Veremos en lo que sigue nuevas variables aleatorias discretas. Estas variables y sus distribuciones se utilizan como modelos en muchas aplicaciones estadísticas.

Distribución Binomial

Consideremos los siguientes experimentos aleatorios

- ε_1 : Una moneda es arrojada 4 veces.
- ε_2 : Una caja contiene 1 bolita roja y 9 bolitas negras. Se realizan 5 extracciones con reposición.
- ε_3 : Un lote contiene artículos que cumplen (E) con las especificaciones y otros que no las cumplen (F).

Los experimentos ε_1 , ε_2 , ε_3 son experimentos binomiales

Definición Un experimento se denomina Binomial si satisface las condiciones siguientes

1. El experimento consta de *n* pruebas o repeticiones, *n* se fija antes de realizar el experimento.
2. Las pruebas son idénticas y en cada prueba puede ocurrir uno de los dos sucesos posibles, que denominamos Éxito (E) y Fracaso (F).
3. Las pruebas son independientes, es decir que el resultado de una prueba no influye sobre el de las otras.
4. La probabilidad de Éxito (P(E)=p) se mantiene constante en todas las pruebas.

Cada prueba se denomina **ensayo de Bernoulli**.

Ejemplos

- ε_1 , Éxito es el suceso "sale cara".
- ε_2 , Éxito es el suceso "se extrae una bolita roja"

Consideremos el siguiente experimento

- ε_4 : Una caja contiene 1 bolita roja y 9 bolitas negras. Se realizan 5 extracciones sin reposición, Éxito es el suceso "se obtiene una bolita roja"

Es ε_4 un experimento Binomial? Sean B_1 el suceso salió roja en la primera extracción y B_2 el suceso salió roja en la segunda extracción

$$P(B_2 \mid B_1) = 0 \neq P(B_2) = \frac{1}{10}$$

- no se verifica la tercera condición de experimento binomial
- tampoco se verifica la segunda ya que las pruebas no son idénticas (la composición de la urna varía)
- la cuarta condición sí se satisface.

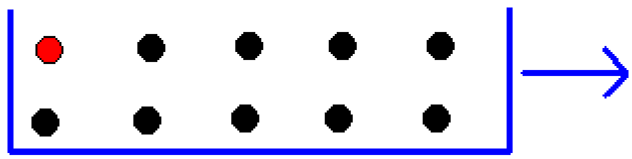
Variable aleatoria binomial

Definición: Sea X = cantidad de éxitos en un experimento Binomial que consta de n pruebas con probabilidad de éxito p. Decimos que X es una variable aleatoria con distribución Binomial de parámetros n y p.

Notación: X ~ Bi (n,p).

Ejemplo: Seguimos con el experimento ε_2

X = cantidad de bolitas rojas extraídas



5 bolitas con reposición

$n = 5 \qquad p = 0.1$

$X \sim Bi(5, 0.1)$, $R_X =$

$P(X = 1) = p_X(1) = P(1 \text{ roja})$

R N N N N
N R N N N

N N R N N 1 roja = 5 sucesos elementales = $\binom{5}{1} = \dfrac{5!}{1!4!}$

N N N R N
N N N N R

P(1 roja) = P (R N N N N) + P (N R N N N) +
P (N N R N N) + P (N N N R N) + P (N N N N R)
= 5 P (R N N N N) = 5 (0.1) $(0.9)^4$

Luego

$$P(X = 1) = \frac{5!}{1!4!}(0.1)^1 (0.9)^{5-1}$$

X = 2 R R N N N   N R R N N   N N R N R
R N R N N   N R N R N   N N N R R
R N N R N   N R N N R

R N N N R   N N R R N $\binom{5}{2} = \frac{5!}{2!3!} = 10$

P (X = 2) = $\binom{5}{2}$ $(0.1)^2\,(0.9)^{5-2}$

P (X = k) = $\binom{5}{k}$ $(0.1)^k\,(0.9)^{5-k}$ k: 0, 1, 2, 3, 4, 5

¿Cuál es el espacio muestral asociado a ε_2?

¿Es de equiprobabilidad?

<u>En general:</u>

Si X ~ Bi (n,p)

i) $p_X(k) = \binom{n}{k}\,(p)^k\,(1-p)^{n-k}$ k: 0, 1, . . ., n

Verifiquemos que $\sum_{k=0}^{n} p_X(k) = 1.$

$$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \left(p + (1-p)\right)^n = 1^n = 1.$$

Hemos usado la fórmula del Binomio de Newton: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$

ii) E (X) = n p

iii) Var (X) = n p (1- p)

Ejemplo:

Se fabrican diariamente N = 50 000 unidades de un producto. Se extrae una muestra de n = 10 unidades sin reposición. Llamamos éxito (E) al suceso ‘la unidad es buena’

¿ Es un experimento binomial? ....

Si $n / N < 0.05$, si se extrae menos del 5% de la población, se puede usar la aproximación binomial aunque el muestreo se realice sin reposición

Supongamos que 40 000 de las 50 000 son buenas

P (4ª buena) = 0.8

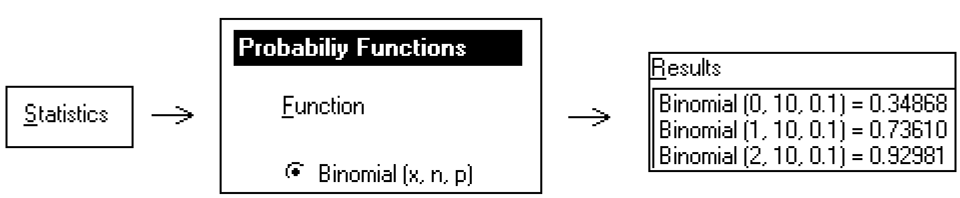
P (4ª buena / salieron 1ª, 2ª, 3ª malas) = $\frac{40000}{49997}$ = 0.800048

P (4ª buena / salieron 1ª, 2ª, 3ª buenas) = $\frac{39997}{49997}$ =0.799988

Falla “un poco” la condición de independencia.

Ejemplo: $X \sim Bi(10, 0.1)$

Tablas $P(X \leq x) = F_X(x)$ Función de Distribución Acumulada



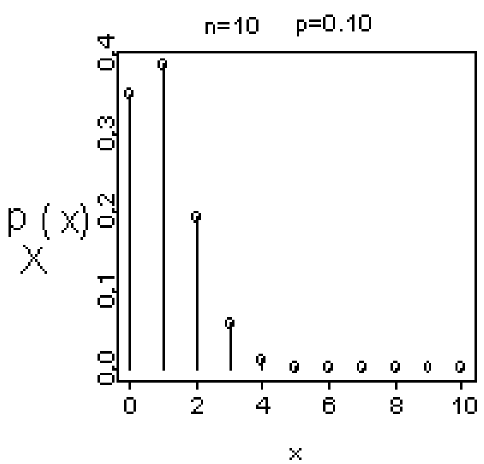
$p_X(0) = P(X \leq 0) = 0.349$

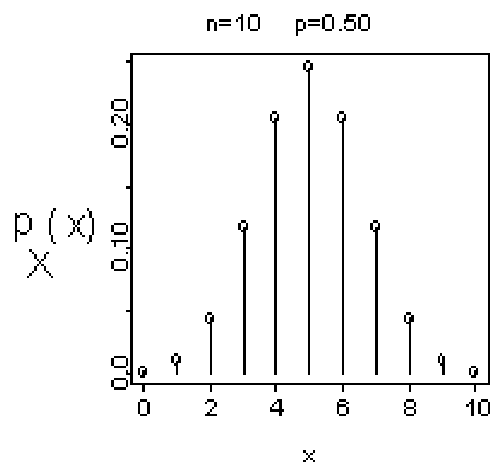
$p_X(1) = P(X \leq 1) - P(X \leq 0) = 0.736 - 0.349 = 0.387$

$p_X(2) = P(X \leq 2) - P(X \leq 1) = 0.930 - 0.736 = 0.194$

....

....





$X \sim Bi(10, 0.5)$

$p_X(0) = 0.001 = p_X(10)$

$p_X(1) = 0.010 = p_X(9)$

$p_X(2) = 0.044 = p_X(8)$

La v.a. X se distribuye en forma simétrica alrededor de ....

Ejemplo.

Un lote contiene tornillos, se eligen 10 tornillos al azar con reposición y se observa que 5 son defectuosos.

a) Suponiendo que el lote tiene 50% de tornillos defectuosos. Calcule la probabilidad de que 5 o más tornillos sean defectuosos.

X = cantidad de tornillos defectuosos entre los 10 elegidos

$X \sim Bi\ (10,\ 0.5)$

$P\ (\ X \geq 5\ ) = 1 - P\ (\ X \leq\ 4) = 1 - 0.377 = 0.623$

b) ídem a) suponiendo que el lote tiene 20% de tornillos defectuosos

$X \sim Bi\ (10,\ 0.2)$

$P\ (\ X \geq 5\ ) = 1 - P\ (\ X \leq\ 4) = 1 - 0.967 = 0.033$

c) ídem a) suponiendo que el lote tiene 10% de tornillos defectuosos

$X \sim Bi\ (10,\ 0.1)$

$P\ (\ X \geq 5\ ) = 1 - P\ (\ X \leq\ 4) = 1 - 0.998 = 0.002$

Distribución Hipergeométrica

Hemos visto que la distribución binomial es un modelo probabilístico aproximado para el caso de un muestreo sin reemplazo.

La distribución hipergeométrica es el modelo probabilístico exacto para un muestreo sin reemplazo de una población finita dicotómica (E-F)

M = tamaño de la población
- N cantidad de Éxitos
- M - N cantidad de Fracasos

Se elige al azar, de la población, una muestra de tamaño n sin reemplazo.

X = cantidad de éxitos en la muestra

X ~ H (M, N, n)

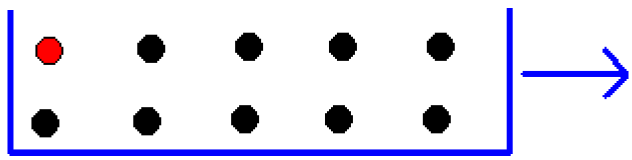
$$p_X(x) = \frac{\binom{N}{x}\binom{M-N}{n-x}}{\binom{M}{n}} \qquad x : (\max\{0, n-(M-N)\}, \cdots, \min\{n, N\})$$

Si en la muestra hay 0 éxitos significa que n son fracasos, pero el total de fracasos en la población es M-N.

Luego si el tamaño de la muestra n > M-N, habrá como mínimo n - (M-N) > 0 éxitos y no podrá ser cero.

Para que en la muestra haya n éxitos, n $\leq$ N.

Ejemplo: Seguimos con el experimento ε_4



5 extracciones sin reposición

X = cantidad de bolitas rojas de la muestra

$R_X = \{\ 0, 1\ \}$

$$p_X(0) = \frac{\binom{1}{0}\binom{10-1}{5-0}}{\binom{10}{5}} = \frac{\binom{1}{0}\binom{9}{5}}{\binom{10}{5}} = \frac{1}{2} \qquad\qquad p_X(1) = \frac{\binom{1}{0}\binom{9}{4}}{\binom{10}{5}} = \frac{1}{2}$$

Esperanza y varianza de la distribución hipergeométrica

$$E(X) = n\frac{N}{M} \qquad\qquad \mathrm{Var}(X) = \left(\frac{M-n}{M-1}\right) n\frac{N}{M}\left(1-\frac{N}{M}\right)$$

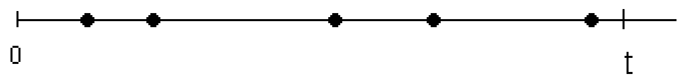
Comparación con la distribución binomial

Se extraen con reposición n elementos de una población de tamaño M, con N éxitos, p = P (éxito) = $\frac{N}{M}$

$E(X) = n\frac{N}{M}$ igual que la distr. hipergeométrica

$\mathrm{Var}(X) = n\frac{N}{M}\left(1-\frac{N}{M}\right)$ mayor que la distr. hipergeométrica

Proceso de Poisson



Sea X_t = cantidad de pulsos radioactivos recibidos por un contador Geiger en un intervalo de longitud t.

Vamos a realizar las siguientes suposiciones sobre la forma en que pueden ocurrir los pulsos:

1) Para cualquier intervalo de longitud pequeña, Δt la probabilidad de recibir exactamente 1 pulso es proporcional a la longitud del intervalo, a menos de un término $o(\Delta t)$ tal que $\lim_{h \to 0} \frac{o(\Delta t)}{\Delta t} = 0$ es decir:

$$P(\text{ocurra 1 pulso en } \Delta t) = \theta\, \Delta t + o(\Delta t)$$

2) La probabilidad de recibir más de 1 pulso en un intervalo longitud pequeña Δt es prácticamente cero:

$$P(\text{recibir más de un pulso en } \Delta t) = o(\Delta t)$$

3) El número de pulsos recibidos en un intervalo de longitud Δt es independiente del número de pulsos recibidos anteriormente a ese intervalo de tiempo.

¿Cómo se interpreta la cantidad θ?

θ = valor esperado o valor medio de pulsos en la unidad de tiempo: intensidad

Teorema: Si se cumplen las suposiciones 1, 2 y 3 entonces

$$P(X_t = k) \quad = p_{X_t}(k) = \frac{(\theta\, t)^k e^{-\theta t}}{k!} \qquad\qquad k: 0, 1, \cdots$$

Se dice que X_t sigue un Proceso de Poisson.

Si nos interesa un intervalo de longitud t fija, decimos que X_t tiene distribución Poisson de parámetro θ t

$$\boxed{X_t \sim \mathcal{P}(\theta\, t)}$$

Tenemos una distribución de probabilidades válida, en efecto si llamamos $\lambda = \theta$ t, $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$ k : 0, 1, 2, ...

i) $p_X(k) \geq 0 \qquad \forall k$.

ii) $\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$

El modelo de Poisson suele ser adecuado en los casos en que interesa estudiar:

- el número de imperfecciones que presenta un rollo de tela que se produce en forma continua

- el número de glóbulos rojos que se observan en un porta objetos

- número de partículas emitidas por una fuente radioactiva

- número de automóviles que pasan por una esquina

- número de llamadas telefónicas recibidas por una central

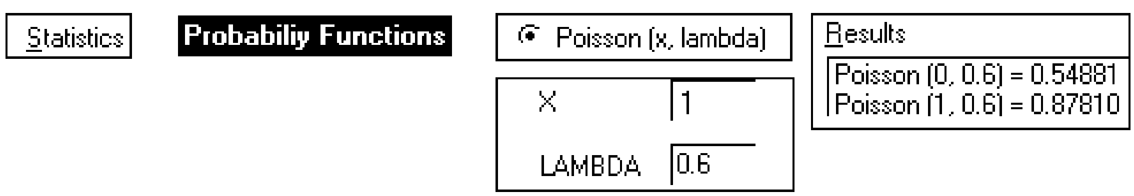
Ejemplo. En un proceso electrolítico continuo de aplicación de estaño, se descubren en promedio 0.2 imperfecciones por minuto ¿Cuál es la probabilidad de descubrir

a) una imperfección en 3 minutos?

X_3 = número de imperfecciones en 3 minutos
$t = 3$ minutos , $\theta = 0.2$ imperfecciones / minuto
$\theta\, t = 0.6$ imp

$X_3 \sim \mathcal{P}(0.6)$

Nuevamente podemos hallar la función de distribución <u>acumulada</u>:



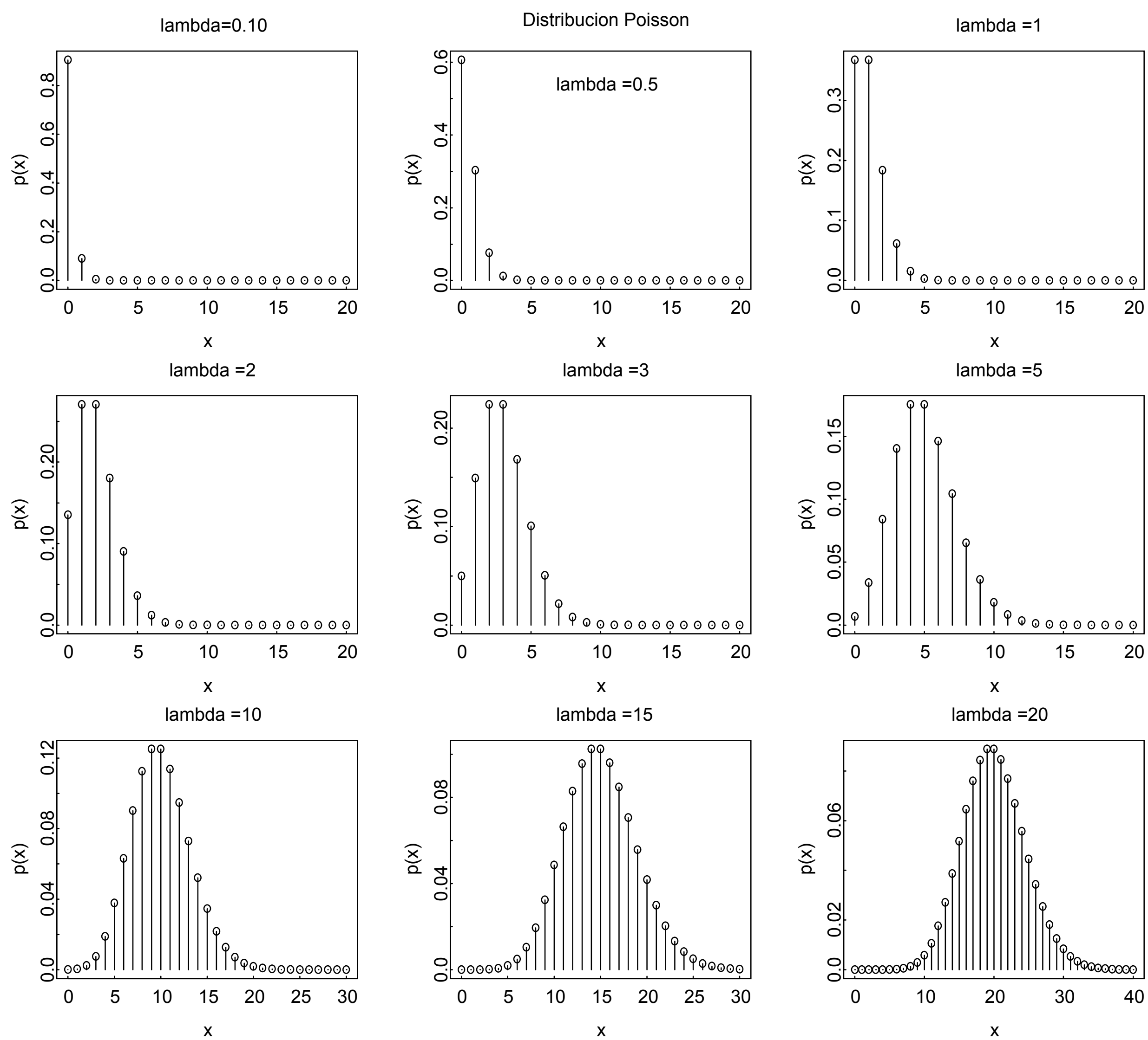
Luego $P(X_3 = 1) = P(X_3 \leq 1) - P(X_3 \leq 0)$

$$= 0.878 - 0.549 = 0.329$$

b) al menos dos imperfecciones en 5 minutos?

Forma de la distribución Poisson

Función de probabilidad puntual, para la distribución de Poisson para distintos valores de λ.



La distribución se simetriza alrededor de λ a medida que este parámetro crece.

Aproximación de la distribución Binomial por la distr. Poisson

Proposición: Sea X ~ Bi(n,p) y sea Y ~ $\mathcal{P}$ (λ), supongamos que $n \to \infty$ y $p \to 0$, de manera que $n \cdot p = \lambda$ (fijo), entonces :

$$p_X(k) \longrightarrow p_Y(k) \qquad \forall k \in N_o = N \cup \{0\}$$

Dem:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)...(n-k+1)}{n^k}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!}$$

$$= \left[\frac{n}{n}\frac{n-1}{n}....\frac{n-k+1}{n}\right]\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!}.$$

Observemos que:

$$1 \cdot \frac{n-1}{n}....\frac{n-k+1}{n} \xrightarrow[n\to\infty]{} 1$$

$$\left(1-\frac{\lambda}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\lambda}$$

$$\left(1-\frac{\lambda}{n}\right)^{-k} \xrightarrow[n\to\infty]{} 1$$

Entonces, $p_X(k) \longrightarrow \dfrac{e^{-\lambda}\lambda^k}{k!} = p_Y(k)$, como queríamos demostrar.

Ejemplo. Se tiran 2 dados equilibrados 100 veces y se registra

X = n^{o} de veces que sale el doble seis

$X \sim Bi\ (100,\ 1/36)$ $p = 1/36 = 0.0278$

$\lambda = n\ p = 100\ .\ 1\ /\ 36 = 2.78$

n grande
p pequeño   aproximo la binomial por $\mathcal{P}\ (2.78)$

k	Prob. exacta (Binomial)	Prob. Poisson
0	0.0598	0.0622
1	0.1708	0.1727
2	0.2416	0.2399
5	0.0857	0.0857
8	0.0049	0.0055
9	0.0014	0.0017
10	0.0004	0.0005
11	0.0001	0.0001

¿Cuándo podemos utilizar la aproximación?

<u>REGLAS PRÁCTICAS</u>

- $n \geq 20$ admisible
- $p \leq 0.05$ $n \geq 100$ $n\ p \leq 10$ excelente
- $p \leq 0.01$ $n \geq 100$ $n\ p \leq 20$ excelente

$\boxed{\text{Proposición: } X \sim \mathcal{P}(\lambda) \text{ entonces } \quad E(X) = \lambda \qquad Var(X) = \lambda}$