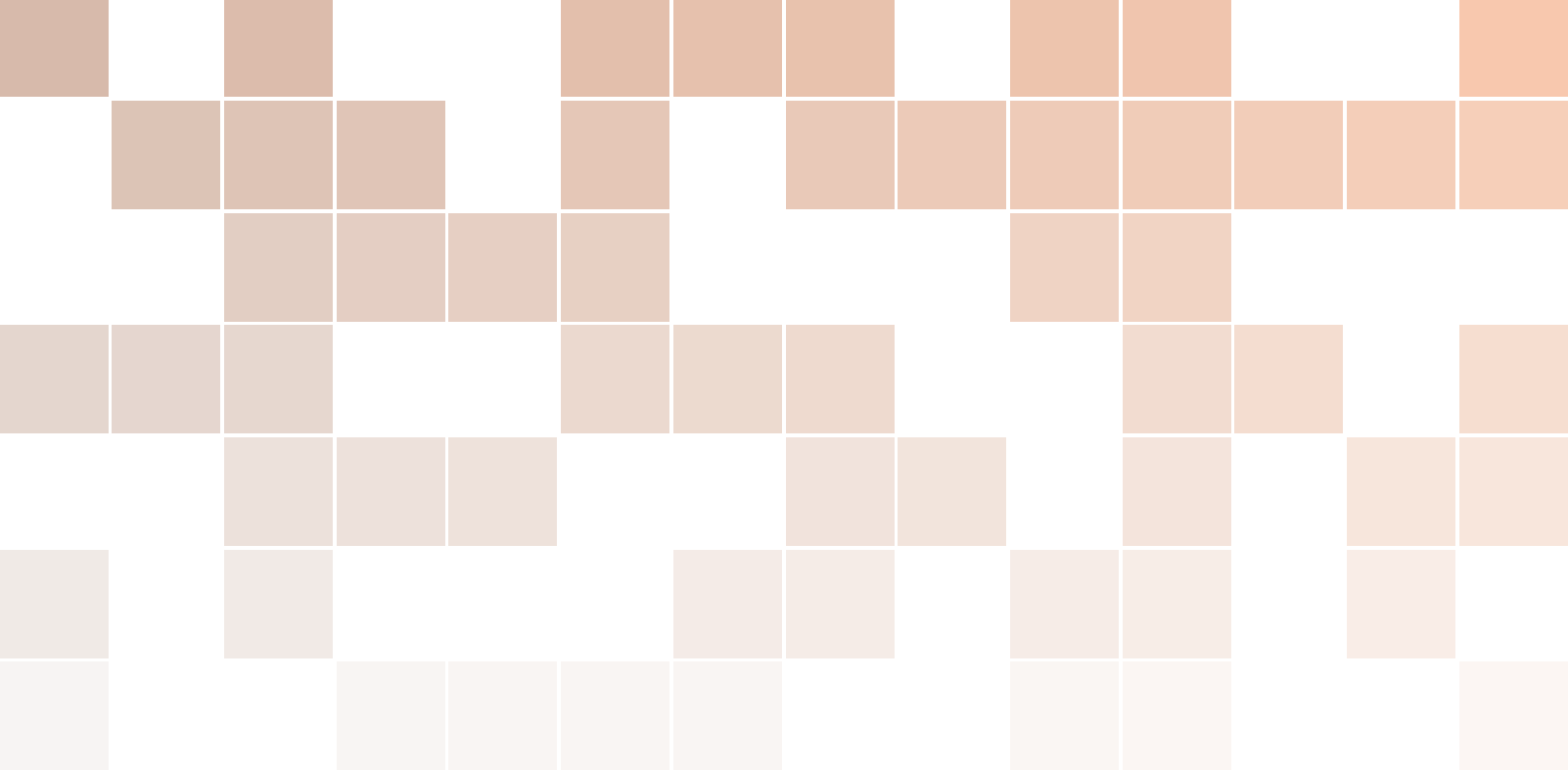


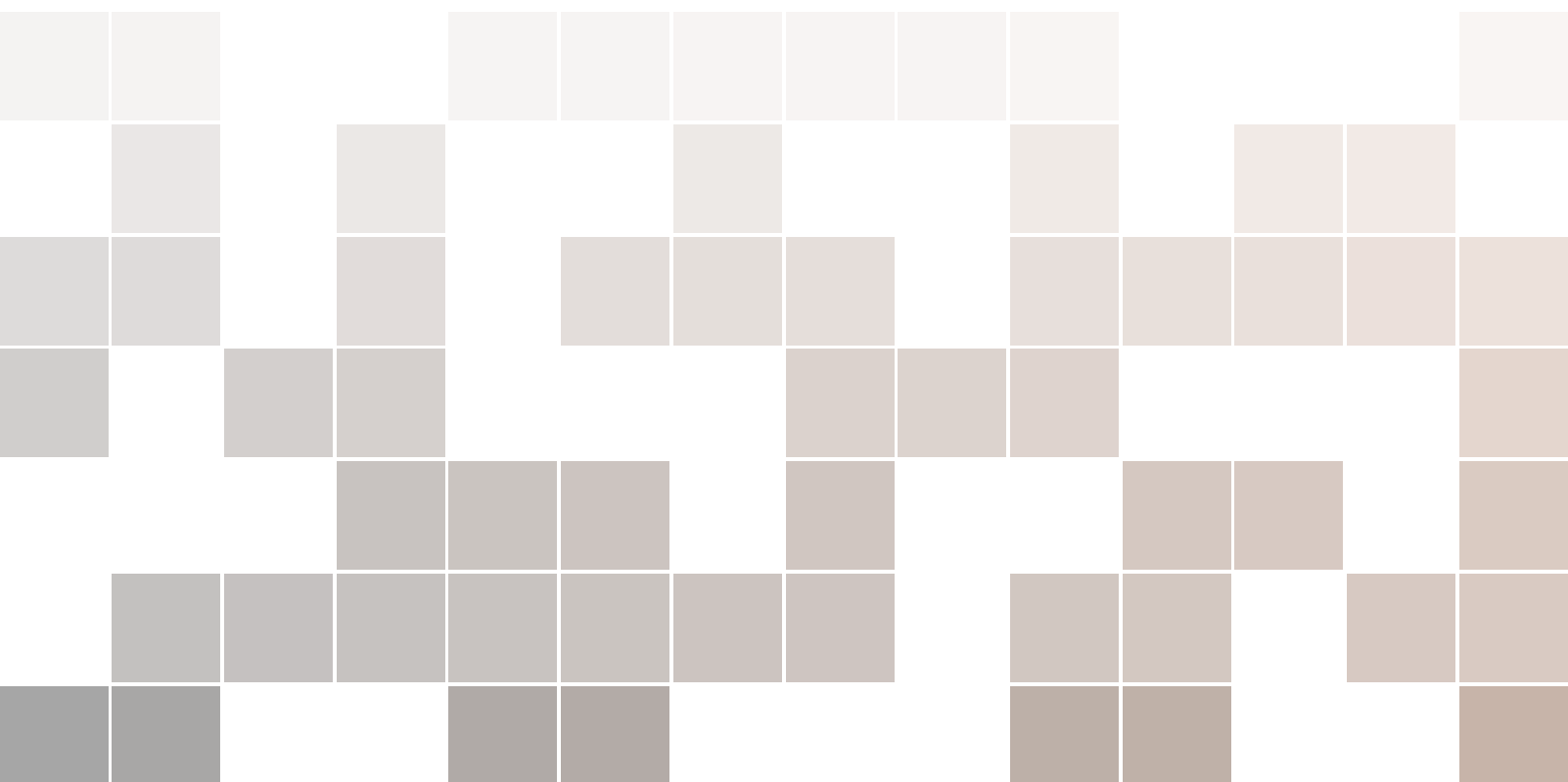


Matemáticas III

Apuntes

Manuel E. Gegúndez Arias

Manuel Maestre Hachero

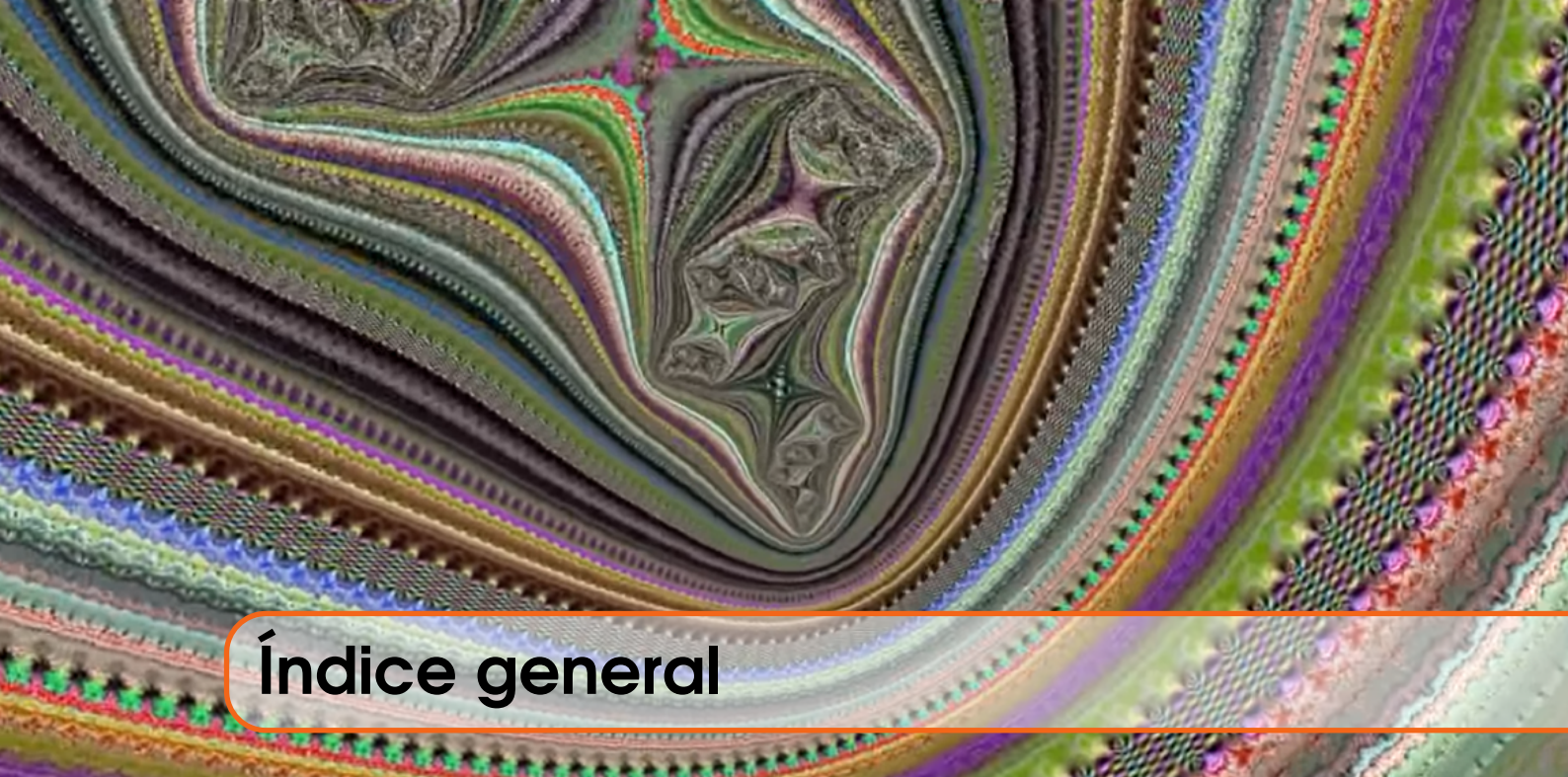


Copyright © 2013 Nombre del Autor

Este texto fue publicado originalmente en la web en septiembre de 2013.

Está licenciado bajo una licencia **Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)**.

Para más información, consulta: <https://creativecommons.org/licenses/by-nc/4.0/>



Índice general

I

Part One

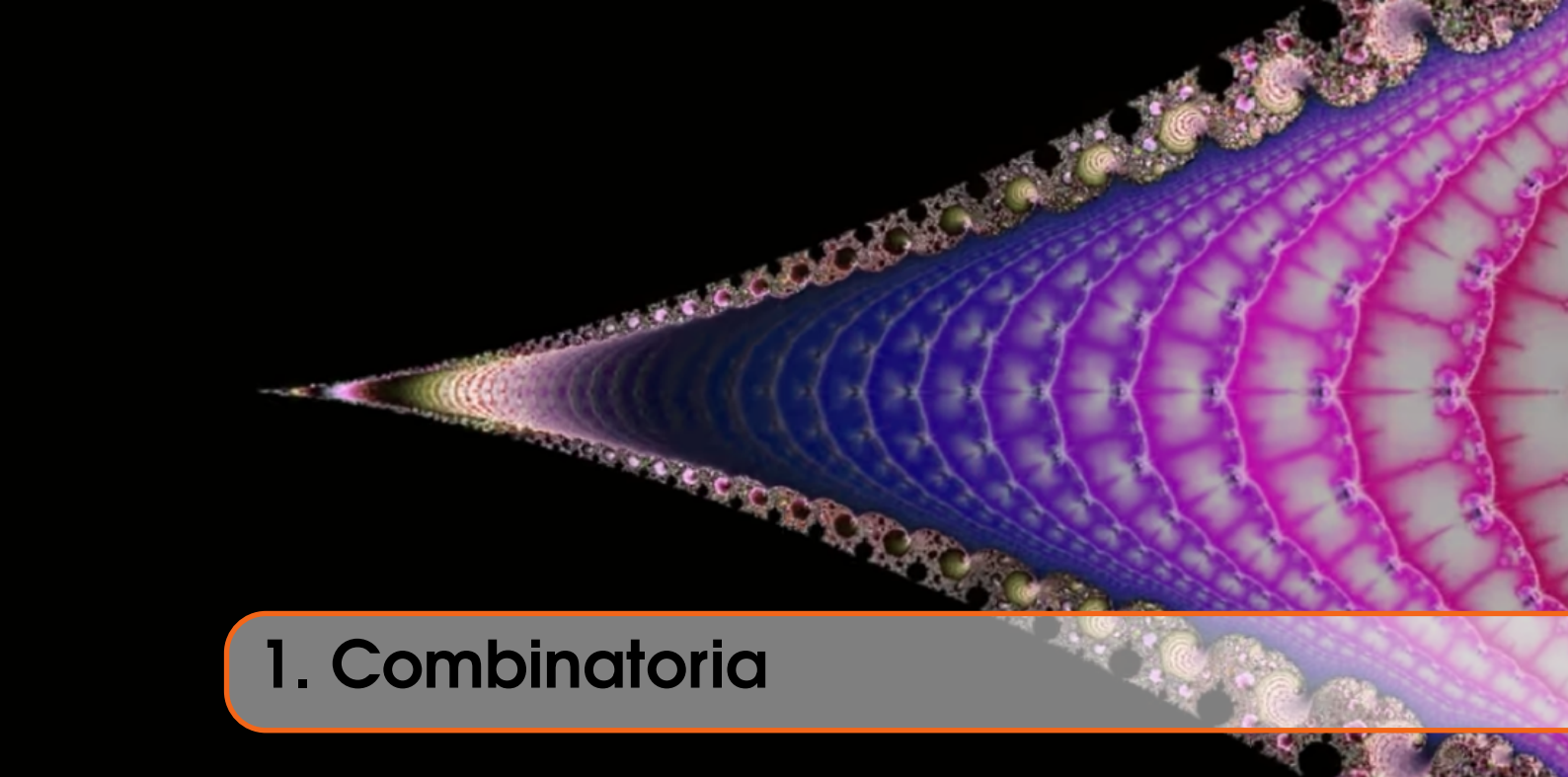
1	Combinatoria	9
1.1	Principios básicos	9
1.2	Variaciones (o arreglos) y permutaciones	10
1.3	Combinaciones	12
1.4	Permutaciones, variaciones y combinaciones generalizadas	12
1.4.1	Permutaciones con repetición	13
1.4.2	Variaciones con repetición	13
1.4.3	Combinaciones con repetición	13
1.5	Coeficientes binomiales e identidades combinatorias	14
1.6	Cuadro Resumen	15
1.7	Ejercicios	15
2	Probabilidad	21
2.1	Espacios muestrales y sucesos	21
2.2	Algunas relaciones de la teoría de conjuntos	22
2.3	Axiomas y propiedades de la probabilidad	23
2.3.1	Propiedades de la probabilidad	24
2.4	Asignación de probabilidades	25
2.4.1	Probabilidad Geométrica	26

2.5	Probabilidad condicionada	27
2.5.1	El teorema de Bayes	28
2.6	Independencia de sucesos	29
2.6.1	Independencia de dos o más sucesos	30
2.7	Ejercicios	30
2.7.1	Ejercicios de exámenes de años anteriores	37
3	Variables aleatorias y distribuciones de prob.	43
3.1	Variables aleatorias	43
3.2	Distribuciones de probabilidad para variables aleatorias discretas	44
3.3	Variables aleatorias continuas y funciones de densidad.	47
3.4	Características asociadas a una variable aleatoria	48
3.5	Modelo Uniforme Discreto	52
3.6	Modelo de Bernoulli	52
3.7	Modelo binomial	52
3.8	Modelo geométrico	53
3.9	Modelo de Poisson	54
3.10	Modelo Uniforme Continuo	55
3.11	Modelo exponencial	55
3.12	Modelo Normal	56
3.12.1	Modelo Normal reducida o tipificada	56
3.12.2	Normal Generalizada	57
3.13	Aproximaciones de modelos de variables	58
3.13.1	Aproximación de una v.a. binomial por una v.a. Poisson	58
3.13.2	Aproximación de una v.a. binomial por una v.a. normal	59
3.13.3	Aproximación de una v.a. Poisson por una v.a. normal	59
3.13.4	Relación entre una v.a. Poisson y una v.a. exponencial negativa	59
3.13.5	Corrección por continuidad	60
3.14	Ejercicios	60
3.14.1	Ejercicios de exámenes de años anteriores	65
4	Muestreo y Estimación	73
4.1	Muestreo	73
4.1.1	Estadísticos y sus Distribuciones	73
4.1.2	Distribución de la media muestral	75
4.1.3	Teorema Central del Límite	75
4.1.4	Distribución de una comb. lineal	76
4.2	Estimación puntual	77
4.2.1	Algunos conceptos generales de la estimación puntual	77
4.2.2	Métodos de estimación puntual	80

4.3	Estimación por intervalos de confianza	82
4.3.1	Propiedades básicas	83
4.3.2	Interpretación de los intervalos de confianza	83
4.3.3	Obtención de un intervalo de confianza	84
4.3.4	Int. de confianza en poblaciones normales	85
4.4	Ejercicios	87
4.4.1	Ejercicios de exámenes de años anteriores	93
5	Contraste de Hipótesis	103
5.1	Hipótesis y procedimientos de prueba	103
5.1.1	Tipos de errores en un contraste	104
5.1.2	Procedimientos de prueba	105
5.2	Prueba para μ con σ conocida en una población normal	107
5.2.1	En función de la región crítica	107
5.2.2	En función del estadístico de contraste	108
5.3	Contraste de hipótesis mediante el p-valor	110
5.4	Ejercicios	111
5.4.1	Ejercicios de exámenes de años anteriores	115
6	Modelos de Regresión	123
6.1	Modelo de regresión lineal simple	123
6.1.1	Modelo probabilístico lineal	124
6.2	Estimación de los parámetros del modelo	125
6.2.1	Estimación de β_0 y β_1	125
6.2.2	Estimación de σ^2 y σ	126
6.3	Coeficiente de Determinación	127
6.3.1	Coeficiente de variación	128
6.4	Inferencias sobre el parámetro de la pendiente β_1	128
6.4.1	Intervalo de confianza para β_1	129
6.4.2	Procedimiento de contraste de hipótesis	129
6.5	Inferencia relacinada con $E[Y _{x^*}]$ y predicción de valores futuros	130
6.6	Correlación	130
6.7	El modelo de regresión lineal múltiple.	131
6.7.1	Mod. lineal general y proc. de los mínimos cuadrados	131
6.7.2	Método de mínimos cuadrados	132
6.8	Regresión no lineal	133
6.9	Ejercicios	134
	Combinatoria	



1.6	Exercises	
1.7	Exercises	
2	Probabilidad	21
2.1	Espacios muestrales y sucesos	
2.2	Algunas relaciones de la teoría de conjuntos	
2.3	Axiomas y propiedades de la probabilidad	
2.4	Asignación de probabilidades	
2.5	Probabilidad condicional	
2.6	Independencia de sucesos	
2.7	Ejercicios	
3	Variables aleatorias y distribuciones de prob.	43
3.1	Variables aleatorias	
3.2	Distribuciones de probabilidad para variables aleatorias discretas	
3.3	Variables aleatorias continuas y funciones de densidad.	
3.4	Características asociadas a una variable aleatoria	
3.5	Modelo Uniforme Discreto	
3.6	Modelo de Bernoulli	
3.7	Modelo binomial	
3.8	Modelo geométrico	
3.9	Modelo de Poisson	
3.10	Modelo Uniforme Continuo	
3.11	Modelo exponencial	
3.12	Modelo Normal	
3.13	Aproximaciones de modelos de variables	
3.14	Ejercicios	
4	Muestreo y Estimación	73
4.1	Muestreo	
4.2	Estimación puntual	
4.3	Estimación por intervalos de confianza	
4.4	Ejercicios	
5	Contraste de Hipótesis	103
5.1	Hipótesis y procedimientos de prueba	
5.2	Prueba para μ con σ conocida en una población normal	
5.3	Contraste de hipótesis mediante el p -valor	
5.4	Ejercicios	
6	Modelos de Regresión	123
6.1	Modelo de regresión lineal simple	
6.2	Estimación de los parámetros del modelo	
6.3	Coefficiente de Determinación	
6.4	Inferencias sobre el parámetro de la pendiente β_1	
6.5	Inferencia relacionada con $E[Y x^*]$ y predicción de valores futuros	
6.6	Correlación	
6.7	El modelo de regresión lineal múltiple.	
6.8	Regresión no lineal	
6.9	Ejercicios	



1. Combinatoria

Muchos problemas de probabilidad requieren que contemos el número de formas en la que un determinado suceso puede ocurrir. La combinatoria es la ciencia que estudia cómo contar sin enumerar directamente todos los casos posibles utilizando técnicas de ordenación, colocación, selección, etc., de objetos. Por ello, estudiaremos algunos conceptos básicos de combinatoria como son las permutaciones o variaciones y las combinaciones (generalizadas o no).

Antes de comenzar con estos conceptos es necesario introducir algunos resultados básicos de conteo así como los conceptos de pares, ternas y tuplas.

1.1 Principios básicos

Definición 1.1.1 Una **tupla** es una lista ordenada de elementos. Una n -tupla es una secuencia (o lista ordenada) de n elementos, siendo n un número natural. El caso de una tupla con 2 elementos, 2-tupla, se denomina también par ordenado. Las tuplas se denotan por la secuencia ordenada de sus elementos entre corchetes ($[]$), separados por comas. Por ejemplo, $[2, 7, 4, 1, 7]$ denota una 5-tupla.

Definición 1.1.2 Un **conjunto** es una colección de elementos distintos. Normalmente se caracterizan por compartir alguna propiedad. Para que un conjunto esté bien definido debe ser posible discernir si un elemento arbitrario pertenece o no al conjunto. Los conjuntos se denotan por la enumeración de sus elementos entre llaves ($\{ \}$), separados por comas. Por ejemplo, $\{2, 7, 4, 1\}$ y $\{2, 4, 1, 7\}$ definen el mismo conjunto.

Nota 1.1.1 Los elementos de un conjunto no están ordenados, aunque vengan especificados secuencialmente. Los conjuntos pueden definirse de manera explícita, citando todos los elementos de los que consta, $A = \{1, 2, 3, 4, 5\}$, o implícita, dando una o varias características que determinen completamente si un elemento dado pertenece o no al conjunto, $A = \{\text{números naturales del 1 al 5}\}$.

Proposición 1.1.2 — Regla del producto. Si el primer elemento de un par ordenado se puede seleccionar en n_1 formas, y para cada una de estas n_1 formas se puede seleccionar el segundo elemento del par en n_2 formas, entonces el número de pares ordenados es $n_1 \times n_2$.

■ **Ejemplo 1.1** Para reformar una vivienda, requiere los servicios de un arquitecto y de un aparejador. En la población en la que vive hay 12 arquitectos y 9 aparejadores. ¿De cuántas formas se puede seleccionar un equipo con un arquitecto y un aparejador?

Solución Si representaremos cada equipo (solución del problema) como un par ordenado $[ar, ap]$, dónde el primer elemento lo ocupa un arquitecto y el segundo elemento lo ocupa un aparejador, el conjunto de posibles soluciones sería

$$\{[arq_i, apar_j], \text{ con } i = 1, \dots, n_1 \text{ y } j = 1, \dots, n_2\}$$

Como hay $n_1 = 12$ arquitectos y $n_2 = 9$ aparejadores, el número de formas en que puede hacer la elección es:

$$n_1 \times n_2 = 12 \times 9 = 108.$$

■

Proposición 1.1.3 — Regla del producto o principio de la multiplicación (para k -tuplas). Si para construir una k -tupla hay n_1 opciones posibles para el primer elemento, hay n_2 elecciones posibles del segundo elemento, ... y n_k elecciones posibles para el k -ésimo elemento. Entonces hay $n_1 \times \dots \times n_k$ k -tuplas posibles.

■ **Ejemplo 1.2** Un inspector de control de calidad desea seleccionar un artículo para la inspección de cada uno de cuatro recipientes diferentes que contienen 4, 3, 5 y 4 artículos, respectivamente. ¿De cuántas maneras diferentes se pueden escoger los cuatro artículos?

Solución

El número total de maneras es

$$4 \times 3 \times 5 \times 4 = 240.$$

■

1.2 Variaciones (o arreglos) y permutaciones

Definición 1.2.1 Cualquier secuencia ordenada de n objetos, “ n -tupla ordenada” tomada de un conjunto de n objetos distintos se llama **permutación**. El número de permutaciones de tamaño n que se puede construir a partir de n objetos se denota P_n .

Teorema 1.2.1 Existen $n!$ permutaciones de n elementos.

■ **Ejemplo 1.3** ¿De cuántas maneras diferentes se pueden presentar al público los cinco jugadores titulares de un equipo de baloncesto?

Solución

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \quad \text{maneras}$$

■

Definición 1.2.2 Una **variación** (r -permutación) de n elementos distintos $x_1, \dots, x_n$ es un ordenamiento de un subconjunto r elementos distintos de $[x_1, \dots, x_n]$. El número de variaciones de n elementos tomados de r en r se nota V_n^r o $V_{n,r}$.

Teorema 1.2.2 El número de variaciones de n elementos tomados de r en r es

$$V_n^r = \frac{n!}{(n-r)!} = n(n-1)(n-2) \cdots (n-r+1), \quad r \leq n$$

■ **Ejemplo 1.4** De entre los 24 miembros de un club se sacan 4 nombres para los puestos de presidente, vicepresidente, tesorero y secretario. ¿De cuántas formas diferentes se puede hacer?

Solución

$$V_{24}^4 = \frac{24!}{20!} = 24 \times 23 \times 22 \times 21 = 255024$$

■

Definición 1.2.3 Las permutaciones que ocurren cuando los objetos se arreglan en un círculo se denominan **permutaciones circulares**.

Nota 1.2.3 Dos permutaciones circulares se consideran iguales (y se cuentan sólo una vez) si son iguales salvo por el giro, i.e. los objetos correspondientes en los dos arreglos tienen los mismos objetos a su izquierda y su derecha. Por ejemplo, si cinco personas se sentasen en una mesa redonda, no obtenemos una permutación diferente si todos se cambian a la silla que tienen a su derecha.

Teorema 1.2.4 El número de permutaciones circulares de n elementos es $(n - 1)!$.

■ **Ejemplo 1.5** Cinco personas se disponen a sentarse alrededor de una mesa circular. ¿De cuántas formas distintas pueden hacerlo?

Solución

La cinco personas pueden sentarse de $(5 - 1)! = 4! = 4 \times 3 \times 2 \times 1 = 24$ formas distintas. ■

1.3 Combinaciones

Definición 1.3.1 Una **combinación** (r -combinación) de n elementos distintos $x_1, \dots, x_n$ es una selección no ordenada de r elementos de $\{x_{i_1}, \dots, x_{i_r}\} \subseteq \{x_1, \dots, x_n\}$, cumpliendo que $i_a \neq i_b$ si $a \neq b$. El número de combinaciones de n elementos tomados de r en r , se denota C_n^r o $C_{n,r}$.

En contraposición a las variaciones o arreglos, en las que el hecho diferenciador es el orden interno de la tupla, están las combinaciones en ellas el hecho diferenciador es el contenido de la misma, esto es, los elementos que la componen. En este sentido, podemos considerar una combinación como un conjunto ya que, como con los conjuntos, al describir una combinación no importa el orden en el cual se escriben sus elementos. Así, por ejemplo, la combinación $\{c_1, c_2\}$ es la misma que la combinación $\{c_2, c_1\}$.

Teorema 1.3.1 El número de combinaciones de n elementos tomados de r en r es

$$C_n^r = \binom{n}{r} = \frac{V_n^r}{r!} = \frac{n!}{(n-r)!r!}, \quad r \leq n$$

■ **Ejemplo 1.6** Una compañía de autobuses tiene en su plantilla 10 conductores. ¿Durante cuántos días puede escoger un grupo de diferente de 6 conductores si cada día se forma un grupo de trabajo?

Solución

Dado que no importa el orden en que se eligen, sólo importa si forman o no el mismo conjunto de 6 personas, son *combinaciones*

$$C_{10,6} = C_{10}^6 = \binom{10}{6} = \frac{10!}{6! \cdot 4!} = 210$$

1.4 Permutaciones, variaciones y combinaciones generalizadas

Hasta ahora los tanto los elementos de las variaciones como los de las combinaciones han sido distintos, por ejemplo cada uno de los 12 arquitectos del Ejemplo 1.1 es una persona/objeto distinto y distinguible de los demás, pero esto no tiene por qué ser siempre así. En ocasiones tendremos como objetivo la ordenación de una colección de objetos entre los que se encuentran objetos repetidos o indistinguibles.

1.4.1 Permutaciones con repetición

Proposición 1.4.1 Si en una colección S de n elementos, se tienen n_1 objetos exactamente iguales entre sí de tipo 1, n_2 objetos de tipo 2, ... y n_t objetos de tipo t , entonces el número de ordenamientos de S es

$$P_n^{n_1, n_2, \dots, n_t} = \frac{n!}{n_1! \dots n_t!}$$

■ **Ejemplo 1.7** Cuántos números diferentes pueden escribirse utilizando todos los dígitos de la colección $\langle 1, 3, 6, 2, 1, 3, 1 \rangle = \langle 1, 1, 1, 2, 3, 3, 6 \rangle$

Solución

Es obvio que son permutaciones ya que se están ordenando todos los elementos de la colección. Ahora bien si nos fijamos en el número 1631321 e intercambiamos el tercer y quinto dígitos, obtenemos el mismo número. Se trata, pues, de *permutaciones con repetición* dado que algunos de ellos están repetidos.

$$P_7^{3,1,2,1} = \frac{7!}{3! \times 1! \times 2 \times 1!} = 420$$

1.4.2 Variaciones con repetición

En ocasiones necesitamos conocer el número de ordenaciones diferentes que se puede hacer de un conjunto de objetos, permitiendo la repetición de los componentes. En este caso estamos ante un problema de Variaciones con repetición

Proposición 1.4.2 El número de ordenaciones diferentes de n objetos tomados de m en m es $VR_{n,k} = n^k$.

Nota 1.4.3 Obsérvese que en este caso el número de objetos a seleccionar m puede ser mayor que el número de objetos existentes n .

■ **Ejemplo 1.8** ¿De cuántas maneras diferentes se puede rellenar una quiniela de fútbol?

Para rellenar una quiniela de futbol, hay que seleccionar un resultado de entre tres posibles: 1, X o 2 en cada uno de los 15 partidos de fútbol que la organización escoge.

Solución

Tenemos, pues, tres resultados posibles que, podrían repetirse hasta 15 veces, es decir,

$$VR_{3,15} = 3^{15}$$

maneras diferentes. ■

1.4.3 Combinaciones con repetición

Proposición 1.4.4 El número de selecciones no ordenadas de una colección de n elementos, permitiendo repeticiones es

$$CR_{n,k} = \binom{n+k-1}{k}$$

■ **Ejemplo 1.9** Al regresar a casa 7 siete estudiantes de bachillerato se detienen en un restaurante de comida rápida, donde cada uno puede comer lo siguiente: una hamburguesa, un *hot dog*, un taco o un sandwich de atún. ¿Cuántas compras diferentes son posibles?

Solución

Del enunciado se desprende que interesa el número de artículos comprados de cada clase. se trata, pues, de combinaciones. Como se puede repetir (por ejemplo, todos pueden pedir una hamburguesa), son combinaciones con repetición.

$$CR_4^7 = C_{4+7-1,7} = \binom{10}{7}$$

■

1.5 Coeficientes binomiales e identidades combinatorias

Teorema 1.5.1 — El Teorema del binomio. Si a y b son números reales y n es un entero positivo, entonces

$$(a+b)^n = \sum_{k=0}^n C_{n,k} a^{n-k} b^k = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Teorema 1.5.2

1. $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$
2. $\sum_{k=0}^n \binom{n}{k} = 2^n$

1.6 Cuadro Resumen

	No importa orden Indistinguibles	Sí importa orden Distinguibles
Con repetición Con reemplazamiento	$CR_{n,m} = \binom{n+m-1}{m}$	$VR_{n,m} = n^m$
Sin repetición Sin reemplazamiento	$C_{n,m} = \binom{n}{m}$ ($m \leq n$)	$V_{n,m} = n(n-1) \cdots (n-m+1)$ ($m \leq n$)

Tabla resumen de combinaciones y variaciones, con o sin repetición, de n objetos tomados de m en m .

1.7 Ejercicios

Ejercicio 1.1 La junta médica de un hospital cuenta con un representante de cada una de las cinco especialidades médicas (cardiología, neurología, pediatría, dermatología y cirugía). ¿De cuántas maneras se puede:

- (a) seleccionar coordinador y vocal de la junta?
- (b) seleccionar coordinador, vocal y secretario?
- (c) seleccionar dos integrantes para el consejo de gobierno?

Solución: Antes de comenzar la resolución, es necesario reflexionar sobre qué estamos seleccionando y cómo se estructura. En el primer caso, se trata de asignar nombres a cargos específicos dentro de la junta médica. Es decir, estamos ubicando a 5 personas en 2 cargos distintos.

- (a) El orden importa, ya que no es lo mismo ser coordinador que vocal, y no se permite que una misma persona ocupe ambos cargos.
Esto implica una estructura del tipo $[,] = [\dots]_2$ con dos posiciones (coordinación y vocalía), y la solución es $V_{5,2} = 5!/(5-2)!$
- (b) Se aplica el mismo razonamiento anterior, pero ahora con tres cargos distintos. La solución es $V_{5,3}$
- (c) En este caso, el orden no importa ya que ambos seleccionados tendrán la misma función dentro del consejo de gobierno, y no se permite repetir. Es decir, no se puede asignar a la misma persona dos veces.

Esto corresponde a una estructura del tipo $\{ , \} = \{ \dots \}_2$, por lo que la solución es $C_{5,2} = \binom{5}{2}$ ■

Ejercicio 1.2 Tras la evaluación del examen de la asignatura de “Ciencias de los materiales”, se observa que 8 de los 25 estudiantes han suspendido.

- (a) ¿De cuántas formas se puede seleccionar una muestra de 5 estudiantes de los 25 para una revisión detallada del examen?
- (b) ¿De cuántas formas una muestra de 5 estudiantes incluye exactamente 4 que hayan suspendido?

Solución: Tenemos $n = 25$ estudiantes en dos grupos: 17 aprobados y 8 suspensos.

- (a) Como no importa el orden y no se permite repetir, la respuesta es $C_{25,5} = \binom{25}{5}$
- (b) Sigue sin importar el orden y sin repetición. Necesitamos seleccionar 4 estudiantes que hayan suspendido, $C_{8,4}$, y 1 estudiante que haya aprobado, $C_{17,1}$. Aplicando la regla del producto, la solución es: $C_{8,4} \times C_{17,1}$ ■

Ejercicio 1.3 Una caja de cierto almacén contiene 4 impresoras láser monocromo básicas, cinco impresoras láser color estándar y seis impresoras láser color avanzada. Suponga que se seleccionan al azar 3 impresoras.

Nota: Este problema puede interpretarse de dos formas según si consideramos que las impresoras son distinguibles (como si fueran de diferentes marcas o modelos) o indistinguibles (como si fueran cartuchos de colores).

- (a) ¿En cuántas selecciones hay exactamente 2 impresoras láser color avanzada?
- (b) ¿En cuántas selecciones las 3 impresoras tienen el mismo tipo?
- (c) ¿De cuántas formas se puede seleccionar una impresora de cada tipo? ■

Ejercicio 1.4 El profesor Carmelo Cotón tiene 20 alumnos en la asignatura de Cálculo, 15 en la de Álgebra y 10 en la de Estadística. Un inspector selecciona a 6 de estos alumnos para realizarles entrevistas de evaluación.

- (a) ¿Cuántas selecciones incluyen exactamente 6 alumnos de Cálculo?
- (b) ¿Cuántas selecciones incluyen a 6 alumnos de una misma asignatura?
- (c) ¿En cuántas selecciones están representadas al menos 2 asignaturas distintas?

Solución: Tenemos $n = 45$ alumnos repartidos en tres grupos: 20 en Cálculo, 15 en Álgebra y 10 en Estadística.

- (a) No importa el orden y no se puede repetir. Seleccionamos 6 alumnos de los 20 de Cálculo, así que la solución es $C_{20,6}$.
- (b) Este caso es similar al anterior, pero considerando las tres asignaturas. Cualquier grupo de 6 alumnos pertenecientes al mismo grupo es una configuración válida. La solución es: $C_{20,6} + C_{15,6} + C_{10,6}$.
- (c.1) Si se quiere resolver directamente, se considerarían las combinaciones en que al menos dos asignaturas están representadas. Por ejemplo:

* 1 alumno de Cálculo y 1 alumno de Álgebra, y los 4 restantes de cualquier forma: $20 \times 15 \times C_{38,4}$

* 1 alumno de Cálculo y 1 de Estadística, más 4 cualquiera: $20 \times 10 \times C_{38,4}$

* 1 alumno de Álgebra y 1 de Estadística, más 4 cualquiera: $15 \times 10 \times C_{38,4}$

Por lo tanto, la solución total sería: $20 \times 15 \times C_{38,4} + 20 \times 10 \times C_{38,4} + 15 \times 10 \times C_{38,4}$

(c.2) Otra forma más eficiente es aplicar el principio del complementario. Calculamos el total de grupos posibles sin restricción: $C_{45,6}$, y le restamos las configuraciones en las que todos los alumnos pertenecen a una misma asignatura (calculado en el apartado b). La solución es: $C_{45,6} - (C_{20,6} + C_{15,6} + C_{10,6})$



Ejercicio 1.5 Un inspector de calidad desea programar una cita con cada uno de los 8 trabajadores seleccionados, 4 del turno A y 4 del turno B, para realizar entrevistas individuales. ¿En cuántas programaciones estará, por lo menos, un trabajador del turno B entre los 3 primeros entrevistados?

Solución: Tenemos $n = 8$ trabajadores: 4 del turno A y 4 del turno B.

Al igual que en el ejercicio 1.4, este problema se resuelve más fácilmente mediante el uso del complementario. Es decir, de todas las programaciones posibles, restaremos aquellas en las que “no hay ningún trabajador del turno B” entre los tres primeros entrevistados.

Primero, calculamos el número total de programaciones posibles. Como importa el orden y no se puede repetir, se trata de permutaciones: $V_{8,8} = P_8$.

Ahora, consideramos las programaciones en las que los tres primeros trabajadores entrevistados pertenecen exclusivamente al turno A. Debemos seleccionar y ordenar 3 trabajadores de los 4 del turno A: $V_{4,3}$, y luego ordenar los 5 trabajadores restantes (1 del turno A y 4 del turno B): $V_{5,5} = P_5$.

Por tanto, el número total de programaciones con al menos un trabajador del turno B entre los tres primeros es:

$$V_{8,8} - (V_{4,3} \times V_{5,5})$$



Ejercicio 1.6 Con los dígitos $\{1, 2, 3, 4, 5, 6\}$, ¿cuántos números de tres cifras distintas se pueden escribir?



Ejercicio 1.7

(a) ¿De cuántas formas se pueden sentar en una fila de 10 butacas 5 personas?

(b) ¿Y si fuesen 10 personas?

Solución: Obviamente, sea cual sea el planteamiento, está claro que importa el orden en el que finalmente se sientan las personas y que no se pueden repetir (ni personas ni butacas). La dificultad de este problema está en decidir si lo que establece el orden son las 5 personas $[..]_5$ o los 10 asientos $[..]_{10}$.

(a) En este caso lo que establece el orden son las personas, por lo que estamos

- repartiendo 5 butacas de las 10 que hay. Por tanto la solución es $V_{10,5}$.
- (b) En el caso en que fuesen 10 personas la solución sería $V_{10,10} = P_{10}$



Ejercicio 1.8 ¿De cuántas maneras se pueden repartir tres piezas de sandía y cuatro de melón entre siete personas de forma que cada una reciba una pieza?

Solución: Tenemos un total de $7 = 4 + 3$ piezas de fruta y 7 personas. No se puede repetir, ya que cada persona recibe una pieza, y de algún modo importa el orden, ya que no es lo mismo recibir sandía que melón. Este es el mismo caso que el del ejercicio 1.7; el orden lo ponen las personas. Luego estamos ante un caso de permutaciones con repetición de $7 = 4 + 3$ elementos y la solución es $P_{7,(4,3)}$.



Ejercicio 1.9 Un hospital cuenta con 25 enfermeras y 15 enfermeros.

- (a) ¿Cuántos turnos de guardia de cinco personas se pueden formar?
- (b) ¿Cuántos turnos de guardia están constituidos por tres enfermeras y dos enfermeros?



Ejercicio 1.10 Dadas 10 consonantes y las cinco vocales, ¿cuántas palabras de 5 letras, con y sin sentido, pueden formarse escogiendo 3 consonantes y 2 vocales?



Ejercicio 1.11 El Rector de la universidad desea organizar una campaña educativa y cuenta con cuatro vicerrectores: Beatriz, Carmen, María y Mónica. Se planea ofrecer un total de 10 charlas en distintos colegios. Averiguar de cuántas formas posibles se pueden distribuir esas charlas entre los vicerrectores en los siguientes casos:

- (a) De manera genérica (algún vicerrector puede no participar en ninguna charla).
- (b) Cada vicerrector debe participar, al menos, en una charla.
- (c) Cada vicerrector debe participar, al menos, en una charla y Mónica, vicerrectora ejecutiva, debe encargarse de al menos 5 charlas.

Solución: El objetivo es distribuir $n = 10$ charlas entre 4 personas. La cuestión más importante del problema es: ¿qué estamos repartiendo realmente? En lugar de repartir las 10 charlas como objetos, lo que hacemos es seleccionar, 10 veces, a uno de los vicerrectores (pudiendo repetir). El orden no importa, solo cuántas veces participa cada uno. Por tanto, hablamos de “combinaciones con repetición”.

- (a) Repartimos 10 charlas entre 4 personas, pudiendo repetir. La solución es $CR_{4,10}$.
- (b) Si cada uno debe participar en al menos una charla, damos una charla a cada uno de ellos y repartimos las 6 restantes libremente. Por tanto, la solución es $CR_{4,6}$.
- (c) En este caso, damos 5 charlas a Mónica y una a cada uno de los otros 3 vicerrectores ($5+3=8$), quedando 2 charlas por repartir sin restricción. La solución es entonces $CR_{4,2}$.



Ejercicio 1.12 ¿De cuántas maneras podemos distribuir 7 manzanas y seis naranjas entre cuatro niños, de modo que cada uno reciba al menos una manzana? ■

2. Probabilidad

2.1 Espacios muestrales y sucesos

1

Definición 2.1.1 Un **experimento (aleatorio)** es cualquier acción o proceso cuyo resultado está sujeto a incertidumbre.

■ Ejemplo 2.1

- El lanzamiento de una moneda.
- El lanzamiento de un dado.
- El resultado de la lotería primitiva.

■

Definición 2.1.2 El **espacio muestral** de un experimento, denotado por Ω , es el conjunto formado por todos los posibles resultados de ese experimento.

■ Ejemplo 2.2

- En el experimento “*lanzamiento de una moneda*”
 $\Omega = \{C, X\} = \{\text{cara, cruz}\}.$
- En el experimento: “*lanzamiento de un dado*”
 $\Omega = \{1, 2, 3, 4, 5, 6\}$

■

Definición 2.1.3 Un **suceso** es cualquier colección (subconjunto) de resultados contenida en el espacio muestral Ω . Se dice que un suceso es **simple** (o elemental) si consiste en exactamente un resultado, y **compuesto** si consta de más de uno.

■ **Ejemplo 2.3** En el experimento: “lanzamiento de un dado”

- El suceso $A = \{\text{“obtener un 2”}\}$ es *simple*
- El suceso $B = \{\text{“obtener número par”}\}$ es *compuesto*

■

Definición 2.1.4 Un espacio muestral Ω se dice que es de tipo **discreto** si el número de resultados posibles es finito o infinito numerable. Se dice que es **continuo** si Ω es un conjunto no numerable.

Cuando se lleva a cabo un experimento, se dice que **ocurre** (o se verifica) un suceso particular A si el resultado experimental está contenido en A .

Definición 2.1.5

Ω como subconjunto de sí mismo representa al **suceso seguro**.
 $\emptyset$ representa al **suceso imposible**.

2.2 Algunas relaciones de la teoría de conjuntos

Un suceso no es otra cosa que un conjunto, así que las relaciones y resultados de la teoría elemental de conjuntos se puede usar para estudiar sucesos.

Definición 2.2.1

- La **unión** de dos sucesos A y B , denotada por $A \cup B$ es el suceso que consiste en los resultados que están *ya sean en A , en B o en ambos sucesos*.
- La **intersección** de dos sucesos A y B , que se denota por $A \cap B$ es el suceso que consiste en los resultados que están tanto en A como en B .
- El **complemento** de un suceso A , denotado por $\bar{A}$ (o A^C), es el conjunto de todos los resultados de Ω que no están en A .
- La **diferencia** entre dos sucesos A y B ($A \setminus B$) es el suceso formado por todos los resultados de A que no están en B ($A \setminus B = A \cap \bar{B}$).

Proposición 2.2.1 (Leyes de De Morgan). Se verifica:

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

■ **Ejemplo 2.4** Consideremos el espacio muestral $\Omega \equiv$ lanzamiento de un dado y sea A el suceso “la puntuación obtenida es par” y B el suceso “la puntuación obtenida es múltiplo de 3”. Calcular $A \cup B$, $A \cap B$ y $\overline{A}$.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{2, 4, 6\}$$

$$B = \{3, 6\}$$

$$A \cup B = \{2, 3, 4, 6\}$$

$$A \cap B = \{6\}$$

$$\overline{A} = \{1, 3, 5\}$$

■

Definición 2.2.2 Cuando dos sucesos A y B no tienen resultados en común ($A \cap B = \emptyset$), se dice que son sucesos **mutuamente excluyentes** o **disjuntos**.

■ **Ejemplo 2.5** Consideremos el espacio muestral $\Omega \equiv$ lanzamiento de un dado y sea A el suceso “la puntuación obtenida es par” y B el suceso “la puntuación obtenida es “3””.

Obviamente, $A \cap B = \emptyset$ es decir, son *mutuamente excluyentes* o *disjuntos*. ■

Definición 2.2.3 Los sucesos $A_1, A_2, \dots, A_n$ se dice que son **exhaustivos** si $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$.

Definición 2.2.4 Dado un espacio muestral Ω se llama **espacio de sucesos** (o **partes de Ω**) al conjunto formado por todos los posibles sucesos (elementales y compuestos) incluidos el *suceso seguro* y el *suceso imposible*.

Se suele denotar $\mathcal{P}(\Omega)$ y tiene estructura de σ -álgebra.

2.3 Axiomas y propiedades de la probabilidad

Dados un experimento y un espacio muestral Ω , el objetivo de la probabilidad es asignar a cada suceso A un número $P(A)$, denominado **probabilidad** del suceso A .

Definición 2.3.1 Sea Ω un espacio muestral asociado a un experimento aleatorio y $\mathcal{P}(\Omega)$ el correspondiente espacio de sucesos. Diremos que la aplicación $P : \Omega \rightarrow \mathbb{R}$ es una **función de probabilidad** si verifica los siguientes axiomas:

Axioma 1 Para cualquier suceso A , $P(A) \geq 0$.

Axioma 2 $P(\Omega) = 1$

Axioma 3 Si $A_1, A_2, \dots, A_n, \dots$ es una colección infinita de sucesos mutuamente excluyentes, entonces

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

Nota 2.3.1 En el axioma 3 de la definición anterior, si tuviésemos una colección finita de sucesos, sería un caso particular del anterior ya que bastaría con añadirle, a la colección finita de sucesos, una cantidad infinita de conjuntos vacíos.

Definición 2.3.2 A la terna $(\Omega, \mathcal{P}(\Omega), P)$ se le denomina **espacio de probabilidad**

2.3.1 Propiedades de la probabilidad

Proposición 2.3.2 Para cualquier suceso $P(\bar{A}) = 1 - P(A)$.

Demostración. $1 \underset{Ax,2}{=} P(\Omega) = P(A \cup \bar{A}) \underset{Ax,3}{=} P(A) + P(\bar{A}) \Rightarrow P(\bar{A}) = 1 - P(A)$. ■

Proposición 2.3.3 Si A y B son mutuamente excluyentes, entonces $P(A \cap B) = 0$ (o lo que es lo mismo, $P(\emptyset) = 0$).

Proposición 2.3.4 Para dos sucesos cualesquiera $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proposición 2.3.5 Si $A \subseteq B \Rightarrow P(A) \leq P(B)$.

Proposición 2.3.6 Para tres sucesos cualesquiera

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

2.4 Asignación de probabilidades

Si Ω es de tipo discreto bastará con asignar probabilidad a los sucesos elementales $\{\omega_i\}$ y por el axioma 3 se tendrá que:

$$P(A) = \sum_{\omega_k \in A} P(\omega_k)$$

En ocasiones, cuando un experimento presenta N resultados posibles, puede ser lógico asumir que cada uno de esos sucesos elementales tiene la misma probabilidad. En tal situación, se dice que los sucesos son **equiprobables**.

Sea $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ un espacio muestral discreto formado por n sucesos **equiprobables**. En este caso, $P(\omega_k) = p \quad \forall k$, así que

$$1 = P(\Omega) = \sum_{i=1}^n P(\omega_i) = n \cdot p \quad \Rightarrow \quad p = \frac{1}{N}$$

Para un suceso $A = \{\omega_1, \omega_2, \dots, \omega_k\}$

$$P(A) = \frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n} = \frac{k}{n} = \frac{\text{casos favorables}}{\text{casos posibles}}$$

resultado que se conoce como **Regla de Laplace**.

Nota 2.4.1 Si los resultados no fuesen equiprobables, la naturaleza del problema nos indicará como determinar cada $P(\omega_k)$.

■ **Ejemplo 2.6** Una compañía de inversiones ofrece a sus clientes varios fondos: uno del mercado monetario (MM), tres fondos de bonos diferentes (corto –BC–, mediano –BM– y largo –BL– plazo), dos de acciones (riesgo moderado –AM– y alto –AA–) y uno equilibrado (FE). Entre los clientes que poseen participaciones de un sólo fondo, los porcentajes de clientes en los diferentes fondos son como sigue: MM (20%), BC (15%), BM (10%), BL (5%), AA (18%), AM (25%) y FE (7%).

Se selecciona al azar un cliente que tiene acciones en un sólo fondo.

- Describe el espacio muestral Ω .
- ¿Cuál es la probabilidad de que el individuo elegido tenga participaciones del fondo equilibrado?
- ¿Cuál es la probabilidad de que el individuo un fondo de bonos?
- ¿Cuál es la probabilidad de que el individuo seleccionado no posea acciones?

Solución

Llamemos FE al suceso “el individuo tiene acciones de un fondo equilibrado”, FB al suceso “el individuo tiene acciones de un fondo de bonos” y FA al suceso “el individuo tiene acciones de un fondo de acciones”

- a. $\Omega = \{MM, BC, BM, BL, AM, AA, FE\}$.
 b. $P(FE) = 0'07$.
 c. $P(FB) = P(BC \cup BI \cup BL) = P(BC) + P(BI) + P(BL) = 0'15 + 0'1 + 0'05 = 0'3$.
 (*)

* Al ser clientes con participaciones en un sólo fondo: $BC \cap BI = \emptyset$, $BC \cap BL = \emptyset$ y $BI \cap BL = \emptyset$

- d. $P(\bar{A}) = 1 - P(A) = 1 - (0'18 + 0'25) = 0'57$

■ **Ejemplo 2.7** Un almacén de una universidad recibió un envío de 25 impresoras, de las cuales 10 son modelos láser y 15 de inyección de tinta. Si las impresoras son distinguibles y un técnico elige al azar 6 de ellas para comprobar su funcionamiento, ¿cuál es la probabilidad de que exactamente 3 de las elegidas sean impresoras láser?

Solución

Obviamente, si se eligen exactamente 3 impresoras láser, se eligen también 3 de inyección.

El número de posibilidades de elegir 3 impresoras láser es: $\binom{10}{3}$

El número de posibilidades de elegir 3 de inyección es: $\binom{15}{3}$

Aplicando la regla multiplicativa: para elegir 3 láser y 3 de inyección: $\binom{10}{3} \cdot \binom{15}{3}$.

EL número de elecciones de 6 impresoras cualesquiera de entre las 25 es: $\binom{25}{6}$

Aplicando la regla de Laplace, la probabilidad pedida es: $\frac{\binom{10}{3} \cdot \binom{15}{3}}{\binom{25}{6}} = 0,3083$ ■

2.4.1 Probabilidad Geométrica

Nota 2.4.2 Si el espacio muestral es de tipo continuo no es posible proceder de la misma manera y conduce al cálculo, generalmente mediante integrales, de las longitudes o volúmenes asociados a los sucesos y al espacio muestral.

■ **Ejemplo 2.8** Si el lugar de impacto sobre una diana de 25 cm. de radio fuese aleatorio, ¿Cuál sería la probabilidad de dar en su círculo central si este tiene 3 cm. de radio?

Solución

$$P(\text{dar en el centro}) = \frac{\text{área favorable}}{\text{área posible}} = \frac{3^2\pi}{25^2\pi} = 0'144.$$

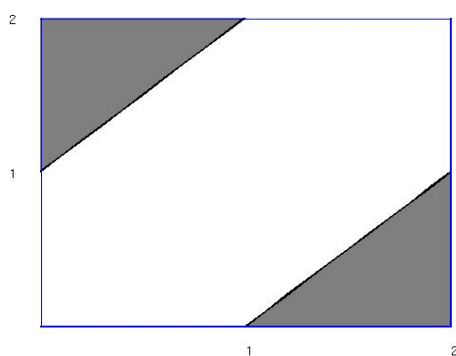
■ **Ejemplo 2.9** Un ordenador elige de forma aleatoria dos números reales del intervalo $[0, 2]$. ¿Cuál es la probabilidad de que la diferencia entre ambos sea mayor de uno?. ¿Y que el producto de ambos sea menor que dos?

Solución

$$x, y \in [0, 2] \Rightarrow 0 \leq x \leq 2; \quad 0 \leq y \leq 2.$$

$$P(\text{diferencia} > 1) = P(|x - y| > 1) = P(-1 < x - y < 1).$$

Representamos gráficamente el conjunto solución:

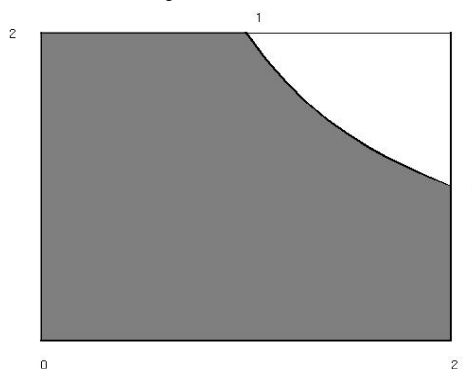


La probabilidad pedida es $P(R_1 \cup R_3) = \frac{1}{4}$.

Para que el producto de ambos sea menor que 2.

$P(xy < 2)$.

Representamos gráficamente el conjunto solución:



El área favorable es: $\text{rea} = 2 + \int_1^2 \frac{2}{x} dx = 2 + 2 \ln x \Big|_1^2 = 2 + 2 \ln 2$.

Luego la probabilidad es: $P(xy < 2) = \frac{2+2\ln 2}{4} = \frac{1+\ln 2}{2}$. ■

2.5 Probabilidad condicionada

Definición 2.5.1 Para dos sucesos cualesquiera A y B con $P(B) > 0$, la **probabilidad condicional de A dado que ocurrió B** , que denotaremos por $P(A|B)$ se define mediante la ecuación

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

Proposición 2.5.1 $P(\cdot|A)$, verifica los tres axiomas de probabilidad, así como las propiedades que se deducen de los axiomas.

Teorema 2.5.2 — Regla del Producto. Dados dos sucesos A y B se verifica

$$P(A \cap B) = P(A|B) \cdot P(B)$$

- **Ejemplo 2.10** Supongamos que lanzamos un dado. Calcular la probabilidad de:
- (a) Obtener un “2”.
 - (b) Obtener un “2” sabiendo que la puntuación que se ha obtenido es par.

Solución

Sea A el suceso “obtener un 2”.

Sea P el suceso “la puntuación obtenida es par”.

(a) $P(A) = \frac{1}{6}$

(b) La vamos a hacer de dos maneras:

(b.1) Restringiendo el espacio de probabilidad.

$$\Omega_A \equiv \{2, 4, 6\}, \text{ con lo que } P(A) = \frac{1}{3}$$

(b.2) Aplicando la fórmula anterior

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

■

El teorema puede generalizarse para una colección finita de sucesos:

Teorema 2.5.3 Dados n sucesos $A_1, \dots, A_n$

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) \cdots P\left(A_n|\bigcap_{i=1}^{n-1} A_i\right).$$

2.5.1 El teorema de Bayes

Teorema 2.5.4 — Ley de la probabilidad total. Sean $A_1, A_2, \dots, A_n$ sucesos mutuamente excluyentes y exhaustivos. Entonces para cualquier otro suceso B

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i).$$

Teorema 2.5.5 — Teorema de Bayes. Sea $A_1, A_2, \dots, A_n$ un sistema de n sucesos mutuamente excluyentes y exhaustivos con $P(A_i) > 0 \quad \forall i = 1, \dots, n$ Entonces para cualquier otro suceso B para el que $P(B) > 0$

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}, \quad k = 1, 2, \dots, n$$

Demostración. Sea k con $0 \leq k \leq n$. Por la definición de probabilidad condicionada

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{P(B)}$$

Ahora bien, si aplicamos el teorema 2.5.4 a $P(B)$ se obtiene

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{P(B)} = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(B|A_i)}, \quad k = 1, 2, \dots, n$$

■

■ **Ejemplo 2.11** Una cadena de tiendas de ordenadores instala en sus ordenadores tres sistemas operativos distintos: *Windows*, *Linux* y *McOS*. De sus ventas de ordenadores, 50 % lleva instalado sistema operativo *Windows*, 30 % llevan instalado *Linux* y el 20 % sistema operativo *McOS*. Independientemente del sistema operativo instalado, ofrece un año de asistencia técnica. Se sabe que 25 % de los ordenadores con *Windows* requieren asistencia técnica, en tanto que los porcentajes correspondientes para las marcas *Linux* y *McOS* 20 % y 10 %, respectivamente.

- ¿Cuál es la probabilidad de que un comprador elegido al azar haya comprado un ordenador con sistema operativo *Windows* que necesite asistencia técnica durante el primer año?
- ¿Cuál es la probabilidad de que un comprador elegido al azar tenga un ordenador que necesite asistencia técnica durante el primer año?
- Si un cliente vuelve a la tienda con un ordenador que necesite asistencia técnica, ¿cuál es la probabilidad de que sea un ordenador con sistema operativo *Windows*? ¿Un ordenador con *Linux*? ¿Un ordenador con *McOS*?

Solución

Sean W el suceso “el ordenador lleva instalado *Windows*”, L , “el ordenador lleva instalado *Linux*” y M “el ordenador lleva instalado *McOS*”.

Sea A el suceso “el ordenador requiere asistencia técnica durante el primer año”.

De los datos del enunciado: $P(W) = 0.5$, $P(L) = 0.3$, $P(M) = 0.2$

$P(A|W) = 0.25$, $P(A|L) = 0.2$, $P(A|M) = 0.1$

$$(a) P(W \cap A) = P(A|W) \cdot P(W) = 0.25 \cdot 0.5 = 0.125.$$

$$(b) P(A) = P(A \cap (W \cup L \cup M)) = P((A \cap W) \cup (A \cap L) \cup (A \cap M)) =$$

[al ser disjuntos] $= P(A \cap W) + P(A \cap L) + P(A \cap M) =$

$$P(A|W) \cdot P(W) + P(A|L) \cdot P(L) + P(A|M) \cdot P(M) = 0.25 \cdot 0.5 + 0.2 \cdot 0.3 + 0.1 \cdot 0.2 = 0.205$$

$$(c) P(W|A) = \frac{P(W \cap A)}{P(A)} = \frac{0.125}{0.205} = 0.61$$

$$P(L|A) = \frac{P(L \cap A)}{P(A)} = \frac{0.06}{0.205} = 0.29$$

$$P(M|A) = P(M \cap A)/P(A) = \frac{0.02}{0.205} = 0.10$$

■

2.6 Independencia de sucesos

Definición 2.6.1 Dos sucesos A y B son **independientes** si $P(A|B) = P(A)$. En caso contrario son **dependientes**.

Proposición 2.6.1 Si B es independiente de A , entonces A es independiente de B .

Proposición 2.6.2 Si A y B son sucesos independientes, también lo son A y $\bar{B}$, $\bar{A}$ y B , y $\bar{A}$ y $\bar{B}$.

Proposición 2.6.3 A y B son independientes si y sólo si

$$P(A \cap B) = P(A) \cdot P(B).$$

2.6.1 Independencia de dos o más sucesos

Definición 2.6.2 Los sucesos $A_1, A_2, \dots, A_n$ son **mutuamente independientes** si para toda k ($k = 2, \dots, n$) y todo subconjunto de índices $i_1, i_2, \dots, i_k$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k})$$

2.7 Ejercicios

Ejercicio 2.1 Sean A y B dos sucesos tales que $P(A) = 1/3$ y $P(\bar{B}) = 1/4$. Son A y B mutuamente exclusivos? ■

Ejercicio 2.2 Sean A y B dos sucesos incompatibles tales que $P(A) = \frac{1}{5}$ y $P(B) = \frac{2}{3}$. Calcular $P(\bar{A} \cap \bar{B})$. ■

Ejercicio 2.3 Consideremos los sucesos A , que el individuo seleccionado tenga una tarjeta de crédito Visa, y B , el suceso análogo para una MasterCard. Suponga que $P(A) = 0,5$, $P(B) = 0,4$ y $P(A \cap B) = 0,25$.

- Calcule la probabilidad de que el individuo seleccionado tenga al menos una de las dos tarjetas.
- ¿Cuál es la probabilidad de que el individuo elegido no tenga ninguna de esas tarjetas?
- Calcular la probabilidad de que el alumno seleccionado tenga una tarjeta Visa, pero no una MasterCard.

Solución:

- $P[A \cup B] = P[A] + P[B] - P[A \cap B] = \dots$
- $P[\overline{A \cup B}] = 1 - P[A \cup B] = \dots$
- $P[A \setminus (A \cap B)] = \dots = P[A] - P[(A \cap B)] = \dots$

■

Ejercicio 2.4 Una compañía eléctrica ofrece una reducción de la factura a cualquier familia cuyo consumo de electricidad sea menor de 240 kWh/mes . Sea A el suceso en el que una familia elegida al azar no rebasa el consumo de 240 kWh durante el mes de enero, y sea B el equivalente para el mes de julio (A y B se refieren a la misma familia). Supóngase que $P(A) = 0,8$, $P(B) = 0,7$ y $P(A \cup B) = 0,9$. Calcule lo siguiente:

- $P(A \cap B)$.

- b. Describir en términos de A y B que el consumo reducido sea rebasado en exactamente uno de los dos meses. Calcular su probabilidad.

Ejercicio 2.5 Un candidato a guardia de seguridad debe superar dos pruebas en el proceso de selección: una prueba física inicial y un examen psicotécnico posterior. La probabilidad de superar la primera prueba es 0,4, la probabilidad de aprobar la segunda es 0,5, y la probabilidad de superar al menos una de las dos es 0,6. ¿Cuál es la probabilidad de que el candidato:

- apruebe ambas pruebas?
- supere la prueba física pero no el psicotécnico?
- supere exactamente una de las dos pruebas?

Solución: Para resolver este problema, definimos el espacio muestral como el conjunto de candidatos que se presentan al proceso de selección:

$$\Omega = \{\text{candidatos al proceso de selección}\}$$

Y definimos los siguientes sucesos:

$$A = \{\text{candidato aprueba la prueba física}\}$$

$$B = \{\text{candidato aprueba el examen psicotécnico}\}$$

Sabemos que $P[A] = 0,4$, $P[B] = 0,5$ y $P[A \cup B] = 0,6$.

- a.) La probabilidad de aprobar ambas pruebas es:

$$P[A \cap B] = P[A] + P[B] - P[A \cup B] = 0,4 + 0,5 - 0,6 = 0,3$$

- b.) La probabilidad de aprobar la prueba física pero no el psicotécnico es:

$$P[A \setminus B] = P[A] - P[A \cap B] = 0,4 - 0,3 = 0,1$$

- c.) La probabilidad de aprobar exactamente una de las dos pruebas es:

$$P[(A \cup B) \setminus (A \cap B)] = P[A \cup B] - P[A \cap B] = 0,6 - 0,3 = 0,3$$

Ejercicio 2.6 Las tres opciones principales en cierto tipo de automóvil nuevo son cambio automático (A), techo corredizo (B) y GPS incorporado (C). Si el 70% de los compradores piden A , el 80% B , el 75% C , el 85% A o B , el 90% A o C , el 95% B o C y el 98% A o B o C , calcule las probabilidades de los siguientes sucesos.

- El siguiente comprador solicita una de las tres opciones.
- El siguiente comprador no selecciona ninguna de las tres opciones.
- El siguiente comprador sólo selecciona un automóvil con cambio automático y ninguna de las otras dos opciones.
- El siguiente comprador selecciona exactamente una de las tres opciones.

Ejercicio 2.7 En una empresa agrícola, los clientes pueden elegir, además, tres aditivos para enriquecer su sustrato: enriquecido con nitrógeno (N), con potasio (P) o con humus de lombriz (H). Se sabe que el 70% de los compradores seleccionan N , el 80% seleccionan P , y el 75% eligen H . Además, el 85% seleccionan N o P , el 90% seleccionan N o H , el 95% seleccionan P o H , y el 98% seleccionan al menos uno de los tres tipos (N, P o H). Calcule las probabilidades de los siguientes sucesos:

- El próximo cliente selecciona al menos uno de los tres tipos de fertilizante.
- El próximo cliente no selecciona ninguno de los tres tipos de fertilizante.
- El próximo cliente selecciona sólo fertilizante con nitrógeno y ninguno de los otros dos tipos.
- El próximo cliente selecciona exactamente uno de los tres tipos de fertilizante.

Ejercicio 2.8 Se extraen dos cartas de una baraja de 52 cartas. Calcular la probabilidad de que ambas sean copas o ambas sean reyes.

Solución: De las 52 cartas hay 13 de cada palo y 4 reyes. El espacio muestral de este problema son todos los posibles pares de cartas. Teniendo en cuenta el espacio muestral y que todos los pares de cartas son igualmente probables podemos abordar la solución de ambas cuestiones según la regla de Laplace (CF/CP).

El número de casos posibles CP es, teniendo en cuenta que no importa el orden y que no se pueden repetir cartas en el par, $CP = C_{52}^2 = \binom{52}{2}$.

El número de pares en las que ambas son copas es $CP = C_{13}^2 = \binom{13}{2}$ y el número en el que ambas son reyes $CP = C_4^2 = \binom{4}{2}$

Por tanto la respuesta a la primera pregunta es

$$P[\text{ambas sean copas}] = \frac{CF}{CP} = \frac{\binom{13}{2}}{\binom{52}{2}} = \frac{78}{1326} = 0,0586 \text{ y}$$

la respuesta a la segunda es

$$P[\text{ambas sean reyes}] = \frac{CF}{CP} = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{6}{1326} = 0,0045$$

Ejercicio 2.9 Hay 10 piezas en una caja, 4 de las cuales están pintadas. Un montador toma 3 piezas al azar. Encuentra la probabilidad de que al menos una de las piezas esté pintada.

Ejercicio 2.10 De una baraja con 48 cartas se extraen dos cartas a la vez. Hallar la probabilidad de que:

- Ambas sean “copas”.

- (b) Por lo menos una sea “copas”.
- (c) Una sea “copas” y la otra “espadas”.

Ejercicio 2.11 De una baraja de 48 cartas (4 palos y cada palo numerados del 1 al 12) se eligen 4 cartas al azar. Calcular la probabilidad de que sean de palos distintos y números consecutivos.

Ejercicio 2.12 Se considera la ecuación $ax + b = 0$, donde el coeficiente a ha sido elegido aleatoriamente en el intervalo $(1, 2)$, y b en el intervalo $(-1, 1)$. Determinar la probabilidad de que la solución de la ecuación sea mayor que $1/4$.

Ejercicio 2.13 Dos barcos llegan a un puerto el mismo día, siendo los momentos de llegada igualmente probables a lo largo del día e independientes el uno del otro. Determinar la probabilidad de que el tiempo de llegada del primer barco (contando desde las 00:00 horas de ese día) sea superior al tiempo que transcurre entre la llegada de ambos barcos.

Ejercicio 2.14 Suponga que de todos los individuos que compran determinada cámara digital, 60 % incluyen una tarjeta de memoria opcional en su compra, 40 % incluyen una batería y 30 % una tarjeta y una batería. Considere elegir al azar un comprador y sea $A = \{\text{compró una tarjeta de memoria}\}$ y $B = \{\text{compró una batería}\}$. Calcular la probabilidad de que un individuo que haya comprado una batería, compre también una tarjeta de memoria opcional.

Ejercicio 2.15 Sean A, B sucesos tales que $P(A) = 0,2$, $P(A|B) = 0,25$ y $P(B|A) = 0,5$. Calcular $P(\bar{A}|\bar{B})$.

Ejercicio 2.16 De un grupo de cuatro individuos que respondieron a una petición de un banco de sangre ninguno de ellos había donado con anterioridad, así que se desconoce su tipo de sangre. Suponga que sólo se desea el tipo $O+$ y en realidad sólo uno de los cuatro tiene este tipo. Si se eligen al azar los posibles donadores para la tipificación, ¿cuál es la probabilidad de que por lo menos se deba tipificar a tres individuos para obtener el tipo deseado?

¿Cuál es la probabilidad de que el tercer tipo sea $O+$?

Ejercicio 2.17 Sólo 1 de 1000 adultos se ve afectado por una enfermedad rara para la cual se ha desarrollado una prueba diagnóstica. La prueba es tal, que cuando un individuo en realidad tiene la enfermedad, ocurre un resultado positivo 99 % de las veces, en tanto que un individuo sin la enfermedad presenta un resultado positivo sólo 2 % de las veces. Si se aplica la prueba a un individuo seleccionado al azar y el resultado es positivo, ¿cuál es la probabilidad de que el individuo tenga la enfermedad?

Ejercicio 2.18 Una caja contiene seis bolas rojas y cuatro verdes, y una segunda caja contiene siete bolas rojas y tres verdes. Se escoge al azar una bola de la primera caja y se coloca en la segunda caja. Después se selecciona al azar una bola de la segunda caja y se pone en la primera caja.

- ¿Cuál es la probabilidad de que se seleccione una bola roja de la primera caja y una bola roja de la segunda caja?
- En la conclusión del proceso, ¿cuál es la probabilidad de que los números de bolas rojas y verdes de la primera caja sean idénticos a los del comienzo?

Solución: En este caso Ω es el conjunto de todos los pares (ordenados) de bolas rojas y verdes. Teniendo esto en cuenta, notamos por R_i (V_i respect.) al conjunto de pares que tienen una bola Roja (Verde respect.) de la caja i -ésima.

- $P[R_1 \cap R_2] = P[R_2|R_1] * P[R_1] = 8/11 * 6/10$
- $P[(R_1 \cap R_2) \cup (V_1 \cap V_2)] \underset{\text{disj.}}{=} P[R_1 \cap R_2] + P[V_1 \cap V_2] = \dots$

■

Ejercicio 2.19 En cierta gasolinera, 40% de los clientes utilizan gasolina normal (A_1), 35% usan gasolina extra (A_2), y 25% gasolina premium (A_3). De los clientes que consumen gasolina normal, sólo 30% llenan sus tanques (suceso B). De los que compran gasolina extra, 60% llenan sus tanques, en tanto que quienes llevan gasolina premium, 50% llenan sus tanques.

- ¿Cuál es la probabilidad de que el siguiente cliente pida gasolina extra y llene su tanque ($A_2 \cap B$)?
- ¿Cuál es la probabilidad de que el siguiente cliente llene el tanque?
- Si el siguiente cliente llena el tanque, ¿cuál es la probabilidad de que pida gasolina regular? ¿Extra? ¿Premium?

■

Ejercicio 2.20 El 70% de las avionetas que desaparecen mientras vuelan son rescatados después. De las avionetas rescatadas, el 60% tienen localizador de emergencia, mientras que el 90% de las no rescatadas no tienen un localizador. Suponed que desaparece una avioneta.

- Si tiene localizador, ¿cuál es la probabilidad de que no sea rescatado?
- Si no tiene localizador, ¿cuál es la probabilidad de que sea rescatado?

■

Ejercicio 2.21 — El problema del caballero De Meré. El caballero De Meré, aficionado a llevar un largo registro de los resultados de sus timbas, había observado que era ventajoso jugar a los dados tirando uno y apostando a la par que sacaría al menos un seis en cuatro tiradas. En cambio, le era desventajoso extender el juego echando dos dados y apostando a sacar al menos un doble seis en veinticuatro tiradas. Sin embargo, él razonaba que siendo las proporciones entre el número de seises a obtener y las tiradas las mismas, la probabilidad debería ser también la misma.

¿Tenía razón en su razonamiento?

■

Ejercicio 2.22 Sean A y B dos sucesos cualesquiera.

- a) Si $P[A] = 1/3$, $P[B] = 1/5$ y $P[A|B] + P[B|A] = 2/3$. Calcular $P[\overline{A} \cap \overline{B}]$.
 b) Si $P[A \cup B] = 0,8$, $P[A \cap B] = 0,3$ y $P[B \cap \overline{A}] = 0,4$, calcular $P[B|A]$.

Solución:

$$a) P[\overline{A} \cap \overline{B}] = P[\overline{A \cup B}] = 1 - P[A \cup B] = 1 - (P[A] + P[B] - P[A \cap B]) = **$$

Conocemos dos de los tres sumandos, nos falta saber cuánto vale $P[A \cap B]$. Para calcularlo utilizaremos el dato $P[A|B] + P[B|A] = 2/3$.

$$2/3 = P[A|B] + P[B|A] = \frac{P[A \cap B]}{P[B]} + \frac{P[B \cap A]}{P[A]}$$

si multiplicamos todo por $P[A] * P[B]$ tenemos que

$$2/3 * P[A] * P[B] = P[A \cap B] * P[A] + P[B \cap A] * P[B]$$

y despejando $P[A \cap B]$ tenemos que

$$P[A \cap B] = 2/3 * \frac{P[A] * P[B]}{P[A] + P[B]} = 2/3 * \frac{1/3 * 1/5}{1/3 + 1/5} = 1/12$$

por lo que basta con sustituir en la primera línea para obtener el resultado.

$$P[\overline{A} \cap \overline{B}] = 1 - (P[A] + P[B] - P[A \cap B]) = 1 - (1/3 + 1/5 - 1/12)$$

- b) $0,4 = P(B \cap \overline{A}) = P[B] - P[A \cap B]$; despejando obtenemos que $P[B] = 0,7$
 Utilizando, ahora, la regla del producto tenemos que

$$P[B|A] = \frac{P[A \cap B]}{P[A]} = 3/7$$

■

Ejercicio 2.23 Sean A , B y C sucesos independientes tales que $P[A] = 0,3$, $P[B] = 0,6$ y $P[C] = 0,7$. Calcular $P[A \cap (\overline{B} \cup \overline{C})]$.

Solución:

$$\begin{aligned} P[A \cap (\overline{B} \cup \overline{C})] &= P[A \cap (\overline{B \cap C})] = \\ &= P[A] - P[A \cap B \cap C] = \quad \quad \quad [\text{indep.}] \\ &= P[A] - P[A] \cdot P[B] \cdot P[C] \end{aligned}$$

■

Ejercicio 2.24 Una empresa que fabrica móviles produce un modelo básico y uno de lujo. El año pasado, 40% de las móviles vendidos fueron del modelo básico. De los compradores del modelo básico, 30% compran una garantía ampliada, en tanto que 50% de los compradores del modelo de lujo hacen lo mismo. Si se sabe que un

comprador seleccionado al azar tiene una garantía ampliada, ¿qué tan probable es que tenga un modelo básico? ■

Ejercicio 2.25 La junta de un avión requiere 25 remaches. La junta se tiene que volver a hacer si está defectuoso cualquiera de los remaches. Suponga que cada uno de los remaches tiene la misma probabilidad de estar defectuoso y que no depende de los demás. Si el 20 % de las juntas se tiene que volver a hacer, ¿cuál es la probabilidad de que un remache esté defectuoso?

Solución: Si nos fijamos bien en el enunciado vemos que nos dan tanto la probabilidad de que un remache sea defectuoso como de que una junta haya que revisarla. En este contexto, puede que en una primera impresión planteemos que Ω son los remaches $\Omega = \{1, 2, \dots, 25\}$ pero con este espacio muestral no podemos expresar que una junta tenga que ser revisada o que no tenga que serlo. Sin embargo si pensamos en Ω como el conjunto de las diferentes juntas fabricadas, es fácil pensar en el conjunto de juntas que estan bien (o mal)

$$C = \{\text{juntas correctas}\}$$

y también podemos plantear el hecho de que un determinado remache, el remache i , sea defectuoso, R_i

$$R_i = \{\text{juntas en las que el remache } i \text{ es defectuoso}\}$$

Aproximación 1

$$0,2 = P[\bar{C}] = P[\text{juntas con algún remache (o varios) defectuosos}]$$

El conjunto de juntas defectuosas tiene una expresión muy compleja(extensa) en términos de R_i , ya que sería

$$\begin{aligned} \bar{C} = \{ & R_1 \cup R_2 \cup R_3 \cup \dots \cup R_{25} \cup \\ & (R_1 \cap R_2) \cup (R_1 \cap R_2) \cup \dots \cup (R_{24} \cap R_{25}) \cup \\ & (R_1 \cap R_2 \cap R_3) \cup (R_1 \cap R_2 \cap R_4) \cup \dots \cup (R_{23} \cap R_{24} \cap R_{25}) \cup \\ & \dots \cup \\ & (R_1 \cap \dots \cap R_{25}) \} \end{aligned}$$

Los cálculos serían muy largos y tediosos, así que lo abordaremos mediante otra aproximación.

Aproximación 2

$$0,8 = P[C] = P[\text{juntas correctas}] = P[\bar{R}_1 \cap \dots \cap \bar{R}_{25}] = *$$

pero los remaches fallan de forma independiente, luego

$$* = P[\overline{R_1}] \cdot \dots \cdot P[\overline{R_{25}}] = *$$

y todos tienen la misma probabilidad de fallar

$$* = P[\overline{R_1}] \cdot \dots \cdot P[\overline{R_1}] = P[\overline{R_1}]^{25}$$

Resolviendo, llegamos a que

$$P[\overline{R_1}] = \sqrt[25]{0,8} = 0,9911$$

y por tanto

$$P[R_i] = 0,0081$$



Ejercicio 2.26 El 60 % de los automóviles examinados en un centro de verificación pasan la prueba. Si se supone que automóviles sucesivos pasan la prueba, o no, sin dependencia mutua. Si observamos tres automóviles contiguos, calcular las siguientes probabilidades:

- $P(\text{pasan los tres automóviles})$.
- $P(\text{por lo menos es rechazado uno de los tres automóviles inspeccionados})$.
- $P(\text{exactamente pasa uno de los tres automóviles inspeccionados})$.
- $P(\text{a lo sumo pasa uno de los tres automóviles inspeccionados})$.
- Dado que por lo menos uno de los tres automóviles pasa la verificación, ¿cuál es la probabilidad de que pasen los tres (una probabilidad condicional)?



Ejercicio 2.27 El profesor Pérez puede tomar una de dos rutas en su camino de regreso a casa desde el trabajo. En la primera ruta hay cuatro cruces de ferrocarril. La probabilidad de que se detenga ante el paso de un tren en cualquiera de los cruces es 0.1, y los trenes operan de manera independiente en los cuatro cruces. La otra es más larga pero sólo hay dos cruces, independientes entre sí, con la misma probabilidad de detención para cada uno que en la primera ruta. En determinado día, el profesor Pérez tiene una junta en casa programada para cierta hora. Sin importar qué ruta tome, calcula que llegará tarde si el tren se cruza en su camino por lo menos en la mitad de los cruces.

- ¿Qué ruta debe tomar para reducir la probabilidad de llegar tarde a la junta?
- Si lanza al aire una moneda para determinar la ruta y llega tarde, ¿cuál es la probabilidad de que haya tomado la ruta con cuatro cruces?



2.7.1 Ejercicios de exámenes de años anteriores

Ejercicio 2.28 Un médico aplica un test para detectar una enfermedad cuya incidencia (probabilidad a priori de padecer la enfermedad) es del 10 %. La probabilidad de que el

test dé positivo al aplicarlo sobre un enfermo es del 80% mientras que la probabilidad de que el test dé negativo sobre una persona sana es del 75%.

- (a) Calcular la probabilidad de que el test dé positivo al aplicarlo sobre un niño elegido al azar.
- (b) Calcular la probabilidad de que el test suministre un resultado incorrecto.

Ejercicio 2.29 En un laboratorio se diseña un test para detectar la presencia de una bacteria en el agua. Para probar el test se considera un gran número de probetas con agua que pueden, o no, contener la bacteria. La probabilidad de que una probeta escogida al azar contenga la bacteria es de 0.2. Por otra parte, si una probeta contiene la bacteria, el test da positivo en el 90% de los casos. En cambio, si una probeta no contiene la bacteria, el test da positivo en el 5% de los casos.

- (a) Al escoger al azar una probeta, ¿cuál es la probabilidad de que dé positivo el test?
- (b) Si una probeta ha dado positivo en el test, ¿cuál es la probabilidad de que contenga la bacteria?
- (c) Entre las probetas que han dado negativo en el test, ¿cuál es la proporción de probetas que contienen la bacteria?

Ejercicio 2.30 En una granja avícola existen tres razas de gallinas: A , B y C . De la raza A hay 300 aves, de la B 600 y de la C 100. La probabilidad de que una gallina de la raza A ponga huevos de calidad inferior a la tolerada es igual a 0.2 y en la raza B es 0.15. Se desconoce cuál es esta probabilidad en la raza C ; sin embargo, se han establecido tres hipótesis sobre la probabilidad de que, tomando una raza al azar y siendo el huevo de baja calidad, proceda de la raza C ; estas probabilidades son 0.6, 0.5 y 0.3. Verificar qué hipótesis puede/n ser correcta/s.

Ejercicio 2.31 En una determinada población el número de fumadores respecto de los que no fuman está en la proporción de $3/5$. Por los datos experimentales se sabe que la probabilidad de que un fumador padezca una enfermedad determinada es de 0.3 mientras que de padecerla un no fumador es solo del 0.025.

- (a) Si se eligen tres individuos al azar, ¿Cuál es la probabilidad de que alguno de ellos padezca la enfermedad?
- (b) Una determinada persona padece dicha enfermedad, ¿Cuál es la probabilidad de que sea fumador?

Ejercicio 2.32 El profesor Pérez olvida poner su despertador 3 de cada 10 días. Además, ha comprobado que uno de cada 10 días en los que pone el despertador acaba no levantándose a tiempo de dar su primera clase, mientras que en 2 de cada 10 días en los que se olvida poner el despertador, llega a tiempo a dar su primera clase.

- (a) ¿Cuál es la probabilidad de que el profesor Pérez llegue a tiempo a dar su primera clase?
- (b) Si un día no ha llegado a tiempo, ¿qué probabilidad hay de que olvidase poner el despertador la noche anterior?

Ejercicio 2.33 En un sistema de alarma, la probabilidad de que se produzca un peligro es de 0.1. Si éste se produce, la probabilidad de que la alarma funcione es 0.95. La probabilidad de que funcione la alarma sin haber peligro es 0.03.

- (a) Hallar la probabilidad de que habiendo funcionado la alarma, no haya habido peligro.
- (b) Hallar la probabilidad de que haya un peligro y la alarma no funcione.
- (c) Hallar la probabilidad de que no habiendo funcionado la alarma, haya un peligro.

Ejercicio 2.34 Un sistema de comunicaciones dispone de un transmisor que envía los símbolos 0, 1 y 2 a través de un canal hasta un receptor. Este canal puede cambiar el símbolo transmitido. La probabilidad de que se reciba el mismo símbolo que se ha transmitido es de 0.8. Por otra parte la probabilidad de recibir de forma defectuosa un símbolo es igual sea cual sea el símbolo transmitido. Las probabilidades de transmisión de los símbolos 0, 1 y 2 son de 0.5, 0.3 y 0.2 respectivamente.

- (a) Calcular las probabilidades de recibir 0, 1 y 2.
- (b) Si se ha recibido un 1, calcular la probabilidad de que realmente se haya enviado un 1.

Ejercicio 2.35 Se lanza un dado y se toman tantas monedas como indique el resultado obtenido. A continuación se lanzan las monedas. Se pide:

- (a) ¿Cuál es la probabilidad de obtener tres caras?
- (b) Sabiendo que se han obtenido tres caras, ¿cuál es la probabilidad de que se hubiera obtenido un cuatro en el lanzamiento del dado?

Ejercicio 2.36 Un dado está trucado de manera que la probabilidad de que salga un número es proporcional a dicho número. Se lanza dicho dado, si sale par se saca una bola de una urna con 3 bolas blancas, 2 rojas y 1 verde, si sale impar se saca una bola de otra urna que tiene 4 bolas blancas 3 rojas y 2 verdes.

- (a) Obtener la probabilidad de sacar una bola roja.
- (b) Si se extrae una bola verde, ¿cuál es la probabilidad de que el número obtenido en el lanzamiento del dado sea impar?

Ejercicio 2.37 Se tienen 2 urnas: la urna A que contiene 3 bolas rojas, 6 blancas y 7 negras; la urna B contiene 5 bolas rojas y 4 bolas blancas. Para decidir de qué urna se va a hacer la extracción se lanzan dos dados corrientes y si se obtiene una suma de 6 entre los dos dados se escoge una bola de la urna B y en caso contrario se escoge de la urna A . Hallar la probabilidad de que:

- (a) Se obtenga una bola roja;
- (b) Si se saca una bola y es roja, ¿cuál es la probabilidad de que proceda de la urna B ?

Ejercicio 2.38 Por un estudio encargado por el partido “*Seguro Que Ganamos*” (SQG) se obtiene la siguiente información: el 17 % de la población tiene estudios superiores, el 44 % estudios medios, el 30 % estudios primarios y el 9 % no tiene estudios. De entre los de estudios superiores el 25 % votan al partido SQG, entre los de estudios medios el 35 %, entre los de estudios primarios el 22 % y entre los que no tienen estudios votan partido SQG el 18 %. Si se extrae un sujeto al azar, obtenga las siguientes probabilidades:

- (a) Que sea titulado superior sabiendo que vota al SQG.
- (b) Que sea persona sin estudios que vota al SQG;
- (c) Que sea una persona con estudios primarios o que no vote al SQG.

Ejercicio 2.39 Existen dos métodos A y B para enseñar a los trabajadores de una empresa cierto procedimiento industrial. El porcentaje de fracasos es del 20 % para el método A y del 10 % para B . Sin embargo, B es más costoso que A y se utiliza sólo en el 30 % de las ocasiones, mientras que A se utiliza en el 70 % restante. Se entrenó a un trabajador según uno de los métodos pero no logró aprenderlo correctamente. ¿Cuál es la probabilidad de que recibiera el entrenamiento con el método A ?

Ejercicio 2.40 Sean A y B dos sucesos cualesquiera.

- (a) Supongamos que $P(A) = 1/3$, $P(B) = 1/5$ y $P(A|B) + P(B|A) = 2/3$. Calcular $P(\overline{A} \cap \overline{B})$.
- (b) Si $P(A \cup B) = 0,8$, $P(A \cap B) = 0,3$ y $P(B \cap \overline{A}) = 0,4$. Calcular $P(B|A)$.

Ejercicio 2.41 Hay tres cajas A , B y C con 20 piezas cada una, conteniendo 20, 15 y 10 piezas buenas respectivamente. La probabilidad de elegir la caja A es igual a la de elegir la caja B , y la de elegir la caja C es igual a la suma de esas dos probabilidades. Eligiendo al azar una caja se extraen con reposición dos piezas. Calcular:

- (a) la probabilidad de que ambas piezas sean buenas.
- (b) la probabilidad de que si las dos piezas resultan ser buenas provengan de la caja A .

Ejercicio 2.42 Cierta test para determinar el factor Rh tiene probabilidad 0.95 de dar el resultado correcto cuando se aplica sobre muestras de sangre Rh^- y probabilidad 0.8 de dar el resultado correcto cuando se aplica sobre muestras Rh^+ . Si la población es 84 % Rh^+ , hallar la probabilidad de que si el resultado del test

- (a) es Rh^+ , el individuo lo sea.
- (b) es Rh^- , el individuo también lo sea.

Ejercicio 2.43 En una granja avícola se utilizan 2 tipos de piensos, A y B , para alimentar a las aves. El 25 % de las aves son alimentadas exclusivamente con el pienso A , el 35 % son alimentadas exclusivamente con el pienso B y el resto de las aves son alimentadas con una mezcla de ambos tipos de pienso. Se sabe que la probabilidad de que el engorde de las aves sea superior a 1 Kg. cuando se utiliza solamente el tipo A es de 0.86, cuando se utiliza solamente el pienso B es de 0.58 y cuando se utilizan ambos tipos de pienso es del 0.92. Se decide escoger al azar una de las aves de la granja:

- (a) Determinar la probabilidad de que el engorde del ave sea superior a 1 Kg.
- (b) Si se comprueba que el engorde del ave ha superado 1 Kg., determinar qué tipo de alimentación es más probable que haya seguido y con qué probabilidad.

Ejercicio 2.44 Un dado está trucado de manera que la probabilidad de que salga un número es proporcional a dicho número. Se lanza dicho dado, si sale par se saca una bola de una urna con 3 bolas blancas, 2 rojas y 1 verde, si sale impar se saca una bola de otra urna que tiene 4 bolas blancas 3 rojas y 2 verdes.

- (a) Obtener la probabilidad de sacar una bola roja.
- (b) Si se extrae una bola verde, ¿cuál es la probabilidad de que el número obtenido en el lanzamiento del dado sea impar?



3. Variables aleatorias y distribuciones de

3.1 Variables aleatorias

Definición 3.1.1 Dado un espacio de muestral Ω , una **variable aleatoria** (v.a.) es toda función $X : \Omega \rightarrow \mathbb{R}$; esto es, toda función cuyo dominio es el espacio muestral y cuyo recorrido es el conjunto de los números reales.

■ **Ejemplo 3.1**

- Lanzamos una moneda tres veces y definimos una función que a cada resultado le asigna el número de caras obtenidas.
 - Se dispara a una diana y se asigna a cada impacto la distancia desde éste hasta el centro de la diana.
-

Definición 3.1.2 Una v.a. **discreta** es una v.a. cuya imagen es un conjunto finito o bien infinito numerable. Una v.a. es **continua** si su imagen contiene un intervalo de la recta real.

Proposición 3.1.1 Si X es una v.a. sobre Ω y f es una función continua, $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ tal que $Im(X) \subseteq D$, entonces $f \circ X : \Omega \rightarrow \mathbb{R}$ es una v.a. que denotaremos por $f(X)$.

$$f(X)(\omega) = (f \circ X)(\omega) = f(X(\omega)) \quad \forall \omega \in \Omega$$

En particular, si X e Y son vv.aa. sobre Ω y $a, b \in \mathbb{R}$, entonces aX , $aX + bY$ y $X \cdot Y$ son vv.aa.

3.2 Distribuciones de probabilidad para variables aleatorias discretas

Definición 3.2.1 La **función de probabilidad** o **función de distribución de probabilidad** (*fdp*) de una v.a. discreta X es una función que a todo $x \in \mathbb{R}$ asocia $P(x) \in [0, 1]$ con

$$P(x) = P[X = x] = P(\{\omega \in \Omega : X(\omega) = x\})$$

Se verifica que $\sum_i P(x_i) = 1$ y que $P(x) = 0$ si $x \notin X(\Omega)$

■ **Ejemplo 3.2** Consideremos el experimento de tirar un par de dados. Sea Y la variable aleatoria que muestra la cantidad total obtenida en una tirada. Entonces la *fdp* de Y es:

- $p(2) = P[Y = 2] = P[\{(1, 1)\}] = 1/6 \cdot 1/6 = 1/36$
- $p(3) = P[Y = 3] = P[\{(1, 2), (2, 1)\}] = 2 \cdot (1/6 \cdot 1/6) = 1/18$
- $p(4) = P[Y = 4] = P[\{\dots\}] = 3 \cdot (1/6 \cdot 1/6) = 1/12$
- ...
- $p(11) = P[Y = 11] = P[\{(5, 6), (6, 5)\}] = 2 \cdot (1/6 \cdot 1/6) = 1/18$
- $p(12) = P[Y = 12] = P[\{(6, 6)\}] = 1/6 \cdot 1/6 = 1/36$
- $p(x) = 0$ si $x \notin \{2, 3, 4, \dots, 10, 11, 12\}$

■

Definición 3.2.2 Sea $p(x; v)$ una fdp que depende de una cantidad $v \in \vartheta$. Ese valor v se denomina **parámetro de la distribución** y la colección de las diferentes distribuciones generadas por los parámetros $v \in \vartheta$ se denomina **familia de distribuciones de probabilidad**.

■ **Ejemplo 3.3** Consideremos el experimento de tirar una moneda hasta que salga cara. La moneda está trucada de modo que la probabilidad de obtener una cara es v . Sea Y la variable aleatoria que muestra la cantidad total de tiradas necesarias. Entonces la *fdp* de Y es

- $p(1) = P[Y = 1] = P[\{c\}] = v$
- $p(2) = P[Y = 2] = P[\{x, c\}] = (1 - v) \cdot v$
- $p(3) = P[Y = 3] = P[\{x, x, c\}] = (1 - v) \cdot (1 - v) \cdot v$
- $p(4) = P[Y = 4] = P[\{x, x, x, c\}] = (1 - v)^3 \cdot v$
- ...

$$p(n) = P[Y = n] = \begin{cases} (1 - v)^{(n-1)} \cdot v & \text{si } n \in \{1, 2, 3, \dots\} \\ 0 & \text{si } n \notin \{1, 2, 3, \dots\} \end{cases}$$

y v es el **parámetro de la distribución**. ■

Definición 3.2.3 Dada una v.a. discreta X , la función $F_X : \mathbb{R} \rightarrow [0, 1]$,

$$F_X(x) = P[X \leq x] = \sum_{-\infty < y \leq x} P(y)$$

se denomina **función de distribución** o **función de acumulación** (*fda*) asociada a X . Para cualquier número x , $F(x)$ es la probabilidad de que el valor observado de X sea menor o igual que x .

■ **Ejemplo 3.4** Si retomamos el ejemplo anterior con $p = 1/2$, la *fdp* de Y es

$$p(x) = \begin{cases} 0 & \text{si } x \notin \mathbb{N} \\ (1/2)^x & \text{si } x \in \mathbb{N} \end{cases}$$

y por tanto la *fda* de Y es:

- Si $x < 1$, $F_Y(x) = 0$
 - Si $1 \leq x < 2$, $F_Y(x) = 1/2$
 - Si $2 \leq x < 3$, $F_Y(x) = 1/2 + 1/4$
 - Si $3 \leq x < 4$, $F_Y(x) = 1/2 + 1/4 + 1/8$
 - ...
 - Si $n \leq x < n + 1$, $F_Y(x) = \sum_{k=1}^n (1/2)^k$
-

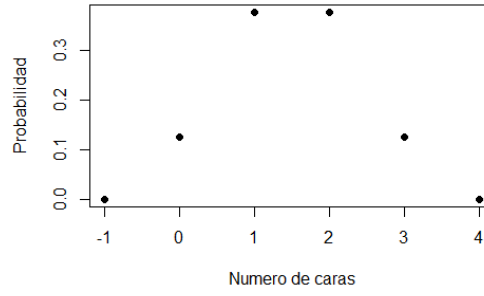


Figura 3.1: Función de probabilidad del Ejemplo 3.5

Proposición 3.2.1 Sea F_X la fda asociada a una v.a. aleatoria X y sean $a, b \in \mathbb{R}$ con $a \leq b$.

- F_X es escalonada y los puntos de salto son los $x_i \in X(\Omega)$.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$; $\lim_{x \rightarrow \infty} F_X(x) = 1$.
- F_X es no decreciente.
- F_X es continua por la derecha.
- $P[a \leq X \leq b] = F(b) - F(a^-)$.

■ **Ejemplo 3.5** Se lanza una moneda tres veces y anotamos el número de caras.

Si lanzamos una moneda tres veces y anotamos el número de caras que pueden aparecer, tenemos las siguientes posibilidades: $\{0, 1, 2, 3\}$. Es decir, nuestra variable aleatoria X cuyo valor es el *número de caras obtenidas* toma los valores: $X = \{0, 1, 2, 3\}$.

La función de probabilidad es:

- $p(0) = P[X = 0] = PR_3^3 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{8}$
- $p(1) = P[X = 1] = PR_3^{2,1} \cdot \left(\frac{1}{2}\right)^3 = \frac{3}{8}$
- $p(2) = P[X = 2] = PR_3^{1,2} \cdot \left(\frac{1}{2}\right)^3 = \frac{3}{8}$
- $p(3) = P[X = 3] = PR_3^3 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{8}$
- $p(k) = P[X = k] = 0 \quad \forall k \notin \{0, 1, 2, 3\}$

La función de acumulación es:

$$F_X(x) = P[X \leq x] = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{8} & \text{si } 0 \leq x < 1 \\ \frac{1}{2} & \text{si } 1 \leq x < 2 \\ \frac{7}{8} & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$

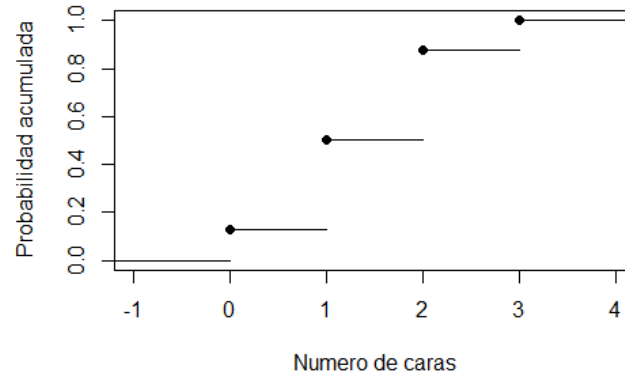


Figura 3.2: Función de acumulación del Ejemplo 3.5

La gráfica de la función de distribución puede verse en la Figura 3.2 y la gráfica de la función de probabilidad puede verse en la Figura 3.1.

■

3.3 Variables aleatorias continuas y funciones de densidad.

Definición 3.3.1 Sea X v.a. continua. Entonces la **distribución de probabilidad** o **función de densidad** de probabilidad (*fdd*) de X es una función $f_X : \mathbb{R} \rightarrow \mathbb{R}$ tal que

- $f_X(x) \geq 0 \quad \forall x \in \mathbb{R},$
- $\int_{-\infty}^{\infty} f_X(x) dx = 1,$
- $P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad , \forall a, b \in \mathbb{R} \text{ con } a \leq b.$

Definición 3.3.2 Una v.a. X con *fdp*

$$f_X(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{en caso contrario.} \end{cases}$$

se dice que tiene una **función de distribución uniforme** en el intervalo $[A, B]$; donde A y B son los parámetros de la familia de distribuciones.

Proposición 3.3.1 Sea X es una v.a. continua, entonces para cualquier número c se tiene que $P(X = c) = 0$. Además, para dos números cualesquiera a y b con $a < b$, se cumple que

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

Definición 3.3.3 Dada una v.a. continua X y f_X su *fdp*, la función

$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f_X(t) dt$$

se denomina **función de distribución acumulada** o **función de acumulación** (*fda*) de la v.a.

Proposición 3.3.2 Sea X es una v.a. continua con *fda* F_X . Entonces para cualquier par de números a y b con $a < b$, se tiene que

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$,
- $\lim_{x \rightarrow \infty} F_X(x) = 1$,
- $P[X > a] = 1 - F_X(a)$ y
- $P[a \leq X \leq b] = F_X(b) - F_X(a)$.

Proposición 3.3.3 Sea X es una v.a. continua con *fdp* f_X y *fda* F_X . Entonces se verifica que F_X es continua no decreciente y que para todo x real, $f_X(x) = F'_X(x)$.

Ejercicio 3.1 Supongamos que la *fdd* de X viene dada por

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en caso contrario} \end{cases}$$

- Representar $f(x)$.
- Calcular y representar la *fda* $F(x)$.
- Calcular $P(1 \leq x \leq 1,5)$.
- Calcular $P(x > 1)$.

■

3.4 Características asociadas a una variable aleatoria

Definición 3.4.1 El **valor esperado, valor medio, media o esperanza** de X , denotado por $E[X]$ o μ_X , es

Caso discreto: si X es una v.a. que toma un conjunto D de valores posibles y tiene una fdp $p(x) = P[X = x]$,

$$\mu_X = E[X] = \sum_{x \in D} x \cdot P[X = x].$$

Caso continuo: si X es una v.a. con fdp $f(x)$,

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

■ **Ejemplo 3.6** Calcular $E[X]$ de la variable aleatoria X con fdp

x	3	4	5	6	7	8	9
$p(x)$	0,01	0,03	0,13	0,25	0,39	0,17	0,02

Solución:

$$E[X] = \sum_{i=1}^7 (i+2) \cdot p(i) = 3 \cdot 0,01 + \dots + 9 \cdot 0,02 = 6,57$$

■

■ **Ejemplo 3.7** Calcular $E[X]$ para la v.a. X con fdd

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Solución:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^1 x \cdot \frac{3}{2}(1-x^2) dx = \\ &= \frac{3}{2} \int_0^1 (x-x^3) dx = \left[\frac{3}{2} \left(\frac{x^2}{2} - \frac{x^4}{4} \right) \right]_{x=0}^{x=1} = 3/8 \end{aligned}$$

■

Definición 3.4.2 Dadas la v.a. X y una función $g(X)$ de dicha v.a., la esperanza de $g(X)$ ($E[g(X)]$) es

$$E[g(X)] = \begin{cases} \sum_{x \in D} g(x) \cdot p(x) & X \text{ discreta} \\ \int_{-\infty}^{\infty} g(x) \cdot f(x) dx & X \text{ continua} \end{cases}$$

Propiedades

1. $E[aX + b] = aE[X] + b.$
2. $E[g(X) + h(X)] = E[g(X)] + E[h(X)].$

$$3. E[aX + bY] = aE[X] + bE[Y].$$

Definición 3.4.3 Sea X una v.a. con fdp $p(x)$ (fdd $f(x)$) y esperanza μ . Entonces la **varianza** de X , denotada por $Var(X)$ o σ_X^2 es

Caso discreto: $Var(X) = E[(X - \mu)^2] = \sum_{i \geq 1} (x_i - \mu)^2 \cdot P[X = x_i]$

Caso continuo: $Var(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$

La **desviación estandar** (DE) de X es $\sigma_X = \sqrt{\sigma_X^2}$

Proposición 3.4.1

$$Var(X) = E[X^2] - (E[X])^2 = E[X^2] - \mu_X^2.$$

Propiedades

1. $Var[X + b] = Var[X]$.
2. $Var[aX + b] = a^2 Var[X]$.
3. Si X e Y son independientes entonces $Var[X + Y] = Var[X] + Var[Y]$.
4. $Var(h(X)) = \sigma_{h(X)}^2 = E[(h(X) - \mu(h(X)))^2]$.

Definición 3.4.4 La **moda** (M_o) de una v.a. X es:

Caso discreto: $M_o = x_k$ si $P[X = x_k] = \max P[X = x_i]$

Caso continuo: $M_o = x_0$ si $f(x_0) = \max f(x)$

Definición 3.4.5 La **mediana** de una v.a. X es:

Caso discreto: $M_e = x_k$ si $F(x_{k-1}) < \frac{1}{2}$ y $F(x_k) \geq \frac{1}{2}$

Caso continuo: $M_e = x_0$ si $F(x_0) = \frac{1}{2} \left(= \int_{-\infty}^{x_0} f(t) dt \right)$

■ **Ejemplo 3.8** Calcular $V[X]$ de la variable aleatoria del ejemplo 3.6

x	3	4	5	6	7	8	9
$p(x)$	0,01	0,03	0,13	0,25	0,39	0,17	0,02

Solución:

$$\begin{aligned}
 V[X] &= E[X^2] - (E[X])^2 = \sum_{i=3}^9 i^2 \cdot p(i) - (E[X])^2 = \\
 &= 3^2 \cdot 0,01 + 4^2 \cdot 0,03 + \dots + 9^2 \cdot 0,02 - (6,57)^2 = \\
 &= 44,43 - 43,1649 = 1,2651
 \end{aligned}$$

■ **Ejemplo 3.9** Calcular $V[X]$ de la variable aleatoria del ejemplo 3.7

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Solución:

$$\begin{aligned}
 V[X] &= E[X^2] - (E[X])^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - (3/8)^2 = \\
 &= \int_0^1 x^2 \cdot \frac{3}{2}(1-x^2) dx - \left(\frac{3}{8}\right)^2 = \\
 &= (1/5) - (3/8)^2 = 0,059
 \end{aligned}$$

Ejercicio 3.2 Sea X la v.a. que representa el número de defectos importantes en un determinado electrodoméstico cuya *fdp* es

x	0	1	2	3	4
$p(x)$	0,08	0,15	0,45	0,27	0,05

Calcule lo siguiente

- $E(X)$.
- $V(X)$ directamente de la definición.
- La desviación estándar de X .
- $V(X)$ con la fórmula abreviada.

3.5 Modelo Uniforme Discreto

Definición 3.5.1 Una v.a. X se distribuye según un **modelo uniforme discreto** de parámetro N , y se denota por $X \sim U(N)$, si su *fdp* es

$$P[X = k] = \begin{cases} \frac{1}{N} & \text{si } k \in \{1, 2, \dots, N\} \\ 0 & \text{en caso contrario} \end{cases}$$

Proposición 3.5.1 Si $X \sim U(N)$, entonces

$$E[X] = \frac{N+1}{2} \quad \text{y} \quad \text{Var}[X] = \frac{(N+1)(N-1)}{12}.$$

- **Ejemplo 3.10** Experimentos con función de probabilidad uniforme discreta:
- Observar el resultado de lanzar un dado equilibrado.
 - Observar el color de una bola extraída de una urna con 4 bolas rojas, 4 bolas azules y 4 bolas verdes.
 - Observar las dos últimas cifras de la matrícula del próximo vehículo que entre en la autovía.
-

3.6 Modelo de Bernoulli

Definición 3.6.1 Un experimento de Bernoulli es aquél que sólo puede dar lugar a dos resultados: “éxito”(E), con probabilidad p o “fracaso”(F), con probabilidad $q = 1 - p$.

Definición 3.6.2 Una v.a. X se distribuye según un **modelo de Bernoulli** de parámetro p , y se escribe $X \sim Be(p)$, si su *fdp* viene dada por:

$$P[X = k] = \begin{cases} p^k(1-p)^{1-k} & \text{si } k \in \{0, 1\} \\ 0 & \text{en otro caso} \end{cases}$$

Proposición 3.6.1 Si $X \sim Be(p)$, entonces $E[X] = p$ y $\text{Var}[X] = pq$.

- **Ejemplo 3.11** Experimentos con función de probabilidad de tipo Bernoulli:
- Observar el resultado de lanzar una moneda trucada.
 - Observar el color de una bola extraída de una urna con 4 bolas rojas y 8 bolas azules.
 - Observar si el próximo vehículo que entre en la autovía es un turismo o un camión.
-

3.7 Modelo binomial

Definición 3.7.1 Sean $X_1, X_2, \dots, X_n$ vv.aa. independientes e idénticamente distribuidas según un modelo $Be(p)$. La v.a. $X = \sum_{i=1}^n X_i$ se distribuye según un **modelo binomial** de parámetros n y p , y se denota $X \sim B(n, p)$.

$X \sim B(n, p)$, X representa el número de éxitos obtenidos en una sucesión de n experimentos de Bernoulli.

Proposición 3.7.1 Sea $X \sim B(n, p)$, entonces

1. La fdp de X es

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

2. $E[X] = np$ y $Var[X] = np(1-p)$.

■ **Ejemplo 3.12** Experimentos con función de probabilidad de tipo Binomial:

- Observar el número de caras obtenidas al lanzar una moneda trucada 10 veces.
- Observar el número de personas de pelo castaño en muestras de tamaño 20.
- Observar el número de turismos que entran en la autovía entre los próximos 15 vehículos.

■

Ejercicio 3.3 Si el 25 % de los automovilistas se detienen por completo en un determinado cruce. ¿Cuál es la probabilidad de que de 20 automovilistas seleccionados al azar, ...

- a. a lo sumo 6 se detengan por completo?
- b. exactamente 6 se detengan por completo?
- c. al menos 6 se detengan por completo?
- d. De los próximos 20 automovilistas, ¿Cuántos se puede esperar que se detengan por completo?

■

3.8 Modelo geométrico

Definición 3.8.1 Dada una sucesión de experimentos de Bernoulli con probabilidad de éxito p . La v.a. X definida como el número de realizaciones del experimento hasta obtener el primer éxito se distribuye según un **modelo geométrico** de parámetro p , y se denota por $X \sim Ge(p)$.

Proposición 3.8.1 Si $X \sim Ge(p)$, entonces

1. La fdp de X es:

$$P[X = k] = \begin{cases} q^{k-1}p & \text{si } k = 1, 2, \dots \\ 0 & \text{en otro caso} \end{cases}$$

2. $E[X] = \frac{1}{p}$ y $Var[X] = \frac{q}{p^2}$.

■ **Ejemplo 3.13** Experimentos con función de probabilidad de tipo Geométrica:

- Observar el número de lanzamientos realizados de una moneda hasta obtener la primera cara.
- Observar el número de donaciones necesarias hasta obtener un donante del tipo 0+.
- Observar el número de coches que entran en la autovía hasta que pase le primer coche verde.

■

Ejercicio 3.4 Dada una moneda trucada con $\frac{1}{4}$ de posibilidades de que salga cara. Calcular

- a. La probabilidad de obtener una cara antes del tercer intento.
- b. La probabilidad de obtener una cara en el tercer intento.
- c. Sean necesarios al menos 3 intentos para obtener una cara.

■

3.9 Modelo de Poisson

Definición 3.9.1 Una v.a. discreta X se distribuye según un **modelo de Poisson** de parámetro $\lambda > 0$, y se denota $X \sim \mathcal{P}(\lambda)$ si su fdp es:

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

Proposición 3.9.1 Sea $X \sim \mathcal{P}(\lambda)$ entonces

1. $E[X] = \lambda$ y $Var[X] = \lambda$.
2. **Reproductividad:** Dadas $X_1, X_2, \dots, X_n$ vv.aa. independientes, con $X_i \sim \mathcal{P}(\lambda_i)$, entonces

$$X = X_1 + X_2 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

Ejercicio 3.5 Suponga que la cantidad de conductores que realizan un determinado desplazamiento en el periodo vacacional sigue una distribución de Poisson con parámetro

- $\lambda = 10$. ¿Cuál es la probabilidad de que el número de conductores
- sea a lo sumo 10?
 - sea mayor que 20?
 - esté entre 10 y 20 (ambos inclusive)? Y ¿estrictamente entre 10 y 20?
 - esté dentro de dos desviaciones estandar respecto de la media?

Ejercicio 3.6 El número de errores de escritura en un determinado tipo de disco duro sigue una distribución de Poisson con parámetro $\lambda = 0,2$.

- ¿Cuál es la probabilidad de que en el disco haya exactamente un error?
- ¿Cuál es la probabilidad de que haya por lo menos dos errores?
- Si se eligen dos discos al azar. ¿Cuál es la probabilidad de que ninguno contenga errores?

3.10 Modelo Uniforme Continuo

Definición 3.10.1 Una v.a. continua se distribuye según un modelo **uniforme continuo** en el intervalo (a, b) , $X \sim U(a, b)$, si su *fdp* es constante dentro del intervalo, y nula fuera de él, esto es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Proposición 3.10.1 Si $X \sim U(a, b)$, entonces

$$E[X] = \frac{a+b}{2} \quad \text{y} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

3.11 Modelo exponencial

Definición 3.11.1 Una v.a. X se distribuye según un modelo exponencial de parámetro $\lambda > 0$, y se denota $X \sim \text{Exp}(\lambda)$, si su *fdp* es:

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \lambda e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

Proposición 3.11.1 Si $X \sim \text{Exp}(\lambda)$, entonces $E[X] = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$ y

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Ejercicio 3.7 Sea X la v.a. que representa el tiempo que tarda un cliente en repostar en una determinada gasolinera. Si X sigue una distribución exponencial con $\lambda = 1$. Calcule lo siguiente:

- El tiempo esperado de repostaje.
- La DE de repostaje.
- $P(X \leq 4)$.
- $P(2 \leq X \leq 5)$.

Proposición 3.11.2 Si la v.a. X que modela el número de eventos que ocurren por unidad de tiempo sigue una fdp de Poisson de parámetro α (i.e. $X \sim \mathcal{P}(\alpha)$), entonces la distribución de tiempos Y entre eventos sigue una fdp exponencial de parámetro α (i.e. $Y \sim \text{Exp}(\alpha)$).

Proposición 3.11.3 Si la v.a. X sigue una fdp exponencial de parámetro λ (i.e. $X \sim \text{Exp}(\lambda)$) entonces X tiene la propiedad de "amnesia o falta de memoria"; esto es

$$P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$$

Ejercicio 3.8 Sea X la distancia (m) que recorre un animal desde que nace hasta que encuentra un territorio desocupado. Suponga que X sigue una distribución exponencial con parámetro $\lambda = 0,01386$.

- ¿Cuál es la probabilidad de que la distancia sea a lo sumo 100m? ¿A lo sumo 200m? ¿Entre 100 y 200m?
- ¿Cuál es la probabilidad de que la distancia sea mayor que la distancia promedio más dos DE?
- ¿Cuál es el valor de la distancia mediana?

3.12 Modelo Normal

3.12.1 Modelo Normal reducida o tipificada

Definición 3.12.1 Una v.a. Z tiene una distribución **normal** de parámetros 0 y 1, **normal reducida** o **normal tipificada**, y la denotamos por $Z \sim N(0, 1)$, si su fdp es:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \forall x \in \mathbb{R}$$

Proposición 3.12.1 Sea $Z \sim N(0, 1)$, y $f(x)$ su fdp, entonces $f(x)$ cumple las siguientes propiedades:

- $f(-x) = f(x)$ (simétrica respecto del eje OY).
- Es creciente para $x < 0$ y decreciente para $x > 0$.
- Tiene un máximo absoluto en $x = 0$.
- Tiene asíntota horizontal en $y = 0$.
- Los puntos $x = -1$ y $x = 1$ son puntos de inflexión.

Proposición 3.12.2 Características asociadas a una v.a. $Z \sim N(0, 1)$:

- $E[Z] = 0$,
- $Var[Z] = 1$,
- $M_e(Z) = M_o(Z) = 0$,
- $F_Z(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx$
- Se denota por z_α al valor tal que

$$P[Z \leq z_\alpha] = \alpha.$$

- Por simetría de la fdp se tiene que:

$$P[Z \leq -z] = P[Z \geq z] = 1 - P[Z \leq z]$$

3.12.2 Normal Generalizada

Definición 3.12.2 Una v.a. X se dice que se distribuye según un **modelo Normal** de media $\mu \in \mathbb{R}$ y varianza $\sigma^2 \in \mathbb{R}^+$, y se denota por $X \sim N(\mu, \sigma^2)$ si su fdp es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in \mathbb{R}$$

Proposición 3.12.3

- f es simétrica respecto de $x = \mu$: $f(\mu - x) = f(\mu + x)$.
- Tiene un máximo en $x = \mu$,
- Presenta puntos de inflexión en $\mu - \sigma$ y $\mu + \sigma$.
- Presenta la asíntota horizontal $y = 0$.
- La fdp de probabilidad de X está concentrada en torno a μ y en el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$.

Proposición 3.12.4 Si $X \sim N(\mu, \sigma^2)$, entonces

1. $E[X] = \mu$ y $Var(X) = \sigma^2$.
2. $\frac{X-\mu}{\sigma} \sim N(0, 1)$ (**Tipificación**).
3. Dadas $X_1, X_2, \dots, X_n$ independientes con $X_i \sim N(\mu_i, \sigma_i^2)$, entonces

$$b + a_1 X_1 + \dots + a_n X_n \sim N\left(b + \sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

4. La mediana y la moda valen μ .

Ejercicio 3.9 La fuerza que actúa sobre una columna tiene una distribución normal con media 15.0 kips y desviación estandar de 1.25 kips. ¿Cuál es la probabilidad de que la fuerza

- a. sea a lo sumo 18 kips?
- b. se encuentre entre 10 y 12 kips?
- c. difiera de 15.0 kips en a lo sumo 2 DE?

Ejercicio 3.10 La dureza de cierta aleación sigue una distribución normal con $\mu = 70$ y $\sigma = 3$.

- a. Si es de dureza aceptable sólo si su dureza está entre 67 y 75, ¿cuál es la probabilidad de que un espécimen seleccionado al azar sea aceptable?
- b. ¿Cuál es el número esperado de especímenes aceptables de entre 10 elegidos al azar?
- c. ¿Cuál es la probabilidad de que a lo sumo 8 de 10 especímenes elegidos al azar tengan una dureza menor que 73.84?
- d. Si el intervalo de dureza aceptable fuera $(70-c, 70+c)$, ¿para qué valor de c , el 95 % de los especímenes tendrán dureza aceptable?

3.13 Aproximaciones de modelos de variables

3.13.1 Aproximación de una v.a. binomial por una v.a. Poisson

Proposición 3.13.1 Sea $X \sim B(n, p)$ si hacemos crecer $n \rightarrow \infty$ y hacemos $p \rightarrow 0$ de modo que $p = \lambda/n$. Entonces $X \rightsquigarrow P(\lambda)$.

La aproximación anterior es aplicable si $n > 30$, $p < 0,05$.

■ **Ejemplo 3.14** En un banco, la probabilidad de recibir un cheque sin fondos es de 0.4. Si en un día se reciben 100 cheques, ¿cuál es la probabilidad de encontrar más de cinco cheques sin fondo?

Solución

$$X \sim \text{Bi}(100, 0,04)$$

Como $n > 30$ y $p < 0,05$ podemos aproximar por una variable Poisson de parámetro $\lambda = n \cdot p = 4$, es decir, $X \approx P(4)$

$$P[X < 5] = 1 - P[X \leq 5] = 1 - 0,785 = 0,215. \quad \blacksquare$$

3.13.2 Aproximación de una v.a. binomial por una v.a. normal

El siguiente teorema, que se conoce como teorema del límite de De Moivre–Laplace, asegura una proximación adecuada mediante la distribución normal de las probabilidades binomiales si n es suficientemente grande.

Teorema 3.13.2 — de De Moivre-Laplace. Sea $X \sim B(n, p)$, entonces si se verifica que $np > 5$ y $nq > 5$ se cumple que $F_X(x) = P[X \leq x]$ se puede aproximar por $F_Y(x) = P[Y \leq x]$, donde $Y \sim N(np, npq)$ (distribución normal de media np y varianza npq).

Observación 3.1 La aproximación será mejor cuanto mayor sea n y cuanto más se aproxime p a 0,5.

3.13.3 Aproximación de una v.a. Poisson por una v.a. normal

Proposición 3.13.3 Sea $X \sim \mathcal{P}(\lambda)$ entonces si λ es grande (> 20) se cumple que $F_X(x) = P[X \leq x]$ se puede aproximar por $F_Y(x) = P[Y \leq x]$, donde $Y \sim N(\lambda, \lambda)$ (distribución normal de media λ y varianza λ).

Observación 3.2 La aproximación será mejor cuanto mayor sea el valor de λ .

Nota 3.13.4 Al ser una aproximación de una variable aleatoria discreta o una continua, hemos de aplicar también, al igual que en la Binomial, la **corrección de continuidad**.

3.13.4 Relación entre una v.a. Poisson y una v.a. exponencial negativa

Proposición 3.13.5 Si la v.a. X que modela el número de eventos que ocurren por unidad de tiempo sigue una fdp de Poisson de parámetro λ (i.e. $X \sim \mathcal{P}(\lambda)$), entonces la distribución de tiempos Y entre eventos sigue una fdp exponencial de parámetro λ (i.e. $Y \sim \text{Exp}(\lambda)$).

Distribución binomial	Distribución normal
$P[X_{Bi} = a]$	$P[a - 0,5 < X_N < a + 0,5]$
$P[X_{Bi} < a]$	$P[X_N < a - 0,5]$
$P[X_{Bi} \leq a]$	$P[X_N < a + 0,5]$
$P[X_{Bi} > a]$	$P[X_N > a + 0,5]$
$P[X_{Bi} \geq a]$	$P[X_N \geq a - 0,5]$
$P[a < X_{Bi} < b]$	$P[a + 0,5 < X_N < b - 0,5]$
$P[a < X_{Bi} \leq b]$	$P[a + 0,5 < X_N < b + 0,5]$
$P[a \leq X_{Bi} < b]$	$P[a - 0,5 < X_N < b - 0,5]$
$P[a \leq X_{Bi} \leq b]$	$P[a - 0,5 < X_N < b + 0,5]$

Cuadro 3.2: Situaciones posibles en la aproximación de una variable binomial por una variable normal (a, b son números enteros).

3.13.5 Corrección por continuidad

Cuando aproximamos una distribución discreta mediante una distribución continua en la variable aleatoria continua la probabilidad de que la variable tome un valor concreto cualquiera es cero, al contrario que en los valores de la discreta. Para evitar este problema en la aproximación de los valores, éstos se corrigen (**corrección de continuidad**) sustituyendo el valor del punto en la discreta por el valor de la continua sobre todo un intervalo centrado en el punto.

En la Tabla 3.2 se muestran todas las situaciones posibles para el caso de aproximar una binomial X_{Bi} por una normal X_N .

3.14 Ejercicios

Ejercicio 3.11 Un distribuidor de electrodomésticos vende tres modelos diferentes de congeladores con capacidad de 13,5, 15,9 y 19,1 unidades cúbicas. Sea X la capacidad que compra el siguiente cliente y suponga que X tiene la siguiente *fdp*

x	13,5	15,9	19,1
$p(x)$	0,2	0,5	0,3

- Calcule $E(X)$, $E(X^2)$ y $V(X)$.
- Si el precio de un congelador con capacidad de X unidades cúbicas es $25X - 8,5$, ¿cuál es el precio esperado que paga un cliente por un congelador?
- ¿Cuál es la varianza del precio $25X - 8,5$ que paga el siguiente cliente?
- Suponga que si bien la capacidad nominal de un congelador es X , la capacidad real es $h(X) = X - 0,01X^2$. ¿Cuál es la capacidad real esperada del congelador que compra el siguiente cliente?

Ejercicio 3.12 Cierta aparato registra el nivel de saturación de la red eléctrica en una comarca. El error relativo porcentual de la medida dada por el aparato es una variable aleatoria continua X con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - (1 - x)^3 & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1. \end{cases}$$

- (a) Calcular la probabilidad de que una medida registrada por el aparato tenga un error entre el 0.1 % y el 0.2 %
- (b) Calcular el error relativo medio.

■

Ejercicio 3.13 La *fda* de la duración X de tiempo de préstamo de un libro es

$$F(x) = \begin{cases} 0 & x < 0, \\ x^2/4 & 0 \leq x < 2, \\ 1 & 2 \leq x \end{cases}$$

Calcule:

- a. $P(X \leq 1)$, $P(0,5 \leq X \leq 1)$ y $P(X > 1)$.
- b. La mediana $\tilde{\mu}$ (resuelva $0,5 = F(x)$).
- c. La función de densidad $f(x) = F'(x)$.
- d. $E(X)$, $V(X)$ y $DE(X)$.
- e. Si se carga una cantidad $h(X) = X^2$ cuando la duración del préstamo es X , calcule el cargo esperado $[E(h(X))]$.

■

Ejercicio 3.14 Dada una variable aleatoria X con *fda*

$$f(x) = \begin{cases} 2 \cos(2x) & \text{si } 0 < x < \frac{\pi}{4}, \\ 0 & \text{caso contrario} \end{cases}$$

Calcule:

- a. La media de la distribución.
- b. La varianza de la distribución.
- c. La mediana de la distribución.
- d. La moda de la distribución.

■

Ejercicio 3.15 El tiempo de avance de dos automóviles elegidos al azar en una determinada carretera sigue una distribución X . Suponga que la distribución de tiempo de

avance es:

$$f(x) = \begin{cases} 0 & x \leq 1, \\ k/x^4 & x > 1, \end{cases}$$

- Determine k para que $f(x)$ sea una *fdp* y obtenga $F(x)$.
- Determine $P[X > 2]$ y $P[2 < X < 3]$.
- Obtenga la media y la desviación estandar de X .
- ¿Cuál es la probabilidad de que el avance esté dentro de una desviación estandar del promedio?

Ejercicio 3.16 La demanda semanal de Gas Propano (en miles de galones) es una va X con *fdp*

$$f(x) = \begin{cases} 2 \cdot \left(1 - \frac{1}{x^2}\right) & \text{si } 1 \leq x \leq 2, \\ 0 & \text{en caso contrario,} \end{cases}$$

- Calcule la *fda* de X .
- Obtenga una expresión para el percentil (100p). ¿Cuál es el valor de $\tilde{\mu}$?
- Calcule $E(X)$ y $V(X)$.
- Si hay en existencia 1500 galones al comienzo de la semana y durante la semana no es conveniente ningún nuevo suministro, ¿cuánto de los 1500 galones se espera que quede al final de la semana?

Ejercicio 3.17 Sea X una va que tiene *fdp* la función de Pareto

$$f(x; k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta, \\ 0 & x < \theta, \end{cases}$$

- Si $k > 1$, calcule $E(X)$. ¿Qué se puede decir de $E(X)$ si $k = 1$?
- Si $k > 2$, demuestre que $V(X) = \frac{k\theta^2}{(k-1)^2(k-2)}$
- Si $k = 2$, ¿qué se puede decir sobre $V(X)$.
- ¿Qué condiciones en k son necesarias para asegurar que $E(X^n)$ es finita?

Ejercicio 3.18 Un tipo de raqueta de tenis se fabrica en dos tamaños: mediano (M) y grande (G). El 60% de los clientes prefiere la versión (G).

- Entre 10 clientes que compran este tipo de raqueta, ¿cuál es la probabilidad de que al menos 6 prefieran la versión (G)?
- De 10 clientes seleccionados al azar, ¿cuál es la probabilidad de que el número que prefiere (G) esté dentro de una desviación estandar del valor medio?
- La tienda tiene en existencias 7 raquetas de cada tipo. ¿cuál es la probabilidad de

que los siguientes 10 clientes obtengan, con las existencias actuales, la versión de raqueta que buscan?

Ejercicio 3.19 El 20 % de los teléfonos de un determinado tipo son llevados a reparar cuando aún está vigente la garantía. De éstos, el 60 % se reparan y el 40 % debe ser sustituido por uno nuevo.

Si una empresa compra 10 teléfonos, ¿cuál es la probabilidad de que exactamente 2 terminen siendo reemplazados dentro del periodo de garantía?

Ejercicio 3.20 Suponga que sólo 0.10 % de los ordenadores de un cierto tipo experimentan un fallo durante la garantía. Consideremos una muestra de 10000 ordenadores.

- ¿Cuál es el valor esperado y la desviación estandar del número de ordenadores de la muestra que experimentarán un fallo en el periodo de garantía?
- ¿Cuál es la probabilidad aproximada de que más de 10 ordenadores muestren un fallo?
- ¿Cuál es la probabilidad aproximada de que ninguno de los ordenadores muestreados tengan defecto?

Ejercicio 3.21 La concentración de sustrato (mg/cm^3) de flujo entrante a un reactor tiene una distribución normal con $\mu = 0,30$ y $\sigma = 0,06$

- ¿Cuál es la probabilidad de que la concentración sea mayor que 0.25?
- ¿Cuál es la probabilidad de que la concentración sea a lo sumo 0.10?
- ¿Cómo caracterizaría el 5 % más grande de los valores de la concentración?

Ejercicio 3.22 El diámetro de los árboles (pulg.) de un bosque sigue una distribución normal con $\mu = 8,8$ y $\sigma = 2,8$.

- ¿Cuál es la probabilidad de que el diámetro de un árbol sea por lo menos 10 pulg?
- ¿Cuál es la probabilidad de que el diámetro de un árbol sea mayor que 20 pulg?
- ¿Cuál es la probabilidad de que el diámetro de un árbol esté entre 5 y 10 pulg?
- ¿Qué valor de c es tal que el intervalo $(8.8-c, 8.8+c)$ incluya el 98 % de los valores de diámetro?
- Si se elige a cuatro árboles de forma independiente, ¿cuál es la probabilidad de que por lo menos uno tenga un diámetro mayor que 10 pulg.?

Ejercicio 3.23 El tiempo hasta que un tipo de ventiladores de un fallo sigue una distribución exponencial. Suponga que el tiempo medio hasta presentar el fallo es de 25000h. ¿Cuál es la probabilidad de que

- un ventilador seleccionado al azar dure al menos 20000h.? ¿A lo sumo 30000? ¿Entre 20000 y 30000h?
- la vida útil de un ventilador exceda el valor medio en más de 2 DE? ¿En más de 3

DE?

Ejercicio 3.24 Si X sigue una distribución exponencial con parámetro λ , deduzca una expresión general para el percentil (100p) de la distribución. Después especifique cómo obtener la mediana.

Ejercicio 3.25 Las capacidades de los condensadores de una serie en una cadena de fabricación siguen una distribución normal, con una media de 1.3 nF (nano-Faradios), y una desviación típica de 0.04 nF .

- Escogido un condensador al azar, calcular la probabilidad de que su capacidad está comprendida entre 1.28 nF y 1.30 nF .
- Calcular entre qué dos valores (simétricamente distribuidos en torno a la media) se encuentran las capacidades del 80 % de los condensadores de esta serie.
- Se selecciona al azar una muestra de 16 condensadores. Calcular la probabilidad de que su capacidad media está comprendida entre 1.28 nF y 1.30 nF .

Ejercicio 3.26 La probabilidad de fabricar una pieza defectuosa y rechazable por el cliente es 0.005. Tenemos un lote de 500 piezas. ¿Cuál es la probabilidad de que al menos el 2 % de las piezas sean rechazadas por el cliente?

Ejercicio 3.27 En un proceso de fabricación de componentes electrónicos, a veces se producen defectos que los hacen inservibles. Supongamos que 2 de cada 1000 piezas salen defectuosas en promedio, ¿cuál es la probabilidad de que en una muestra aleatoria de 5000 piezas menos de 9 sean inservibles?

- Expresar el resultado en términos de la variable correspondiente.
- Expresar el resultado aproximando por una variable aleatoria Poisson.
- Expresar el resultado aproximando por una variable aleatoria normal.
 - Sin aplicar la corrección de continuidad.
 - Aplicando la corrección de continuidad.

Ejercicio 3.28 El gerente de una fábrica sabe que el 2 % de los artículos que fabrica son defectuosos. Para hacer una prueba de control de calidad se seleccionan 1.000 artículos aleatoriamente, ¿Cuál es la probabilidad de que el número de artículos defectuosos,

- sea mayor o igual a 14?
- sea menor que 10?

Aplicar en ambos casos la corrección de continuidad.

Ejercicio 3.29 En una fábrica la probabilidad de que se produzcan n piezas defectuosas sigue una distribución de Poisson de media 3 diarias. Determinar la probabilidad de que en 200 días el número de defectuosas esté comprendida entre 600 y 690 (sin incluir los extremos).

Ejercicio 3.30 Tras un estudio de mercado, la empresa “El burrito veloz”, ha determinado que el número medio de llamadas en la sucursal de Huelva sigue una distribución de Poisson de media 5 llamadas a la hora mientras que la sucursal de Sevilla sigue una distribución de Poisson de media 7 llamadas a la hora.

- Determinar la probabilidad de que entre ambas sucursales se reciban más de 15 llamadas en la siguiente hora.
- Determinar la probabilidad de que se reciba exactamente una llamada, en la sucursal de Huelva, en el siguiente cuarto de hora.
- Si, en la sucursal de Sevilla, hace media hora que se recibió la última llamada, ¿qué probabilidad hay de que la próxima llamada se reciba entre los 20 y los 40 minutos siguientes.

Ejercicio 3.31 Se sabe que el número de veces que una persona lleva su coche a lavar sigue una distribución de Poisson con media de 2.1 veces por semana.

- Hallar la probabilidad de que en un día determinado la persona lleve su coche a lavar más de una vez.
- Hallar la probabilidad de que transcurra más de 5 días entre dos lavados consecutivos del coche.

3.14.1 Ejercicios de exámenes de años anteriores

Ejercicio 3.32 El tiempo de reparación (en horas) de un tipo de máquina es una variable aleatoria X con función de distribución:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0,5x & 0 \leq x < 1 \\ 0,5 & 1 \leq x < 2 \\ 0,25x & 2 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

- Determinar la función de densidad de X .
- Calcular el tiempo medio de reparación de este tipo de máquina.
- Si se realizan 10 reparaciones, ¿cuál es la probabilidad de que al menos 3 de ellas requieran más de 3 horas?

Ejercicio 3.33 Las calificaciones que obtienen los alumnos en una asignatura siguen una distribución normal de media 4.7 puntos y desviación típica 1.5 puntos.

- Calcular la probabilidad de que un alumno apruebe la asignatura.
- Generalmente acuden a la revisión el 5% de los alumnos que obtienen una calificación menor o igual que 4, el 80% de los que tienen entre 4 y 5 puntos y el 2% de los que aprueban. Si un alumno acude a revisión, ¿cuál es la probabilidad

de que haya aprobado?

Ejercicio 3.34 En un bosque que conviven dos tipos de escarabajos, A y B , el 60 % de la población es de tipo A y el resto de tipo B . La longitud, en cm , de la población de escarabajos sigue una distribución Normal de media 1.5 cm y desviación típica 0.25 cm para el tipo A y media 1.6 cm y desviación típica 0.2 cm para el tipo B . Si tomamos un escarabajo al azar,

- (a) ¿Cuál es la probabilidad de que mida menos de 1.8 cm ?
- (b) Si mide menos de 1.8 cm , ¿cual es la probabilidad de que pertenezca al tipo B ?
- (c) Si se ha tomado una muestra de 20 individuos del tipo B , ¿cuál será el número esperado de escarabajos cuya longitud está entre 1.5 cm y 1.8 cm ?

Ejercicio 3.35 La dimensión de ciertas piezas de hormigón siguen una distribución normal de media 150 y desviación típica 0.4 . Se consideran aceptables todas aquellas piezas cuya longitud se encuentre dentro del intervalo $(149,2, 150,4)$. Se pide:

- (a) El porcentaje de piezas defectuosas.
- (b) Proporcionar un intervalo centrado en 150 que contenga el 95 % de los valores de la dimensión.
- (c) Estamos interesados en reducir la variabilidad de los valores de la dimensión de modo que el 99 % de estos valores se sitúen estén en el intervalo $(149,485, 150,515)$. >¿Qué valor de σ debemos conseguir?

Nota: si el valor de la probabilidad no aparece en la tabla, obtener el valor de z como la media aritmética entre el valor por defecto y el valor por exceso.

Ejercicio 3.36 En la elaboración de un determinado medicamento en forma de comprimido interviene un producto químico cuya cantidad sigue aproximadamente una distribución Normal de media 3 grs. y desviación típica 0.05 grs.

- (a) Calcular la probabilidad de que un comprimido pese más de 3.025 grs.
- (b) Un comprimido se considera defectuoso cuando su peso difiere de la media en más de una desviación típica. Calcular la proporción de comprimidos defectuosos que se fabrican.
- (c) Estos comprimidos se envasan en cajas de 10 unidades. Si un envase contiene 2 o más comprimidos defectuosos se elimina del mercado. Determinar el porcentaje de cajas que se retiran del mercado. (Para mayor comodidad, redondear la probabilidad obtenida y trabajar con un decimal).

Ejercicio 3.37 Una fábrica produce un tipo de componente electrónico en dos calidades diferentes. El 60 % de la producción es de calidad A y el resto de calidad B . La duración X de un componente de calidad A sigue una distribución normal de media 4 años y desviación típica 2 , mientras que los componentes de calidad B tienen una duración Y que sigue una distribución exponencial de media 4 años .

- (a) ¿Cuál es la probabilidad de que un componente de calidad A dure más de 2 años? ¿Y si es de calidad B?
- (b) Si tomamos un componente de calidad A y otro de calidad B, ¿cuál es la probabilidad de que sólo el de calidad A dure más de 2 años?
- (c) Se toma un componente al azar de toda la producción y se observa que dura más de dos años, ¿cuál es la probabilidad de que fuera de calidad A?

Ejercicio 3.38 La cantidad demandada diariamente de un determinado producto, generado en cierto proceso, sigue una distribución normal de media 1000 Kg y desviación típica 100 kg.

- (a) Determinar la probabilidad de que un día cualquiera la demanda sea inferior a 990 Kg.
- (b) Calcular la cantidad que se debe producir diariamente para satisfacer plenamente la demanda con probabilidad 0.9.
- (c) Calcular un intervalo centrado en 1000, que contenga el 95 % de las cantidades demandadas diariamente.

Ejercicio 3.39 Un estudio de la DGT estima que el número de horas prácticas necesarias para la obtención del permiso de conducir sigue una distribución normal $N(24, 3^2)$.

- (a) ¿Qué probabilidad hay de obtener el permiso de conducir con 20 horas de prácticas o menos?
- (b) ¿Cuántas horas de prácticas ha necesitado un conductor para obtener el permiso si el 68 % de los conductores ha necesitado más horas prácticas que él.

La autoescuela D'ESPACIO ingresa por alumno una parte fija de 250 euros, más 23 euros por hora práctica.

- (c) Calcular el ingreso esperado por alumno.
- (d) Calcular la desviación típica del ingreso por alumno.

Ejercicio 3.40 La producción diaria de un producto químico, en miles de kg, es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} \frac{2}{9} x (3 - x) & \text{si } x \in (0, 3) \\ 0 & \text{si } x \notin (0, 3) \end{cases}$$

- (a) Calcular la producción media diaria.
- (b) Un día se considera “malo” si se producen menos de 800 kg, “normal” si se producen entre 800 y 2600 kg y “excelente” si se superan los 2600 kg. ¿Cuál es la probabilidad de que un día resulte “malo”? ¿y “excelente”?
- (c) La probabilidad de satisfacer la demanda un día “excelente” es 1, si el día es “normal” dicha probabilidad es de 0,9 y si es “malo” es 0,5. Si un día elegido al azar no se ha cubierto la demanda, ¿cuál es la probabilidad de que fuera un día

“normal”?

Ejercicio 3.41 En un proceso químico se pueden utilizar dos tipos de catalizadores. El catalizador *A* es más caro y sólo se utiliza en el 30 % de los procesos, y en el resto se utiliza el catalizador *B*. El rendimiento de un proceso cuando se utiliza el catalizador *A* sigue una distribución $N(92,7,8,9)$, mientras que con el catalizador *B* sigue una distribución $N(92,3,5,7)$.

- (a) Calcular la probabilidad de que el rendimiento de un proceso sea superior a 92.5 cuando se utiliza el catalizador *A*.
- (b) Si en un proceso el rendimiento es inferior a 92.5, ¿cuál es la probabilidad de que se haya empleado el catalizador *B*?

Ejercicio 3.42 Un fabricante de hormigón proporciona dos calidades diferentes del mismo: *A* y *B*, cuyas características de resistencia se distribuyen normalmente con medias 200 kg/cm^2 y 210 Kg/cm^2 y desviación típicas 5 Kg/cm^2 y 10 Kg/cm^2 respectivamente. Determinar:

- (a) La resistencia que para cada calidad puede garantizarse con una probabilidad de 0'95.
- (b) La probabilidad de que una muestra de 36 unidades de hormigón de tipo *A* tenga una resistencia media superior a la correspondiente de una muestra equivalente de 36 unidades de hormigón de tipo *B*.
- (c) Tomando el mismo número de unidades de hormigón de ambos tipos, ¿cuántos elementos se deben muestrear a fin de tener asegurado con probabilidad 0'001 que la resistencia media de la muestra del tipo *A* es mayor que la media del tipo *B*?

Ejercicio 3.43 En una fábrica el número de accidentes por semana sigue una distribución de Poisson de media 2.

- (a) Hallar la probabilidad de que en una semana hay algún accidente.
- (b) Hallar la probabilidad de que haya 4 accidentes en el transcurso de dos semanas.
- (c) Si nos dicen que ha habido algún accidente, hallar la probabilidad de que en esa semana no haya habido más de tres accidentes.

Ejercicio 3.44 El coeficiente intelectual (CI) de una población de individuos se distribuye de manera aproximadamente normal con media 100 y desviación estándar 16. ¿Cuál es la probabilidad de que un individuo seleccionado aleatoriamente tuviera un CI

- (a) de más de 140?
- (b) entre 90 y 110?
- (c) ¿Qué intervalo de valores, centrado en 100, incluiría al 70 % de los individuos?

Ejercicio 3.45 Varios test de inteligencia aplicados a los individuos de cierta población, dieron una puntuación que sigue una ley normal con media 100 y desviación típica 15.

- (a) ¿Qué intervalo centrado en 100 contiene al 90 % de la población?
- (b) Se escogen 10 individuos al azar, ¿qué probabilidad hay de que al menos 4 de ellos tengan coeficiente superior a 120? (Redondear el valor de p a la primera cifra decimal).

■

Ejercicio 3.46 La memoria RAM para un ordenador se puede recibir de dos fabricantes A y B con igual probabilidad. Si la memoria proviene del fabricante A, la probabilidad de que falle antes del tiempo especificado por la garantía es $P[X \leq 1]$ donde la variable X sigue una ley exponencial de parámetro $\lambda = 0,2$; si la memoria proviene del fabricante B, la probabilidad de que falle antes del tiempo especificado por la garantía es $P[|Y| < 2]$ donde Y tiene una distribución normal de media $\mu = 4$ y varianza $\sigma^2 = 4$.

- (a) ¿Cuál es la probabilidad de que una memoria RAM falle antes del tiempo especificado por la garantía?
- (b) Si se ha observado que la memoria RAM ha fallado, ¿cuál es la probabilidad de que proceda del fabricante A?
- (c) Si se tienen 100 memorias RAM, ¿cuál es la probabilidad de que al menos el 90 % de ellas duren más que el tiempo especificado por la garantía?

■

Ejercicio 3.47 Un atleta realiza saltos cuya longitud sigue una distribución normal de media 7.5 metros y desviación estándar 0.5 metros. Por cada salto menor de 6.86 metros que realiza en una competición, su entrenador le multa con 1 euro. Se pide:

- (a) La probabilidad de que el salto sea merecedor de multa.
- (b) Sabiendo que ha sido multado en el primer salto, ¿cuál es la probabilidad de que el salto haya sido mayor de 6.5 metros?
- (c) ¿Cuál es el valor esperado de dinero que tendrá que pagar en multas en 100 saltos de competición?

Nota: Redondear todas las probabilidades en este problema al segundo decimal.

■

Ejercicio 3.48 El porcentaje de contaminante presente en una muestra de aire es una variable aleatoria con función de densidad dada por

$$f(x) = \begin{cases} a + bx^2 & 0 < x < 1 \\ 0 & \text{caso contrario} \end{cases}$$

- (a) Si $E[X] = \frac{3}{5}$, calcular a y b para que f sea función de densidad.
- (b) Calcular la varianza.
- (c) Calcular la probabilidad de que el porcentaje de contaminante en una muestra de aire sea superior a 0.6.

■

Ejercicio 3.49 Sea X una variable aleatoria que sigue una distribución normal de media 3. Sabiendo que $P[X \leq 1] = 0,0228$ y que $P[X \geq 4] = 0,1587$, calcular $P[2 \leq X \leq 5]$. ■

Ejercicio 3.50 Sea X una variable aleatoria que mide la longitud de cierta pieza producida por un determinado aparato y que tiene por función de densidad

$$f(x) = k(x-1)(3-x) \quad 1 \leq x \leq 3$$

- (a) Calcular, razonadamente, k para que $f(x)$ sea, efectivamente una función de densidad.
- (b) Una pieza sólo es válida si su longitud varía entre 1.7 y 2.4. ¿Cuál es la probabilidad de que esto ocurra?
- (c) Las piezas se empaquetan en lotes de 5 unidades, y se acepta el lote si contiene menos de 2 piezas defectuosas. ¿Cuál es la probabilidad de que un determinado lote sea rechazado?

Ejercicio 3.51 La concentración del soluto en % (m/m) de una disolución salina a una cierta temperatura, se modela a través de una variable aleatoria X cuya función de densidad viene dada por

$$f(x) = \begin{cases} 3x^2 & \text{si } x \in (0, 1) \\ 0 & \text{si } x \notin (0, 1) \end{cases}$$

Calcular la función de distribución, la esperanza y la varianza de la concentración del soluto. ■

Ejercicio 3.52 Una fábrica produce un tipo de componente electrónico en dos calidades diferentes. El 60% de la producción es de calidad A y el resto de calidad B. La duración X de un componente de calidad A sigue una distribución normal de media 5 años y desviación típica 2, mientras que la duración de los componentes de calidad B es una variable aleatoria Y que sigue una distribución normal de media 2 años y desviación típica 0.5.

- (a) ¿Cuál es la probabilidad de que un componente dure más de 3 años si es de calidad A? ¿Y si es de calidad B?
- (b) Si tomamos un componente de calidad A y otro de calidad B, ¿cuál es la probabilidad de que sólo el de calidad A dure más de 3 años?
- (c) Se toma un componente al azar de toda la producción y se observa que dura más de 3 años. ¿Cuál es la probabilidad de que sea de calidad A?

Ejercicio 3.53 Sea X una v.a. continua que mide la longitud, en centímetros, de una

determinada pieza, cuya función de densidad es:

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \leq x \leq b \\ 0 & \text{caso contrario} \end{cases}$$

- (a) Calcular b y la función de distribución $F_X(x)$ de X .
- (b) Se dice que la pieza es larga si su longitud es superior a 1.6 cm . y corta si no supera a $\sqrt{0,2} \text{ cm}$. Calcular la probabilidad de elegida una pieza al azar sea larga o corta.
- (c) Se cogen 20 piezas al azar de esta distribución, ¿cuál es la probabilidad de que haya menos de 5 piezas cortas? ¿Cuál es el número esperado de piezas cortas?

Ejercicio 3.54 La capacidad de unos determinados envases sigue una distribución Normal de media 100 cl y desviación típica 0.61 cl . Según una norma de calidad, se consideran aceptables todas aquellos envases cuya capacidad esté comprendida dentro del intervalo $(99, 101)$.

- (a) Determinar el porcentaje de envases que cumplen la norma.
- (b) Determinar un intervalo centrado en la media donde se encuentren el 95 % de las capacidades de los envases.
- (c) Supongamos que los envases se empaquetan en lotes de 12 unidades, y un lote se rechaza si contiene más de 2 envases defectuosos. Determinar la proporción de lotes que se rechazarán.

Nota: Por comodidad, redondear la probabilidad intermedia a la primera cifra decimal.



4. Muestreo y Estimación

4.1 Muestreo

Cuando realiza un estudio estadístico sobre una población, lo habitual es que no se pueda acceder a todos los individuos que la componen ya sea porque no son accesibles o por simple falta de recursos materiales o técnicos. Cuando esto ocurre se hace necesario elegir una muestra, y realizar el estudio sobre ella con la esperanza de que los datos obtenidos sobre la muestra sean extrapolables a toda la población. Este proceso permite ahorrar recursos obteniendo resultados parecidos a los que se alcanzarían si se realizase un estudio de toda la población.

En general para que el muestreo sea válido solo se pueden hacer estimaciones sobre la población sino también se pueden estimar los márgenes de error asociados a dichas estimaciones. Es importante observar que nunca podremos estar enteramente seguros de que el resultado sea correcto y que se haya obtenido de una muestra representativa, pero sí podemos estar seguros de que los resultados obtenidos son correctos con una alta probabilidad”. Para ello la muestra ha de elegirse de manera que sea lo más representativa posible de la población bajo estudio. El proceso de extracción de la muestra se denomina muestreo.

4.1.1 Estadísticos y sus Distribuciones

De una misma población, cada vez que se muestrea, se extrae una muestra diferente. Esto es, dadas dos muestras de una misma población, los valores de ambas muestras $x_1^1, x_2^1, \dots, x_n^1$ y $x_1^2, x_2^2, \dots, x_n^2$ son por lo general (y con bastante probabilidad) distintos. Antes de obtener los datos, hay incertidumbre acerca del valor de cada x_i ya que lo único que se conoce es que son datos observaciones de una determinada población. Debido a esta incertidumbre, antes que estén disponibles los datos se ve cada observación como una variable aleatoria y la muestra completa se representa como $X_1, X_2, \dots, X_n$.

Se cumple, por tanto, que teniendo en cuenta que los valores de cada una de las

observaciones muestrales varían de una muestra a otra, el valor de cualquier cantidad calculada a partir de la muestra difícilmente coincidirá de forma exacta con el valor que está siendo estimado.

Definición 4.1.1 Un **estadístico** es cualquier cantidad cuyo valor se calcula a partir de los datos de una muestra. Se denota siempre por una letra mayúscula, denotando por la correspondiente letra minúscula el valor calculado del estadístico en función de la muestra que se utiliza para calcular su valor.

■ **Ejemplo 4.1** La media muestral, considerada como estadístico, se denota por $\bar{X}$, es decir,

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

La varianza muestral

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Para una muestra concreta de valores: $n = 4$, $x_1 = 2$, $x_2 = 5$, $x_3 = 1$, $x_4 = 2$ se tiene

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{2 + 5 + 1 + 2}{4} = 2,5$$

$$s^2 = \frac{(2 - 2,5)^2 + (5 - 2,5)^2 + (1 - 2,5)^2 + (2 - 2,5)^2}{3} = 3$$

■

Observación 4.1 Como antes de obtener una muestra existe incertidumbre de cual va a ser el valor que tome, **un estadístico es una v.a.**

La *fdp* de un estadístico depende no sólo de la *fdp* de la v.a. original y del tamaño de la muestra, sino que depende también del método de muestreo. Por ello, la distribución de probabilidad de un estadístico también puede denominarse **distribución de muestreo**.

Definición 4.1.2 La secuencia de v.a. $X_1, \dots, X_n$ forman una **muestra aleatoria (simple)** (m.a.s.) de tamaño n si:

1. Las X_i son v.a. independientes.
2. Todas las X_i tienen la misma *fdp*.

Las condiciones 1 y 2 pueden resumirse diciendo que las X_i son independientes e idénticamente distribuidas (*iid*).

Nota 4.1.1 Las condiciones anteriores se satisfacen (aprox.) si $n \ll N$. En la práctica si a lo sumo se muestrea el 5 % de la población se puede suponer que las X_i forman una m.a.s.

4.1.2 Distribución de la media muestral

La importancia de la media muestral $\bar{X}$ surge de su uso para sacar conclusiones de la media poblacional μ .

Proposición 4.1.2 Sea $X_1, \dots, X_n$ una m.a.s. de una *fdp* cualquiera con media μ y desviación estandar σ . Entonces

1. $E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \mu_{\bar{X}} = \mu$.
2. $Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \sigma_{\bar{X}}^2 = \sigma^2/n$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

Además, si $T_0 = X_1 + \dots + X_n$ (el total de la muestra), $E(T_0) = n\mu$ y $Var(T_0) = n\sigma^2$.

Como vemos en la proposición anterior, promediar concentra la probabilidad en la media en tanto que totalizar la dispersa.

■ **Ejemplo 4.2** En una prueba de fatiga a la tensión en un espécimen de titanio, el número esperado de ciclos hasta la primera emisión acústica X es $\mu_X = 28000$ y la desviación estándar es $\sigma_X = 5000$. Sea $X_1, X_2, \dots, X_{25}$ una muestra de tamaño 25. Calcular $E[\bar{X}]$ y $V[\bar{X}]$.

Solución

$$E[\bar{X}] = \mu_{\bar{X}} = \mu_X = 28000 \quad \text{y} \quad V[\bar{X}] = \sigma_{\bar{X}}^2 = \sigma_X^2/n = 5000^2/25 = 1000^2. \quad \blacksquare$$

4.1.3 Teorema Central del Límite

Teorema 4.1.3 — Teorema central del Límite (TCL). Sea $X_1, \dots, X_n$ una m.a.s. de una v.a. X con media μ y varianza σ^2 . Entonces si n es suficientemente grande, $\bar{X} \simeq N(\mu, \sigma^2/n)$ y $T_0 \simeq N(n\mu, n\sigma^2)$. Mientras mayor sea el valor de n , mejor es la aproximación.

Nota 4.1.4 A efectos de aplicación práctica en esta asignatura, si $n > 30$ se puede usar el Teorema Central del Límite

4.1.4 Distribución de una comb. lineal

Definición 4.1.3 Dada una colección de n v.a. $X_1, \dots, X_n$ y n constantes numéricas $a_1, \dots, a_n$, a la v.a.

$$Y = a_1X_1 + \dots + a_nX_n = \sum_{i=1}^n a_iX_i$$

se denomina **combinación lineal** de las X_i .

Proposición 4.1.5 Sean $X_1, \dots, X_n$ con medias $\mu_1, \dots, \mu_n$ y varianzas $\sigma_1^2, \dots, \sigma_n^2$ respectivamente y sea Y una combinación lineal de las X_i en las condiciones de la definición anterior. Entonces

1. Para cualquier n -upla (independientes o no) se cumple que

$$\begin{aligned} E(Y) &= E(a_1X_1 + \dots + a_nX_n) &= \\ &= a_1E(X_1) + \dots + a_nE(X_n) &= \\ &= a_1\mu_1 + \dots + a_n\mu_n \end{aligned}$$

$$Var(Y) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j)$$

2. Si las X_i son independientes se cumple que

$$\begin{aligned} Var(Y) &= Var(a_1X_1 + \dots + a_nX_n) &= \\ &= a_1^2 Var(X_1) + \dots + a_n^2 Var(X_n) &= \\ &= a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2 \end{aligned}$$

Corolario 4.1.6 Se cumple que $E(X_1 - X_2) = E(X_1) - E(X_2)$ y, si X_1 y X_2 son independientes, se cumple que $Var(X_1 - X_2) = Var(X_1) + Var(X_2)$.

Ejercicio 4.1 Dada una v.a. X , demostrar que $Var(X + X) \neq Var(2X)$. ¿Por qué ocurre esto? ■

Proposición 4.1.7 Si $X_1, \dots, X_n$, con $X_i \sim N(\mu_i, \sigma_i^2)$, son v.a. independientes con medias y varianzas posiblemente distintas, entonces toda comb. lin. de las X_i sigue una distribución normal con media y varianza según la proposición anterior.

Esto es particularmente aplicable al caso de diferencia de dos variables aleatorias.

Proposición 4.1.8 Sea $X_1, \dots, X_n$ una m.a.s. de una v.a. $X \sim N(\mu, \sigma^2)$. Entonces para todo $n \in \mathbb{N}$, se cumple que

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \text{y} \quad T_0 \sim N(n\mu, n\sigma^2)$$

4.2 Estimación puntual

Mediante las técnicas de estimación puntual se pretende obtener conclusiones generales a partir de un conjunto limitado de datos. Cuando se infiere no se tiene ninguna garantía ni de la corrección ni de lo ajustada que es realmente la conclusión a la que se llega. Sin embargo mediante las técnicas de estimación puntual sí podemos tener algún conocimiento sobre las cualidades del error asociado a la estimación.

De forma general, las distribuciones de probabilidad están definidas por cierto número de parámetros ($N(\mu, \sigma^2)$, $Exp(\lambda)$, etc.) y son, entre otros, esos parámetros los que de forma general estamos interesados en estimar a partir de los datos. El objetivo de la estimación puntual es usar una muestra para obtener valores que, en algún sentido, sean los que mejor representan a los verdaderos valores de los parámetros de interés.

4.2.1 Algunos conceptos generales de la estimación puntual

Dado un parámetro de interés, como por ejemplo la media poblacional μ , el objetivo de la estimación puntual es usar los valores de una muestra para calcular un número que represente en cierto sentido una buena estimación del valor real del parámetro.

Definición 4.2.1 Una **estimación puntual** de un parámetro θ es un sólo número, $\hat{\theta}$, que se puede considerar como valor aproximado (en el sentido probabilístico, no de distancia) de θ . Una estimación puntual, $\hat{\theta}$, se obtiene al seleccionar un estadístico adecuado, $\hat{\theta}$, y calcular su valor a partir de los datos muestrales proporcionados. El estadístico seleccionado se llama **estimador puntual** de θ .

■ **Ejemplo 4.3** Si $X \sim N(\mu, \sigma)$ entonces el estadístico

$$\hat{\mu}(X_1, X_2, \dots, X_n) = \frac{X_1 + \dots + X_n}{n}$$

es un estimador para μ y por tanto, para una muestra concreta $x_1, \dots, x_n$, el valor

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

es una estimación de μ . ■

En el mejor de los casos, podríamos encontrar un estimador $\hat{\theta}$ para el cual siempre se tiene que $\hat{\theta} = \theta$. Sin embargo, $\hat{\theta}$ es una función de las X_i , muestrales, por lo que es una variable aleatoria. En general, $\hat{\theta}$ produce un valor distinto para cada muestra que obtengamos (y distinto a θ). Si escribimos

$$\hat{\theta} = \theta + \text{error de estimación}$$

entonces un estimador preciso sería uno que dé como resultado errores de estimación pequeños de modo que los valores estimados estarán cerca del valor real.

Estimadores Insesgados

Definición 4.2.2 Se dice que un estimador puntual $\hat{\theta}$ es un **estimador insesgado** de θ si $E(\hat{\theta}) = \theta$. Si $\hat{\theta}$ no es insesgado, a la diferencia $E(\hat{\theta}) - \theta$ la denominamos **sesgo** de $\hat{\theta}$.

Es decir, $\hat{\theta}$ es insesgado si su distribución de probabilidad (es decir, muestreo) siempre está “centrada” en el valor verdadero del parámetro.

Proposición 4.2.1 Cuando X es una variable aleatoria binomial con parámetros n y p , la proporción muestral $\hat{p} = X/n$ es un estimador insesgado de p .

El **Principio de Estimación Insesgada (PEI)** nos dice que al elegir entre varios estimadores de θ es en general preferible seleccionar uno que sea insesgado.

Proposición 4.2.2 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2 , entonces el estimador

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

es un estimador insesgado de σ^2 .

Proposición 4.2.3 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución con media μ , entonces $\bar{X}$ es un estimador insesgado de μ . Si además la distribución es continua y simétrica, entonces cualquier media recortada es también un estimador insesgado de μ .

Supongamos que $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores de θ , ambos insesgados. Entonces, aunque la distribución de cada estimador se centra en el valor verdadero de θ , las dispersiones de las distribuciones con respecto al valor verdadero podrían ser diferentes.

Definición 4.2.3 — Eficiencia. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados para estimar θ . Diremos que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$.

El **Principio de Estimación Inesgada con Varianza Mínima** nos dice que entre los estimadores insesgados de θ , es preferible elegir aquel que tiene varianza mínima; denominándose el estimador resultante **estimador insesgado de varianza mínima (EIVM)**.

Teorema 4.2.4 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución normal de parámetros μ y σ . Entonces el estimador $\hat{\mu} = \bar{X}$ es el EIVM para μ .

Observación 4.2

El teorema anterior no dice que al estimar una media poblacional μ , el estimador $\bar{X}$ se debe usar sin tomar en cuenta la distribución que está siendo muestreada.

En general, el mejor estimador (en el sentido del de menor varianzas) para μ depende mucho de qué distribución esté siendo muestreada.

4.2.2 Métodos de estimación puntual

Hasta ahora hemos hablado de qué son, cómo funcionan y qué propiedades tienen los estimadores puntuales sin embargo nada se ha dicho sobre cómo encontrar un estimador de un parámetro que tenga alguna probabilidad de ser un estimador razonable; esto es que al menos tenga sentido en función de las características de las observaciones que se utilizarán para su cálculo.

En lo que sigue describiremos dos métodos para calcular estimadores puntuales, el método de los momentos y el método de máxima verosimilitud. El primero, se basa en hacer corresponder tantos momentos muestrales y poblacionales como sean necesarios para su cálculo y el segundo se basa en maximizar la probabilidad del conjunto de datos observado.

Método de los momentos

La idea básica del **Método de los momentos** es igualar ciertas características muestrales, con los correspondientes valores esperados poblacionales. La resolución de estas ecuaciones acaba generando los estimadores.

Definición 4.2.4 Dada una v.a. X con función de probabilidad $f(x)$, el **k-ésimo momento poblacional** o **k-ésimo momento de la distribución $f(x)$** , es $E(X^k)$, con $k = 1, 2, 3, \dots$

Definición 4.2.5 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución con fdp $f(x)$, el **k-ésimo momento muestral** es $\frac{1}{n} \sum_{i=1}^n X_i^k$, con $k = 1, 2, 3, \dots$

■ **Ejemplo 4.4** Así, dadas $X_1, X_2, \dots, X_n$ una muestra aleatoria, el primer momento poblacional de X es $E(X) = \mu$ y el primer momento muestral es $\sum_{i=1}^n \frac{X_i}{n} = \bar{X}$.

Los segundos momentos poblacionales y muestrales son $E(X^2)$ y $\sum_{i=1}^n \frac{X_i^2}{n}$, respectivamente. ■

Definición 4.2.6 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución fdp $f(x; \theta_1, \theta_2, \dots, \theta_m)$, donde $\theta_1, \theta_2, \dots, \theta_m$ son parámetros cuyos valores se desconocen. Entonces los **estimadores de momento** $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ se obtienen igualando los primeros m momentos muestrales con los primeros m momentos poblacionales correspondientes y resolviendo para $\theta_1, \theta_2, \dots, \theta_m$.

■ **Ejemplo 4.5** Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de tiempos de servicio de n clientes en cierta instalación. Cada X_i sigue una distribución exponencial con parámetro λ . Estimar dicho parámetro por el método de los momentos.

Solución

Como sólo hay que estimar un parámetro, basta igualar los primeros momentos (muestral y poblacional)

$$\text{Por una parte } \mu = E[X] = \frac{1}{\lambda}$$

$$\text{Por otra parte } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\text{Igualando se tiene } \hat{\lambda} = \frac{1}{\bar{X}}.$$

■

Método de la máxima verosimilitud

Definición 4.2.7 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria con fdp conjunta

$$V(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) \quad (4.1)$$

donde los parámetros $\theta_1, \dots, \theta_m$ tienen valores desconocidos. Cuando $x_1, \dots, x_n$ son los valores muestrales observados y (4.1) es considerada como una función de $\theta_1, \dots, \theta_m$ se llama **función de verosimilitud**. Las estimaciones de máxima verosimilitud (emv) $\hat{\theta}_1, \dots, \hat{\theta}_m$ son los valores de las θ_i que maximizan la función de verosimilitud, así que

$$V(x_1, \dots, x_n; \hat{\theta}_1, \dots, \hat{\theta}_m) \geq V(x_1, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) \quad \text{para toda } \theta_1, \dots, \theta_m$$

Cuando las x_i se sustituyen por las X_i obtenemos los **estimadores de máxima verosimilitud**.

La función de verosimilitud señala qué tan probable es la muestra observada como función de los valores de parámetros posibles. Al maximizar la probabilidad se obtienen los valores de parámetros para los cuales la muestra observada tiene más probabilidades de haber sido generada; es decir, los valores de parámetros que “*mejor justifican*” los datos observados.

Proposición 4.2.5 Sea X una variable aleatoria con función de densidad $f(x; \theta)$ o distribución de probabilidad $P[X = x; \theta]$ y $X_1, X_2, \dots, X_n$ una muestra aleatoria simple. La distribución conjunta de $X_1, X_2, \dots, X_n$ será:

$$V(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta) \quad \text{ó}$$

$$P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta] = \prod_{i=1}^n P[X_i = x_i; \theta]$$

según que la variable aleatoria sea continua o discreta.

Estimar funciones de parámetros

Proposición 4.2.6 — Principio de Invarianza. Sean $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ los estimadores de máxima verosimilitud de los parámetros $\theta_1, \theta_2, \dots, \theta_m$. Entonces la estimación de máxima verosimilitud de cualquier función $h(\theta_1, \theta_2, \dots, \theta_m)$ de estos parámetros es la función $h(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)$ de las estimaciones de máxima verosimilitud.

■ **Ejemplo 4.6** Si conociéramos el E.M.V. para la varianza, $\hat{\sigma}^2$, el E.M.V. para la desviación típica sería $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$. ■

Proposición 4.2.7 En condiciones muy generales de la distribución conjunta de la muestra, si el tamaño de la muestra n es grande, el estimador de máxima verosimilitud de cualquier parámetro θ es aproximadamente insesgado [$E(\hat{\theta}) \approx \theta$] y tiene varianza que es casi tan pequeña como se pueda lograr mediante cualquier estimador. Expresado de otra manera, la estimación de máxima verosimilitud de $\hat{\theta}$ es aproximadamente el estimador insesgado de varianza mínima.

4.3 Estimación por intervalos de confianza

Como en la sección anterior en esta vamos a estudiar herramientas para estimar un determinado parámetro de una población como puede ser la media, la varianza, etc. a partir de una muestra de tamaño N . Sin embargo, a diferencia de la estimación puntual donde mediante la estimación determinábamos un valor concreto, con la estimación por intervalos el planteamiento muy diferente. Lo que haremos es determinar un intervalo donde afirmaremos que en su interior se encontrará el parámetro a estimar, con una probabilidad de acertar previamente fijada y que trataremos que sea la mayor posible, es decir próxima a 1.

De este modo las conclusiones de los resultados del estudio presentarán en vez de un valor concreto un intervalo de valores al cual se espera que pertenezca el parámetro con un nivel de confianza predeterminado.

4.3.1 Propiedades básicas

Definición 4.3.1 Sean $X_1, X_2, \dots, X_n$ una m.a.s. de variables aleatorias cuya función de distribución depende de un parámetro θ , y sean $l_1(X_1, X_2, \dots, X_n)$ y $l_2(X_1, X_2, \dots, X_n)$ dos estadísticos tales que $l_1 \leq l_2$. Un **intervalo** $(l_1(X_1, X_2, \dots, X_n), l_2(X_1, X_2, \dots, X_n))$, cuyos límites dependen de la muestra observada $x_1, x_2, \dots, x_n$, se denomina **intervalo aleatorio** de límites l_1 y l_2 .

Definición 4.3.2 Un intervalo aleatorio $(l_1(X_1, X_2, \dots, X_n), l_2(X_1, X_2, \dots, X_n))$ se denomina **intervalo de confianza** para θ con nivel de confianza $100(1 - \alpha)\%$, ($0 < \alpha < 1$), si

$$P(l_1(X_1, \dots, X_n) < \theta < l_2(X_1, \dots, X_n)) = 1 - \alpha.$$

$l_1(X_1, X_2, \dots, X_n)$ y $l_2(X_1, X_2, \dots, X_n)$ son los límites de confianza inferior y superior. En la definición se observa que α es la probabilidad de que θ no esté incluido en el intervalo.

4.3.2 Interpretación de los intervalos de confianza

Un nivel de confianza al $100(1 - \alpha)\%$ no es una declaración acerca de algún intervalo particular, sino de que si se construyera un número muy grande de intervalos con los límites estimados según esos estimadores, $(l_1(X_1, X_2, \dots, X_n), l_2(X_1, X_2, \dots, X_n))$, el valor del parámetro poblacional a estimar θ estaría en el interior de dichos intervalos en el $100(1 - \alpha)\%$ de las ocasiones.

■ **Ejemplo 4.7** Dada una muestra $x_1, x_2, \dots, x_n$, de una *fdp* con desviación típica σ , el intervalo de confianza al 95 % para estimar la media poblacional μ es

$$\left(\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}} \right).$$

Este intervalo, que cambia con cada muestra, cumple que si calculamos un número grande intervalos de confianza para distintas muestras, la media poblacional μ estará contenida en aproximadamente el 95 % de esos intervalos de confianza. ■

Definición 4.3.3 Los intervalos de la forma $(-\infty, l_2)$ (resp. $(l_1, +\infty)$) cumpliendo que $P(\theta < l_2) = 1 - \alpha$ (resp. $P(\theta > l_1) = 1 - \alpha$) se denominan **intervalos de confianza unilaterales**.

4.3.3 Obtención de un intervalo de confianza

En general para construir un intervalo de confianza para un parámetro θ se procede de la siguiente manera:

- Se busca una función (estadístico) T que dependa de la muestra pero no del parámetro θ (ni de ningún otro parámetro desconocido).
- Se determinan dos números reales ε_1 y ε_2 tales que $P(\varepsilon_1 < T < \varepsilon_2) = 1 - \alpha$.
- A partir de ese intervalo y por transformaciones algebraicas obtenemos el intervalo $(l_1(X_1, X_2, \dots, X_n), l_2(X_1, X_2, \dots, X_n))$ tal que $P(l_1(X_1, \dots, X_n) < \theta < l_2(X_1, \dots, X_n)) = 1 - \alpha$.
 $(l_1(X_1, X_2, \dots, X_n), l_2(X_1, X_2, \dots, X_n))$, es el intervalo de confianza para θ con un nivel de confianza de $1 - \alpha$ o $100(1 - \alpha)\%$.

■ Ejemplo 4.8 — Intervalos de confianza para μ – muestras grandes.

Sea $X_1, X_2, \dots, X_n$ una m.a.s. de una población que tiene media μ y desviación estándar σ . Si n es grande, el Teorema Central del Límite nos dice que $\bar{X}$ sigue una distribución aproximadamente normal, de media μ y desviación estándar $\sigma/\sqrt{n}$.

- Como $\bar{X}$ es un estimador para μ y además $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, se tiene que $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ es un estadístico que depende de la muestra y no depende de ningún parámetro.
- Fijado el nivel de confianza $1 - \alpha$, podemos determinar ε_1 y ε_2 tales que $P(\varepsilon_1 < Z < \varepsilon_2) = 1 - \alpha$. Los números ε_1 y ε_2 no son únicos pero determinan un intervalo de amplitud mínima si $\varepsilon_1 = -\varepsilon_2$. Así pues, determinaremos $\varepsilon_2 / P(-\varepsilon_2 < Z < \varepsilon_2) = 1 - \alpha$

Por la simetría de la distribución $N(0, 1)$, Si $P(Z < \varepsilon_2) = 1 - \frac{\alpha}{2}$ entonces $\varepsilon_2 = z_{1-\frac{\alpha}{2}}$.

- Por tanto se cumple que,

$$\begin{aligned} P(-z_{1-\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}) &= 1 - \alpha \Leftrightarrow \\ \Leftrightarrow P(-z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \Leftrightarrow \\ \Leftrightarrow P(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \end{aligned}$$

Se obtiene, finalmente que el intervalo de confianza para μ es

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

■

Nota 4.3.1 Si la muestra es grande y σ es desconocida, podemos utilizar el estadístico

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

y el intervalo de confianza para μ sería

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \right)$$

Nota 4.3.2 Estas fórmulas son válidas independientemente de la distribución de la población.

4.3.4 Int. de confianza en poblaciones normales

Si la muestra es grande y σ es desconocida, el estadístico

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

sigue una distribución normal. Ahora bien si n es pequeña, no es probable S que esté lo cerca de σ .

Proposición 4.3.3 Cuando $\bar{X}$ es la media de una muestra aleatoria de tamaño n de una distribución normal con media μ , la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución de probabilidad llamada **t de Student** con $n - 1$ grados de libertad.

Teorema 4.3.4 — Propiedades de las distribuciones t .

Sea t_n la curva de la función de densidad de una t de Student con n grados de libertad. Entonces:

- Cada curva t_n es en forma de campana y está centrada en el valor 0.
- Toda curva t_n es más dispersa que la curva normal estándar (z).
- Cuando aumenta n disminuye la dispersión de la curva t_n correspondiente.
- Cuando $n \rightarrow \infty$, la secuencia de curvas t_n se aproxima a la normal estándar.

Definición 4.3.4 Sea $t_{\alpha,n}$ el número en el eje de abcisas para el cual el área bajo la curva t_n a la izquierda de $t_{\alpha,n}$ es α , es decir, $P[T \leq t_{\alpha,n}] = \alpha$.

Int. de confianza para μ con σ desconocida

La variable estandarizada T tiene una distribución t con $n - 1$ grados de libertad, y el área bajo la curva de densidad t correspondiente entre los valores $-t_{1-\alpha/2,n-1}$ y $t_{1-\alpha/2,n-1}$ es $1 - \alpha$. Por tanto

$$P(-t_{1-\frac{\alpha}{2},n-1} < T < t_{1-\frac{\alpha}{2},n-1}) = 1 - \alpha$$

Si $X \sim N(\mu, \sigma)$ y $X_1, X_2, \dots, X_n$ es una m.a.s. de X , sabemos que $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$.

Proposición 4.3.5 Sean $\bar{x}$ y s la media y desviación estándar muestrales calculadas a partir de los resultados de una muestra aleatoria de una población normal con media μ . Entonces un **intervalo de confianza** del $100(1 - \alpha)\%$ para μ es

$$\left(\bar{x} - t_{1-\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \right)$$

o, bien, en forma más compacta $\bar{x} \pm t_{1-\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$.

Intervalo de confianza para la varianza

Teorema 4.3.6 Sea $X_1, X_2, \dots, X_n$ una m.a.s. de una distribución normal con parámetros μ y σ^2 . Entonces la variable aleatoria

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2}$$

tiene una distribución de probabilidad **chi-cuadrada** (χ^2) con $n - 1$ grados de libertad.

Definición 4.3.5 Sea $\chi_{\alpha,n}^2$, denominado **valor crítico chi-cuadrada**, el número en el eje de abcisas para el cual el área bajo la curva χ^2 con n grados de libertad a la derecha de $\chi_{\alpha,n}^2$ es α , es decir, $P[\chi^2 \leq \chi_{\alpha,n}^2] = \alpha$.

Teorema 4.3.7 Un **intervalo de confianza** de $100(1 - \alpha) \%$ para la varianza σ^2 de una población normal es

$$\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}} \right)$$

Un **intervalo de confianza para σ** tiene por límites inferior y superior las raíces cuadradas de los límites inferior y superior correspondientes a la varianza.

4.4 Ejercicios

Ejercicio 4.2 El diámetro de un anillo de pistón seleccionado al azar es una variable aleatoria con valor medio de 12 cm y desviación estándar de 0.04 cm.

- Si $\bar{X}$ es el diámetro medio muestral para una muestra aleatoria de $n = 16$ anillos, ¿dónde está centrada la distribución de muestreo de $\bar{X}$ y cuál es la desviación estándar de la distribución $\bar{X}$?
- Conteste las preguntas planteadas en el inciso (a) para un tamaño de muestra $n = 64$ anillos.
- ¿Para cuál de las dos muestras aleatorias, la del inciso (a) o la del inciso (b), $\bar{X}$ tiene más probabilidad de estar entre 0.01 cm de 12 cm? Explique su razonamiento.

Ejercicio 4.3 La resistencia a la ruptura de un remache tiene un valor medio de 10000 lb/pulg² y una desviación estándar de 500 lb/pulg².

- ¿Cuál es la probabilidad de que la resistencia media muestral a la ruptura para una muestra aleatoria de 40 remaches esté entre 9900 y 10200?
- Si el tamaño muestral hubiera sido 15 en vez de 40, ¿podría haberse calculado la probabilidad pedida en el inciso (a) a partir de la información suministrada?

Ejercicio 4.4 El tiempo que tarda una persona seleccionado al azar en rellenar los impresos de una hipoteca tiene una distribución normal con valor medio de 10 minutos y desviación estándar de dos minutos. Si en un día determinado cinco individuos rellenan un impreso y seis en otro día, ¿cuál es la probabilidad de que la cantidad de tiempo promedio de la muestra de cada día sea a lo sumo 11 minutos?

Ejercicio 4.5 Sean X_1 , X_2 y X_3 los tiempos necesarios para llevar a cabo tres tareas de reparación sucesivas en cierto taller. Suponga que los tiempos son variables aleatorias independientes, normales, con valores esperados μ_1 , μ_2 y μ_3 y varianzas σ_1^2 , σ_2^2 y σ_3^2 , respectivamente.

- Si $\mu_1 = \mu_2 = \mu_3 = 60$ y $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 15$, calcule $P(X_1 + X_2 + X_3 \leq 200)$ y $P(150 \leq X_1 + X_2 + X_3 \leq 200)$
- Por medio de las μ_i , y las σ_i del inciso (a), calcule $P(55 \leq \bar{X})$ y $P(58 \leq \bar{X} \leq 62)$.

- c. Con las μ_i , y las σ_i del inciso (a), calcule $P(-10 \leq X_1 - 0,5X_2 - 0,5X_3 \leq 5)$.
- d. Si $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ y $\sigma_3^2 = 14$, calcule $P(X_1 + X_2 + X_3 \leq 160)$ y $P(X_1 + X_2 \geq 2X_3)$.

Ejercicio 4.6 Cinco automóviles del mismo tipo serán conducidos en un viaje de 300 millas. Los primeros dos utilizarán una marca económica de gasolina, y los otros tres automóviles, gasolina de marca registrada. Sean X_1, X_2, X_3, X_4 y X_5 los rendimientos de combustible observados (mpg) para los cinco automóviles. Suponer que estas variables son independientes y tienen una distribución normal con $\mu_1 = \mu_2 = 20$, $\mu_3 = \mu_4 = \mu_5 = 21$ y $\sigma^2 = 4$ para la marca económica y 3.5 para la de marca registrada. Definimos la variable Y mediante

$$Y = \frac{X_1 + X_2}{2} - \frac{X_3 + X_4 + X_5}{3}$$

de modo que Y es una medida de la diferencia de rendimiento entre la gasolina económica y la de marca registrada. Se pide:

- (a) $P(0 \leq Y)$.
- (b) $P(-1 \leq Y \leq 1)$.

Ejercicio 4.7 La manufactura de cierto componente requiere tres operaciones consecutivas. El tiempo para cada operación tiene una distribución normal y los tres tiempos son independientes entre sí. Los valores medios son 15, 30 y 20 minutos, respectivamente y las desviaciones estándar son 1, 2 y 1.5 minutos. ¿Cuál es la probabilidad de que tome a lo sumo una hora producir un componente seleccionado al azar?

Ejercicio 4.8 Una pieza de tubo de PVC se va a insertar dentro de otra pieza. La longitud de la primera tiene una distribución normal con valor medio de 20 pulgadas, y desviación estándar de 0.5 pulgadas. la longitud de la segunda pieza es una variable aleatoria normal con media y desviación estándar de 15 y 0.4 pulgadas, respectivamente. La cantidad de traslape tiene una distribución normal con valor medio de una pulgada y desviación estándar de 0.1 pulgada. Si se supone que las longitudes y cantidad de traslape son independientes entre sí, ¿cuál es la probabilidad de que la longitud total después de la inserción esté entre 34.5 y 35 pulgadas?

Ejercicio 4.9 Considere una muestra aleatoria $X_1, \dots, X_n$ de la fdp

$$f(x; \theta) = 0,5(1 + \theta x) \quad -1 \leq x \leq 1$$

donde $-1 \leq \theta \leq 1$. Demuéstrese que $\hat{\theta} = 3\bar{X}$ es un estimador insesgado de θ .
Sugerencia: Primero determine $\mu = E(X) = E(\bar{X})$.

Ejercicio 4.10 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución de Rayleigh

con fdp

$$f(x; \theta) = \frac{x}{\theta} e^{-x^2/(2\theta)} \quad x > 0$$

- Demostrar que $E(X^2) = 2\theta$ y utilizar este hecho para construir un estimador insesgado de θ con base en $\sum X_i^2$ (y use las reglas de valor esperado para demostrar que es insesgado).
- Estime θ de las siguientes $n = 10$ observaciones sobre esfuerzo vibratorio de una paleta de turbina bajo condiciones específicas:

16.88	10.23	4.59	6.66	13.68
14.23	19.87	9.40	6.51	10.95

Ejercicio 4.11 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución exponencial con parámetro λ . Calcular el estimador de máxima verosimilitud para λ .

Ejercicio 4.12 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una distribución normal. Estimar μ y σ^2 por el método de la máxima verosimilitud.

Ejercicio 4.13 Suponga que el crecimiento promedio real μ de un tipo de planta durante un año es idéntico al de un segundo tipo, pero la varianza de crecimiento para el primer tipo es σ^2 , mientras que para el segundo tipo la varianza es $4\sigma^2$. Sean $X_1, \dots, X_m$, m observaciones de crecimiento independientes en el primer tipo [entonces, $E(X_i) = \mu$, $V(X_i) = \sigma^2$], y sean $Y_1, \dots, Y_n$, n observaciones de crecimiento independientes en el segundo tipo [$E(Y_i) = \mu$, $V(Y_i) = 4\sigma^2$].

- Demuestre que para cualquier δ entre 0 y 1, el estimador $\hat{\mu} = \delta \bar{X} + (1 - \delta) \bar{Y}$ es insesgado para μ .
- Para m y n , calcule $V(\hat{\mu})$, y luego determine el valor de δ que minimice a $V(\hat{\mu})$.

Ejercicio 4.14 Sea que X indica la proporción de tiempo asignado que un estudiante seleccionado al azar emplea trabajando en cierta prueba de aptitud. Suponga que la fdp de X es

$$f(x; \theta) = \begin{cases} (\theta + 1)x^\theta & 0 \leq x \leq 1, \\ 0 & \text{caso contrario} \end{cases}$$

donde $-1 < \theta$. Una muestra aleatoria de diez estudiantes produce los datos $x_1 = 0,92$, $x_2 = 0,79$, $x_3 = 0,90$, $x_4 = 0,65$, $x_5 = 0,86$, $x_6 = 0,47$, $x_7 = 0,73$, $x_8 = 0,97$, $x_9 = 0,94$, $x_{10} = 0,77$.

- Utilice el método de momentos para obtener un estimador de θ , y después calcule la estimación para estos datos.
- Obtenga el estimador de máxima verosimilitud de θ , y después calcule la estimación para los datos proporcionados.

Ejercicio 4.15 Se determina la resistencia al corte de cada una de diez soldaduras de punto, obteniéndose los datos siguientes (lb/pulg²)

392 376 401 367 389 362 409 415 358 375

- Si se supone que la resistencia al corte tiene una distribución normal, estime la resistencia al corte promedio real y su desviación estándar con el método de máxima verosimilitud.
- Suponga de nuevo una distribución normal y estime el valor de resistencia abajo del cual 95 % de las soldaduras tendrán sus resistencias (*Sugerencia: ¿cuál es el percentil 95 en términos de μ y σ ? Ahora utilice el principio de invarianza*).

Ejercicio 4.16 Una muestra aleatoria de 110 relámpagos dio como resultado una duración de eco de radar promedio muestral de 0.81 s y una desviación estándar muestral de 0.34 s. Calcular un intervalo de confianza (bilateral) de 99 % para la duración de eco promedio real μ e interprete el intervalo resultante.

Ejercicio 4.17 La resistencia a la fractura (en MPa) de $n = 169$ barras de cerámica cocidas en determinado horno da una media $\bar{x} = 89,10$ y una desviación estándar $s = 3,73$.

- Calcular un intervalo de confianza (bilateral) para la resistencia a la fractura promedio real usando un nivel de confianza del 95 %. ¿Parece que se estimó con precisión la resistencia a la fractura promedio real?
- Suponer que los investigadores creían, a priori, que la desviación estándar de la población era 4 MPa. Con esta suposición, ¿qué tan grande se habría requerido una muestra para estimar μ hasta dentro de 0.5 MPa con confianza de 95 %?

Ejercicio 4.18 Con una muestra aleatoria de $n = 8$ especímenes de prueba de cierto tipo de fibra de vidrio E se obtuvo un límite elástico cortante interfacial promedio muestral de 30.2 y una desviación estándar muestral de 3.1. Si se supone que el límite elástico cortante interfacial está distribuido normalmente, calcular un intervalo de confianza de 95 % para el límite elástico promedio real.

Ejercicio 4.19 Con una muestra aleatoria de $n = 14$ especímenes de uniones de cierto tipo se obtuvo un esfuerzo límite proporcional medio muestral de 8.48 MPa y desviación estándar muestral de 0.79 MPa.

- Calcular e interpretar un límite de confianza inferior de 95 % para el esfuerzo límite proporcional promedio real de todas uniones. ¿Qué hipótesis, si es que alguna resultó pertinente, planteó acerca de la distribución del esfuerzo límite proporcional?
- Calcular e interpretar un límite inferior de predicción de 95 % para el esfuerzo proporcional límite de una sola unión de este tipo.

Ejercicio 4.20 Se determinó la cantidad de expansión lateral (milésimas de pulgada) para una muestra de $n = 9$ soldaduras de arco gas-metal activado por impulsos que se utilizan en buques tanque que transportan gas natural licuado. La desviación muestral resultante fue $s = 2,81$ milésimas de pulgada. Suponiendo que se trata de una distribución normal, obtener un intervalo de confianza de 95 % para σ^2 y para σ . ■

Ejercicio 4.21 Las observaciones siguientes se hicieron en relación con la resistencia a la fractura de una placa base de 18 % de acero *maragizado* con níquel.

69.5	71.9	72.6	73.1	73.3	73.5	75.5	75.7
75.8	76.1	76.2	76.2	77.0	77.9	78.1	79.6
79.7	79.9	80.1	82.2	83.7	93.7		

Calcular un intervalo de confianza de 99 % para σ . ¿Es válido este intervalo cualquiera que sea la naturaleza de la distribución? ■

Ejercicio 4.22 Se ha llevado a cabo un análisis de nicotina en una marca de cigarrillos y se ha encontrado en una muestra de 100 cigarrillos una media $\bar{x} = 26$ mg. Si se sabe que la desviación típica de la marca es 8 mg. , calcular un intervalo de confianza del 99 % para el contenido medio en nicotina de la marca.

¿Cuál debería ser el tamaño de la muestra para que el intervalo disminuya en 2 mg.? (Suponer normalidad) ■

Ejercicio 4.23 Determinar un intervalo con un nivel de confianza del 95 % para la media de una distribución normal con varianza 9, usando una muestra de tamaño $n=100$ valores con media $\bar{x} = 5$.

¿Qué valor debe tomar n si deseamos obtener un intervalo con un nivel de confianza del 95 % y longitud $L=0.4$? ■

Ejercicio 4.24 Calcular qué tamaño de muestra se necesita para producir un intervalo de confianza al nivel 99 % para la media de una distribución normal de longitud

- (a) 2σ ,
- (b) σ ,
- (c) $0,2\sigma$,
- (d) $0,02\sigma$

Ejercicio 4.25 Una muestra de 200 personas tiene una estatura media de 1,72 m. y varianza de 0,52 m. Suponiendo que proceden de una población normal , obtener un intervalo de confianza para la media al 98 %. ■

Ejercicio 4.26 Calcular un intervalo de confianza para la media de una normal con desviación típica de 3 sabiendo que los valores de una muestra de dicha población son : 3.3 , 0.3 , 0.6 , 0.9. Calcular un intervalo en el caso de que la desviación típica sea

desconocida. ■

Ejercicio 4.27 Se sabe que la longitud media de una determinada especie de lagartija es de 12 cm. Para estimar la varianza de la longitud de la especie se toma una muestra de 30 lagartijas para las cuales se obtuvo una varianza de 3.4. Obtener un intervalo de confianza para la varianza poblacional al 96 %. (Suponer normalidad) ■

Ejercicio 4.28 Encontrar un intervalo de confianza para la media de una distribución uniforme, usando una muestra de tamaño 100 con media 0.5 y desviación típica 0.3. ■

Ejercicio 4.29 Una central de productos lácteos recibe diariamente la leche de dos granjas A y B. Deseando estudiar la calidad de los productos recibidos se extraen muestras al azar y se analiza el contenido en materia grasa, obteniéndose los siguientes resultados en tantos por cientos:

$$\begin{array}{llll} \text{granja A} & \bar{X}_A = 8,7\% & s_A^2 = 1,02\% & n_A = 33 \\ \text{granja B} & \bar{X}_B = 10,9\% & s_B^2 = 1,73\% & n_B = 27 \end{array}$$

Se pide construir un intervalo de confianza al 95 % para la diferencia del contenido medio en grasa de la leche de ambas granjas. ■

Ejercicio 4.30 Se quiere probar la efectividad de un antitérmico (reductor de la fiebre). Con tal fin se tomó la temperatura a 10 niños afectados de gripe, antes y después de la administración del antitérmico resultando una diferencia de medias de $\bar{X} - \bar{Y} = 1,58$ y una cuasidesviación típica de la diferencia $S_{cd} = 0,71$. Suponiendo que la característica en estudio es normal, calcular un intervalo de confianza del 90 % para la diferencia de medias. ¿Qué tamaño de muestra habría que tomar para tener una confianza del 95 % de que la disminución media de la fiebre estimada y la verdadera se diferencien en 10^{-2} ? ■

Ejercicio 4.31 Se sabe que la demanda diaria de un determinado medicamento sigue una distribución normal $N(\mu, \sigma^2)$. En un hospital, con objeto de garantizar las necesidades de dicho medicamento, se ha registrado durante 50 días el consumo diario de éste, resultando, para este período, una media de 60 unidades diarias con una varianza de 16 unidades.

- Obtener un intervalo de confianza para la media y otro para la varianza a un nivel del 95 %.
- En vista de los datos anteriores, ¿con qué cantidad mínima de dicho medicamento debe contarse cada día para garantizar con una probabilidad de 0,99 que las necesidades estarán cubiertas? ■

Ejercicio 4.32 En un estudio antropométrico sobre la longitud del fémur se han tomado dos muestras de tamaños 20 y 30 respectivamente, con mediciones de dicha longitud extraídas de dos poblaciones. Supuesto que la característica en estudio sigue distribu-

ciones normales en ambas poblaciones, calcular un intervalo de confianza al 95 % para el cociente de varianzas poblacionales, sabiendo que las varianzas de las muestras son $S_X^2 = 57$ y $S_Y^2 = 45$. ■

Ejercicio 4.33 Se desea estudiar el tiempo que tarda en hacer efecto dos tratamientos, A y B. Una m.a. compuesta por 10 pacientes sobre la que se aplicó el tratamiento A dio una media de 12.3 min. y una varianza de 6.5, mientras que los resultados para el tratamiento B fueron 10.1 min. y 4.1. Hallar un intervalo de confianza al 90 % para σ_A/σ_B . ■

Ejercicio 4.34 El Ayuntamiento de una ciudad se encuentra interesado en estimar la cantidad promedio de dinero que la población se gasta, al cabo del mes, en espectáculos. Se realizó una encuesta a 16 personas elegidas al azar obteniéndose (en euros) los siguientes datos: 15.00, 17.50, 16.30, 14.80, 14.20, 18.90, 13.50, 17.40, 16.80, 15.20, 15.80, 18.40, 13.40, 14.60, 15.50, 16.30. Si se supone que la cantidad de dinero gastada en un mes es una variable aleatoria que sigue una distribución normal, obtener los intervalos de confianza para la varianza con nivel de confianza del 98 %. ■

Ejercicio 4.35 La dirección de una clínica desea estimar el número promedio de días necesarios para el tratamiento de los pacientes. En una muestra aleatoria de 500 pacientes se observó una media de 5.4 días y una desviación típica de 3.1 días. Suponiendo normalidad, obtener un intervalo con nivel de confianza del 95 %, para el promedio de días de estancia en la clínica. ■

Ejercicio 4.36 En una muestra aleatoria simple de 50 programadores se ha observado una media de la diferencia entre el sueldo más bajo y el más alto por hora pagado a un mismo programador de 90 euros. Si se supone normalidad con varianza $\sigma^2 = 0'1$ euros², calcular un intervalo de confianza para la diferencia de sueldos al 95 %. ¿Cuál debe ser el tamaño de la muestra para que, con una probabilidad de 0'99 la media muestral se encuentre a no más de 0'1 euros del verdadero salario por hora promedio? ■

Ejercicio 4.37 Dos universidades tienen métodos diferentes para matricular a sus alumnos. En cada universidad se anotaron los tiempos de inscripción para 40 alumnos seleccionados al azar. Las medias y desviaciones típicas observadas fueron las siguientes: $\bar{X}_1 = 50'2$, $\bar{X}_2 = 52'9$, $S_1 = 4'8$ y $S_2 = 5'5$. Si se supone que el muestreo se llevó a cabo sobre dos poblaciones normales, obtener un intervalo de confianza al 90 % para la diferencia de medias supuesto que las varianzas de los tiempos de matriculación son iguales en ambas universidades. ■

4.4.1 Ejercicios de exámenes de años anteriores

Ejercicio 4.38 Sea que $X_1, X_2, \dots, X_n$ representa una muestra aleatoria de la distribu-

ción de Rayleigh con la función de densidad

$$f(x; \theta) = \frac{x}{\theta} e^{-x^2/(2\theta)} \quad x > 0$$

Determine

- a. El estimador de máxima verosimilitud de θ y después calcule la estimación para los datos de esfuerzo vibratorio

16.88	10.23	4.59	6.66	13.68
14.23	19.87	9.40	6.51	10.95

¿Este estimador es el mismo que el insesgado sugerido en el Ejercicio 15?

- b. La estimación de máxima verosimilitud de la mediana de la distribución del esfuerzo vibratorio. (*Sugerencia:* primero exprese la mediana en términos de θ).

Ejercicio 4.39 Sea $X_1, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria discreta X con función de probabilidad

$$P[X = x] = \left(\frac{\theta}{2}\right)^{|x|} (1 - \theta)^{1-|x|} \quad x \in \{-1, 0, 1\}$$

donde $\theta \in [0, 1]$ es un parámetro desconocido.

- (a) Determinar el estimador de máxima verosimilitud para estimar θ .
 (b) Comprobar que $E[X] = 0$ y $V[X] = \theta$.

Ejercicio 4.40 El tiempo promedio necesario para realizar el trabajo A es de 75 minutos, con una desviación típica de 15 minutos y para realizar el trabajo B hace falta en promedio 100 minutos, con una desviación típica de 10 minutos. Se supone que los tiempos de realización de los trabajos son independientes y se distribuyen normalmente.

- (a) Si comienzan a realizarse ambos trabajos a la vez, ¿cuál es la probabilidad de que se finalice el trabajo A más de 15 minutos antes que el B ?
 (b) Si se efectúan los trabajos uno a continuación de otro, con una pausa intermedia de 10 minutos, ¿cuál es la probabilidad de terminar la labor antes de 3 horas?

Ejercicio 4.41 En cierta población el peso de los hombres y el de las mujeres se distribuye normalmente, con distribuciones $N(70, 8^2)$ y $N(60, 6^2)$ respectivamente. Elegidos aleatoria e independientemente un varón y una mujer, ¿cuál es la probabilidad de que:

- (a) el hombre pese menos de 65 Kg y la mujer más de 62 Kg?
 (b) el hombre pese menos que la mujer?
 (c) el hombre pese más de 5 Kg más que la mujer?

Ejercicio 4.42 Se desea analizar las diferencias entre dos grupos de alumnos mediante la distribución en el muestreo de medias. Se toman muestras aleatorias independientes de ambas poblaciones para lo que dispondremos de la siguiente información distribuida en la tabla adjunta que se presenta a continuación:

	Tamaño de la muestra	Media poblacional	Desviación estándar poblacional
Grupo A	100	4.10	1.55
Grupo B	81	5.18	1.95

- Determinar la probabilidad de que la diferencia de medias muestrales sea mayor que 1.
- Determinar la probabilidad de que la diferencia de medias muestrales sea menor que 2.

Ejercicio 4.43 Una empresa dedicada al suministro de acceso a internet realizó un estudio acerca de la demanda del servicio por parte de sus clientes los fines de semana. Del estudio se dedujo que el número de usuarios atendidos por el servidor sigue una distribución $N(1500, 175^2)$ durante los sábados y $N(2000, 175^2)$ los domingos. Supuesto que el número de usuarios atendidos en sábado es independiente del número de usuarios atendidos en domingo:

- Determinar la probabilidad de que en domingo se atiendan a más usuarios que en sábado.
- Calcular la probabilidad de que el número de usuarios atendidos durante un fin de semana sea a lo sumo 3800.

Ejercicio 4.44 Un examen consta de tres ejercicios. Se estima que el tiempo, en minutos, que necesita un alumno para resolver el primer, segundo y tercer ejercicio siguen distribuciones normales $N(40, 10^2)$, $N(50, 20^2)$ y $N(60, 20^2)$, respectivamente

- ¿Cuál es la probabilidad de que un alumno resuelva los tres ejercicios en menos de dos horas?
- Si al examen se presentan 30 alumnos, ¿cuál es la probabilidad de que, a lo sumo, 3 alumnos tarden más en realizar el primer ejercicio que el tercero?

Nota: Por comodidad, redondear la probabilidad intermedia a la primera cifra decimal.

Ejercicio 4.45 El método utilizado por una compañía para la creación de un DVD consta de dos etapas. La primera consiste en la creación del menú mediante un programa de tratamiento de imágenes digitales. En la segunda etapa se añade el menú al DVD creándose la película final con el formato deseado. El tiempo, en minutos, que se requiere para realizar la primera etapa es una variable aleatoria $X \sim N(60, 10^2)$,

mientras que el tiempo, en minutos, requerido para realizar la segunda etapa es una variable aleatoria $Y \sim N(40, 5^2)$, siendo ambas variables aleatorias independientes. Se pide:

- Determinar la probabilidad de que sólo una de las dos etapas requiera más de 45 minutos.
- Determinar la probabilidad de que al crear un DVD se tarde en la primera etapa más tiempo que en la segunda.

Ejercicio 4.46 Dos cadenas de montaje A y B , convergen a un único taller de acabado que durante cinco minutos trata a una sola unidad. El tiempo que las unidades están en cada cadena es aleatorio y con distribución normal, independiente de una cadena a la otra, con parámetros

	min. estancia en Cadena A	min. estancia en Cadena B
Media	12	9
Desviación típica	4	3

Si a las 12 horas entra a la cadena A una unidad y ocho minutos después entra a la cadena B otra unidad

- ¿cuál es la probabilidad de que la unidad de B llegue antes al taller de acabado que la unidad de A ?
- ¿cuál es la probabilidad de que cuando la unidad de B llega al taller pueda ser tratada sin demora?

Ejercicio 4.47 Un examen de tipo test consta de diez preguntas con cuatro respuestas alternativas cada una, siendo sólo una de ellas correcta. Un alumno responde al azar a todas las preguntas del examen.

- ¿Cuál es la probabilidad de que haya acertado menos de 3 preguntas?
- Si cada respuesta correcta suma 1 punto y cada fallo resta 0.5 puntos, determinar la calificación esperada de este alumno.
- Calcular la desviación típica de la calificación.

Ejercicio 4.48 Sea $X_1, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de probabilidad:

$$P[X = x] = p^2 \cdot (1 - p)^{x-2} \cdot (x - 1) \quad x = 2, 3, 4, \dots$$

donde $p \in (0, 1)$ es un parámetro desconocido. Se sabe que $E[X] = \frac{2}{p}$ y $\text{var}(X) = \frac{2(1-p)}{p^2}$.

- Determinar el estimador de máxima verosimilitud de p .

- (b) Para muestras de tamaño $n = 3$ se consideran los estimadores

$$T_1 = \frac{X_1 + X_2 + X_3}{6} \quad T_2 = \frac{X_1 + 2X_2 + 3X_3}{12}$$

Demostrar que T_1 y T_2 son estimadores insesgados de $\frac{1}{p}$. ¿Cuál es más eficiente?

Ejercicio 4.49 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x) = 2\theta e^{-2\theta x} \quad \text{si } x > 0$$

donde $\theta > 0$ es un parámetro desconocido.

- (a) Determinar el estimador de máxima verosimilitud de θ .

$$2 \sum_{i=1}^n X_i^2$$

- (b) ¿Es el estimador $T = \frac{2 \sum_{i=1}^n X_i^2}{n}$ insesgado para estimar θ^{-2} ?

Ejercicio 4.50 Sea X una variable aleatoria discreta con función de probabilidad

$$P[X = x] = \frac{2 \cdot \theta^x}{x! \cdot (2-x)!} \cdot (1-\theta)^{2-x} \quad x \in \{0, 1, 2\}$$

donde $\theta \in (0, 1)$ es un parámetro desconocido.

- (a) Calcular $E[X]$ y $\text{Var}(X)$.
 (b) Determinar el estimador de máxima verosimilitud para estimar θ .
 (c) ¿Es el estimador de máxima verosimilitud insesgado para estimar θ ?

Ejercicio 4.51 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x) = \frac{1}{\sqrt{\theta}} e^{-x/\sqrt{\theta}} \quad \text{si } x > 0$$

donde $\theta > 0$ es un parámetro desconocido.

- (a) Determinar el estimador de máxima verosimilitud de θ .
 (b) Para muestras de tamaño 3 se consideran los estimadores $T_1 = \frac{X_1 + 2X_2 + 3X_3}{6}$ y $T_2 = \frac{X_1 + 3X_2 + 7X_3}{11}$
 (b.1) ¿Son insesgados para estimar $\sqrt{\theta}$?
 (b.2) ¿Cuál es más eficiente?

Ejercicio 4.52 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de probabilidad

$$f(x; \theta) = \frac{x}{\theta^2} \cdot e^{-x/\theta} \quad x \geq 0$$

donde $\theta > 0$ es un parámetro desconocido.

- Calcular un estimador de θ por el método de los momentos.
- Determinar el estimador de máxima verosimilitud para estimar θ .
- ¿Son los estimadores anteriores insesgados para estimar θ ?

■

Ejercicio 4.53 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x) = \frac{\theta \cdot 3^\theta}{x^{\theta+1}} \quad \text{si } x \geq 3$$

donde $\theta > 1$ es un parámetro desconocido.

- Determinar el estimador de máxima verosimilitud de θ .
- Determinar el estimador para estimar θ por el método de los momentos.

■

Ejercicio 4.54 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x, \theta) = \frac{2\theta}{1-\theta} \cdot x^{\frac{3\theta-1}{1-\theta}}, \quad 0 \leq x \leq 1,$$

donde $\theta > 0$ es un parámetro desconocido.

- Determinar el estimador de máxima verosimilitud para estimar θ .
- Obtener un estimador para estimar θ por el método de los momentos.

■

Ejercicio 4.55 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x, \theta) = \theta \left(\frac{1}{x} \right)^{\theta+1}, \quad x \geq 1,$$

donde $\theta > 1$ es un parámetro desconocido.

- Determinar el estimador de máxima verosimilitud para estimar θ .
- Obtener un estimador para estimar θ por el método de los momentos.
- Calcular las estimaciones de los parametros obtenidos en los apartados (a) y (b) para los valores de la muestra: $\{2, 3, 2, 2, 5, 1, 2, 2, 3, 1, 4, 6, 3, 4, 4\}$.

■

Ejercicio 4.56 El tiempo de vida en días X de los individuos de una población afectados de una nueva enfermedad es una variable aleatoria continua con función de densidad

$$f(x; \theta) = 2\theta^2 x^{-3}, \quad \text{si } x > \theta, \quad f(x; \theta) = 0, \quad \text{si } x \leq \theta$$

siendo $\theta > 0$ un parámetro desconocido.

- Calcular el estimador para θ por el método de la máxima verosimilitud.
- Obtener el estimador del parámetro θ por el método de los momentos.

■

Ejercicio 4.57 Sea X una variable aleatoria con función de probabilidad:

$$P[X = x] = e^{-(\lambda-1)^2} \cdot \frac{(\lambda-1)^{2x}}{x!} \quad \text{para } x = 0, 1, 2, \dots$$

donde $\lambda > 1$. Sea $X_1, x_2, \dots, X_n$ una muestra aleatoria simple procedente de X . Calcular el estimador de máxima verosimilitud para λ .

■

Ejercicio 4.58 Mediante estudios se sabe que el tiempo de espera, en minutos, para poder ser atendidos en los cajeros de una determinada entidad financiera, sigue una ley de probabilidad *Gamma*, cuya función de densidad viene dada por $f(x) = \begin{cases} \frac{1}{2\theta^3} x^2 e^{-x/\theta} & x > 0, \theta > 0 \\ 0 & \text{en el resto} \end{cases}$ siendo su valor esperado $E[X] = 3\theta$ y $V[X] = 3\theta^2$.

- Obtener el estimador de máxima verosimilitud para estimar θ .
- Una muestra de 20 clientes proporciona el siguiente tiempo de espera, en minutos, de cada uno de ellos: $\{12, 15, 18, 14, 13, 2, 15, 14, 8, 5, 10\}$. Se pide: Determinar la estimación del parámetro θ .

■

Ejercicio 4.59 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f(x) = \frac{1}{\theta} x^{\frac{1}{\theta}-1} \quad 0 \leq x \leq 1,$$

donde $\theta \neq 0$ es un parámetro desconocido.

- Determinar el estimador de máxima verosimilitud de θ .
- Obtener el estimador para θ por el método de los momentos.

■

Ejercicio 4.60 El tiempo X que un servidor de internet tarda en encontrar una determi-

nada información se distribuye según una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{1}{2}\lambda^3 x^2 e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

donde $\lambda > 0$ es un parámetro desconocido. Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de la variable X . Se pide:

- Determinar el estimador de máxima verosimilitud para estimar λ .
- Determinar el estimador por el método de los momentos.

■

Ejercicio 4.61 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de la variable X con función de densidad

$$f(x) = \frac{x}{\beta^2} e^{-x/\beta} \quad \text{si } x > 0$$

donde $\beta > 0$ es un parámetro desconocido.

- Determinar el estimador de máxima verosimilitud para estimar β .
- Demostrar que $T = \frac{1}{6n} \sum_{i=1}^6 X_i^2$ es un estimador insesgado de β^2 .

■

Ejercicio 4.62 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X continua con función de densidad

$$f(x; \theta) = \theta(1-x)^{\theta-1}, \quad 0 \leq x \leq 1; \quad f(x; \theta) = 0 \text{ en caso contrario}$$

donde $\theta > 0$ un parámetro desconocido.

- Calcular el estimador para θ por el método de la máxima verosimilitud.
- Obtener el estimador del parámetro θ por el método de los momentos.

■

Ejercicio 4.63 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple procedente de una variable aleatoria con función de densidad:

$$f(x; \theta) = \begin{cases} \frac{1}{2\theta - 1} \cdot e^{-x/(2\theta-1)} & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases}$$

donde $\theta > 0,5$ es un parámetro desconocido.

- Calcular el estimador de máxima verosimilitud para θ .
- ¿Es el estimador anterior insesgado para estimar θ ?

■

Ejercicio 4.64 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de probabilidad densidad

$$P[X = x] = \frac{\theta^x}{(1 + \theta)^{x+1}}; \quad x = 0, 1, 2, \dots$$

donde $\theta > 0$ es un parámetro desconocido. Se sabe que $E[X] = \theta$.

(a) Comprobar que P es una función de probabilidad.

Nota: Se recuerda que si $0 < r < 1$, $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$.

(b) Determinar el estimador de máxima verosimilitud para estimar θ .

(c) Para muestras de tamaño $n = 3$ se consideran los estimadores insesgados

$$T_1 = \frac{X_1 + 2X_2 + 3X_3}{6} \quad T_2 = \frac{X_1 + 3X_2 + 5X_3}{9}$$

¿Cuál es más eficiente?

Ejercicio 4.65 Sea $X_1, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad:

$$f(x) = \frac{\theta \cdot 2^\theta}{x^{\theta+1}} \quad \text{si } x > 2$$

donde $\theta > 0$ es un parámetro desconocido.

(a) Determinar el estimador de máxima verosimilitud de θ .

(b) Determinar el estimador para estimar θ por el método de los momentos.

Ejercicio 4.66 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad $f(x; \theta) = \frac{\theta}{x^{\theta+1}}$ para $x > 1$ y donde $\theta > 0$ es un parámetro desconocido.

(a) Obtener el estimador para estimar θ por el método de la máxima verosimilitud.

(b) Obtener el estimador para estimar θ por el método de los momentos.

Ejercicio 4.67 Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una variable X con función de densidad

$$f(x) = \begin{cases} \frac{x}{\beta^2} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

donde $\beta > 0$ es un parámetro desconocido. Se sabe que $E[X] = 2\beta$ y que $\text{Var}[X] = 2\beta^2$. Se pide:

(a) Determinar el estimador de máxima verosimilitud para estimar β .

- (b) ¿Es $T = \frac{1}{6n} \sum_{i=1}^n X_i^2$ un estimador insesgado de β^2 ?

Intervalos de Confianza

Ejercicio 4.68 Una persona desde su móvil llama a otro móvil. El tiempo que transcurre desde la llamada hasta la primera señal del otro sigue una distribución Normal de media μ y una varianza $\sigma^2 = 16$. Dicha persona ha llamado a n móviles elegidos al azar e independientemente. Sea $\bar{X}$ la media que se obtendrá de las n llamadas. Se pide:

- Si $n = 100$ calcular A para que con probabilidad 0.9 se tenga que $\mu \leq \bar{X} + A$.
- Calcular el tamaño n que debe tener la muestra para que con probabilidad 0.95 se tenga que $\mu \leq \bar{X} + 0.32$.
- Sean $n = 64$, $\bar{X} = 15$ y supongamos que σ es desconocida y que $S^2 = 9$. Calcular un intervalo de confianza al 95 % para el tiempo medio.



5. Contraste de Hipótesis

El método del contraste de hipótesis utiliza tests de significación para determinar la probabilidad de que una determinada hipótesis (a menudo relacionadas con la media o la varianza de una distribución) sea cierta y de con qué probabilidad, como estadísticos, la aceptaríamos como cierta.

5.1 Hipótesis y procedimientos de prueba

En lo que sigue entenderemos primero qué es y cómo realizar un contraste de hipótesis. Cómo formularlo y las diferentes formas de resolverlo. Comenzaremos formulando la hipótesis que queremos comprobar denominada la hipótesis alternativa en la que basaremos toda la carga de la prueba del test frente la hipótesis nula (la negación de la hipótesis alternativa). Finalmente estableceremos el nivel de probabilidad α –también denominado nivel de significación del test– que se corresponderá con el nivel de probabilidad máximo con el que estamos dispuestos a equivocarnos en caso de rechazar la hipótesis nula siendo esta la correcta.

Definición 5.1.1 Una **hipótesis estadística** es una afirmación acerca del valor de un sólo parámetro, característica de la población o característica de una distribución de probabilidad, o bien acerca de los valores de varios parámetros o de la forma de una distribución de probabilidad completa.

En este tema nos limitaremos al estudio afirmaciones o conjeturas torno a los **parámetros** de una distribución (**hipótesis paramétricas**).

Definición 5.1.2 La **hipótesis nula**, denotada con H_0 , es la afirmación que se supone al principio como cierta (la afirmación de “*creencia previa*”). La **hipótesis alternativa**, denotada por H_1 es la afirmación niega H_0 .

El **procedimiento de prueba** consiste en rechazar H_0 en favor de H_1 sólo si la evidencia muestral hace pensar que H_0 es falsa. Si la muestra obtenida no contradice de forma “contundente” H_0 , se continúa con la suposición de que la hipótesis H_0 es verdadera. Las dos conclusiones posibles de un análisis de prueba de hipótesis son entonces rechazar H_0 o no rechazar H_0 .

Notación 5.1.1 H_0 se expresará siempre como una afirmación de igualdad, $H_0 : \theta = \theta_0$; al valor θ_0 se le denomina *valor nulo* del parámetro.

La hipótesis alternativa, H_1 , se asemejará a una de las tres siguientes afirmaciones:

- (1) $H_1 : \theta > \theta_0$, (*contraste unilateral*)
- (2) $H_1 : \theta \neq \theta_0$, (*contraste bilateral*)
- (3) $H_1 : \theta < \theta_0$, (*contraste unilateral*)

En el primer caso la hipótesis nula implícita es realmente $\theta \leq \theta_0$ y en el tercero $\theta \geq \theta_0$

■ **Ejemplo 5.1** Sea $X \sim N(\mu, \sigma^2)$.

Hipótesis nulas	Hipótesis alternativas
$\mu = 10$	$\mu < 10$
$\sigma^2 = 9$	$\sigma^2 \neq 9$
$\sigma = 1$	$\sigma > 1$

■

5.1.1 Tipos de errores en un contraste

Al analizar un contraste existen dos posibles decisiones con respecto a H_0 (*rechazar H_0 o no rechazar H_0*). Ahora bien, cualquiera de las dos decisiones se hacen en función de los datos contenidos en una muestra y, por tanto, pueden dar lugar a un tipo de error.

Definición 5.1.3

1. El **error de tipo I** es el error que se comete al rechazar la hipótesis nula (H_0) cuando en realidad es verdadera.
2. El **error de tipo II** es el error que se comete al no rechazar la hipótesis nula (H_0) cuando en realidad es falsa.

Las probabilidades de error tipo I y tipo II se denominan, α y β , respectivamente.

$$\alpha = P[\text{Rechazar } H_0 | H_0 \text{ es cierta}]$$

$$\beta = P[\text{No Rechazar } H_0 | H_0 \text{ es falsa}]$$

Hay que notar que tanto α como β son probabilidades condicionadas. No pueden obtenerse las probabilidades de los errores de tipo I y tipo II en un sentido absoluto debido a que el estado de la naturaleza no es conocido.

5.1.2 Procedimientos de prueba

Definición 5.1.4 Toda regla, con respecto a alguna característica desconocida de la población, para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población se denomina **prueba** de una *hipótesis estadística*.

Definición 5.1.5 Un *procedimiento de prueba* se especifica mediante:

1. Un **estadístico de prueba**, una función de los datos muestrales en la que se basará la decisión (rechazar o no rechazar H_0).
2. Una **región de rechazo** (o **región crítica**), el conjunto de los valores del estadístico de prueba para los que se rechazará H_0 .

Entonces, la hipótesis nula será rechazada si y sólo si el valor del estadístico de prueba calculado pertenece a la región de rechazo.

■ **Ejemplo 5.2** Supongamos que queremos contrastar que la media μ de una población es menor o igual que 10. Para ello establecemos como hipótesis nula $H_0 : \mu = 10$.

Supongamos también que para un tamaño dado n de la muestra, se decide rechazar H_0 si se observa un valor de $\bar{X}$ que sea más grande que 12.

En estas condiciones, $\bar{X}$ es el estadístico de prueba, el valor $\bar{X} = 12$ se denomina **valor crítico**, y el conjunto de valores mayores que 12, es decir, el intervalo $(12, +\infty)$, es la **región de rechazo** o **región crítica** de la prueba. ■

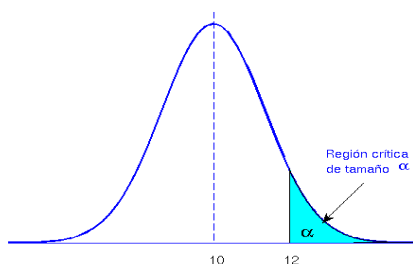


Figura 5.1: Función de densidad de $\bar{X}$ bajo H_0 supuesto n grande. $\alpha = P[\bar{X} > 12 | \mu = 10]$.

El procedimiento de prueba se construye con objetivo de decidir si se rechaza (o no) la hipótesis nula. Sin embargo, con la inclusión de la hipótesis alternativa, puede ser más descriptivo decir que probar una hipótesis estadística es proporcionar una decisión entre H_0 y H_1 .

En los contrastes de hipótesis, seguiremos la siguiente **Regla para formular Contrastes de Hipótesis**: definir la hipótesis nula H_0 en forma tal que ésta se rechace sólo si la evidencia muestral apoya esta decisión. En otras palabras, lo que realmente se desea es, mediante evidencias muestrales, rechazar H_0 sólo si hay evidencias de que la hipótesis alternativa H_1 es la correcta.

De esta forma, cuando se prueban hipótesis estadísticas, se juega un papel parecido al de un fiscal en su intento de proporcionar la suficiente evidencia para rechazar la hipótesis nula (todo el mundo es inocente mientras no se demuestre lo contrario).

En otras palabras, en las pruebas de hipótesis estadísticas el enfoque general es aceptar la premisa de que el error de tipo I es mucho más serio (grave, costoso, peligroso) que el error de tipo II, y formular las hipótesis nula y alternativa de acuerdo a dicho enfoque.

En la práctica lo que se suele hacer es fijar α y tratar que β sea lo más pequeño posible. Se fija α pequeño para que, en caso de que H_0 sea cierta, la probabilidad de rechazarla erróneamente sea pequeña.

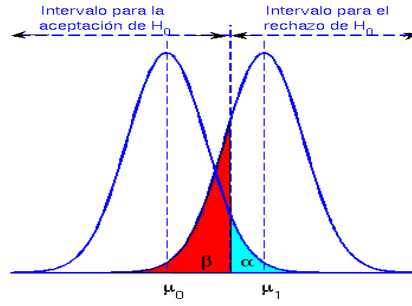


Figura 5.2: Interpretación de los errores tipo I y tipo II en un contraste unilateral.

Notación 5.1.2

- α se conoce como **nivel de significación de la prueba**.
- El procedimiento de prueba correspondiente se denomina **prueba de nivel α**

5.2 Prueba para μ con σ conocida en una población normal

5.2.1 En función de la región crítica

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una v.a. normal de media μ y varianza σ^2 ; entonces se cumple que $\bar{X} \sim N(\mu, \sigma^2/n)$ y que, si H_0 es cierta, $\mu = \mu_0 = \mu_{\bar{X}}$.

En estas condiciones, consideremos

1. el estadístico $\bar{X}$
2. el contraste de hipótesis $\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0 \end{cases}$
3. la $P(\text{error tipo I}) = \alpha$

Cuando H_0 es verdadera, $\mu_{\bar{X}} = \mu_0$, así que un valor de $\bar{x}$ próximo a μ_0 no es contradictorio. Por tanto, una región de rechazo debe tener la forma

$$(-\infty, \mu_0 - c) \cup (\mu_0 + c, +\infty) \quad \text{o bien} \quad \mathcal{R} \setminus (\mu_0 - c, \mu_0 + c)$$

esto es para rechazar la hipótesis nula tendrá que haber evidencias de que $\mu_{\bar{X}} \neq \mu_0$, o lo que es lo mismo tendrá que haber evidencias de que el estadístico es muy distinto de μ_0 . Obviamente el valor de corte c (la frontera de la región “distinto”) hay que determinarlo adecuadamente.

$$\begin{aligned}
\alpha &= P(\text{error tipo I}) &= \\
&= P(\text{se rechaza } H_0 \text{ cuando es cierta}) &= \\
&= P\left(\bar{X} \notin (\mu_0 - c, \mu_0 + c) \mid \bar{X} \sim N(\mu_0, \sigma^2/n)\right) &= \\
&= 1 - P\left(\mu_0 - c \leq \bar{X} \leq \mu_0 + c \mid \bar{X} \sim N(\mu_0, \sigma^2/n)\right) &= \\
&= 1 - P\left[\frac{\mu_0 - c - \mu_0}{\sigma/\sqrt{n}} \leq Z \leq \frac{\mu_0 + c - \mu_0}{\sigma/\sqrt{n}}\right] &= \\
&= 1 - P\left[\frac{-c}{\sigma/\sqrt{n}} \leq Z \leq \frac{c}{\sigma/\sqrt{n}}\right] &= \\
&= 2\left(1 - P\left[Z \leq \frac{c}{\sigma/\sqrt{n}}\right]\right), &=
\end{aligned}$$

de donde se obtiene que $1 - \frac{\alpha}{2} = P\left[Z \leq \frac{c}{\sigma/\sqrt{n}}\right],$

luego $z_{1-\alpha/2} = \frac{c}{\sigma/\sqrt{n}},$

por lo que $c = \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$

La región de rechazo es por tanto:

$$\mathcal{R} \setminus \left(\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \right)$$

o lo que es lo mismo

$$\left(-\infty, \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \right) \cup \left(\mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, +\infty \right)$$

5.2.2 En función del estadístico de contraste

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de una v.a. normal de media μ y varianza σ^2 ; entonces se cumple que $\bar{X} \sim N(\mu, \sigma^2/n)$ y por tanto se cumple que $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = Z \sim N(0, 1)$. En estas condiciones sabemos que $\mu_Z = 0$, por lo que en vez de realizar el contraste para ver si $\mu = \mu_0 = \mu_{\bar{X}}$ podemos hacerlo para ver si realmente la muestra desmiente que $\mu_Z = 0$, lo que significaría que $\mu_{\bar{X}} = \mu \neq \mu_0$.

En estas condiciones, consideremos

1. el estadístico $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ y
2. el contraste de hipótesis $\begin{cases} H'_0 : \mu_Z = 0 \\ H'_1 : \mu_Z \neq 0 \end{cases}$
3. la $P(\text{error tipo I}) = \alpha$

Sabemos que según el contraste de hipótesis en función de la región crítica la región de rechazo tiene la forma

$$\mathcal{R} \setminus \left(\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \right)$$

estandarizando obtenemos

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = Z \notin (-z_{1-\alpha/2}, z_{1-\alpha/2}) \iff Z \in (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, +\infty)$$

esto es para rechazar la hipótesis nula tendrá que haber evidencias de que $Z \neq 0$, o lo que es lo mismo tendrá que haber evidencias de que el estadístico es muy distinto de 0.

Resumen del procedimiento de prueba

Razonando de la misma forma para los contrastes unilaterales, se tiene el siguiente resumen del procedimiento:

Hipótesis nula: $H_0 : \mu = \mu_0$

Valor del estadístico de prueba: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Hip. alternativa	Est. contraste	R. de rechazo
$H_1 : \mu > \mu_0$	$z \in (z_{1-\alpha}, +\infty)$	$\bar{x} \in \left(\mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}, +\infty \right)$
$H_1 : \mu < \mu_0$	$z \in (-\infty, -z_{1-\alpha})$	$\bar{x} \in \left(-\infty, \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha} \right)$
$H_1 : \mu \neq \mu_0$	$z \in \mathcal{R} \setminus (-z_{1-\alpha/2}, z_{1-\alpha/2})$	$\bar{x} \in \mathcal{R} \setminus \left(\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \right)$

■ **Ejemplo 5.3** Un fabricante de sistemas rociadores utilizados para protección contra incendios en edificios de oficinas afirma que la temperatura promedio real de activación del sistema es 130°F. Una muestra de $n = 9$ sistemas, cuando se prueban, produce una temperatura promedio muestral de activación de 131.08°F. Si la distribución de los tiempos de activación es normal con desviación estándar de 1.5°F, ¿contradicen los datos la afirmación del fabricante al nivel de significación $\alpha = 0,01$?

Solución – mediante el estadístico –

- El estadístico de prueba es $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ y el valor del estadístico es $z = \frac{131,08 - 130}{0,5} = 2,16$
- El contraste es $\begin{cases} H'_0 : \mu = 0, \\ H'_1 : \mu \neq 0 \end{cases}$
- La región de rechazo es $\mathcal{R} \setminus (-z_{1-\alpha/2}, z_{1-\alpha/2})$ por lo que si $z \leq -z_{1-\alpha/2}$ ó $z \geq z_{1-\alpha/2}$ se tendrá que rechazar la hipótesis nula.

- Leyendo en las tablas de la normal vemos que $z_{1-\alpha/2} = z_{0,995} = 2,58$
- Como el valor calculado no cae en la región de rechazo (no cae lo suficientemente lejos de 0), no se puede rechazar H'_0 al nivel de significación $\alpha = 0,01$ y por tanto no se puede rechazar H_0 al nivel de significación $\alpha = 0,01$. Esto es, los datos no respaldan de manera significativa la afirmación de que el promedio real difiere del valor de diseño de 130.

Solución – mediante la region de rechazo –

- El estadístico de prueba es $\bar{x}$ y el valor del estadístico es $\bar{x} = 131,08$
- El contraste es $\begin{cases} H_0 : \mu = 130, \\ H_1 : \mu \neq 130 \end{cases}$
- $\bar{X} \in (-\infty, \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}) \cup (\mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}, +\infty)$ por lo que si $\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$ ó $\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$ se tendrá que rechazar la hipótesis nula.
- Utilizando las tablas vemos que $\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} = 128,71$ y que $\mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} = 131,29$
- Como $\bar{x} = 131,08$ no cae en la región de rechazo (no cae lo suficientemente lejos de 130), no se puede rechazar H_0 al nivel de significación $\alpha = 0,01$. Esto es, los datos no respaldan de manera significativa la afirmación de que el promedio real difiere del valor de diseño de 130.

■

5.3 Contraste de hipótesis mediante el p -valor

Una manera de informar el resultado de un análisis de prueba de hipótesis, es decir simplemente si se rechazó la hipótesis nula al nivel de significación especificado. Este tipo de enunciados no dice nada acerca de si el valor calculado del estadístico de prueba apenas cae en la región de rechazo, o si excede el valor crítico por una gran cantidad.

Hasta ahora, para realizar un contraste de hipótesis –tanto con el estadístico cómo con la región de rechazo– utilizábamos la probabilidad de error Tipo I permitida (α) para encontrar la región de rechazo del estadístico y determinar si rechazamos o no la hipótesis nula. El flujo de deducción era $\alpha \Rightarrow RR \Rightarrow estadístico$. Sin embargo también podemos utilizar el estadístico para encontrar cual sería la mínima probabilidad de error (p -valor) para la cual se rechazaría la hipótesis nula y podemos utilizar ese mínimo de modo que si α es mayor rechazaríamos H_0 . El flujo de deducción en este caso sería $estadístico \Rightarrow p\text{-valor} \Rightarrow \alpha$.

Definición 5.3.1 El **valor p** (o *nivel de significación observado*) es la probabilidad, dado que H_0 es cierta, de que el estadístico de prueba toma un valor mayor o igual que el calculado en base en la muestra aleatoria. Una vez determinado el valor p , la conclusión a algún nivel α particular resulta de comparar el valor p con α :

- (1) Si el valor $p \leq \alpha$ se rechaza H_0 al nivel α .
- (2) Si el valor $p > \alpha$ no se rechaza H_0 al nivel α .

■ **Ejemplo 5.4** En el ejemplo anterior, calcular el valor p .

Solución – mediante el p -valor –

- El estadístico de prueba es $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ y el valor del estadístico es $z = \frac{131,08 - 130}{0,5} = 2,16$
- Utilizando las tablas de la normal calculamos $p = P[|z| \geq 2,16]$:
 $p = P[|z| \geq 2,16] = P[z \leq -2,16] + P[z \geq 2,16] = 2P[z \geq 2,16] = 2(1 - P[z \leq 2,16]) = 0,0308$
- Como $p = 0,0308 > 0,01 = \alpha$ significa que el estadístico no cae en la región de rechazo y por tanto no se puede rechazar H_0 al nivel de significación $\alpha = 0,01$.

De hecho con el p -valor no sólo sabemos si se rechaza la hipótesis nula o no sino que sabemos cuál sería el máximo α con el que no rechazaríamos H_0 . En este ejemplo concreto, se observa que tendríamos que aumentar en más de 3 veces α (el riesgo de equivocarnos) para poder rechazar la hipótesis nula. ■

5.4 Ejercicios

Ejercicio 5.1 Se midió la temperatura de fusión de cada una de 16 muestras de cierta marca de aceite vegetal hidrogenado, y se encontró que $\bar{x} = 94,32$. Suponer que la distribución de la temperatura de fusión es normal con $\sigma = 1,20$.

- (a) Probar $H_0 : \mu = 95$ contra $H_1 : \mu \neq 95$ por medio de una prueba de dos extremos con nivel 0.01.
- (b) Si se emplea una prueba con nivel 0.01, ¿cuál es $\beta(94)$, la probabilidad de error de tipo II cuando $\mu = 94$?

Ejercicio 5.2 En una muestra de 26 observaciones en relación con el tiempo de evacuación (medido en segundos) para trabajadores petroleros en un ejercicio simulado, se ha obtenido una media muestral $\bar{x} = 370,69$ y una desviación estándar muestral $s = 24,36$. Supongamos que los investigadores tienen la creencia previa de que el tiempo de escape promedio sería a lo sumo seis minutos. ¿Los datos contradicen esta creencia previa? Suponiendo normalidad, probar la hipótesis apropiada usando un nivel de significación $\alpha = 0,05$

Ejercicio 5.3 El dolor de espalda baja (DEB) es un problema grave de salud en muchos entornos industriales. En cierto artículo se dan dos muestras sobre el avance lateral de movimiento (grados), una muestra de trabajadores sin antecedentes de DEB y otra muestra con antecedentes de este padecimiento.

Condición	Tamaño de la muestra	Media muestral	DE de la muestra
Sin DEB	28	91.5	5.5
Con DEB	31	88.3	7.8

Suponiendo normalidad, calcular un intervalo de confianza al 90 % para la diferencia entre el alcance medio poblacional del movimiento lateral para las dos condiciones. ¿El intervalo indica que el movimiento promedio lateral poblacional es distinto para las dos condiciones? ¿La conclusión es distinta si se usa un nivel de confianza del 95 %? ■

Ejercicio 5.4 Las mediciones de diferencias de potencial para una muestra de conectores cableados con aleación de aluminio y otra muestra cableada con aluminio EC. ¿Los siguientes datos indican que la caída de potencial promedio real para conexiones de aleación (tipo 1) es mayor que para conexiones con EC (tipo 2)? Suponer normalidad y usar un nivel de significación $\alpha = 0,01$.

Tipo	Tamaño de la muestra	Media muestral	DE de la muestra
1	20	17.5	0.55
2	20	16.9	0.49

Ejercicio 5.5 El codo de tenista se agrava por el impacto experimentado al golpear la pelota. En un estudio realizado, se describe la fuerza (N) sobre lamano justo después del impacto en un golpe de revés con una sola mano para seis jugadores expertos y ocho jugadores intermedios.

Tipo de jugador	Tamaño de la muestra	Media muestral	DE de la muestra
1. Avanzado	6	40.3	11.3
2. Intermedio	8	21.4	8.3

Suponiendo normalidad, calcular un intervalo de confianza al 95 % para la diferencia entre la fuerza promedio real para jugadores avanzados (μ_1) y la fuerza promedio real de los jugadores intermedios (μ_2). ¿Su intervalo es una prueba para concluir que las dos medias son distintas? ¿Se habría llegado a la misma conclusión si se hubiera calculado un intervalo de confianza para $\mu_2 - \mu_1$? ■

Ejercicio 5.6 De una población $N(\mu, 0'3)$ se obtiene una muestra de tamaño $n=100$ de media $\bar{X} = 1'70$. Se desea contrastar la hipótesis

$$\begin{cases} H_0 : \mu = 1'68 \\ H_1 : \mu \neq 1'68 \end{cases}$$

con un nivel de significación de 0.05.

Para este nivel de significación, calcular $\beta(171)$ y $\beta(169)$. ■

Ejercicio 5.7 Se desea contrastar la hipótesis

$$\begin{cases} H_0 : \mu = 150 \\ H_1 : \mu > 150 \end{cases}$$

Para ello se toma una muestra de la población $N(\mu, 225)$ de tamaño 36 y cuyo valor medio es $\bar{X} = 155$. Considerar:

- (a) $\alpha = 0.05$
- (b) $\alpha = 0.01$
- (c) Calcular $\beta(152)$ para ambos niveles de significación.

■

Ejercicio 5.8 Dada una población $N(\mu, \sigma^2)$ se toma una muestra de tamaño 10, de media $\bar{X} = 15$ y varianza $S^2 = 9$. Realizar, con nivel de significación del 5 %, el contraste de hipótesis

$$\begin{cases} H_0 : \sigma^2 = 15 \\ H_1 : \sigma^2 \neq 15 \end{cases}$$

■

Ejercicio 5.9 Se desea contrastar la hipótesis de la varianza de una población normal que es desconocida. Para ello se toma una muestra de tamaño 15 cuya desviación típica es $S = 5$. Se pide:

- (a) Contrastar $H_0 : \sigma^2 = 49$ frente a $H_1 : \sigma^2 < 49$
- (b) Contrastar $H_0 : \sigma^2 = 49$ frente a $H_1 : \sigma^2 > 49$
- (c) Contrastar $H_0 : \sigma^2 = 49$ frente a $H_1 : \sigma^2 \neq 49$

Considerar un nivel de significación del 5 % en todos los casos.

■

Ejercicio 5.10 Un laboratorio afirma que un calmante de su fabricación quita la jaqueca en 14 minutos en los casos corrientes. Con el fin de comprobar estadísticamente esta afirmación, se eligen al azar 18 pacientes con jaqueca y se toma como variable de respuesta en el experimento el tiempo transcurrido entre la administración del calmante y el momento en que desaparece la jaqueca. Los resultados en la muestra de pacientes han sido $\bar{X} = 19$ y $S = 7$. ¿Se puede rechazar, al 5 % de significación, la afirmación del laboratorio? (Suponer normalidad)

■

Ejercicio 5.11 De los datos experimentales se deduce que la duración de una pila sigue una distribución normal. El fabricante asegura que el tiempo medio de vida es de 15 horas de duración. Temiendo que la duración pueda ser menor se toma una muestra de 10 pilas, obteniendo los siguientes datos (duración en horas):

14 14.5 15 15.5 13 16.5 10 15.5 14 14.

Decidir si, en base a los datos muestrales, debemos rechazar o no la afirmación del comerciante con un nivel de significación de 0.05.

■

Ejercicio 5.12 El índice de dureza Rockwell para el acero se determina al rayar el acero con un punzón de diamante y medir la profundidad del surco. En 50 muestras de cierto tipo de acero, el índice de dureza tuvo una media de 62 y desviación 8. El productor afirma que este acero tiene un índice de dureza promedio de, por lo menos, 64. ¿Hay suficiente evidencia para refutar la información del productor con un nivel de significación del 1 %? (Suponer normalidad). ■

Ejercicio 5.13 Distintas mediciones de la cantidad de cierto producto obtenidas para dos tipos de suelo dieron los siguientes resultados:

Suelo tipo A	Suelo tipo B
$n_A = 30$	$n_B = 35$
$\bar{Y}_A = 1'65$	$\bar{Y}_B = 1'43$
$S_A = 0'26$	$S_B = 0'22$

¿Difieren los dos suelos con respecto a la cantidad promedio del producto? (Suponer normalidad y un nivel de significación del 1 %). ■

Ejercicio 5.14 Una compañía farmacéutica realizó un experimento para comprobar los tiempos promedios (en días) necesarios para recuperarse del resfriado común. En este experimento se compararon personas que tomaron una dosis diaria de vitamina C, con personas que tomaron un placebo. El estudio se hizo a 35 personas en cada grupo, obteniéndose los siguientes datos:

Tratamiento	Con placebo	Con vitamina C
Media	7.3	4.6
Desviación	2.9	1.2

Supuesto normalidad, y con un nivel de significación de 1 %, ¿se puede afirmar que el tiempo de recuperación promedio cuando se toma placebo supera en más de un día al tiempo de recuperación promedio cuando se ingiere vitamina C? ■

Ejercicio 5.15 Una fábrica compra un ingrediente a dos proveedores diferentes. El nivel promedio de impurezas en la materia prima es aproximadamente el mismo en ambos proveedores, pero interesa la baja variabilidad de impurezas de un envío a otro. Para comparar la variación se seleccionan diez envíos de cada uno de los proveedores midiendo el porcentaje de impurezas en cada cargamento. Las medias y las varianzas muestrales se presentan en la tabla siguiente:

Proveedor A	Proveedor B
$n_A = 10$	$n_B = 10$
$\bar{Y}_A = 1'89$	$\bar{Y}_B = 1'85$
$S_A = 0'273$	$S_B = 0'094$

¿Presentan los datos suficiente evidencia para afirmar una diferencia en la variabilidad de los niveles de impurezas de los envíos para los dos proveedores? (Suponer norma-

lidad y utilizar $\alpha = 0'01$). Basándose en los resultados de la prueba, ¿cuál sería su recomendación para el productor? ■

Ejercicio 5.16 Los siguientes datos fueron recabados para verificar si existe diferencia sistemática en los pesos obtenidos con dos balanzas diferentes:

Peso en gramos	Balanza I	Balanza II
Muestra de roca 1	11.23	11.27
Muestra de roca 2	14.36	14.41
Muestra de roca 3	8.33	8.35
Muestra de roca 4	10.50	10.52
Muestra de roca 5	23.42	23.41
Muestra de roca 6	9.15	9.17
Muestra de roca 7	13.47	13.52
Muestra de roca 8	6.47	6.46
Muestra de roca 9	12.40	12.45
Muestra de roca 10	19.38	19.35

Suponiendo normalidad probar, con nivel de significación $\alpha = 0'05$, si la diferencia de medias de los pesos obtenidos es significativa. ■

Ejercicio 5.17 Mientras realizan su trabajo, el ritmo cardíaco de 25 pilotos se incrementó en un promedio de 18.4 pulsaciones por minuto con una desviación típica de 4.9 pulsaciones por minuto. Suponiendo normalidad, contrastar si el ritmo cardíaco de los pilotos sufre una desviación estándar de 6 frente a menor de 6. (Suponer normalidad y utilizar $\alpha = 0'05$). ■

Ejercicio 5.18 Se obtienen los siguientes datos con respecto a la resistencia de dos dispositivos diferentes:

Dispositivo A: 15.8, 12.7, 13.2, 16.9, 10.6, 18.8, 11.1, 14.3, 17.0, 12.5

Dispositivo B: 24.9, 23.6, 19.8, 22.1, 20.4, 21.6, 21.8, 22.5

Probar si es razonable suponer que las muestras provienen de poblaciones (normales) con la misma varianza. (Utilizar $\alpha = 0'05$) ■

Ejercicio 5.19 Se requiere determinar si existe menos variabilidad en el plateado realizado por la compañía A que en el efectuado por la compañía B. Si dos muestras aleatorias independientes de tamaño 12 del trabajo desempeñado por ambas compañías presentaron desviaciones típicas $S_A = 0'035 \text{ mm.}$ y $S_B = 0'062 \text{ mm.}$, comprobar la hipótesis nula de que $\sigma_A = \sigma_B$ contra la hipótesis alternativa de que $\sigma_A < \sigma_B$ con un nivel de significación de 0'05. (Suponer normalidad). ■

5.4.1 Ejercicios de exámenes de años anteriores

Ejercicio 5.20 Los pesos de los rodamientos fabricados en un proceso siguen una distribución normal con media 250 g. y desviación típica 5 g. Tras reajustar el mismo, el encargado sospecha que el peso promedio ha aumentado, pero su desviación típica no ha cambiado. Se selecciona una muestra aleatoria simple de 16 rodamientos, con un peso medio de 251.9 g.

- ¿Tiene razón el encargado? Lleva a cabo el contraste para un nivel de significación del 5%.
- Supuesto que la media verdadera fuese $\mu = 256$, calcular el error de tipo II asociado, es decir, calcular $\beta(256)$.

Ejercicio 5.21 Una línea de producción funciona con un peso de llenado de 16 gr por envase. De acuerdo con datos anteriores, se sabe que el peso sigue una distribución normal. El exceso o defecto de peso en el llenado son problemas graves, y la línea de producción deber parar si se presenta alguno de ellos. Un inspector de calidad toma una muestra de 30 artículos y de acuerdo con los resultados toma la decisión de parar la línea o dejarla trabajando. Utilizando un nivel de significación $\alpha = 0,01$, si se encuentra que $\bar{x} = 15,82$ gr y $s_X = 0,8$ gr., ¿qué acción recomendarías?

- Razonando mediante región crítica.
- Razonando mediante el p -valor.

Ejercicio 5.22 Una fábrica requiere ácido sulfúrico como materia prima. La calidad de la producción depende en gran parte de la pureza de esta sustancia, la cual es provista por dos proveedores diferentes A y B. Se toman muestras de ambos proveedores obteniéndose los siguientes datos de pureza:

	tamaño muestral	media	varianza
proveedor A	9	0.987	0.02
proveedor B	16	0.882	0.0135

Supuesto que las variables bajo estudio siguen distribuciones normales independientes,

- Contrastar la igualdad de las varianzas poblacionales. (Tómese $\alpha = 0,05$)
- ¿Hay evidencia para afirmar, al 5% de significación, que el ácido sulfúrico procedente del proveedor A tiene, en promedio, una mayor pureza que el procedente de B?

Ejercicio 5.23 En la fabricación de chip para circuitos integrados hay una variable denominada *amplitud de ventana*, que está relacionada con los procedimientos de interconexión entre los circuitos. Se desea estudiar el efecto que tiene sobre la amplitud de ventana una determinada reacción química que se produce durante la fabricación de los chips. Para ello se toman muestras antes y después de la reacción obteniéndose los siguientes resultados (en milimicras):

	tamaño muestral	media	desviación típica
antes de la reacción	10	2.385	0.380
después de la reacción	8	3.376	0.603

Supuesto que las variables bajo estudio siguen distribuciones normales independientes, ¿se puede afirmar, al 5 % de significación, que después de la reacción aumenta la amplitud media de ventana? ■

Ejercicio 5.24 Dos tipos de soluciones químicas *A* y *B* fueron ensayadas para estudiar su *pH*. El análisis de 6 soluciones de tipo *A* dieron un *pH* promedio de 7.59 con una desviación típica de 0.032. El análisis de 5 soluciones de tipo *B* dieron un *pH* promedio de 7.52 con una desviación típica de 0.024. Supuesto que las variables bajo estudio siguen distribuciones normales independientes, ¿se puede afirmar al 5 % de significación que el *pH* de la solución de tipo *A* supera, en promedio, al *pH* de la solución tipo *B*? ■

Ejercicio 5.25 Una empresa fabrica discos duros cuyos tiempos de acceso se ajustan a una distribución normal de media 8 ms. El fabricante está diseñando un nuevo modelo de disco que mejore el tiempo medio de acceso. Realiza 80 pruebas con los nuevos prototipos obteniendo un tiempo medio de acceso de 5.805 ms. y una desviación típica de 0.98 ms.

- ¿Se puede afirmar, al 1 % de significación, que el nuevo disco responde a las expectativas del fabricante? Razónese mediante regiones críticas.
- Razonar el resultado anterior mediante el *p*-valor.

Ejercicio 5.26 Se determinaron los niveles de colesterol de un grupo de 25 personas de una ciudad europea y de un grupo de 30 personas de una ciudad japonesa de características similares. Los resultados fueron:

	Media	Desv. Típica
Ciudad europea	175	11.5
Ciudad japonesa	179	9.3

- ¿Es menos variable el nivel de colesterol en los japoneses?
- ¿Tienen, por término medio, los japoneses el colesterol más alto?

Ejercicio 5.27 Se quiere medir el nivel de satisfacción de graduados de dos universidades españolas. El cuestionario se administra aleatoriamente a 11 estudiantes de la Universidad *X* y a 15 estudiantes de la Universidad *Y*. El resumen resultados es:

Universidad	Media	Varianza
<i>X</i> :	83.4545	70.6727
<i>Y</i> :	76	89.2857

¿Podemos concluir a la vista de los resultados que el grado de satisfacción medio de los alumnos de la Universidad X excede en más de 4 unidades al nivel de satisfacción medio de los alumnos de la Universidad Y . (Suponer normalidad y utilizar un nivel de significación del 5 %.) ■

Ejercicio 5.28 Se sabe que la duración del sueño en cierta población se distribuye normalmente con media igual a 7 horas. Intentamos ver si cierta dosis de un sedante S alarga la duración del sueño. Para ello elegimos una muestra aleatoria de 25 personas y les aplicamos dicha dosis. El resultado es que la duración media del sueño de estas 25 personas ha sido de 8 horas con una desviación típica de 4.84 horas. ¿Puede decirse con un nivel de confianza del 5 % que tuvo efectos positivos la utilización del sedante? ■

Ejercicio 5.29 Se lleva a cabo un experimento para comparar el desgaste por abrasivo de dos diferentes materiales laminados. Se prueban 12 piezas del material 1 mediante la exposición de cada pieza a una máquina para medir el desgaste. Diez piezas del material 2 se prueban de manera similar. En cada caso, se mide la profundidad del desgaste. Las muestras del material 1 dan como desgaste promedio (codificado) de 85 unidades con una desviación estándar de 4, mientras que las muestras del material 2 dan un promedio de 81 y una desviación estándar de 5. ¿Podemos concluir con un nivel de significación del 5 % que el desgaste abrasivo del material 1 excede el del material 2 en más de dos unidades? Suponer que las poblaciones son normales. ■

Ejercicio 5.30 Un estudio revela que en los hogares que tienen conexión a internet, el tiempo de conexión siguen una distribución normal 15 horas semanales, en término medio. Una operadora de servicios de internet sospecha que el tiempo de conexión es más alto. Para ello realiza una encuesta a una muestra aleatoria de 60 hogares obteniendo una media de 17.34 horas de conexión a internet semanales con una desviación típica de 7.54 horas.

¿Presentan estos datos evidencia suficiente para apoyar la sospecha de la operadora? (utilizar un nivel de significación del 10 %).

- (a) Razonar mediante la región crítica.
- (b) Razonar mediante el p -valor.

Ejercicio 5.31 Las puntuaciones en un test que mide la creatividad siguen, en la población general de adolescentes, una distribución Normal de media 11.5. En un centro escolar que ha implantado un programa de estimulación de la creatividad una muestra de 30 alumnos ha proporcionado las siguientes puntuaciones:

11	9	12	17	8	11	9	4	5	9	14	9	17	24	19
10	17	17	8	23	8	6	14	16	6	7	15	20	14	15

A un nivel de confianza del 95 % ¿puede afirmarse que el programa es efectivo?

- (a) Razonado mediante la región crítica.
- (b) Razonando mediante el p -valor.

Ejercicio 5.32 La concentración media de dióxido de carbono en el aire en una determinada zona no es habitualmente mayor de 355 p.p.m.v. (partes por millón de volumen). Se sospecha que esta concentración es más alta en la capa de aire más próxima a la superficie. Para contrastar esta hipótesis se analiza aire en 50 puntos elegidos aleatoriamente a una misma altura del suelo. Los datos recogidos tienen una media muestral de 550 p.p.m.v. y una desviación típica de 180.

Suponiendo que las mediciones siguen una distribución normal, ¿podemos afirmar a un nivel de significación del 1 %, que los datos proporcionan suficiente evidencia estadística de que la concentración es mayor cerca del suelo?

- (a) Razonar mediante la región crítica.
- (b) Razonar mediante el p -valor.

Ejercicio 5.33 Como parte de un estudio de calidad del servicio, se recogen los tiempos de acceso (en décimas de segundo) a dos servidores *ftp*, obteniendo los siguientes resultados:

Servidor A	72, 77, 84, 78, 69, 70, 77, 73, 65, 75, 72
Servidor B	71, 75, 65, 69, 73, 66, 68, 71, 74

Supuesto que los tiempos de acceso siguen distribuciones normales independientes, ¿se puede afirmar a un nivel de significación del 5 % que el tiempo medio de acceso al servidor A supera en más de 2 décimas de segundo al tiempo de acceso medio al servidor B?

Ejercicio 5.34 En un estudio sobre el porcentaje de disco duro ocupado por los videojuegos en los ordenadores, se han recogido los siguientes datos muestrales en dos poblaciones A y B:

	Tamaño muestral	media	varianza
muestra de A	25	44	64
muestra de B	19	32	46

Supuesto que las características bajo estudio siguen distribuciones normales independientes, ¿se puede afirmar, a un nivel de significación del 5 %, que el porcentaje promedio ocupado por los videojuegos en los ordenadores de la población A supera en más de 6 unidades al ocupado en los ordenadores de la población B?

Ejercicio 5.35 La velocidad X de un procesador, en GHz, sigue una distribución $N(\mu, \sigma^2)$. Se han tomado 40 medidas de la velocidad del procesador observándose un valor medio de 1.4 GHz y desviación típica 0.17 GHz.

- Se construye un intervalo de confianza para la varianza poblacional resultando el intervalo $(L_1, 0.0539528)$. Calcular razonadamente L_1 .
- A partir de los datos muestrales, ¿tenemos evidencia para afirmar, al 1 % de significación, que la velocidad media del procesador es superior a 1.3 GHz? (Plantear y resolver un contraste de hipótesis adecuado).

■

Ejercicio 5.36 Con el fin de estudiar la cantidad de radiación, debida a la presencia de I-129, en una fuente destinada para el consumo de la población, la Agencia para la Protección Ambiental ha tomado una muestra de 16 mediciones que arrojan una media de 0.31 Bq/l con una desviación típica de 0.06 B1/l. Se supone que la cantidad de radiación sigue una distribución normal.

- Si la cantidad máxima permitida es, en promedio, de 0.27 Bq/l, ¿hay evidencia para concluir a partir de los datos muestrales que existe un exceso de radiación en el agua? (Utilizar un nivel de significación del 5 %)
- Calcular un intervalo de confianza al 90 % para la varianza de la cantidad de radiación.

■

Ejercicio 5.37 Un fabricante de automóviles debe elegir entre un determinado tipo de piezas de acero suministradas por un proveedor A y otras suministradas por otro proveedor B . Para proceder a la elección se ha analizado la resistencia a la tracción de las piezas suministradas por ambos proveedores, tomando una muestra de 10 piezas de A y otra muestra de 12 piezas de B . La resistencia media de la muestra A es de 54000 unidades y de la muestra B es de 49000 unidades, siendo las desviaciones típicas muestrales $S_A = 2100$ y $S_B = 1900$. Las resistencias de las piezas de ambos proveedores se distribuyen normalmente. Las piezas del proveedor B son más baratas que las del proveedor A , por lo que estas últimas sólo son rentables si tienen una resistencia media superior a la de B en más de 2000 unidades y la misma variabilidad. ¿A qué proveedor habría que comprar las piezas en vista de los resultados muestrales? (Tómese $\alpha = 0,05$).

■

Ejercicio 5.38 Los resultados que se muestran a continuación se refieren al pH de muestras de lluvia tomadas en dos regiones industrializadas:

	tamaño muestral	media	varianza
Región I	6	3.72	1.29
Región II	5	4.16	1.20

Supuesto que las características bajo estudio siguen distribuciones normales independientes, ¿hay evidencia para afirmar, al 5 % de significación, que en la región II el nivel

medio de pH de la lluvia es superior al de la región I? ■

Ejercicio 5.39 Se instala un nuevo dispositivo de filtrado en una unidad química. Antes de su instalación una muestra aleatoria proporcionó la siguiente información acerca del porcentaje de impurezas:

tamaño muestral	media	varianza
9	13.5	34.73

Después de la instalación se tomó otra muestra obteniéndose:

tamaño muestral	media	varianza
8	9.4	28.84

Supuesto que las características bajo estudio siguen distribuciones normales independientes, ¿hay evidencia para afirmar, al 10% de significación, que el nuevo dispositivo de filtrado reduce el porcentaje medio de impurezas? ■

Ejercicio 5.40 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	Varianza
Proveedor A	5459	33703
Proveedor B	5162	39928

Supuesto que ambas poblaciones son normales e independientes ¿a qué proveedor debemos comprar los cartuchos de tóner? (Tomar $\alpha = 0,05$) ■



6. Modelos de Regresión

Uno de los aspectos más relevantes de la Estadística es el análisis de las relaciones entre variables. En general es interesante conocer el efecto que una o varias variables pueden causar sobre otra, e incluso predecir en mayor o menor grado valores en una variable a partir de otra.

El modelo de regresión es un procedimiento estadístico que permite estimar la relación lineal que relaciona dos o más variables entre sí pudiendo utilizarlo para explicar y predecir los valores de una de las variables a partir de valores observados en las demás (ya sean una o varias). Esta relación resumiría como cambia una de ellas en función del cambio de las restantes. El modelo permitirá además obtener conclusiones sobre la significación estadística del mismo, esto es, si la relación entre las variables puede ser debida exclusivamente al azar o por el contrario, es funcional.

6.1 Modelo de regresión lineal simple

En esta sección se estudiará el modelo de regresión de dos variables; en él, una de las variables asumirá el estatus de variable independiente y la otra de variable dependiente. La primera, la variable independiente, se considerará a partir de ese momento cómo la generadora de los cambios en la variable dependiente.

En principio no se sabe si las variables en cuestión están relacionadas o no, o si en caso de haber dependencia es significativa o no. De haber entre ellas una dependencia lineal significativa, podríamos expresar la variable dependiente Y a partir de la variable independiente X mediante una recta.

La relación matemática más simple entre dos variables x e y es una relación lineal

$$y = \beta_0 + \beta_1 x;$$

esto es, el conjunto de pares (x, y) para los cuales $y = \beta_0 + \beta_1 x$ determina una recta.

En cualquier caso, si las dos variables no se relacionan de modo determinístico, entonces para un valor fijo de x , el valor de la segunda variable es aleatorio.

La variable X cuyo valor x lo fija el experimentador se llama **variable independiente, predictora o explicativa**. Una vez fijada la x , la segunda variable será aleatoria Y y su valor observado y . Esta variable Y se denomina **variable dependiente o de respuesta**.

En la práctica se dispondrá de un conjunto de pares de observaciones $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ experimentales. Un primer paso en el análisis de regresión en el que intervienen dos variables es construir una **gráfica de dispersión** de los datos, que consiste en representar cada (x_j, y_j) con $j = 1, \dots, n$ como un punto en un sistema de coordenadas.

Ejercicio 6.1 En un determinado artículo se ilustra un diagrama de dispersión, junto con la recta de mínimos cuadrados, de $x =$ volumen de lluvia (m^3) e $y =$ volumen de escurrimiento (m^3) para determinado lugar. Los valores siguientes se tomaron de la gráfica:

x	5	12	14	17	23	30	40	47	55	67	72	81	96	112	127
y	4	10	13	15	15	25	27	46	38	46	53	70	82	99	100

- Dibujar el diagrama de dispersión.

6.1.1 Modelo probabilístico lineal

Dado el modelo $y = \beta_0 + \beta_1 x$, el modelo probabilístico da por hecho que *el valor esperado de Y es una función lineal de x* , pero, la variable Y difiere de su valor esperado por una cantidad aleatoria.

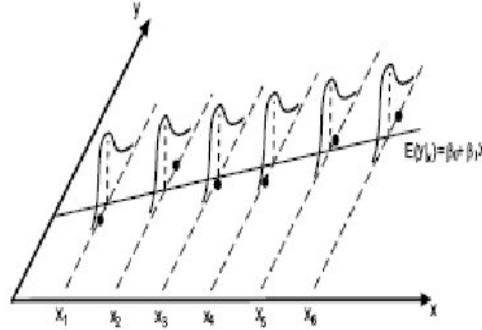
Modelo de regresión lineal simple

Hay parámetros β_0 , β_1 y σ^2 tal que para cualquier valor fijo de la variable independiente x , la variable dependiente se relaciona con x por medio de la **ecuación modelo**

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (6.1)$$

La cantidad ε es una variable aleatoria, que se supone tiene una distribución normal con $E[\varepsilon] = 0$ y $Var[\varepsilon] = \sigma^2$; esto es, $\varepsilon \sim N(0, \sigma^2)$.

- ε se conoce como **desviación aleatoria o término de error aleatorio**.
- La recta $y = \beta_0 + \beta_1 x$ se llama **recta de regresión verdadera (o poblacional)**.
- **Interpretación del modelo:** para un valor fijo x^* de X ,
 $E[Y|_{x^*}] = E[\beta_0 + \beta_1 x^* + \varepsilon] = \beta_0 + \beta_1 x^* + E[\varepsilon] = \beta_0 + \beta_1 x^*$,
 $Var[Y|_{x^*}] = Var[\beta_0 + \beta_1 x^* + \varepsilon] = Var[\beta_0 + \beta_1 x^*] + Var[\varepsilon] = 0 + \sigma^2$.



6.2 Estimación de los parámetros del modelo

En la práctica se dispone de un conjunto con n pares de observaciones experimentales

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

y el objetivo será encontrar estimaciones $\hat{\beta}_0$ y $\hat{\beta}_1$ de β_0 y β_1 , así como, en segundo término, encontrar una estimación $\hat{\sigma}^2$ (ó s^2) de σ^2 para conocer la naturaleza del error. De este modo, podremos escribir el modelo como $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $i = 1, 2, \dots, n$

Nota 6.2.1 Durante todo el tema supondremos que los x_i son valores no aleatorios (observados o medidos con un error despreciable) mientras que los ε_i son v.a. independientes, con media 0, varianza constante σ^2 .

6.2.1 Estimación de β_0 y β_1

Nuestro objetivo será buscar una recta de la forma $y = b_0 + b_1 x$ que estime a la recta verdadera $y = \beta_0 + \beta_1 x$ y proporcione un buen ajuste a los puntos de datos observados. Para ello vamos a buscar un ajuste por el *principio de los mínimos cuadrados*, mediante el cual una recta proporciona un buen ajuste para los datos si las distancias verticales (desviaciones) de los puntos observados a la recta son pequeños.

La desviación vertical del punto (x_i, y_i) de la recta $y = b_0 + b_1 x$ es:

$$y_i - (b_0 + b_1 x_i)$$

Entonces la suma de desviaciones verticales al cuadrado de los puntos $(x_1, y_1), \dots, (x_n, y_n)$ a la recta es

$$f(b_0, b_1) = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2 \quad (6.2)$$

Las estimaciones puntuales de β_0 y β_1 , denotadas por $\hat{\beta}_0$ y $\hat{\beta}_1$ y conocidas como **estimaciones de mínimos cuadrados**, son los valores que minimizan $f(b_0, b_1)$ y por tanto la **recta de regresión estimada** o **recta de mínimos cuadrados** es la recta cuya ecuación es $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

Derivando e igualando a cero la Ecuación 6.2 obtenemos que

$$b_1 = \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (6.3)$$

y que

$$b_0 = \hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \cdot \bar{x} \quad (6.4)$$

donde

$$S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i) / n \quad (6.5)$$

y

$$S_{xx} = \sum x_i x_i - (\sum x_i)(\sum x_i) / n = \sum x_i^2 - (\sum x_i)^2 / n \quad (6.6)$$

En consecuencia, la recta estimada es

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}}(x - \bar{x}) \quad (6.7)$$

Proposición 6.2.2 $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$

Ejercicio 6.2 — Continuación del Ejercicio 6.1. Con los datos del ejercicio anterior,

- calcular una estimación de la recta de regresión poblacional. Dibujar la recta de regresión junto con el diagrama de dispersión.
- calcular una estimación puntual del volumen de escurrimiento promedio real cuando el volumen de lluvia es 50.

6.2.2 Estimación de σ^2 y σ

Definición 6.2.1 Los **valores ajustados** $\hat{y}_1, \dots, \hat{y}_n$ se obtienen al sustituir en forma sucesiva $x_1, \dots, x_n$ en la ecuación de la recta de regresión estimada: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \dots, n$ y los **residuos** $e_i = y_i - \hat{y}_i$, $i = 1, \dots, n$ son las desviaciones verticales respecto a la recta estimada.

Definición 6.2.2 La **suma de cuadrados del error** (o suma de cuadrados del residuo) denotada por SSE , es

$$SSE = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right]^2$$

y la estimación de σ^2 es

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} = \frac{SSE}{n - 2} \quad (6.8)$$

El divisor $n - 2$ es el número de grados de libertad relacionados con la estimación (esto es debido a que al tener que estimar previamente los dos parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$ da como resultado una pérdida de 2 grados de libertad).

Ejercicio 6.3 — Continuación del Ejercicio 6.1. Con los datos anteriores,

- calcular una estimación puntual de la desviación estándar σ .

Al sustituir cada y_i por Y_i se obtiene el estimador S^2 , que es un estimador insesgado para σ^2

Un modo más sencillo de calcular SSE es según la siguiente

Proposición 6.2.3

$$SSE = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

6.3 Coeficiente de Determinación

La suma de cuadrados del error SSE puede interpretarse como una medida de la variación en Y que el modelo deja sin explicar; es decir, cuánto no se puede atribuir a una relación lineal.

Una medida cuantitativa de la cantidad total de variación en los valores observados de Y la proporciona la **suma total de cuadrados** $SST = \sum (y_i - \bar{y})^2$.

y una forma más sencilla de calcular SST es según la siguiente

Proposición 6.3.1

$$SST = S_{yy} \quad \left(= \sum y_i^2 - (\sum y_i)^2 / n \right)$$

Definición 6.3.1 — 514. El **coeficiente de determinación**, representado por R^2 , de determina mediante:

$$R^2 = 1 - \frac{SSE}{SST}$$

Se interpreta como la proporción de variación de y observada que es explicada por el modelo de regresión lineal simple.

Obviamente $0 < R^2 < 1$. Cuanto mayor sea R^2 , preciso es el modelo de regresión lineal simple en explicar la variación de y .

Si introducimos la **suma de cuadrados de regresión**, SSR , como $SSR = SST - SSE$, se verifica

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SST - SSE}{SST} = \frac{SSR}{SST}$$

Ejercicio 6.4 — Continuación del Ejercicio 6.1. Con los datos anteriores,

- ¿qué proporción de la variación observada en el volumen de escurrimiento se puede atribuir a la relación de regresión lineal simple entre escurrimiento y lluvia?

6.3.1 Coeficiente de variación

El coeficiente de variación, CV , es un criterio razonable para estudiar la calidad del ajuste y medir la dispersión de los datos alrededor de la recta de regresión y se define como:

$$CV = \frac{S}{|\bar{Y}|} \cdot 100\%$$

Su uso viene motivado por el hecho de que la estimación de la desviación estandar del error, S , no es generalmente satisfactoria para medir la calidad del ajuste al no estar libre de escala.

6.4 Inferencias sobre el parámetro de la pendiente β_1

β_1 (pendiente de la recta de regresión poblacional) es el cambio promedio real en la variable y relacionada con un incremento unitario en la variable x . $\hat{\beta}_1$, es una estimación puntual de β_1 . Los estimadores para β_0 y β_1 se obtienen al reemplazar y_i por Y_i en sus fórmulas correspondientes:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \frac{\sum Y_i - \hat{\beta}_1 \sum x_i}{n}$$

El estimador para σ^2 es

$$\hat{\sigma}^2 = S^2 = \frac{\sum Y_i^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum x_i Y_i}{n - 2}$$

$\hat{\beta}_1$ es una función lineal de las variables independientes $Y_1, \dots, Y_n$.

1. $E[\hat{\beta}_1] = \beta_1$.
2. $Var(\hat{\beta}_1) = \sigma_{\hat{\beta}_1}^2 = \frac{\sigma^2}{S_{xx}}$. Si reemplazamos σ por su estimación, s , se obtiene una estimación para $\sigma_{\hat{\beta}_1}$, $s_{\hat{\beta}_1} = \hat{\sigma}_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}}$.
3. $\hat{\beta}_1$ tiene una distribución normal.

6.4.1 Intervalo de confianza para β_1

Teorema 6.4.1 Las suposiciones del modelo de regresión lineal simple implican que la variable estandarizada

$$T = \frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}$$

tiene una distribución t con $n - 2$ grados de libertad.

Teorema 6.4.2 Un intervalo de confianza al $100(1 - \alpha) \%$ para la **pendiente** β_1 de la regresión lineal verdadera es

$$\hat{\beta}_1 \pm t_{1-\alpha/2, n-2} \cdot s_{\hat{\beta}_1}$$

Ejercicio 6.5 — Continuación del Ejercicio 6.1. Continuando con el ejercicio anterior,

- calcular un intervalo de confianza para β_1 al 95 %.

6.4.2 Procedimiento de contraste de hipótesis

La hipótesis nula en una prueba respecto de β_1 será un enunciado de igualdad ($\beta_1 = \beta_{1_0}$). El estadístico de prueba resulta de reemplazar β_1 en la variable T estandarizado por un valor nulo β_{1_0} .

Nota 6.4.3 — Estandarizar el estimador de β_1 supuesto que H_0 es verdadera.

Hipótesis nula: $H_0 : \beta_1 = \beta_{1_0}$

Estadístico: $t = \frac{\hat{\beta}_1 - \beta_{1_0}}{s_{\hat{\beta}_1}}$

Hip. alternat. Reg. crítica para prueba nivel α

$H_1 : \beta_1 > \beta_{1_0} \quad t \geq t_{1-\alpha, n-2}$

$H_1 : \beta_1 < \beta_{1_0} \quad t \leq -t_{1-\alpha, n-2}$

$H_1 : \beta_1 \neq \beta_{1_0} \quad \{t \leq -t_{1-\alpha/2, n-2}\} \cup \{t \geq t_{1-\alpha/2, n-2}\}$

Definición 6.4.1 Una **prueba de utilidad del modelo** es una prueba de $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$ en cuyo caso el valor del estadístico de prueba es $t = \hat{\beta}_1 / s_{\hat{\beta}_1}$

Ejercicio 6.6 — Continuación del Ejercicio 6.1. Continuando con el ejercicio anterior:

- ¿podemos afirmar al 95 % que $\beta_1 = 0,85$?

6.5 Inferencia relacionada con $E[Y|x^*]$ y predicción de valores futuros

Intervalo de confianza para el **valor esperado** de Y cuando $x = x^*$ con nivel de confianza $100(1 - \alpha) \%$

$$\hat{y} \pm s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} \cdot t_{1-\alpha/2, n-2} \quad \text{donde} \quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

Intervalo de confianza para la predicción de un **valor futuro** de Y cuando $x = x^*$ con nivel de confianza $100(1 - \alpha) \%$

$$\hat{y} \pm s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} \cdot t_{1-\alpha/2, n-2} \quad \text{donde} \quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

Ejercicio 6.7 — Continuación del Ejercicio 6.1. Siguiendo con los datos anteriores,

- Obtener un intervalo de confianza para el valor medio de y cuando $x^* = 60$.
- Obtener un intervalo de confianza para el valor futuro de y cuando $x^* = 150$.

6.6 Correlación

Definición 6.6.1 El **coeficiente de correlación muestral** para los n pares $(x_1, y_1), \dots, (x_n, y_n)$ es $r = \frac{S_{xy}}{\sqrt{S_{xx}} \cdot \sqrt{S_{yy}}}$

Propiedades de r

1. El valor de r no depende de cuál de las dos variables se marque como dependiente y cuál como independiente.
2. El valor de r es independiente de las unidades en que se midan las variables.
3. $-1 \leq r \leq 1$
4. Si $r = 1$ los pares están sobre la bisectriz del primer al tercer cuadrante. Si $r = -1$, dichos pares están sobre la bisectriz del segundo al cuarto cuadrante.
5. $r^2 = R^2$ (coeficiente de determinación)

Ejercicio 6.8 — Continuación del Ejercicio 6.1. Para finalizar el ejercicio, calcular el coeficiente de correlación. ■

6.7 El modelo de regresión lineal múltiple.

Supongamos que tenemos que considerar más de una variable regresora. El modelo desarrollado anteriormente se puede generalizar a este caso.

Consideremos un experimento en el que los datos generados son del tipo

y	x_1	x_2	$\dots$	x_k
y_1	x_{11}	x_{21}	$\dots$	x_{k1}
y_2	x_{12}	x_{22}	$\dots$	x_{k2}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{1n}	x_{2n}	$\dots$	x_{kn}

donde la variable y es la “respuesta” observada y $x_1, \dots, x_k$ representan magnitudes que se pueden medir de forma precisa.

En la fila i -ésima aparecen representados los datos observados en el experimento i -ésimo para la variable respuesta y para las variables regresoras.

Si suponemos que cada y_i está relacionada de forma lineal con las variables $x_1, x_2, \dots, x_n$, tenemos el modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad i = 1, 2, \dots, n; \quad n \geq k + 1$$

ε_i es un error incorrelado de observación en observación, con media 0 y varianza σ^2 . Los x_{ij} son no aleatorios y medidos con error despreciable.

Un modelo es lineal cuando lo es en los parámetros β_i .

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \varepsilon$$

$$\ln y = \beta_0 + \beta_1 \left(\frac{1}{x_1}\right) + \beta_2 \left(\frac{1}{x_2}\right) + \varepsilon$$

y son no lineales

$$y = \beta_0 + \beta_1 x_1^{\beta_2} + \beta_3 x_2^{\beta_4} + \varepsilon$$

$$y = \frac{\beta_0}{1 + e^{-(\beta_1 x_1 + \beta_2 x_2)}} + \varepsilon$$

6.7.1 Mod. lineal general y proc. de los mínimos cuadrados

El modelo descrito anteriormente se puede escribir de la forma

$$Y = X\beta + \varepsilon$$

donde

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ε e Y son vectores aleatorios, X es una matriz $n \times p$ de datos, con $p = k + 1$, que es el número total de parámetros del modelo, y recibe el nombre de matriz de diseño.

6.7.2 Método de mínimos cuadrados

Se trata de buscar el estimador b del vector β que satisfaga:

$$\frac{\partial}{\partial b}[\varepsilon'\varepsilon] = \frac{\partial}{\partial b}[(Y - Xb)'(Y - Xb)] = 0$$

Si X es una matriz de rango máximo, $b = (X'X)^{-1}X'Y$ y un estimador insesgado de σ^2 viene dado por

$$s^2 = \frac{(Y - Xb)'(Y - Xb)}{n - p} = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - p}$$

donde p es el número de parámetros que deben ser estimados e $\hat{y}_i$ es la respuesta estimada en el i -ésimo dato.

Utilizando que $E[\varepsilon] = \Theta$ (vector nulo) se puede comprobar que b es un estimador insesgado de β y si además suponemos que los errores ε_i son incorrelados, entonces $Var[b] = \sigma^2(X'X)^{-1}$

Teorema 6.7.1 — Gauss-Markov. Si $E(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2 \cdot I$ (matriz de varianzas y covarianzas), entonces el estimador de mínimos cuadrados de β es de mínima varianza entre todos los insesgados.

Al igual que ocurría en el caso de la regresión lineal simple, se tiene la siguiente relación entre SS_{Tot} , SS_{Reg} y SS_{Res} :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde $\hat{y}_i$ es la respuesta estimada para la observación i -ésima, y viene dado por

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki}$$

El coeficiente de determinación se interpreta del mismo modo que en el caso de la regresión lineal simple, y nuevamente puede calcularse como:

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Tot}}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Por otra parte, y aunque no entraremos en ese terreno, al igual que ocurría en el caso de la regresión lineal simple, es posible calcular intervalos de confianza así como realizar contrastes de hipótesis acerca de los parámetros que intervienen en el modelo (siempre bajo la hipótesis de normalidad de los errores).

6.8 Regresión no lineal

Hasta ahora hemos considerado modelos con estructura lineal en los parámetros. En ocasiones aparecen situaciones experimentales que requieren el uso de modelos no lineales. Hay muchos ejemplos de modelos no lineales y ya vimos algunos en la sección 6.7.

Aplicar el método de mínimos cuadrados directamente sobre estos modelos puede conducir a ecuaciones complicadas que no se pueden resolver con herramientas algebraicas y requieren el uso de métodos numéricos.

Por este motivo, no vamos a profundizar en la resolución de estos problemas y veremos, únicamente, cómo un cambio de variable adecuado puede ayudar a obtener una solución aproximada cuando se trata de ajustar determinados modelos. Esta solución no deja de ser una aproximación pues con el cambio de variable, generalmente, es posible que no respetemos ciertas hipótesis como los errores aditivos o la hipótesis de normalidad.

Ejercicio 6.9 El director de un restaurante tiene la siguiente distribución de frecuencias sobre el coste (en euros) de las comidas solicitadas por sus clientes en un fin de semana:

Número de comidas sol.	30	36	40	43	50	55	60	71	75	79	62
Coste por comida	24	22	21	20	19	17	16	15	13	10	12

Hacer un ajuste potencial ($y = A \cdot x^B$). ■

6.9 Ejercicios

Ejercicio 6.10 Ajustar una recta a través de los cinco puntos siguientes. Obtener las estimaciones para β_0 y β_1 . Representar gráficamente los puntos así como la recta obtenida.

x	-2	-1	0	1	2
y	3	2	1	1	0,5

¿Presentan los datos suficiente evidencia para indicar que la pendiente β_1 difiere de cero? (Utilizar un nivel de significación del 5 %). Obtener un intervalo de confianza al 95 % para β_0 y para β_1 . ■

Ejercicio 6.11 Los experimentos diseñados para medir valores de CL50 en peces se efectúan con dos métodos diferentes. En un método el agua fluye continuamente y el otro método tiene condiciones de agua en reposo. A fin de establecer los criterios para sustancias tóxicas, la Agencia para la protección ambiental pretende ajustar todos los resultados a la condición dinámica, por lo que se requiere un modelo para relacionar los dos tipos de observaciones.

P. tóxico.	1	2	3	4	5	6	7	8	9	10
CL50 din.	23.00	22.30	9.40	9.70	0.15	0.28	0.75	0.51	28.00	0.39
CL50 est.	39.00	37.50	22.20	17.50	0.64	0.45	2.62	2.36	32.00	0.77

Ajustar el modelo $y = \beta_0 + \beta_1 x + \varepsilon$. Interpretar los resultados. Estimar el valor dinámico para un producto tóxico con un valor estático de CL50 de $x = 12$ partes por millón. ■

Ejercicio 6.12 Los promedios de los precios de venta de nuevas casas para una sola familia durante un periodo de 8 años se indican en la tabla siguiente. Sea y el promedio de los precios de venta ($\times 1000$) y x el año. Ajustar el modelo $y = \beta_0 + \beta_1 x + \varepsilon$, aplicando números enteros, $1, \dots, 8$ para los años. ¿Qué se puede concluir de los resultados?

Año	72	73	74	75	76	77	78	79
Precio prom.	27.6	32.5	35.9	39.3	44.2	48.8	55.7	62.9

Ejercicio 6.13 Se realizó un experimento para observar el efecto de un incremento de la temperatura en la efectividad de un antibiótico. Se almacenaron tres porciones de una onza del antibiótico durante el mismo periodo de tiempo a cada una de las siguientes temperaturas: 30^0 , 50^0 , 70^0 y 90^0 . Las lecturas de la efectividad observadas a la temperatura del periodo experimental fueron:

Efect. (y)	38, 43, 29	32, 26, 33	19, 27, 23	14, 19, 21
Temp. (x)	30^0	50^0	70^0	90^0

(a) Mediante el método de mínimos cuadrados, ajusta una recta a los datos observados.

- (b) Representar gráficamente los puntos y la recta obtenida.
 (c) Calcular S^2 .

Ejercicio 6.14 Ajustar un modelo del tipo $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ a los datos de la siguiente tabla. Representar gráficamente los puntos así como la parábola ajustada.

y	−3	−2	−1	0	1	2	3
x	1	0	0	−1	−1	0	0

Ejercicio 6.15 Las estadísticas de enfermedades en Florida para la década que terminó en 1976 demuestran que la hepatitis infecciosa tenía las tasas de incidencia que se presentan en la siguiente tabla (expresadas en casos por población de 100.000).

x	67	68	69	70	71	72	73	74	75	76
y	10,5	18,5	22,6	27,2	31,2	33,0	44,9	49,4	35,0	27,6

- (a) Sea y la tasa de incidencia y x el año “codificado” (0 para 1967, 1 para 1968, hasta 9 para 1976). Ajustar un modelo del tipo $y = \beta_0 + \beta_1 x + \varepsilon$.
 (b) Ajustar los datos al modelo $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$.

Calcular en cada caso una medida de la bondad del ajuste realizado y de la dispersión relativa de los datos respecto a la curva ajustada.

Hemos visto en ejercicios anteriores que en ocasiones es conveniente, para simplificar los cálculos, tener los valores de x espaciados uniformemente y que estén distribuidos simétricamente alrededor de cero. Se puede cambiar la escala de los x sin pérdida de información en el análisis estadístico.

Ejercicio 6.16 A partir de los datos del ejercicio “venta de nuevas casas”: “

- (a) Determinar SS_{Res} y S^2 .
 (b) Aplicar a los valores de x (dados originalmente en una escala de 1 a 8) el cambio de variable:

$$x^* = \frac{x - 4,5}{0,5}$$

y ajustar a los datos resultantes el modelo $y = \beta_0^* + \beta_1^* x^* + \varepsilon$. Calcular entonces SS_{Res} . Comparar SS_{Res} con el valor obtenido en el apartado (a).

Ejercicio 6.17 Los datos muestrales siguientes fueron registrados por una compañía de mudanzas sobre los pesos de seis equipos, las distancias a que fueron trasladados y los daños sufridos:

Peso (1000 Kg.)	Distancia (1000 Km.)	Daño (1000 Pts.)
4.0	1.5	16
3.0	2.2	11.2
1.6	1.0	6.9
1.2	2.0	9
3.4	0.8	12.3
4.8	1.6	18.6

- (a) Ajustese una ecuación de la forma $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
 (b) Estimar los daños que sufrirá un equipo que pesa 2400 Kg. cuando se traslada 1200 Km.

Ejercicio 6.18 Se realiza un experimento para determinar la duración de vida (y) de ciertos circuitos electrónicos en función de dos variables de agrupación x_1 y x_2 , con los siguientes resultados.

y	11	8	73	21	46	30
x_1	-10	0	10	-10	0	10
x_2	0	-5	5	0	5	-5

- (a) Ajustar a los datos un modelo del tipo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.
 (b) Calcular e interpretar una medida de la bondad del ajuste realizado.

Ejercicio 6.19 Con objeto de determinar la relación entre la razón agua/cemento y la resistencia de las vigas, se realizó una serie de pruebas en las que se recogieron los siguientes resultados:

agua/cemento	1.21	1.29	1.37	1.46	1.62	1.79
Resistencia	1.302	1.231	1.061	1.040	0.803	0.711

- (a) Ajustar el modelo $y = \beta_0 + \beta_1 x + \varepsilon$.
 (b) Probar $H_0 : \beta_1 = 0$ frente a $H_1 : \beta_1 < 0$, con $\alpha = 0,05$. ¿Que se podría decir cuando $x = 0$?
 (c) Encontrar un intervalo de confianza al 90 % para la resistencia esperada de la viga cuando se utiliza una relación agua/cemento de 1,5. ¿Qué ocurre con el intervalo de confianza si tratamos de estimar promedios de resistencia para razones de agua/cemento de 0,3 o 2,7?

Ejercicio 6.20 En la siguiente tabla la variable “ x ” representa la fuerza de tensión aplicada a una barra de acero en miles de libras e “ y ” es el estiramiento resultante en milésimas de pulgada:

x	1	2	3	4	5	6
y	14	33	40	63	76	85

- (a) Representar gráficamente los datos para comprobar si es razonable suponer que la regresión de y sobre x es lineal.

- (b) Encontrar la recta de mínimos cuadrados y predecir el estiramiento cuando la tensión es de 3,5 miles de libras. Dar un intervalo de confianza al 95 % para el estiramiento medio de la barra cuando se le aplica esa tensión.
- (c) Construir un intervalo de confianza para la pendiente con un nivel de confianza del 95 %.

Ejercicio 6.21 Las mediciones de la humedad relativa en el lugar de almacenamiento y la humedad en una muestra de las materias primas allí almacenadas dieron los siguientes resultados:

Humedad (x)	42	35	50	43	48	62	31	36	44	39	55	48
Hum. Muestra (y)	12	8	14	9	11	16	7	9	12	10	13	11

- (a) Determinar la recta de regresión de y sobre x . Dar una medida de la bondad del ajuste realizado.
- (b) ¿Es razonable suponer que la humedad del almacén influye de forma lineal en la humedad de la fibra? ($\alpha = 0.05$)
- (c) Encontrar un intervalo de confianza del 99 % para el contenido medio de humedad de las materias primas cuando la humedad del lugar de almacenamiento es del 40 %.

Ejercicio 6.22 Una compañía de seguros desea relacionar la cantidad y (miles de euros) de daños producidos por el fuego, en incendios residenciales, con la distancia x (en Km.), entre la residencia y el parque de bomberos más próximo. Una muestra de 10 incendios arroja los siguientes datos:

x	3.4	1.8	4.6	2.3	3.1	5.5	0.7	3	2.6	4.3
y	26.2	17.8	31.3	23.1	27.5	36	14.1	22.3	19.6	31.3

Ejercicio 6.23 Mediante el método de los mínimos cuadrados, ajustar una función de la forma $y = kx^b$, donde $k > 0$ y $b > 0$, a la nube de puntos:

X	1.1	1.5	2	2.5	2.9
Y	1.4	2.3	4.2	6.4	9

Determinar una medida de la bondad del ajuste realizado.

Ejercicio 6.24 La siguiente tabla muestra los valores experimentales de la presión P (libras/pulgada cuadrada), de una masa de gas, correspondientes a diferentes valores de su volumen V (pulgadas cúbicas).

Volumen	54.3	61.8	72.4	88.7	118.6	194
Presión	61.2	49.5	37.6	28.4	19.2	10.1

Según los principios de la termodinámica, existe una relación entre las variables que tiene la forma $P \cdot V^\lambda = C$, donde λ y C son constantes.

- (a) Hallar los valores de λ y C .
- (b) Hallar la ecuación que liga a P y V .
- (c) Estimar P cuando $V = 1000$ pulgadas cúbicas.

■

Ejercicio 6.25 El número de cm. y que una estructura recién construida se hunde en el suelo, viene dado por

$$y = 3 - \beta e^{-\alpha x}$$

donde x es el número de meses que lleva construido.

Del estudio de seis estructuras se recogen los siguientes datos

x	2	4	6	12	18	24
y	1.07	1.88	2.26	2.78	2.97	2.99

Estimar α y β . (Sugerencia: obsérvese que la relación entre $\ln(3 - y)$ y x es lineal)

■

Ejercicio 6.26 Un estudio psicológico proporcionó los siguientes datos sobre el número de hijos en la familia, X , y el coeficiente intelectual promedio de esa familia, Y :

Número de hijos en fam.	1	2	3	4	5	6	8	9	10
Coef. intelec. medio familiar	105	102	104	100	97	101	93	97	88

Se pide, hacer un ajuste hiperbólico ($y = \frac{1}{Ax+B}$).

■