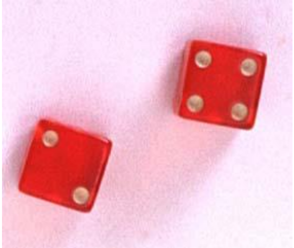
Biometría

Distribuciones de probabilidad para variables aleatorias discretas

(Binomial, Hipergeométrica y Poisson)

Variable aleatoria



- El resultado de un experimento aleatorio puede ser descripto en ocasiones como una cantidad numérica
- En estos casos aparece la noción de variable aleatoria
  - es una descripción numérica del resultado de un experimento aleatorio
  - es imposible conocer de antemano el valor que tomará en una repetición cualquiera de dicho experimento
  - El rango de posibles valores que puede tomar se conoce como dominio

Tipos de variables aleatorias

- **V.A. discretas**: toman valores numéricos, contables, finitos o infinitos

- **V.A. continuas:** toman valores numéricos, no contables e infinitos

Identificar en los siguientes experimentos aleatorios, la VA y su dominio

1) Responder por azar un examen choice de 20 preguntas y contar el número de respuestas correctas. Cada pregunta tiene 4 ítems.
2) Contar el número de larvas muertas luego de la aplicación de un herbicida en un lote de 50 larvas; la tasa de mortalidad es del 5%
3) Contar los incendios forestales que se producen en el Parque Nacional Nahuel Huapi en un año. Las estadísticas indican uno cada 6 meses en promedio
4) De un curso de 70 alumnos, en donde sólo trabajan 12, se tomó una muestra de 5 y se contó el número de alumnos que trabaja
5) Ídem 4, pero en el curso sólo trabajan 3 alumnos (no 12)

Distribución de probabilidad de una V.A discreta

- Conocer la distribución de probabilidad de una V.A. discreta X implica conocer para cada valor de X la correspondiente probabilidad P(x) asociada a dicho valor
- Se denomina **función de probabilidad** al procedimiento, fórmula o regla utilizado para obtener el valor de la probabilidad.

$$f(x) = P(X = x) = P(x)$$

$$\text{Se debe verificar que} \quad 0 \le P(x) \le 1 \ \text{ y } \ \sum P(x) = 1$$

- Se puede definir también la **función de distribución de probabilidad** F(x) como

$$F(\mathrm{x}) = \mathrm{P}(\mathrm{X} \le \mathrm{x}) = \sum_{\mathrm{mín}}^{\mathrm{x}} P(x)$$

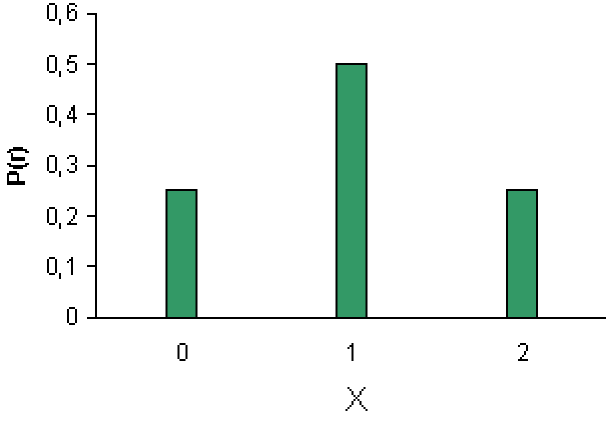
Ejemplo

- EA : tirar dos monedas equilibradas
- X = cantidad de caras observadas cuando se lanzan dos monedas equilibradas

P [de no observar caras]	=	P [XX]	=	P[X=0]	=	¼
P [de observar una cara]	=	P [CX o XC]	=	P[X=1]	=	$^2/_4$
P [de observar dos caras]	=	P [CC]	=	P[X=2]	=	¼

DISTRIBUCIÓN DE PROBABILIDADES

x	P(x)	F(x)
0	0,25	0,25
1	0,50	0,75
2	0,25	1,0
	1,0	



Valor esperado o esperanza de una v.a. discreta

- Se representa mediante $E(x)$ o μ
- Es el equivalente a la media
- ¿Cómo calculábamos el valor promedio?

$$\mu = \frac{\sum x_i FA_i}{N}$$

ahora: $$E(x) = \mu = \sum x_i p_i$$

Varianza de una v.a. discreta

- Se representa mediante *Var(x)* o $\sigma^2(x)$
- Resume la variabilidad en los valores de la V.A.
- ¿Cómo calculábamos la varianza?

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 FA_i}{N}$$

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

- Se llama desviación estándar o típica a σ

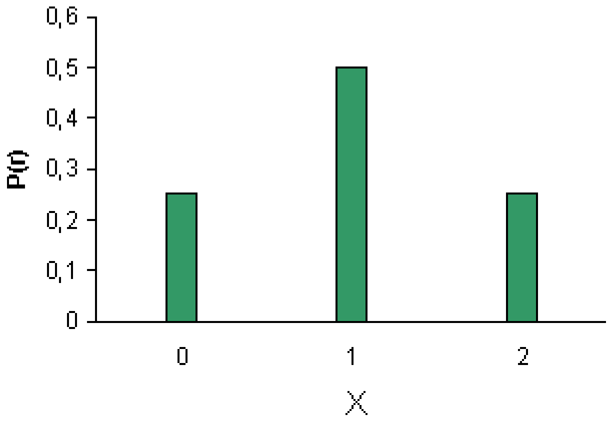
$$\sigma = \sqrt{\sigma^2}$$

Ejemplo

- X = cantidad de caras observadas cuando se lanzan dos monedas equilibradas

DISTRIBUCIÓN DE PROBABILIDADES

x	P(x)	F(x)
0	0,25	0,25
1	0,50	0,75
2	0,25	1,0
	1,0	



$$E(x) = \mu = \sum x_i p_i =$$

$$\sigma(x) = \sqrt{\sum (x_i - \mu)^2 p_i} =$$

¿Cómo averiguamos la distribución de probabilidades de una VA discreta?

- En forma empírica
- A través de modelos teóricos
  - Modelo o distribución Bernoulli
  - Modelo o distribución binomial
  - Modelo o distribución hipergeométrica
  - Modelo o distribución de Poisson

Distribución Bernoulli

- La probabilidad de que un recién nacido sea varón es: 0.51
- Supongamos que nos interesan las familias con **un** hijo. ¿Cuántos serán varón?

- X=
- Dominio de X =
- ¿Distribución de probabilidades de X?

Características de la distribución Bernoulli

- El experimento consiste en 1 ensayo
- Dicho ensayo tiene 2 resultados posibles (éxito o fracaso)
- La probabilidad de éxito del ensayo es p
- La probabilidad de fracaso es $q = 1-p$
- Estamos interesados en x = cantidad de éxitos en 1 ensayo

X es una variable aleatoria con distribución Bernoulli de parámetro p

Función de probabilidad

- La probabilidad de encontrar exactamente ***k*** éxitos en ***1*** ensayo se calcula como:

$$P(X = k) = p^k q^{1-k}$$

$$X \rightsquigarrow \mathbf{Ber}(p) \iff X = \begin{cases} 0 & \longrightarrow & P(X = 0) = p^0 q^{1-0} = q \\ 1 & \longrightarrow & P(X = 1) = p^1 q^{1-1} = p \end{cases}$$

- El dominio de la variable es

$$0 \leq x \leq 1$$

Distribución Binomial

- La probabilidad de que un recién nacido sea varón es: 0.51
- Supongamos que nos interesan las familias de 3 hijos. ¿Cómo será la distribución de sexos?

- X=
- Dominio de X =
- ¿Distribución de probabilidades de X?

Características de la distribución binomial

- El experimento consiste en n ensayos idénticos
- Cada ensayo tiene 2 resultados posibles (éxito o fracaso)
- La probabilidad de éxito de un ensayo simple es p y permanece constante de ensayo en ensayo
- La probabilidad de fracaso es $q = 1 - p$
- Los ensayos son independientes
- Estamos interesados en x = cantidad de éxitos en n ensayos

X es una variable aleatoria con distribución binomial de parámetros n y p

Función de probabilidad

- La probabilidad de encontrar exactamente ***k*** éxitos en ***n*** ensayos o repeticiones se calcula como:

$$P(X=k) = {}_nC_k\, p^k q^{n-k}$$

- Siendo ${}_nC_k$ la cantidad de combinaciones distintas de n elementos con k elementos iguales ("n elementos tomados de a k")

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

- El dominio de la variable es

$$0 \le x \le n$$

En el ejemplo

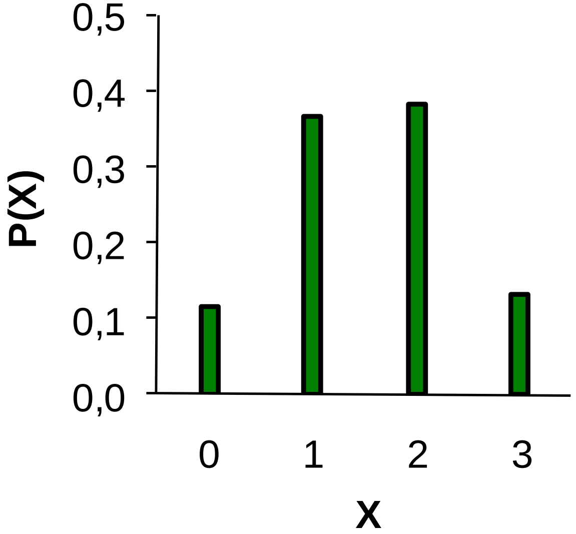
X=cantidad de varones en familias con 3 hijos

Parámetros: $n =$ $p =$

Dominio:

DISTRIBUCIÓN DE PROBABILIDADES

X	P(x)	F(x)
0	0,1176	0,1176
1	0,3674	0,4850
2	0,3823	0,8673
3	0,1327	1
	1	



Esperanza y varianza en la distribución binomial

El valor esperado de X puede calcularse como:

$$E(x) = \mu = \sum x_i p_i$$

Alternativamente puede calcularse como:

$$\boxed{E(x) = \mu = np}$$

De la misma manera, la varianza puede calcularse como:

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

O bien:

$$\boxed{\sigma^2(x) = npq}$$

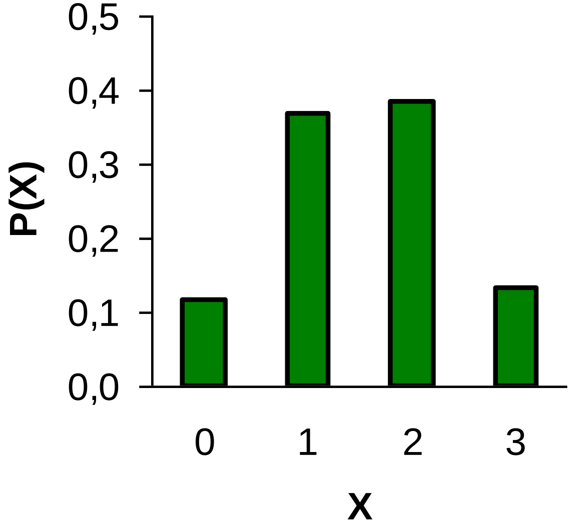
En el ejemplo

X=cantidad de varones en familias con 3 hijos

Parámetros: $n = 3$ $p = 0.51$

DISTRIBUCIÓN DE PROBABILIDADES

X	P(x)	F(x)
0	0,1176	0,1176
1	0,3674	0,4850
2	0,3823	0,8673
3	0,1327	1
	1	

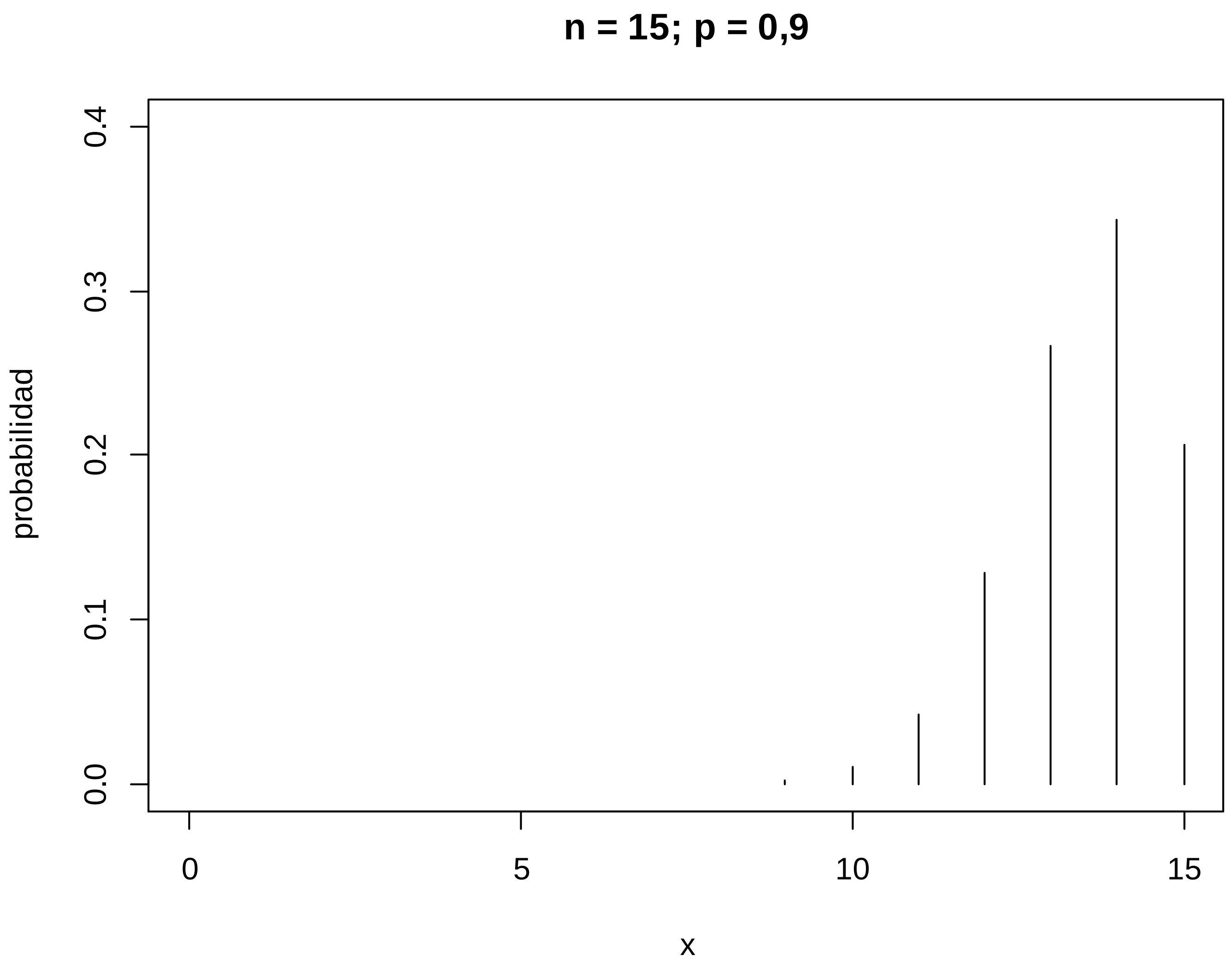


$$E(x) = \mu = np = 3x0.51 = 1.53\ varones$$

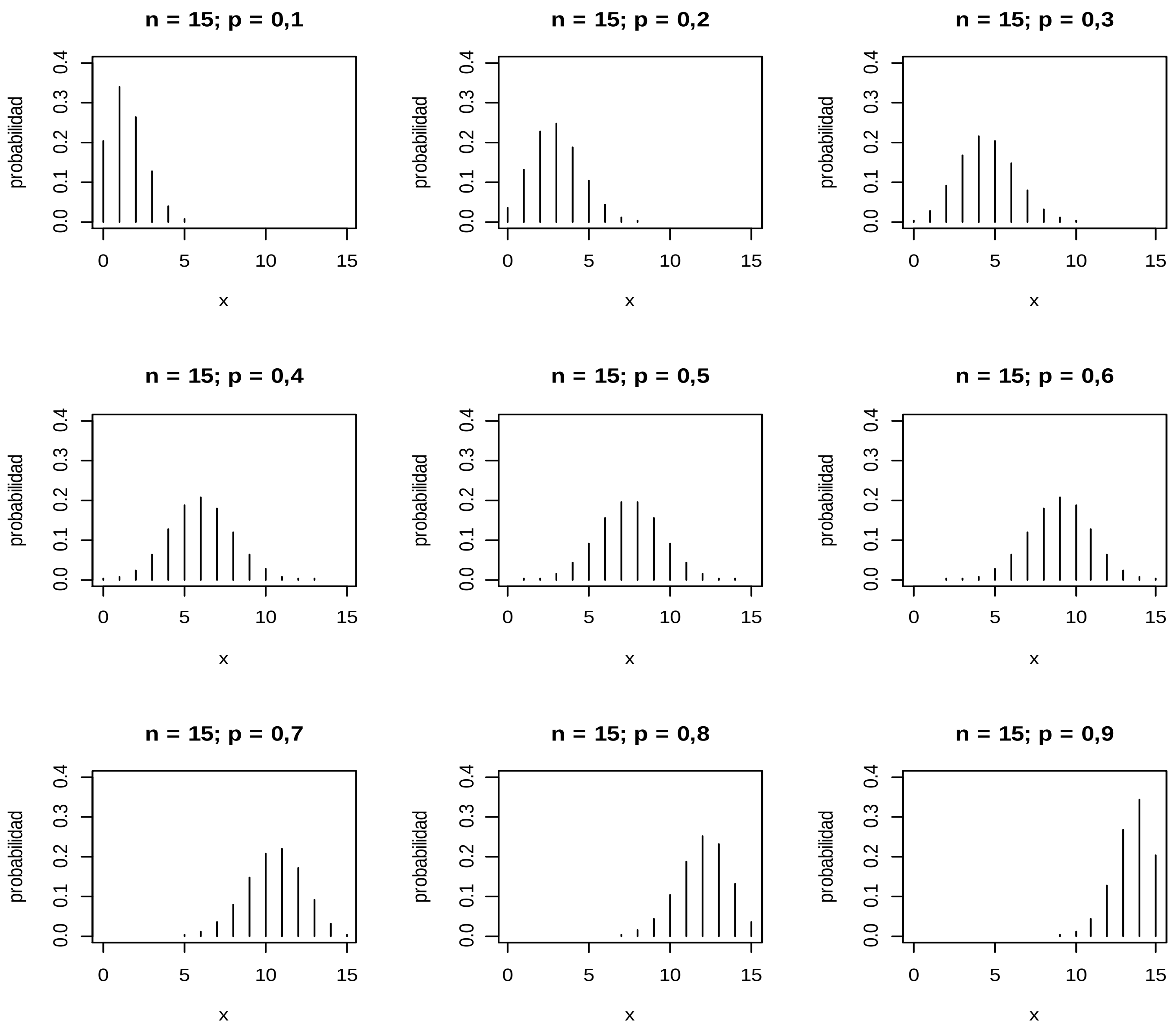
$$\sigma(x) = \sqrt{npq} = \sqrt{3x0.51x0.49} = 0.87\ varones$$

$$CV = \frac{\sigma}{\mu} = \frac{0.87}{1.53} = 0.57 = 57\%$$

Distribución Binomial con diferentes valores de p (n constante=15)



Distribución Binomial con diferentes valores de p (n constante=15)



¿Es binomial o no?

- En una pecera hay 5 ejemplares de carpa dorada (*Carassius auratus*). Tres de estos ejemplares están parasitados por el copépodo *Ergasilus sp*.
- Se sacan 2 ejemplares de la pecera (sin reposición)
- Sea x= cantidad de carpas parasitadas en la muestra de 2
- ¿Sigue x una distribución binomial?
- ¿Y si la extracción fuese con reposición?

¿Es binomial o no?

- En la práctica, el requisito de extracción con reposición rara vez se cumple
- Se sabe que el 4% de la población es portador del gen causante de la fenilcetonuria
- Se eligen 10 personas de la población
  - Para la primera persona, p = P(gen) = 0.04
  - Para la segunda persona, p ~ P(gen) = 0.04, aunque una persona ha sido removida de la población…
  - Para la décima persona, p ~ P(gen) = 0.04

<u>Regla</u>: Si ***n/N < 0.05***, equivale prácticamente a una extracción con reposición y por lo tanto el experimento se considera Binomial

Volviendo al ejemplo

- En una pecera hay 5 ejemplares de carpa dorada (*Carassius auratus*). Tres de estos ejemplares están parasitados por el copépodo *Ergasilus sp*.
- Se sacan 2 ejemplares de la pecera (sin reposición)
- Sea x= cantidad de carpas parasitadas en la muestra de 2
- ¿Cuál es la probabilidad de que solo uno esté parasitado?

Distribución hipergeométrica

- El experimento consiste en n ensayos
- Cada ensayo tiene 2 resultados posibles (éxito o fracaso)
- La probabilidad de éxito cambia de ensayo en ensayo; los ensayos son dependientes
- Se definen:
  - N = tamaño de la población
  - D = cantidad de éxitos en la población
  - n = tamaño de la muestra
  - X = cantidad de éxitos en la muestra

X es una variable aleatoria con distribución hipergeométrica de parámetros N, D y n

Función de probabilidad

- La probabilidad de encontrar exactamente k éxitos en n ensayos o repeticiones se calcula como:

$$P(x = k) = \frac{C_k^D C_{n-k}^{N-D}}{C_n^N}$$

- El dominio de la variable es

$$máx(0, n + D - N) \leq x \leq mín(n, D)$$

Esperanza y varianza en la distribución hipergeométrica

- La esperanza y la varianza de una variable hipergeométrica recuerdan a las de una variable con distribución binomial:
- **Valor esperado o esperanza**

$$E(x) = \mu = n\left(\frac{D}{N}\right)$$

- **Varianza**

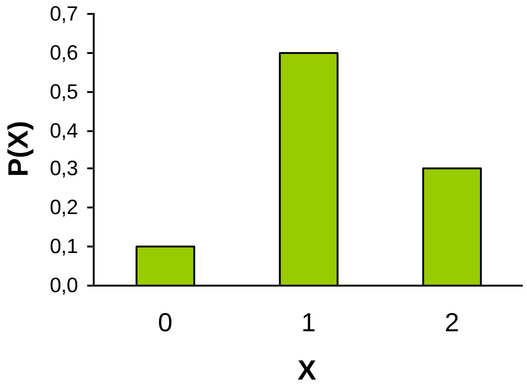
$$\sigma^2 = n\left(\frac{D}{N}\right)\left(\frac{N-D}{N}\right)\left(\frac{N-n}{N-1}\right)$$

En el ejemplo

X=cantidad de carpas parasitadas en una muestra de 2
Parámetros: $N = 5$ $D= 3$ $n= 2$
Dominio: [0-2]

DISTRIBUCIÓN DE PROBABILIDADES

x	P(x)	F(x)
0	0,1	0,1
1	0,6	0,7
2	0,3	1,0
	1,0	



$$E(x) = \mu = n\left(\frac{D}{N}\right) = 2x\frac{3}{5} = 1.2\, carpas\ parasitadas$$

$$\sigma = \sqrt{n\left(\frac{D}{N}\right)\left(\frac{N-D}{N}\right)\left(\frac{N-n}{N-1}\right)} = 0.6\, carpas\ parasitadas$$

Distribución Poisson

- Describe la distribución de la variable aleatoria

  x = cantidad de acontecimientos o eventos puntuales que se presentan en un continuo
- Los eventos son independientes entre sí
- Los eventos se producen al azar dentro del continuo
- Se definen
  - λ = cantidad esperada de eventos en el continuo bajo análisis

X es una variable aleatoria con distribución Poisson de parámetro λ

Ejemplos

- número de insectos por parcela en un determinado cultivo
- número de partículas emitidas por algún elemento radiactivo en un período de tiempo
- cantidad de accidentes por hora
- la superficie, el volumen, el tiempo, etc. son los espacios continuos en los que está registrada la cantidad de eventos

Función de probabilidad

- La probabilidad de encontrar exactamente k eventos en el continuo se calcula como:

$$P(x=k)=\frac{e^{-\lambda}(\lambda)^{k}}{k!}$$

- El dominio de la variable es

$$x \geq 0$$

- El valor esperado es λ
- La varianza es λ

Ejemplo

- Se estima que cada 100 años, en promedio, cae a la tierra un meteorito de entre 10 y 50 m de diámetro
- ¿Cuál es la probabilidad de que en los próximos 50 años se produzca sólo un impacto de estas características?
- Sea x=cantidad de impactos en 50 años

un impacto / 100 años= 0,01 impactos/año
50 años=0,5 impactos

$$\mu = \lambda = 0{,}5_{\text{impactos}}$$

$$P(x=1) = \frac{e^{-0{,}5}(0{,}5)^1}{1!} = 0{,}3033$$

En el ejemplo

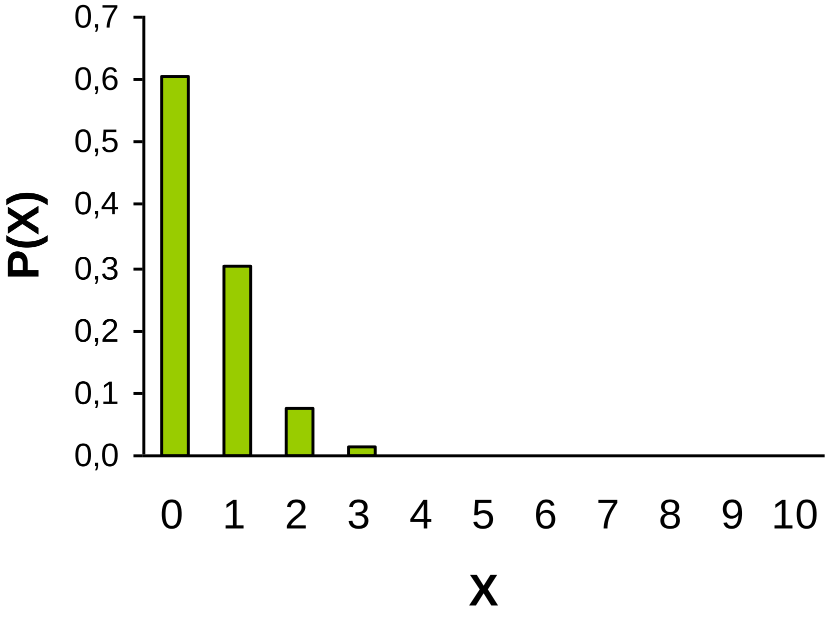
x=cantidad de impactos en 50 años

Parámetros: $\mu = \lambda = 0{,}5$ impactos

Dominio:

DISTRIBUCIÓN DE PROBABILIDADES

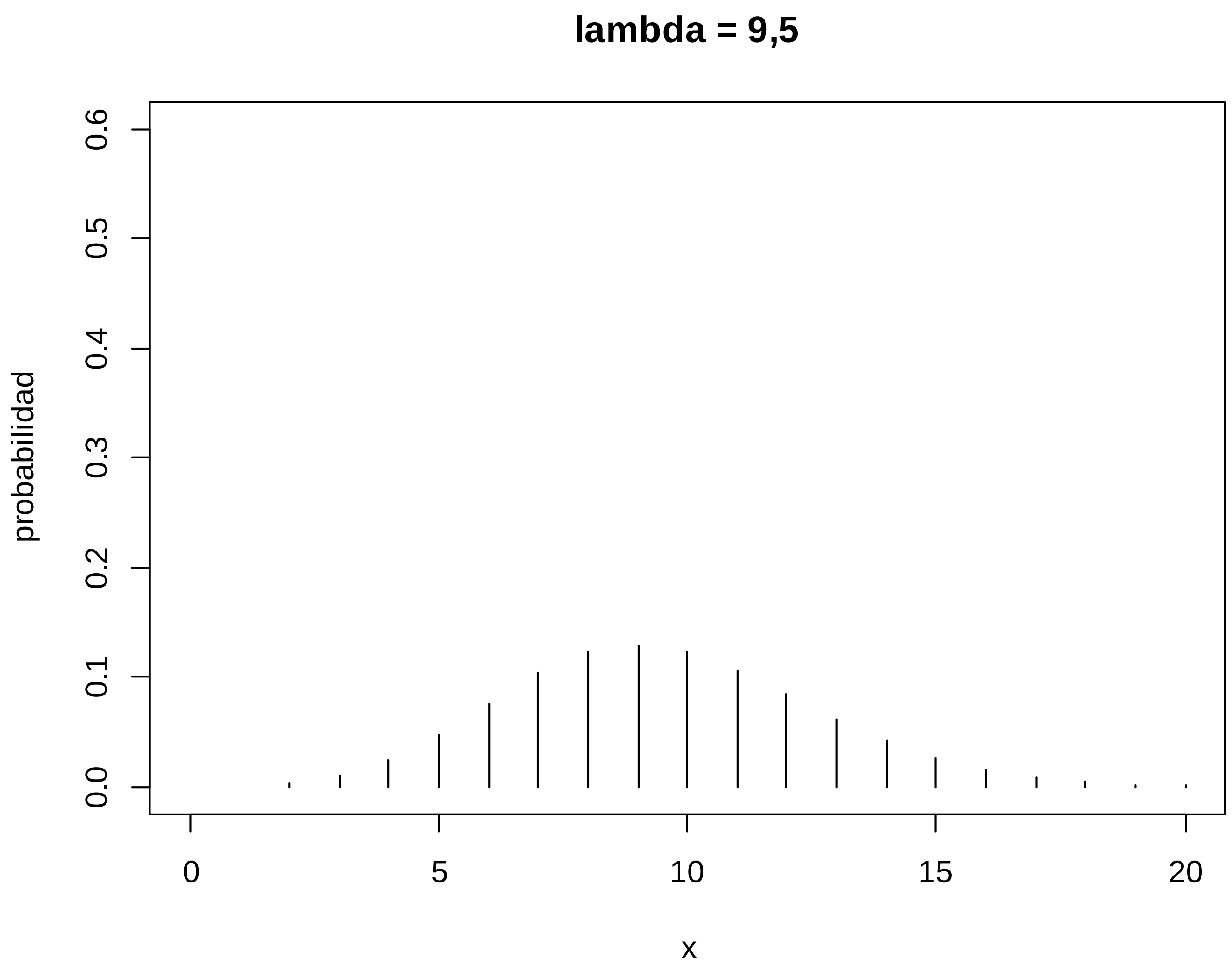
x	P(x)	F(x)
0	0,60653	0,60653
1	0,30327	0,90980
2	0,07582	0,98561
3	0,01264	0,99825
4	0,00158	0,99983
5	0,00016	0,99999
6	0,00001	0,99999
7	0,00000	0,99999
8	0,00000	0,99999
9	0,00000	0,99999
10	0,00000	0,99999
...	...	0,99999
	1,0	



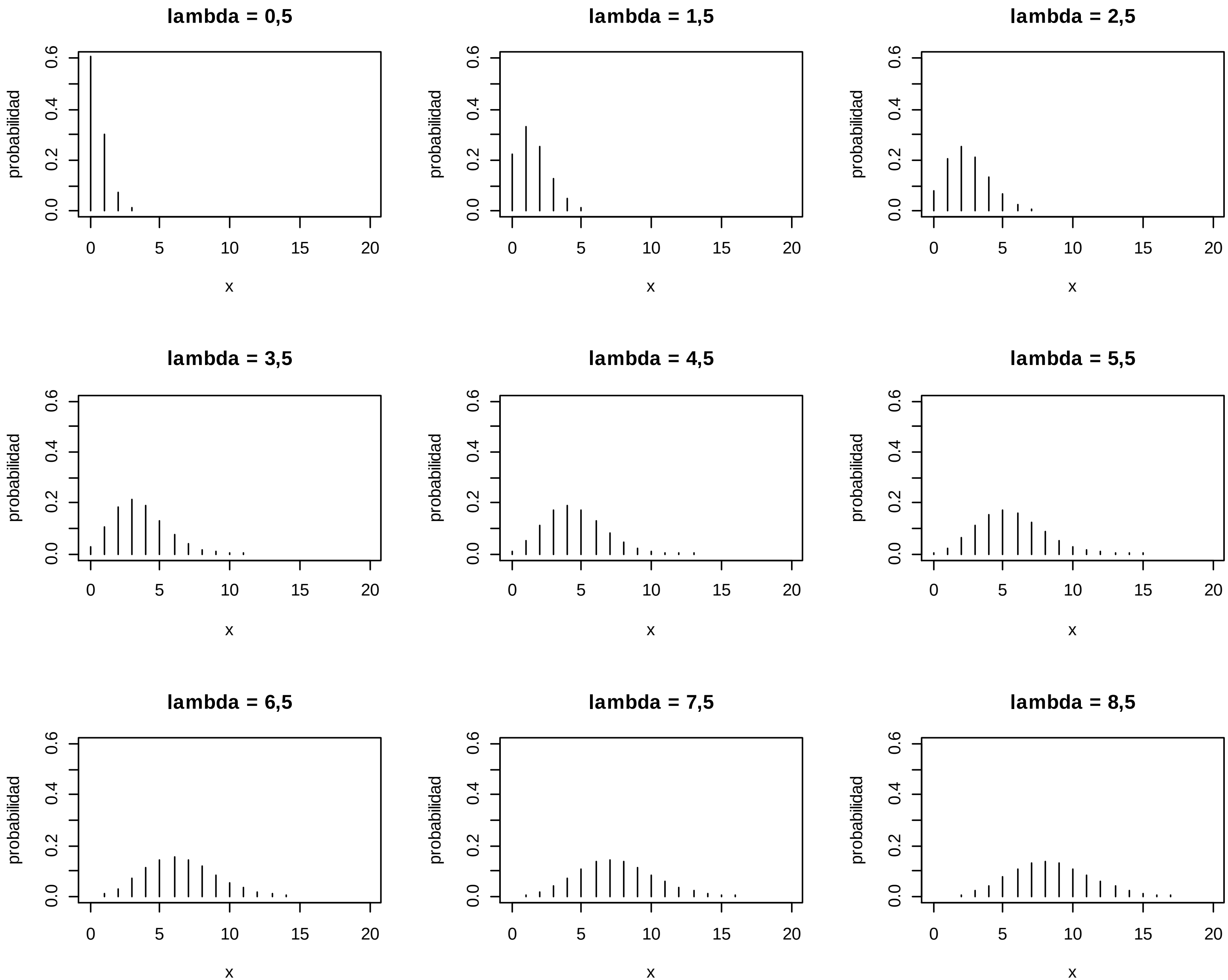
$$E(x) = \mu = 0.5 \text{ impactos}$$

$$\sigma(\mathrm{x}) = \sqrt{\lambda} = \sqrt{0.5} = 0.71\, impactos$$

Distribución Poisson con diferentes valores de lambda



Distribución Poisson con diferentes valores de lambda



Otro ejemplo

Cierta enfermedad tiene una probabilidad muy baja de ocurrir, *p=1/100.000.*
Calcular la probabilidad de que en una ciudad con 500.000 habitantes haya más de 3 personas con dicha enfermedad. Calcular el número esperado de habitantes que la padecen.

X: $\leq x \leq$

$$X \sim Bi\ (500.000, 1/100.000)$$

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

P(x>3)= P(x=4)+P(x=5)+…………………+ P(x=500.000)

P(x>3)= 1-P(x≤3)= 1-(P(x=0)+P(x=1)+P(x=2)+P(x=3))

P(x=0)=

P(x=1)=

P(x=2)=

$$P(x=3)=\binom{500.000}{3}\left(\frac{1}{100.000}\right)^{3}\left(\frac{99.999}{100.000}\right)^{499.997}$$

P(x=3)=

Aproximación Binomial a Poisson

Sea X una variable con distribución **Binomial** con parámetros n y p, donde *n* es muy alto y *p* es muy bajo tal que $nxp<6$

$$X \sim Bi\,(n, p)$$

n= ↑↑
p= ↓↓
$n_x p<6$

La distribución de probabilidades de X se aproximará a la de una variable con distribución **Poisson**

$$\boxed{X \sim Bi\,(n, p) \approx X \sim P(\lambda)}$$

$$E(x) = n_x p \qquad E(x) = (\lambda)$$

$$n_x p = \lambda$$

En el ejemplo

Cierta enfermedad tiene una probabilidad muy baja de ocurrir, *p=1/100.000.*
Calcular la probabilidad de que en una ciudad con 500.000 habitantes haya más de 3 personas con dicha enfermedad. Calcular el número esperado de habitantes que la padecen.

$$X \sim Bi\ (500.000, 1/100.000) \approx X \sim P(5)$$

$$E(x) = 500.000 \text{ x } 1/100.000 = 5$$