MD.PlantillaTexto(02)Esp.dot

DISTRIBUCIONES DE PROBABILIDAD

UD000578_V(02)

ÍNDICE

MOTIVACIÓN

¿Sabías que los estudios en los que se dice que una población o un grupo de personas se comportan de una manera o de otra se realizan gracias a las aplicaciones de las distribuciones de probabilidad? Efectivamente, gracias a las distribuciones de probabilidad podemos hacernos una idea de cómo se comporta una población o muestra y de qué posibilidades existen para que lo haga de una manera o de otra.

Dependiendo de la variable a estudiar se realizará con una distribución de probabilidad u otra.

Existen diversas distribuciones de probabilidad para variables discretas y para variables continuas. En la unidad didáctica que vamos a tratar estudiaremos la distribución binomial en variables discretas y la distribución normal para variables continuas.

PROPÓSITOS

Cuando finalices el estudio de esta unidad didáctica, conseguirás:

- Distinguir entre variable aleatoria discreta y variable aleatoria continua.
- Conocer y calcular la función de distribución y la función de densidad de una variable aleatoria.
- Interpretar los parámetros de una variable aleatoria binomial.
- Interpretar los parámetros de una variable aleatoria normal.
- Entender la tipificación de variables normales y la distribución normal estándar.
- Utilizar las tablas para las distribuciones de probabilidad.

PREPARACIÓN PARA LA UNIDAD

En esta unidad didáctica desarrollaremos el concepto de variable aleatoria distinguiendo entre variables aleatorias discretas y continuas. Nos centraremos en la variable aleatoria binomial como ejemplo de variable aleatoria discreta, y en la variable aleatoria normal como ejemplo de variable aleatoria continua. Además, aprendemos a realizar diversas transformaciones y aproximaciones entre variables, lo cual nos permitirá resolver algunos problemas con mayor facilidad.

A menudo se confunden unas variables con otras, pero como se podrá observar, cada tipo de variable posee unas características y parámetros específicos, que nos van a permitir distinguirlas.

1. VARIABLE ALEATORIA

Una **variable aleatoria** es la cualidad que se desea conocer en un estudio estadístico. No podemos conocer a priori el valor que va adoptar; para hacernos una idea, una variable es cualquier cosa que se pueda medir. Puede tomar distintos valores.

Las variables aleatorias sirven para poder representar, en el "papel", un fenómeno o realidad.

Por ejemplo, la temperatura ambiental es un fenómeno climático, una realidad que nosotros conseguimos representar mediante una variable aleatoria que sería, por ejemplo, el número de grados centígrados que hay en un determinado lugar, en un determinado momento.

El número de grados centígrados es una variable aleatoria en cuanto a que es medible, y toma distintos valores según el momento y el lugar.

Una primera clasificación de las variables aleatorias sería la siguiente:

- **Cuantitativas:** hacen referencia a un aspecto que podemos medir y se puede cuantificar.
- **Cualitativas:** hacen referencia a una cualidad del individuo; son aspectos que no podemos cuantificar a priori: sexo, profesión, etc.

Las variables aleatorias se clasifican también en:

- **Discretas:** toman valores enteros, por ejemplo, la variable aleatoria "número de puntos obtenidos al lanzar un dado", puesto que solo toma los valores enteros 1, 2, 3, 4, 5 y 6.
- **Continuas:** toman valores enteros, decimales, positivos o negativos, por ejemplo, la variable temperatura.

1.1. ESTUDIO DE DISTRIBUCIONES DE PROBABILIDAD DE VARIABLES ALEATORIAS DISCRETAS

Aunque el concepto de función de probabilidad es el mismo independientemente de si la variable estadística es discreta o continua, esta función presenta características muy diferentes según nos encontremos en una u otra situación.

1.1.1. FUNCIÓN DE PROBABILIDAD

La función de probabilidad es la función asociada a una variable. Esta función establece una relación entre cada valor que toma la variable y su probabilidad.

Dado que la función es discreta, la variable solo toma un conjunto de valores enteros, por lo que la función de distribución no será continua.

Reto

Completa y realiza el siguiente ejercicio.

Una variable aleatoria es ______[1] cuando solo puede tomar unos ciertos valores _____[2]. Una **variable aleatoria** es toda función que asocia a cada elemento del espacio ______[3] un número ____[4].

Solución:

[1] Discreta.
[2] Enteros.
[3] Muestral.
[4] Real.

1.1.2. Media y varianza de una variable aleatoria discreta

Se llama media de una variable aleatoria X, que toma los valores $x_1, x_2..., x_n$, con probabilidades $p_1, p_2..., p_n$, respectivamente, al valor de la siguiente expresión:

$$\mu = \sum_{i=1}^{n} x_i \cdot p_i$$

A la media también se la llama esperanza matemática, y se representa con E[X].

Nota

Fíjate que la definición de media es análoga a la definición para una distribución de frecuencias.

Se llama varianza de una variable aleatoria X, que toma los valores $x_1, x_2..., x_n$, con probabilidades $p_1, p_2..., p_n$, respectivamente, al valor de la siguiente expresión:

$$\sigma^2 = \sum x_i^2 \cdot p_i - \mu^2 = \sum (x_i - \mu)^2 \cdot p_i$$

Reto

Realiza el siguiente ejercicio.

¿Qué otro nombre tiene también la media?

Solución:

Esperanza matemática.

La raíz cuadrada positiva de la varianza se llama desviación típica, y se representa por σ.

Recuerda

Recuerda que tienes a tu disposición un servicio de tutorías para resolver cualquier duda que te pueda surgir, bien mediante carta, Campus Virtual, fax o llamada telefónica.

Ejemplo:

La ley de probabilidad de una variable aleatoria discreta viene dada por la tabla siguiente:

x_i	0	1	2	3	4
p_i	0,1	a	b	c	0,2

Sabiendo que $p(X \leq 2) = 0{,}7$ y $p(X \geq 2) = 0{,}75$, calcular: la esperanza matemática y la desviación típica.

En primer lugar, calculemos los valores de a, b y c, teniendo en cuenta que la suma de todas las probabilidades es igual a la unidad.

$$0{,}1 + a + b + c + 0{,}2 = 1$$

De $p(X \leq 2) = 0{,}7$, se deduce que: $0{,}1 + a + b = 0{,}7$.

De $p(X \geq 2) = 0{,}75$, se deduce que: $b + c + 0{,}2 = 0{,}75$.

Resolviendo el sistema:

$$\begin{cases} 0{,}1 + a + b + c + 0{,}2 = 1 \\ 0{,}1 + a + b = 0{,}7 \\ b + c + 0{,}2 = 0{,}75 \end{cases}$$

$$\Rightarrow \; a = 0{,}15,\; b = 0{,}45,\; c = 0{,}1$$

Para el cálculo de la esperanza y la varianza formaremos la siguiente tabla:

x_i	p_i	$x_i \cdot p_i$	$x_i^2 \cdot p_i$
0	0,1	0	0
1	0,15	0,15	0,15
2	0,45	0,9	1,8
3	0,1	0,3	0,9
4	0,2	0,8	3,2
	1	2,15	6,05

Luego la esperanza matemática es: μ = 2,15.

Desviación típica es:

$$\sigma = \sqrt{6{,}05 - 2{,}15^2} = 1{,}19.$$

1.1.3. FUNCIÓN DE DISTRIBUCIÓN

La función de distribución asocia a cada valor de la variable la probabilidad acumulada hasta él:

$$F(x) = p(\,X \leq x\,)$$

La función de distribución verifica las siguientes propiedades:

1. Es creciente en todo su dominio.
2. Si la variable es discreta, la función de distribución es escalonada ya que toma valor constante entre dos valores consecutivos de la variable.
3. El recorrido de la función de distribución es el intervalo [0, 1].
4. Para valores inferiores al menor valor de la variable, la función de distribución toma valor cero.
5. Para valores superiores al mayor valor de la variable, la función de distribución toma valor uno.

Ejemplo

Para el ejemplo anterior vamos a obtener la función de distribución.

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

- Para las $x < 0$, se tiene $F(x) = 0$.
- Para las $x \in [0, 1)$, se tiene $F(x) = 0{,}1$.
- Para las $x \in [0, 2)$, se tiene $F(x) = 0{,}1 + 0{,}15 = 0{,}25$.
- Para las $x \in [0, 3)$, se tiene $F(x) = 0{,}1 + 0{,}15 + 0{,}45 = 0{,}70$.
- Para las $x \in [0, 4)$, se tiene $F(x) = 0{,}1 + 0{,}15 + 0{,}45 + 0{,}1 = 0{,}8$.
- Para las $x \geq 4$, se tiene $F(x) = 0{,}1 + 0{,}15 + 0{,}45 + 0{,}1 + 0{,}2 = 1$.

2. LEY BINOMIAL: FÓRMULA DE BERNOUILLI

Vamos a estudiar a continuación la distribución de Bernouilli y como generalización de esta, la distribución binomial.

2.1. DISTRIBUCIÓN DE BERNOUILLI

Denominamos **experimento de Bernouilli** a aquel experimento aleatorio que da lugar únicamente a dos posibles resultados a los que daremos el nombre de suceso éxito A y suceso fracaso A^C.

Como consecuencia tendremos una probabilidad de éxito y una de fracaso:

$P(A) = p \Rightarrow$ probabilidad de éxito.

$P(A^C) = 1 - p \Rightarrow$ probabilidad de fracaso.

Recuerda

La distribución de Bernouilli consiste en realizar un experimento aleatorio tan solo una vez y comprobar si ocurre o no determinado suceso, siendo p la probabilidad de que así sea (éxito) y $q = 1 - p$, que no lo sea (fracaso).

Definimos la **variable aleatoria de Bernouilli** como la variable X dada de la siguiente forma:

$$X = \begin{cases} 0 & \text{si ocurre el suceso } A^C \\ 1 & \text{si ocurre el suceso } A \end{cases}$$

Se denota por $X \to B(p)$ y se tiene:

$$P(X = 0) = 1 - p = q$$

$$P(X = 1) = p$$

2.2. DISTRIBUCIÓN BINOMIAL

Cuando un experimento de Bernouilli lo repetimos varias veces obtenemos la **variable binomial.** Es decir, tendremos:

x = "número de éxitos en las n repeticiones del experimento".

Seguiremos teniendo una probabilidad de éxito p y otra de fracaso q, pero ahora la variable ya no va a tomar solamente los valores 0 y 1, sino todos los valores entre 0 y n, donde n es el número de veces que repetiremos el experimento. Se denota por $X \to B(n,p)$.

Un fenómeno aleatorio sigue una distribución binomial cuando cumple las siguientes condiciones:

- **a)** Existe un número n fijo de ensayos repetidos, siendo cada uno de ellos un acierto o un fracaso.
- **b)** La probabilidad de éxito (acierto) es p. Es la misma en cada ensayo. La probabilidad de fracaso es q = 1 – p.
- **c)** Los ensayos son independientes, el resultado de uno no influye en el siguiente.

Atención

La distribución de Bernouilli es una distribución binomial con n = 1.

2.2.1. LA FUNCIÓN DE PROBABILIDAD

Veamos la probabilidad de k aciertos en n pruebas:

$$F(k) = P\ [X = k] = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

Donde F es la función de distribución binomial.

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

Donde n! es un número factorial que se calcula como: n! = n·(n-1)·(n-2)·...3·2·1.

Sea X una variable aleatoria Bin (n, p), se tiene lo siguiente:

Media= μ = np.

Varianza = σ^2= n·p·q.

Desviación típica = $\sigma = \sqrt{npq}$.

2.2.2. AJUSTE MEDIANTE UNA BINOMIAL

En algunas situaciones se tienen unos datos que se sospecha que responden a un comportamiento binomial y se desea confirmar esta intuición. Lo que se hace es igualar la media de la muestra con la de una distribución binomial, que como sabemos es n·p; así:

$$\overline{X} = n \cdot p \Rightarrow p = \frac{\overline{x}}{n}$$

2.2.3. EJERCICIOS RESUELTOS

Ejercicio 1:

En una oposición una de las pruebas es un test compuesto por 10 preguntas con 4 opciones de respuesta cada una, de las cuales solo una es correcta. Si un opositor contesta al azar, teniendo en cuenta que para aprobar es necesario contestar correctamente, al menos, 7 preguntas, ¿cuál será la probabilidad de que supere dicha prueba?

Sea X la variable aleatoria discreta que expresa el número de respuestas correctas, sigue una distribución binomial de parámetros B(10, 0,25).

Nos piden la probabilidad siguiente:

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) =$$

$$= \binom{10}{7} \cdot 0{,}25^7 \cdot 0{,}75^3 + \binom{10}{8} \cdot 0{,}25^8 \cdot 0{,}75^2 + \binom{10}{9} \cdot 0{,}25^9 \cdot 0{,}75 + \binom{10}{10} \cdot 0{,}25^{10} =$$

$$= 0{,}0036$$

Luego solo superará la prueba, por término medio, 36 veces de cada 10.000 que se presentará a la oposición.

Ejercicio 2:

Un representante realiza 15 visitas al día a los comercios de su ramo y, por su experiencia anterior, sabe que la probabilidad de que le hagan un pedido en cada visita es del 0,4. ¿Cuál es la probabilidad de que en un día le hagan 3 o más pedidos?

Obtenemos del enunciado una variable binomial de parámetros n = 15 y p = 0,4.

La probabilidad que nos piden es:

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = (\text{ consultar en las tablas}) =$$

$$= 1 - 0{,}0271 = 0{,}9729$$

Nota

El cálculo de probabilidades de la función de distribución binomial puede ser algo tedioso, por ello existen tablas para algunos valores de n y p que facilitan el trabajo. Entra en el Campus Virtual y encontrarás dichas tablas.

3. DISTRIBUCIONES CONTINUAS

Como ejemplos de distribuciones aleatorias continuas tenemos la estatura, la longitud de piezas fabricadas, las temperaturas, etc.

Puesto que una función de probabilidad de una variable aleatoria discreta asigna a cada valor la probabilidad de que este ocurra, esto no tiene sentido con una variable aleatoria continua, dado que habría que asignar a cada valor de la variable la probabilidad nula.

3.1. FUNCIÓN DE DENSIDAD

Se llama función de densidad de una variable aleatoria continua a aquella función f(x), cuya gráfica es el límite de los polígonos de frecuencias relativas de todas las posibles muestras.

Nos informa de cuáles son las zonas donde están los valores más probables, de las más densas en probabilidad.

- Propiedades de la función de densidad:
    - **a)** $f(x) \geq 0$.
    - **b)** El área que la función delimita con el eje OX vale 1.

      También se puede expresar:

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

(por ser la suma de toda la probabilidad).

3.2. FUNCIÓN DE DISTRIBUCIÓN

La función de distribución es el límite del polígono acumulativo de probabilidades, cuando el tamaño del intervalo de clase tiende a cero.

Dicho de otro modo, la función de distribución asocia a cada valor de la variable la probabilidad acumulada hasta él:

$$F(x) = P[\, X \leq x \,]$$

La función integral obtenida a partir de la de densidad se llama función de distribución de la variable continua X.

- Propiedades de la función de distribución:
  - **a)** Es una función monótona creciente.
  - **b)** Es continua.
  - **c)** $\lim_{x \to +\infty} F(x) = 1$
  - **d)** $\lim_{x \to -\infty} F(x) = 0$

3.3. MEDIA Y VARIANZA DE UNA VARIABLE ALEATORIA CONTINUA

$$\text{Media} = EX = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

$$\text{Varianza} = \sigma 2 = \int_{-\infty}^{+\infty} (x - EX)^2 \cdot f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2$$

TABLA COMPARATIVA:

Variable estadística	Variable aleatoria discreta	Variable aleatoria continua
$\overline{x} = \sum_{i=1}^{n} x_i \cdot f_i$	$\mu = \sum_{i=1}^{n} x_i \cdot p_i$	$\mu = \int_{-\infty}^{+\infty} x \cdot f(x)dx$
$S^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot f_i$	$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$	$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)dx$

Caso práctico

La función de densidad de una variable aleatoria continua viene definida por:

$$f(x) = \begin{cases} 2x & \text{para } 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Halla la función de distribución F(x).

b) Halla la media.

c) Halla la varianza.

d) Halla la desviación típica.

Solución:

a) Partiendo de la igualdad $F'(x) = f(x)$, obtenemos la función de distribución, que será:

$$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ x^2 & \text{para } 0 \leq x \leq 1 \\ 1 & \text{para } x > 1 \end{cases}$$

b) Media:

$$\mu = \int_0^1 x \cdot 2x \; dx = \left[\frac{2x^3}{3} \right]_0^1 = \frac{2}{3}$$

c) Varianza:

$$\sigma^2 = \int_0^1 \left(x - \frac{2}{3} \right)^2 \cdot 2x dx = \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x \right) dx =$$

$$= \left[\frac{2x^4}{4} - \frac{8x^3}{9} + \frac{4x^2}{9} \right]_0^1 = \frac{2}{4} - \frac{8}{9} + \frac{4}{9} = 0{,}055$$

d) Desviación típica:

$$\sigma = \sqrt{0{,}055} = 0{,}235$$

3.4. FRECUENCIA RELATIVA DE ÉXITO EN N PRUEBAS

Sea X = número de éxitos en n pruebas.

Sea $X_0 = \dfrac{X}{n}$ = frecuencia relativa de éxito en n pruebas.

Entonces X_0 tiene:

$$\mu = p \qquad \text{y} \qquad \sigma^2 = \frac{pq}{n}$$

3.5. DISTRIBUCIÓN NORMAL

Es la más importante de las distribuciones continuas y se presenta en multitud de problemas.

Todas las distribuciones estadísticas se aproximan a la normal a medida que aumenta el número de observaciones (teorema central del límite, TCL). Este teorema ocupa un puesto importante en la probabilidad y en la estadística y sus posibilidades de aplicación son enormes.

La función de densidad de la distribución normal fue descubierta por Gauss, y viene dada por la siguiente expresión:

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

Denotaremos a la normal con N (μ, σ), y si μ = 0 y σ = 1, la distribución está normalizada y obtenemos la normal estándar Z = N (0, 1).

Nota

La distribución normal también se denomina distribución de Gauss o distribución gaussiana, ya que Carl Gauss fue quien desarrolló y formuló la ecuación de la curva, que por su forma suele denominarse **campana de Gauss.**

3.5.1. DEFINICIÓN

Una variable aleatoria continua X tiene una distribución normal de media μ y desviación típica σ, si cumple las siguientes condiciones:

1. La variable recorre toda la recta real.

2. La función de densidad, que es la expresión en términos de ecuación matemática de la curva de Gauss, es decir:

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

Nota

La función de densidad de una distribución normal es simétrica y con forma de campana, lo que hace que se utilice para numerosos modelos de variables.

- Propiedades de la gráfica de la función de densidad:
    1. Su dominio son los números reales.
    2. Es simétrica respecto a la ordenada relativa a la media. Este es el valor máximo de la función de densidad.
    3. Coinciden la moda, la mediana y la media.
    4. Posee una asíntota: $y = 0$.
    5. La función crece hasta $x = \mu$, y decrece a partir de $x = \mu$.
    6. Es cóncava en $(\mu-\sigma, \mu+\sigma)$ y convexa en el resto. Luego tiene dos puntos de inflexión: $x = \mu - \sigma$ y $x = \mu + \sigma$.
    7. El área encerrada por la función es 1.
    8. En cualquier distribución normal se cumple lo siguiente:
        - En $(\mu-\sigma, \mu+\sigma)$ se encuentra el 68% del área total.
        - En $(\mu-2\sigma, \mu+2\sigma)$ se encuentra el 95% del área total.

3.6. TIPIFICACIÓN DE LA VARIABLE NORMAL

La variable normal estándar N(0, 1) se encuentra tabulada, lo cual permite un cálculo rápido de las probabilidades asociadas a esta distribución.

Por otra parte, es obvio que no se pueden construir tablas para todos los posibles valores de μ y σ; luego recurriremos al proceso de transformar una variable aleatoria normal N (μ, σ), en otra variable Z que siga una distribución N (0, 1). Esta transformación se conoce con el nombre de **tipificación** de la variable.

Para llevar a cabo la tipificación realizaremos dos pasos:

1. Centrar, es decir, trasladar la media de la distribución al origen de coordenadas. Esto equivale a hacer $\mu = 0$.
2. Reducir la desviación estándar a 1. Esto equivale a dilatar o contraer la gráfica de la distribución.

Podemos realizar simultáneamente los dos pasos con el siguiente cambio de variable:

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

3.7. USO DE LAS TABLAS DE LA DISTRIBUCIÓN NORMAL

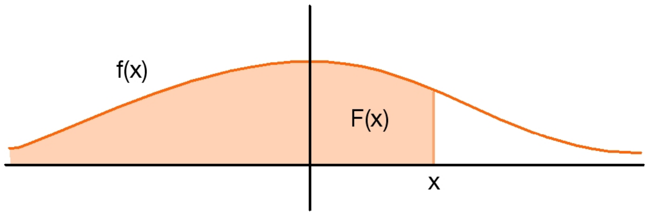
Para el cálculo de probabilidades correspondientes a fenómenos aleatorios que siguen una distribución normal, se utilizan las tablas de la distribución normal, pues los cálculos directos pueden ser complicados y laboriosos.

Para trabajar con las tablas debemos tener en cuenta lo siguiente:

1. Si la variable no está tipificada, debemos tipificarla para convertirla en normal estándar.
2. Si aparece un valor negativo, debemos pasarlo a positivo por simetría.
3. Si la tabla nos da las probabilidades de las colas izquierdas, deberemos dejar la probabilidad en función de F(x).

3.7.1. ESTUDIO DE CASOS PRINCIPALES

P[Z ≤ x] = F(x).



P[Z ≤ -x] = P[Z ≥ x] = 1 – P[Z < x].

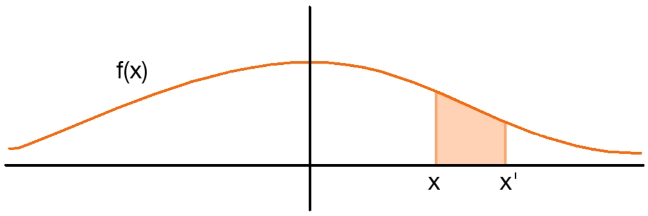
Se debe tener en cuenta que en una variable continua P[Z < x] = P[Z ≤ x].

Recuerda

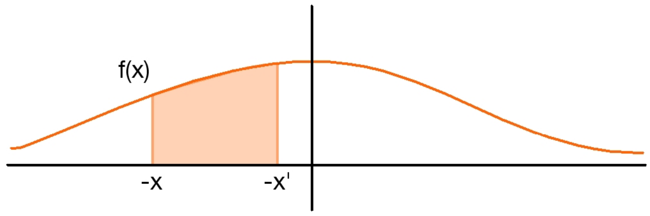
El símbolo **≤** significa “menor o igual que”.

El símbolo **≥** significa “mayor o igual que”.

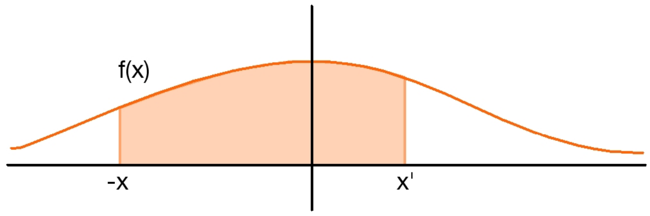
P[x ≤ Z≤ x’] = P[Z ≤ x’] - P[Z ≤ x] = F(x’)-F(x).



P[-x ≤ Z ≤ -x’] = P[Z ≤ -x’]-P[Z ≤ -x] = P[Z ≥ x’]- P[Z ≥ x] = 1 - P[Z < x’] – (1 - P[Z < x]) = P[Z < x] - P[Z < x’].



P[-x ≤ Z ≤ x'] = P[Z ≤ x'] - P[Z ≤ -x] = P[Z ≤ x] - P[Z ≥ x] = P[Z ≤ x] –(1 - P[Z < x]) = P[Z ≤ x] + P[Z < x] – 1.



3.7.2. APLICACIONES PRÁCTICAS

A veces tiene interés saber qué proporción de individuos se distribuye en intervalos de la forma (μ - k·σ, μ + k·σ), siendo k un número entero.

Supongamos que tenemos una variable aleatoria normal X que sigue una distribución N (μ, σ). Veamos qué proporción de individuos se encontrarán en el intervalo (μ-σ, μ+σ):

$$P(\mu-\sigma < X \leq \mu+\sigma) = P\left(\frac{\mu-\sigma-\mu}{\sigma} < Z \leq \frac{\mu+\sigma-\mu}{\sigma}\right) = P(-1 < Z \leq 1) =$$

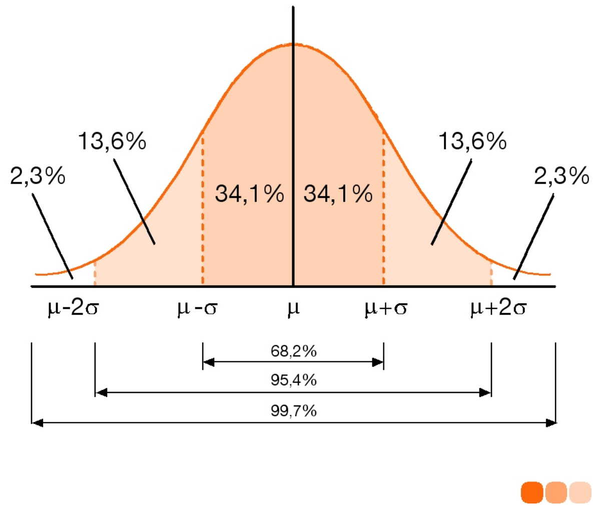
$$= 2\cdot[P(Z \leq 1) - 0{,}5] = 2\cdot(0{,}8413 - 0{,}5) = 0{,}6826$$

Es decir, el 68,26% de las observaciones (o de los individuos) se encuentra en el intervalo (μ-σ, μ+σ).

Análogamente, se puede calcular que:

El 95,4% de las observaciones está en (μ - 2σ, μ + 2σ).

El 99,7% de las observaciones está en (μ - 3σ, μ + 3σ).



Proporción de individuos que se distribuyen en intervalos de la forma (μ - k·σ, μ +k·σ)

En ocasiones nos interesa saber entre qué valores se encuentra el 50% de los individuos. Para ello hay que manejar las tablas en forma inversa, ya que en los casos anteriores se conocía el valor de la abscisa y se trataba de hallar el valor de la probabilidad, pero ahora se conoce el valor de la probabilidad y se trata de hallar el valor de la abscisa.

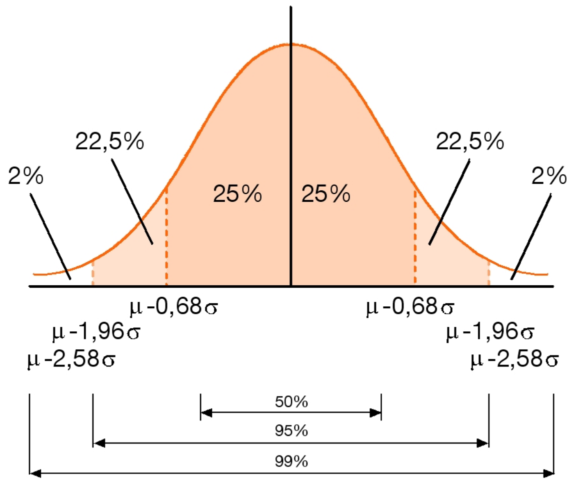
El caso propuesto anteriormente equivale a calcular un valor de la abscisa que deje a su izquierda el 75% de los individuos de la población, es decir, el precentil 75.

Si X es una variable aleatoria N (μ, σ), se verifica que:

$$P\left(\frac{X-\mu}{\sigma} \leq Z\right) = 0{,}750.$$

Buscando en las tablas una probabilidad lo más próxima a 0,75, se obtiene para z el valor 0,68. Luego:

$$\frac{X-\mu}{\sigma} \leq \mathbf{0{,}68} \Rightarrow X - \mu \leq 0{,}68\cdot\sigma \Rightarrow X \leq 0{,}68\cdot\sigma + \mu.$$



Y por simetría se tiene que el 50% de los individuos está en el intervalo (μ - 0,68σ, μ + 0,68σ).

Análogamente, podemos calcular:

El 95% de los individuos está en el intervalo (μ - 1,96σ, μ + 1,96σ).

El 99% de los individuos está en el intervalo (μ - 2,58σ, μ + 2,58σ).

Ejemplo:

En un almacén de patatas el peso de los sacos vendidos en el último mes se distribuye normalmente con media 70 kg y desviación típica 6 kg. Si se han vendido 60 sacos, calcula cuántos de ellos tendrán un peso entre 64 y 76 kg.

Se trata de una distribución normal N(70, 6). Según lo visto anteriormente, el 68,2% de los sacos está entre (70 – 6, 70 + 6); por tanto, habrá que calcular el 68,2% de 2.000, es decir:

$$68{,}2 \cdot \frac{2.000}{100} = 1{,}364.$$

Luego se espera que haya 1.364 sacos con peso comprendido entre 64 y 76 kg.

Caso práctico

Aplicamos un test de hábitos de estudio a 300 alumnos y se observa que se distribuyen normalmente con media 30 y desviación típica 12. Calcula:

a) La proporción de alumnos con una puntuación entre 20 y 35.

b) El número de alumnos con una puntuación superior a 42.

Solución:

a) Se trata de una distribución N(30, 12). Calculemos las probabilidades:

$$p(20 < X \leq 35) = p\left(\frac{20-30}{12} < Z \leq \frac{35-30}{12}\right) =$$

$$p(-0{,}83 < Z \leq 0{,}42) = p(Z \leq 0{,}42) - [1 - p(Z \leq 0{,}83)] = 0{,}6628 -$$

$$(1 - 0{,}7967) = 0{,}4595.$$

b) Es decir, aproximadamente el 46% de los alumnos tiene una puntuación en el test entre 20 y 35.

$$P(X > 42) = 1 - P(X \leq 42) = 1 - p\left(Z \leq \frac{42-30}{12}\right) = 1 - p(Z \leq 1)$$

$$= 1 - 0{,}8413 = 0{,}1587.$$

3.8. APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL

La función de densidad de la distribución normal se obtiene como límite de la normal de probabilidad binomial si se cumple:

$$P = \frac{1}{2} \qquad \text{y} \qquad n \to \infty$$

Este hecho puede usarse para calcular probabilidades de experimentos o sucesos binomiales, que, para valores de n grandes resultan complicadas de calcular a través de las tablas de la distribución normal.

Esta aproximación está indicada cuando $n > 10$. La bondad del ajuste es tanto mayor cuanto más grande sea n y cuanto más próximo esté p a $\frac{1}{2}$.

Si p es muy grande o muy pequeño (próximo a cero o uno), se emplea con mejores resultados la distribución de Poisson.

Recordemos que la media, la varianza y la desviación típica de una distribución binomial B(n, p) son:

Media: $\mu = np$.

Varianza: $\sigma^2 = n \cdot p \cdot q$.

Desviación típica: $\sigma = \sqrt{npq}$.

+ Info

De Moivre demostró que bajo determinadas condiciones la distribución binomial B(n, p) se puede aproximar mediante la distribución normal N (np, $\sqrt{npq}$), y esta, a su vez, se puede tipificar y convertir en una normal estándar.

Las condiciones para aplicar la aproximación de De Moivre son que $\mathbf{n \cdot p \geq 5}$ **y** $\mathbf{n \cdot q \geq 5}$.

Así, cuanto mayor es el valor de n y más próximo a 0,5 sea el valor de p, mejor será la aproximación realizada.

3.8.1. Observación

La distribución binomial es de variable aleatoria discreta y, por tanto, tiene sentido calcular probabilidades puntuales. Sin embargo, la distribución normal es de variable aleatoria continua y, por tanto, no tiene sentido calcular probabilidades puntuales, pues son todas nulas.

Basta considerar los valores de la variable aleatoria discreta como marcas de clase de intervalos. Debemos hacer una corrección *(corrección de Yates):* cada valor entero x, se considera como un intervalo (x – 0,5, x + 0,5).

Ejemplo

Realizado un sondeo en una cierta población, se observa que únicamente el 15% de la misma está en desacuerdo con la prohibición de fumar en lugares públicos. Elegida una muestra al azar de 50 individuos, calcular:

a) La probabilidad de que haya más de cinco personas favorables a fumar en lugares públicos.
b) La probabilidad de que, a lo sumo, haya seis personas favorables.

Se trata de una distribución binomial. Consideremos A como el suceso "ser favorable":

$p = p(A) = 0{,}15, \qquad q = 1 - p = 1 - 0{,}15 = 0{,}85.$

Sea X la variable aleatoria que expresa el número de personas de la muestra favorable a fumar en público, se verifica entonces que X sigue una distribución Bin (50, 0,15).

Veamos si se cumplen las condiciones para aproximar la variable binomial a una normal:

$n \cdot p = 50 \cdot 0{,}15 \geq 5 \qquad y \qquad n \cdot q = 50 \cdot 0{,}85 \geq 5.$

Ejemplo:

Luego aplicando el método de De Moivre aproximamos la variable X a una nueva variable X´, que seguirá la siguiente distribución:

$$N\left(50 \cdot 0{,}15,\ \sqrt{50 \cdot 0{,}15 \cdot 0{,}85}\right) = N(7{,}5,\ 2{,}52).$$

$$p(X > 5) = p(X' > 4.5) = p\left(Z > \frac{4.5 - 7.5}{2.52}\right)$$

$$= p(Z > -1{,}19) = 1 - p(Z \leq 1{,}19) = 1 - 0{,}8830 = 0{,}1170.$$

$$p(X \leq 6) = p(X' \leq 6{,}5) = p\left(Z \leq \frac{6{,}5 - 7{,}5}{2{,}52}\right) = p(Z \leq -0{,}4) =$$

$$= 1 - p(Z \leq 0{,}4) = 1 - 0{,}6554 = 0{,}3446.$$

CONCLUSIONES

Las distribuciones de probabilidad se agrupan en distribuciones de probabilidad discretas y continuas, dependiendo de si las variables son discretas o continuas.

La más importante para las variables discretas es la distribución binomial y para las variables continuas, la distribución normal. Esta última es utilizada en multitud de estudios.

RECAPITULACIÓN

	BINOMIAL Bin (n, p)	NORMAL N (μ, σ)
Variable	Aleatoria discreta $X \in \{0, 1..., n\}$	Aleatoria continua $Z \in (-\infty, \infty)$
Parámetros	n = n.º de pruebas p = probabilidad de éxito	μ = media aritmética σ = desviación típica
Función de probabilidad	$f(x) = \binom{n}{x} p^x \cdot q^{n-x}$	$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
Probabilidad puntual	$P(X = r) = \binom{n}{r} \cdot p^r \cdot q^{n-r}$	$P(X = r) = 0$
Función de distribución	$F(r) = p(X \le r) =$ $= p(X = 0) + p(X = 1) + ... + p(X = r)$	$F(r) = p(X \le r) =$ $= \int_{-\infty}^{r} f(x)dx$
Parámetros	$\mu = n \cdot p$ y $\sigma^2 = n \cdot p \cdot q$	μ y σ

AUTOCOMPROBACIÓN

1. **Si nos fijamos en la gráfica de una variable aleatoria normal:**

   a) Coinciden la media y la moda.

   b) Coinciden la mediana y la moda.

   c) Coinciden la mediana y la media.

   d) Ninguna es falsa.

2. **Si el cociente intelectual de los alumnos de la clase se distribuye normalmente como N (100, 20), ¿qué porcentaje de alumnos tiene un cociente intelectual superior a 130?**

   a) 93,3%.

   b) 75,8%.

   c) 6,7%.

   d) 15,3%.

3. **Consideremos la variable normalizada. Es falso que:**

   a) (-2, 2) contiene más del 75% de los datos normalizados.

   b) (-3, 3) contiene más del 88% de los datos normalizados.

   c) El rango de los datos normalizados depende del tipo de datos.

   d) Si la población es normal las probabilidades están tabuladas.

4. **La variable tipificada posee:**

   a) Esperanza y varianza cero.

   b) Varianza cero y esperanza uno.

   c) Esperanza y varianza uno.

   d) Varianza uno y esperanza cero.

5. **Sea X la variable aleatoria y F su función de distribución, entonces se cumple que:**

   a) F es continua.

   b) F toma solo valores en [0,1].

   c) F es estrictamente creciente.

   d) Todas son falsas.

6. **Respecto a la distribución binomial, es verdadero que:**

   a) No la usaremos para n grandes por ser muy lenta.

   b) Cumple que $B(n, p) + B(m, p) = B(n + m, 2p)$.

   c) Es simétrica para valores de p cercanos a cero.

   d) Es simétrica para valores de p cercanos a uno.

7. **Decimos que X es una variable aleatoria continua si se tiene que:**

   a) Su función de distribución es continua por la derecha.

   b) Su función de distribución es continua.

   c) Existe una función real positiva f tal que $F(x) = \int_{-\infty}^{x} f(t)dt$.

   d) Existe una función real f tal que $F(x) = \int_{-\infty}^{x} f(t)dt$.

8. **Indica la afirmación errónea:**

   a) Toda variable aleatoria es discreta o continua.

   b) La función de densidad encierra bajo ella un área de uno.

   c) Si f es continua y F derivable se cumple que $F'(x) = f(x)$.

   d) Una de las anteriores no es cierta.

9. **Consideremos la función de densidad normal. Es cierto que:**

   a) Su gráfica posee dos puntos de inflexión en μ-2σ y μ+2σ.

   b) Es simétrica pero no respecto de la media.

   c) No tiene extremos relativos.

   d) Todas son falsas.

10. **La expresión de la función de densidad de la distribución normal estándar es la siguiente:**

   a) f(x) = $\frac{1}{\sqrt{2\pi}} e^{x^2/2}$ si x $\in$ R.

   b) f(x) = $e^{-x/2}$ si x $\in$ R.

   c) f(x) = $\frac{1}{\sqrt{2\pi}} e^{-x^2}$ si x $\in$ R.

   d) f(x) = $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ si x $\in$ R.

SOLUCIONARIO

1.	d	**2.**	c	**3.**	c	**4.**	d	**5.**	b
6.	a	**7.**	b	**8.**	a	**9.**	d	**10.**	d

PROPUESTAS DE AMPLIACIÓN

Te proponemos que compruebes si tienes materiales complementarios, o clases grabadas dentro de la unidad. Si es así, descárgalos para ampliar la información sobre el tema y **recuerda marcar he terminado.**

Te proponemos también que entres **en la sección de agenda** y compruebes qué clases en directo y/o talleres tienes disponibles, para complementar tus estudios, o tu preparación a la hora de afrontar los exámenes.

BIBLIOGRAFÍA

- VV.AA. EULER. *Matemáticas I.* Madrid: S.M, 2001.
- VV.AA. EULER. *Matemáticas II*. aplicadas a las ciencias sociales. Madrid: SM, 2000.
- COLERA, J y otros. *Matemáticas. En tus manos.* Madrid: Anaya, 2002.
- BESCÓS, E y PEÑA, Z. *Proyecto Exedra. Matemáticas I*. Madrid: Oxford, 2001.
- VV.AA. *Matemáticas I.* Madrid: Edelvives, 2003.
- VV.AA. *Matemáticas II.* Madrid: Edelvives, 2003.