Matemáticas Aplicadas a las Ciencias Sociales I: 1º Bachillerato

Capítulo 8: Distribuciones de probabilidad

www.apuntesmareaverde.org.es

Autora: Raquel Caro

Revisores: David Miranda y Luis Carlos Vidal

Ilustraciones: Banco de Imágenes de INTEF

Índice

Resumen

Muchos de los eventos que ocurren en la vida diaria no pueden ser predichos con exactitud a priori por diversas razones, pues la mayoría de ellos están influidos por factores externos. Además, existen sucesos que están directamente afectados por el azar, es decir, por procesos en los que no se está seguro de lo que va a ocurrir. La teoría de la probabilidad nos permite acercarnos a estos sucesos y estudiarlos, ponderando sus posibilidades de ocurrencia y proporcionando métodos para realizar estas ponderaciones.

En los capítulos anteriores has utilizado frecuencias. ahora vamos a asignar probabilidades y al estudiar los fenómenos aleatorios mediante distribuciones de probabilidad podremos construir modelos que reflejen la realidad y afirmar, con tal probabilidad. lo que va a ocurrir.

Además la teoría de la probabilidad es una herramienta necesaria para abordar la *Inferencia Estadística*. Esta agrupa un conjunto de métodos y técnicas que permiten extraer conclusiones generales de una población a partir de la observación de una muestra obtenida de ella. Además, también intenta obtener indicadores sobre la significación de las conclusiones obtenidas; es decir, sobre la confianza que merecen.

1. DISTRIBUCIONES DE PROBABILIDAD

1.1. Distribuciones de probabilidad: Media y desviación típica.

Cuando se analiza un fenómeno observable aparece una serie de resultados que han de ser tratados convenientemente, de manera que se puedan comprender mejor tanto los resultados como la característica objeto de estudio correspondiente a dicho fenómeno. Para este fin ya sabes realizar una primera descripción de los datos, histograma de frecuencias absolutas o relativas y polígono de frecuencias absolutas o acumuladas, y determinar sus parámetros: la media, varianza, desviación típica…

En ese caso, los propios resultados del experimento son numéricos como en el caso en el que se mide la velocidad de un vehículo, o la altura de un individuo. En cambio, en otras ocasiones, los resultados del experimento no proporcionan dicha información adecuadamente, como puede ser en el caso de los juegos de azar (ruleta, lotería, etc.).

En estos casos, se puede utilizar una variable aleatoria, que es una función que asigna un número real a cada resultado posible del espacio muestral del fenómeno bajo estudio. Por ejemplo, en los juegos de azar se puede asignar a cada resultado la ganancia o pérdida que produce en el jugador.

Las variables aleatorias se denotan mediante una letra mayúscula y pueden ser discretas (cuando pueden tomar un número finito o infinito numerable de valores) o continuas (cuando pueden tomar cualquier valor dentro de un rango).

En cuanto a las variables aleatorias discretas, éstas son las que pueden tomar un número finito o infinito numerable (como el conjunto N de los números naturales) de valores. Dado que la variable aleatoria puede tomar diferentes valores dependiendo de los resultados del experimento aleatorio al que está asociado, su valor no se podrá predecir de manera exacta. Así pues, para describir una variable aleatoria es necesario conocer su distribución de probabilidad.

Conocer la distribución de probabilidad de una variable aleatoria X discreta consiste en asignar una probabilidad a cada uno de los resultados posibles de dicha variable aleatoria. Es decir, se trata de saber calcular o asignar los valores $P[X = x]$, para todos los posibles valores x que puede tomar la variable aleatoria X.

Actividad resuelta

- *Se lanzan dos monedas y contamos el número de caras. La distribución de probabilidad es:*

Al hacer un diagrama en árbol calculamos las probabilidades:

Número de caras (x):	0	1	2
Probabilidad ($p(x)$):	1/4	1/2	1/4
Función de distribución $F(x)$:	1/4	3/4	4/4

Por un lado, tenemos la función $p(x)$, que es la probabilidad puntual o función de cuantía o función masa de probabilidad.

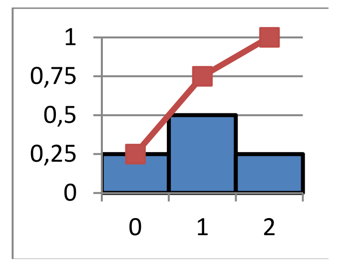
Por otro lado podemos calcular lo que sería equivalente a las frecuencias acumuladas. La función $F(x)$, a la que se denomina función de distribución, nos indica la probabilidad de que $F(x) = P(X \leq x)$, es decir, calcula la probabilidad de que se tomen valores menores a x.

Autora: Raquel Caro
Revisores: David Miranda y Luis Carlos Vidal
Ilustraciones: Banco de Imágenes de INTEF

La tabla que hemos presentado es una **distribución de probabilidad**, donde hemos definido una función que asigna a la variable aleatoria x una probabilidad:

$$x \to p(x)$$

y es el resultado que nos ayudará a hacer predicciones sobre un experimento aleatorio.



También podemos representar la tabla anterior mediante un histograma para la función de cuantía, en el que las áreas de cada rectángulo son ahora probabilidades, en lugar de frecuencias relativas, y podemos representar con una línea la función de distribución.

Observamos que siempre se verifican las siguientes propiedades.

Función de probabilidad o función de cuantía:

1) $p(x) \geq 0$
2) $\Sigma\, p(x) = 1$.

Función de distribución:

1) $0 \leq F(x) \leq 1$
2) $F(x)$ es una función creciente
3) $F(x_{Máximo}) = 1$.

Actividades propuestas

1. Se lanzan 3 monedas y contamos el número de caras que salen. Haz un diagrama en árbol. Escribe en una tabla la función de cuantía y la función de distribución. Representa la función de cuantía en un histograma y con una línea la función de distribución.

2. Se lanzan 2 dados y contamos el número de 5 que aparecen. Haz un diagrama en árbol, escribe en una tabla la función de cuantía y la función de distribución, y represéntalas gráficamente.

Actividad resuelta

- *Se lanzan dos dados. Por cada 5 que aparezca ganas 20 euros y pierdes 10 euros en caso contrario. ¿Te conviene ese juego? ¿Cuánto esperas ganar o perder en 60 jugadas?*

En cada lanzamiento puedes perder 10 euros o ganar 40 euros o ganar 20 euros. Esos son los valores de una variable aleatoria que podemos llamar ganancia, cuyas probabilidades calculamos, haciendo un diagrama de árbol, y escribimos en la siguiente tabla:

Ganancia (euros) (x):	–10	20	40
Probabilidad ($p(x)$):	25/36	10/36	1/36

Por tanto en 36 jugadas "esperamos" perder 10 euros en 25 de ellas, ganar 20 euros en 10, y ganar 40 euros en una jugada.

Ahora la variable aleatoria, que es discreta, es la ganancia.

Podemos calcular la **media** o **esperanza matemática** $E(x)$ con la expresión:

$$E(x) = \sum_i x_i \cdot p(x_i)$$

La **esperanza matemática** es una media teórica, de ahí el nombre de esperanza. Indica que si repetimos el experimento varias veces se espera que la media de los valores obtenidos se aproxime a esta esperanza calculada.

Para distinguir la media de una distribución de frecuencias de la esperanza de una distribución de probabilidad, se suele utilizar para las frecuencias la letra m o $\bar{x}$, mientras que para la esperanza matemática se utiliza la letra griega μ (que se lee "mu") o $E(x)$.

Un juego es equitativo si la esperanza matemática de la ganancia es 0, es ventajoso si $E(x) > 0$, y es desventajoso si $E(x) < 0$.

En la actividad propuesta anteriormente calculamos la media o esperanza matemática:

Esperanza matemática = $E(x)$ = –10(25/36) + 20(10/36) + 40(1/36) = (–250 + 200 + 40)/36 = –10/36. Como $E(x) < 0$, el juego es desventajoso.

Esto sería como calcular lo que "esperas" perder en 60 jugadas.

Actividades propuestas

3. Se lanzan 3 monedas. Por jugar nos cobran 1 euro, y por cada cara que aparezca ganamos 1 euro. Escribe una distribución de probabilidad y representa el histograma. ¿Cuánto esperas ganar o perder en 100 lanzamientos?

4. Una persona apuesta 10 euros a un juego de tirar una moneda y que salga cara o cruz (o similar). Si gana se retira y deja de jugar. Si pierde, apuesta el doble, 20 euros. Si gana se retira. Si pierde apuesta el doble, 40 euros. Y así sucesivamente. Con esta estrategia siempre acaba ganando 10 euros, pero tiene un defecto, ¡que no lleve suficiente dinero para seguir jugando hasta ganar! Imagina que lleva 500 euros. A) Haz un diagrama de árbol y calcula todas las posibilidades y sus probabilidades. B) La distribución de probabilidad: Ganancia (x) $\rightarrow$ Probabilidad (x). C) ¿Es un juego ventajoso? ¿Y para nuestro jugador? D) Calcula la probabilidad de ganar 10 euros y la de perder 500 euros.

5. Lanzamos dos dados. Si apostamos al 7 y sale, recuperamos tres veces lo apostado. Si apostamos que sale menor que 7 y sale, recuperamos lo apostado, y lo mismo si apostamos que sale mayor que 7. ¿Cuál es la mejor estrategia?

Por ejemplo, asegurar un coche a todo riesgo es un juego desventajoso, pero nos asegura que no habrá pérdidas grandes.

Para saber si los valores son próximos a la media, ya sabes que se utiliza la **desviación típica**. Lo mismo en las distribuciones de probabilidad. Para medir esa dispersión se utiliza la expresión:

$$\sigma^2 = \sum_i (x_i - \mu)^2 \cdot p(x_i) = E(x^2) - E^2(x).$$

$$\sigma = \sqrt{E(x^2) - E^2(x)}$$

Ahora, cuando se refiere a distribuciones de probabilidad se utiliza la letra griega σ para indicar la desviación típica, y σ^2 para la varianza. Recuerda, con frecuencias utilizábamos s y s^2.

La desviación típica y la varianza son teóricas ya que se refieren a una distribución de probabilidad.

Las propiedades que verificaba la media y la desviación típica de las frecuencias se continúan verificando para la esperanza matemática y la desviación típica de las probabilidades.

Actividades resueltas

- *Se lanza una moneda 3 veces y contamos el número de caras. Calcula la desviación típica de la distribución de probabilidad.*

Hacemos un diagrama de árbol y comprobamos que la distribución de probabilidad es:

Número de caras (x):	0	1	2	3
Probabilidad $p(x)$:	1/8	3·(1/8)	3·(1/8)	1/8

Completamos la tabla con las filas siguientes:

					Suma
$x \cdot p(x)$:	0	3/8	6/8	3/8	3/2
x^2:	0	1	4	9	
$x^2 \cdot p(x)$:	0	3/8	12/8	9/8	3

De donde deducimos que:

$$E(x) = \mu = 3/2.$$

$$E(x^2) = 3.$$

$$\sigma^2 = E(x^2) - E^2(x) = 3 - \left(\frac{3}{2}\right)^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

$$\sigma = \sqrt{\frac{3}{4}} = \sqrt{0.75} = 0.87.$$

1.2. Distribución binomial

DISTRIBUCION BINOMIAL

https://www.youtube.com/watch?v=W8y1LhHtuzU

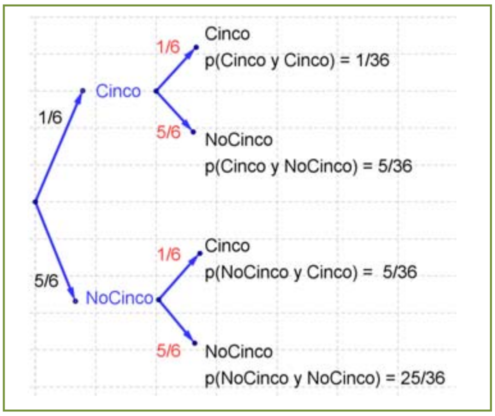
Este apartado está dedicado a describir y caracterizar matemáticamente algún modelo utilizado para variables aleatorias discretas que se repiten con frecuencia en las aplicaciones prácticas. Nos referimos al modelo de probabilidad discreto con más aplicaciones prácticas: la distribución binomial.

Antes de estudiarla vamos a ver dos ejemplos que ya conoces:

Actividad resuelta

- *Se lanza un dado. Llamamos "éxito" a que salga un* 5. *Escribe la distribución de probabilidad.*

Número de "*éxitos*":	0	1
Probabilidad:	5/6	1/6



- *Lanzamos ahora 2 dados.*

Número de "*éxitos*":	0	1	2
Probabilidad:	25/36	5/36 + 5/36 = 10/36	1/36

Observa que:

$$p(0) = \left(\frac{5}{6}\right)^2 = 1 \cdot \left(\frac{5}{6}\right)^2 = \binom{2}{2} \cdot \left(\frac{5}{6}\right)^2$$

$$p(1) = 2 \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right) = \binom{2}{1} \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right) \qquad p(2) = \left(\frac{1}{6}\right)^2 = 1 \cdot \left(\frac{1}{6}\right)^2 = \binom{2}{0} \cdot \left(\frac{1}{6}\right)^2$$

donde $\binom{2}{0}, \binom{2}{1}, \binom{2}{2}$ son los números combinatorios.

Actividad resuelta

- *Se lanza una moneda 3 veces. Llamamos "éxito" a que salga cara. Escribe la distribución de probabilidad.*

Dibujamos el diagrama en árbol y calculamos las probabilidades:

Número de "*éxitos*":	0	1	2	3
Probabilidad:	1/8	3·(1/8)	3·(1/8)	1/8

Observa que:

$$p(0) = \frac{1}{8} = \left(\frac{1}{2}\right)^3 = 1 \cdot \left(\frac{1}{2}\right)^3 = \binom{3}{0} \cdot \left(\frac{1}{2}\right)^3 \qquad p(1) = \frac{3}{8} = 3 \cdot \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^2 = \binom{3}{1} \cdot \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2$$

$$p(2) = \frac{3}{8} = 3 \cdot \left(\frac{1}{2}\right)^3 = \binom{3}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right) \qquad p(3) = \frac{1}{8} = \left(\frac{1}{2}\right)^3 = 1 \cdot \left(\frac{1}{2}\right)^3 = \binom{3}{3} \cdot \left(\frac{1}{2}\right)^3$$

donde $\binom{3}{0}, \binom{3}{1}, \binom{3}{2}, \binom{3}{3}$ son los números combinatorios.

Los dos ejemplos anteriores son de una distribución binomial.

La **distribución binomial** se caracteriza porque puede ser interpretada como un experimento en el que se consideran sucesos dicotómicos, es decir, el de tener "éxito" y el de no tenerlo, de probabilidades p y $q = 1 - p$ respectivamente. Se realiza el experimento n veces, todas independientes y con la misma probabilidad p.

Se representa a la distribución binomial de parámetro p, probabilidad de "éxito", para n, número de pruebas como $B(n, p)$.

En los ejemplos anteriores hemos obtenido que la probabilidad de tener x éxitos en n pruebas repetidas en una distribución binomial $B(n, p)$ es:

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Esta distribución es importante pues aparece en muchas aplicaciones.

Actividades propuestas

6. Se ha comprobado que la distribución de probabilidad del sexo de un recién nacido es:

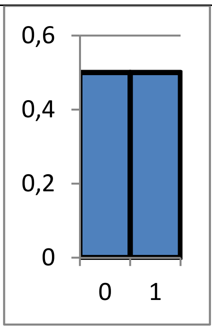
Sexo del recién nacido:	chica	chico
Probabilidad:	0.485	0.515

En un hospital van a nacer hoy 10 bebés. Escribe la expresión de probabilidad de que nazcan 7 chicas.

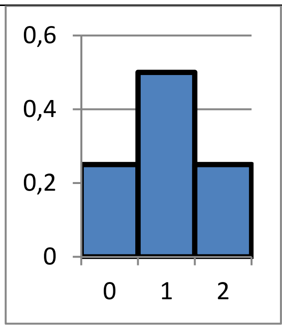
7. Se estima que el porcentaje de hogares que utiliza una determinada marca de tomate frito es del 12 %. En una muestra de 20 hogares, ¿qué probabilidad hay de encontrar entre 6 y 15 que lo utilicen? (No lo calcules, sólo plantea como lo calcularías).

Actividades resueltas

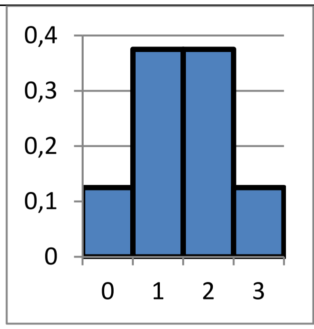
- *Volvemos al problema de lanzar una moneda* n veces. *Dibujamos los histogramas de la distribución de probabilidad binomial.* En este caso $p = q = 1/2$. Y n = 1, 2, 3, 5, 10, 15 y 20.



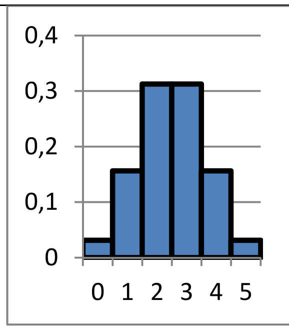
n = 1, B(1, 1/2)



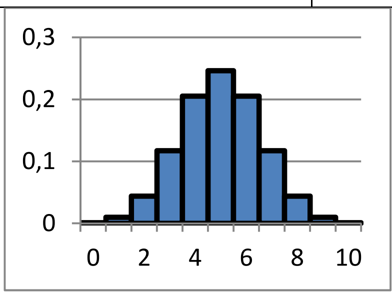
n = 2, B(2, 1/2)



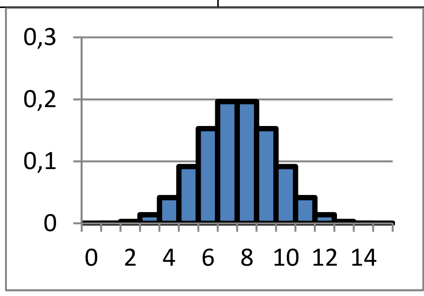
n = 3, B(3, 1/2)



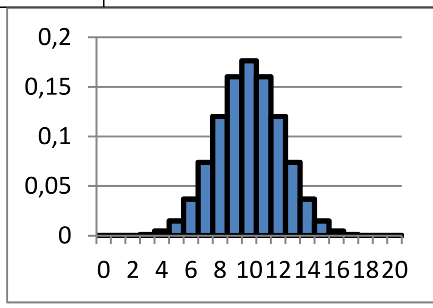
n = 5, B(5, 1/2).



n = 10, B(10, 1/2)



n = 15, B(15, 1/2)

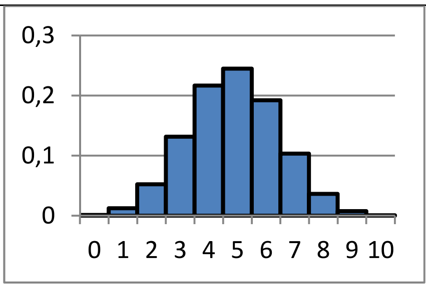


n = 20, B(20, 1/2)

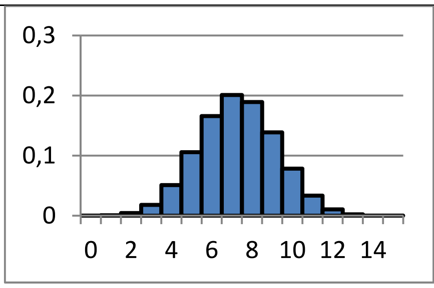
¿Observas alguna diferencia entre los histogramas para n par y para n impar?

En este caso los histogramas son simétricos pues $p = q = 1/2$.

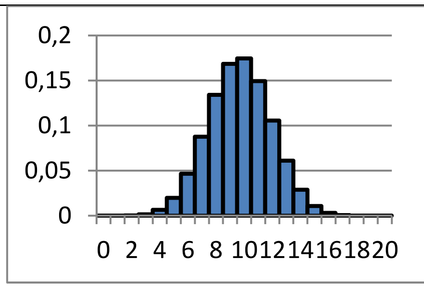
- *Dibujamos los histogramas de la distribución de probabilidad binomial para* n = 10, 15 y 20 *del sexo* *de un recién nacido, donde p* = 0.485 *y por tanto q* = 0.515.



n = 10, B(10, 0.485).



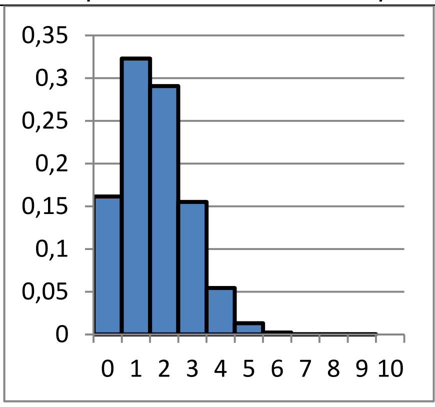
n = 15, B(15, 0.485).



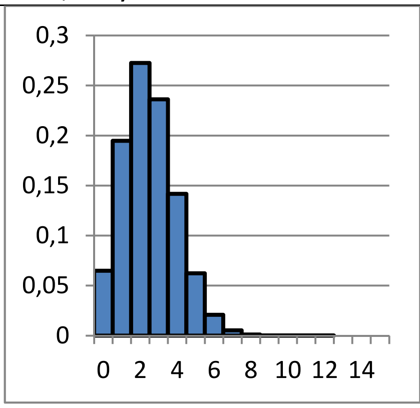
n = 20, B(20, 0.485).

Observa como ahora el histograma no es simétrico.

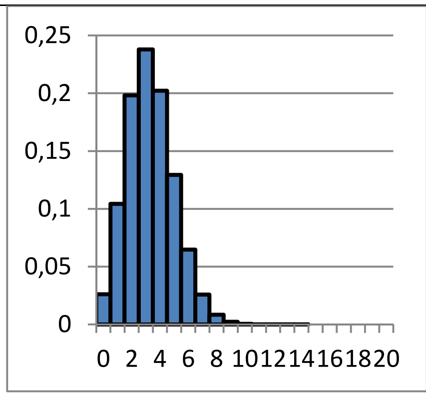
- *Se lanza un dado. Llamamos "éxito" a que salga un 5. Dibujamos los histogramas de la distribución de probabilidad binomial para n* = 10, 15 y 20.



n = 10, B(10, 1/6).



n = 15, B(15, 1/6).



n = 20, B(20, 1/6).

La probabilidad viene indicada por el área bajo es histograma.

Distribución binomial (Ejercicio resuelto). En este video resolveremos un ejercicio utilizando distribución binomial, obteniendo probabilidades de éxito y fracaso, y a partir de la fórmula y de calcular el coeficiente binomial, determinaremos la probabilidad para cierto número de éxitos. Mate Fácil

https://www.youtube.com/watch?v=EisaSQ1j_Kk

Actividades propuestas

8. Lanzamos dos monedas y contamos el número de caras. Calcula la media y la desviación típica de dicho experimento.

9. Observa el histograma del experimento de lanzar una moneda 3 veces. Indica las siguientes probabilidades. A) Probabilidad de que el número de caras sea menor que 1. B) Probabilidad de que el número de caras sea menor o igual a 1.

10. Observa el histograma del experimento de lanzar una moneda 5 veces. Indica las siguientes probabilidades. A) Probabilidad de que el número de caras sea menor que 3. B) Probabilidad de que el número de caras sea menor o igual a 3.

11. Escribe la expresión (no lo calcules) de la probabilidad de que al lanzar una moneda 15 veces el número de caras sea menor que 5.

12. Escribe la expresión (no lo calcules) de la probabilidad de que al lanzar un dado 15 veces el número de cincos sea mayor que 10.

Parámetros de la distribución binomial

Vamos a dar la expresión de los parámetros de una distribución binomial, su media, varianza y desviación típica. No vamos a demostrar sus expresiones, pero si vamos a calcularlas para algunos casos particulares, que generalizaremos.

Imagina una distribución binomial para $n = 1$, con probabilidad de éxito p: $B(1, p)$. Entonces la distribución de probabilidad es:

X	$p(x)$	$x \cdot p(x)$	$x^2 \cdot p(x)$
0	q	0	0
1	p	p	p
Suma	1	$\mu = p$	$E(x^2) = p$

Luego $\mu = p$ y $\sigma^2 = E(x^2) - E^2(x) = p - p^2 = p(1 - p) = p \cdot q$.

Hacemos lo mismo para $n = 2$, con probabilidad de éxito p: $B(2, p)$.

X	$p(x)$	$x \cdot p(x)$	$x^2 \cdot p(x)$
0	q^2	0	0
1	$2\,q \cdot p$	$2\,q \cdot p$	$2\,q \cdot p$
2	p^2	$2\,p^2$	$4\,p^2$
Suma	1	μ	$E(x^2)$

Luego $\mu = 2p$ y $\sigma^2 = E(x^2) - E^2(x) = 2\,q \cdot p + 4\,p^2 - (2p)^2 = 2\,p \cdot q$.

Y ahora para $n = 3$, con probabilidad de éxito p: $B(3, p)$.

X	$p(x)$	$x \cdot p(x)$	$x^2 \cdot p(x)$
0	q^3	0	0
1	$3\,q^2 \cdot p$	$3\,q^2 \cdot p$	$3\,q^2 \cdot p$
2	$3\,q \cdot p^2$	$6\,q \cdot p^2$	$12\,q \cdot p^2$
3	p^3	$3\,p^3$	$9\,p^3$
Suma	1	μ	$E(x^2)$

Luego $\mu = 3\,q^2\,p + 6\,q\,p^2 + 3\,p^3 = 3\,p(q^2 + 2\,q\,p + p^2) = 3\,p(q + p)^2 = 3\,p$ y
$\sigma^2 = E(x^2) - E^2(x) = 3\,q^2\,p + 12\,q\,p^2 + 9\,p^3 - (3p)^2 = 3\,p(q^2 + 2\,q\,p + p^2) + 6\,q\,p^2 + 6\,p^3 - 9p^2 =$
$3\,p(q + p)^2 + 6\,p^2\,(q + p) - 9p^2 = 3\,p - 3p^2 = 3\,p\,(1 - p) = 3\,p \cdot q$.

¡Piensa! Queremos generalizar estos resultados para cualquier valor de n. ¿Cuánto crees que valdrá la media y la varianza?

En efecto:

En una distribución binomial $B(n, p)$ la media vale siempre $E(x) = \mu = n \cdot p$, la varianza $\sigma^2 = n{\cdot}p{\cdot}q$ y la desviación típica $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$.

Actividades propuestas

13. En el control de calidad de bombillas de bajo consume de una fábrica se ha comprobado que el 90 % son buenas. Se toma una muestra de 500 bombillas. Por término medio, ¿cuántas serán de buena calidad? Calcula la media, varianza y desviación típica.

14. En el estudio sobre una nueva medicina para la hepatitis C se ha comprobado que produce curaciones completas en el 80 % de los casos tratados. Se administra a mil nuevos enfermos, ¿cuántas curaciones esperaremos que se produzcan?

1.3. Distribuciones de probabilidad continuas

La distribución binomial se utiliza para describir fenómenos aleatorios discretos: número de caras, número de curaciones, número de bombillas de buena calidad… No tendría sentido decir que se habían obtenido 0.3 cincos al tirar unos dados. Ya sabes que otras variables aleatorias pueden ser continuas, como la estatura de una persona, la medida de una pieza de fabricación… Vamos a estudiar una distribución de probabilidad continua adecuada para estos casos. Hay más distribuciones de probabilidad discretas y continuas, pero la distribución binomial para variables discretas y la distribución normal para variables continuas son las más importantes, las más utilizadas.

Ya hemos analizado las propiedades de las funciones de cuantía de las variables discretas. Las **funciones de densidad de las variables continuas** $f(x)$ deben verificar también una serie de propiedades que estudiarás con más rigor el próximo curso.

Propiedades de la función de densidad $f(x)$:

1) $f(x) \geq 0$. Es natural, pues estamos midiendo probabilidades.
2) El área total bajo la curva debe medir 1. Ya que la probabilidad del suceso seguro es 1.

Propiedades de la función de distribución $F(x)$:

1) $0 \leq F(x) \leq 1$
2) Es una función creciente en todo su dominio de definición
3) $F(x_{\text{máximo}}) = 1$.

Algo que puede sorprenderte es que la probabilidad de que una persona mida exactamente 1.8 metros es 0. ¿Por qué? La razón es que se debe calcular el área de un rectángulo de base 0, y esa área es 0. Es una situación nueva pues hasta ahora parecía que si la probabilidad era nula el suceso era imposible y no es así, lo que se verifica es que si el suceso es imposible entonces la probabilidad es nula.

Tendríamos que calcular esa área en un intervalo, por ejemplo, entre 1.79 y 1.81. Ya sabes que toda medida lleva implícita una cierta imprecisión. Si decimos que Juan mide 1.8 metros como habrá una imprecisión de por ejemplo ± 0.01, estaremos en un cierto intervalo. No estamos diciendo que no sea posible que Juan mida exactamente 1.8, sino que su probabilidad es nula.

Como consecuencia de lo anterior se tiene que:

$$P(c \leq x \leq d) = P(c < x \leq d) = P(c \leq x < d) = P(c < x < d).$$

Para calcular una probabilidad debemos calcular el área bajo la curva $y = f(x)$ función de densidad. Las frecuencias relativas acumuladas se corresponden con lo que denominamos función de distribución de probabilidad, $y = F(x)$.

La función: $F(t) = P(a < x < t)$ es la función de distribución.

Conocida una podemos calcular la otra.

Autora: Raquel Caro
Revisores: David Miranda y Luis Carlos Vidal
Ilustraciones: Banco de Imágenes de INTEF

1.4. Distribución normal

La distribución normal es la distribución más importante tanto en lo que se refiere a la teoría estadística (debido a sus múltiples aplicaciones en inferencia) como en lo que se refiere a sus aplicaciones prácticas. Esta distribución fue propuesta independientemente por *Pierre Simon de Laplace* y *Carl Friedrich Gauss* a finales del siglo XVIII y principios del XIX. Por este motivo, también se la conoce como *distribución de Gauss*. En algunas ocasiones se refiere a ella como *campana de Gauss*, debido a la forma de campana de su función de densidad. Aunque se dice (en broma) que los físicos creen que fue descubierta por un matemático y que los matemáticos opinan que la descubrió un físico.

La expresión de su función de densidad y de su función de distribución es complicada:

$$N(\mu,\sigma) = \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \Phi(x) = \int_{-\infty}^{x} \phi(x) \cdot dx = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

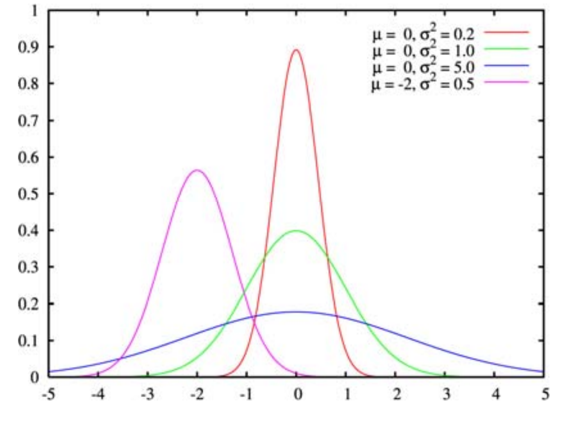
Donde μ es la media y σ la desviación típica. Para denotar que la variable aleatoria X sigue una distribución normal de parámetros μ y σ^2 se escribe $N(\mu, \sigma)$.

¡No te asustes! ¡No vamos a usar integrales! Son expresiones demasiado complicadas, y además, la integral que aparece no es posible resolverla. Y entonces, ¿qué hacemos? Por ejemplo se podría tabular $N(\mu, \sigma)$, pero serían necesarias infinitas tablas, una para cada uno de los posibles valores de μ y de σ.

Utilizando las propiedades de la esperanza matemática y de desviación típica podemos comprobar que basta con tabular una de ellas, la normal de media 0 y desviación típica 1, $N(0, 1)$, que vamos a denominar **distribución normal estándar**. Por tanto, como la función de distribución no puede calcularse analíticamente, hace que los cálculos de probabilidades en la distribución normal se tengan que realizar utilizando tablas que encontraras más adelante.

Dada una variable aleatoria x, de media μ y desviación típica σ se llama **variable aleatoria tipificada** a la variable z, obtenida por $z = \frac{x-\mu}{\sigma}$, con lo que se obtiene una variable aleatoria de media 0 y desviación típica 1.

Observaciones:



Curvas normales. Fuente: Wikipedia

1) La transformación, tipificación, supone una traslación, que cambia el origen de μ a 0, y una contracción o dilatación.
2) Se conservan las áreas bajo ambas curvas, una vez que usemos las variables tipificadas.
3) La variable aleatoria tipificada es adimensional, pues se obtiene dividiendo magnitudes de la misma dimensión, lo que permite poder comparar variables aleatoria diferentes, como estaturas de una población, y pesos de recién nacidos.
4) En la figura del margen puedes observar varias curvas normales, la dibujada en verde es la tipificada. Observa que todas las curvas normales son simétricas, de eje de simetría $x = \mu$ (o $x = 0$ en el caso de $N(0, 1)$). Tienen la media, la moda y la mediana iguales. En los puntos de abscisa $x = \mu - \sigma$ y $x = \mu + \sigma$ tienen un punto de inflexión. Son crecientes hasta $x = \mu$, en ese punto se alcanza un máximo, y decrecientes de $x = \mu$ en adelante.
5) La expresión de la función de densidad tipificada es: $N(0, 1) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$

ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR. N(0, 1)

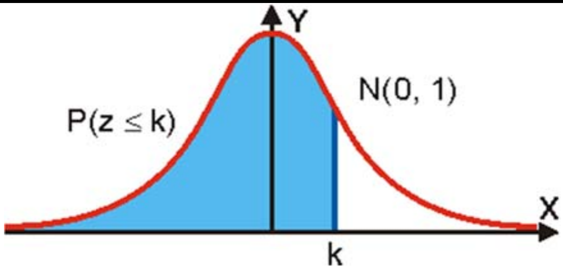


Tabla de la uam: Universidad Autónoma de Madrid

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Vamos ahora a observa con cuidado la tabla para aprender a calcular, con ella, probabilidades.

No están todos los valores. Como el área total bajo la curva es 1, y la curva es simétrica $\phi(-z) = 1 - \phi(z)$.

Introducción a la distribución normal desde cero (Bachillerato). En este vídeo introduzco cómo se calculan las probabilidades con una distribución normal tabulada, como la que aparece en los libros de Bachillerato. La tabla de la distribución normal N (0, 1) aparece al final del vídeo. Carmela Oroa.

https://www.youtube.com/watch?v=-vtLUeGNENY

Actividades resueltas

- *Utiliza la tabla para calcular las probabilidades:* a) $P(z \leq 1)$; b) $P(z \leq 2.46)$; c) $P(z \geq 1)$; d) $P(z \leq -1)$; e) $P(0.5 < z < 1.5)$.

a) $P(z \leq 1)$: Buscamos en la primera columna el 1, y como no tenemos cifras decimales, buscamos en la primera fila el 0. Obtenemos que $P(z \leq 1)$ = 0.8413.

b) $P(z \leq 2.46)$: Hacemos lo mismo, buscamos el 2.4 en la primera columna y el 0.06 en la primera fila. Obtenemos $P(z \leq 2.46)$ = 0.9931

c) $P(z \geq 1)$: Como el área total es 1 y la curva es simétrica. $P(z \geq 1)$ = 1 – $P(z \leq 1)$ = 1 – 0.8413 = 0.1587.

d) $P(z \leq -1)$: Como el área total es 1 y la curva es simétrica. $P(z \leq -1)$ = 1 – $P(z \leq 1)$ =1 – 0.8413 = 0.1587.

e) $P(0.5 < z < 1.5)$: Calculamos $P(0.5 < z < 1.5)$ = $P(z < 1.5)$ – $P(z < 0.5)$. Buscamos en la tabla y obtenemos $P(0.5 < z < 1.5)$ = $P(z < 1.5)$ – $P(z < 0.5)$ = 0.9332 – 0.6915 = 0.2417.

Actividades propuestas

15. Utiliza la tabla de la normal tipificada para calcular:

a) $P(z \leq 0.37)$; b) $P(z < 1.51)$; c) $P(z \geq 0.87)$; d) $P(z \leq -0.87)$; e) $P(0.32 < z < 1.24)$.

Para calcular probabilidades en una $N(\mu, \sigma)$ basta tipificar las variables y buscar las probabilidades en la tabla de $N(0, 1)$.

Actividad resuelta

- *El consumo familiar diario de electricidad (en kW) en cierta ciudad se puede aproximar por una variable aleatoria con distribución normal de media* 5.7 *kW y desviación típica* 1.1 *kW. Calcula la probabilidad de que al tomar una persona al azar su consumo esté comprendido entre* 5 *kW y* 6 *kW.*

Debemos calcular $P(5 < x < 6)$ en una distribución $N(5.7, 1.1)$. Tipificamos las variables:

$z = \frac{x-\mu}{\sigma} = \frac{x-5.7}{1.1}$. por tanto. $z = \frac{x-\mu}{\sigma} = \frac{5-5.7}{1.1} = \frac{-0.7}{1.1} = -0636$ y $z = \frac{x-\mu}{\sigma} = \frac{6-5.7}{1.1} = \frac{0.3}{1.1} = 0.2727$.

Entonces: $P(5 < x < 6)$ = $P(-0.636 < z < 0.2727)$ = $P(z < 0.2727)$ – $P(z < -0.636)$ =

$P(z < 0.2727)$ – (1 – $P(z < 0.636)$) = $P(z < 0.2727)$ – 1 + $P(z < 0.636)$ = 0.6064 – 1 + 0.7389 = 0.3453.

Actividades propuestas

16. Se trata a pacientes con trastorno del sueño con un tratamiento que modela el número de días con una distribución normal de media 290 días y desviación típica 30. Calcula la probabilidad de que al tomar una persona al azar su tratamiento dure más de 300 días.

17. En una estación meteorológica que las precipitaciones anuales de lluvia tienen una media de 450 mm/m^2 con una desviación típica de 80 mm/m^2. Suponemos que la variable aleatoria sigue una distribución normal. Calcula la probabilidad de que: a) Este próximo año la precipitación exceda los 500 mm/m^2. b) La precipitación esté entre 400 y 510 mm/m^2. c) La precipitación sea menor de 300 mm/m^2.

18. En el caso del problema anterior de una $N(450, 80)$ determina la probabilidad de que la variable esté en los intervalos $(\mu - \sigma, \mu + \sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$, $(\mu - 3\sigma, \mu + 3\sigma)$.

El resultado es el mismo para cualquier normal, verificándose que:

$$P(\mu - \sigma < x < \mu + \sigma) = P(-1 < z < 1) = 0.6826;$$

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = P(-2 < z < 2) = 0.9544;$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3) = 0.9974$$

como puedes comprobar calculándolo con la tabla pues $P(-a < x < a) = 2\,P(x < a) - 1$.

En una distribución normal los valores comprendidos entre $[\mu - \sigma, \mu + \sigma]$ se consideran "normales" (desde el punto de vista estadístico. Un año con precipitaciones entre $[\mu + \sigma, \mu + 2\sigma]$ se considera lluvioso. Un año con precipitaciones entre $[\mu + 2\sigma, \mu + 3\sigma]$ se considera muy lluvioso. Un año con precipitaciones entre $[\mu - 2\sigma, \mu - \sigma]$ se considera seco. Un año con precipitaciones entre $[\mu - 3\sigma, \mu - 2\sigma]$ se considera muy seco. Y esto mismo se generaliza para cualquier distribución normal.

video

Distribución normal ejemplo

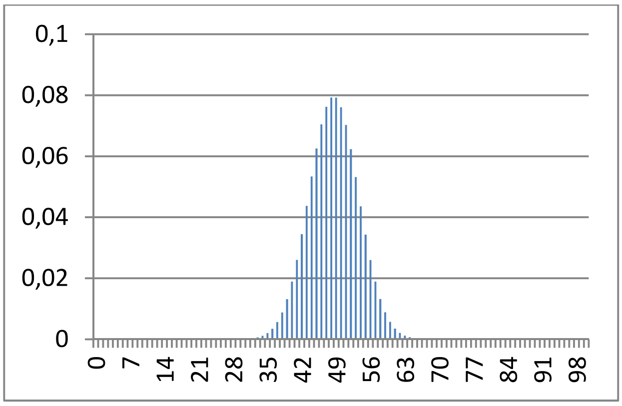
https://www.youtube.com/watch?v=K6arYpJP_k4

Actividades propuestas

19. En una fábrica de coches se hacen pruebas para conocer el tiempo que tardan sus vehículos en alcanzar la velocidad punta. Se considera que esa variable aleatoria tiempo se distribuye según una distribución normal de media 20 s y desviación típica 2 s. Calcula las probabilidades siguientes: a) Que un vehículo alcance su velocidad punta a los 25 s. b) Alcance su velocidad punta en menos de 25 s. c) La alcance entre 18 s y 22s. d) ¿Qué velocidad punta consideras que tendrán los vehículos rápidos? e) ¿Y los lentos?

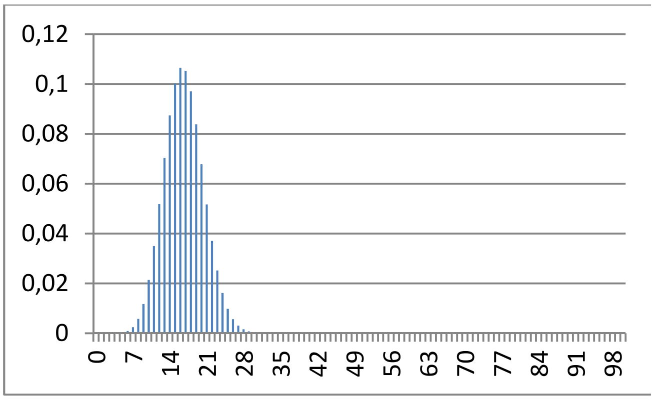
1.5. Aproximación de la binomial a la normal

Hemos visto que la distribución binomial $B(n, p)$ tiene una media $\mu = np$, y una varianza $\sigma = npq$. Queremos analizar en este apartado si la distribución binomial *"se ajusta bien"* a una normal de igual media y desviación típica. Entenderemos que el ajuste es bueno cuando el área bajo la normal en un cierto intervalo sea *casi igual* al área de los rectángulos de la binomial.



B(100. 0.485)

Al estudiar la distribución binomial representamos muchos histogramas de distintas binomiales donde puedes observar que, incluso para valores de n bajos, el ajuste no es malo. Representamos el histograma de $B(100, 0.485)$ sobre el sexo de los bebés y parece que el ajuste es muy bueno. Al margen puedes observar el histograma del experimento tirar 100 dados y contar el número de cincos: $B(100, 1/6)$ que resultaba muy asimétrico. ¿Qué opinas? ¿Se ajustan a la normal?



B(100. 1/6)

Otra forma de hacer la comparación podría ser comparar las áreas en determinados intervalos entre la curva normal y el histograma de la distribución binomial. Por ejemplo para $B(3, 1/2)$ para $x = 1$ calculamos el área bajo el histograma para el intervalo (0.5, 1.5) que es 0.38. La media es $\mu = 3/2 = 1.5$ y $\sigma = \sqrt{3 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{\frac{3}{4}} = 0.866$. Tipificamos la normal $N(3/2, 0.866)$ y calculamos:

$$P(0.5 < x < 1.5) = P(\frac{0.5-1.5}{0.866} < z < \frac{1.5-1.5}{0.866}) = P(-1.1547 < z < 0) = P(z < 0) - P(z < -1.1547) =$$

$$P(z < 0) - (1 - P(z < 1.1547)) = P(z < 0) + P(z < 1.1547) - 1 = 0.5 + 0.8749 - 1 = 0.3749.$$

Hasta en este caso tan desfavorable el ajuste es bueno.

Se puede demostrar que el ajuste es bueno entre binomial y normal cuando $npq \geq 9$.

Al estudiar la distribución binomial no hicimos los cálculos en muchos de los ejercicios pues eran muy laboriosos. Sin embargo, mirar la tabla de la normal es bastante más rápido y sencillo.

Observa también que no hemos tomado el valor $x = 1$ pues para tomar intervalos le hemos restado a 1 y sumado a 1 la longitud del intervalo: 0.5, y hemos tomado el intervalo (0.5, 1.5).

Actividad resuelta

- *En una determinada población se divide la población activa en dos grupos, los que trabajan en agricultura y servicios que son un 44 %, y el resto. Se elige al azar una muestra de 200 personas entre la población activa, ¿qué probabilidad hay de que haya entre 80 y 100 personas del primer grupo?*

Es un problema de distribución binomial $B(200, 0.44)$ pues una persona o pertenece a dicho grupo, o no pertenece. Sabemos que:

$$p(x) == \binom{n}{x} \cdot p^x \cdot q^{n-x} = \binom{200}{x} \cdot 0.44^x \cdot 0.56^{200-x}$$

Y deberíamos calcular:

$$P(80 \le x \le 100) = \sum_{x=80}^{x=100} p(x) = \sum_{x=80}^{x=100} \binom{200}{x} \cdot 0.44^x \cdot 0.56^{200-x}.$$

Habíamos advertido que el cálculo era laborioso, pero ahora podemos utilizar el ajuste de la binomial a la normal. Calculamos la media y la desviación típica:

$$\mu = np = 200 \cdot 0.44 = 88$$

$$\sigma = \sqrt{200 \cdot 0.44 \cdot 0.56} = 7.02\ .$$

por lo que ajustamos con la normal $N(88, 7.02)$.

Como la longitud de cada intervalo es 1, se añade a cada valor 0.5 para ir desde el extremo del intervalo, y no desde el centro.

$$P(80 - 0.5 \le x \le 100 + 0.5)$$

Ahora tipificamos:

$$P(\frac{80 - 88 - 0.5}{7.02} \le z \le \frac{100 - 88 + 0.5}{7.02}) = P(z \le 1.78) + P(z \le 1.21) - 1 = 0.9625 + 0.8869 - 1 = 0.8494.$$

En el 85 % de los casos habrá entre 80 y 100 personas del primer grupo.

Como npq = 49.28 ≥ 9. el ajuste es bueno.

Actividades propuestas

20. Se lanza una moneda mil veces, ¿cuál es la probabilidad de que el número de caras obtenidas esté entre 400 y 600? ¿Y de que sea mayor que 800?

21. En una fábrica de bombillas de bajo consumo se sabe que el 70 % de ellas tienen una vida media superior a 1 000 horas. Se toma una muestra de 50 bombillas, ¿cuál es la probabilidad de que haya entre 20 y 30 cuya vida media sea superior a mil horas? ¿y la probabilidad de que haya más de 45 cuya vida media sea superior a 1 000 horas?

22. Una compañía aérea ha estudiado que el 5 % de las personas que reservan un billete para un vuelo no se presentan, por lo que venden más billetes que las plazas disponibles. Un determinado avión de la compañía tiene 260 plazas (con lo que suelen reservar hasta 270). Calcula la probabilidad de que lleguen 260 pasajeros. En 500 vuelos de dicho avión, ¿en cuántos consideras que habrá exceso de pasajeros?

1.6. Intervalos de confianza

Queremos ahora resolver otro tipo de problema. En lugar de calcular la probabilidad de un intervalo dado queremos encontrar intervalos con una probabilidad dada.

Utilizaremos una actividad anterior.

Actividades resueltas

- *En una determinada población se divide la población activa en dos grupos, los que trabajan en agricultura y servicios que son un 44 %, y el resto. Se elige al azar una muestra de 200 personas entre la población activa y queremos conocer cuántas pertenecerán al primer grupo con una probabilidad del 0.99.*

Habrá muchos intervalos que resuelvan el problema, pero nos van a interesar intervalos simétricos con respecto a la media. Recuerda $\mu = np = 200 \cdot 0.44 = 88$ y $\sigma = \sqrt{200 \cdot 0.44 \cdot 0.56} = 7.02$, por lo que ajustamos con la binomial $B(200, 0.44)$ con la normal $N(88, 7.02)$.

Vamos a tener en cuenta que la longitud de cada intervalo de la binomial es 1, luego vamos a añadir 0.5 a cada lado.

$$0.99 = P(88 - 0.5 - k \leq x \leq 88 + 0.5 + k) = P(\frac{88 - 0.5 - k - 88}{7.02} \leq z \leq \frac{88 + 0.5 + k - 88}{7.02}) =$$
$$P(\frac{-0.5 - k}{7.02} \leq z \leq \frac{0.5 + k}{7.02}) = P(z \leq \frac{0.5 + k}{7.02}) + P(z \leq \frac{-0.5 - k}{7.02}) - 1 = 2P(z \leq \frac{0.5 + k}{7.02}) - 1$$

Despejamos:

$$P(z \leq \frac{0.5+k}{7.02}) = \frac{0.99+1}{2} = 0.995.$$

Buscamos ese valor en la tabla de la curva normal estándar, y obtenemos 2.58, por lo tanto, $\frac{0.5+k}{7.02} = 2.58$ de donde $k = 17.61 \approx 18$, por lo que el intervalo buscado es:

$$(88 - 18, 88 + 18) = (70, 106).$$

Volvemos al problema de las encuestas de votos.

Actividad resuelta

- *En una población de 8 millones de votantes elegimos una muestra aleatoria de 2 000 de la que 700 personas nos afirman que van a votar a un determinado partido. ¿Qué podemos asegurar sobre el número de votos que recibirá dicho partido?*

Como 700/2 000 = 35, una primera respuesta podría ser que 0.35 · 8 000 000 = 2 800 000 votos, pero, ¿qué confianza podemos tener de ese resultado.

Fijamos un nivel de significación α, o un grado de confianza, $1 - \alpha$. Sea $\alpha = 0.05$ y $1 - \alpha = 0.95$.

Sea p la proporción de votantes al partido estudiado. Tenemos una distribución binomial de media $\mu = np = 2\,000 \cdot p$ y $\sigma = \sqrt{npq} = \sqrt{2\,000 \cdot p(1-p)}$. Calculamos la probabilidad de que el número de votantes al partido estudiado de la muestra sea:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 0.95$$

Pasamos de la distribución binomial a la normal para calcular k y p:

$$P(\mu - k\sigma - 0.5 \leq X \leq \mu + k\sigma + 0.5) \geq 0.95$$

Tipificamos:

$$P(\frac{-k\sigma-0.5}{\sigma} \leq Z \leq \frac{+k\sigma+0.5}{\sigma}) \geq 0.95.$$

Obtenemos que $z = \frac{+k\sigma+0.5}{\sigma} \geq 1.96$, por lo que $k\sigma + 0.5 \geq 1.96\sigma$. Debemos sustituir μ y α en función de p como se hizo anteriormente y se obtiene que: $0.3280 \leq p \leq 0.3719$, es decir que la proporción de votantes debe estar entre el 33 % y el 37 %.

Actividades propuestas

23. Rehaz los cálculos de la actividad anterior para un nivel de confianza del 99 %.

24. Se investigan los hábitos de consumo de una población de dos millones de personas. Se pasa una encuesta a mil personas y se les pregunta si en su domicilio se cocina con gas, de los que 600 responden afirmativamente. Qué puedes afirmar sobre el número de personas en las que en su domicilio se usa gas con un nivel de confianza del 95 %.

25. Se lanza 600 veces un dado y contamos el número de 5s. a) ¿Cuál es el intervalo simétrico respecto de la media con una probabilidad de 0.99? b) Lo mismo con una probabilidad del 0.6.

26. En una actividad anterior vimos que en una compañía aérea se había estudiado que el 5 % de las personas que reservan un billete para un vuelo no se presentan. Un determinado avión de la compañía tiene 260 plazas. ¿Qué número de reservas n puede aceptar la compañía admitiendo una probabilidad del 0.02 para que el número de reservas supere al número de plazas.

(*Ayuda*: Busca una binomial tal que $p(x > 260) < 0.02 \Rightarrow p(x \leq 260) = 1 - p(x > 260) \geq 0.98$).

CURIOSIDADES. REVISTA

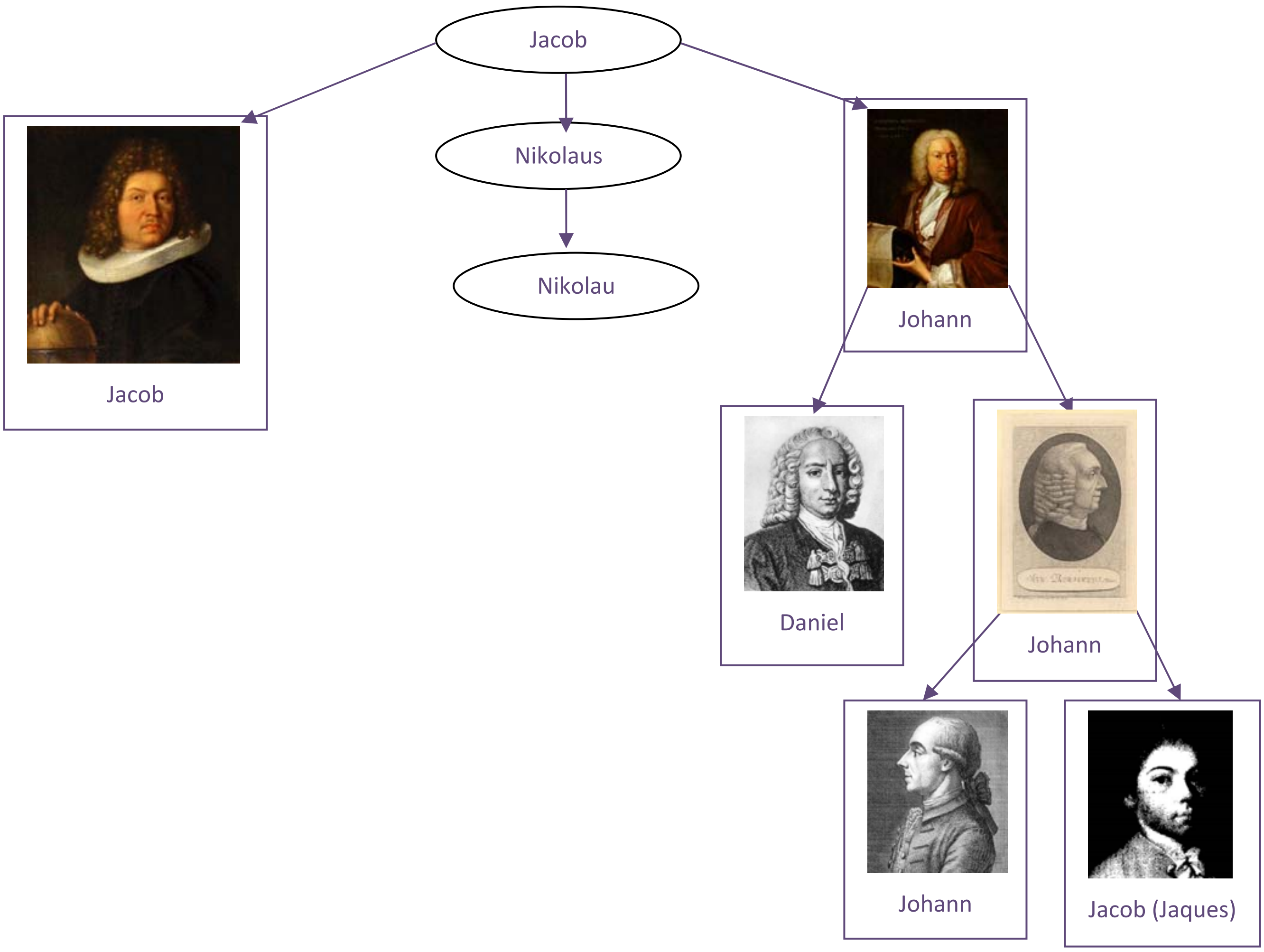
La saga de los *Bernoulli*

Si te dicen: "*Bernoulli hizo esto o descubrió aquello*" tu puedes preguntar:

- ¿Cuál *Bernoulli*?

Y es que hubo una familia suiza del siglo XVII en la que hubo padres, hijos, tíos y sobrinos, muchos de ellos matemáticos y físicos con importantes descubrimientos.

El primero de ellos, Jacob *el viejo*, nació en Amberes, Bélgica, pero huyendo de una persecución religiosa pues era hugonote, se trasladó a vivir a Basilea (Suiza) el año 1622. Tuvo un único hijo, *Nikolaus*, que tuvo varios hijos, dos de ellos matemáticos famosos, *Jacob* (1654–1705) y *Johann* (1667–1748). El primero dio su nombre a los *números de Bernoulli*, y el segundo trabajó en cálculo infinitesimal. Otro de los hijos de *Nikolaus*, de nombre *Nikolau* (1687–1759), también fue matemático. *Johann* tuvo varios hijos, entre ellos, *Daniel* (1700–1782) que desarrolló en *principio de Bernoulli*, y *Johann* (1710–1790), que a su vez también tuvo varios hijos matemáticos, como *Johann* (1744–1807) y *Jacob* (1759–1789), también conocido como *Jaques*, del que recibe el nombre la *distribución de Bernoulli*.



Distribución de *Bernoulli*

Se llama distribución de *Bernoulli* a una distribución con sólo dos posibilidades, "éxito" o "no éxito". Por ejemplo:

- Tirar una moneda y ver si sale cara
- Tirar un dado y ver si sale un 5.
- Tirar al blanco…

Distribución de Binomial

Si consideramos n variables aleatorias idénticas que siguen una distribución de *Bernoulli*, la variable aleatoria suma sigue una distribución binomial. Por ejemplo:

- Tirar una moneda 100 veces y contar el número de caras.
- Tirar un dado mil veces y contar el número de cincos.
- Tirar al blanco 20 veces y contar el número de éxitos.

Debe verificarse que la probabilidad sea siempre la misma y que los sucesos sean independientes.

No es una distribución binomial contar el número de bolas rojas que sacamos en 5 extracciones de una bolsa con 10 bolas rojas y 12 bolas de otro color. si la extracción es SIN reemplazamiento. pues la probabilidad va cambiando.

Distribución Normal

La **importancia** de esta distribución se debe a que se utiliza para modelar numerosos fenómenos naturales, médicos y sociales. Son fenómenos en los que influyen muchas variables difíciles de controlar, por lo que podemos suponer que es suma de distintas causas independientes.

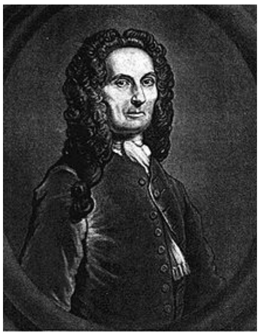
Ejemplos clásicos de fenómenos que se distribuyen según una normal son:

- Fenómenos morfológicos como la estatura o el peso
- Fisiológicos como los efectos de un fármaco
- Sociológicos como los de consumo
- Psicológicos como el cociente intelectual
- El ruido en las telecomunicaciones
- Los errores cometidos al medir una magnitud…

La **historia** de la distribución normal: Aparece por primera vez con *Abraham de Moivre* en un artículo publicado en 1733, sobre la distribución binomial para valores grandes de n.

El resultado fue trabajado por *Laplace* en su libro sobre la teoría de las probabilidades trabajando sobre errores.

También sobre errores la utilizó *Gauss*, analizando datos astronómicos. En su honor también se denomina a la curva normal, *campana de Gauss*.

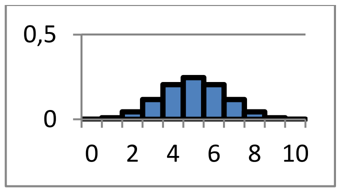
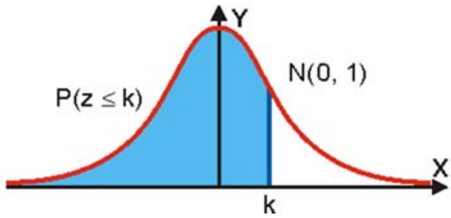


Moivre

Laplace

Gauss

RESUMEN

<table>
<tr><td>Propiedades de función de cuantía</td><td>1) $p(x) \geq 0$
2) $\Sigma\, p(x) = 1$.</td><td rowspan="2">Lanzamos dos monedas y contamos el número de caras:<table><tr><td>Número de caras (x):</td><td>0</td><td>1</td><td>2</td></tr><tr><td>Función de cuantía ($p(x)$):</td><td>1/4</td><td>1/2</td><td>1/4</td></tr><tr><td>Función de distribución $F(x)$:</td><td>1/4</td><td>3/4</td><td>4/4</td></tr></table></td></tr>
<tr><td>Propiedades de función de distribución</td><td>1) $0 \leq F(x) \leq 1$
2) $F(x)$ es una función creciente
3) $F(x_{Máximo}) = 1$</td></tr>
<tr><td>Esperanza matemática</td><td>$$E(x) = \mu = \sum_i x_i \cdot p(x_i)$$</td><td>$\mu = 0 \cdot (1/4) + 1 \cdot (1/2) + 2 \cdot (1/4) = 1$</td></tr>
<tr><td>Varianza y desviación típica</td><td>$$\sigma^2 = \sum_i (x_i - \mu)^2 \cdot p(x_i) = E(x^2) - E^2(x)$$ $$\sigma = \sqrt{E(x^2) - E^2(x)}$$</td><td>$\sigma^2 = (0-1)^2 \cdot (1/4) + (1-1)^2 \cdot (1/2) + (2-1)^2 \cdot (1/4) = 1/2$.
$$\sigma = \sqrt{1/2}$$</td></tr>
<tr><td>Distribución binomial</td><td>$$B(n,p) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$ $$E(x) = \mu = n \cdot p.$$ $$\sigma^2 = n \cdot p \cdot q = n \cdot p \cdot (1-p)$$</td><td>
$B(10, 1/2)$.</td></tr>
<tr><td>Distribución normal</td><td>$$N(\mu,\sigma) = \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$</td><td></td></tr>
<tr><td>Aproximación de la binomial a la normal</td><td>Una binomial con $npq \geq 9$ se considera se ajusta bien a una normal de igual media y desviación típica.</td><td></td></tr>
</table>

Autora: Raquel Caro
Revisores: David Miranda y Luis Carlos Vidal
Ilustraciones: Banco de Imágenes de INTEF

EJERCICIOS Y PROBLEMAS

1. Se lanza un dado tres veces y se cuanta el número de treses que aparecen. Dibuja el histograma, la función de cuantía y la función de distribución. Calcula la media y la desviación típica.

2. Lanzamos 4 monedas. Por cada cara que salga ganamos 5 euros, pero debemos pagar 3 euros por jugar. ¿Cuánto esperas ganar en una jugada? ¿Y en 20 jugadas? ¿Y en 100 jugadas?

3. Disponemos de dos urnas, la primera con 6 bolas idénticas numeradas del 1 al 6; la segunda con 4 bolas idénticas numeradas del 1 al 4. Sacamos a la vez una bola de cada urna, y consideramos la variable aleatoria, "suma de puntos obtenidos". A) Calcula la distribución de probabilidad y dibuja el histograma correspondiente. B) Si sacamos más de 5 puntos ganamos 10 euros, y en caso contrario perdemos la misma cantidad. ¿Es un juego equitativo?

4. La población activa de un cierto país se puede dividir en los que tienen estudios superiores y los que no los tienen, siendo el primero de un 20 %. Elegimos 10 personas de la población activa al azar. Escribe la expresión de todas las posibilidades y sus probabilidades. Calcula la probabilidad de que haya 9 o 10 que tengan estudios superiores.

5. Si $p(x)$ es la probabilidad de tener x éxitos en una distribución binomial $B(n, p)$, y $p(x + 1)$ es la de obtener x+1 éxitos, comprueba que se verifica la siguiente relación recurrente:

$$p(x+1) = \frac{p(x)}{x+1}(n-x)\frac{p}{q}$$

6. En una ruleta hay 37 números numerados del 0 al 36, de los cuales 18 son pares y 18 impares. Si sale el 0 gana la banca. Jugamos al dos por 1 a impar, apostamos 10 euros a impar, y la banca nos paga 20 euros si sale un impar, y se queda con nuestros 10 euros si no sale. ¿Te parece un juego equitativo?

7. Juego de San Petersburgo: Se lanza una moneda no trucada hasta que aparece cara. Si sale en el primer lanzamiento, se ganan 10 euros, si en el segundo, 20, si en el tercero, 40, … y en el n-ésimo, $10 \cdot 2^{n+1}$. Calcula la ganancia media si sólo se puede lanzar 5 veces la moneda. ¿Y si se puede lanzar 10 veces?

8. Lanzamos un dado no trucado mil veces y contamos el número de 5, ¿qué número de éxitos esperamos con una probabilidad no inferior al 0.95, es decir, en el intervalo media menos dos veces la desviación típica y media más dos veces la desviación típica?

9. En una distribución binomial $B(10, 0.3)$ calcula $P(x = 0)$, $P(x \neq 0)$, $P(x = 10)$ y $P(x = 7)$. Determina también la media y la desviación típica.

10. Lanzamos 5 monedas, calcula las probabilidades de obtener:

a) 0 caras. b) 1 cara. c) 2 caras. d) 3 caras.

11. Calcula en una distribución normal estándar las probabilidades siguientes:

a) $P(z = 0)$. b) $P(z < 0)$. c) $P(z = 1.82)$. d) $P(z > 1.82)$.

12. Calcula en una distribución normal estándar las probabilidades siguientes:

a) $P(z > 4)$. b) $P(z < 4)$. c) $P(z > 1)$. d) $P(z < 1)$.

Autora: Raquel Caro
Revisores: David Miranda y Luis Carlos Vidal
Ilustraciones: Banco de Imágenes de INTEF

13. Calcula en una distribución normal estándar las probabilidades siguientes:

a) $P(1 < z < 2)$. b) $P(-1.3 < z < 4)$. c) $P(-0.2 < z < 2.34)$. d) $P(-1 < z < 1)$.

14. Calcula en una distribución normal $N(1, 2)$ las probabilidades siguientes:

a) $P(x > 4)$. b) $P(x < 4)$. c) $P(x > 1)$. d) $P(x < 1)$.

15. Calcula en una distribución normal $N(0.5, 0.2)$ las probabilidades siguientes:

a) $P(x > 4)$. b) $P(x < 4)$. c) $P(x > 1)$. d) $P(x < 1)$.

16. Calcula en una distribución normal $N(1, 1/2)$ las probabilidades siguientes:

a) $P(1 < x < 2)$. b) $P(-1.3 < x < 4)$. c) $P(-0.2 < x < 2.34)$. d) $P(-1 < x < 3)$.

17. En una distribución binomial $B(10, 0.3)$ calcula la media y la desviación típica, y mediante la aproximación a la normal determina $P(x = 0)$, $P(x \neq 0)$, $P(x = 10)$ y $P(x = 7)$. Compara con los resultados obtenidos en el ejercicio 9.

18. En una distribución binomial $B(100, 0.4)$ calcula la media y la desviación típica, y mediante la aproximación a la normal determina $P(x > 40)$, $P(x \leq 50)$, $P(x \geq 50)$ y $P(40 \leq x \leq 50)$.

19. En una distribución binomial $B(1000, 0.5)$ calcula la media y la desviación típica, y mediante la aproximación a la normal determina $P(x < 200)$, $P(x = 150)$, $P(x < 150)$ y $P(50 \leq x \leq 150)$.

20. En una distribución binomial $B(1000, 0.05)$ calcula la media y la desviación típica, y mediante la aproximación a la normal determina $P(x > 200)$, $P(x = 200)$, $P(x < 200)$ y $P(50 \leq x \leq 200)$.

21. Una fábrica de móviles ha comprobado que el 1 % de los que fabrica son defectuosos. En un control de calidad se toman 10 móviles al azar. Calcula la media y la desviación típica. Calcula la probabilidad de que haya más de 2 móviles defectuosos.

22. La probabilidad de que María gane a Raquel en una partida es de 0.4. Juegan 6 partidas. Calcula la probabilidad de que:

a) María gane alguna vez.
b) Raquel gane al menos una vez.
c) Raquel gane más de la mitad de las partidas.
d) María gane 2 partidas.

23. Las estaturas de las personas de una cierta población se distribuyen según una normal de media 180 cm y desviación típica 15 cm. Determina las probabilidad de que:

a) Una persona tenga una estatura superior a 190 cm.
b) Una persona tenga una estatura menor a 160 cm.
c) ¿Qué proporción de personas tienen una estatura comprendida entre 160 cm y 190 cm?

24. En un examen para entrar en un cuerpo del Estado se sabe que los puntos obtenidos se distribuyen según una normal de media 100 y desviación típica 10 puntos. Determina la probabilidad de que:

a) Un opositor obtenga 120 puntos.
b) Si para aprobar es necesario tener más de 120 puntos. ¿Qué porcentaje de opositores aprueban?
c) Si aprueban únicamente los que están entre el 20 % de los mejores, ¿cuántos puntos de obtener un opositor para aprobar?

AUTOEVALUACIÓN

1. Se lanza un dado tres veces y se anota el número de cuatros que aparecen. La distribución de probabilidad que tenemos es:

a) $B(4, 1/6)$ b) $B(4, 1/4)$ c) $B(3, 1/6)$ d) $B(3, 5/6)$

2. En la distribución anterior, la media es:

a) $\mu = 4/6$ b) $\mu = 1/2$ c) $\mu = 15/6$ d) $\mu = 1$

3. Y la varianza es:

a) $\sigma^2 = 15/12$ b) $\sigma^2 = 5/6$ c) $\sigma^2 = 1/36$ d) $\sigma^2 = 5/12$

4. Utiliza la tabla de la distribución normal estándar para calcular la probabilidad $P(z \leq 2.02)$, que vale:

a) $P(z \leq 2.02) = 0.0217$ b) $P(z \leq 2.02) = 0.9772$ c) $P(z \leq 2.02) = 0.0228$ d) $P(z \leq 2.02) = 0.9783$

5. Utiliza la tabla de la distribución normal estándar para calcular la probabilidad $P(0.5 < z < 1.5)$. que vale:

a) 0.3417 b) 0.9332 c) 0.6915 d) 0.2742

6. Sin mirar la tabla, ni tipificar la variable, la probabilidad de $P(x < \mu)$ es:

a) –0.4 b) 0.5 c) 0.6 d) No puede saberse

7. En una distribución binomial $B(10, 0.3)$ el valor de $P(x = 0)$ es:

a) 0.11 b) 0.0198 c) 0.00001024 d) 0.8

8. El 2 % de las pastillas de freno fabricadas se sabe que son defectuosas. En una caja con 2 000 pastillas, la probabilidad de que haya menos de 50 defectuosas es:

a) 0.6011 b) 0.7635 c) 0.9357 d) 0.8655

9. Una fábrica de ordenadores ha comprobado que el 5 % de los que fabrica son defectuosos. En un control de calidad se toman 10 ordenadores al azar. Determina si la probabilidad de que no haya ninguno defectuoso es:

a) 0.5987 b) 0.4027 c) 0.9357 d) 0.8074

10. La probabilidad de que María gane a Raquel en una partida es 2/3. Juegan 4 partidas. Determina si la probabilidad de que María gane alguna vez es:

a) 0.0123 b) 0.5 c) 0.8972 d) 0.9877