Tema 5. Variables aleatorias discretas

Resumen del tema

5.1. Definición de variable aleatoria discreta

5.1.1. Variables aleatorias

Una **variable aleatoria** es una función que asigna un número a cada suceso elemental de un experimento aleatorio.

Cualquier variable estadística cuantitativa estudiada en los temas 1 a 3 podría considerarse variable aleatoria con la condición de que esté observada en todos los individuos de una población.

La **media** de una variable aleatoria X se denota por μ_x. En el caso en el que no exista la posibilidad de confusión respecto de la variable aleatoria con la que estamos trabajando, la media se denotará solamente por μ. A la media de una variable aleatoria X también se le llama **esperanza matemática** de X, denotándola entonces por $E(X)$.

La **varianza** de una variable aleatoria X se denota por $\mathrm{Var}(X)$, por σ_x^2 o simplemente por σ^2.

Por tanto, la **desviación típica** de una variable aleatoria X se denota por σ_x o por σ.

La **función de distribución** de una variable aleatoria X se denota por F_X o simplemente por F y se define de la siguiente forma:

$$F_X(t) = P(X \leq t) \quad \text{para todo } t\,.$$

CLASIFICACIÓN DE LAS VARIABLES ALEATORIAS:

- ⋆ **Variable aleatoria discreta**: sólo puede tomar valores numéricos aislados (fijados dos consecutivos, no puede existir ninguno intermedio).
- ⋆ **Variable aleatoria continua**: puede tomar cualquier valor numérico dentro de un intervalo, de modo que entre cualesquiera dos de ellos siempre existe otro posible valor.

5.1.2. Variables aleatorias discretas

- **Identificación de una variable aleatoria discreta** X: es preciso conocer el conjunto de los posibles resultados de X:

  $$\{x_1, x_2, \ldots, x_k, \ldots\}$$

  y el conjunto de las probabilidades siguientes:

  $$\begin{aligned} p_1 &= P(X = x_1) \\ p_2 &= P(X = x_2) \\ &\vdots \\ p_k &= P(X = x_k) \\ &\vdots \end{aligned}$$

- **Función de probabilidad** de una variable aleatoria discreta X:

$$p_i = P(X = x_i) \quad \text{para } i = 1, 2, \ldots, k, \ldots$$

- **Propiedades importantes de la función de probabilidad**:
  - ⋆ $p_i \geq 0$ para todo i.
  - ⋆ $\sum p_i = 1$.

- La **representación gráfica de la función de probabilidad** de una variable aleatoria discreta es análoga al diagrama de barras de frecuencias relativas de una variable estadística discreta.

- **Función de distribución** de una variable aleatoria discreta X:

$$F_X(t) = P(X \leq t) \quad \text{para todo } t\,.$$

- La **representación gráfica de la función de distribución** de una variable aleatoria discreta es análoga al gráfico de frecuencias relativas acumuladas de una variable estadística discreta.

- **Media** de una variable aleatoria discreta X:

$$\mu = E(X) = \sum x_i\, p_i\,.$$

- **Varianza** de una variable aleatoria discreta X:

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2\, p_i = \left(\sum x_i^2\, p_i\right) - \mu^2\,.$$

- **Desviación típica** de una variable aleatoria discreta X:

$$\sigma = \sqrt{\mathrm{Var}(X)}\,.$$

5.2. Distribuciones discretas importantes

De las distribuciones discretas importantes solamente vamos a estudiar la distribución Binomial.

Una variable aleatoria X tiene una distribución **Binomial** de parámetros n y p si cuenta el número de veces que ocurre un suceso denominado *éxito* cuando se repite n veces un experimento cuyos únicos resultados son el suceso *éxito* o su contrario, permaneciendo constante la probabilidad (p) del suceso *éxito*.

El suceso contrario de *éxito* se suele llamar *fracaso* y su probabilidad es $q = 1 - p$.

Son equivalentes las dos afirmaciones siguientes: "X tiene una distribución Binomial de parámetros n y p" y "X es una variable aleatoria Binomial de parámetros n y p".

La variable aleatoria Binomial de parámetros n y p será denotada por: $\mathcal{B}(n, p)$.

La afirmación "X tiene una distribución Binomial de parámetros n y p" será denotada por: $X \equiv \mathcal{B}(n, p)$.

Los **posibles resultados** de $\mathcal{B}(n, p)$ son:

$$\{0, 1, 2, \ldots, n\}\,.$$

Si $X \equiv \mathcal{B}(n, p)$, su **función de probabilidad** es:

$$P(X = k) = \frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k}, \qquad \text{para } k = 0, 1, 2, \ldots, n\,.$$

Los resultados de la función de probabilidad de $\mathcal{B}(n,p)$ se pueden determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de n y p) y con las tablas de la función de probabilidad (para algunos valores de n y p).

Si $X \equiv \mathcal{B}(n,p)$, su **función de distribución** se define como se definía para cualquier variable aletoria; es decir:

$$F_X(t) = P(X \leq t) \quad \text{para todo } t\,.$$

Los resultados de la función de distribución de $\mathcal{B}(n,p)$ se pueden determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de n y p) y con las tablas de la función de distribución (para algunos valores de n y p).

A la función de distribución se le llama también *función acumulativa* o *función de probabilidades acumuladas* debido a que, para un número natural t, se verifica:

$$\begin{aligned} P(X \leq t) &= P\left[(X=0) \cup (X=1) \cup \cdots \cup (X=t)\right] \\ &= P(X=0) + P(X=1) + \cdots + P(X=t)\,, \end{aligned}$$

es decir, si t es un número natural entonces para obtener $P(X \leq t)$ se van sumando (acumulando) las probabilidades de los sucesos $(X=0)$, $(X=1)$, ... y $(X=t)$.

A partir de los resultados de la función de distribución podemos determinar los valores de la función de probabilidad ya que:

$$\begin{aligned} P(X=0) &= P(X \leq 0) = F_X(0)\,, \\ P(X=k) &= P[(X \leq k) - (X \leq k-1)] = P(X \leq k) - P(X \leq k-1) \\ &= F_X(k) - F_X(k-1), \qquad \text{para } k = 1, 2, \ldots, n\,. \end{aligned}$$

Si $X \equiv \mathcal{B}(n,p)$, su **media** es:

$$\mu = np\,,$$

su **varianza** es:

$$\sigma^2 = npq\,,$$

y su **desviación típica** es:

$$\sigma = \sqrt{npq}\,.$$

Ejemplos que se van a resolver en clase

Ejemplo 5.1. Se sabe que el 4 % de los libros que se prestan en una biblioteca escolar se devuelven con retraso. Se realiza el experimento que consiste en observar si la devolución de un libro se ha hecho con retraso o no.

a) Determinar la función de probabilidad y hacer su representación gráfica.

b) Calcular la función de distribución y hacer su representación gráfica.

c) Hallar la media y la varianza.

Ejemplo 5.2. Sea X una variable aleatoria discreta cuya función de probabilidad viene dada por:

x_i	1	3	5	7	9
$p_i = P(X = x_i)$	$0'2$	$0'1$	$0'3$	$0'3$	$0'1$

Calcular:

a) La expresión de la función de distribución $F(t)$ para todo valor de t.

b) La esperanza matemática de X.

c) La varianza y la desviación típica de X.

Ejemplo 5.3. Sea $X \equiv \mathcal{B}(10, 0'4)$. Calcular:

a) $P(X \leq 5)$

b) $P(X > 5)$

c) $P(X < 5)$

d) $P(X \geq 5)$

e) $P(X = 5)$

f) $P(2 \leq X \leq 8)$

g) $P(2 \leq X < 8)$

h) $P(2 < X \leq 8)$

i) $P(2 < X < 8)$

Ejemplo 5.4. Sea $X \equiv \mathcal{B}(10, 0'6)$. Calcular $P(X \leq 5)$.

Ejemplo 5.5. Supongamos que el 30 % de la población de todos los usuarios de un centro de documentación tiene un título de licenciado. Supongamos, también, que la población es suficientemente grande como para que al elegir un usuario al azar y apartarlo, no se altere dicho porcentaje. Realizamos el experimento que consiste en elegir al azar tres usuarios de dicho centro de documentación y observar la variable aleatoria X=*número de usuarios del centro de documentación que tiene un título de licenciado, entre los tres elegidos al azar.*

a) Hallar la función de probabilidad de X y hacer su representación gráfica.

b) Determinar la función de distribución de X y hacer su representación gráfica.

c) Calcular la media y la desviación típica de X.

Ejemplo 5.6. Supongamos, igual que en el ejemplo anterior, que el 30 % de la población de todos los usuarios de un centro de documentación tiene un título de licenciado. Supongamos, también, que la población es suficientemente grande como para que al elegir un usuario al azar y apartarlo, no se altere dicho porcentaje. Se eligen al azar 20 usuarios de dicho centro de documentación. Calcular:

a) La probabilidad de que ninguno de ellos tenga un título de licenciado.

b) La probabilidad de que menos de 5 tengan un título de licenciado.

c) La probabilidad de que 10 de ellos o más tengan un título de licenciado.

Problemas propuestos

Problema 5.1. De los usuarios de un centro de documentación, el 23 % pertenece al grupo I de edad (menos de 20 años). Supongamos, también, que la población es suficientemente grande como para que al elegir un usuario al azar y apartarlo, no se altere dicho porcentaje. Realizamos el experimento que consiste en elegir al azar tres usuarios del centro de documentación y observar la variable aleatoria X=*número de usuarios que pertenecen al grupo I de edad, entre los tres elegidos al azar.*

a) Hallar el conjunto de los posibles resultados de la variable aleatoria X, así como su función de probabilidad.

b) Hallar la probabilidad de que el número de usuarios que pertenecen al grupo I sea menor que dos.

c) Determinar la función de distribución de X y hacer su representación gráfica.

d) Calcular la media y la desviación típica de X.

Problema 5.2. De un total de 500 libros, 50 son científicos. Extraemos al azar un primer libro entre los 500 y lo reponemos en la población de libros antes de realizar una nueva extracción; volvemos a extraer al azar un segundo libro entre los 500 y lo reponemos antes de hacer una nueva extracción; ...; finalmente, extraemos un quinto libro entre los 500. Consideramos la variable aleatoria X=*número de libros científicos, entre los 5 elegidos al azar con reposición.*

a) Hallar la función de probabilidad de X y hacer su representación gráfica.

b) Determinar la función de distribución de X y hacer su representación gráfica.

c) A partir de la función de distribución de X, calcular la probabilidad de que el número de libros científicos sea mayor que 3.

d) Calcular la media y la desviación típica de X.

Problema 5.3. Los libros que salen de una imprenta se clasifican en defectuosos (si tienen defectos de impresión) y no defectuosos (si no tienen defectos de impresión). Se supone que la cantidad de libros que salen de dicha imprenta es tan grande, que puede considerarse infinita. Por tanto, si elegimos y apartamos un libro, esto no altera el porcentaje de libros no defectuosos, que es 95 %.

a) Si se eligen al azar 20 libros, ¿cuál es la probabilidad de que 18 de ellos sean no defectuosos?

b) Si se eligen al azar 25 libros, ¿cuál es la probabilidad de que el número de libros no defectuosos sea mayor o igual que 21?

Problema 5.4. Se sabe que el 4 % de los libros que se prestan en una biblioteca escolar se devuelven con retraso. Se realiza el experimento que consiste en observar si la devolución de cada libro se ha hecho con retraso o no. Se eligen al azar 12 libros prestados.

a) ¿Cuál es la probabilidad de que se devuelvan con retraso 2 libros?

b) ¿Cuál es la probabilidad de que se devuelvan con retraso más de 2 libros?

Problema 5.5. Supongamos que el 1 % de la población de todos los usuarios de un centro de documentación tiene menos de 10 años. Supongamos, también, que la población es suficientemente grande como para que al elegir un usuario al azar y apartarlo, no se altere dicho porcentaje. Se eligen al azar 15 usuarios de dicho centro de documentación. Calcular:

a) La probabilidad de que ninguno de ellos tenga menos de 10 años.

b) La probabilidad de que tengan menos de 10 años 3 usuarios o menos.

c) La probabilidad de que tengan menos de 10 años menos de 3 usuarios.

d) La probabilidad de que tengan menos de 10 años más de 2 usuarios.

e) La probabilidad de que tengan menos de 10 años 2 usuarios o más.

f) La probabilidad de que el número de usuarios con menos de 10 años esté comprendida entre 2 (incluido) y 10 (incluido).

g) El número medio de usuarios con menos de 10 años.

Soluciones de los problemas propuestos

Solución del problema 5.1. $X \equiv \mathcal{B}(n=3, p=0'23)$

a) La función de probabilidad de X es:

$$P(X=k) = \frac{n!}{k!(n-k)!}\, p^k\, q^{n-k} \qquad \text{para } k=0,1,2,3$$

siendo $n=3$, $p=0'23$ y $q=1-p=0'77$.

Explícitamente, la función de probabilidad viene dada en la siguiente tabla:

x_i	$p_i = P(X=x_i)$
0	$0'456533$
1	$0'409101$
2	$0'122199$
3	$0'012167$

b) $P(X<2) = 0'865634$

c) La función de distribución, para todo valor de t, es la siguiente:

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t<0 \\ 0'456533 & \text{si} \quad 0 \le t < 1 \\ 0'865634 & \text{si} \quad 1 \le t < 2 \\ 0'987833 & \text{si} \quad 2 \le t < 3 \\ 1 & \text{si} \quad t \ge 3 \end{cases}$$

Su representación gráfica es similar al gráfico de frecuencias acumuladas relativas de una variable estadística cuantitativa discreta (con datos no agrupados en intervalos).

d) Media= $\mu = n\,p = 3 \cdot 0'23 = 0'69$ usuarios, Varianza= $\sigma^2 = n\,p\,q = 3 \cdot 0'23 \cdot 0'77 = 0'5313$ usuarios^2, Desviación típica= $\sigma = 0'7289$ usuarios.

Solución del problema 5.2. $X \equiv \mathcal{B}(n=5, p=0'1)$

a) La función de probabilidad de X es:

$$P(X=k) = \frac{n!}{k!(n-k)!}\, p^k\, q^{n-k} \qquad \text{para } k=0,1,2,3,4,5$$

siendo $n=5$, $p=0'1$ y $q=1-p=0'9$.
Explícitamente, la función de probabilidad viene dada en la siguiente tabla:

x_i	$p_i = P(X=x_i)$
0	$0'59049$
1	$0'32805$
2	$0'0729$
3	$0'0081$
4	$0'00045$
5	$0'00001$

Su representación gráfica es similar al diagrama de barras de frecuencias relativas de una variable estadística cuantitativa discreta (con datos no agrupados en intervalos).

b) La función de distribución, para todo valor de t, es la siguiente:

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ 0'59049 & \text{si} \quad 0 \le t < 1 \\ 0'91854 & \text{si} \quad 1 \le t < 2 \\ 0'99144 & \text{si} \quad 2 \le t < 3 \\ 0'99954 & \text{si} \quad 3 \le t < 4 \\ 0'99999 & \text{si} \quad 4 \le t < 5 \\ 1 & \text{si} \quad t \ge 5 \end{cases}$$

Su representación gráfica es similar al gráfico de frecuencias acumuladas relativas de una variable estadística cuantitativa discreta (con datos no agrupados en intervalos).

c) $P(X > 3) = 1 - F_X(3) = 0'00046$.

d) Media$= \mu = n\,p = 5 \cdot 0'1 = 0'5$ libros científicos, Varianza$= \sigma^2 = n\,p\,q = 5 \cdot 0'1 \cdot 0'9 = 0'45$ (libros científicos)^2, Desviación típica$= \sigma = 0'6708$ libros científicos.

Solución del problema 5.3.

a) Sea X=*número de libros no defectuosos, entre los 20 elegidos al azar*. Entonces $X \equiv \mathcal{B}(n = 20, p = 0'95)$. Por tanto, $P(X = 18) = F_X(18) - F_X(17) = 0'188677$.

b) Sea X=*número de libros no defectuosos, entre los 25 elegidos al azar*. Entonces $X \equiv \mathcal{B}(n = 25, p = 0'95)$. Por tanto, $P(X \ge 21) = 1 - F_X(20) = 0'992835$.

Solución del problema 5.4. Sea X=*número de libros que se devuelven con retraso, entre los 12 libros prestados elegidos al azar*. Entonces $X \equiv \mathcal{B}(n = 12, p = 0'04)$.

a) $P(X = 2) = F_X(2) - F_X(1) = 0'070206$.

b) $P(X > 2) = 1 - F_X(2) = 0'010729$.

Solución del problema 5.5. Sea X=*número de usuarios con menos de 10 años, entre los 15 elegidos al azar*. Entonces $X \equiv \mathcal{B}(n = 15, p = 0'01)$.

a) $P(X = 0) = F_X(0) = 0'860058$.

b) $P(X \le 3) = F_X(3) = 0'999988$.

c) $P(X < 3) = F_X(2) = 0'999584$.

d) $P(X > 2) = 1 - F_X(2) = 0'000416$.

e) $P(X \ge 2) = 1 - F_X(1) = 0'000963$.

f) $P(2 \le X \le 10) = F_X(10) - F_X(1) = 0'000963$.

g) $E(X) = np = 0'15$ usuarios con menos de 10 años.