

## Capítulo 3

# Introducción a las variables aleatorias

- 
1. Introducción. La variable aleatoria como modelo de variabilidad
  2. Variables aleatorias univariantes discretas
  3. Variables aleatorias univariantes continuas
  4. Medidas características de las variables aleatorias
  5. Variables aleatorias multivariantes (avanzado)
- 

---

<sup>0</sup>Apuntes realizados por Ismael Sánchez. Universidad Carlos III de Madrid.

### 3.1. Introducción

En el tema anterior utilizamos la denominación de **experimento aleatorio** para nombrar a aquellos experimentos en los que no se obtiene siempre el mismo resultado de la variable de interés. Al definir un experimento hay que fijar sus **condiciones de experimentación**, de forma que si obtenemos un nuevo dato manteniendo constantes dichas condiciones, estaremos repitiendo el mismo experimento. Cuando las condiciones de experimentación no engloban a todos los factores que influyen en la variable de interés, las repeticiones del experimento podrán dar lugar a valores diferentes. Esta posible variabilidad en el resultado es debida a la influencia de esos factores que no mantenemos constantes. Por contra, el resultado de un experimento determinista es una constante, y por tanto perfectamente predecible.

Llamaremos **variable aleatoria** al **resultado de un experimento aleatorio**. Por ejemplo, el resultado de lanzar un dado es una variable aleatoria, o el resultado de medir el tiempo que un ordenador tarda en acceder a una red, etc. Según esta definición, una variable aleatoria no tendrá, en general, siempre el mismo valor. Depende de cómo influya el azar en cada una de las repeticiones del experimento. Para que una **variable aleatoria** esté bien definida, necesitamos conocer dos elementos fundamentales:

1. **El espacio muestral**, que consiste en la unión de todos los sucesos elementales; es decir, de todos los resultados observables posibles.
2. **Una regla de asignación de probabilidades** para los distintos sucesos en que estemos interesados. A esta regla de asignación se le suele denominar **modelo de probabilidad**.

Por ejemplo, el resultado de lanzar una moneda es una variable aleatoria; su espacio muestral es  $\{\text{cara}, \text{cruz}\}$ , y el **modelo de probabilidad** que establece la probabilidad de cada uno de esos dos resultados es  $P(\text{cara})=P(\text{cruz})=0.5$ . Una variable aleatoria es pues algo abstracto, matemático. Al valor observado de una variable aleatoria al ejecutar el experimento se le llama **realización**. Por ejemplo, si lanzamos una moneda cinco veces obtendremos cinco realizaciones de la variable aleatoria. Por ejemplo, podríamos haber obtenido (cara, cara, cruz, cara, cruz, cara). Es importante notar que la proporción observada de cada resultado en una muestra concreta de resultados no tiene por qué coincidir con la probabilidad de cada uno de ellos, pues la probabilidad es la frecuencia relativa si se repite el experimento **indefinidamente**.

Se denomina **población de valores de una variable aleatoria al conjunto de realizaciones que se obtendrían de repetir el experimento todas las veces posibles**. Se puede decir entonces que las características de la población son las de la variable aleatoria que la genera.

En este tema aprenderemos a manejar los **modelos de probabilidad** que describen los posibles resultados de una variable aleatoria, asignando probabilidades a los diferentes sucesos que nos interesen. En este tema trataremos los modelos de probabilidad a nivel muy general, analizando sus características generales y viendo cómo extraer información (probabilidades) de ellos. En el tema siguiente, mostraremos un pequeño catálogo de modelos de probabilidad que se utilizan con mucha frecuencia en el campo de la ingeniería. En temas posteriores, aprenderemos a seleccionar, a partir del análisis de un conjunto de datos, un modelo de probabilidad concreto que ayude a describir un problema real. Las variables aleatorias que estudiaremos en este tema serán cuantitativas, y las clasificaremos en discretas o continuas.

**NOTACIÓN:** Las variables aleatorias las denotaremos habitualmente con letras mayúsculas del final del alfabeto. Lo más habitual es emplear las letras  $X$  e  $Y$ . A las realizaciones de dichas variables, es decir, a los valores finalmente observados, los denotaremos por la respectiva

letra minúscula. Así, por ejemplo, los distintos valores que se puedan observar de la variable aleatoria  $X$  serán  $x_1, x_2, \dots$ . La probabilidad de que la variable aleatoria  $X$  tome el valor  $x$  o  $x_i$  se denota entonces como  $P(X = x)$  o  $P(X = x_i)$  respectivamente. Es frecuente, sin embargo, no usar notación diferente para nombrar a la variable aleatoria y a sus valores observados, y en muchos textos se emplea únicamente la notación en minúsculas.

## 3.2. Variables aleatorias univariantes discretas

Son las variables aleatorias que toma valores (cuantitativos) discretos. Ejemplos: resultado de lanzar un dado o una moneda, número de defectos que tiene cada artículo manufacturado por la máquina  $M$ . Número de clientes que llegan a cierto puesto de servicio por unidad de tiempo, número de desconexiones de una red por unidad de tiempo, número de peticiones de acceso a un servidor por unidad de tiempo, etc.

Para construir un **modelo de probabilidad** que asigne probabilidades a los diferentes sucesos en una variable aleatoria discreta se usan las llamadas **función de probabilidad** y **función de distribución**, que definimos a continuación

### 3.2.1. Función de probabilidad

La función de probabilidad se define como la función  $p(x)$  de la variable aleatoria discreta  $X$  que asigna a cada valor diferente de  $X : x_1, x_2, \dots, x_K$  la probabilidad de ser obtenido en una repetición del experimento aleatorio.

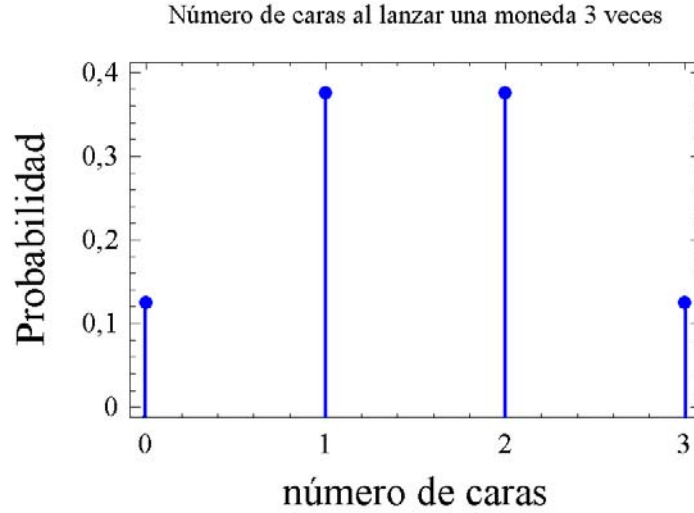
**Ejemplo 1** Sea la variable aleatoria  $X =$  resultado de lanzar un dado, se tiene que  $x = 1, \dots, 6$  y al ser cada valor igual de probable  $p(x) = 1/6$  para  $x = 1, \dots, 6$ .

A este tipo de variables aleatorias en las que cada valor es igual de probable se les denomina **variables aleatorias uniformes discretas**. Por ejemplo, una variable aleatoria uniforme discreta definida entre 1 y  $N$  tendrá como sucesos elementales los valores  $1, 2, \dots, N$  y su función de probabilidad será  $p(x) = 1/N$ , para  $x = 1, \dots, N$ .

**Ejemplo 2** El experimento consiste en lanzar una moneda tres veces y sumar el número de caras que se obtienen. La variable aleatoria es  $X =$  número de caras al lanzar una moneda tres veces, y tomará los valores  $x = \{0, 1, 2, 3\}$ . Si llamamos  $C$  al suceso cara y  $+$  al suceso cruz, los posibles resultados en los tres lanzamientos son  $CCC, CC+, C+C, +CC, C++, +C+, ++C, +++$ . Por tanto, podemos deducir la función de probabilidad siguiente

$$p(x) = \begin{cases} 1/8 & \text{si } x = 0 \\ 3/8 & \text{si } x = 1 \\ 3/8 & \text{si } x = 2 \\ 1/8 & \text{si } x = 3 \end{cases}$$

La siguiente figura representa gráficamente esta función de probabilidad



El valor que toma la función de probabilidad para un valor concreto de  $x$  se escribe como  $p(x_i)$ . Es decir,  $p(x_i) = P(X = x_i)$ . Por ejemplo, en el caso anterior de sumar las caras obtenidas en tres lanzamientos se tiene que  $p(2) = 3/8$ . Se comprueba entonces que si  $x$  puede tomar  $K$  valores diferentes,  $x_1, \dots, x_K$  se ha de verificar que

$$\sum_{i=1}^K p(x_i) = 1.$$

Para construir una función de probabilidad es necesario conocer las propiedades del sistema que genera los diferentes datos, bien a través de sus propiedades teóricas, como en el ejemplo anterior de las monedas, o conociendo las propiedades de toda la población, como en el ejemplo siguiente.

**Ejemplo 3** *Un artículo manufacturado está compuesto por dos piezas,  $P1$  y  $P2$ . Se sabe por información histórica que el estado de cada pieza, defectuosa o aceptable, es independiente la una de la otra. Dicha información histórica revela además que el 98 % de las piezas  $P1$  son aceptables, y el 99 % de las piezas  $P2$  son aceptables. Construyamos la función de probabilidad de la variable  $X = \text{número de piezas defectuosas del artículo}$ .*

*La variable aleatoria  $X$  puede tomar 3 valores,  $x = 0, 1, 2$ . Si llamamos  $D_1$  al suceso: la pieza  $P1$  es defectuosa, y  $D_2$  al suceso: la pieza  $P2$  es defectuosa tenemos que:*

$$\begin{aligned} P(X = 0) &= P(\text{ninguna pieza defectuosa}) = P(\bar{D}_1 \cap \bar{D}_2) \\ &= P(\bar{D}_1)P(\bar{D}_2) = 0,98 \times 0,99 = 0,9702. \end{aligned}$$

Por otra parte tenemos que

$$P(X = 1) = P(\text{sólo una pieza defectuosa}) = P[(D_1 \cap \bar{D}_2) \cup (\bar{D}_1 \cap D_2)].$$

Nótese que no es lo mismo el que **una** pieza sea defectuosa con que el que **alguna** pieza sea defectuosa, que sería  $D_1 \cup D_2$  (¿por qué?). Aplicando que  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ,

se tiene que

$$\begin{aligned}
 P[(D_1 \cap \bar{D}_2) \cup (\bar{D}_1 \cap D_2)] &= P(D_1 \cap \bar{D}_2) + P(\bar{D}_1 \cap D_2) - P(D_1 \cap \bar{D}_2 \cap \bar{D}_1 \cap D_2) \\
 &= P(D_1 \cap \bar{D}_2) + P(\bar{D}_1 \cap D_2) \\
 &= P(D_1)P(\bar{D}_2) + P(\bar{D}_1)P(D_2) = 0,0198 + 0,0098 \\
 &= 0,0296.
 \end{aligned}$$

Finalmente, tenemos que

$$\begin{aligned}
 P(X = 2) &= P(\text{ambas piezas defectuosas}) = P(D_1 \cap D_2) \\
 &= P(D_1)P(D_2) = 0,0002.
 \end{aligned}$$

Vemos que  $P(X = 0) + P(X = 1) + P(X = 2) = 1$ . La función de probabilidad de  $x$  es por tanto:

$$p(x) = \begin{cases} 0,9702 & \text{si } x = 0 \\ 0,0296 & \text{si } x = 1 \\ 0,0002 & \text{si } x = 2 \end{cases}$$

### 3.2.2. Función de distribución

Otro procedimiento para obtener probabilidades de una variable aleatoria, totalmente equivalente al anterior, es mediante la llamada función de distribución. La función de distribución es una función continua que está definida en toda la recta real. Supongamos que  $x$  es un punto de dicha recta. Entonces el valor de la función de distribución de la variable aleatoria  $X$  en el punto  $x$  se denota por  $F(x)$  y es

$$F(x) = P(X \leq x).$$

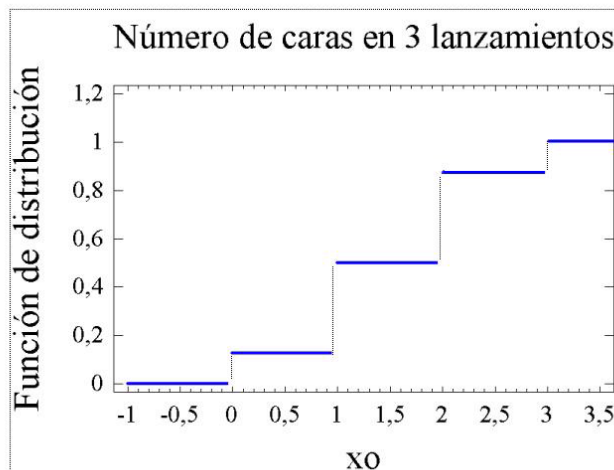
Es por tanto una función que va **acumulando la probabilidad de observar los distintos valores a medida que nos desplazamos desde  $-\infty$  a  $+\infty$** . De esta definición se puede deducir que en el extremo inferior se tiene  $F(-\infty) = 0$ , mientras que en el extremo superior se tiene  $F(+\infty) = 1$ . Para construir la función de distribución se ha de recorrer por tanto todos los puntos de la recta real. Lo más sencillo es comenzar desde el extremo  $-\infty$ . Entonces, en cada punto vamos calculando **la probabilidad de obtener valores menores e iguales que él**.

**Ejemplo 4** La función de probabilidad para el Ejemplo 1 de la sección anterior será:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0, \text{ pues } P(X < 0) = 0 \\ 1/8 & \text{si } 0 \leq x < 1, \text{ pues } P(X < 1) = P(X < 0) + P(X = 0) \\ 4/8 & \text{si } 1 \leq x < 2, \text{ pues } P(X < 2) = P(X < 1) + P(X = 1) \\ 7/8 & \text{si } 2 \leq x < 3, \text{ pues } P(X < 3) = P(X < 2) + P(X = 2) \\ 1 & \text{si } x \geq 3 \end{cases}$$

Según esta función tendremos por ejemplo que  $F(2,5) = 7/8$ , o  $F(1,2) = 4/8$ . La figura siguiente ilustra la construcción de esta función de distribución para el ejemplo de  $X = \text{número de caras al}$

lanzar 3 veces una moneda.



Si 0 es el valor más pequeño que toma la variable aleatoria  $X = \text{número de caras al lanzar 3 veces una moneda}$ , se tendrá que para todo valor  $x$  menor que 0  $F(x) = 0$ . En  $x = 0$  la función de distribución tendrá un escalón pues pasará de valor 0 a valer  $p(0) = 1/8$ . A partir de ese valor  $F(x)$  tendrá un valor constante hasta llegar a  $x = 1$ . En ese punto, la probabilidad acumulada será la probabilidad de sacar 0 ó 1 cara, y por tanto  $F(1) = p(0) + p(1)$ , y así sucesivamente.

### 3.3. Variables aleatorias univariantes continuas

En las variables aleatorias continuas, ya no se tienen un número finito de valores, sino infinito. Por tanto no tiene sentido hablar de la probabilidad de obtener un valor concreto, pues ésta será siempre cero. Por ejemplo, si consideramos el experimento de seleccionar al azar un número real comprendido en el intervalo  $[3,4]$ , ¿cuál es la probabilidad de elegir el número  $\pi$  (3.1416...). Al haber infinitos valores todos con las mismas posibilidades de ser seleccionados, tendremos que  $P(3,1416...) = 1/\infty = 0$ . Y similar resultado se obtiene para cualquier variable continua.

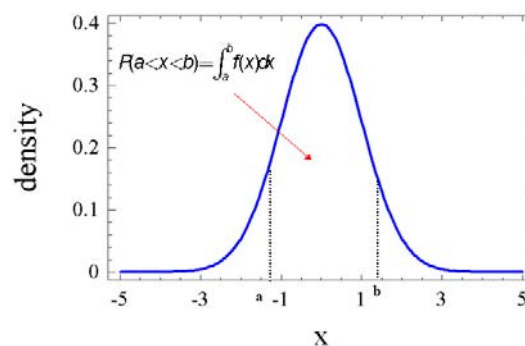
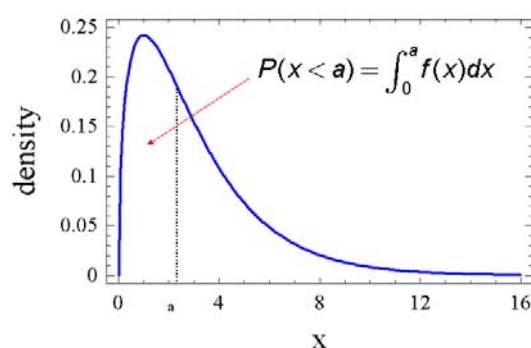
Por tanto, no tiene sentido definir una variable aleatoria continua utilizando la función de probabilidad. Con variables continuas, hablaremos de probabilidades de obtener valores en cierto intervalo. Por ejemplo, la probabilidad de tardar más de 10 minutos en realizar una tarea, o la probabilidad de que una pieza manufacturada tenga una longitud comprendida entre 7.5 y 7.8 centímetros. Para calcular probabilidades con una variable aleatoria continua usaremos la llamada función de densidad así como la función de distribución

#### 3.3.1. Función de densidad

Sea  $X$  una variable aleatoria continua que toma valores en la recta real. La función de densidad, o función de densidad de probabilidad, de  $X$  en el punto  $X = x$  es una función  $f(x) \geq 0$  tal que la probabilidad de obtener un valor entre  $x_1$  y  $x_2$  es

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx.$$

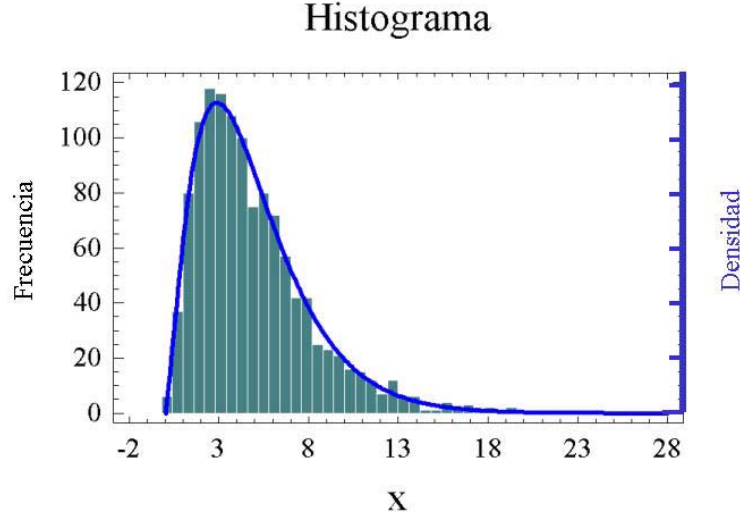
La función de densidad de probabilidad es similar a cualquier otra función de densidad en física. Por ejemplo, la densidad de masa de un cuerpo es su peso por unidad de volumen, mientras que la función de densidad de probabilidad es la probabilidad por unidad de medida de la variable  $X$ . De esta forma, sumando (integrando) la densidad a lo largo de un número de unidades de medida conseguimos la probabilidad. A continuación se muestran algunos ejemplos de funciones de densidad de probabilidad.



No es fácil, en general, construir una función de densidad que se adecúe perfectamente a un problema concreto. Lo que se suele hacer en la práctica es utilizar alguna función de densidad de las ya existentes en la literatura de forma que describa a nuestra variable aleatoria suficientemente. En el capítulo siguiente se propondrán funciones de densidad que pueden usarse como modelos útiles en muchas situaciones reales. En posteriores capítulos aprenderemos a seleccionar una función de densidad para una población a partir del análisis de una muestra de datos. Por tanto, en este capítulo, la función de densidad será una función que nos proporcione un analista, y de la que intentaremos sacar el máximo de información sobre la variable aleatoria que representa.

La función de densidad puede interpretarse como el polinomio de frecuencias (ver Tema 1) que se obtendría si hiciésemos un histograma con todos los valores posibles de la variable aleatoria  $X$ , es decir con los infinitos valores de la población. Al haber tantos datos, los intervalos del histograma se podrían hacer entonces muy estrechos, con lo que el polígono de frecuencias podría ser una función suave. La función de densidad sería entonces el polinomio de frecuencias límite cuando los

intervalos son de ancho  $dx$ , como se ilustra en la siguiente figura.



En esta figura, el eje de ordenadas representa las unidades de medida de la variable aleatoria, por ejemplo los segundos que se tarda en ejecutar una tarea, o los centímetros de un artículo. En este caso, la figura muestra en el eje de ordenadas de la izquierda las frecuencias absolutas de cada rectángulo del histograma, pero también podrían ser frecuencias relativas. Es importante notar que dichas unidades del eje de ordenadas del histograma no deben confundirse con las unidades de la función de densidad (eje de la derecha), que son las de **probabilidad por unidad de medida**. Veamos esta diferencia. Supongamos un histograma realizado con una muestra de  $n$  individuos. Imaginemos un rectángulo de dicho histograma de altura  $n_c$  = número de individuos que hay en dicho rectángulo y anchura  $a$ . Denotemos por  $x_c$  al valor central del rectángulo de forma que el intervalo de anchura  $a$  comprende los valores  $[x_c - a/2; x_c + a/2]$ . Entonces, la probabilidad de que un individuo tenga un valor dentro de dicho intervalo es

$$P(x_c - a/2 \leq X \leq x_c + a/2) = \frac{n_c}{n}, \quad (3.1)$$

mientras que la densidad de probabilidad, o simplemente densidad, de dicho intervalo es la probabilidad de estar en ese intervalo dividido por el número de unidades de medida que lo comprende, es decir,

$$\text{Densidad de probabilidad del intervalo} = \frac{P(x_c - a/2 \leq X \leq x_c + a/2)}{a}. \quad (3.2)$$

La densidad de probabilidad en el punto  $x_c$ , que es lo que proporciona la función de densidad  $f(x_c)$ , será por tanto

$$f(x_c) = \lim_{a \rightarrow 0} \frac{P(x_c - a/2 \leq X \leq x_c + a/2)}{a}.$$

La función de densidad es entonces una función matemática que permite calcular mediante integración la probabilidad de sucesos, siendo estos sucesos intervalos de valores de la variable



aleatoria  $X$ . Por tanto, una función de densidad debe cumplir

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad (3.3)$$

Al tratarse de función de densidad, y no de probabilidad, es posible que una función de densidad pueda valer más de 1 en algún punto.

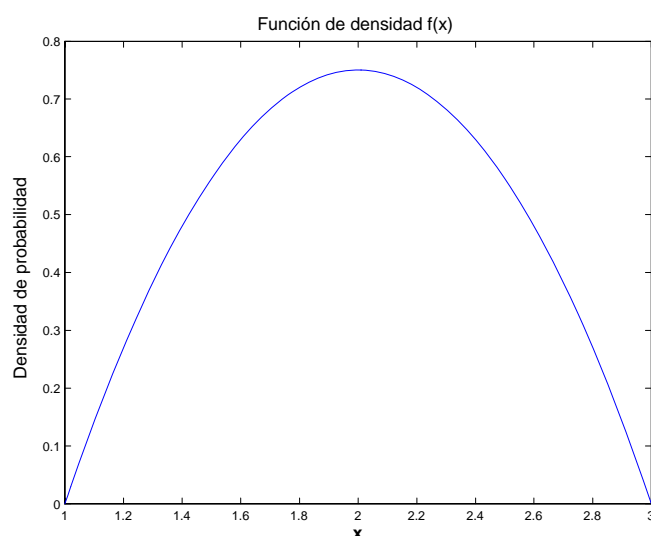
**Ejemplo 5** Sea  $X =$  longitud de cierta pieza, que se distribuye según la siguiente función de densidad

$$f(x) = \begin{cases} k(x-1)(3-x) & \text{si } x \in [1, 3] \\ 0 & \text{resto} \end{cases}$$

¿Cuál debe ser el valor de  $k$  para que  $f(x)$  sea una función de densidad?. Para calcular  $k$  podemos aplicar la restricción (3.3). Se tiene entonces que

$$\begin{aligned} \int_{-\infty}^{\infty} k(x-1)(3-x)dx &= \int_{-\infty}^1 f(x)dx + \int_1^3 f(x)dx + \int_3^{\infty} f(x)dx \\ &= \int_1^3 k(x-1)(3-x)dx = 1 \Rightarrow k = \frac{3}{4}. \end{aligned}$$

La siguiente figura representa esta función de densidad.

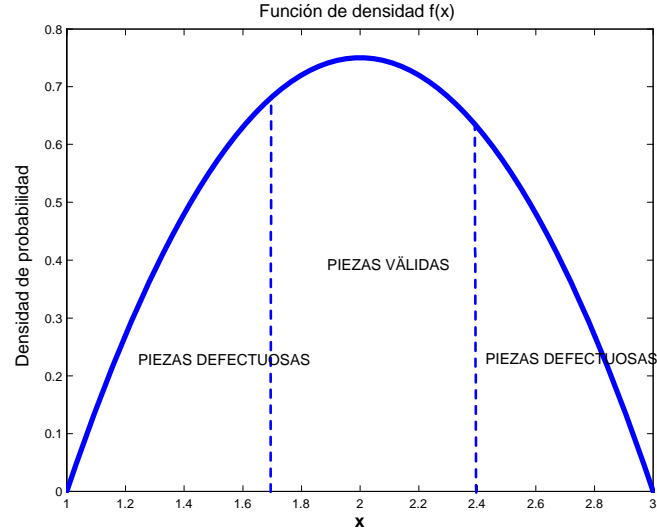


Una vez definida la función de densidad se pueden calcular probabilidades. Estas piezas se consideran válidas si su longitud está comprendida entre 1.7 cm y 2.4 cm. ¿Cuál es la probabilidad de que una pieza sea útil? Esta probabilidad será el área bajo la curva  $f(x)$  que está comprendida en el intervalo  $[1.7; 2.4]$ , es decir

$$P(\text{válida}) = P(1.7 < X < 2.4) = \int_{1.7}^{2.4} \frac{3}{4}(x-1)(3-x)dx = 0.502.$$

Por tanto, algo más del 50 % de las piezas serán válidas. Este resultado indica que el proceso de producción es muy deficiente, pues casi la mitad de las piezas son defectuosas, como puede verse en

la figura siguiente; mientras que en un proceso que tenga un funcionamiento adecuado la proporción de piezas defectuosas es mínima. Nótese que al tratarse de variables continuas, es irrelevante utilizar el signo  $=$  en las desigualdades. Por ejemplo, se cumple que  $P(X \geq x) = P(X > x)$ .



### 3.3.2. Función de distribución

La definición de función de distribución en variables aleatorias continuas es la misma que en el caso de variables discretas. Es decir, la función de distribución en el punto  $x$  es  $F(x) = P(X \leq x)$ . Usando la función de densidad para calcular probabilidades se tiene que

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du.$$

De esta expresión se deduce también que

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

Si conocemos la función de distribución, es también sencillo calcular la probabilidad de obtener un valor en cierto intervalo:

$$\begin{aligned} P(x_1 < X < x_2) &= \int_{x_1}^{x_2} f(x) dx = \int_{-\infty}^{x_2} f(x) dx - \int_{-\infty}^{x_1} f(x) dx \\ &= F(x_2) - F(x_1). \end{aligned}$$

**Ejemplo 6** Sea  $X$  una variable aleatoria con función de densidad

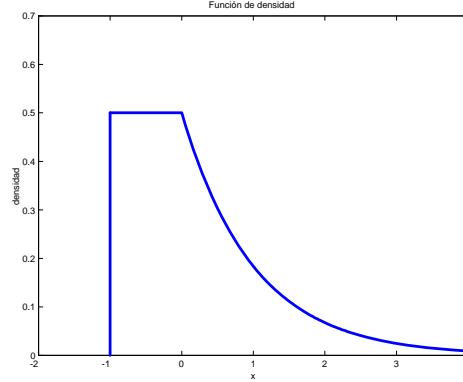
$$f(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 0 \\ ae^{-x} & x > 0 \end{cases}$$

siendo  $a$  una constante. Calcularemos en primer lugar el valor de  $a$  para que  $f$  sea una función de densidad. Posteriormente calcularemos la función de distribución de probabilidad.

**Ejemplo 7** Como  $\int_{-1}^{\infty} f(x)dx = 1$ , entonces

$$\begin{aligned}\int_{-1}^{\infty} f(x)dx &= \int_{-1}^0 \frac{1}{2}dx + \int_0^{\infty} ae^{-x}dx \\ &= \frac{1}{2} + (-ae^{-x})\Big|_0^{\infty} = \frac{1}{2} + a = 1 \iff a = \frac{1}{2}.\end{aligned}$$

La siguiente figura muestra esta densidad  $f(x)$



Como la función de densidad está definida por tramos, habrá que calcular  $F(x)$  en cada uno de esos tramos.

*Tramo 1:*  $x < -1$ . Para  $x < -1$ , se tiene que  $f^{(Tramo\ 1)}(x) = 0$ . Por tanto

$$F(x) = P(X \leq x) = \int_{-\infty}^x f^{(Tramo\ 1)}(u)du = \int_{-\infty}^x 0du = 0; x < -1.$$

*Tramo 2:*  $-1 \leq x \leq 0$ . En este tramo se tiene que  $f^{(Tramo\ 2)}(x) = 1/2$ . Entonces

$$\begin{aligned}F(x) &= P(X \leq x) = P(X < -1) + P(-1 \leq X \leq x) \\ &= \int_{-\infty}^{-1} f^{(Tramo\ 1)}(u)du + \int_{-1}^x f^{(Tramo\ 2)}(u)du = \\ &= \int_{-\infty}^{-1} 0du + \int_{-1}^x \frac{1}{2}du \\ &= 0 + \frac{1}{2}(x+1) ; -1 \leq x \leq 0.\end{aligned}$$

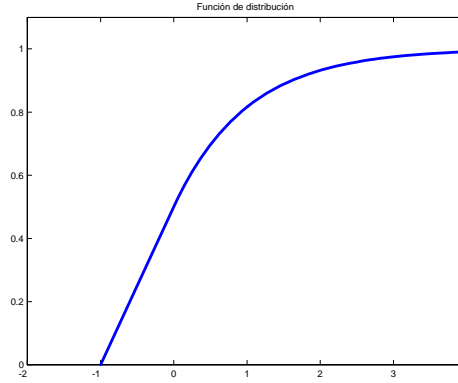
*Tramo 3:*  $x > 0$ . En este tramo se tiene que  $f^{(Tramo\ 3)} = (1/2)e^{-x}$ . Entonces

$$\begin{aligned}F(x) &= P(X \leq x) = P(X < -1) + P(-1 \leq X \leq 0) + P(0 \leq X \leq x) \\ &= \int_{-\infty}^{-1} f^{(Tramo\ 1)}(u)du + \int_{-1}^0 f^{(Tramo\ 2)}(u)du + \int_0^x f^{(Tramo\ 3)}(u)du \\ &= \int_{-\infty}^{-1} 0du + \int_{-1}^0 \frac{1}{2}du + \int_0^x \frac{1}{2}e^{-u}du \\ &= 0 + \frac{1}{2} + \int_0^x \frac{1}{2}e^{-u}du = 1 - \frac{1}{2}e^{-x}; x > 0\end{aligned}$$

En resumen:

$$F(x) = \begin{cases} 0 & x \leq -1 \\ \frac{1}{2}(x+1) & -1 < x \leq 0 \\ 1 - \frac{1}{2}e^{-x} & x > 0 \end{cases}$$

La siguiente figura muestra el dibujo de esta función de distribución.



### 3.4. Medidas características de las variables aleatorias

Una vez que tenemos construida la función de probabilidad o de densidad de una variable aleatoria (o la de distribución), la variable aleatoria queda totalmente definida. Podemos entonces calcular la probabilidad de cualquier suceso que nos interese. En esta sección definiremos medidas numéricas que resuman aspectos importantes de dicha variable aleatoria que nos ayuden a entender mejor su comportamiento. Estas medidas características son las análogas a las definidas en el Tema 1 para un conjunto de datos, sólo que ahora serán un resumen de toda la población. Dichas medidas características se calcularán a partir de las funciones de probabilidad o densidad.

#### 3.4.1. Medidas de centralización

**Media o esperanza matemática,  $\mu$**

Para variables aleatorias discretas con  $K$  valores distintos  $x_1, \dots, x_K$  es

$$E(X) = \sum_{i=1}^K x_i p(x_i) \equiv \mu. \quad (3.4)$$

y es una constante. Esta expresión nos proporcionará el centro de gravedad de la función de probabilidad. Esta expresión es la misma que la definición de media aritmética cuando había datos repetidos, solo que ahora estamos promediando los infinitos valores de la población.

Para variables continuas la esperanza matemática es

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx. \quad (3.5)$$

Nótese que (3.5) es similar a (3.4) en el sentido de sumar cada valor por su peso en la población total. En (3.5) estamos sumando infinitos valores de peso infinitesimal. Se verifica además que

$$E(g(X)) = \begin{cases} \sum_{i=1}^K g(x_i)p(x_i) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{si } X \text{ es continua} \end{cases} \quad (3.6)$$

La esperanza es un operador lineal. Por tanto se verifica que, para dos variables aleatorias  $X$  e  $Y$  se tiene que

$$E(aX + bY) = aE(X) + bE(Y).$$

Es frecuente referirse a la esperanza de una variable aleatoria como la media poblacional, para distinguirla de la media muestral que se obtendría de promediar un conjunto de datos observados de dicha variable.

**Ejemplo 8** Sea  $X$  la variable aleatoria continua definida en el intervalo  $(0,1)$  y de función de densidad  $f(x) = 12x^2(1-x)$ . La esperanza de esta variable aleatoria será

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 12x^3(1-x)dx = 12 \left( \frac{x^4}{4} - \frac{x^5}{5} \right) \Big|_0^1 = \frac{12}{20} = \frac{3}{5}$$

### Moda

Es el valor de mayor probabilidad  $p(x_i)$  o densidad  $f(x)$

### Mediana

Es el valor que deja a cada lado un 50 % de la probabilidad. Es decir, es el valor  $x_m$  tal que

$$F(x_m) = 0,5$$

**Ejemplo 9** A partir de la función de distribución del ejemplo 4, tenemos que

$$F(x_m) = \frac{1}{2} (x_m + 1) = 0,5 \Rightarrow x_m = 0.$$

**Ejemplo 10** Sea  $X$  una variable aleatoria con función de densidad  $f(x) = 2x$ , y definida en el intervalo  $(0,1)$ . La mediana será

$$\int_0^{x_m} 2x dx = 0,5 \Rightarrow \frac{2x^2}{2} \Big|_0^{x_m} = x_m^2 = 0,5 \Rightarrow x_m = \frac{1}{\sqrt{2}}$$

En variables con distribución simétrica, la mediana coincidirá con la media. Al igual que ocurre con las medidas características de un conjunto de datos, en las distribuciones asimétricas, media y mediana no coinciden, estando la media desplazada en la dirección de la cola de la distribución.

## 3.4.2. Medidas de dispersión

### Varianza

El concepto de varianza de una variable aleatoria es similar al de la varianza de un conjunto de datos, solo que extendido a los infinitos elementos que componen la población que representa

dicha variable aleatoria. Para un conjunto de datos, la varianza es el promedio de las desviaciones cuadráticas de cada dato respecto a la media de los datos. Extrapolando esa idea a variables aleatorias, la varianza será la esperanza de las desviaciones cuadráticas a la esperanza, es decir,

$$\text{var}(X) \equiv \sigma^2 = E[(X - \mu)^2]$$

donde  $\mu = E(X)$ . Aplicando (3.6) se tiene que

$$\text{var}(X) \equiv \sigma^2 = \begin{cases} \sum_{i=1}^K (x_i - \mu)^2 p(x_i) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Operando se puede obtener que

$$\sigma^2 = E[(X - \mu)^2] = E(X^2 + \mu^2 - 2X\mu) = E(X^2) + E(\mu^2) - E(2X\mu)$$

Como  $\mu$  es una constante, se tiene que  $E(\mu) = \mu$ . Por tanto

$$\sigma^2 = E(X^2) + \mu^2 - 2\mu^2 = E(X^2) - \mu^2,$$

relación que puede ser útil para calcular  $\sigma^2$ . La raíz cuadrada de la varianza será  $\sigma$  y es la desviación típica de la variable. El coeficiente de variación de una variable aleatoria se define entonces como

$$CV(X) = \frac{\sigma}{|\mu|}.$$

**Ejemplo 11** (continuación) La varianza de la variable aleatoria del Ejemplo 5 mostrada más arriba será

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

Aplicando (3.6) tenemos que

$$E[X^2] = \int_0^1 x^2 f(x) dx = \int_0^1 12x^4(1-x) dx = 12 \left( \frac{x^5}{5} - \frac{x^6}{6} \right) \Big|_0^1 = \frac{12}{30} = \frac{2}{5}$$

por lo que, usando los resultados obtenidos anteriormente para  $E(X)$

$$\text{var}[X] = E[X^2] - (E[X])^2 = \frac{2}{5} - \left( \frac{3}{5} \right)^2 = \frac{1}{25}$$

### Cuartiles

Para calcular los cuartiles aplicamos la misma idea que para calcular la varianza. El cuartil  $Q_1$  será el valor que verifique

$$F(Q_1) = 0,25,$$

mientras que el cuartil 3 verifica

$$F(Q_3) = 0,75,$$

y la distancia  $Q_3 - Q_1$  es el rango intercuartílico.

### 3.4.3. La desigualdad de Chebyshev

La desigualdad de Chebyshev es un resultado sobre la dispersión de los valores de una variable aleatoria alrededor de su media. Dice lo siguiente

**Desigualdad de Chebyshev:** Para cualquier variable  $X$  de media  $\mu$  y varianza  $\sigma^2$  se cumple que

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

y por tanto

$$P(|X - \mu| < c\sigma) > 1 - \frac{1}{c^2}$$

para  $c > 0$ . La desigualdad de Chebyshev sirve para tener unos márgenes alrededor de la media, dentro de los cuales estarán los valores generados por una variable aleatoria con cierta probabilidad. Por ejemplo, si  $\mu = 0$  y  $\sigma^2 = 1$ , se tiene que

$$P(|X| \geq c) \leq \frac{1}{c^2}$$

o bien

$$P(|X| < c) > 1 - \frac{1}{c^2}$$

y la probabilidad de que  $X$  esté dentro del intervalo  $(-2, 2)$  es superior a  $1 - 1/0.25 = 0.75$ .

### 3.4.4. Efecto de las transformaciones lineales en las medidas características

Como vimos anteriormente, la esperanza es un operador lineal, pues es una suma. Verifica entonces que  $E(a + bX) = a + bE(X)$ , que es un resultado similar al que se obtenía para la media de un conjunto de datos. Para dos variables aleatorias  $X$  e  $Y$  se tiene que  $E(aX + bY) = aE(X) + bE(Y)$ . Asimismo, la varianza de una transformación lineal verifica que

$$\begin{aligned} \text{Var}(a + bX) &= E[(a + bX - a - bE(X))^2] = E[(bX - bE(X))^2] \\ &= b^2 \text{Var}(X), \end{aligned}$$

que es similar al obtenido con la varianza de un conjunto de datos.

Enunciaremos el siguiente resultado sin demostrarlo:

**Resultado:** Sean  $X$  e  $Y$  dos variables aleatorias independientes, es decir, que el valor que tome una de ellas no modifica las propiedades de la otra. Entonces

$$E(XY) = E(X)E(Y).$$

Aplicando este resultado tenemos que si dos variables aleatorias  $X$  e  $Y$  son independientes:

$$\begin{aligned}
 \text{var}(aX + bY) &= E \left\{ [(aX + bY) - E(aX + bY)]^2 \right\} = E \left\{ [aX + bY - aE(X) - bE(Y)]^2 \right\} \\
 &= E \left\{ [aX - aE(X) + bY - bE(Y)]^2 \right\} \\
 &= E \left\{ (a[X - E(X)] + b[Y - E(Y)])^2 \right\} \\
 &= E \left\{ a^2 [X - E(X)]^2 + b^2 [Y - E(Y)]^2 + 2ab [X - E(X)][Y - E(Y)] \right\} \\
 &= a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab E[(X - E(X))(Y - E(Y))] \\
 &= a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab E(XY - XE(Y) - YE(X) + E(X)E(Y)).
 \end{aligned}$$

Aplicando que  $E(X)$  y  $E(Y)$  son constantes, y por tanto  $E(E(X)) = E(X)$  y  $E(E(Y)) = E(Y)$ , y el resultado anterior sobre la independencia de  $X$  e  $Y$ , tenemos que

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y).$$

Análogamente, si  $X$  e  $Y$  son dos variables aleatorias independientes

$$\text{var}(aX - bY) = a^2 \text{var}(X) + b^2 \text{var}(Y)$$

Nótese que en ambos casos, suma o resta, el resultado es la suma de las varianzas.

### 3.5. Variables aleatorias multivariantes (avanzado)

#### 3.5.1. Distribución conjunta

En las secciones anteriores hemos definido las variables aleatorias univariantes. Por ejemplo, si  $X$  es una variable aleatoria discreta (número de artículos defectuosos en un lote de  $n$ , o número de llamadas a una centralita en una hora), usaremos la función de probabilidad o la de distribución para definirla. Por el contrario, si es continua, se usará la función de densidad o de distribución.

En esta sección nos ocuparemos de variables aleatorias de más de una dimensión. A este tipo de variables de más de una dimensión les llamaremos variables aleatorias multidimensionales o multivariantes. Por ejemplo, para describir la longitud  $X$  de un cilindro usaremos una variable aleatoria unidimensional que describa la probabilidad de tener determinadas longitudes utilizando su función de densidad  $f(x)$ . Lo mismo podría decirse para describir el diámetro de dichos cilindros: usaríamos una variable aleatoria  $Y$  que describiese los diferentes diámetros que por azar podremos obtener a través de la función de densidad  $f(y)$ . Las variables univariantes sólo nos proporcionan información de una sola variable cada vez, y no sabremos si ambas variables están o no relacionadas. Por ejemplo, sería interesante saber si la producción de cilindros más largos de lo normal está también acompañada de la obtención de radios alejados de la media, o por el contrario ambas dimensiones son independientes. ¿Cuál es la probabilidad de producir cilindros que al mismo tiempo sean más largos que  $l$  cm pero con diámetro menos que  $d$  centímetros?, es decir

$$P[(X > l) \cap (Y < r)].$$

Para calcular esa **probabilidad conjunta** es necesario conocer las propiedades conjuntas de ese par de variables. En el caso de variables discretas será necesario conocer la función de probabilidad conjunta, y en el caso de variables continuas, la función de densidad conjunta.



### Función de probabilidad conjunta

Sean  $X_1, X_2$  dos variables aleatorias discretas que toman los siguientes valores:  $X_1 : \{x_{11}, \dots, x_{1I}\}$ ;  $X_2 : \{x_{21}, \dots, x_{2J}\}$ . Su función de probabilidad conjunta  $p(x_{1i}, x_{2j})$  es la función que nos da la probabilidad de obtener dichos valores, es decir,

$$p(x_{1i}, x_{2j}) = P[(X_1 = x_{1i}) \cap (X_2 = x_{2j})].$$

Esta definición es similar a la distribución de frecuencias conjunta descrita en el capítulo 1. Veamos un ejemplo:

**Ejemplo 12** *Estamos interesados en conocer la probabilidad de que un vehículo tenga los neumáticos en mal estado así como el estado de las luces. Denotaremos por  $X_1$  al 'número de neumáticos en mal estado que tiene un vehículo ( $x_1 = 0, 1, 2, 3, 4$ ). Denotaremos por  $X_2$  al estado de las luces de un vehículo:  $x_2 = 1$  representa que las luces están en perfecto estado, mientras que  $x_2 = 0$  representa a las luces con alguna deficiencia. Después de analizar los ficheros de un conjunto de talleres donde se realiza la Inspección Técnica de Vehículos (ITV) se puede construir la siguiente tabla que representa la función de probabilidad conjunta de las variables  $X_1$  y  $X_2$ .*

$X_2$		$X_1$					Total $X_2$
		0	1	2	3	4	
1		0.425	0.2125	0.102	0.068	0.0425	0.85
0		0.075	0.0375	0.018	0.012	0.0075	0.15
Total $X_1$		0.50	0.25	0.12	0.08	0.05	1.00

Los valores del interior de esta tabla son los valores de la función de probabilidad conjunta, que nos proporcionan la probabilidad de observar simultáneamente cierto valor de  $X_1$  y de  $X_2$ . Lógicamente, para que una función de probabilidad conjunta esté bien definida debe sumar 1, es decir

$$\sum_{i=1}^I \sum_{j=1}^J p(x_{1i}, x_{2j}) = 1$$

**Ejemplo 13** *Normalmente, el resultado del análisis de la característica de un coche en una ITV toma los valores: sin defecto, defecto leve o defecto grave. Sea  $A$  el resultado de analizar el estado de las luces de un vehículo, con resultados  $A_1 = \text{sin defecto}$  y  $P(A = A_1) = p_1$ ;  $A_2 = \text{defecto leve}$  y  $P(A = A_2) = p_2$  y  $A_3 = \text{defecto grave}$ , con  $P(A = A_3) = p_3$  (donde  $p_3 = 1 - p_1 - p_2$ ). Supongamos que el proceso de aparición de defectos leves o graves es estable e independiente. Es decir, la secuencia de valores  $A_i$ ,  $i = 1, 2, 3$ , que observamos cuando analizamos una sucesión de vehículos no muestra ninguna tendencia. Supongamos que analizamos  $n$  vehículos y definimos la siguiente variable aleatoria*

$$X_i = \text{número de vehículos con luces en estado } A_i, i = 1, 2, 3.$$

Entonces el vector  $\mathbf{X} = (X_1, X_2, X_3)$  es una variable aleatoria tridimensional). Puede demostrarse que su función de probabilidad es

$$P(X_1 = r_1, X_2 = r_2, X_3 = r_3) = \frac{n!}{r_1! r_2! r_3!} p_1^{r_1} p_2^{r_2} p_3^{r_3},$$

donde  $r_1 + r_2 + r_3 = n$  y  $p_1 + p_2 + p_3 = 1$ . Esta función nos dará la probabilidad de cada combinación de valores  $r_1, r_2, r_3$ .

### Función de densidad conjunta

Si las variables aleatorias son continuas necesitaremos de una función de densidad conjunta que nos dé la densidad de cada punto de dimensión multivariante.

**Ejemplo 14** *Un producto químico está formado por dos componentes, que aparecen en cantidades variables. Llamaremos  $X$  a la cantidad de miligramos que un compuesto posee del producto 1, y llamaremos  $Y$  a la cantidad de miligramos del producto 2. La función de densidad conjunta viene expresada por la siguiente función*

$$f(x, y) = \begin{cases} kxy, & \text{si } 0 < x < 1; 0 < y < 1; \\ 0 & \text{en el resto} \end{cases}$$

Podemos entonces calcular la probabilidad de que exista menos de 0.5 gramos de cada producto:

$$P(X < 0,5, Y < 0,5) = \int_0^{0,5} \int_0^{0,5} f(x, y) dx dy.$$

Como no conocemos la constante  $k$  habrá que calcularla usando la información de que la probabilidad total será 1. Por tanto

$$\begin{aligned} \int_0^1 \int_0^1 f(x, y) dx dy &= 1 \\ \int_0^1 \int_0^1 kxy dx dy &= k \int_0^1 x \left[ \frac{y^2}{2} \right]_0^1 dx = \frac{k}{2} \int_0^1 x dx = \frac{k}{2} \left[ \frac{x^2}{2} \right]_0^1 \\ &= \frac{k}{4} = 1 \implies k = 4. \end{aligned}$$

Entonces la probabilidad que necesitamos es

$$P(X < 0,5, Y < 0,5) = \int_0^{0,5} \int_0^{0,5} 4xy dx dy = 4 \int_0^{0,5} x dx \int_0^{0,5} y dy = 4 \left[ \frac{x^2}{2} \right]_0^{0,5} \left[ \frac{y^2}{2} \right]_0^{0,5} = 0,0625$$

### 3.5.2. Distribución marginal

La distribución marginal es simplemente la distribución univariante que se obtiene a partir de la función de probabilidad o densidad conjunta. Se obtiene sumando todos los valores de las variables que no nos interesan. En el caso de la tabla anterior del Ejemplo 12, la distribución marginal de  $X_1$  es la fila inferior, y la marginal de  $X_2$  es la columna de la derecha. Es decir,

$$p(x_{1i}) = P(X_1 = x_{1i}) = \sum_{j=1}^J p(x_{1i}, x_{2j}).$$

Análogamente, para variables continuas, se obtendrán las distribuciones marginales integrando la función de densidad conjunta aquellas variables que no nos interesan. Por ejemplo, la distribución marginal de  $X$ , cuando tenemos  $f(x, y)$  será

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

**Ejemplo 15** En el caso del Ejemplo 14 se tienen las siguientes distribuciones marginales

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 4xy dy = 4x \int_0^1 y dy = 2x \left[ \frac{y^2}{2} \right]_0^1 = 2x, \\ f(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 4xy dx = 4y \int_0^1 x dx = 2y \left[ \frac{x^2}{2} \right]_0^1 = 2y. \end{aligned}$$

### 3.5.3. Distribución condicionada e independencia

Si en el Ejemplo 12 queremos calcular la función de probabilidad de que un vehículo que no tiene ningún neumático defectuoso ( $X_1 = 0$ ) tenga las luces en mal estado ( $X_2 = 0$ ) tendremos que calcular una probabilidad condicionada  $P(X_2 = 0 | X_1 = 0)$ . Usando la fórmula de la probabilidad condicionada se tendrá que

$$P(X_2 = 0 | X_1 = 0) = \frac{P[(X_2 = 0) \cap (X_1 = 0)]}{P(X_1 = 0)},$$

y en general se tendrá que la función de probabilidad condicionada será

$$p(x_{1i} | x_{2j}) = \frac{p(x_{1i}, x_{2j})}{p(x_{2j})}. \quad (3.7)$$

La extensión de este resultado al caso continuo nos proporciona la siguiente función de densidad condicionada

$$f(x|y) = \frac{f(x, y)}{f(y)}. \quad (3.8)$$

**Ejemplo 16** Usando el Ejemplo 14, la probabilidad de que un producto tenga menos de 0.3 miligramos del primer componente cuando del segundo tiene 0.8 miligramos será

$$P(X < 0,3 | Y = 0,8) = \int_0^{0,3} f(x|y = 0,8) dx$$

donde

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{4xy}{2y} = 2x.$$

Por tanto

$$P(X < 0,3 | Y = 0,8) = \int_0^{0,3} f(x|y = 0,8) dx = \int_0^{0,3} 2x dx = 2 \left[ \frac{x^2}{2} \right]_0^{0,3} = 0,09$$

Si las variables aleatorias implicadas son independientes entre si significará que el conocimiento de una de ellas no alterará la incertidumbre que se tenga sobre la otra. Por tanto

$$\begin{aligned} p(x_{1i} | x_{2j}) &= p(x_{1j}), \\ f(x|y) &= f(x). \end{aligned}$$

Por tanto, usando (3.7) y (3.8) se tiene que las variables aleatorias independientes verificarán

$$\begin{aligned} p(x_{1i}, x_{2j}) &= p(x_{1j})p(x_{2j}) \\ f(x, y) &= f(x)f(y), \end{aligned}$$

siendo este resultado válido para cualquier dimensión. Es decir, si un conjunto de  $K$  variables aleatorias son independientes su representación conjunta verifica

$$p(x_1, x_2, \dots, x_K) = \prod_{k=1}^K p(x_k),$$

$$f(x_1, x_2, \dots, x_K) = \prod_{k=1}^K f(x_k).$$

Este resultado nos dice que si un conjunto de variables aleatorias es independiente podemos reconstruir la probabilidad conjunta a partir de la información univariante. Este resultado es del tipo si y sólo si, por lo que suele también usarse como definición de independencia.

**Ejemplo 17** *En el caso del Ejemplo 14 se tiene que*

$$f(x)f(y) = 2x2y = 4xy = f(xy),$$

*por lo tanto ambas variables aleatorias son independientes. Este resultado ya se conocía pues también vimos que  $f(x|y) = 2x = f(x)$ .*

### 3.5.4. Covarianza y correlación

La covarianza entre la variable aleatoria  $X$  y la variable aleatoria  $Y$  es una medida de su dependencia lineal. Se define de la siguiente manera:

$$\begin{aligned} \text{cov}(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

La covarianza tiene la misma interpretación que una covarianza muestral. Si entre  $X$  e  $Y$  hay algún tipo de dependencia lineal positiva, entonces  $\text{cov} > 0$ . Por el contrario, si entre  $X$  e  $Y$  hay algún tipo de dependencia lineal negativa,  $\text{cov} < 0$ . Si no hay ningún tipo de dependencia lineal  $\text{cov} = 0$ . En este caso se dice que ambas variables están incorreladas. Nótese que si  $X$  e  $Y$  son independientes estarán incorreladas, mientras que lo contrario no necesariamente tiene que ser cierto. En el caso de incorrelación se tendrá que

$$0 = E(XY) - E(X)E(Y) \Rightarrow E(XY) = E(X)E(Y).$$

El coeficiente de correlación se define como

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}},$$

y al igual que se expuso para la correlación muestral, es un coeficiente que estará entre -1 y 1, lo que introduce restricciones entre los posibles valores de varianzas y covarianzas. Si la relación entre  $X$  e  $Y$  es del tipo  $Y = aX + b$  (o de forma  $X = aY + b$ ), se dice que la relación lineal es exacta. Se puede comprobar que en ese caso el coeficiente de correlación es, en valor absoluto, igual a la unidad.