos en esta sección una introducción a las gráficas de control, en una versión algo simplificada, pero que permite ilustrar sus fundamentos.

Las gráficas de control permiten comprobar de manera continua que se mantiene constante la calidad de una producción, favoreciendo la intervención rápida en el caso en que se detecta que ésta se deteriora.

V.7.1. Gráfica de control \bar{X} .

Consideremos el contexto siguiente: una empresa identifica la concentración en CaCO_3 como una característica importante de la calidad de su producto. Idealmente esta concentración debería ser igual a 55, pero la variabilidad es inevitable. Sin embargo se asume que, en condiciones normales de producción los valores de la concentración se distribuyen según una distribución aproximadamente Normal con desviación típica $\sigma = 8$. Para controlar la calidad de la producción, analiza 4 envases de producto, calculando a continuación la media de los cuatro valores obtenidos. En la tabla siguiente, se recogen los datos correspondientes a veinte controles.

Muestra n ^o	\bar{x}	Muestra n ^o	\bar{x}
1	54.0	11	53.1
2	59.1	12	61.1
3	54.0	13	61.5
4	56.5	14	67.7
5	60.5	15	64.9
6	56.0	16	67.6
7	47.3	17	66.9
8	51.7	18	67.1
9	62.9	19	73.5
10	64.7	20	66.4

¿Cómo comprobar que la calidad de la producción sigue conforme con los criterios fijados? es decir, ¿cómo detectar que el instrumento de producción se ha desajustado por ejemplo? Si representamos la secuencia de los valores calculados para \bar{x} en los controles consecutivos, obtenemos la gráfica de la Figura V.7, donde también se ha dibujado una línea horizontal para indicar la concentración ideal 55. Parece sin duda que la tensión de los monitores va aumentando y alejándose del objetivo 55, pero ¿cómo definir una regla que nos sirva de señal de alarma?

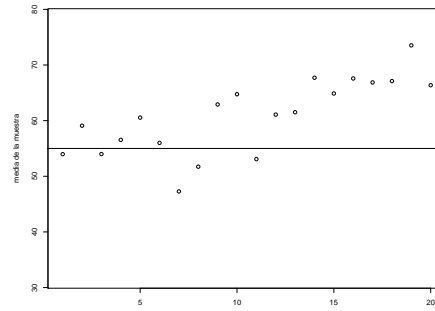


Figura V.7: Valores consecutivos de \bar{x} , ejemplo de la concentración en NaCO_3 .

Formalicemos el contexto: consideramos la v.a $X = \text{“concentración de NaCO}_3\text{”}$. Sabemos que $X \sim \mathcal{N}(\mu, \sigma^2)$ con $\sigma = 8$. También sabemos que en condiciones normales de producción, se debe cumplir que $\mu = 55$. Si escojemos al azar cuatro monitores en la producción de una hora, y llamamos \bar{X} la media de las tensiones correspondientes, sabemos que los valores de \bar{X} se distribuyen según una Normal de media μ y de desviación típica $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, es decir $8/2 = 4$. En particular si μ es efectivamente igual a 55, se espera que el 99,7% de los valores de \bar{X} se encontrarán entre $\mu - 3\sigma_{\bar{X}}$ y $\mu + 3\sigma_{\bar{X}}$, es decir entre 60.4 y 49.6.

Por consiguiente, si para una muestra, observamos un valor de \bar{X} fuera de este rango de valores, es razonable pensar que el proceso de producción se ha desajustado, puesto que sólo había una probabilidad de 3 entre 1000 que esto ocurriera, siendo el proceso bien ajustado (es decir siendo μ igual a 55).

Realizar una gráfica de control \bar{X} consiste por lo tanto, suponiendo que los valores de la variable que queremos controlar siguen aproximadamente una Normal y que conocemos su desviación típica, en representar en una gráfica los valores de \bar{X} que vamos obteniendo, junto con tres líneas horizontales:

- la línea objetivo, en nuestro caso $\mu = 55$,
- el límite de control superior en $\mu + 3\sigma/\sqrt{n}$, en nuestro caso, 60.4.
- el límite de control inferior en $\mu - 3\sigma/\sqrt{n}$, en nuestro caso, 49.6.

En la Figura V.8, se representa la gráfica de control para este ejemplo. A partir de la muestra número 14 se detecta que el proceso está fuera de control, y que la calidad se ha deteriorado.

V.7.2. Gráfica de control \hat{p}

En algunas situaciones, la calidad de la producción no se mide a través de una variable X sino a través de la proporción de defectuosos producidos. En estos casos se monitorea la calidad utilizando una gráfica de control \hat{p} .

Para llevar a cabo el control utilizando las mismas ideas que para la gráfica de control \bar{X} , recurrimos a la distribución muestral de \hat{p} . Sabemos que si $np > 5$ y $n(1-p) > 5$, ésta se puede aproximar por una Normal:

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right), \text{ aproximadamente.}$$

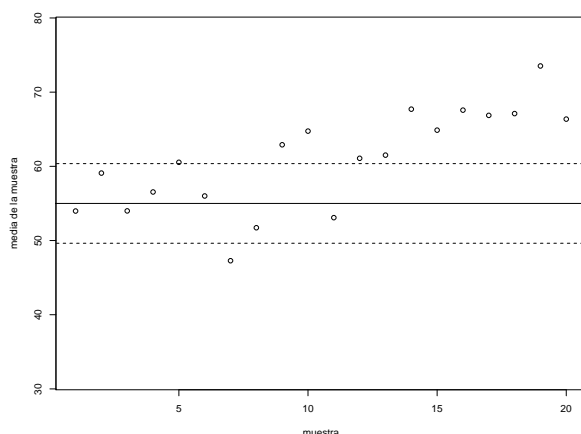


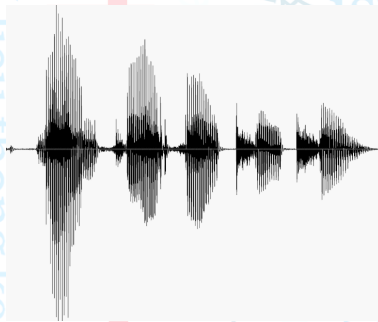
Figura V.8: Ejemplo de gráfica de control \bar{x} .

La gráfica de control \hat{p} se realizará por lo tanto dibujando en la gráfica tres líneas horizontales:

- la línea objetivo,
- el límite de control superior en $p + 3\frac{\sqrt{p(1-p)}}{\sqrt{n}}$,
- el límite de control inferior en $p - 3\frac{\sqrt{p(1-p)}}{\sqrt{n}}$, en nuestro caso.

V.7.3. Otra señal de alarma

Existen otras posibles señales de alarma para decidir si un proceso está fuera de control. Una de ellas corresponde a dibujar la línea objetivo y concluir que la máquina está mal ajustada si se observan nueve puntos consecutivos por debajo(o por encima) de la línea objetivo. La probabilidad de falsa alarma, es decir concluir erróneamente que el proceso está fuera de control es del orden de 2 entre 1000.



2013: Año Internacional de la Estadística. ¿Sabías qué...?

El reconocimiento del habla implementado en nuestros smartphones manda por internet la señal digitalizada de nuestra voz a servidores donde la pasan por un modelo estadístico del lenguaje, basado en millones de muestras de fragmentos de voz, que devuelve el texto escrito más probable asociado a la señal de voz analizada.

Fuente: Entrevista ITConversations el 19-07-2011, a Mike Cohen, Responsable (2004-2012) de tecnología del habla de Google.

statistics

Introducción a la teoría de la estimación

VI.1. Introducción

Consideramos un experimento aleatorio para el cual estamos dispuestos a escoger un modelo, posiblemente con uno o varios parámetros que tendremos que ajustar. Por ejemplo, queremos realizar una medición con un aparato, la variable que nos interesa es X “valor proporcionado por el aparato”, pensamos que la distribución de los valores que puede tomar X se puede aproximar por una distribución Normal. Nos falta “ajustar” los valores de la media y de la varianza de esta distribución normal, para disponer de un modelo completamente especificado que nos permitirá realizar cálculos de probabilidad, predicciones etc... Para ajustar los parámetros que nos faltan, repetiremos el experimento varias veces y sacaremos información - se dice inferir - sobre estos parámetros a partir de los valores obtenidos de X . El primer tipo de información que podemos intentar sacar es acerca de su valor. Estimar un parámetro consiste en obtener una aproximación de su valor en base a los datos de la variable correspondientes a varias realizaciones del experimento. Recordar que vimos en el tema anterior que los datos provenientes de varias realizaciones del experimento constituyen una **muestra** de la distribución de X .

VI.2. Estimación puntual

VI.2.1. Definición

Consideramos un experimento aleatorio, con una v.a X , y un modelo para la distribución de X . Este modelo incluye parámetros desconocidos. Disponemos de una muestra de la distribución de X .

Definición VI.2.1 *Cualquier estadístico (es decir, cualquier función de las observaciones de la muestra) diseñado para aproximar el valor de un parámetro θ del modelo, se llama estimador puntual del parámetro θ .*

En la tabla siguiente se presentan algunos parámetros usuales y los estimadores asociados:

θ	Estimador
μ	\bar{X} , media muestral
σ^2	S^2 , varianza muestral
p	\hat{p} , proporción muestral

Un aspecto fundamental de un estimador es que es una variable aleatoria: su valor concreto depende de la muestra escogida. Utilizaremos los resultados del tema anterior sobre distribuciones muestrales para deducir propiedades de las distribuciones de los estimadores más usados.

VI.2.2. Propiedades deseables para un estimador

VI.2.2.1. Estimador insesgado

Una primera propiedad deseable para un estimador es que el centro de la distribución de los valores que puede tomar coincida con el valor del parámetro que queremos aproximar. Si éste es el caso, decimos que el estimador es insesgado. Así, si $\hat{\theta}$ es un estimador del parámetro θ , decimos que $\hat{\theta}$ es un estimador insesgado de θ si

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Comprobemos si los estimadores más usados son insesgados:

- La media muestral \bar{X} : hemos visto en el tema 5 que, sea cual sea la distribución de X , se cumple que $\mathbb{E}[\bar{X}] = \mu_X$. Deducimos que \bar{X} es un estimador insesgado de μ_X .
- La varianza muestral S^2 . Tenemos que

$$S^2 = \frac{n}{n-1} [\overline{X^2} - (\bar{X})^2].$$

Por lo tanto,

$$\mathbb{E}[S^2] = \frac{n}{n-1} [\mathbb{E}[\overline{X^2}] - \mathbb{E}[(\bar{X})^2]].$$

Necesitamos calcular por una parte $\mathbb{E}[\overline{X^2}]$ y por otra parte $\mathbb{E}[(\bar{X})^2]$. Al ser $\overline{X^2}$ la media muestral de la variable X^2 , sabemos por el tema 5 que $\mathbb{E}[\overline{X^2}] = \mathbb{E}[X^2] = \text{var}(X) + \mu_X^2$. Por otra parte, $\mathbb{E}[(\bar{X})^2] = \text{var}(\bar{X}) + (\mathbb{E}[\bar{X}])^2 = \frac{\sigma^2}{n} + \mu_X^2$. Deducimos que

$$\mathbb{E}[S^2] = \frac{n}{n-1} \left[\sigma^2 - \frac{\sigma^2}{n} \right] = \sigma^2.$$

Hemos por lo tanto comprobado que la varianza muestral es un estimador insesgado de la varianza. De hecho, este resultado constituye la justificación de que la varianza muestral se defina con el factor $n/(n-1)$, para que el estimador resulte insesgado.

- Proporción muestral \hat{p} : en el tema 5, hemos obtenido la caracterización de \hat{p} como N/n donde N es el número de elementos en la muestra con la característica de interés, y hemos visto que $N \sim \mathcal{B}(n, p)$. Deducimos que

$$\mathbb{E}[\hat{p}] = \frac{\mathbb{E}[N]}{n} = \frac{np}{n} = p.$$

En este caso también, la proporción muestral resulta ser un estimador insesgado de la proporción.

VI.2.2.2. Estimador consistente

Si un estimador es insesgado, nos interesa que la dispersión de los valores que puede tomar sea la más pequeña posible, para que la precisión de la estimación sea la mayor posible. Por consiguiente, una buena propiedad adicional de un estimador insesgado es que su varianza tienda a cero si el número de observaciones n crece hacia infinito. En este caso, se dice que el estimador es consistente.

De la misma manera que en el apartado anterior, podemos deducir, utilizando los resultados del tema 5, que

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \text{var}(\hat{p}) = \text{var}\left(\frac{N}{n}\right) = \frac{1}{n^2}\text{var}(N) = \frac{p(1-p)}{n}.$$

Es fácil comprobar que, en efecto tanto $\text{var}(\bar{X})$ como $\text{var}(\hat{p})$ tienden a cero si n tiende a infinito, es decir que son dos estimadores consistentes.

VI.2.3. Métodos de construcción de estimadores

En los ejemplos de las secciones anteriores, los estimadores propuestos están basados en estadísticos naturales para los parámetros de interés: la media muestral para estimar la media, la proporción muestral para estimar la proporción, etc... En modelos más sofisticados es útil disponer de métodos generales de construcción de estimadores razonables.

VI.2.3.1. Estimadores de momentos

Es el método más antiguo de construcción de estimadores y se debe a Karl Pearson a principios del siglo XX.

Consideremos una v.a. X y un modelo para la distribución de sus valores, que consiste en la especificación de $x \mapsto f_X(x; \theta)$, siendo f_X la función puntual de probabilidad, o la función de densidad según si X es una variable discreta o continua.

El parámetro θ es posiblemente multidimensional, llamamos p su dimensión, es decir que p es el número de parámetros desconocidos en el modelo. Para un entero k , consideramos el momento μ_k de orden k de la distribución de X :

$$\mu_k = \mathbb{E}[X^k].$$

Cabe destacar que la expresión de μ_k depende del parámetro θ . Para enfatizar esta dependencia, escribiremos $\mu_k(\theta)$ para denotar el momento de orden k del modelo descrito por $x \mapsto f_X(x; \theta)$. De manera paralela, definimos el momento *muestral* de orden k :

$$m_k = \overline{X^k} = \frac{X_1^k + \dots + X_n^k}{n}.$$

Para un parámetro de dimensión p , los estimadores de los momentos se obtienen igualando los p primeros momentos del modelo para la distribución de X con sus

equivalentes muestrales:

$$\begin{aligned}\mu_1(\theta) &= \bar{X}, \\ \mu_2(\theta) &= \overline{X^2}, \\ &\vdots = \vdots, \\ \mu_k(\theta) &= \overline{X^k}.\end{aligned}$$

Calculemos para ilustrar el método los estimadores de momentos en los modelos siguientes:

- $X \sim \mathcal{N}(\mu, \sigma^2)$, donde $\theta = (\mu, \sigma^2)$. Necesitamos igualar los dos primeros momentos con sus equivalentes muestrales. Los dos primeros momentos de la distribución $\mathcal{N}(\mu, \sigma^2)$ son

$$\begin{aligned}\mu_1(\theta) &= \mu \\ \mu_2(\theta) &= \mathbb{E}[X^2] = \text{Var}(X) + (\mathbb{E}[X])^2 = \sigma^2 + \mu^2.\end{aligned}$$

Deducimos que los estimadores de los momentos son solución del sistema:

$$\begin{aligned}\mu &= \bar{X} \\ \sigma^2 + \mu^2 &= \overline{X^2},\end{aligned}$$

es decir

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \overline{X^2} - (\bar{X})^2.$$

- Modelo de Bernoulli: $X \sim \text{Bernoulli}(p)$, donde desconocemos p . Sólo necesitamos igualar el primer momento con su equivalente muestral, obtenemos

$$\hat{p} = \bar{X},$$

puesto que X_1, \dots, X_n sólo pueden tomar el valor 1 o el valor 0, su media es igual a la proporción muestral de 1. El estimador de momentos de la proporción p en un modelo de Bernoulli es la proporción muestral.

VI.2.3.2. Método de máxima verosimilitud

El método de máxima verosimilitud es sin dudas el método más utilizado de construcción de un estimador puntual.

a). Verosimilitud Sea X una v.a. con distribución especificada por $x \mapsto f_X(x; \theta)$, donde θ es el vector de parámetros, de dimensión p . Repetimos el experimento n veces y consideramos la muestra aleatoria simple de la distribución de X : (X_1, \dots, X_n) . La distribución de la v.a. n -dimensional (X_1, \dots, X_n) está descrita a través de la relación

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = f_{X_1}(x_1, \theta) \dots f_{X_n}(x_n, \theta),$$

puesto que las v.a. X_1, \dots, X_n son independientes. En esta última igualdad, f representa o bien la función puntual de probabilidad o bien la función de densidad.

Para un valor concreto de (X_1, \dots, X_n) , que denotamos por (x_1, \dots, x_n) , consideramos la función de θ :

$$L_n : \begin{cases} \mathbb{R}^p \rightarrow \mathbb{R} \\ \theta \mapsto L_n(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta). \end{cases}$$

La función L_n asocia a cada valor de θ el valor de la densidad (o de la función puntual de probabilidad) de las observaciones (X_1, \dots, X_n) evaluada en (x_1, \dots, x_n) , los valores concretos observados.

Ejemplo. Consideremos la tirada de una moneda y asociamos la v.a. X que valga 1 si sale cara y 0 si sale cruz. Utilizamos un modelo de Bernoulli de parámetro p entre 0 y 1. Tiramos 10 veces la moneda y obtenemos la secuencia de valores siguiente: 0, 0, 1, 0, 1, 1, 1, 1, 1, 1. La verosimilitud asocia a cada valor posible de p , la cantidad

$$\mathbb{P}(X_1 = 0; X_2 = 0; X_3 = 1; X_4 = 0; X_5 = 1; X_6 = 1; X_7 = 1; X_8 = 1; X_9 = 1; X_{10} = 1).$$

Deducimos que $L_n(p) = (1-p)(1-p)p(1-p)(1-p)^6 = (1-p)^3 \cdot p^7$. Se representa la gráfica de la función $L_n(p)$ en la Figura VI.1

La verosimilitud nos indica para qué valor de p , la probabilidad de haber observado la secuencia 0, 0, 1, 0, 1, 1, 1, 1, 1, 1 es la más alta.

b). Estimador de máxima verosimilitud

Definición VI.2.2 *Dados (x_1, \dots, x_n) los valores observados de una muestra, consideramos la verosimilitud $\theta \mapsto L_n(\theta)$.*

El estimador de máxima verosimilitud $\hat{\theta}$ de θ es cualquier valor de θ que maximiza $\theta \mapsto L_n(\theta)$,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L_n(\theta).$$

La maximización se realiza sobre todos los valores admisibles para el parámetro θ .

Ejemplo. Consideramos $X \sim \text{Bernoulli}(p)$. Observamos x_1, \dots, x_n una realización de la muestra aleatoria simple (X_1, \dots, X_n) . Puesto que si $x = 0, 1$, $f_X(x) = \mathbb{P}(X = x) = p^x \cdot (1-p)^{(1-x)}$, la verosimilitud es

$$L_n(p) = p^{x_1} \cdot (1-p)^{(1-x_1)} \dots p^{x_n} \cdot (1-p)^{(1-x_n)} = p^{\sum x_i} (1-p)^{n - \sum x_i}.$$

Los candidatos a alcanzar el máximo se obtienen derivando la verosimilitud, o de manera equivalente y más sencilla, su logaritmo (llamado log-verosimilitud):

$$\frac{d \log L_n(p)}{dp} = (n - \sum x_i) \left(-\frac{1}{1-p} \right) + \frac{\sum x_i}{p} = 0.$$

Despejamos p y encontramos $\hat{p} = (\sum x_i)/n$. Comprobamos además que la derivada segunda de L_n es negativa, lo que implica que \hat{p} es efectivamente un máximo global. Deducimos que el estimador de máxima verosimilitud de p es la proporción muestral.

Ejemplo. Consideramos $X \sim \mathcal{N}(\mu, \sigma^2)$. Observamos x_1, \dots, x_n una realización de la muestra aleatoria simple (X_1, \dots, X_n) . La verosimilitud se obtiene a partir de la expresión de la densidad de X :

$$L_n(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}.$$

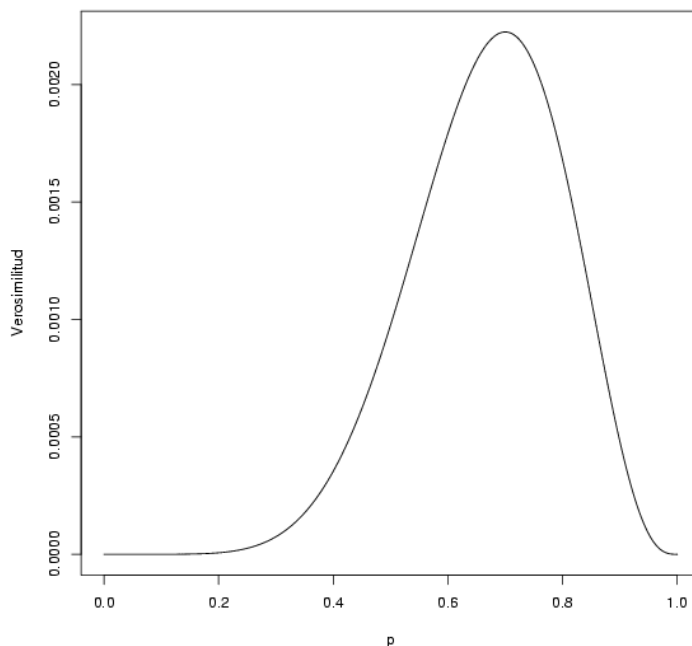


Figura VI.1: Verosimilitud correspondiente al ejemplo de 10 tiradas de una moneda.

La log-verosimilitud es

$$\log L_n(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}.$$

Para encontrar el máximo, calculamos las derivadas parciales de $\log L_n$ respecto de μ y σ^2 :

$$\begin{aligned} \frac{\partial}{\partial \mu} \log L_n(\theta) &= \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} \\ \frac{\partial}{\partial \sigma^2} \log L_n(\theta) &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2}. \end{aligned}$$

Resolvemos $\frac{\partial}{\partial \mu} L_n = 0$ y $\frac{\partial}{\partial \sigma^2} L_n = 0$, y encontramos que los dos candidatos a máximo son

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}, \quad \widehat{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{n} = \frac{n}{n-1} s^2.$$

Para comprobar que son efectivamente máximos globales, podemos fijarnos en la expresión de la log-verosimilitud:

$$\log L_n(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}.$$

Sea cual sea el valor de σ^2 , la función $\mu \mapsto \log L_n(\mu, \sigma^2)$ alcanza su máximo cuando $\sum_{i=1}^n (x_i - \mu)$ es mínimo, es decir cuando $\mu = (\sum_{i=1}^n x_i)/n$. El máximo de $(\mu, \sigma^2) \mapsto \log L_n(\mu, \sigma^2)$ corresponderá por lo tanto al máximo de la función $\sigma^2 \mapsto \log L_n(\hat{\mu}, \sigma^2)$. Es fácil comprobar que $\sigma^2 \mapsto \log L_n(\hat{\mu}, \sigma^2)$ alcanza su máximo en $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{n} = \frac{n}{n-1} s^2$.

Los estimadores de máxima verosimilitud de μ y σ^2 son por lo tanto la media muestral y la llamada varianza muestral sesgada $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{n} = \frac{n}{n-1} s^2$. En un apartado anterior hemos visto como la varianza muestral s^2 es un estimador insesgado, por lo tanto $\mathbb{E}[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2$. Es un ejemplo en él que el método de máxima verosimilitud proporciona un estimador sesgado.

VI.3. Estimación por intervalos

No queremos limitarnos a dar un valor para aproximar un parámetro sino proporcionar también una medida del error que pensamos cometer. Para ello, calcularemos un intervalo en él que pensamos que se encuentra el parámetro.

VI.3.1. Idea básica

Supongamos que queremos estimar la media μ de una v.a. X cuya distribución es Normal con una desviación típica igual a 2 unidades, es decir $X \sim \mathcal{N}(\mu, 4)$. Para ello, extraigo una muestra de tamaño 4, y estimo μ por el valor de \bar{X} . Por el tema 5, ver V.3.2.1, sabemos que la distribución de \bar{X} es $\mathcal{N}(\mu, \sigma^2/n)$ es decir $\mathcal{N}(\mu, 1)$. Por la propiedad de la distribución Normal, ver b), deducimos que el 95 % de las muestras proporcionan un valor de \bar{X} que se encuentra a menos de 2 unidades de la media μ .

Invertamos ahora la situación: sé donde está \bar{X} , ¿donde está μ ? Por la misma regla, se encuentra, para el 95 % de las muestras, a menos de 2 unidades de \bar{X} , es decir que μ se encuentra en el intervalo $[\bar{X} - 2, \bar{X} + 2]$. Dicho de otra manera, para el 95 % de las muestras, el intervalo aleatorio $[\bar{X} - 2, \bar{X} + 2]$ captura el valor del parámetro μ .

VI.3.2. Intervalo de confianza para la media μ de una distribución Normal con varianza conocida

VI.3.2.1. Construcción

Consideramos la variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Suponemos que conocemos el valor de σ^2 . La construcción del intervalo de confianza para la media μ se realiza siguiendo los siguientes pasos.

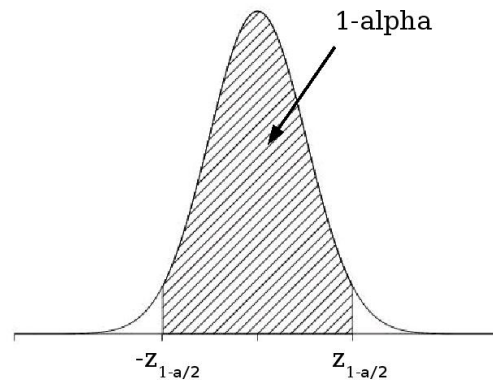
- Nos fijamos el llamado “nivel de riesgo”, α un número entre 0 y 1. La cantidad $1 - \alpha$ expresada en porcentaje se llama **nivel de confianza**. Los valores más utilizados de α son 0,1, 0,05, y 0,01, lo que corresponde con niveles de confianza del 90 % ,95 % y 99 % respectivamente.

- Escogemos el estadístico \bar{X} para estimar μ . Su distribución en su forma tipificada es

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Para $0 \leq u \leq 1$, utilizamos la notación z_u para denotar el cuantil u de la distribución Normal estándar, es decir el valor que cumple $\mathbb{P}(Z \leq z_u) = u$, o dicho de otra manera, el valor que deja a su izquierda un área igual a u debajo de la curva de la densidad Normal estándar. En particular usaremos de manera repetida los cuantiles siguientes: $z_{0,95}$, $z_{0,975}$ y $z_{0,995}$. Para conocer sus valores, podemos buscar en la tabla de la Normal estándar, los valores 0,95, 0,975 y 0,995 en la columna de las probabilidades $\phi(t)$ y apuntar los valores correspondientes de t . Encontramos $z_{0,95} = 1,64$, $z_{0,975} = 1,96$ y $z_{0,995} = 2,56$.

- Dibujo en la densidad del estadístico $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, una región central que represente el $100(1 - \alpha)\%$ del área total, tal como viene ilustrado en la figura siguiente



- Deducimos

$$\mathbb{P}(-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}) = 1 - \alpha.$$

Despejamos μ en las desigualdades

$$\begin{aligned} \mathbb{P}(-z_{1-\alpha/2}\sigma/\sqrt{n} \leq \bar{X} - \mu \leq z_{1-\alpha/2}\sigma/\sqrt{n}) &= 1 - \alpha \\ \Leftrightarrow \mathbb{P}(-\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq -\mu \leq -\bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}) &= 1 - \alpha \\ \Leftrightarrow \mathbb{P}(\bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n} \geq \mu \geq \bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}) &= 1 - \alpha \\ \Leftrightarrow \mathbb{P}(\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}) &= 1 - \alpha \end{aligned}$$

- El intervalo de confianza al $100(1 - \alpha)\%$ para μ es

$$\mu \in [\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}].$$

Se escribe también de otra manera equivalente:

$$\mu = \bar{X} \pm z_{1-\alpha/2}\sigma/\sqrt{n},$$

el término $z_{1-\alpha/2}\sigma/\sqrt{n}$ se llama término de error.

VI.3.2.2. Interpretación

El intervalo $[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}]$ es un intervalo aleatorio, puesto que sus extremos dependen de la muestra escogida. Por su construcción, sabemos que este intervalo aleatorio tiene una probabilidad de $100(1 - \alpha)\%$ de capturar el valor de μ . Es decir que, al extraer una muestra, tengo una probabilidad igual a $1 - \alpha$ de que el intervalo que calcularé efectivamente capture el valor μ que busco. También tengo una probabilidad α de que, al afirmar que μ se encuentra en $[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}; \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}]$, me equivoque. Sin embargo, esta probabilidad α , el riesgo de equivocarme, se fija en general bastante pequeño, por ejemplo $\alpha = 0,05$.

Para ilustrar esta interpretación, he simulado 20 veces el proceso de extraer una muestra de tamaño 4 de una distribución $X \sim \mathcal{N}(\mu_X, 1)$. He representado en la Figura VI.2 en el eje Ox el número de la muestra y en el eje Oy el intervalo de confianza asociado. Además una línea horizontal representa el valor de μ que se pretende estimar, en este caso $\mu = 2$. La gran mayoría de los intervalos capturan el valor correcto de μ , pero hay un intervalo, el correspondiente a la muestra número 13 que no lo hace: este intervalo es erróneo, y esta muestra forma parte del 5% de las muestras “malas”, es decir las que proporcionan intervalos equivocados.

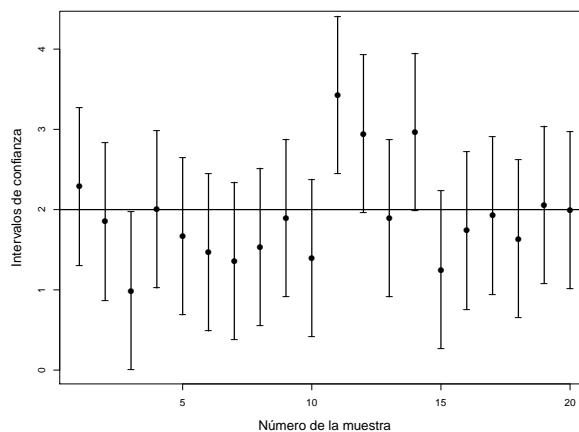


Figura VI.2: Los intervalos de confianza al 95% correspondientes a 20 muestras de tamaño 4. La media que se busca estimar es $\mu = 2$.

VI.3.2.3. Ejemplo

Supongamos que queremos estimar la longitud media de un artículo producido por una máquina. Por experiencia, sabemos que es razonable modelizar la distribución de los valores de la longitud de los artículos producidos por una distribución Normal con media μ y desviación típica igual a 0.05. Para estimar μ extraemos una muestra de 5 artículos y construimos un intervalo de confianza al 90%. Supongamos que los datos que se obtienen son los siguientes:

20.1, 20.05, 20.01, 19.95, 19.99.

El intervalo de confianza es $\mu \in [\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}]$. Necesitamos \bar{X} , es fácil comprobar que $\bar{X} = 20,02$, por otra parte, al haber escogido 90% de

confianza, fijamos $\alpha = 0,1$. Deducimos de la tabla Normal que $z_{1-\alpha/2} = z_{0,95} = 1,64$. Concluimos que el intervalo buscado será

$$\left[20,02 - 1,64 \frac{0,05}{\sqrt{5}}, 20,02 + 1,64 \frac{0,05}{\sqrt{5}}\right],$$

es decir $\mu \in [19,98, 20,06]$, o de forma equivalente $\mu = 20,02 \pm 0,04$.

VI.3.3. Comentarios importantes

a) La construcción del intervalo de confianza está basada en la hipótesis de que la distribución de la v.a. X es Normal, puesto que utilizamos

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Si la distribución de X no es Normal, el intervalo no es válido, es decir que no podemos garantizar que la confianza especificada sea cierta. Sin embargo, en el caso en que la muestra es grande, podemos recurrir al Teorema Central del Límite, ver V.3.1, y sabemos que

$$\text{aproximadamente, } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

lo que posibilita que los intervalos sean aproximadamente válidos: la confianza especificada no será exacta pero casi...

¿A partir de cuantas observaciones consideramos una muestra como grande? No hay respuesta universal, depende mucho de lo alejado que está la distribución de X de una distribución Normal. En general, se suele considerar en práctica que $n \geq 30$ es suficiente para que los intervalos construidos sean aproximadamente válidos.

b) Factores que afectan a la precisión de la estimación.

Recordar que en la estimación por un intervalo, el margen de error es $\pm z_{1-\alpha/2} \sigma / \sqrt{n}$. Deducimos en particular que

- cuanto mayor sea n , más precisa será la estimación, es decir que más pequeño será el intervalo de confianza.
- cuanto menor sea σ , mayor precisión en la estimación.
- cuanto mayor sea la confianza, peor será la precisión de la estimación: si queremos garantizar con gran confianza que el intervalo proporcionado captura μ , no hay más remedio que proporcionar un intervalo grande...

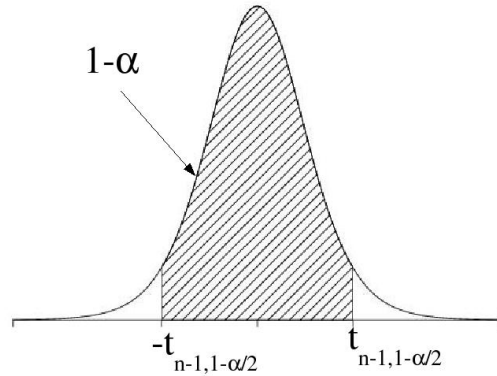
c) La hipótesis de que σ es conocida no es realista: en general también hay que estimarla a partir de la muestra. La distribución del estadístico que resulta de sustituir σ por S , la desviación típica muestral, $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ es una t de Student con $n - 1$ grados de libertad. Podemos repetir los pasos de construcción del intervalo de confianza para μ basándonos en el estadístico $\frac{\bar{X} - \mu}{S/\sqrt{n}}$:

- Nos fijamos el “nivel de riesgo”, α .

- Escogemos el estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

- Dibujo en la densidad del estadístico T una región central que represente el $100(1-\alpha)\%$ del área total, tal como viene ilustrado en la figura siguiente



- Deducimos

$$\mathbb{P}(-t_{n-1, 1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1, 1-\alpha/2}) = 1 - \alpha,$$

donde hemos utilizado la notación $t_{n-1, 1-\alpha/2}$ para denotar el cuantil $1 - \alpha/2$ de la distribución t_{n-1} , es decir el punto que deja un área igual a $1 - \alpha/2$ a su izquierda. Los valores de los cuantiles más usados de la distribución t están recogidos en una tabla en el apéndice de este capítulo. Despejamos μ en las desigualdades y obtenemos

$$\mathbb{P}(\bar{X} - t_{n-1, 1-\alpha/2}S/\sqrt{n} \leq \mu \leq \bar{X} + t_{n-1, 1-\alpha/2}S/\sqrt{n}) = 1 - \alpha.$$

- El intervalo de confianza al $100(1 - \alpha)\%$ para μ es

$$\mu \in [\bar{X} - t_{n-1, 1-\alpha/2}S/\sqrt{n}; \bar{X} + t_{n-1, 1-\alpha/2}S/\sqrt{n}].$$

Se escribe también

$$\mu = \bar{X} \pm t_{n-1, 1-\alpha/2}S/\sqrt{n},$$

el término $t_{n-1, 1-\alpha/2}S/\sqrt{n}$ es el término de error.

VI.3.4. Determinación del tamaño muestral

VI.3.4.1. Planteamiento

Si estoy en condiciones de diseñar el experimento que quiero realizar para estimar la media μ , puedo intentar decidir del número de observaciones en la muestra que serán necesarias para garantizar, con una confianza dada, que el margen de error sea menor que una cantidad prefijada. Es decir, que me fijo una cantidad *max*, y me pregunto cuál deberá de ser el valor de n para que

$$z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \text{max}.$$

Es fácil obtener n despejándolo de la desigualdad.

VI.3.4.2. Ejemplo

La medición de la conductividad de un material sigue una distribución que modelizamos por una Normal con desviación típica $\sigma = 0,5$. Quiero construir un intervalo de confianza al 95 % para el valor promedio proporcionado de la conductividad pero quiero que el error cometido sea menor de 0.3. ¿cuántas veces deberé repetir la medición?

Busco n tal que $z_{1-\alpha/2}\sigma/\sqrt{n} \leq 0,3$, sabiendo que $\sigma = 0,5$, y $\alpha = 0,05$. Obtengo

$$1,96 \frac{0,5}{\sqrt{n}} \leq 0,3,$$

es decir que

$$n \geq \left(\frac{1,96 \cdot 0,5}{0,3} \right)^2 \simeq 10,67.$$

Habrà por lo tanto que realizar 11 mediciones.

Cuantiles de la distribución t de Student

Valores de los cuantiles de la distribución t de Student con k grados de libertad: para un $0 \leq p \leq 1$, el valor $t_{k,p}$ satisface $\mathbb{P}(t \leq t_{k,p}) = p$.

k	$t_{k,0,995}$	$t_{k,0,99}$	$t_{k,0,975}$	$t_{k,0,95}$	$t_{k,0,90}$	$t_{k,0,80}$	$t_{k,0,70}$	$t_{k,0,60}$	$t_{k,0,50}$
1	63,657	31,821	12,706	6,314	3,078	1,376	0,727	0,325	0,158
2	9,925	6,965	4,303	2,92	1,886	1,061	0,617	0,289	0,142
3	5,841	4,541	3,182	2,353	1,638	0,978	0,584	0,277	0,137
4	4,604	3,747	2,776	2,132	1,533	0,941	0,569	0,271	0,134
5	4,032	3,365	2,571	2,015	1,476	0,92	0,559	0,267	0,132
6	3,707	3,143	2,447	1,943	1,44	0,906	0,553	0,265	0,131
7	3,499	2,998	2,365	1,895	1,415	0,896	0,549	0,263	0,13
8	3,355	2,896	2,306	1,86	1,397	0,889	0,546	0,262	0,13
9	3,25	2,821	2,262	1,833	1,383	0,883	0,543	0,261	0,129
10	3,169	2,764	2,228	1,812	1,372	0,879	0,542	0,26	0,129
11	3,106	2,718	2,201	1,796	1,363	0,876	0,54	0,26	0,129
12	3,055	2,681	2,179	1,782	1,356	0,873	0,539	0,259	0,128
13	3,012	2,65	2,16	1,771	1,35	0,87	0,538	0,259	0,128
14	2,977	2,624	2,145	1,761	1,345	0,868	0,537	0,258	0,128
15	2,947	2,602	2,131	1,753	1,341	0,866	0,536	0,258	0,128
16	2,921	2,583	2,12	1,746	1,337	0,865	0,535	0,258	0,128
17	2,898	2,567	2,11	1,74	1,333	0,863	0,534	0,257	0,128
18	2,878	2,552	2,101	1,734	1,33	0,862	0,534	0,257	0,127
19	2,861	2,539	2,093	1,729	1,328	0,861	0,533	0,257	0,127
20	2,845	2,528	2,086	1,725	1,325	0,86	0,533	0,257	0,127
21	2,831	2,518	2,08	1,721	1,323	0,859	0,532	0,257	0,127
22	2,819	2,508	2,074	1,717	1,321	0,858	0,532	0,256	0,127
23	2,807	2,5	2,069	1,714	1,319	0,858	0,532	0,256	0,127
24	2,797	2,492	2,064	1,711	1,318	0,857	0,531	0,256	0,127
25	2,787	2,485	2,06	1,708	1,316	0,856	0,531	0,256	0,127
26	2,779	2,479	2,056	1,706	1,315	0,856	0,531	0,256	0,127
27	2,771	2,473	2,052	1,703	1,314	0,855	0,531	0,256	0,127
28	2,763	2,467	2,048	1,701	1,313	0,855	0,53	0,256	0,127
29	2,756	2,462	2,045	1,699	1,311	0,854	0,53	0,256	0,127
30	2,75	2,457	2,042	1,697	1,31	0,854	0,53	0,256	0,127
40	2,704	2,423	2,021	1,684	1,303	0,851	0,529	0,255	0,126
60	2,66	2,39	2	1,671	1,296	0,848	0,527	0,254	0,126
120	2,617	2,358	1,98	1,658	1,289	0,845	0,526	0,254	0,126
>120	2,576	2,326	1,960	1,645	1,282	0,842	0,524	0,253	0,126



2013: Año Internacional de la Estadística. ¿Sabías qué...?

Con el estudio de la actividad de los genes en las secuencias de ADN en el núcleo de nuestras células está en juego el anticipar y prevenir el desarrollo de determinadas enfermedades. La estadística, combinada con la informática, juega un papel fundamental en identificar, a partir de datos de experimentos de microarrays, la diferencia de actividad génica entre personas sanas y personas con una determinada enfermedad.

Fuente: Rivas-López M. J, Sánchez-Santos J.M & De las Rivas, J. “Estructura y análisis de microarrays”, BEIO (2005), 22, 10-15

statistics

إحصائيات

statistică

statistikë

आँकड़े

統計

statistiek

סטטיסטיקה

statistieke

statistika

số liệu thống kê

estatística

statistiko

statistikat

statistic

statistika

Introducción a los contrastes de hipótesis

VII.1. Introducción

En el tema anterior, hemos aprendido cómo estimar, es decir, aproximar el valor de un parámetro basándonos en las observaciones de una muestra. Hay situaciones en las que más que conocer el valor concreto del parámetro, queremos tomar una decisión acerca de éste. Formularemos una hipótesis sobre el valor del parámetro y la contrastaremos con los datos de la muestra para comprobar si éstos la apoyan o la desmienten.

Para ilustrar los conceptos relacionados con los contrastes de hipótesis, retomamos el ejemplo visto al final del tema 5 cuando describimos la gráfica de control \bar{X} : una empresa controla la concentración de CaCO_3 en su producto. El valor ideal de esta concentración es 55. Si llamamos X la concentración de CaCO_3 medida en un envase, sabemos que es razonable modelizar la distribución de X por una distribución Normal de media μ y desviación típica 8. En el tema 5, vimos cómo la empresa puede realizar un control de la calidad de su producción gracias a una gráfica \bar{X} : cada hora toma una muestra de 4 envases, mide la concentración de CaCO_3 en cada caso y calcula su media. Basándose en este valor decide si el proceso de producción está en condiciones de correcto funcionamiento, es decir si $\mu = 55$.

Para decidir si $\mu = 55$ o $\mu \neq 55$, la empresa se fija una regla: si $\bar{X} > 60,4$ ó $\bar{X} < 49,6$, decide que $\mu \neq 55$ y para la producción para ajustar el proceso de fabricación.

Este ejemplo contiene todos los ingredientes del contraste de hipótesis y pasamos a describirlos en un contexto más general.

VII.2. Planteamiento general

VII.2.1. Hipótesis estadística

Una hipótesis estadística es una proposición acerca del valor de un parámetro en el modelo considerado. La formulación de un contraste de hipótesis pasa siempre por el planteamiento de dos hipótesis:

$$\begin{cases} H_0 : \mu = 55, & \text{Hipótesis nula} \\ H_1 : \mu \neq 55, & \text{Hipótesis alternativa} \end{cases}$$

Habrán casos en los que nos interesará decidir si el parámetro es mayor (o menor) que un valor dado, entonces cambiaremos la formulación de la hipótesis alternativa, pero seguiremos, para simplificar, considerando la igualdad en la hipótesis nula. Por ejemplo si queremos contrastar si μ es mayor que 55, plantearemos el contraste:

$$\begin{cases} H_0 : \mu = 55, \\ H_1 : \mu > 55, \end{cases}$$

mientras que si queremos decidir si μ es menor que 55, plantearemos

$$\begin{cases} H_0 : \mu = 55, \\ H_1 : \mu < 55, \end{cases}$$

De los tres contrastes, el primero se llama contraste bilateral, puesto que la hipótesis alternativa comprende tanto valores mayores como valores menores que 55, mientras que los dos últimos se llaman contrastes unilaterales.

VII.2.2. Regla de decisión

Basándonos en un estadístico $T(X_1, \dots, X_n)$, es decir en una función de las observaciones, determinaremos una región de rechazo R . Para mi muestra calcularé el valor concreto de $T(X_1, \dots, X_n)$; si este valor pertenece a R , rechazaremos H_0 , es decir afirmaremos que los datos apoyan la hipótesis alternativa H_1 .

En cambio si el valor de $T(X_1, \dots, X_n)$ no pertenece a R , aceptaremos H_0 , diremos que los datos no presentan argumentos en contra de la hipótesis nula.

En el ejemplo de los monitores de ordenador, la regla de decisión que se había fijado la empresa es: basándose en el estadístico $T(X_1, \dots, X_n) = \bar{X}$, la región de rechazo es $R = \{x < 49,6\} \cup \{x > 60,4\}$.

VII.2.3. Evaluación del error

Al tomar la decisión acerca de la veracidad de H_0 , podemos cometer dos tipos de error:

VII.2.3.1. Error de tipo I

Podemos afirmar que H_0 es falsa, cuando en realidad es cierta, es decir que los datos nos llevan a rechazar H_0 cuando ésta es cierta. Este tipo de error se llama error de tipo I, y, una vez fijada una regla de decisión, la probabilidad de cometerlo se denota por α , (la letra griega "alfa"). Tenemos por lo tanto

$$\alpha = \mathbb{P}_{H_0}(\text{Rechazar } H_0) = \mathbb{P}_{H_0}(T(X_1, \dots, X_n) \in R),$$

donde con la notación \mathbb{P}_{H_0} , nos referimos a la probabilidad suponiendo que H_0 es cierta.

En el ejemplo de la concentración de CaCO_3 , podemos calcular la probabilidad de error de tipo I:

$$\alpha = \mathbb{P}_{H_0}(\text{Rechazar } H_0) = \mathbb{P}_{\mu=55}((\bar{X} < 49,6) \cup (\bar{X} > 60,4)).$$

Pero, precisamente, los límites de control en la gráfica \bar{X} se fijaron para que, si la máquina está bien ajustada, es decir si $\mu = 55$, sólo el 3 por 1000 de las muestras deben llevar a un valor de \bar{X} fuera de los límites. Deducimos que $\alpha = 0,003$.

VII.2.3.2. Error de tipo II

El segundo tipo de error se comete cuando admitimos H_0 cuando en realidad es falsa. Una vez fijada la regla de decisión, la probabilidad de cometer un error de tipo II se denota por β (la letra griega “beta”). Tenemos

$$\beta = \mathbb{P}_{H_1}(\text{Aceptar } H_0) = \mathbb{P}_{H_1}(T(X_1, \dots, X_n) \notin R).$$

El cálculo de β sólo se puede hacer si especificamos un valor concreto de μ en la hipótesis alternativa. Para el ejemplo de la concentración de CaCO_3 , podemos por ejemplo calcular β cuando en realidad $\mu = 65$. Tenemos $\beta = \mathbb{P}_{\mu=65}(49,6 \leq \bar{X} \leq 60,4)$, y sabemos que $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ es decir $\bar{X} \sim \mathcal{N}(\mu, (4)^2)$. Tipificamos \bar{X} para calcular β :

$$\beta = \mathbb{P}_{\mu=65}\left(\frac{49,6 - 65}{4} \leq \frac{\bar{X} - 65}{4} \leq \frac{60,4 - 65}{4}\right) \simeq \phi(-2,3) - \phi(-7,7) \simeq 0,13.$$

VII.2.4. Procedimiento

Para llevar a cabo un contraste de hipótesis, tendremos que

- Formular las hipótesis H_0 y H_1 .
- Fijarnos la probabilidad de error de tipo I, α . Al igual que para los contrastes de hipótesis, los valores de α más comunes son 0.05, 0.01 o 0.1. (95 %, 99 % ó 90 % de confianza respectivamente).
- Escogemos el estadístico de prueba $T(X_1, \dots, X_n)$ basado generalmente en un estimador del parámetro. Describimos su distribución muestral bajo la hipótesis de que H_0 es cierta.
- Determinamos la región de rechazo R de tal manera que la probabilidad de rechazar H_0 cuando ésta es cierta coincida con el valor prefijado de α , es decir

$$\mathbb{P}_{H_0}(T(X_1, \dots, X_n) \in R) = \alpha.$$

- Para nuestra muestra, calculamos el valor concreto del estadístico de prueba $T(X_1, \dots, X_n)$. Si este valor cae en la región R , rechazamos H_0 y afirmamos H_1 , mientras que si no cae en la región R , admitimos H_0 .

VII.3. Contraste de hipótesis para la media μ de una distribución Normal con varianza conocida.

Consideramos una variable X , suponemos que su distribución ha sido modelizada por una Normal con media μ y varianza σ^2 . Suponemos además que conocemos el valor de la varianza σ^2 .

Queremos llevar a cabo un contraste sobre μ , para ello, extraeremos una muestra de tamaño n de la distribución de X : X_1, \dots, X_n .

VII.3.1. Hipótesis bilateral

Para construir el contraste para μ en el caso en que formulamos una hipótesis alternativa bilateral, ver el apartado VII.2.1, seguimos los pasos descritos en la sección VII.2.4:

- Formulamos las hipótesis:

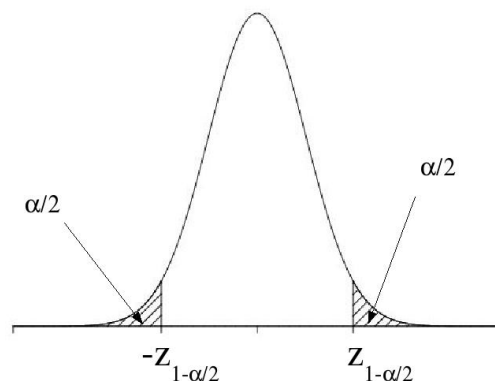
$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0, \end{cases}$$

donde μ_0 representa el valor concreto con el que queremos comparar μ . En el ejemplo de los monitores, μ_0 vale 55.

- Nos fijamos el valor de α .
- El estadístico de prueba es la versión tipificada de \bar{X} , sabemos por el tema 5 que

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{si } H_0 \text{ es cierto.}$$

- Podemos ahora especificar la región de rechazo. La probabilidad de que el estadístico de prueba Z_0 caiga en R cuando H_0 es cierta debe coincidir con el valor de α que nos hemos fijado. Además queremos que Z_0 caiga en R cuando μ es distinto de μ_0 (H_1 cierta), es decir que corresponderá a valores grandes positivos o negativos de Z_0 . Por consiguiente fijamos la región de rechazo de la manera siguiente:



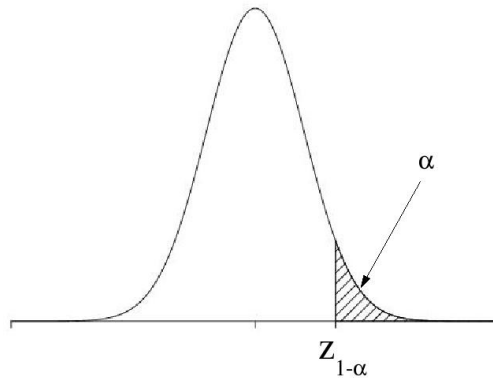
La región R está formada por los valores menores que $-z_{1-\alpha/2}$ o mayores que $z_{1-\alpha/2}$.

- Nos queda calcular, para nuestra muestra, el valor concreto del estadístico de prueba Z_0 . Si pertenece a R , rechazaremos H_0 y afirmaremos H_1 , mientras que si no pertenece a R , admitiremos H_0 .

VII.3.2. Hipótesis unilateral

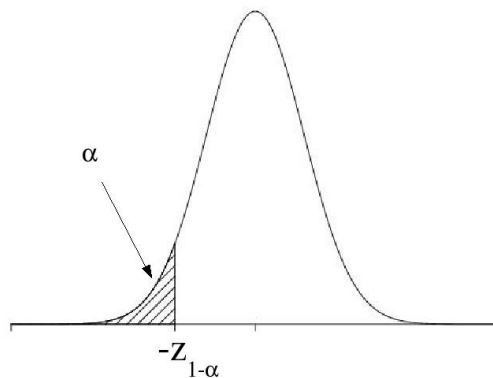
En el caso en que hemos planteado una hipótesis unilateral, los pasos que seguimos son los mismos que en el apartado anterior con la salvedad de la determinación de R :

- Si la hipótesis alternativa es $H_1 : \mu > \mu_0$, la región de rechazo será



es decir que se rechazará H_0 si el valor del estadístico de prueba Z_0 es mayor de $z_{1-\alpha/2}$.

- Si la hipótesis alternativa es $H_1 : \mu < \mu_0$, la región de rechazo será



es decir que se rechazará H_0 si el valor del estadístico de prueba Z_0 es menor de $-z_{1-\alpha/2}$.

VII.3.3. Ejemplos

VII.3.3.1. Hipótesis alternativa bilateral

En un proceso de producción, la longitud de los artículos producidos se modeliza a través de una distribución Normal con media μ . Por experiencia acerca del proceso, se cuantifica su desviación típica en $\sigma = 1$. En condiciones de funcionamiento correcto, se espera que la longitud media de los artículos sea 50mm. Para comprobar la calidad se decide tomar una muestra de 10 artículos que resultan tener una longitud media \bar{X} igual a 51mm. Basándonos en esta muestra, ¿qué podemos decir acerca del funcionamiento del proceso?

La variable que introducimos asociada al experimento “producir una pieza”, es $X =$ “longitud de la pieza producida”. Planteamos las hipótesis

$$\begin{cases} H_0 : \mu = 50, \\ H_1 : \mu \neq 50. \end{cases}$$

Decidimos trabajar al 95% de confianza, que es el nivel estándar de confianza, es decir que nos fijamos $\alpha = 0,05$.

El estadístico de prueba es $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, que sigue una distribución Normal estándar si H_0 es cierta.

Las fronteras de la región de rechazo son $-z_{1-\alpha/2} = -z_{0,975} = -1,96$ y $-z_{1-\alpha/2} = 1,96$.

Basándonos en la muestra, calculamos el valor de Z_0 :

$$Z_0 = \frac{51 - 50}{1/\sqrt{10}} \simeq 3,162.$$

Puesto que Z_0 pertenece a R , rechazamos H_0 y afirmamos al 95% de confianza que el proceso está desajustado.

VII.3.3.2. Hipótesis alternativa unilateral

Creo que un aparato de medición de una señal sobrevalora su valor real. Para comprobarlo pienso realizar 5 mediciones de una señal simple cuyo valor sé es igual a 10000. Considerando que la distribución de los valores medidos se puede modelizar por una Normal con desviación típica igual a 500, llevar a cabo el contraste para comprobar si el valor central de los valores medidos es superior a 10000, si he encontrado un valor promedio de 10300 para las 5 mediciones de la muestra.

El experimento aleatorio es “realizar la medición de la señal”, y la v.a $X =$ “valor proporcionado por el aparato”. Modelizamos X por una distribución $\mathcal{N}(\mu, \sigma^2)$ con $\sigma = 500$.

Planteamos las hipótesis

$$\begin{cases} H_0 : \mu = 10000, \\ H_1 : \mu > 10000, \end{cases}$$

El estadístico es Z_0 , al igual que en el ejemplo anterior, pero la región de rechazo está constituida por los valores mayores que $z_{1-\alpha} = z_{0,95} = 1,64$.

Para mi muestra, el valor de Z_0 es

$$Z_0 = \frac{10300 - 10000}{500/\sqrt{5}} \simeq 1,34.$$

Deducimos que Z_0 no pertenece a R , por lo que no podemos rechazar H_0 : los datos no contradicen H_0 .

VII.4. Concepto de p-valor

En el ejemplo VII.3.3.1, para el contraste

$$\begin{cases} H_0 : \mu = 50, \\ H_1 : \mu \neq 50, \end{cases}$$

Hemos encontrado que el valor del estadístico de prueba era $z_0 = 3,162$, y hemos rechazado al 95 % de confianza la hipótesis nula.

¿Cuál habría sido nuestra decisión si, en lugar de habernos fijado el 95 % de confianza, hubieramos escogido 90 % de confianza?

Por la forma en la que hemos construido la región de rechazo, ésta contiene el 5 % del área total, y la región de aceptación, es decir el complementario de R , contiene el 95 % del área total. Deducimos por lo tanto que la región de rechazo que corresponde al 90 % de confianza es más grande que la región de rechazo que corresponde la 95 % de confianza. Será más fácil rechazar H_0 al 90 % que al 95 % de confianza.

Esto corresponde a un hecho general: si rechazamos H_0 a un nivel de confianza dado, también la rechazaremos para cualquier nivel de confianza menor...

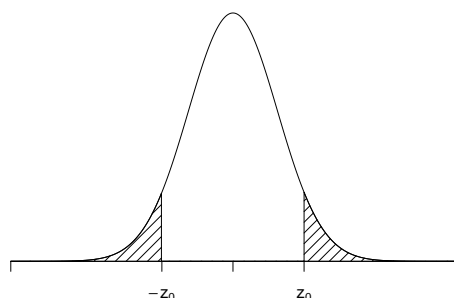
En cambio, si nos preguntamos cuál habría sido nuestra decisión al 99 % de confianza? La región de rechazo mengua, y para saber si seguimos rechazando H_0 necesitamos comprobar si el valor de nuestro estadístico de prueba sigue encontrándose dentro de la nueva región de rechazo. En nuestro ejemplo VII.3.3.1, las fronteras de la región de rechazo al 99 % de confianza son $-z_{1-\alpha/2} = -z_{0,995} = -2,56$ y $z_{0,995} = 2,56$, puesto que Z_0 toma el valor 3.162, también rechazamos H_0 al 99 % de confianza.

Planteados un contraste, y para un valor concreto del estadístico de prueba, podemos preguntarnos cuál habría sido la confianza máxima con la que rechazaríamos H_0 para estos datos. Equivalentemente, podemos calcular el valor más pequeño de α que nos lleve a rechazar H_0 .

Definición VII.4.1 *El valor de α más pequeño que nos lleve a rechazar H_0 se llama el p-valor de la prueba, y lo denotaremos por α_0 .*

Para determinar α_0 , tendremos que considerar la región de rechazo que haga de frontera entre las dos decisiones: rechazar H_0 y aceptar H_0 . Si en la gráfica de la distribución del estadístico Z_0 , empezamos primero por señalar el valor de z_0 obtenido para la muestra, esta región de rechazo se obtendrá al hacer coincidir una de sus fronteras con z_0 : para una región de rechazo más grande (es decir un α más grande) se rechazará H_0 mientras que para una región de rechazo más pequeña (es decir un α más pequeño) tendremos que aceptar H_0 . El valor de α correspondiente a esta región R es α_0 .

Lo ilustramos para el ejemplo en él que $z_0 = 3,162$ en la gráfica siguiente:



Para calcular α_0 , deducimos del dibujo anterior que

$$\alpha_0/2 = \mathbb{P}(Z \geq 3,162),$$

es decir que $\alpha_0 = 2(1 - \phi(3,162)) \simeq 0,00156$.

Deducimos que para el ejemplo, la confianza máxima con la que podríamos haber rechazado es

$$100(1 - \alpha_0) = 100(0,99844) = 99,84\%.$$

Este resultado es coherente con las decisiones que hemos tomado al 95 % y al 99 % de confianza.

Cualquier programa de estadística que permita llevar a cabo un contraste de hipótesis no solicita del usuario que especifique la confianza, sino que directamente le proporciona el p-valor, dejando en sus manos la decisión de rechazar o aceptar H_0 . En general se suele considerar que un p-valor menor de 0.1 nos lleva a rechazar H_0 aunque el estándar correspondería realmente a un p-valor menor que 0,05. Si el p-valor es mayor de 0.2, se admite H_0 . Si el p-valor está comprendido entre 0.1 y 0.2, no permite concluir de manera muy segura y deberíamos intentar coleccionar más datos.

VII.5. Potencia del test

VII.5.1. Definición

Hemos visto que, a la hora de construir un contraste de hipótesis, lo más fácil es controlar la probabilidad de error de tipo I, puesto que la región de rechazo se define para que esta probabilidad coincida con el valor fijado de α . Sin embargo, también es importante saber que, si H_0 es falsa, nuestro contraste lo detectará con bastante probabilidad, es decir que nos llevará a concluir de manera correcta que H_0 es falsa.

Definición VII.5.1 Consideremos H_1 la hipótesis alternativa, y μ_1 un valor concreto de μ incluido en los valores contemplados en H_1 .

La potencia de un test (contraste de hipótesis) contra la alternativa $\mu = \mu_1$, es la probabilidad de rechazar H_0 cuando ésta es falsa y en realidad $\mu = \mu_1$. Es decir

$$Pot(\mu_1) = \mathbb{P}_{\mu=\mu_1}(\text{Rechazar } H_0).$$

Cuanto mayor será la potencia, mejor será el contraste. Se suele considerar suficiente una potencia de al menos 0.8

Recordar que el error de tipo II consiste en aceptar H_0 cuando en realidad ésta es falsa, la relación entre la probabilidad β de error de tipo II y la potencia es por lo tanto

$$\beta = 1 - Pot(\mu_1).$$

VII.5.2. Cálculo de la potencia

Queremos plantear un contraste sobre la media, por ejemplo en su versión bilateral,

$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0, \end{cases},$$

con un cierto nivel de confianza, y planificamos tomar una muestra de n observaciones.

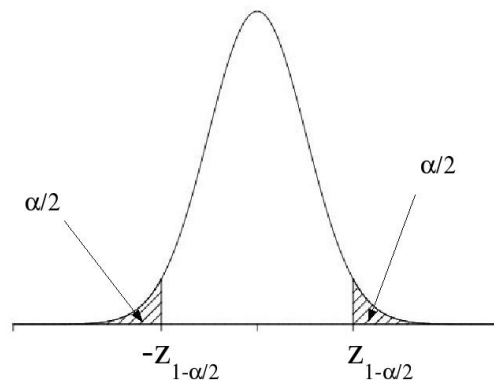
Para calcular la potencia de este contraste contra la alternativa $\mu = \mu_1$, seguimos los pasos de la realización del contraste hasta la definición de la región de rechazo R incluida:

- Por ejemplo

$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0, \end{cases},$$

pero podría ser con hipótesis alternativa unilateral también.

- Nos fijamos α .
- El estadístico de prueba es $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, que sigue una distribución Normal estándar si H_0 es cierta.
- Construimos la región de rechazo según el tipo de hipótesis alternativa que nos hemos planteado. Por ejemplo si es bilateral, la región es



- A partir de aquí, podemos pasar al cálculo de la potencia: sabemos que

$$Pot(\mu_1) = \mathbb{P}_{\mu=\mu_1}(\text{Rechazar } H_0),$$

es decir que

$$Pot(\mu_1) = \mathbb{P}_{\mu=\mu_1}(Z_0 \in R). \quad (\text{VII.1})$$

En el caso de una hipótesis alternativa bilateral, esta probabilidad es

$$Pot(\mu_1) = \mathbb{P}_{\mu=\mu_1}((Z_0 \leq -z_{1-\alpha/2}) \cup (Z_0 \geq z_{1-\alpha/2})).$$

Para calcular la potencia necesitamos por lo tanto conocer la distribución de Z_0 cuando H_0 no es cierta, sino $\mu = \mu_1$. Para ello, utilizamos la relación siguiente

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}.$$

Si $\mu = \mu_1$, la variable $\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}}$ sigue una distribución Normal estándar. Deducimos por lo tanto que

$$\text{Si } \mu = \mu_1, \quad Z_0 \sim \mathcal{N}(\delta, 1),$$

donde δ se llama el parámetro de no-centralidad y se define como

$$\delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}.$$

Ésta es la distribución que utilizaremos para calcular la potencia a partir de la expresión en (VII.1). Para ello bastará con tipificar la variable Z_0 para expresar la probabilidad buscada en términos de ϕ .

VII.5.3. Ejemplo de cálculo de la potencia

Volvamos al ejemplo del apartado VII.3.3.1, en él que estudiamos la longitud media de los artículos producidos. La v.a. introducida es $X = \text{longitud de un artículo producido}$ y hemos supuesto que $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\sigma = 1$.

Queremos comprobar que la longitud media de los artículos producidos no es significativamente distinta de 50mm. Para ello, planificamos llevar a cabo el contraste

$$\begin{cases} H_0 : \mu = 50, \\ H_1 : \mu \neq 50, \end{cases},$$

cogiendo una muestra de 10 piezas, y fijando una confianza del 95 %.

¿Cuál es la probabilidad de que, si en realidad $\mu = 50,5$, y por lo tanto H_0 es falsa, el contraste que hemos planeado nos permita detectar que H_0 es falsa, es decir que nos lleve a rechazar H_0 .

Queremos calcular $Pot(50,5)$. Desarrollamos el contraste hasta la determinación de R .

▪

$$\begin{cases} H_0 : \mu = 50, \\ H_1 : \mu \neq 50, \end{cases}$$

▪ Nos fijamos $\alpha = 0,05$.

▪ El estadístico $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ sigue una distribución Normal estándar si H_0 es cierta.

- La región de rechazo es $R = \{z : z < -z_{1-\alpha/2} \text{ o } z > z_{1-\alpha/2}\}$ es decir $R = \{z : z < -1,96 \text{ o } z > 1,96\}$.
- Ahora

$$Pot(50,5) = \mathbb{P}_{\mu=\mu_1}(Z_0 \in R) = \mathbb{P}_{\mu=\mu_1}((Z_0 \leq -1,96) \cup (Z_0 \geq 1,96)).$$

Sabemos que, si $\mu = \mu_1$, $Z_0 \sim \mathcal{N}(\delta, 1)$. Calculemos δ :

$$\delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} = \frac{50,5 - 50}{1/\sqrt{10}} \simeq 1,58.$$

Deducimos tipificando que

$$\begin{aligned} Pot(50,5) &= \mathbb{P}_{\mu=\mu_1}(Z_0 \leq -1,96) + \mathbb{P}_{\mu=\mu_1}(Z_0 \geq 1,96) \\ &= \mathbb{P}_{\mu=\mu_1}\left(\frac{Z_0 - \delta}{1} \leq \frac{-1,96 - \delta}{1}\right) + \mathbb{P}_{\mu=\mu_1}\left(\frac{Z_0 - \delta}{1} \geq \frac{1,96 - \delta}{1}\right) \\ &= \mathbb{P}(Z \leq -3,54) + \mathbb{P}(Z \geq 0,38) \\ &= \phi(-3,54) + (1 - \phi(0,38)) = 1 - \phi(3,54) - (1 - \phi(0,38)) \simeq 0,35. \end{aligned}$$

Esta potencia es insuficiente, para mejorarla, tendremos que planificar un experimento con más observaciones.

VII.5.4. Factores que influyen la potencia

- Cuanto mayor sea n , mayor será la potencia.
- Cuanto menor sea σ , mayor será la potencia.
- Cuanto mayor sea el nivel de confianza, menor será la potencia: si exigimos más confianza, pagamos un precio...
- Cuanto más diferencia haya entre μ_1 y μ_0 , más fácil será detectar cuando μ no es igual a μ_0 sino a μ_1 , por lo tanto, mayor será la potencia.

VII.6. Inferencia para la media

En la presentación del contraste de hipótesis, hemos considerado el caso en que el modelo es normal con varianza conocida. En el caso más realista en que no se especifica el valor de la varianza como parte del modelo, lo estimaremos a partir de la muestra. A continuación construimos contrastes de hipótesis para la media de una distribución Normal con varianza desconocida.

VII.6.1. Contraste de hipótesis para la media μ de una distribución Normal con varianza desconocida

VII.6.1.1. Construcción

Seguimos los mismos pasos que en el caso en que la varianza es conocida.

- Planteamos las hipótesis. Por ejemplo para una hipótesis alternativa bilateral:

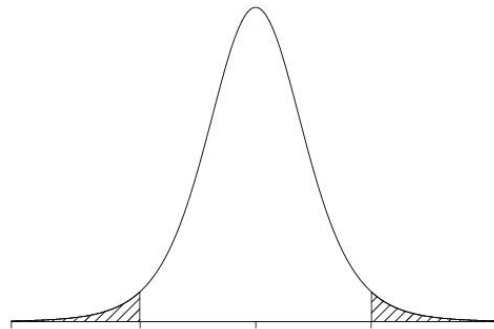
$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0, \end{cases}$$

donde μ_0 representa el valor concreto con el que queremos comparar μ .

- Nos fijamos el valor de α .
- El estadístico de prueba es

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{si } H_0 \text{ es cierto.}$$

- Podemos ahora especificar la región de rechazo.

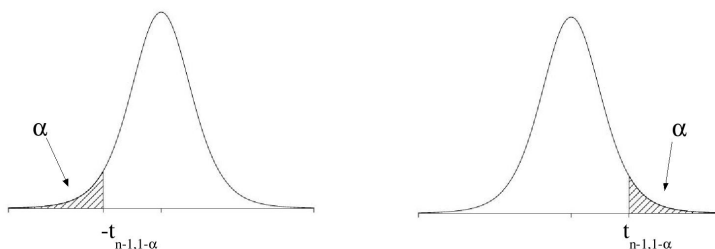


La región R está formada por los valores menores que $-t_{n-1,1-\alpha/2}$ o mayores que $t_{n-1,1-\alpha/2}$.

- Nos queda calcular, para nuestra muestra, el valor concreto del estadístico de prueba T_0 . Si pertenece a R , rechazaremos H_0 y afirmaremos H_1 , mientras que si no pertenece a R , admitiremos H_0 .

En el caso en que la hipótesis alternativa es unilateral lo único que cambia es la región de rechazo:

$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu < \mu_0, \end{cases} \qquad \begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu > \mu_0, \end{cases}$$



VII.6.1.2. Ejemplo

Volvamos al ejemplo de las mediciones visto en la sección anterior, queremos contrastar si el centro de los valores proporcionados por el aparato es mayor que 10.2, basándonos en las mismas tres mediciones.

Planteamos las hipótesis

$$\begin{cases} H_0 : \mu = 10,2, \\ H_1 : \mu > 10,2, \end{cases}$$

Nos fijamos $\alpha = 0,05$, suponiendo que trabajamos con 95 % de confianza. El estadístico de prueba es

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{si } H_0 \text{ es cierto.}$$

La región de rechazo es unilateral : $R = \{t : t > t_{n-1,1-\alpha}\}$, la frontera siendo $t_{2,0,95} = 2,92$.

Para la muestra escogida, el valor del estadístico de prueba es

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{10,24333 - 10,2}{\sqrt{0,0002333}/\sqrt{3}} \simeq 4,913.$$

Este valor pertenece a la región de rechazo por lo que deducimos que al 95 % de confianza rechazamos H_0 .

Notar en particular que deducimos en particular, puesto que hemos rechazado H_0 al 95 % de confianza, que el p-valor es menor que 0.05. En realidad, al igual que en el tema 7, caracterizamos el p-valor como

$$\alpha_0 = \mathbb{P}(t > 4,913),$$

donde t es una distribución t de Student con 2 grados de libertad. Podemos utilizar una calculadora estadística para calcular α_0 de manera precisa. Si sólo tenemos una tabla a mano, podemos ir probando con distintos niveles de confianza para obtener cuotas razonablemente precisas de α_0 .

Por ejemplo, de la tabla de los cuantiles de la distribución t que se encuentra en el apéndice, deduzco que el valor del estadístico de prueba, $T_0 = 4,913$ es mayor que $t_{2,0,975}$ pero menor que $t_{2,0,99}$. Deduzco que rechazaría H_0 al 97.5 % de confianza pero la aceptaría al 99 % de confianza: el p-valor α_0 está comprendido entre 0,025 y 0,01.

VII.7. Inferencia para dos medias

Consideramos ahora situaciones en las que modelizamos dos variables X_1 y X_2 y nos interesa posiblemente comparar sus dos medias, que denotamos respectivamente por μ_1 y μ_2 .

Extraeremos dos muestras: una correspondiente a la primera variable X_1 y otra correspondiente a X_2 . Utilizamos la notación siguiente para designar los valores de estas muestras:

$$\begin{aligned} \text{Muestra 1: } & x_{11}, x_{12}, \dots, x_{1,n_1} \\ \text{Muestra 2: } & x_{21}, x_{22}, \dots, x_{2,n_2} \end{aligned}$$

En particular, hemos supuesto que el tamaño de la muestra 1 es n_1 , mientras que el tamaño de la muestra 2 es n_2 .

Supondremos que hemos modelizado tanto la distribución de X_1 como la distribución de X_2 por Normales,

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

VII.7.1. Estadísticos muestrales

Al pretender comparar μ_1 y μ_2 , nos basaremos en la cantidad $\mu_1 - \mu_2$. El estadístico que utilizaremos para estimar esta cantidad es $\bar{X}_1 - \bar{X}_2$, donde \bar{X}_1 y \bar{X}_2 denotan la media de la primera y de la segunda muestra respectivamente. Introducimos también la notación S_1^2 y S_2^2 para designar las varianzas respectivas de las dos muestras.

Pasamos ahora a presentar distintos estadísticos relacionados con $\bar{X}_1 - \bar{X}_2$ entre los que tendremos que escoger según la situación de modelización en la que nos encontremos: ¿conocemos σ_1^2 y σ_2^2 ?, ¿las desconocemos pero las suponemos iguales? etc...

VII.7.1.1. Caso de varianzas conocidas

Se cumple

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

VII.7.1.2. Caso de varianzas desconocidas

a) Si se suponen las varianzas iguales

Si a la hora de la modelización hemos supuesto $\sigma_1^2 = \sigma_2^2$, podemos estimar la varianza común σ^2 utilizando las dos muestras. Introducimos

$$S_0^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Utilizaremos la distribución

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_0^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}.$$

b) Si NO se suponen iguales

En este caso, no se conoce de manera exacta la distribución muestral del estadístico natural $\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$. Sin embargo, se puede utilizar la aproximación siguiente:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_k, \quad \text{donde } k = \inf(n_1 - 1, n_2 - 1).$$

VII.7.2. Intervalos y contrastes

La construcción de los intervalos y contrastes para $\mu_1 - \mu_2$ se realiza siguiendo los mismos principios que para el caso de una media sólo.

Para ilustrar esta construcción, nos limitamos por lo tanto a tratar dos ejemplos extraídos de problemas de exámenes

a). Ejemplo I. Dos disciplinas de cola para servicio de CPU han sido propuestas por dos diseñadores de sistemas operativos. Para compararlas se instalaron en dos máquinas test iguales y se midieron los tiempos de espera en cada una de ellas de 8 tareas aleatoriamente elegidas:

A	2.41	6.50	3.29	1.22	2.59	2.81	5.35	1.78
B	2.30	5.86	3.71	1.10	2.34	2.24	5.00	1.95

Suponiendo que la distribución que sigue cada variable se puede aproximar por una Normal, calcular el intervalo de confianza para la diferencia entre el tiempo promedio de espera con la disciplina A y el tiempo promedio de espera con la disciplina B.

Solución. *Introduzcamos las variables*

X_A : tiempo de espera de una tarea escogida al azar, procesada por la disciplina A
 X_B : tiempo de espera de una tarea escogida al azar, procesada por la disciplina B

La hipótesis de modelización sobre las distribuciones de X_A y X_B es

$$X_A \sim \mathcal{N}(\mu_A, \sigma_A^2), \quad X_B \sim \mathcal{N}(\mu_B, \sigma_B^2).$$

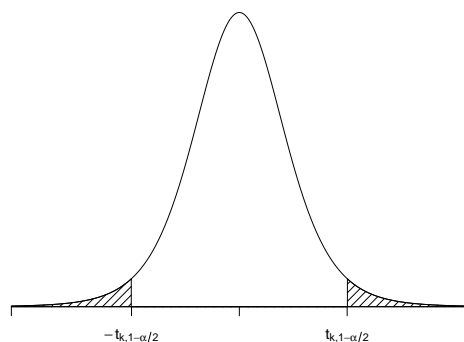
y que son independientes.

Queremos construir un intervalo de confianza para $\mu_A - \mu_B$.

- Nos fijamos el nivel de riesgo $\alpha = 0,05$, es decir una confianza de 95 %.
- El estadístico de prueba, puesto que desconocemos las dos varianzas de X_A y X_B es el descrito en el apartado VII.7.1.2 b)

$$\frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim t_k, \quad \text{donde } k = \inf(n_A - 1, n_B - 1).$$

- Dibujamos una región central con área $1 - \alpha$ en la representación de la densidad del estadístico:



■ *Deducimos que*

$$\mathbb{P}(-t_{k,1-\alpha/2} \leq \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \leq t_{k,1-\alpha/2}) = 1 - \alpha.$$

Despejamos $\mu_A - \mu_B$ y obtenemos

$$\mu_A - \mu_B = \bar{X}_A - \bar{X}_B \pm t_{k,1-\alpha/2} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}.$$

Por otra parte, calculamos

$$\begin{aligned} \bar{X}_A &= 3,24375 & S_A^2 &= 3,227 \\ \bar{X}_B &= 3,0625 & S_B^2 &= 2,695 \end{aligned}$$

Tenemos $n_A = n_B = 8$, y finalmente necesitamos $t_{k\alpha/2} = t_{7,0,975} = 2,365$

Al sustituir obtenemos

$$\mu_A - \mu_B = 0,18125 \pm 2,0349.$$

b). Ejemplo II. Una determinada empresa de material fungible puede adquirir los cartuchos de tóner de impresora de dos proveedores distintos. Con el fin de determinar a que proveedor comprar se toma una muestra de tamaño 12 de cada uno de los proveedores obteniendo los siguientes resultados (número de hojas impresas):

	Media muestral	varianza muestral
<i>Proveedor A</i>	5459	111736
<i>Proveedor B</i>	5162	145258

Si suponemos que las poblaciones son normales con varianzas iguales:

- (a) Construir un intervalo de confianza para la diferencia entre el número medio de hojas que imprime el cartucho de cada proveedor. (tomar $\alpha = 0,05$).

Solución: *Introducimos las variables*

X_A : duración de un cartucho de tóner del proveedor A.

X_B : duración de un cartucho de tóner del proveedor B

Del enunciado sabemos que

$$X_A \sim \mathcal{N}(\mu_A, \sigma^2), \quad X_B \sim \mathcal{N}(\mu_B, \sigma^2),$$

es decir que las dos variables son Normales con varianzas desconocidas pero iguales.

Para construir el intervalo de confianza al 95 %, seguimos los mismos pasos que en el ejemplo anterior, pero ahora el estadístico es

$$\frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{\sqrt{S_0^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} \sim t_{n_A+n_B-2},$$

con $S_0^2 = \frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A+n_B-2}$. Obtenemos por lo tanto que el intervalo de confianza para $\mu_A - \mu_B$ es

$$\mu_A - \mu_B = \overline{X}_A - \overline{X}_B \pm t_{n_A+n_B-2, 1-\alpha/2} \sqrt{S_0^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}.$$

Necesitamos calcular S_0^2 :

$$S_0^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(11)111736 + 11 \cdot 145258}{22} \simeq 128497$$

Deducimos sustituyendo que el intervalo al 95 % de confianza es

$$\mu_A - \mu_B = 297 \pm 302,9.$$

- (b) Razonar qué tipo de contraste se debe de realizar con el fin de decidir si la duración media de los cartuchos del proveedor A es mayor que la de los cartuchos del proveedor B. Realizar este contraste. (tomar $\alpha = 0,05$).

Solución: Queremos plantear el contraste

$$\begin{cases} H_0 : \mu_A = \mu_B, \\ H_1 : \mu_A > \mu_B, \end{cases}$$

es decir

$$\begin{cases} H_0 : \mu_A - \mu_B = 0, \\ H_1 : \mu_A - \mu_B > 0, \end{cases}$$

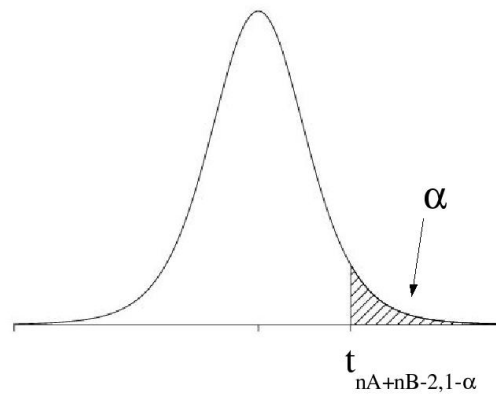
Nos fijamos $\alpha = 0,05$, el estadístico de contraste es

$$\frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{\sqrt{S_0^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}},$$

bajo H_0 , $\mu_A - \mu_B = 0$, y este estadístico se simplifica:

$$T_0 = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{S_0^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} \sim t_{n_A+n_B-2}, \quad \text{si } H_0 \text{ es cierta.}$$

La región de rechazo es unilateral y es de la forma



Su frontera es $t_{n_A+n_B-2, 1-\alpha/2} = t_{22, 0,95} = 1,717$. Nos falta calcular el valor concreto del estadístico de contraste

$$T_0 = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_0^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} = \frac{5459 - 5162}{\sqrt{128497 \left(\frac{1}{12} + \frac{1}{12} \right)}} = 2,0295.$$

El valor de T_0 pertenece a la región de rechazo, deducimos que podemos rechazar H_0 al 95 % de confianza: afirmamos que la duración media de los cartuchos del proveedor A es significativamente mayor que la de los cartuchos del proveedor B.

Apéndice

Distribución t de Student

Valores de los cuantiles de la distribución t de Student con k grados de libertad: para un $0 \leq p \leq 1$, el valor $t_{k,p}$ satisface $\mathbb{P}(t \leq t_{k,p}) = p$.

k	$t_{k,0,995}$	$t_{k,0,99}$	$t_{k,0,975}$	$t_{k,0,95}$	$t_{k,0,90}$	$t_{k,0,80}$	$t_{k,0,70}$	$t_{k,0,60}$	$t_{k,0,50}$
1	63,657	31,821	12,706	6,314	3,078	1,376	0,727	0,325	0,158
2	9,925	6,965	4,303	2,92	1,886	1,061	0,617	0,289	0,142
3	5,841	4,541	3,182	2,353	1,638	0,978	0,584	0,277	0,137
4	4,604	3,747	2,776	2,132	1,533	0,941	0,569	0,271	0,134
5	4,032	3,365	2,571	2,015	1,476	0,92	0,559	0,267	0,132
6	3,707	3,143	2,447	1,943	1,44	0,906	0,553	0,265	0,131
7	3,499	2,998	2,365	1,895	1,415	0,896	0,549	0,263	0,13
8	3,355	2,896	2,306	1,86	1,397	0,889	0,546	0,262	0,13
9	3,25	2,821	2,262	1,833	1,383	0,883	0,543	0,261	0,129
10	3,169	2,764	2,228	1,812	1,372	0,879	0,542	0,26	0,129
11	3,106	2,718	2,201	1,796	1,363	0,876	0,54	0,26	0,129
12	3,055	2,681	2,179	1,782	1,356	0,873	0,539	0,259	0,128
13	3,012	2,65	2,16	1,771	1,35	0,87	0,538	0,259	0,128
14	2,977	2,624	2,145	1,761	1,345	0,868	0,537	0,258	0,128
15	2,947	2,602	2,131	1,753	1,341	0,866	0,536	0,258	0,128
16	2,921	2,583	2,12	1,746	1,337	0,865	0,535	0,258	0,128
17	2,898	2,567	2,11	1,74	1,333	0,863	0,534	0,257	0,128
18	2,878	2,552	2,101	1,734	1,33	0,862	0,534	0,257	0,127
19	2,861	2,539	2,093	1,729	1,328	0,861	0,533	0,257	0,127
20	2,845	2,528	2,086	1,725	1,325	0,86	0,533	0,257	0,127
21	2,831	2,518	2,08	1,721	1,323	0,859	0,532	0,257	0,127
22	2,819	2,508	2,074	1,717	1,321	0,858	0,532	0,256	0,127
23	2,807	2,5	2,069	1,714	1,319	0,858	0,532	0,256	0,127
24	2,797	2,492	2,064	1,711	1,318	0,857	0,531	0,256	0,127
25	2,787	2,485	2,06	1,708	1,316	0,856	0,531	0,256	0,127
26	2,779	2,479	2,056	1,706	1,315	0,856	0,531	0,256	0,127
27	2,771	2,473	2,052	1,703	1,314	0,855	0,531	0,256	0,127
28	2,763	2,467	2,048	1,701	1,313	0,855	0,53	0,256	0,127
29	2,756	2,462	2,045	1,699	1,311	0,854	0,53	0,256	0,127
30	2,75	2,457	2,042	1,697	1,31	0,854	0,53	0,256	0,127
40	2,704	2,423	2,021	1,684	1,303	0,851	0,529	0,255	0,126
60	2,66	2,39	2	1,671	1,296	0,848	0,527	0,254	0,126
120	2,617	2,358	1,98	1,658	1,289	0,845	0,526	0,254	0,126
> 120	2,576	2,326	1,960	1,645	1,282	0,842	0,524	0,253	0,126