6. Variables aleatorias

6.1. Variables aleatorias

Cuando en el tema anterior se definió la noción de probabilidad, el espacio muestral podía ser cualquier conjunto. Así, el espacio muestral del lanzamiento de un dado es $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$, mientras que el espacio muestral del lanzamiento de una moneda es $\Omega_2 = \{cara, cruz\}$.

En principio, no hay ninguna diferencia a la hora de definir y calcular probabilidades, de la misma forma que se pueden calcular frecuencias para cualquier tipo de variable estadística, sea cualitativa o cuantitativa. Sin embargo, si queremos extender los conceptos de media, varianza, etc. que se habían definido en la parte de estadística descriptiva para variables cuantitativas, es necesario que los posibles resultados del experimento sean números. Las variables aleatorias son el equivalente para la probabilidad de las variables cuantitativas para la estadística descriptiva.

Que el espacio muestral sea un conjunto de números es muy habitual, pues muchos experimentos aleatorios dan lugar a resultados numéricos. Así, podemos medir la altura, el número de hermanos, ... que son experimentos que dan lugar a resultados numéricos. Por otra parte, en muchos experimentos aleatorios nos interesa, más que el resultado del experimento, una función real de los resultados.

Ejemplo 97.

Supongamos que tenemos el experimento en el que se lanza una moneda y un dado. Entonces tenemos 12 resultados posibles, y el espacio muestral es

$$\Omega = \{(cara, i), (cruz, i) : i = 1, ..., 6\}.$$

https://dx.doi.org/10.5209/docm.003.06

Ahora bien, supongamos que recibiremos un pago de 5 euros si sale cara más el resultado del dado. En este caso, lo que realmente nos interesa es el premio (que es un valor numérico) y no el resultado del experimento. Así, ganamos 6 euros si sale el resultado $(cruz, 6)$ *o si sale el resultado* $(cara, 1)$, *pero en la práctica no nos preocupa cuál de los dos ha ocurrido, sino que hemos ganado 6 euros.*

La definición de variable aleatoria intenta unir estas dos formas de obtener resultados numéricos en un experimento aleatorio. Una **variable aleatoria** X es una aplicación que a cualquier elemento del espacio muestral le asigna un valor numérico. Matemáticamente, esto se escribe

$$X : \Omega \to \mathbb{R}.$$

Ejemplo 98.

Consideremos el experimento que consiste en lanzar dos monedas. Nuestro espacio muestral es

$$\Omega = \{(cara, cruz), (cruz, cara), (cara, cara), (cruz, cruz)\}.$$

Una posible variable aleatoria para este experimento sería $X \equiv$ *número de caras, que viene definida por*

$$\begin{array}{ll} X((cara, cruz)) = 1, & X((cruz, cara)) = 1, \\ X((cruz, cruz)) = 0, & X((cara, cara)) = 2. \end{array}$$

Nótese que podríamos definir muchas otras variables aleatorias a partir de este experimento aleatorio; por ejemplo, otra posible variable aleatoria Y *sería la definida por*

$$\begin{array}{ll} Y((cara, cruz)) = 10, & Y((cruz, cara)) = 1, \\ Y((cruz, cruz)) = 0, & Y((cara, cara)) = 11. \end{array}$$

La variable aleatoria que nos interesa en cada caso viene determinada por lo que queremos hallar.

En realidad, una variable aleatoria nos permite pasar de unos valores cualesquiera a unos nuevos valores que ahora son numéricos. Sobre el nuevo espacio muestral que nos proporciona la variable aleatoria, que ahora es el conjunto de los números reales $\mathbb{R}$, podemos plantearnos calcular probabilidades de sucesos, que serán subconjuntos de $\mathbb{R}$. Esta probabilidad, que denotaremos por P, se construirá a partir de la probabilidad P^* que teníamos sobre Ω y que nos proporciona el experimento inicial, que no era necesariamente numérico. Dado $A \subseteq \mathbb{R}$ un suceso del nuevo espacio muestral, su probabilidad será:

$$P(A) = P^*(\{\omega \in \Omega \text{ tal que } X(\omega) \in A\}).$$

Esto nos da un nuevo sistema de probabilidades sobre los números reales y nos permite olvidar el experimento inicial del que derivan. Esta nueva probabilidad P se llama **probabilidad inducida** por la variable aleatoria X.

Ejemplo 99. *(Continuación del ejemplo 98)*

La probabilidad asociada al experimento que consiste en lanzar dos monedas se puede obtener a partir de la regla de Laplace. Por lo tanto, se tiene

$$\begin{array}{ll} P^*(cruz, cruz) = \frac{1}{4}, & P^*(cruz, cara) = \frac{1}{4}, \\ P^*(cara, cruz) = \frac{1}{4}, & P^*(cara, cara) = \frac{1}{4}. \end{array}$$

Para la variable X definida anteriormente, se tiene:

$$P(X = 0) = P^*((cruz, cruz)) = \frac{1}{4},\ P(X = 2) = P^*((cara, cara)) = \frac{1}{4}.$$

Sin embargo,

$$P(X = 1) = P^*(\{(cara, cruz), (cruz, cara)\}) = \frac{2}{4}.$$

Y también,

$$P(X < 1) = P^*(\{(cruz, cruz), (cara, cruz), (cruz, cara)\}) = \frac{3}{4}.$$

Para la variable Y se tiene por ejemplo:

$$P(Y = 10) = P^*((cara, cruz)) = \frac{1}{4},\ P(Y = 1) = P^*((cruz, cara)) = \frac{1}{4},$$
$$P(Y = 0) = P^*((cruz, cruz)) = \frac{1}{4},\ P(Y = 11) = P^*((cara, cara)) = \frac{1}{4}.$$

6.2. Función de distribución

Nótese que para determinar una probabilidad es necesario conocer la probabilidad de cada suceso. Esto nos obligaba a definir los valores $P^*(A), \forall A \subseteq \Omega$. Al definir una variable aleatoria, el nuevo espacio muestral es $\mathbb{R}$; por tanto, para definir la probabilidad deberíamos definir la probabilidad $P(B)$, $\forall B \subseteq \mathbb{R}$[1]. El número de subconjuntos de $\mathbb{R}$ es infinito. Esto hace imposible escribir la probabilidad de cualquier suceso por enumeración, es decir, haciendo una lista. Sin embargo, es posible dar una definición equivalente de la probabilidad a partir de una función real y que nos permitirá calcular la probabilidad de cualquier suceso. Esta función real se llama la función de distribución y es un equivalente de la frecuencia relativa acumulada en estadística descriptiva.

Dada una variable aleatoria X, la función $F : \mathbb{R} \to \mathbb{R}$ definida por

$$\boxed{F(x) = P^*(\{\omega \in \Omega \text{ tal que } \mathrm{X}(\omega) \leq \mathrm{x}\}),}$$

se denomina **función de distribución**.

En general, denotaremos la expresión anterior de manera abreviada por

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

Ejemplo 100. *(Continuación del ejemplo 98)*

Consideremos el ejemplo del lanzamiento de dos monedas y la variable aleatoria $X \equiv$ número de caras obtenido. Ya habíamos hallado anteriormente la probabilidad de cada uno de los posibles valores (0, 1, 2) de la variable X. En este caso, la función de distribución viene dada por

$$F(x) = \begin{cases} 0 & \text{si x} < 0 \\ \frac{1}{4} & \text{si } 0 \leq \text{x} < 1 \\ \frac{3}{4} & \text{si } 1 \leq \text{x} < 2 \\ 1 & \text{si } 2 \leq \text{x} \end{cases}$$

[1] Aunque no es el objetivo de este curso y no nos dará ningún problema en la práctica, desde un punto de vista matemático, cuando el referencial es $\mathbb{R}$, no podemos considerar como conjunto de sucesos el conjunto de todos los subconjuntos de $\mathbb{R}$. En ese caso debemos considerar como conjunto de sucesos lo que se conoce como σ-álgebra de Borel, que se denota por $\mathfrak{B}_{\mathbb{R}}$. Dentro de este conjunto están todos los intervalos, abiertos, cerrados y semiabiertos, los puntos y prácticamente cualquier conjunto de los números reales que nos podamos imaginar.

La gráfica de esta función de distribución viene dada en la figura 6.1.

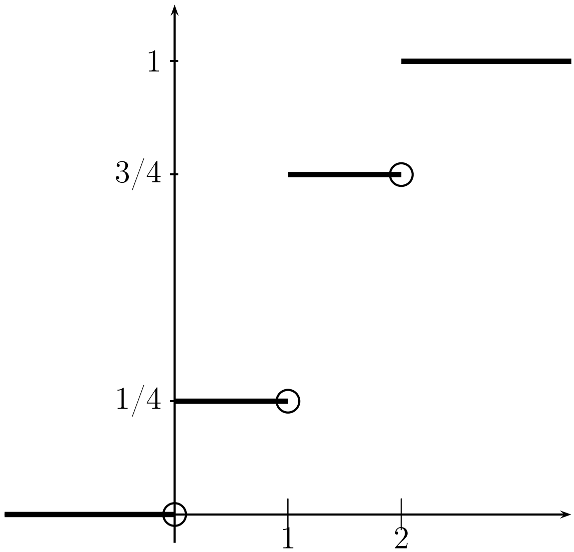

Figura 6.1. Representación gráfica de la función de distribución correspondiente a la variable aleatoria del ejemplo 98.

Por ejemplo,

$$F(1{,}4) = P^*(\{\omega \in \Omega | X(\omega) \leq 1{,}4\}) = P^*(\{(C, X), (X, C), (X, X)\}) = \frac{3}{4}.$$

Es interesante notar la posición de $\leq$ *y* $<$ *en cada una de las partes. Esta posición no es un azar, sino que siempre es así. Nótese que en los puntos de discontinuidad el valor de la función se toma en el trozo de arriba. Esto es lo que representan los círculos en los segmentos inferiores.*

A partir del gráfico anterior puede verse que la función de distribución de una variable aleatoria tiene las siguientes propiedades:

1. F es no decreciente.
2. $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. F es continua por la derecha.

Sin embargo, aunque en este ejemplo la función de distribución es escalonada, esto no es necesario en todos los casos, como veremos más adelante.

Estas cuatro propiedades caracterizan la función de distribución, es decir, dada cualquier función F que verifique estas propiedades, podremos encontrar una variable aleatoria cuya función de distribución sea F; y viceversa, cualquier variable aleatoria tiene una función de distribución que cumple las condiciones anteriores.

A partir de F podemos calcular la probabilidad de cualquier subconjunto real, de la misma forma que con la frecuecia relativa acumulada se podían calcular las frecuencias relativas. Por ejemplo,

$$P(X \in (a, b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

Ejemplo 101. *(Continuación del ejemplo 98)*
En este caso,

$$P(X \in (1{,}5, 3{,}5]) = F(3{,}5) - F(1{,}5) = 1 - \frac{3}{4} = \frac{1}{4}.$$

Finalmente, nótese que

$$P(X < a) = \lim_{x\to a^-} F(x),$$

valor que se denota por $F(a^-)$. Con esto ya podemos hallar la probabilidad de cualquier intervalo a partir de la función de distribución.

Parece que la función de distribución nos complica la teoría, pues en el ejemplo de las monedas nos basta con dar tres valores para conocer la probabilidad de cualquier intervalo. Así, teniendo en cuenta que

$$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{2}{4}, P(X = 2) = \frac{1}{4},$$

podemos calcular por ejemplo que

$$P(X \in (0{,}5, 3{,}5]) = P(X = 1) + P(X = 2) = \frac{3}{4},$$

ya que X solo está dentro de ese intervalo si toma los valores 1 o 2. Todos los demás valores del intervalo son imposibles por la definición de la variable X y entonces no los tenemos en cuenta. Y ahora podemos aplicar las propiedades de la probabilidad para calcular el resultado final.

Sin embargo, como ya hemos indicado, en general deberíamos dar la probabilidad de cualquier suceso, y con experimentos que dan lugar a infinitos resultados podemos tener problemas al aplicar el procedimiento anterior.

En definitiva, la probabilidad es una función de conjunto y, por tanto, difícil de manejar. La función de distribución es, sin embargo, una función real, que es susceptible de ser representada[2]. En otras palabras, la función de distribución simplifica la expresión de la probabilidad. A pesar de ello, como nosotros trataremos solo con tipos especiales de variables aleatorias, no haremos mucho uso de la función de distribución.

6.3. Variables discretas y continuas

Las variables aleatorias se dividen en dos grupos: discretas y continuas (nótese la similitud con la clasificación de las variables estadísticas cuantitativas). Sin embargo, aunque no trataremos ese caso, esta clasificación no es exhaustiva, es decir, pueden construirse variables que no sean ni discretas ni continuas.

[2]Sin entrar en mucho detalle pues no es el objetivo del curso, el cardinal o número de elementos que tiene un conjunto puede ser finito o infinito. En el caso finito tenemos un número natural que nos da el número de elementos del conjunto. En el caso infinito, hay varias clasificaciones en función de lo «grande» que es el infinito. Desde un punto de vista matemático, se puede ver que el cardinal de los subconjuntos de $\mathbb{R}$ es más grande que el cardinal de $\mathbb{R}$, aunque ambos son infinitos.

6.3.1. Variables aleatorias discretas

Una variable se dice **discreta** cuando solo toma un número finito o infinito numerable[3] de valores.

El conjunto de valores que toma la variable se llama **soporte** de la variable aleatoria y se denota por Ω. El soporte jugará el mismo papel que el conjunto de modalidades en estadística descriptiva.

Ejemplo 102. *(Continuación del ejemplo 98)*

El ejemplo del lanzamiento de las dos monedas contando el número de caras que aparecen es una variable discreta cuyo soporte es $\Omega = \{0, 1, 2\}$.

Ejemplo 103.

El número de veces que es necesario lanzar una moneda hasta obtener por primera vez cara es una variable discreta que toma infinitos valores. En concreto, su soporte es

$$\Omega = \{1, 2, 3, ...\}.$$

Veremos más adelante otros ejemplos de variables discretas que toman infinitos valores.

Entonces, tal y como se vio con la función de distribución, esta es una función escalonada con saltos en los valores del soporte de la variable y con valor de salto el valor de probabilidad del punto correspondiente. Por lo tanto, para conocer (es decir, para poder dibujar) la función de distribución solo se necesita conocer los puntos del soporte (donde F es discontinua) y los valores de probabilidad en esos puntos (que son la altura del salto en ese valor).

Entendemos por **función masa de probabilidad** la función

$$P : \Omega \to [0, 1],$$

que a cada punto del soporte le asigna el valor de su probabilidad

$$P(\{\omega \in \Omega | X(\omega) = x\}).$$

[3]En el caso de cardinales infinitos, hay varias clasificaciones en función de lo «grande» que es el infinito, aunque la más habitual es dividir los cardinales en numerables y no numerables o continuos. Los infinitos numerables son los cardinales de los números naturales, enteros o racionales entre otros. Los cardinales no numerables son por ejemplo el de los números reales.

Denotaremos esta probabilidad de forma abreviada por $P(X = x)$. La función masa de proabababilidad es el equivalente de la frecuencia relativa en estadística descriptiva.

Nótese que cualquier función masa de probabilidad verifica las siguientes propiedades:

- $P(x) > 0, \forall x \in \Omega$.
- $\displaystyle\sum_{x\in\Omega} P(x) = 1$.

Estas dos propiedades caracterizan las funciones masa de probabilidad. Nótese que son las mismas propiedades que teníamos en estadística descriptiva para frecuencias relativas. Y en la práctica, la función masa de probabilidad se aplica igual que se haría con las frecuencias relativas.

Como habíamos dicho antes, conocida la función masa de probabilidad se puede conocer la probabilidad de cualquier suceso A sin más que aplicar la fórmula

$$\boxed{P(X \in A) = \sum_{x\in A} P(X = x),}$$

que proviene de la propiedad de aditividad de la probabilidad. En particular, la función de distribución se puede calcular a partir de P de la siguiente manera:

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

Ejemplo 104. *(Continuación del ejemplo 98)*

Consideremos nuevamente el ejemplo del lanzamiento de dos monedas y la variable $X \equiv$ número de caras. Ya habíamos visto anteriormente que X es una variable discreta pues solo toma los valores $\{0, 1, 2\}$. La función masa de probabilidad viene dada por:

$$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{2}{4}, P(X = 2) = \frac{1}{4}.$$

De esta forma se tiene, por ejemplo,

$$P(X \in \{0,1\}) = P(X=0) + P(X=1) = \frac{3}{4},$$
$$P(X \geq 1) = P(X=1) + P(X=2) = \frac{3}{4}.$$

6.3.2. Variables aleatorias continuas

Básicamente, una variable aleatoria **continua** es la que toma una cantidad no numerable de valores[4]. Dicho de otra manera, es una variable que no toma valores aislados, o centrándonos más en nuestro caso, aquellas variables que pueden tomar cualquier valor dentro de un intervalo. Por ejemplo, la altura de los humanos adultos puede tomar cualquier valor en [1.5, 2.5] y es, por tanto, una variable continua; el hecho de que aproximemos las alturas a dos cifras decimales no quiere decir que tome un número finito de valores, sino que se está aproximando; en realidad, todos los valores serían posibles si la precisión fuese suficientemente grande.

Desde un punto de vista matemático, la definición es bastante más abstracta: se dice que una variable aleatoria X con función de distribución F es **absolutamente continua** (o simplemente continua) si existe una función no negativa $f : \mathbb{R} \to \mathbb{R}$ tal que para todo número real x, se cumple que

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

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

Conocida F, es posible calcular f mediante el teorema fundamental del cálculo integral, que establece que

$$F'(x) = f(x).$$

Nótese que por las propiedades de la función de distribución, se tiene

$$1 = \lim_{x\to\infty} F(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{\infty} f(t)dt.$$

[4]Esta definición de variable continua no es exacta, pero intuitivamente es más sencilla de comprender. Más adelante en esta sección daremos la definición matemática.

Esta propiedad, junto a la no negatividad de f, serán las propiedades que nos servirán para comprobar que una función f es función de densidad de una variable aleatoria. Es decir, para comprobar que una función f es función de densidad tiene que cumplirse

- $f(x) \geq 0$, $\forall x \in \mathbb{R}$.
- $\displaystyle\int_{\mathbb{R}} f(x)dx = 1$.

Una diferencia con la función masa de probabilidad es que la función de densidad puede superar el valor 1 en algunos valores reales, lo que no es posible para la probabilidad.

La función de densidad intenta reproducir las propiedades de la función masa de probabilidad para cardinales no numerables. Tiene prácticamente las mismas propiedades, pero ahora sustituimos la suma por la integral. La función de densidad nos indica lo lógico que es que la variable tome un valor (nótese que no hablamos de la probabilidad que toma ese valor); así, si un valor es más lógico que otro, la función de densidad sobre el primero será mayor que sobre el segundo.

Ejemplo 105.

Sea X la variable aleatoria que consiste en escoger un número al azar en el intervalo $(0,1)$. En este caso, X puede tomar cualquier valor en el intervalo $(0,1)$ y es una variable continua. Veamos cuál es la función de densidad de esta variable. Puesto que escogemos un punto al azar, no tenemos ninguna información que haga más probable un punto sobre otro. Esto se debe traducir en que $f(t) = k$, donde k es una constante k sobre todo el intervalo $(0,1)$. Por otra parte, los valores fuera de este intervalo son imposibles, por lo que $f(t) = 0$ fuera de $(0,1)$. Como además $\int_0^1 f(t)dt = 1$ y

$$\int_0^1 kdt = k,$$

se tiene $k = 1$, y la función de densidad es

$$f(t) = \begin{cases} 1 & t \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$

La gráfica de esta función de densidad viene dada en la figura 6.2.

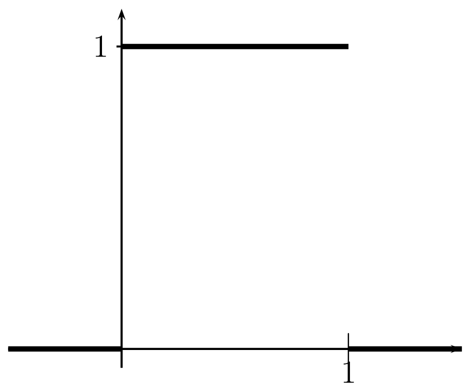

Figura 6.2. Representación gráfica de la función de densidad correspondiente al ejemplo 105.

Si ahora consideramos un intervalo $(a, b]$, se tiene que

$$\begin{aligned} P(X \in (a,b]) &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \\ &= \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_{a}^{b} f(t)dt. \end{aligned}$$

Esto puede extenderse a cualquier suceso A :

$$\boxed{P(X \in A) = \int_A f(t)dt.}$$

Por tanto, aplicando las propiedades de la integral, la probabilidad de un suceso es el área bajo la curva f y limitada por el eje de abscisas.

Como consecuencia de este resultado, si conocemos la función de densidad de una variable continua estaremos en condiciones de hallar la probabilidad de cualquier suceso.

Ejemplo 106. *(Continuación del ejemplo 105)*
Ahora, se tiene por ejemplo:

$$P(X \in \left(\frac{1}{2}, \frac{3}{4}\right)) = \int_{\frac{1}{2}}^{\frac{3}{4}} 1dt = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}.$$

Nótese que para una variable continua, la probabilidad

$$P(X = x) = \int_x^x f(t)dt = 0.$$

Por tanto, la probabilidad de que se tome un valor aislado es 0, a pesar de que este valor es posible. Esto significa que aunque $P(\emptyset) = 0$, pueden existir otros sucesos con probabilidad 0 pero posibles. Este resultado parece que va en contra de toda lógica y que tergiversa completamente el sentido de la probabilidad. En realidad, puede justificarse de la siguiente manera: aunque x sea una valor posible de la variable, si la variable es continua, hay infinitos valores posibles, y como la probabilidad de todos ellos juntos es 1, al repartir punto a punto, esto hace que toquen a probabilidad 0 cada uno de ellos. En realidad, F es como una rampa y el salto entre dos «puntos consecutivos» de la rampa es 0.

Ejemplo 107. *(Continuación del ejemplo 105)*

Para este ejemplo, la figura 6.3 muestra la función de distribución de la variable aleatoria.

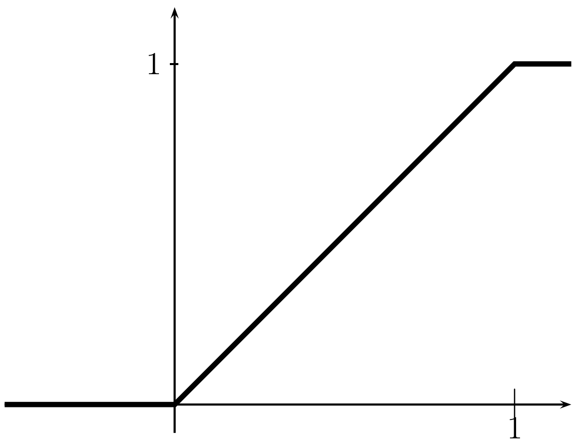

Figura 6.3. Representación gráfica de la función de distribución correspondiente a la variable aleatoria del ejemplo 105.

Como consecuencia del resultado anterior, se tiene que para variables continuas

$$P((a,b]) = P([a,b]) = P((a,b)) = P([a,b)).$$

Nótese que esto no es cierto para las variables discretas.

A modo de conclusión, incluimos en la tabla 6.1 una comparativa de las propiedades de la función masa de probabilidad y de la función de densidad. Como se ha dicho anteriormente y puede verse en esa tabla, la función masa de probabilidad y la función de densidad funcionan de forma muy similar, y la diferencia entre ellas radica en que en las expresiones para la función masa de probabilidad aparecen sumas mientras que para las correspondientes expresiones de la función de densidad aparecen integrales.

	Discretas	Continuas
Condiciones	$P(x_i) > 0$ $\sum_{x\in\Omega} P(x) = 1$	$f(x) \geq 0$ $\int_{\mathbb{R}} f(x)dx = 1$
$F(x)$	$F(x) = \sum_{x_i \leq x} P(X = x_i)$	$F(x) = \int_{-\infty}^{x} f(t)dt$
$P(A)$	$P(X \in A) = \sum_{x\in A} P(X = x)$	$P(X \in A) = \int_A f(t)dt$

Tabla 6.1. Comparativa de las propiedades de la función masa de probabilidad y de la función de densidad: propiedades de caracterización, cálculo de la función de distribución y cálculo de probabilidades.

6.4. Esperanza y varianza de una variable aleatoria

Como se dijo en la introducción de este tema, el uso de variables aleatorias nos permite pasar de espacios muestrales generales a espacios muestrales reales, y esto permite definir muchos de los conceptos de estadística descriptiva. Recuérdese que en la parte dedicada a la estadística descriptiva habíamos estudiado distintas medidas que nos daban información del comportamiento de la variable; sin embargo, para casi todas ellas era necesario que la variable fuese cuantitativa. Entre todas las medidas estudiadas, habíamos visto que las más importantes eran la media como medida de centralización y la varianza (y desviación típica) como medida de dispersión. Pasemos ahora a ver las expresiones de la media y la varianza de una variable aleatoria, aunque este proceso se puede repetir para extender cualquiera de las medidas que se definieron en el capítulo 3. Haremos el estudio por separado para variables discretas y continuas.

6.4.1. Esperanza matemática

Dada una variable aleatoria discreta X con función masa de probabilidad P se define su **media** o **esperanza matemática** como el valor (si existe),

$$\boxed{\mu = E(X) = \begin{cases} \displaystyle\sum_{i=1}^{k} x_i P(X = x_i) & \text{si Soporte(X)} = \{\mathrm{x_1, ..., x_k}\} \\ \displaystyle\sum_{i=1}^{\infty} x_i P(X = x_i) & \text{si Soporte(X)} = \{\mathrm{x_1, ..., x_n, ...}\} \end{cases}}$$

En realidad, esta expresión recuerda mucho a la expresión de la media en variables estadísticas, especialmente si la variable solo toma un número finito de valores. La diferencia radica en cambiar la frecuencia relativa por la probabilidad. Y esto es lógico, puesto que la probabilidad puede verse como el límite de las frecuencias relativas cuando el tamaño de la muestra tiende a infinito. Es por ello que se llama esperanza, pues es lo que se *espera* que pase si pudiésemos realizar el experimento infinitas veces. El caso no finito no es más que la extensión natural del caso finito cuando el soporte tiene cardinal infinito.

La **media** de una variable aleatoria continua X con función de densidad f viene dada por

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

Esta expresión es la misma que para el caso discreto, sin más que sustituir la suma por la integral y la función masa de probabilidad por la función de densidad.

Ejemplo 108. *(Continuación del ejemplo 98)*

Sea la variable aleatoria del lanzamiento de las dos monedas. En este caso,

$$E(X) = 0\frac{1}{4} + 1\frac{1}{2} + 2\frac{1}{4} = 1.$$

Ejemplo 109. *(Continuación del ejemplo 105)*

Sea la variable aleatoria continua considerada en la sección anterior. En este caso,

$$E(X) = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

La esperanza tiene la misma interpretación que la media aritmética en estadística descriptiva, esto es, un valor en mitad de los valores posibles que compensa diferencias por encima y por debajo. Así, en el caso de la variable continua, la esperanza es $\frac{1}{2}$, pues en caso de que todos los valores sean igualmente lógicos, el valor medio debería estar en la mitad del intervalo.

Cabe mencionar, finalmente, que la esperanza de una variable aleatoria tiene las mismas propiedades que habíamos visto para la media aritmética.

6.4.2. Varianza

La **varianza** de una variable aleatoria discreta X con función masa de probabilidad P viene dada por

$$\boxed{V(X) = \sigma^2 = E((X-\mu)^2) = \sum_{x_i \in Soporte(X)} (x_i - \mu)^2 P(X = x_i).}$$

La **varianza** de una variable aleatoria continua viene dada por

$$\boxed{V(X) = \sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.}$$

En ambos casos y al igual que para variables estadísticas, la varianza puede ser calculada por

$$\boxed{\sigma^2 = E(X^2) - E(X)^2,}$$

donde

$$E(X^2) = \begin{cases} \displaystyle\sum_{x \in Soporte(X)} x^2 P(X=x) & \text{si X discreta} \\ \displaystyle\int_{-\infty}^{\infty} x^2 f(x)dx & \text{si X continua} \end{cases}$$

Ejemplo 110. *(Continuación del ejemplo 98)*
Para nuestra variable discreta,

$$E(X^2) = \sum_{x_i} x_i^2 P(X = x_i) = 0\frac{1}{4} + 1\frac{1}{2} + 4\frac{1}{4} = 1{,}5.$$

Luego,

$$V(X) = \sigma^2 = E(X^2) - E(X)^2 = 0{,}5.$$

Ejemplo 111. *(Continuación del ejemplo 105)*
Para nuestra variable continua,

$$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3}.$$

Luego,

$$V(X) = \sigma^2 = E(X^2) - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Finalmente, dada una variable aleatoria X, se define la **desviación típica** (denotada por σ o $D(X)$) como la raíz cuadrada positiva de la varianza. Esta medida aparece para tener una medida de la dispersión que tenga las mismas unidades que X. Las propiedades que habíamos visto para la varianza y la desviación típica para variables estadísticas se cumplen también en el caso de tratar con variables aleatorias.

En la tabla 6.2, se ve una comparativa de las distintas fórmulas para cada situación. Nótese nuevamente el comportamiento similar de la frecuencia relativa en estadística descriptiva, la función masa de probabilidad para variables discretas y la función de densidad para variables continuas.

	$\overline{X}$ o $E(X)$	$\overline{X^2}$ o $E(X^2)$
Descriptiva	$\sum_{i=1}^{k} x_i f_i$	$\sum_{i=1}^{k} x_i^2 f_i$
Discretas	$\sum_{x \in Sop(X)} xP(x)$	$\sum_{x \in Sop(X)} x^2 P(x)$
Continuas	$\int_{-\infty}^{\infty} xf(x)dx$	$\int_{-\infty}^{\infty} x^2 f(x)dx$
	$Var(X)$	$Var(X)$
Descriptiva	$\sum_{i=1}^{k} (x_i - \overline{X})^2 f_i$	$\overline{X^2} - \overline{X}^2$
Discretas	$\sum_{x \in Sop(X)} (x - E(X))^2 P(x)$	$E(X^2) - E(X)^2$
Continuas	$\int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$	$E(X^2) - E(X)^2$

Tabla 6.2. Comparativa de media y varianza para estadística descriptiva, variables discretas y variables continuas. Véanse las similitudes entre todas las fórmulas.

6.5. Ejemplos de distribuciones discretas

Veremos en este sección algunos ejemplos de distribuciones discretas que son sencillos y que aparecen en muchas situaciones en problemas reales. Estos ejemplos nos servirán para ejercitar los conceptos que se definieron en las secciones anteriores. Nos servirán también para ver cómo se deducen las funciones masa de probabilidad en varias situaciones prácticas. Además, el tratar estos casos nos permitirá introducir el concepto de parámetro, que será fundamental en la parte de inferencia estadística que veremos en los próximos capítulos. Finalmente, nos servirán para introducir la notación que se usa para representar esas distribuciones.

6.5.1. Distribución de Bernoulli

Esta variable aparece en aquellos experimentos en los que solo nos interesa saber si ocurrió una determinada situación o no, o si un individuo elegido al azar tiene una determinada característica o no. Si ocurrió la característica, diremos que ha pasado un *éxito*, lo denotaremos por E y le asignaremos valor 1, mientras que en caso contrario diremos que ha pasado un *fracaso*, lo denotaremos por F y le asignaremos valor 0. Los experimentos de este tipo, en el que solo se repite una vez y solo nos interesa si pasa o no algo se llaman *pruebas de Bernoulli*.

Ejemplo 112.

Consideremos el experimento consistente en lanzar un dado y supongamos que lo único que nos interesa es saber si salió 5 o no. Entonces, solo hay dos resultados posibles para el experimento, SI (al que le asociamos el valor 1) o NO (al que le asociamos el valor 0).

Desde un punto de vista matemático, se dice que una variable aleatoria discreta X tiene **distribución de Bernoulli** de parámetro p, donde p es un valor en [0,1], si

$$\boxed{Soporte(X) = \{0,1\},\ P(X=1) = p,\ P(X=0) = 1-p.}$$

Si X tiene distribución de Bernoulli de parámetro p lo denotaremos por $X \sim \mathcal{B}(p)$. Nótese que p es la probabilidad de obtener un éxito al realizar la prueba de Bernoulli.

Como hemos visto, p es el parámetro de la distribución. En general, un parámetro de una distribución es un valor que necesitamos conocer para poder determinar los valores del soporte y de probabilidad de la distribución. Así, por ejemplo, para una distribución de Bernoulli, el soporte siempre es el mismo, y la única diferencia entre distintas distribuciones de Bernoulli es debido a diferencias entre las probabilidades de éxito. Por eso, el único parámetro de la distribución es ese valor de probabilidad.

Ejemplo 113. *(Continuación del ejemplo 112)*

En el caso anterior, la variable sigue una distribución $\mathcal{B}(\frac{1}{6})$*, puesto que la probabilidad de obtener 5 (el éxito) es* $\frac{1}{6}$*, tal y como se ve al aplicar la regla de Laplace.*

La media y la varianza de esta variable son:

$$E(X) = p \cdot 1 + (1 - p) \cdot 0 = p.$$

$$E(X^2) = p \cdot 1^2 + (1 - p) \cdot 0^2 = p \Rightarrow$$
$$\Rightarrow V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p).$$

La distribución de Bernoulli es fundamental en el modelado de proporciones, como veremos más adelante. En concreto, p representa la probabilidad de que pase algo o, en otras palabras, la proporción de individuos de la población que tienen una característica.

Ejemplo 114.

Consideremos por ejemplo la población de los españoles adultos y consideremos el experimento que consiste en seleccionar al azar un individuo y ver si mide más de 1,70. En este caso, nuestra variable tiene distribución de Bernoulli y el parámetro p *denota la probabilidad de que un individuo mida más de 1,70. Y esta probabilidad viene dada por la proporción de individuos con altura superior a 1,70 sin más que aplicar la definición clásica de probabilidad: todos los individuos tienen la misma posibilidad de ser seleccionados, y viene dada por* $\frac{1}{n}$*, donde* n *es el número total de españoles adultos; si hay* r *individuos con altura superior a 1,70, la probabilidad de seleccionar un individuo en estas condiciones es* $\frac{r}{n}$*. Este es el valor de* p*.*

6.5.2. Distribución binomial

Supongamos que se repite n veces un mismo experimento de manera que cada repetición se realiza independientemente de las otras. En cada realización experimental se observa si ocurre un determinado suceso, habitualmente llamado *éxito* (E) o, si por el contrario, no ocurre tal suceso, usualmente llamado *fracaso* (F). Es decir, supongamos que realizamos n experimentos de Bernoulli de manera independiente. Supondremos que la probabilidad de éxito $p = P(E)$ se mantiene constante durante los n experimentos. Consideremos la variable aleatoria

$X \equiv$ número de éxitos obtenidos tras las n realizaciones del experimento.

Una variable aleatoria X en estas condiciones se dice entonces que sigue una **distribución binomial** de parámetros n y p, y lo denotaremos por $X \sim \mathcal{B}(n, p)$.

Ejemplo 115. *(Continuación del ejemplo 112)*

Consideremos nuevamente el ejemplo del dado. Supongamos que lanzamos el dado 20 veces. La variable $X \equiv$ «número de cincos obtenidos» sigue una distribución $\mathcal{B}(20, \frac{1}{6})$.

Veamos las características de esta variable. En primer lugar, es claro que el soporte es $Soporte(X) = \{0, ..., n\}$.

Veamos ahora cuál es su función masa de probabilidad. Es posible comprobar aplicando lo que ya hemos visto de cálculo de probabilidades que para $k \in \{0, ..., n\}$, se tiene

$$\boxed{P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},}$$

donde

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

y $0! = 1$ por convenio. Veamos cómo deducir esta fórmula. En efecto, si $X = k$, eso es porque ha habido k éxitos; si, por ejemplo, tenemos que los éxitos fueron los k primeros experimentos, la probabilidad de este resultado sería

$$\begin{aligned} P(E_1 \cap ... \cap E_k \cap F_{k+1} \cap ... \cap F_n) &= P(E_1) \cdots P(E_k) P(F_{k+1}) \cdots P(F_n) \\ &= p^k (1-p)^{n-k}, \end{aligned}$$

sin más que aplicar la independencia. Sin embargo, esta es solo una posibilidad para obtener k éxitos. En realidad, hay otras combinaciones que llevan a k éxitos y $n-k$ fracasos (por ejemplo, que los $n-k$ primeros experimentos sean fracasos y los k últimos sean éxitos). Todas estas combinaciones tienen la misma probabilidad ($p^k(1-p)^{n-k}$). Por lo tanto, la probabilidad de que X tome el valor k será $p^k(1-p)^{n-k}$ multiplicado por el número de posibilidades. Y el número de posibilidades son el número de formas que hay de elegir los k experimentos con éxito. Este valor es, tal y como se puede ver en el apéndice sobre combinatoria, el número combinatorio $\binom{n}{k}$.

En definitiva, para determinar el soporte y la función masa de probabilidad de una distribución binomial necesitamos conocer el número de experimentos de Bernoulli para conocer el soporte (es el parámetro n) y además el valor de la probabilidad de éxito para determinar los distintos valores de probabilidad (es el parámetro p).

Como caso particular de la distribución binomial, si X sigue distribución de Bernoulli $\mathcal{B}(p)$, entonces X sigue una distribución binomial $\mathcal{B}(1,p)$.

Además, por la construcción de la variable binomial, sabemos que si $X \sim \mathcal{B}(n,p)$, entonces X puede escribirse de la forma

$$X = X_1 + ... + X_n,$$

donde $X_i \sim \mathcal{B}(p)$, $\forall i = 1, ..., p$ independientes entre sí. En realidad, X_i representa el resultado de la i-ésima repetición del experimento, observándose si hubo éxito o no. Por eso, la asignación de los valores 1 y 0 para éxito y fracaso en la distribución de Bernoulli no son arbitrarios y no pueden cambiarse.

Esta forma de escribir una variable binomial como suma de variables de Bernoulli nos permite calcular fácilmente la media y la varianza de X:

$$E(X) = E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n) = p + ... + p = np.$$

$$\begin{aligned} V(X) &= V(X_1 + ... + X_n) = V(X_1) + ... + V(X_n) \\ &= p(1-p) + ... + p(1-p) = np(1-p). \end{aligned}$$

6.5.3. Distribución de Poisson

Esta variable aparece cuando se está midiendo el número de veces que pasa algo por unidad de medida. Por ejemplo, es la variable que se aplica para medir el número de llamadas por hora a una centralita, el número de bacterias por centímetro cúbico de agua, etc.

Desde un punto de vista matemático, una variable aleatoria X tiene **distribución de Poisson** de parámetro λ (donde λ es un número real positivo) y lo denotaremos $X \sim \mathcal{P}(\lambda)$ si su soporte y su función masa de probabilidad vienen dados por

$$\boxed{Soporte(X) = \{0, 1, ...\},\ P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, ...}$$

Veamos cómo deducir esta expresión. Supongamos que queremos saber el número de llamadas a una centralita en una hora. Esto se puede aproximar de la siguiente manera: dividimos la hora en dos partes de media hora cada una y vemos si hay alguna llamada en cada una de ellas o no. Esto sigue una distribución binomial de parámetros $\mathcal{B}(2, p_2)$. Pero nótese que podría ocurrir que tuviésemos dos llamadas en alguno de estos dos tramos, y entonces no saldría el valor que nos interesa. Para conseguir una mejor aproximación, también podemos dividir la hora en cuatro partes y entonces considerar una distribución binomial de parámetros $\mathcal{B}(4, p_4)$. Aquí, p_4 es más pequeño que p_2 porque mide la ocurrencia de algún éxito en menor tiempo. Y así sucesivamente. Al final tenemos un número de partes n que es muy grande (tiende a infinito) y p_n muy pequeño (tiende a 0). Si np_n se aproxima a λ, entonces

$$\lim_{\substack{n\to\infty \\ p_n\to 0 \\ np_n\to\lambda}} P(k) = \lim_{\substack{n\to\infty \\ p_n\to 0 \\ np_n\to\lambda}} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Como en los otros casos, para diferenciar entre distintas distribuciones de Poisson basta conocer el valor de λ. Por eso esta distribución solo tiene un parámetro.

Nótese que en este caso el dominio tiene infinitos valores posibles. A pesar de ello esta variable es discreta, pues toma valores aislados (en realidad porque toma una cantidad infinita numerable de valores).

Puede comprobarse sumando las correspondientes series matemáticas que

$$E(X) = \lambda, V(X) = \lambda.$$

Para recordar estas expresiones, pude ser útil comparar estos valores con los correspondientes de la distribución binomial. La esperanza de la distribución binomial es np, y sabemos que np tiende a λ en el caso de la distribución de Poisson. La varianza de la binomial era $np(1-p)$, pero np tiende a λ y $(1-p)$ tiende a 1.

Ejemplo 116.

Supongamos que un teléfono de reclamaciones recibe una media de 3 llamadas por minuto. Entonces, la variable $X \equiv$ *número de llamadas por minuto sigue una distribución* $\mathcal{P}(3)$, *pues el parámetro de la distribución coincide con su media.*

6.6. Distribuciones continuas

Veamos ahora algunos ejemplos de distribuciones continuas. Seguiremos el mismo proceso que en el caso discreto, estudiando las situaciones en las que aparecen estas distribuciones y sus características (soporte y función de densidad) de cada una de ellas. Veremos también los parámetros de cada una de ellas. Pero antes de empezar hay que tener en cuenta dos cosas:

- Aunque establezcamos que una distribución aparece en una determinada situación, en muchas ocasiones otras familias de distribuciones son váidas para representar ese experimento. Es decir, dado un problema concreto, no podemos asignar directamente la distribución.

- En la sección anterior hemos deducido las distintas funciones masa de probabilidad. Sin embargo, esto no es tan sencillo para la función de densidad, con lo que nos limitaremos a escribirla y no haremos su deducción.

6.6.1. Distribución uniforme

Una variable aleatoria X se dice que tiene **distribución uniforme** de parámetros a y b con $a < b$ si su función de densidad viene dada por:

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

Denotaremos la distribución uniforme en (a, b) por $\mathcal{U}(a, b)$. La distribución uniforme representa la ignorancia total sobre el experimento, en el sentido de que solo se sabe que se ha obtenido un valor en el intervalo (a, b), sin que podamos dar ninguna información adicional sobre qué valor puede haber salido. Por ello, no podemos asignar una densidad mayor a un punto del intervalo que a otro y esto conlleva que la función de densidad sea la misma para todos los puntos del dominio. Nótese que para conocer el soporte y la función de densidad de una distribución uniforme basta conocer el intervalo en el que está palicada. Por eso, los parámetros de la distribución son precisamente los extremos de ese intervalo.

La gráfica de esta función de densidad viene dada en la figura 6.4.

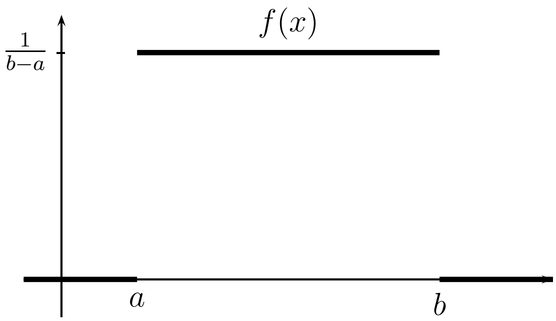

Figura 6.4. Representación gráfica de la función de densidad de una distribución $\mathcal{U}(a, b)$.

Como era de esperar, se tiene que $E(X) = \frac{b+a}{2}$, que coincide con el punto medio del segmento (a, b). Esto es fácil de justificar intuitivamente, pues todos los puntos del intervalo son igualmente lógicos (como puede verse en la representación gráfica de la función de densidad) y este punto es el que compensa diferencias por encima y por debajo. Veamos el cálculo matemático de dicha esperanza.

$$
\begin{aligned}
E(X) &= \int_a^b x\frac{1}{b-a}dx = \left.\frac{x^2}{2}\frac{1}{b-a}\right]_a^b \\
&= \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}.
\end{aligned}
$$

Pasamos ahora a calcular la varianza de esta distribución.

$$
\begin{aligned}
E(X^2) &= \int_a^b x^2\frac{1}{b-a}dx = \left.\frac{x^3}{3}\frac{1}{b-a}\right]_a^b = \frac{b^3-a^3}{3(b-a)} \\
&= \frac{(b-a)(b^2+ba+a^2)}{3(b-a)} = \frac{b^2+ba+a^2}{3}.
\end{aligned}
$$

Como $V(X) = E(X^2) - E(X)^2$, se tiene

$$
\begin{aligned}
V(X) &= \frac{b^2+ba+a^2}{3} - \frac{(b+a)^2}{4} \\
&= \frac{4(b^2+ba+a^2) - 3(b^2+2ab+a^2)}{12} \\
&= \frac{(b-a)^2}{12}.
\end{aligned}
$$

6.6.2. Distribución exponencial

La variable exponencial aparece cuando hay un proceso de Poisson y queremos estudiar el tiempo que transcurre entre dos ocurrencias consecutivas de un suceso. Es decir, la distribución exponencial se puede definir como el tiempo que transcurre entre dos eventos consecutivos de una distribución de Poisson. En la práctica se utiliza para medir el tiempo de vida de una persona, el tiempo antes de que falle una máquina, etc.

Ejemplo 117.

Supongamos que el número de llamadas por hora a una centralita de teléfono sigue distribución de Poisson. Entonces, el tiempo entre dos llamadas consecutivas sigue distribución exponencial.

Desde un punto de vista matemático, se dice que X tiene distribución **exponencial** de parámetro λ (con $\lambda > 0$) si su función de densidad es:

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

Denotaremos la distribución exponencial de parámetro λ por $\mathcal{E}xp(\lambda)$. La gráfica de la función de densidad puede verse en la figura 6.5.

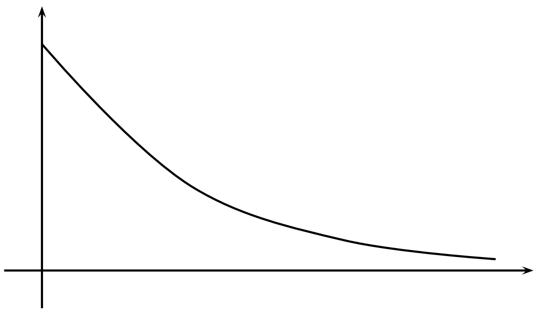

Figura 6.5. Representación gráfica de la función de densidad de una distribución $\mathcal{E}xp(\lambda)$.

Nótese que para conocer una distribución exponencial, el dominio es siempre el mismo y lo que diferencia a distintas distribuciones exponenciales es el valor λ que aparece en la función de densidad. Por eso esta distribución solo tiene un parámetro. Veamos ahora cuáles son los valores de la media y la varianza de esta distribución:

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big]_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 - \frac{1}{\lambda} e^{-\lambda x}\Big]_0^\infty = \frac{1}{\lambda}.$$

Por tanto, cuanto mayor sea λ menos tiempo de vida esperado y más rápido es de esperar que ocurra el suceso. Pasamos ahora a calcular $E(X^2)$.

$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x}\Big]_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Luego

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

6.6.3. Distribución normal

Esta distribución es sin duda la más importante de todas las distribuciones. Es una distribución que aparecerá casi siempre en los capítulos posteriores de inferencia paramétrica. La distribución normal se utiliza además para aproximar muchos fenómenos aleatorios como por ejemplo alturas, pesos, etc. No obstante, no hay condiciones en las que podamos asegurar que un fenómeno concreto siga distribución normal, con lo que deberemos conocerlo por estudios anteriores o comprobarlo en cada caso.

Una variable aleatoria se dice que tiene **distribución normal** de parámetros μ y σ, con $\sigma > 0$ si su función de densidad viene dada por

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

Entonces, para determinar una distribución normal solo tenemos que fijarnos en la función de densidad, puesto que el dominio es toda la recta real para cualquier normal. Para determinar la función de densidad, necesitamos conocer los valores de μ y σ, por lo que estos son los parámetros de la distribución normal. Así, denotaremos la distribución normal por $\mathcal{N}(\mu, \sigma)$.

Su representación gráfica puede verse en la figura 6.6.

La función de densidad de la distribución normal es simétrica respecto al parámetro μ; esto es muy útil a la hora de calcular valores de

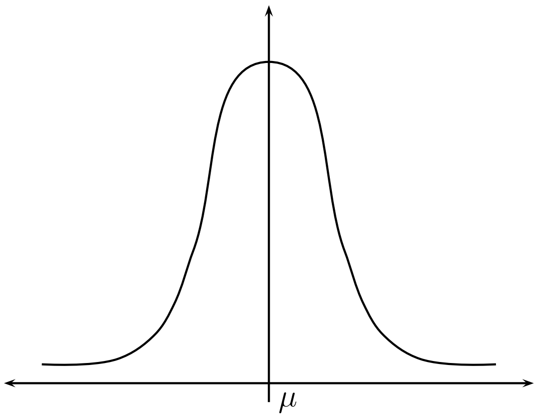

Figura 6.6. Representación gráfica de la función de densidad de una distribución $\mathcal{N}(\mu, \sigma)$.

probabilidad de esta distribución. Por otra parte, tiene forma de campana (de hecho, esta distribución es también conocida como *campana de Gauss*).

Debido a la simetría, es claro que $E(X) = \mu$, pues el punto de simetría compensa diferencias por encima y por debajo. Esto también puede verse calculando directamente la integral. Además, puede comprobarse integrando por partes que $V(X) = \sigma^2$:

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\ &= -\sigma^2 (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \sigma^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\ &= 0 + \sigma^2. \end{aligned}$$

En resumen, el valor de μ nos dice el punto en que se alcanza el máximo de la función de densidad y σ nos indica si este pico es muy apuntado o no; valores grandes de σ implican poco apuntamiento, mientras que valores pequeños de σ implican mucha concentración de probabilidad alrededor de μ y, por tanto, un pico de la campana de Gauss muy pronunciado en el punto medio de la distribución.

La función de densidad de la distribución normal no puede ser integrada mediante métodos matemáticos elementales y hay que recurrir a métodos numéricos para aproximar los diferentes valores de probabilidad. Sin embargo, existen tablas a partir de las cuales pueden obtenerse valores aproximados para la distribución $\mathcal{N}(0,1)$. Esta distribución se conoce como **normal estándar** y se denota por la letra Z. La tabla con los valores de probabilidad de la distribución normal estándar aparecen en la tabla B.1 del apéndice B. Denotaremos por z_α con $\alpha \in (0,1)$ el valor tal que

$$P(Z \geq z_\alpha) = \alpha.$$

Por ejemplo, a partir de la tabla B.1 $z_{0{,}025} = 1{,}96$. Esta notación será muy importante para comprender las fórmulas que aparecen en los capítulos posteriores.

Veamos algunas propiedades de la distribución normal.

- Si $X \sim \mathcal{N}(\mu, \sigma)$, entonces $aX + b \sim \mathcal{N}(a\mu + b, |a|\sigma)$.

  En este resultado, lo único novedoso es que si X sigue distribución normal, también lo hace $aX+b$, pues los valores de los parámetros pueden obtenerse aplicando las propiedades vistas anteriormente para la media y la desviación típica.

  En particular, si tenemos una distribución $\mathcal{N}(\mu, \sigma)$, para poder calcular sus valores de probabilidad tendremos que pasar a una distribución $\mathcal{N}(0,1)$. Esto se consigue aplicando que

  $$X \sim \mathcal{N}(\mu, \sigma) \Rightarrow \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1),$$

  proceso que se conoce con el nombre de *tipificación*. Por ello, a la distribución normal estándar se le llama también distribución normal tipificada.

  Ejemplo 118.

  Supongamos que $X \sim \mathcal{N}(4,2)$ y que queremos hallar la probabilidad de que $X \leq 8$. En este caso

  $$P(X \leq 8) = P\left(\frac{X-4}{2} \leq \frac{8-4}{2}\right) = P(\mathcal{N}(0,1) \leq 2) = 0{,}0228.$$

- Si $X \sim \mathcal{N}(\mu_1, \sigma_1)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2)$, siendo estas variables independientes, entonces

    - $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
    - $X - Y \sim \mathcal{N}(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

  En estos resultados es necesaria la independencia entre las variables. Si esta condición no se tiene, el resultado es falso. Hay que tener cuidado en la fórmula de la diferencia, pues las varianzas se suman; nótese que si se restasen, podríamos obtener valores negativos para la varianza, lo que sería imposible.

6.7. Distribuciones derivadas de la normal

Asociadas a la distribución normal tenemos otras tres distribuciones continuas, que veremos brevemente a continuación. Estas distribuciones están tabuladas y su uso se reducirá a la parte de inferencia estadística, por lo que no necesitamos conocer muchos detalles de las mismas.

6.7.1. Distribución χ^2 de Pearson

Esta distribución está asociada a la distribución normal porque la distribución χ_1^2 es la distribución de una normal estándar al cuadrado. En general, dadas $X_1, ..., X_n$ variables aleatorias normales estándar independientes entre sí, diremos que la variable

$$Y = X_1^2 + ... + X_n^2$$

sigue una distribución χ^2. En consecuencia, la distribución χ^2 solo toma valores positivos. Depende de un parámetro llamado *grados de libertad*, que es el número de variables normales consideradas en la definición, por lo que usaremos la notación $X \sim \chi_n^2$, donde n son sus grados de libertad. Los grados de libertad son un número entero positivo[5]. La gráfica de su función de densidad puede verse en la figura 6.7.

[5]Los grados de libertad pueden tomar cualquier valor positivo, no necesariamente entero, pero nosotros siempre aproximaremos al valor entero más cercano.

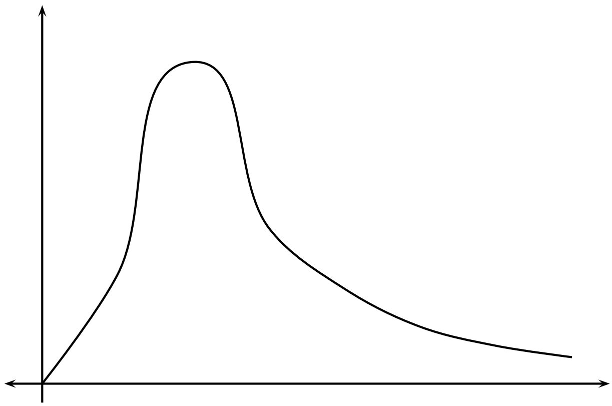

Figura 6.7. Representación gráfica de la función de densidad de una distribución χ^2.

Al igual que con la distribución normal, los valores de probabilidad de la distribución χ^2_n están tabulados, por lo que no necesitaremos conocer su función de densidad (que es bastante complicada) para calcular sus valores de probabilidad. La tabla correspondiente a la distribución χ^2 se incluye en el apéndice B, tabla B.2. Denotaremos por $\chi^2_{n;\alpha}$ con $\alpha \in (0,1)$ el valor tal que

$$P(\chi^2_n \geq \chi^2_{n;\alpha}) = \alpha.$$

Por ejemplo, a partir de la tabla B.2, $\chi^2_{7;0{,}025} = 16{,}01$.

6.7.2. Distribución *t* de Student

Esta distribución[6] proviene del cociente entre una normal estándar y la raíz cuadrada de una χ^2 dividida por sus grados de libertad, siendo estas variables independientes. Es decir, dada $X_1 \sim \mathcal{N}(0,1), X_2 \sim \chi^2_n$ independientes entre sí, entonces

$$Y = \frac{X_1}{\sqrt{X_2/n}},$$

[6]Esta distribución fue desarrollada por W. Gosset, pero utilizaba este pseudónimo porque su empresa (una cervecera) no le permitía realizar estos estudios.

sigue una distribución t. Esta distribución es muy similar a la distribución normal tipificada y de hecho es simétrica respecto a 0. Se diferencia en que las colas de la distribución t están ligeramente más elevadas. La gráfica de su función de densidad puede verse en la figura 6.8

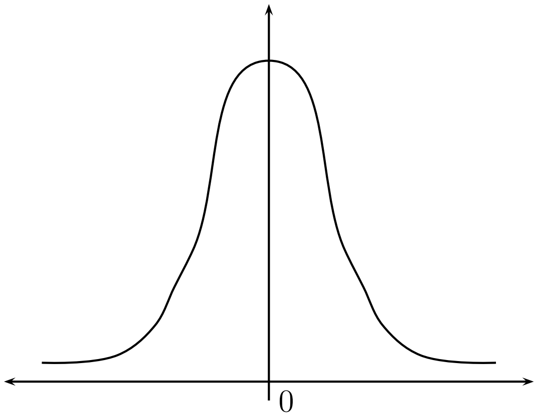

Figura 6.8. Representación gráfica de la función de densidad de una distribución t.

La distribución t depende de un parámetro n llamado *grados de libertad* que coincide con los grados de libertad de la distribución χ^2 que aparece en su definición. Así, usaremos la notación $Y \sim t_n$. Como ocurría anteriormente, los valores de probabilidad de la distribución t están tabulados, por lo que no es necesario conocer su función de densidad para calcular sus valores de probabilidad. La tabla correspondiente a la distribución t se incluye en el apéndice B, tabla B.3. Denotaremos por $t_{n;\alpha}$ con $\alpha \in (0,1)$ el valor tal que

$$P(t_n \geq t_{n;\alpha}) = \alpha.$$

Por ejemplo, a partir de la tabla B.3, $t_{7;0{,}025} = 2{,}365$.

6.7.3. Distribución *F* de Snedecor

Esta distribución proviene del cociente de dos distribuciones independientes χ^2 entre sus grados de libertad. En otras palabras, si $X_1 \sim \chi^2_n, X_2 \sim \chi^2_m$ independientes entre sí, entonces

$$Y = \frac{X_1/n}{X_2/m}$$

sigue una distribución F. Al igual que la distribución χ^2, la distribución F solo toma valores positivos. Depende de dos parámetros, m, n que toman valores enteros y que coinciden con los grados de libertad de las distribuciones χ^2 que aparecen en su definición. Por lo tanto, usaremos la notación $Y \sim F_{n,m}$. La gráfica de su función de densidad puede verse en la figura 6.9.

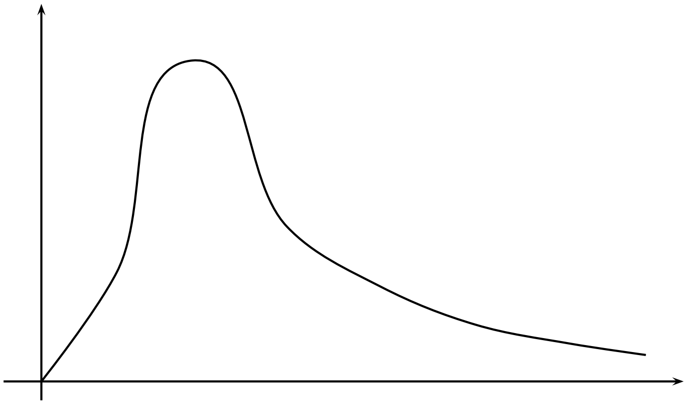

Figura 6.9. Representación gráfica de la función de densidad de una distribución F.

Como ocurría anteriormente, los valores de probabilidad de la distribución F están tabulados, por lo que no es necesario conocer su función de densidad para calcular sus valores de probabilidad. La tabla correspondiente a la distribución F se incluye en el apéndice B, tabla B.4. Denotaremos por $F_{n,m;\alpha}$ con $\alpha \in (0,1)$ el valor tal que

$$P(F_{n,m;\alpha} \geq F_{n,m;\alpha}) = \alpha.$$

Por ejemplo, a partir de la tabla B.4 $F_{3,7;0,05} = 4{,}35$.

Finalmente, esta distribución tiene la propiedad de que si $X \sim F_{n,m}$, entonces se cumple que $\frac{1}{X} \sim F_{m,n}$, propiedad que se utiliza para mirar los valores en las tablas. Por ejemplo, si queremos hallar $F_{7,3;0,95}$ tenemos que proceder de la siguiente manera:

$$
\begin{aligned}
P(F_{7,3} \geq F_{7,3;0,95}) = 0{,}95 \quad &\Rightarrow \quad P(F_{7,3} \leq F_{7,3;0,95}) = 0{,}05 \\
&\Rightarrow \quad P(\frac{1}{F_{7,3}} \geq \frac{1}{F_{7,3;0,95}}) = 0{,}05.
\end{aligned}
$$

Por lo tanto,

$$
\frac{1}{F_{7,3;0,95}} = F_{3,7;0,05} = 4{,}35 \Rightarrow F_{7,3;0,95} = \frac{1}{4{,}35} = 0{,}23.
$$

6.8. Teorema central del límite

En esta sección veremos el que tal vez es el resultado más importante de la estadística. Para comprender el alcance del teorema central del límite[7] debemos tener en cuenta que la distribución de una combinación de variables es un problema que puede ser complicado; por ejemplo, se puede ver que la suma de dos distribuciones binomiales independientes con el mismo valor del parámetro p sigue una distribución binomial; sin embargo, la diferencia de distribuciones binomiales no sigue una distribución binomial. De la misma forma, el producto de una binomial por una constante no sigue distribución binomial. Esto sucede para la mayor parte de las distribuciones de probabilidad y se complica si consideramos tres, cuatro distribuciones, etc. El teorema central del límite nos permitirá hallar estas distribuciones de probabilidad de forma aproximada, incluso cuando el número de variables involucradas sea muy grande. Este resultado es el que justifica que la distribución normal sea la que más aparece en situaciones prácticas.

6.8.1. Distribución de sumas de normales

Ya hemos visto en la sección de variables aleatorias continuas que la suma y la diferencia de distribuciones normales sigue una distribución normal; también sigue una distribución normal el producto de una constante por una distribución normal y la suma de una constante y una distribución normal. Por otra parte, los parámetros que determinan una

[7] Aunque en realidad deberíamos llamarlo teorema del límite central.

distribución normal son la media y la desviación típica (o la varianza). Si juntamos entonces todos estos resultados se obtiene:

Teorema 1.

Sean $X_1, ..., X_n$ *distribuciones normales,* $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ *y sean unas constantes* $a_1, ..., a_n$*. Entonces, la variable* $X = \sum_{i=1}^{n} a_i X_i$ *sigue una distribución* $\mathcal{N}(\mu, \sigma)$ *donde*

$$\mu = a_1\mu_1 + ... + a_n\mu_n.$$

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + \sum_{i<j} 2a_i a_j Cov(X_i, X_j).$$

El problema que tiene este resultado es que necesitamos conocer las covarianzas entre las variables. Sin embargo, en muchas situaciones prácticas tenemos variables que son independientes entre sí. Esto es por ejemplo lo que nos va a pasar en la parte de la inferencia estadística. En este caso tenemos el siguiente corolario:

Teorema 2.

Sean $X_1, ..., X_n$ *distribuciones normales independientes entre sí,* $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ *y sean* $a_1, ..., a_n$ *constantes. Entonces, la variable* $X = \sum_{i=1}^{n} a_i X_i$ *sigue una distribución* $\mathcal{N}(\mu, \sigma)$ *donde*

$$\mu = a_1\mu_1 + ... + a_n\mu_n.$$

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Ejemplo 119.

Se construye una pieza a partir de otras tres piezas X_1, X_2, X_3*. Posteriormente se corta una parte de la pieza total* X_4 *de forma que la pieza final no incluye este trozo. Si las secciones se distribuyen* $X_1 \sim \mathcal{N}(2, 0.02), X_2 \sim \mathcal{N}(2.4,\ 0.25), X_3 \sim \mathcal{N}(2.1,\ 0.05)$ *y* $X_4 \sim \mathcal{N}(1, 0.01)$*, determinar la proporción de piezas con una longitud superior a los 6 cm.*

Veamos cómo resolver este problema. Nótese que necesitamos conocer la distribución de la pieza final, que denotaremos por X. Como estamos con distribuciones normales, se tiene que $X = X_1 + X_2 + X_3 - X_4$ sigue distribución normal por los resultados anteriores. Tenemos que hallar los parámetros, que nuevamente por los resultados anteriores y aplicando que las longitudes de cada pieza son independientes entre sí, son:

$$\begin{aligned} \mu &= 1.6 + 2.4 + 2.3 - 1 = 5.3, \\ \sigma &= \sqrt{0.02^2 + 0.25^2 + 0.05^2 + 0.01^2} = 0.256. \end{aligned}$$

Por lo tanto,

$$P(\mathcal{N}(5.3,\ 0.256) > 6) = P(\mathcal{N}(0,1) > 2.73) = 0.0032.$$

6.8.2. El teorema central del límite

En la sección anterior hemos visto cómo calcular la distribución de sumas y restas de distribuciones normales. Sin embargo, en muchas ocasiones no podemos afirmar que todas las distribuciones consideradas sean normales y esto hace que no podamos aplicar los resultados anteriores. Es aquí donde entra en funcionamiento el teorema central del límite, que nos da la distribución aproximada de cualquier combinación de variables aleatorias independientes entre sí. Daremos aquí una versión simplificada de este resultado más orientada a la práctica y que permite evitar el estudio de resultados de convergencia de variables aleatorias.

Teorema 3. *(Teorema central del límite)*

Sean $\{X_1, ..., X_n, ...\}$ una sucesión de variables aleatorias independientes con medias μ_n y varianzas σ_n^2 respectivamente. Entonces, se tiene que

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \to \mathcal{N}(0,1).$$

Este resultado parece muy abstracto. Veamos las conclusiones que se pueden extraer.

- Lo primero que nos dice es que si tenemos una sucesión de variables aleatorias, la distribución de la suma sigue en el límite una distribución normal, sin importar la distribución de las variables consideradas (podrían ser incluso discretas) ni que estas distribuciones no sean iguales. En otras palabras, nos dice que si tenemos una variable aleatoria que es suma de muchas variables, entonces sigue aproximadamente una distribución normal. Esto es lo que hace que la distribución normal sea tan común, ya que cualquier variable que dependa de muchos factores sigue aproximadamente una distribución normal. Así, variables como la altura o el peso de una persona siguen una distribución normal porque los valores de esas variables dependen de muchos factores (genética, alimentación, estilo de vida...).

- Por la propia definición de límite, esto implica que a partir de un número n suficientemente grande, la suma ya estará suficientemente cerca de la distribución normal como para que la podamos considerar normal sin cometer un grave error; claro está, este valor n es desconocido, pero simulaciones han permitido tomar el valor $n = 30$ como suficientemente grande en muchas situaciones prácticas.

- Por otra parte, nótese que la variable considerada en el teorema central del límite es una normal tipificada; por ello, si tenemos un número finito de variables se tiene que podemos considerar

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right).$$

- En muchos problemas estaremos en la situación de un experimento que se repite muchas veces y se nos pregunta por la variable media muestral; supongamos que en cada experimento la media es μ y la desviación típica es σ. Por el punto anterior tenemos que

$$\frac{\sum_{i=1}^{n} X_i}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Volveremos a hablar de este resultado más adelante.

Veamos un ejemplo de cómo se aplica este resultado, de forma que se pueda ver la enorme simplificación que supone para resolver problemas en los que intervienen muchas variables aleatorias.

Ejemplo 120.
Se ha estimado que la demanda de energía eléctrica en una zona es una variable aleatoria de media 10 y desviación típica 2.5.

- *¿Cuál es la probabilidad de que en un período de 30 días se demande menos de 190?*
- *Si para simplificar suponemos que las facturas vienen en bloques de 30 días (que coinciden más o menos con los meses), ¿qué cantidad es demandada al menos el 20% de las veces?*
- *¿Cuál es la probabilidad de que durante 12 períodos la demanda sea superior a 280?*

En este problema, la demanda en cada período es la suma de las demandas de cada uno de los días. Llamaremos X_i a la demanda en el día i; además, todas estas variables son independientes entre sí. Aplicando ahora el teorema central del límite se tiene que la demanda en el período, que denotaremos por Y, sigue aproximadamente una distribución normal de parámetros

$$\mu_Y = \mu_{X_1} + ... + \mu_{X_{30}} = 300, \sigma_Y^2 = \sigma_{X_1}^2 + ... + \sigma_{X_{30}}^2 = 6.25 \times 30 = 187.5.$$

Nótese que para conseguir este resultado no necesitamos conocer la distribución de la demanda en cada día, tan solo su media y su varianza. Por lo tanto, ahora podemos concluir que

$$P(Y \leq 190) = P(\mathcal{N}(300, 13.7) \leq 190) = P(\mathcal{N}(0,1) \leq -8.03) = 0.$$

En el segundo apartado nos piden el percentil 20 de la distribución; entonces,

$$P(\mathcal{N}(300, 13.7) \leq P_{20}) = 0.2 \Leftrightarrow P\left(\mathcal{N}(0,1) \leq \frac{P_{20} - 300}{13.7}\right) = 0.2$$

$$\Leftrightarrow P\left(\mathcal{N}(0,1) \leq \frac{300 - P_{20}}{13.7}\right) = 0.8.$$

Por las tablas de la normal se obtiene

$$\frac{300 - P_{20}}{13.7} = 0.84 \Leftrightarrow P_{20} = 288.492.$$

Finalmente, sea $Z \equiv$ *número de períodos con demanda superior a 280. Esta variable sigue una distribución* $\mathcal{B}(12, p)$*, donde* p *es la probabilidad de que un mes concreto tenga una demanda superior a 280. Procediendo como en el primer apartado se tiene que*

$$\begin{aligned} p = P(Y \geq 280) &= P(\mathcal{N}(300, 13.7) \geq 280) \\ &= P(\mathcal{N}(0,1) \geq -1.46) \\ &= 0.9279. \end{aligned}$$

Ahora nos piden la probabilidad de que $Z = 12$*, que es según la función masa de probabilidad de la binomial*

$$P(\mathcal{B}(12, 0.9279) = 12) = 0.9279^{12} = 0.41.$$

6.8.3. Aproximación normal de la distribución binomial

Vamos ahora a ver otra consecuencia del teorema central del límite. Supongamos que tenemos una distribución binomial $X \sim \mathcal{B}(n,p)$

donde n es grande. Entonces, ya conocemos que su función masa de probabilidad es

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},\ k = 0, ..., n.$$

Si nos piden ahora una probabilidad concreta de esta distribución tendríamos que calcular los correspondientes números combinatorios y esto puede ser imposible si n es grande (y ya 100 es un valor muy grande para estos números combinatorios). Por otra parte, ya hemos visto que la distribución binomial X puede escribirse como

$$X = \sum_{i=1}^{n} X_i,$$

donde $X_i \sim \mathcal{B}(p)$ independientes entre sí. En definitiva, se tiene que X es una suma de variables y podemos entonces aplicar el teorema central del límite para concluir que la distribución de X se puede aproximar por

$$X \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right).$$

Además, como todas las variables X_i tienen la misma distribución se tiene que

$$\mu_i = p,\ \sigma_i^2 = p(1 - p),$$

con lo que

$$X \sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right),$$

que son precisamente la media y la varianza de X. Este resultado se conoce como la *aproximación normal de la binomial.*

Ejemplo 121.

Sea una variable $X \sim \mathcal{B}(100, 0.4)$. *Vamos a calcular* $P(X \in [35, 45])$. *Por lo dicho anteriormente, podemos aproximar la distribución de* X *por una* $\mathcal{N}(40, \sqrt{24})$.

$$\begin{aligned} P(X \in [34, 45]) &= P(\mathcal{N}(40, \sqrt{24}) \in [35, 45]) \\ &= P(\mathcal{N}(0, 1) \in [\text{-}1.02,\ 1.02]) \\ &= 0.6922. \end{aligned}$$

Nótese que estamos aproximando una variable discreta por una continua. Entonces, si nos piden la probabilidad de un valor concreto, la aproximación sería 0 por ser una variable continua; por otra parte, este valor es necesariamente positivo sin la aproximación, pues pertenece al dominio de la variable. Para evitar esta contradicción, el valor i de X se identifica con el intervalo $[i - 1/2, i + 1/2]$. La idea de esto es repartir la probabilidad del intervalo $[i, i + 1]$, que en la distribución normal tiene valor positivo, entre los dos valores que realmente pueden ocurrir en la binomial, que son i e $i + 1$. Y lo que se hace es entonces asignar $[i, i + 0{,}5]$ a i y $[i + 0{,}5, i + 1]$ a $i + 1$. Esto se conoce como la *corrección por continuidad* y puede verse la idea de esta aproximación en la figura 6.10.

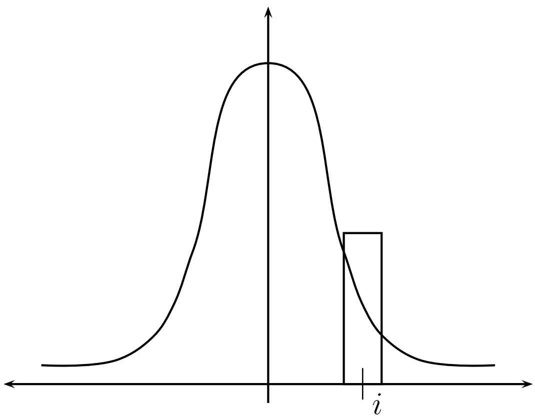

Figura 6.10. Idea intuitiva de la corrección por continuidad de una variable binomial por una distribución normal.

Ejemplo 122. *(Continuación del ejemplo 121)*

Aplicando corrección por continuidad

$$\begin{aligned}
P(X \in [34.5, 45.5]) &= P(\mathcal{N}(40, \sqrt{24}) \in [34.5, 45.5]) \\
&= P(\mathcal{N}(0,1) \in [-1.12,\ 1.12]) \\
&= 0.7372.
\end{aligned}$$

Nótese que nos ha salido una probabilidad parecida a la del ejemplo 121, a pesar de que tenemos una binomial con un parámetro n que no es excesivamente grande.

Esto que se ha hecho para la binomial se puede aplicar también a la distribución de Poisson, pues ya hemos visto que se puede poner como límite de una suma de binomiales. Así, si $X \sim \mathcal{P}(\lambda)$, desarrollando como se hizo para la binomial, se puede ver que podemos aproximar la distribución anterior por $X \sim \mathcal{N}(\lambda, \sqrt{\lambda})$.