Tema 3: PROBABILIDAD - IV

La variable aleatoria binomial. Variables aleatorias discretas.

Biología sanitaria 2017/18. Universidad de Alcalá

M. Marvá. Actualizado: 2017-10-16

Un problema de genética

Ejemplo En una familia, la probabilidad de que un descendiente tenga los ojos verdes es de 0.2. Si hay 5 descendientes, ¿cuál es la probabilidad de que 2 de ellos tengan los ojos verdes?

$P(V) = 0.2$, $P(\bar{V}) = 0.8$

Podemos hacer una lista de casos:

$C_i \equiv$ caso i, $d_j \equiv$ descendiente j

	d_1	d_2	d_3	d_4	d_5
C_1	**V**	**V**			
C_2	**V**		**V**		
C_3	**V**			**V**	
C_4	**V**				**V**
C_5		**V**	**V**		
C_6		**V**		**V**	
C_7		**V**			**V**
C_8			**V**	**V**	
C_9			**V**		**V**
C_{10}				**V**	**V**

Hay $\binom{5}{2} = \dfrac{5!}{3!(5-3)!} = 10$ casos diferentes

En el lenguaje de la probabilidad

X = “número de descendientes con ojos verdes de entre 5 de ellos, con $P(V) = 0.2$”

$$\begin{aligned} P(X=2) &= P(C_1 \cup ... \cup C_{10}) \\ &\quad \text{son incompatibles} \\ &= P(C_1) + ... + P(C_{10}) \end{aligned}$$

Además, para C_1

$$\begin{aligned} P(C_1 2) &= P\left(V \cup V \cup \bar{V} \cup \bar{V} \cup \bar{V} \cup \bar{V}\right) \\ &\text{son independientes} \\ &= P(V)P(V)P(\bar{V})P(\bar{V})P(\bar{V}) \\ &= 0.2 * 0.2 * 0.8 * 0.8 * 0.8 = (0.2)^2 * (0.3)^3 \end{aligned}$$

Resumiendo

$$P(X=2) = \binom{5}{2}(0.2)^2 * (0.3)^3$$

Variable aleatoria

Es cualquier aplicación que asigna un valor numérico a cada evento.

Ejemplo: Experimento, lanzar un dado.

$$X = \begin{cases} 0 & \text{si sale cara} \\ 1 & \text{si sale cruz} \end{cases}$$

Bernoulli

Un **experimento de Bernoulli** es un experimento aleatorio con dos resultados posibles que llamamos, arbitrariamente:

- éxito E con probabilidad p
- fracaso F con probabilidad $q = 1 - p$

La **variable aleatoria de Bernoullí** se define como

$$X = \begin{cases} 0 & \text{fracaso} \\ 1 & \text{éxito} \end{cases}$$

Se usa la notación:

$$P(E) = P(X = 1) = p, \qquad P(F) = P(X = 0) = q = 1 - p$$

Ejemplo: $X =$ "Número de descendientes con ojos verdes de un total 5", con $P(V) = 0.2$.

Entonces, $P(X = 2) = \binom{5}{2}(0.2)^2(0.8)^3$ se calcula con

```
dbinom(2, size = 5, prob = 0.2)
```

```
## [1] 0.2048
```

Observa que

- Se repite un experimento Bernouillí (**dicotómico**) 5 veces.
- Las repeticiones son **independientes**.

Variable aleatoria binomial de parámetros n y p

Es la variable X que representa el número de éxitos al repetir n experimentos independientes de Bernoulli, cada uno con probabilidad p de éxito (y con $q = 1 - p$). Si $X \sim B(n, p)$ se tiene que

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

Ejemplo: $X =$ "Número de descendientes con ojos verdes de un total 5, con $P(V) = 0.2$".

Calculamos la probabilidad de cada valor de X y hacemos una tabla

```
tabla = rbind(0:5, dbinom(0:5, size = 5, prob = 0.2))
rownames(tabla) = c("X", "P(X=x)"); colnames(tabla) = rep(x = "", times = 6
tabla
```

```
##
## X       0.00000 1.0000 2.0000 3.0000 4.0000 5.00000
## P(X=x) 0.32768 0.4096 0.2048 0.0512 0.0064 0.00032
```

hemos obtenido la **función de densidad de probabilidad de** $X \sim B(5, 0.2)$

Ejemplo: Zoo binomial.

Ejemplo: $W =$ "Suma del resultado de lanzar dos dados".

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

Valor de la suma:	2	3	4	5	6	7	8	9	10	11	12
Probabilidad de ese valor:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

hemos obtenido la **función de densidad de probabilidad de** W

Formalicemos lo anterior

Para nuestros propósitos, una **variable aleatoria discreta** es aquella que sólo puede tomar una cantidad finita de valores.

Si X sólo toma una cantidad finita de valores numéricos $x_1, x_2, x_3, \ldots, x_k$, con probabilidades $p_i = P(X = x_i)$, la tabla

Valor:	x_1	x_2	x_3	$\cdots$	x_k
Probabilidad:	p_1	p_2	p_3	$\cdots$	p_k

es la **función de densidad de probabilidad, o función de masa** de la variable aleatoria X.

Ejemplos: las anteriores son variables aleatorias discretas

El valor esperado de una variable aleatoria discreta

Inspirado en la media de una variable estadística discreta que toma valores x_1, x_2,..., x_k con frecuencias absolutas f_1, f_2,..., f_k:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n} = \sum_{i=1}^{k} x_i \cdot \frac{f_i}{n}$$

y aquí $\dfrac{f_i}{n}$ es la frecuencia relativa número i, que se identifica con su probabilidad.

Si X es una variable aleatoria discreta, que toma los valores $x_1, x_2, \ldots, x_k$, con las probabilidades $p_1, p_2, \ldots, p_k$ (donde $p_i = P(X = x_i)$), entonces la **media**, o **valor esperado**, o **esperanza matemática** de X es:

$$\mu = \sum_{i=1}^{k} x_i \cdot P(X = x_i) = x_1\, p_1 + x_2\, p_2 + \cdots + x_k\, p_k.$$

Se usa la notación $E[X] = \mu$.

Ejemplo: X = “Número de descendientes con ojos verdes de un total 5, con $P(V) = 0.2$”.

```
valores = 0:5
probabilidades = dbinom(0:5, size = 5, prob = 0.2)
```

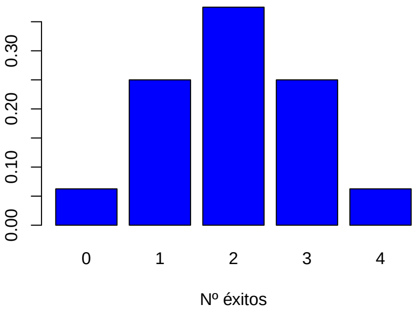
Juego justo

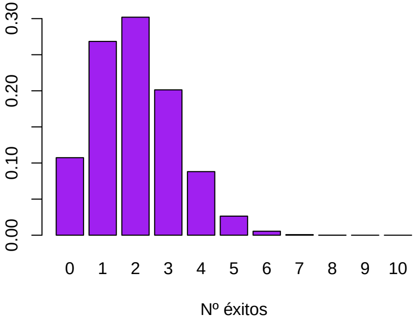
Tenemos una caja con 7 bolas blancas y 4 bolas negras. El juego consiste en lo siguiente.

- Tu pones un euro,
- Yo pongo x euros.
- Sacamos una bola de la caja al azar.
- Si es negra, ganas tú y te quedas con todo el dinero (tu apuesta y la mía).
- Si la bola es blanca, gano yo y me quedo todo el dinero.

¿Cuántos euros debo poner yo para que el juego sea justo?

Ejemplo: Sean $X \sim B(4, 0.5)$ e $Y \sim B(10, 0.2)$,





Las probabilidades P(Y=7), P(Y=8), P(Y=9), P(Y=10) son muy pequeñas.

Ambas cumplen

$$\mu_X = 2, \qquad \mu_Y = 2$$

pero en una los valores están más agrupados que la otra

La varianza de una variable aleatoria discreta

Inspirado en la media de una variable estadística discreta que toma valores x_1, x_2,..., x_k con frecuencias absolutas f_1, f_2,..., f_k:

$$Var(x) = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n} = \sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot \frac{f_i}{n}$$

aquí $\frac{f_i}{n}$ es la frecuencia relativa número i, que se identifica con su probabilidad.

Si X es una variable aleatoria discreta, que toma los valores $x_1, x_2, \ldots, x_k$, con las probabilidades $p_1, p_2, \ldots, p_k$ (donde $p_i = P(X = x_i)$) y valor esperado μ, entonces la **varianza** de X es:

$$\sigma_X^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot P(X = x_i) = (x_1 - \mu)^2 p_1 + \cdots + (x_k - \mu)^2 p_k.$$

y su **desviación típica**

$$\sigma_X = \sqrt{\sum_{i=1}^{k} (x_i - \mu)^2 \cdot P(X = x_i)}.$$

Ejemplo: $X =$ "Número de descendientes con ojos verdes de un total 5, con $P(V) = 0.2$".

```
valores_X = 0:5
probabilidades_X = dbinom(0:5, size = 5, prob = 0.2)
(mu_X= sum(valores_X*probabilidades_X))
```

```
## [1] 1
```

```
(sigma2_X=sum((valores_X-mu_X)^2*probabilidades_X))
```

```
## [1] 0.8
```

Ejemplo: Análogo al anterior, P(V)=0.2

total_descendientes	Valor_esperado	varianza
1	0.2	0.16
2	0.4	0.32
3	0.6	0.48
4	0.8	0.64
5	1.0	0.80

Ejemplo: Y_1 = "Resultado de lanzar 1 dado".

```
valores_Y1 = 1:6
probabilidades_Y1 = rep(1,6)/6
(mu_Y1= sum(valores_Y1*probabilidades_Y1))
```

```
## [1] 3.5
```

```
(sigma2_Y1=sum((valores_Y1-mu_Y1)^2*probabilidades_Y1))
```

```
## [1] 2.916667
```

Ejemplo: Y = "Suma del resultado de lanzar 2 dados".

```
valores_Y2 = 2:12
probabilidades_Y2 = c(1:6,5:1)/36
(mu_Y2= sum(valores_Y2*probabilidades_Y2))
```

```
## [1] 7
```

```
(sigma2_Y2=sum((valores_Y2-mu_Y2)^2*probabilidades_Y2))
```

```
## [1] 5.833333
```

Propiedades de la media y la varianza

Y si X_1, X_2 son dos variables aleatorias, se tiene:

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Si además X_1 y X_2 son *independientes*, entonces

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2).$$

Si X es una variable aleatoria, y a, b son números cualesquiera, entonces

$$E(a \cdot X + b) = a \cdot E(X) + b, \quad \mathrm{Var}(a \cdot X + b) = a^2 \cdot \mathrm{Var}(X).$$

Además, para cualquier variable aleatoria X con media μ y varianza σ^2

la variable aleatoria

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

es una nueva variable aleatoria que cumple:

$$\mu_Z = 0 \qquad \text{y} \qquad \sigma_Z^2 = 1.$$

Se dice que Z es una **variable aleatoria estandarizada**.

Ejemplo: Supongamos que $X \sim Bernoulli(p)$. Entonces

valores	X = 0	X = 1
Probabilidades	P(X = 0) = q = 1-p	P(X = 1) = p

La media es

$$\mu_X = 0 \cdot q + 1 \cdot p = p$$

La varianza es

$$\sigma_X^2 = (0 - p)^2 \cdot q + (1 - p)^2 \cdot p = p \cdot q$$

Ejemplo: Supongamos que $X \sim B(n, p)$. Podemos ver X como la suma de n variables aleatorias Bernoullí *Bernoulli*(p) independientes

$$X = X_1 + X_2 + \cdots + X_n, \qquad X_i \sim Bernoulli(p)$$

por tanto, la media es

$$\begin{aligned} \mu_X = E[X] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] \\ &= p + \cdots + p \\ &= n \cdot p \end{aligned}$$

La varianza es

$$\begin{aligned} \sigma_X^2 = \sigma^2[X] &= \sigma^2[X_1 + \cdots + X_n] \\ &= \sigma^2[X_1] + \cdots + \sigma^2[X_n] \\ &= p \cdot q + \cdots + p \cdot q \\ &= n \cdot p \cdot q \end{aligned}$$

Y la desviación típica

$$\sigma_X = \sqrt{n \cdot p \cdot q}$$

Probabilidades acumuladas de una variable discreta

Inspirados en las frecuencias relativas acumuladas

Ejemplo: $X =$ "Nº de descendientes con ojos verdes de un total de 5, con $P(V) = 0.2$".

Probabilidades de **que haya al menos** x descendientes con ojos verdes de un total de 5.

```
tabla = rbind(0:5, pbinom(0:5, size = 5, prob = 0.2))
rownames(tabla) = c("X", "P(X<=x)")
colnames(tabla) = rep(x = "", times = 6)
tabla
```

```
##
## X        0.00000 1.00000 2.00000 3.00000 4.00000 5
## P(X<=x) 0.32768 0.73728 0.94208 0.99328 0.99968 1
```

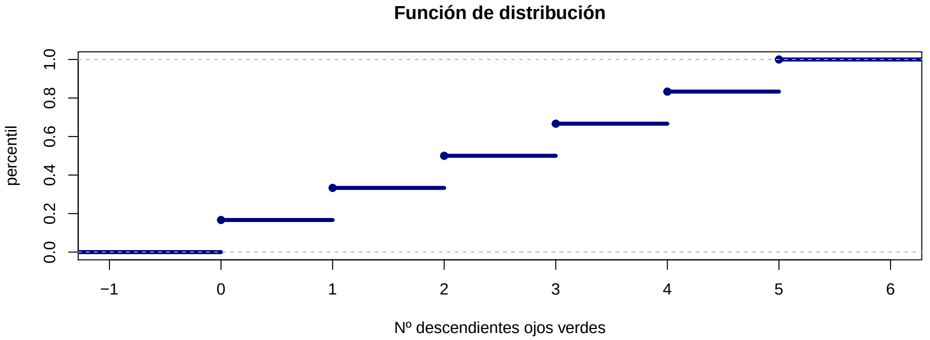
es la **función de distribución de probabilidades** de $X \sim B(5, 0.2)$.

Probabilidades de **que haya al menos** x descendientes con ojos verdes de un total de 5.

```
##
## X        0.00000 1.00000 2.00000 3.00000 4.00000 5
## P(X<=x) 0.32768 0.73728 0.94208 0.99328 0.99968 1
```

Visualización de la **función de distribución**

```r
funcionDistribucion = ecdf(tabla[1, ])
plot(funcionDistribucion, lwd=4, col="navy",
     main="Función de distribución",
     xlab="Nº descendientes ojos verdes", ylab="percentil")
```



Formalicemos lo anterior

Si X es una variable aleatoria que sólo toma una cantidad finita de valores numéricos $x_1, x_2, x_3, \ldots, x_k$, con probabilidades $p_i = P(X = x_i)$, la tabla

Valor:	x_1	x_2	x_3	$\cdots$	x_k
Probabilidad:	p_1	$p_1 + p_2$	$p_1 + p_2 + p_3$	$\cdots$	$p_1 + \cdots + p_k$

es la **función de distribución de probabilidad** de la variable aleatoria X.

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

Unma variable aleatoria continua

Ejercicio: Sea $X =$ "Nº de horas de espera en el servicio de urgencias". A partir de una muestra grande se determina que

Minutos espera	Nº individuos
$[0, 60)$	25
$[60, 120)$	35
$[120, 180)$	20
$[180, 240)$	15
$[240, 300)$	3
Más de 300	2

1. Usa las frecuencias relativas para calcular la probabilidad de esperar entre 0 y 59 minutos, entre 60 y 119,. . .
2. Calcula la probabilidad de esperar entre 60 y 180 minutos.
3. Calcula la probabilidad de esperar entre 60 y 150 minutos.