Manual Básico de Estadística

Alfredo Sánchez Alberca (asalber@ceu.es)

Sep 2017

Departamento de Matemática Aplicada y Estadística
CEU San Pablo

Términos de la licencia

Contenidos

Variables Aleatorias

Variables aleatorias

Cuando seleccionamos una muestra al azar de una población estamos realizando un experimento aleatorio y cualquier variable estadística medida a partir de la muestra será una variable aleatoria porque sus valores dependerán del azar.

Definición (Variable Aleatoria)

Una *variable aleatoria* X es una función que asocia un número real a cada elemento del espacio muestral de un experimento aleatorio.

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

Al conjunto de posibles valores que puede tomar la variable aleatoria se le llama *rango* o *recorrido* de la variable, y se representa por $\operatorname{Ran}(X)$.

En el fondo, una variable aleatoria es una variable cuyos valores provienen de la realización de un experimento aleatorio, y por tanto, tendrá asociada una determinada distribución de probabilidad.

Ejemplo. La variable X que mide el resultado del lanzamiento de un dado

Tipos de variables aleatorias

Las variables aleatorias se clasifican en dos tipos:

Discretas (VAD): Toman valores aislados y su rango es numerable.
Ejemplo. Número de hijos, número de cigarrillos, número de asignaturas aprobadas, etc.

Continuas (VAC): Toman valores en un intervalo real y su rango es no numerable
Ejemplo. Peso, estatura, edad, nivel de colesterol, etc.

Los modelos probabilísticos de cada tipo de variables tienen características diferenciadas y por eso se estudiarán por separado. En este capítulo se estudian los modelos probabilísticos de las variables aleatorias discretas.

Distribución de probabilidad de una variable aleatoria discreta

Como los valores de una variable aleatoria están asociados a los sucesos elementales del correspondiente experimento aleatorio, cada valor tendrá asociada una probabilidad.

Definición (Función de probabilidad)

La *función de probabilidad* de una variable aleatoria discreta X es una función $f(x)$ que asocia a cada valor de la variable su probabilidad

$$f(x_i) = P(X = x_i).$$

Las probabilidades también pueden acumularse, al igual que se acumulaban las frecuencias en las muestras.

Definición (Función de distribución)

La *función de distribución* de una variable aleatoria discreta X es una función $F(x)$ que asocia a cada valor x_i de la variable la probabilidad de que la variable tome un valor menor o igual que x_i,

Distribución de probabilidad de una variable aleatoria discreta

Al rango de la variable, junto a su función de probabilidad o de distribución, se le llama **Distribución de probabilidad** de la variable, y se suele representar en forma de tabla

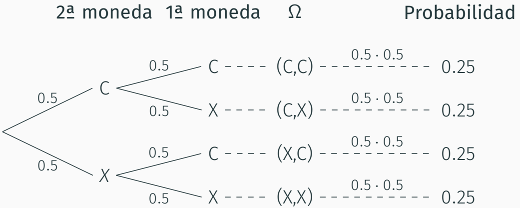
X	x_1	x_2	$\cdots$	x_n	$\sum$
$f(x)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	1
$F(x)$	$F(x_1)$	$F(x_2)$	$\cdots$	$F(x_n) = 1$	

Al igual que la distribución de frecuencias de una variable reflejaba cómo se distribuían los valores de la variable en una muestra, la distribución de probabilidad de una variable aleatoria sirve para reflejar cómo se distribuyen los valores de dicha variable en toda la población.

Distribución de probabilidad de una variable discreta

Ejemplo del lanzamiento de dos monedas

Sea X la variable aleatoria que mide el número de caras en el lanzamiento de dos monedas. El árbol de probabilidad del espacio probabilístico del experimento se muestra a continuación.



y según esto, su distribución de probabilidad es

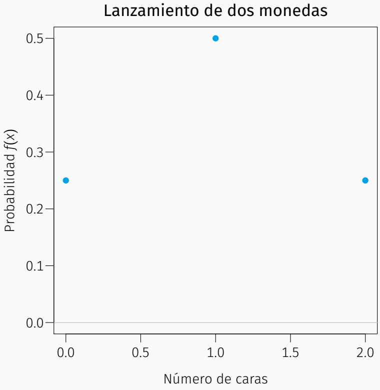
X	0	1	2
$f(x)$	0.25	0.5	0.25
$F(x)$	0.25	0.75	1

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0.25 & \text{si } 0 \leq x < 1 \\ 0.75 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

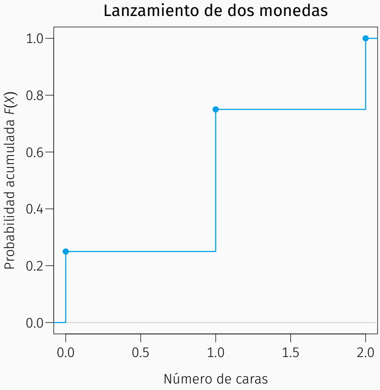
Gráficos de la distribución de probabilidad

Ejemplo del lanzamiento de dos monedas

Función de probabilidad



Función de distribución



Estadísticos poblacionales

Al igual que para describir las variables en las muestras se utilizan estadísticos descriptivos muestrales, para describir las características de las variables aleatorias se utilizan también estadísticos poblacionales.

La definición de los estadísticos poblacionales es análoga a la de los muestrales, pero utilizando probabilidades en lugar de frecuencias relativas.

Los más importantes son[1]:

- Media o esperanza matemática:

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

- Varianza:

$$\sigma^2 = Var(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$$

- Desviación típica:

$$\sigma = +\sqrt{\sigma^2}$$

[1]Para distinguirlos de los muestrales se suelen representar con letras griegas

En el experimento aleatorio del lanzamiento de dos monedas, a partir de la distribución de probabilidad

X	0	1	2
$f(x)$	0.25	0.5	0.25
$F(x)$	0.25	0.75	1

se pueden calcular fácilmente los estadísticos poblacionales

$$\mu = \sum_{i=1}^{n} x_i f(x_i) = 0 \cdot 0.25 + 1 \cdot 0.5 + 2 \cdot 0.25 = 1 \text{ cara},$$

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2 = (0^0 \cdot 0.25 + 1^2 \cdot 0.5 + 2^2 \cdot 0.25) - 1^2 = 0.5 \text{ caras}^2,$$

$$\sigma = +\sqrt{0.5} = 0.71 \text{ caras}.$$

Modelos de distribución de probabilidad

De acuerdo al tipo de experimento en el que se mide la variable aleatoria, existen diferentes modelos de distribución de probabilidad. Los más importantes son

- Distribución Uniforme.
- Distribución Binomial.
- Distribución de Poisson.

Distribución Uniforme $U(a, b)$

Cuando por la simetría del experimento, todos los valores $a = x_1, \ldots, x_k = b$ de una variable discreta X son igualmente probables, se dice que la variable sigue un *modelo de distribución uniforme*.

Definición (Distribución uniforme $U(a, b)$)

Se dice que una variable aleatoria X sigue un *modelo de distribución uniforme* de parámetros a, b, y se nota $X \sim U(a, b)$, si su rango es $Ran(X) = \{a, a+1, \ldots, b\}$, y su función de probabilidad vale

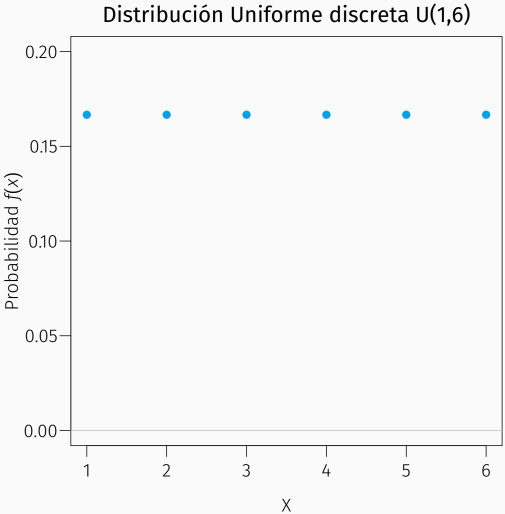
$$f(x) = \frac{1}{b-a+1}.$$

Obsérvese que a y b son el mínimo y el máximo del rango respectivamente.

Su media y varianza valen

$$\mu = \sum_{i=0}^{b-a} \frac{a+i}{b-a+1} = \frac{a+b}{2} \qquad \sigma^2 = \sum_{i=0}^{b-a} \frac{(a+i-\mu)^2}{b-a+1} = \frac{(b-a+1)^2 - 1}{12}$$

La variable que mide el número obtenido al lanzar un dado sigue un modelo de distribución uniforme $U(1, 6)$.



Distribución Binomial

La distribución binomial corresponde a una variable aleatoria medida en un experimento aleatorio con las siguientes características:

- El experimento consiste en una secuencia de n repeticiones de un mismo ensayo aleatorio.
- Los ensayos se realizan bajo idénticas condiciones, y cada uno de ellos tiene únicamente dos posibles resultados conocidos como *Éxito* o *Fracaso*.
- Los ensayos son independientes.
- La probabilidad de éxito es idéntica para todos los ensayos y vale $P(\text{Éxisto}) = p$.

En estas condiciones, la variable aleatoria X que mide le número de éxitos obtenidos en los n ensayos sigue un *modelo de distribución binomial* de parámetros n y p.

Distribución Binomial $B(n, p)$

Definición (Distribución Binomial ($B(n,p)$))

Se dice que una variable aleatoria X sigue un *modelo de distribución binomial* de parámetros n y p, y se nota $X \sim B(n,p)$, si su recorrido es $Ran(X) = \{0, 1, \ldots, n\}$ y su función de probabilidad vale

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

Obsérvese que n es conocido como el número de repeticiones de la prueba o ensayo y p como la probabilidad de Éxito en cada repetición.

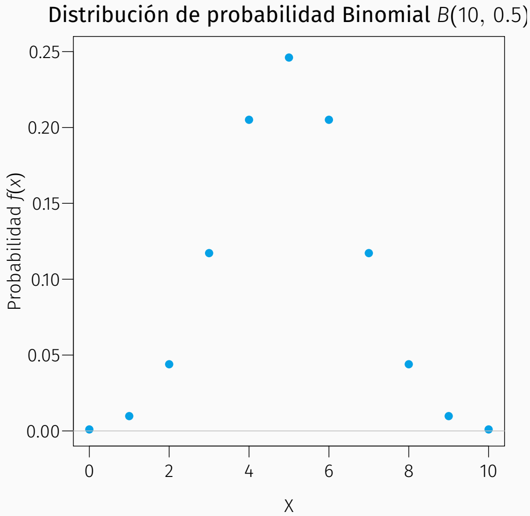
Su media y varianza valen

$$\mu = n \cdot p \qquad \sigma^2 = n \cdot p \cdot (1-p).$$

Distribución Binomial $B(n, p)$

Ejemplo de 10 lanzamientos de una moneda

La variable que mide el número de caras obtenidos al lanzar 10 veces una moneda sigue un modelo de distribución binomial $B(10, 0.5)$.



Sea $X \sim B(10,\ 0.5)$ la variable que mide el número de caras en 10 lanzamientos de una moneda. Entonces,

- La probabilidad de sacar 4 caras es

$$f(4) = \binom{10}{4} 0.5^4 (1-0.5)^{10-4} = \frac{10!}{4!6!} 0.5^4 0.5^6 = 210 \cdot 0.5^{10} = 0.2051.$$

- La probabilidad de sacar dos o menos caras es

$$\begin{aligned} F(2) &= f(0) + f(1) + f(2) = \\ &= \binom{10}{0} 0.5^0 (1-0.5)^{10-0} + \binom{10}{1} 0.5^1 (1-0.5)^{10-1} + \binom{10}{2} 0.5^2 (1-0.5)^{10} \\ &= 0.0547. \end{aligned}$$

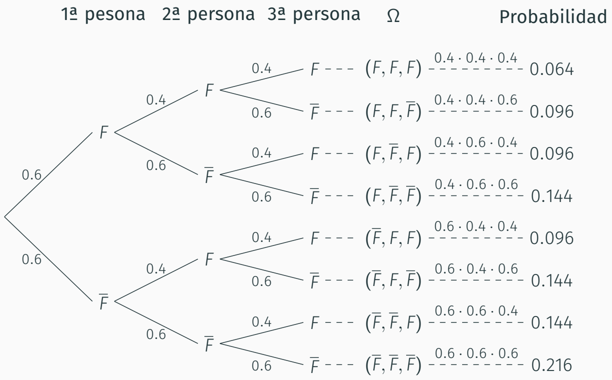
- Y el número esperado de caras es

$$\mu = 10 \cdot 0.5 = 5 \text{ caras}.$$

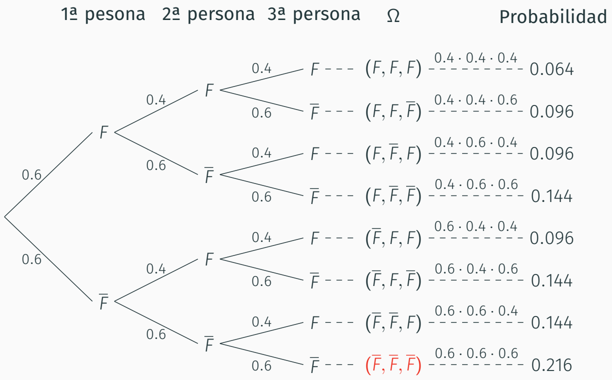
Distribución Binomial $B(n, p)$

Ejemplo de una muestra aleatoria con reemplazamiento

En una población hay un 40% de fumadores. La variable X que mide el número de fumadores en una muestra aleatoria con reemplazamiento de 3 personas sigue un modelo de distribución binomial $X \sim B(3,\ 0.6)$.



En una población hay un 40% de fumadores. La variable X que mide el número de fumadores en una muestra aleatoria con reemplazamiento de 3 personas sigue un modelo de distribución binomial $X \sim B(3,\ 0.6)$.

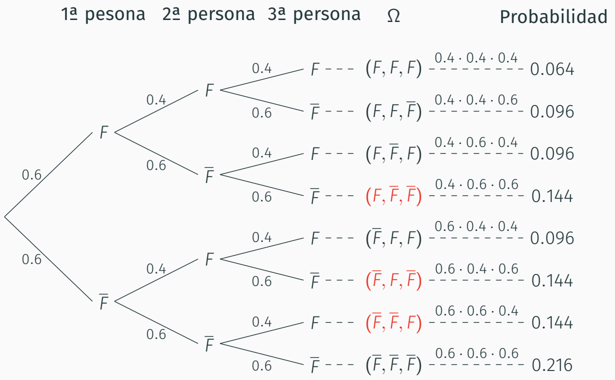


$f(0) = \binom{3}{0} 0.4^0 (1 - 0.4)^{3-0} = 0.6^3,$

Distribución Binomial $B(n, p)$

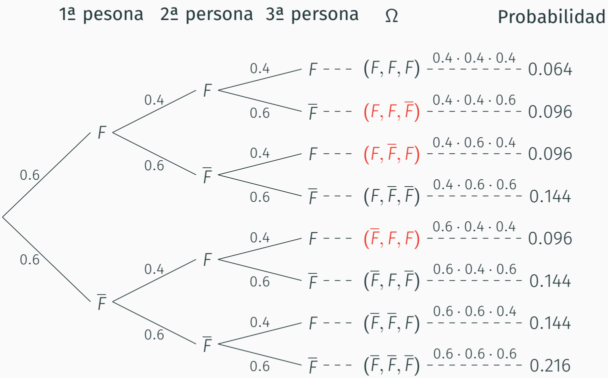
Ejemplo de una muestra aleatoria con reemplazamiento

En una población hay un 40% de fumadores. La variable X que mide el número de fumadores en una muestra aleatoria con reemplazamiento de 3 personas sigue un modelo de distribución binomial $X \sim B(3,\ 0.6)$.



$f(0) = \binom{3}{0} 0.4^0 (1 - 0.4)^{3-0} = 0.6^3,$ $\qquad$ $f(1) = \binom{3}{1} 0.4^1 (1 - 0.4)^{3-1} = 3 \cdot 0.4 \cdot 0.6^2,$

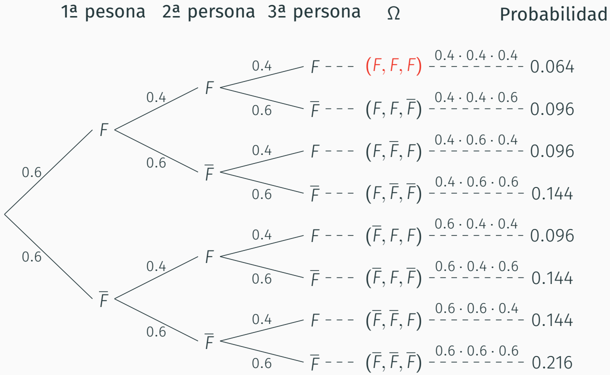
En una población hay un 40% de fumadores. La variable X que mide el número de fumadores en una muestra aleatoria con reemplazamiento de 3 personas sigue un modelo de distribución binomial $X \sim B(3,\ 0.6)$.



$f(0) = \binom{3}{0}0.4^0(1-0.4)^{3-0} = 0.6^3,$ $f(1) = \binom{3}{1}0.4^1(1-0.4)^{3-1} = 3 \cdot 0.4 \cdot 0.6^2,$

$f(2) = \binom{3}{2}0.4^2(1-0.4)^{3-2} = 3 \cdot 0.4^2 \cdot 0.6,$

En una población hay un 40% de fumadores. La variable X que mide el número de fumadores en una muestra aleatoria con reemplazamiento de 3 personas sigue un modelo de distribución binomial $X \sim B(3,\ 0.6)$.



$f(0) = \binom{3}{0} 0.4^0 (1-0.4)^{3-0} = 0.6^3$, $\quad f(1) = \binom{3}{1} 0.4^1 (1-0.4)^{3-1} = 3 \cdot 0.4 \cdot 0.6^2$,

$f(2) = \binom{3}{2} 0.4^2 (1-0.4)^{3-2} = 3 \cdot 0.4^2 \cdot 0.6$, $\quad f(3) = \binom{3}{3} 0.4^3 (1-0.4)^{3-3} = 0.4^3$.

Distribución de Poisson

La distribución de Poisson corresponde a una variable medida en un experimento aleatorio con las siguientes características:

- El experimento consiste en observar la aparición de sucesos puntuales en un intervalo fijo de tiempo o espacio. Por ejemplo, número de nacimientos en un mes, número de correos electrónicos en una hora, número de glóbulos rojos en un volumen de sangre, etc.
- Los sucesos ocurren independientemente.
- El experimento produce, a largo plazo, el mismo número medio de sucesos puntuales λ en el intervalo considerado.

En estas circunstancias, la variable aleatoria X que mide el número de ocurrencias del suceso en el intervalo considerado sigue un *modelo de distribución de Poisson* de parámetro λ.

Distribución de Poisson $P(\lambda)$

Definición (Distribución de Poisson $P(\lambda)$)

Se dice que una variable aleatoria X sigue un *modelo de distribución de Poisson* de parámetro λ, y se nota $X \sim P(\lambda)$, si su rango es $Ran(X) = \{0, 1, \ldots, \infty\}$, y su función de probabilidad vale

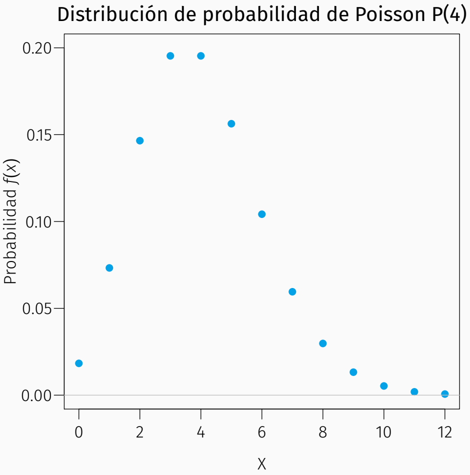
$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Obsérvese que λ es el número medio de sucesos en el intervalo unidad de referencia, y cambiará si se cambia la amplitud del intervalo.

Su media y varianza valen

$$\mu = \lambda \qquad \sigma^2 = \lambda.$$

En una ciudad hay 4 nacimientos diarios por término medio. La variable aleatoria X que mide el número de nacimientos diarios en la ciudad sigue un modelo de distribución de probabilidad de Poisson $X \sim P(4)$.



Sea $X \sim P(4)$ la variable que mide el número de ingresos diarios en un hospital. Entonces,

- La probabilidad de que un día cualquiera se produzcan 5 nacimientos es
$$f(5) = e^{-4}\frac{4^5}{5!} = 0.1563.$$
- La probabilidad de que un día se produzcan menos de 2 nacimientos es
$$F(1) = f(0) + f(1) = e^{-4}\frac{4^0}{0!} + e^{-4}\frac{4^1}{1!} = 5e^{-4} = 0.0916.$$
- La probabilidad de que un día se produzcan más de un 1 nacimiento es
$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0.0916 = 0.9084.$$

El modelo de distribución de Poisson surge a partir del modelo de distribución Binomial, cuando el número de repeticiones del ensayo tiende a infinito y la probabilidad de Éxito tiende a cero.

Teorema (Ley de los casos raros)

La distribución Binomial $X \sim B(n, p)$ tiende a la distribución de Poisson $P(\lambda)$, con $\lambda = n \cdot p$, cuando n tiende a infinito y p tiende a cero, es decir,

$$\lim_{n\to\infty, p\to 0} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}.$$

En la práctica, esta aproximación suele utilizarse para $n \geq 30$ y $p \leq 0.1$.

Se sabe que una vacuna produce una reacción adversa en el 4% de los casos. Si se vacunan una muestra 50 personas, ¿cuál es la probabilidad de que haya más de 2 personas con reacción adversa?

La variable que mide el número de personas con reacción adversa en la muestra sigue un modelo de distribución binomial $X \sim B(50, 0.04)$, pero como $n = 50 > 30$ y $p = 0.04 < 0.1$, se cumplen las condiciones de la ley de los casos raros y se utilizar la distribución de Poisson $P(50 \cdot 0.04) = P(2)$ para realizar los cálculos.

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) = 1 - f(0) - f(1) - f(2) = 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} - e^{-2}\frac{2^2}{2!} = \\
&= 1 - 5e^{-2} = 0.3233.
\end{aligned}
$$