
5. LAS DISTRIBUCIONES MÁS IMPORTANTES



Gauss y la distribución normal

Objetivos

Introducir las distribuciones más importantes y sus usos.

Para leer

Podéis ver los *mini-videos* de Emilio Letón y su equipo sobre Modelos de Probabilidad.

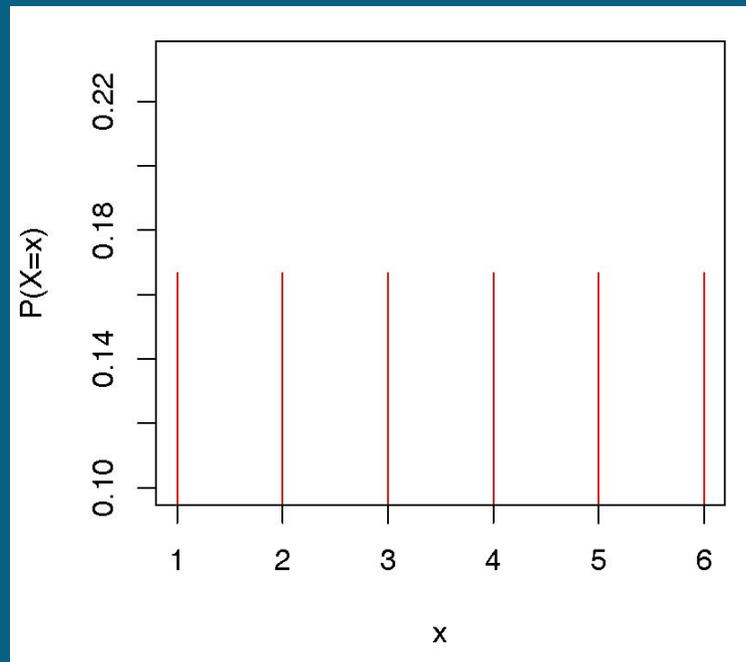
Distribuciones uniformes

Sea X el resultado de una tirada de un dado equilibrado. ¿Cuál es la distribución de X ? Hallar la media y varianza de X .

Distribuciones uniformes

Sea X el resultado de una tirada de un dado equilibrado. ¿Cuál es la distribución de X ? Hallar la media y varianza de X .

$$\text{Tenemos } P(X = x) = \begin{cases} \frac{1}{6} & \text{para } x = 1, 2, \dots, 6 \\ 0 & \text{en otro caso} \end{cases}$$



Además es fácil calcular la media y varianza:

$$\begin{aligned} E[X] &= \sum_{x=1}^6 \frac{1}{6} \times x = 3.5 \\ V[X] &= E[X^2] - E[X]^2 \\ &= \sum_{x=1}^6 \frac{1}{6} \times x^2 - 3.5^2 = 1.25 \end{aligned}$$

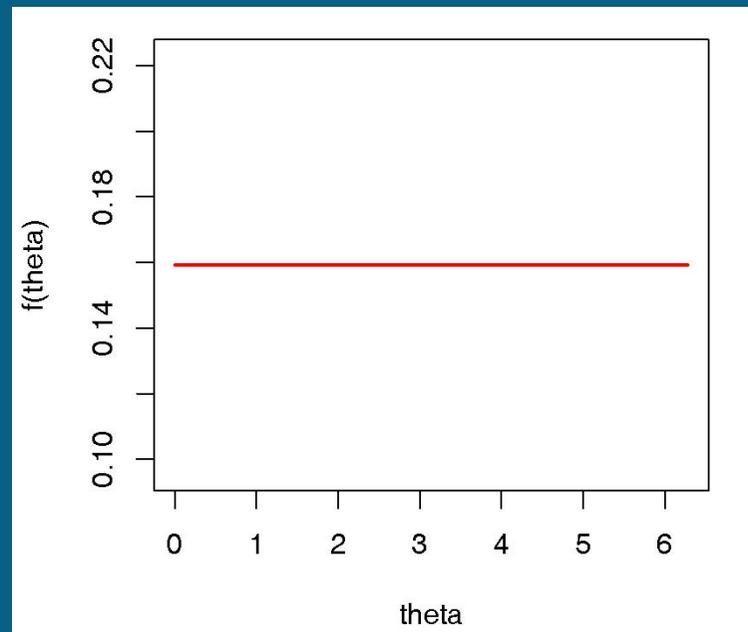
Esto es un ejemplo de una variable con una distribución uniforme discreta.

Giramos una rueda de fortuna y medimos el ángulo, θ , al vertical en radianes. Hallar la distribución de θ , su media y varianza.

Giramos una rueda de fortuna y medimos el ángulo, θ , al vertical en radianes. Hallar la distribución de θ , su media y varianza.

Suponiendo que la rueda es insesgada, la densidad debe ser la misma para cualquier ángulo $0 \leq \theta < 2\pi$. Luego,

$$f(\theta) = \begin{cases} \frac{1}{2\pi} & \text{para } 0 \leq \theta < 2\pi \\ 0 & \text{en caso contrario} \end{cases}$$



Entonces, la media es

$$E[\theta] = \int_0^{2\pi} \theta \times \frac{1}{2\pi} d\theta = \frac{1}{2\pi} \left[\frac{\theta^2}{2} \right]_0^{2\pi} = \pi$$

y la varianza es

$$\begin{aligned} V[\theta] &= E[\theta^2] - E[\theta]^2 \\ E[\theta^2] &= \int_0^{2\pi} \frac{\theta^2}{2\pi} d\theta \\ &= \frac{1}{2\pi} \left[\frac{\theta^3}{3} \right]_0^{2\pi} = \frac{4\pi^2}{3} \\ V[\theta] &= \frac{4\pi^2}{3} - \pi^2 = \frac{\pi^2}{3} \end{aligned}$$

Es un ejemplo de una distribución uniforme continua.

La distribución uniforme discreta

Una variable, X , tiene una distribución *uniforme discreta* con *parámetros* a y $b = a + n - 1$ donde $n \in \mathbb{N}$ si

$$P(X = x) = \begin{cases} \frac{1}{b-a+1} = \frac{1}{n} & \text{para } x = a, a+1, \dots, b-1, b \\ 0 & \text{en caso contrario} \end{cases}$$

En este caso, se escribe $X \sim \mathcal{UD}[a, b]$.

Teorema 17

Si $X \sim \mathcal{DU}[a, b]$ entonces

$$E[X] = \frac{a + b}{2}$$

$$V[X] = \frac{(b - a + 1)^2 - 1}{12} = \frac{n^2 - 1}{12}$$

Para demostrar el teorema, necesitamos dos resultados de matemáticas:

Lema 2

Para $n \in \mathbb{N}$, se tiene

$$\sum_{j=0}^n j = \frac{n(n+1)}{2}$$
$$\sum_{j=0}^n j^2 = \frac{n(n+1)(2n+1)}{6}$$

Demostración Matemáticas



Demostración Observamos que $E[X] = a + E[X - a]$ y que $V[X] = V[X - a]$ y consideramos la variable $Y = X - a$. Obviamente,

$$P(Y = y) = P(X = y + a) = \frac{1}{b - a + 1} \quad \text{para } y = 0, 1, 2, \dots, b - a$$

y luego, $Y \sim \mathcal{UD}[0, b - a]$. Ahora,

$$\begin{aligned} E[Y] &= \sum_{y=0}^{b-a} y \times \frac{1}{b - a + 1} \\ &= \frac{1}{b - a + 1} \sum_{y=0}^{b-a} y \\ &= \frac{1}{b - a + 1} \times \frac{(b - a)(b - a + 1)}{2} = \frac{b - a}{2} \end{aligned}$$

$$\begin{aligned}
E[Y^2] &= \sum_{y=0}^{b-a} y^2 \times \frac{1}{b-a+1} = \frac{1}{b-a+1} \sum_{y=0}^{b-a} y^2 \\
&= \frac{1}{b-a+1} \times \frac{(b-a)(b-a+1)(2(b-a)+1)}{6} \\
&= \frac{(b-a)(2(b-a)+1)}{6} \\
V[Y] &= \frac{(b-a)(2(b-a)+1)}{6} - \frac{(b-a)^2}{4} \\
&= \frac{2(b-a)(2(b-a)+1) - 3(b-a)^2}{12} \\
&= \frac{(b-a)(b-a+1)}{12} = \frac{n^2 - 1}{12}
\end{aligned}$$



La distribución uniforme continua

Una variable X tiene una *distribución uniforme continua* con parámetros $a, b \in \mathbb{R}$ si

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{en caso contrario} \end{cases}$$

En este caso, se escribe $X \sim \mathcal{U}(a, b)$.

Teorema 18

Si $X \sim \mathcal{U}(a, b)$ entonces

$$E[X] = \frac{a + b}{2}$$

$$V[X] = \frac{(b - a)^2}{12}$$

Demostración

$$E[X] = \int_a^b x \times \frac{1}{b-a} dx$$

$$= \left[\frac{x^2}{2(b-a)} \right]_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$E[X^2] = \int_a^b x^2 \times \frac{1}{b-a} dx = \left[\frac{x^3}{3(b-a)} \right]_a^b$$

$$= \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

$$V[X] = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}$$



Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada.

Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada. Ensayo de Bernoulli.
2. El número de cruces en n tiradas.

Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada. Ensayo de Bernoulli.
2. El número de cruces en n tiradas. Distribución binomial.
3. El número de caras antes de ver la primera cruz.

Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada. Ensayo de Bernoulli.
2. El número de cruces en n tiradas. Distribución binomial.
3. El número de caras antes de ver la primera cruz. Distribución geométrica (tipo I).
4. El número de lanzamientos hasta ver la primera cruz.

Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada. Ensayo de Bernoulli.
2. El número de cruces en n tiradas. Distribución binomial.
3. El número de caras antes de ver la primera cruz. Distribución geométrica (tipo I).
4. El número de lanzamientos hasta ver la primera cruz. Distribución geométrica (tipo II).
5. El número de caras antes de ver r cruces y el número de lanzamientos necesarios para ver r cruces.

Distribuciones asociadas con lanzamientos de monedas

Supongamos que vamos a lanzar una moneda con $P(\text{cruz}) = p$ de manera independiente. Luego hay varias variables de interés:

1. El número de cruces en 1 tirada. Ensayo de Bernoulli.
2. El número de cruces en n tiradas. Distribución binomial.
3. El número de caras antes de ver la primera cruz. Distribución geométrica (tipo I).
4. El número de lanzamientos hasta ver la primera cruz. Distribución geométrica (tipo II).
5. El número de caras antes de ver r cruces y el número de lanzamientos necesarios para ver r cruces. Distribuciones binomiales negativas.

Ensayos de Bernoulli

X tiene una *distribución de Bernoulli* con parámetro p si

$$P(X = x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \\ 0 & \text{si no.} \end{cases}$$

Teorema 19

$$\begin{aligned} E[X] &= p \\ V[X] &= 1 - p \end{aligned}$$

Demostración Demostrado anteriormente.



La distribución binomial

X tiene una *distribución binomial* con parámetros n y p si

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, \dots, n \\ 0 & \text{si no.} \end{cases}$$

En este caso, se escribe $X \sim \mathcal{BI}(n, p)$.

Teorema 20

$$E[X] = np$$

$$V[X] = np(1-p)$$

Demostración Demostrado anteriormente.



La distribución binomial como suma de ensayos de Bernoulli

Obviamente, si $X \sim \mathcal{BI}(n, p)$, se tiene $X = X_1 + X_2 + \cdots + X_n$ donde los X_i son ensayos de Bernoulli *independientes y idénticamente distribuidos*.

Observamos que:

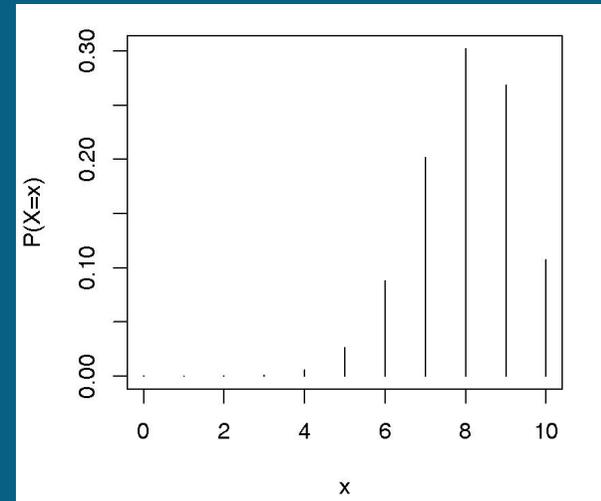
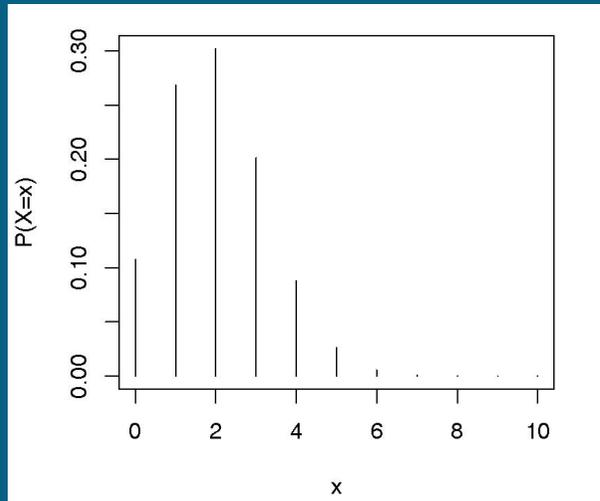
$$E[X] = np = \sum_{i=1}^n E[X_i]$$

$$V[X] = np(1 - p) = \sum_{i=1}^n p(1 - p) = \sum_{i=1}^n V[X_i]$$

$$G_X(s) = (sp + 1 - p)^n = \prod_{i=1}^n G_{X_i}(s).$$

La distribución de $Y = n - X$

Los dos gráficos muestran la función de probabilidad de $X \sim \mathcal{BI}(10, 0.2)$ e $Y \sim \mathcal{BI}(10, 0.8)$ respectivamente.



Se ve que $P(X = x) = P(Y = y)$.

Supongamos ahora que se lanza una moneda con $p = P(\text{cruz})$ n veces.

$$P(x \text{ cruces}) = P(n - x \text{ caras})$$

y entonces si $X \sim \mathcal{BI}(n, p)$, luego $Y = n - X \sim \mathcal{BI}(n, 1 - p)$.

Ejemplos

La probabilidad de que un paciente se recupere de una extraña enfermedad es $.4$. Si se sabe que 15 personas contraen esa enfermedad:

- a. ¿Cuál es la probabilidad de que sobrevivan al menos 10?*
- b. ¿Cuál es la probabilidad de que sobrevivan de 3 a 8?*
- c. Calcule la media y la varianza de esta distribución binomial.*
- d. ¿Es razonable el supuesto de que la distribución sea binomial en este caso? Comentar.*

En una ciudad la necesidad de dinero para comprar drogas se establece como la razón del 75% de los robos. Encuentre la probabilidad de que entre los siguientes cinco casos de robo:

- a. dos resulten de la necesidad de dinero para comprar drogas.*
- b. al menos tres resulten de la necesidad de dinero para comprar drogas.*
- c. Calcule la media y la varianza de esta distribución binomial.*

La distribución geométrica (tipo I)

X tiene una *distribución geométrica* con parámetro p si

$$P(X = x) = \begin{cases} p(1 - p)^x & \text{si } x = 0, 1, \dots, n \\ 0 & \text{si no.} \end{cases}$$

En este caso, escribimos $X \sim \mathcal{GE}(p)$.

Teorema 21

$$E[X] = \frac{1 - p}{p}$$

$$V[X] = \frac{1 - p}{p^2}$$

Demostración Utilizamos la función generadora de probabilidades

$$G(s) = E[s^X] = \sum_{x=0}^{\infty} s^x p(1-p)^x = p \sum_{x=0}^{\infty} [s(1-p)]^x = \frac{p}{1-s(1-p)}$$

$$G'(s) = \frac{p(1-p)}{[1-s(1-p)]^2} \Rightarrow$$

$$E[X] = \frac{p(1-p)}{[1-1 \times (1-p)]^2} = \frac{1-p}{p}$$

$$G''(s) = \frac{2p(1-p)^2}{[1-s(1-p)]^3} \Rightarrow$$

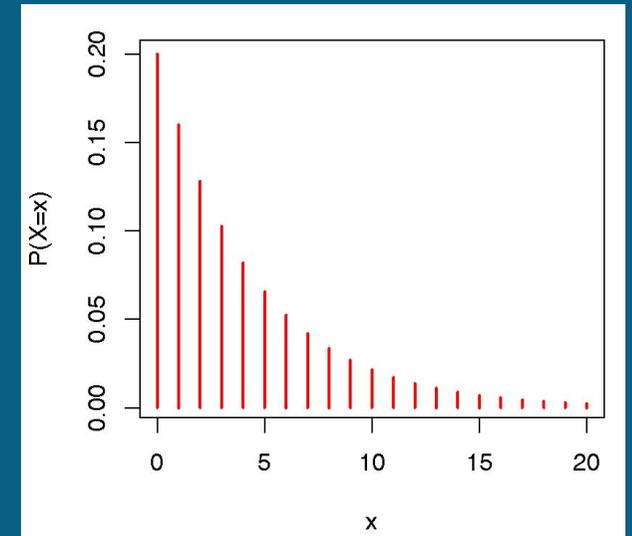
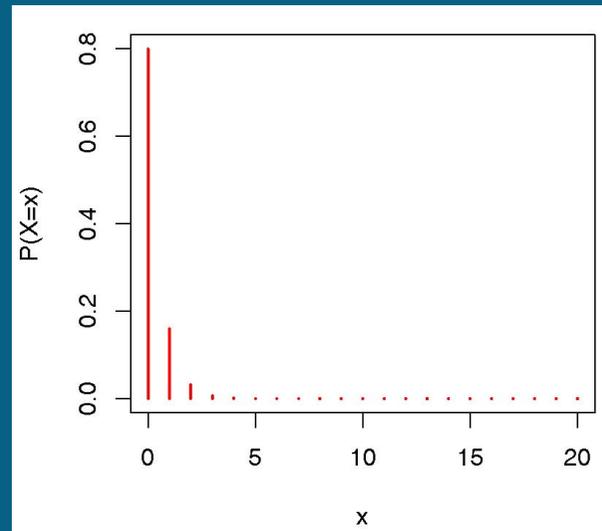
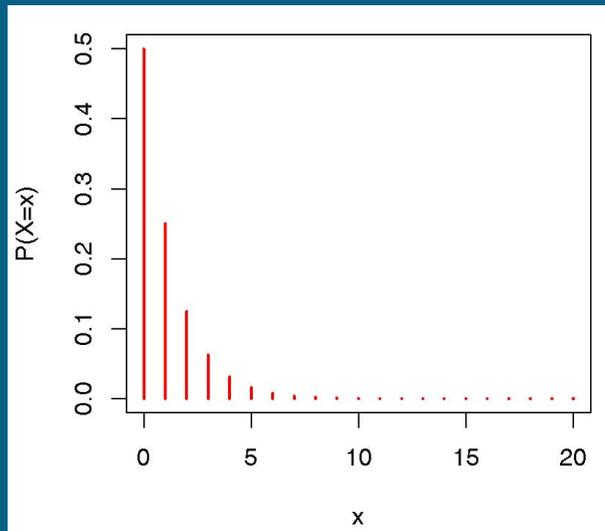
$$E[X(X-1)] = 2 \left(\frac{1-p}{p} \right)^2$$

$$V[X] = 2 \left(\frac{1-p}{p} \right)^2 + \frac{1-p}{p} - \left(\frac{1-p}{p} \right)^2 = \frac{1-p}{p^2}.$$



Distribuciones de probabilidad

Los gráficos muestran las funciones de probabilidad para la distribución geométrica con $p = 0.5$, 0.8 y 0.2 respectivamente.



La distribución geométrica es muy asimétrica y unimodal, con moda 0. Si p es más alto, decae más rápidamente.

La distribución geométrica (tipo II)

Sea $Y = X + 1$ donde $X \sim \mathcal{GE}(p)$. Luego

$$P(Y = y) = P(X = y - 1) = \begin{cases} p(1 - p)^{y-1} & \text{para } y = 1, 2, \dots \\ 0 & \text{en caso contrario} \end{cases}$$

Teorema 22

$$E[Y] = \frac{1}{p}$$
$$V[Y] = \frac{1 - p}{p^2}$$

Demostración Como $Y = X + 1$, luego, $V[Y] = V[X]$ y

$$E[Y] = 1 + E[X] = 1 + \frac{1 - p}{p} = \frac{1}{p}. \quad \square$$

Ejemplos

Los registros de una compañía constructora de pozos, indican que la probabilidad de que uno de sus pozos nuevos, requiera de reparaciones en el término de un año es de 0.20. ¿Cuál es la probabilidad de que el quinto pozo construido por esta compañía en un año dado sea el primero en requerir reparaciones en un año?

Ejemplos

Los registros de una compañía constructora de pozos, indican que la probabilidad de que uno de sus pozos nuevos, requiera de reparaciones en el término de un año es de 0.20. ¿Cuál es la probabilidad de que el quinto pozo construido por esta compañía en un año dado sea el primero en requerir reparaciones en un año?

Sí la probabilidad de que un cierto dispositivo de medición muestre una desviación excesiva es de 0.05, ¿cuál es la probabilidad de que; a) el sexto de estos dispositivos de medición sometidos a prueba sea el primero en mostrar una desviación excesiva?, b) el séptimo de estos dispositivos de medición sometidos a prueba, sea el primero que no muestre una desviación excesiva?

La distribución binomial negativa

Se dice que X tiene una distribución binomial negativa si

$$P(X = x) \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{si no.} \end{cases}$$

En este caso, se escribe $X \sim \mathcal{BN}(r, p)$.

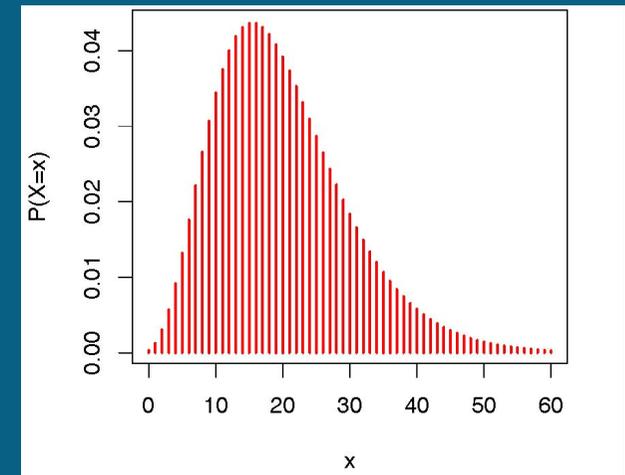
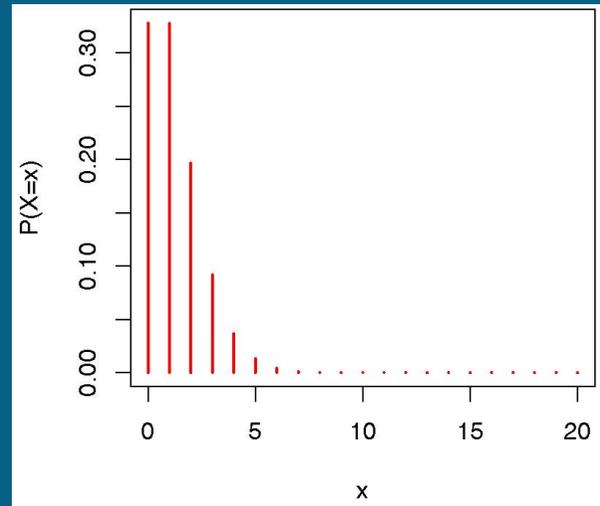
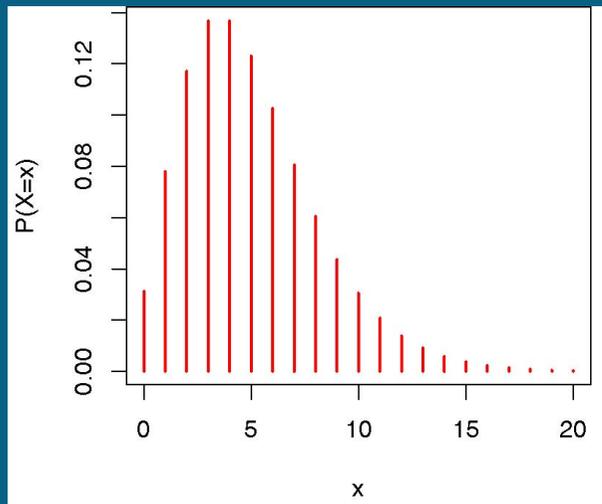
Teorema 23

$$E[X] = r \frac{1-p}{p}$$
$$V[X] = r \frac{1-p}{p^2}$$

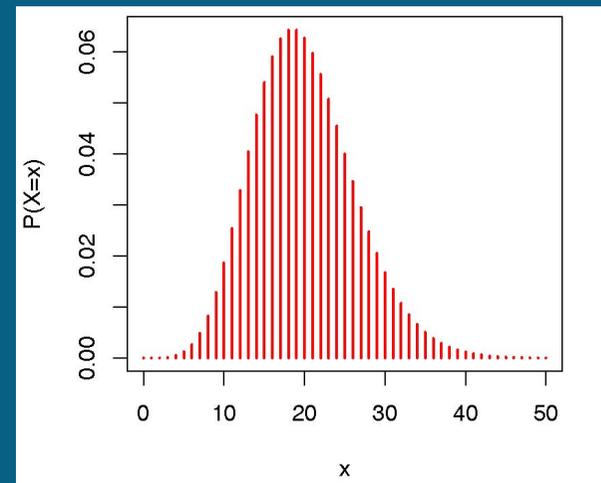
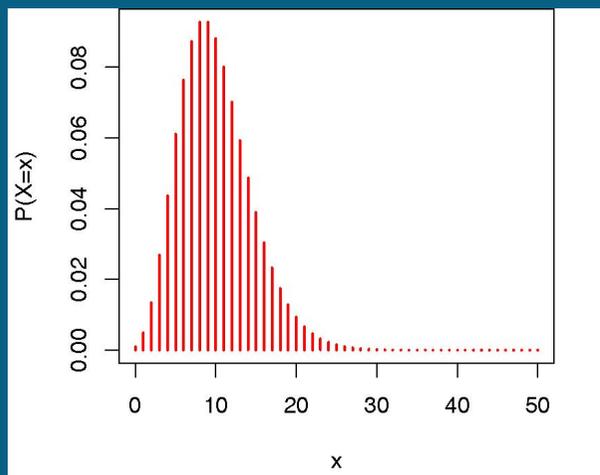
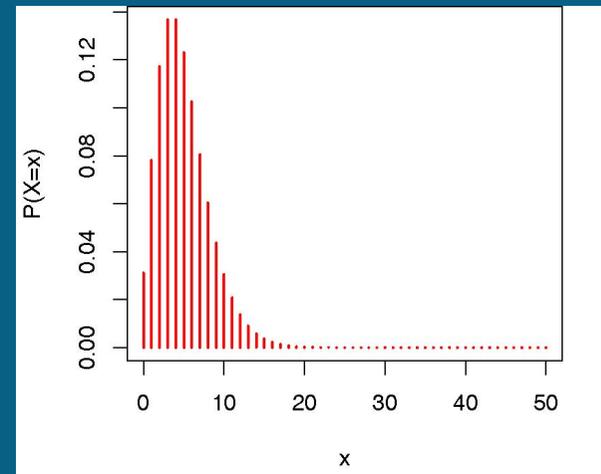
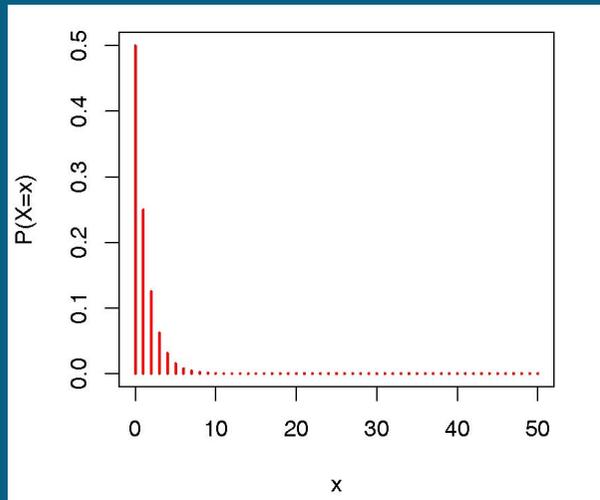
Demostración Ejercicio. Usar la función generadora. □

Distribuciones binomiales negativas

Los gráficos muestran la distribución binomial negativa con $r = 5$ y $p = 0.5, 0.8$ y 0.2 respectivamente.



Los siguientes gráficos muestran la distribución binomial negativa con $p = 0.5$ y $r = 1, 5, 10$ y 20 respectivamente.



La función pierde asimetría como crezca r . El último gráfico tiene forma de campana.

¿Cómo calculamos la probabilidad?

Pensamos en una secuencia de tiradas donde la última es una cruz y hay r cruces y x caras en total, por ejemplo

$$\{\underbrace{CCCC}_{x}\underbrace{XXXX}_{r}\} \quad \{\underbrace{XXXX}_{r-1}\underbrace{CCCC}_{x}X\}$$

La probabilidad de cada una de estas secuencias es $(1-p)^x p^r$ pero ¿cuántas secuencias hay?

La última tirada debe ser cruz y tenemos que arreglar las demás en x caras y $r-1$ cruces $\Rightarrow C_{x+r-1}^x$ secuencias.

Luego, la probabilidad de observar x caras antes de ver r cruces es

$$\text{número de secuencias} \times (1-p)^x p^r = C_{x+r-1}^x (1-p)^x p^r.$$

La distribución binomial negativa como suma de geométricas

Si $X \sim \mathcal{BN}(r, p)$ entonces X representa el número de caras observadas antes de haber visto r cruces. Luego $X = X_1 + X_2 + \cdots + X_r$ donde X_i representa el número de caras observadas entre los cruces $i-1$ y i . Entonces, $X_i \sim \mathcal{GE}(p)$ son variables geométricas (tipo I) independientes.

Observamos que:

$$\begin{aligned} E[X] &= rE[X_i] \\ V[X] &= rV[X_i] \\ G_X(s) &= G_{X_i}(s)^r \\ &= \left(\frac{p}{1 - s(1 - p)} \right)^r \end{aligned}$$

Otra distribución binomial negativa

Sea $X \sim \mathcal{BN}(r, p)$ el número de caras antes de ver el r -ésimo cruz y sea Y el número de tiradas hasta ver el r -ésimo cruz. Entonces,

$$P(Y = y) = P(X = y-r) = \begin{cases} \binom{y-1}{y-r} p^r (1-p)^{y-r} & \text{si } y = r, r+1, r+2, \dots \\ 0 & \text{en caso contrario} \end{cases}$$

Ejemplos

Sí la probabilidad de que un cierto dispositivo de medición muestre una desviación excesiva es de 0.05, ¿cuál es la probabilidad de que; a) el sexto de estos dispositivos de medición sometidos a prueba sea el tercero en mostrar una desviación excesiva?, b) el séptimo de estos dispositivos de medición sometidos a prueba, sea el cuarto que no muestre una desviación excesiva?.

Ejemplos

Sí la probabilidad de que un cierto dispositivo de medición muestre una desviación excesiva es de 0.05, ¿cuál es la probabilidad de que; a) el sexto de estos dispositivos de medición sometidos a prueba sea el tercero en mostrar una desviación excesiva?, b) el séptimo de estos dispositivos de medición sometidos a prueba, sea el cuarto que no muestre una desviación excesiva?.

Los registros de una compañía constructora de pozos, indican que la probabilidad de que uno de sus pozos nuevos, requiera de reparaciones en el término de un año es de 0.20. a) ¿Cuál es la probabilidad de que el sexto pozo construido por esta compañía en un año dado sea el segundo en requerir reparaciones en un año?. b) ¿Cuál es la probabilidad de que el octavo pozo construido por esta compañía en un año dado sea el tercero en requerir reparaciones en un año?.

Jugando a cartas: la distribución hipergeométrica

Supongamos que tenemos una baraja de N cartas, R de ellos siendo rojas y $N - R$ azules. Si se reparten n cartas, entonces el número de cartas rojas repartidas, X sigue una distribución hipergeométrica con parámetros N , R y n

$$P(X = x) = \frac{\binom{R}{x} \binom{N - R}{n - x}}{\binom{N}{n}} \quad \text{para } x = 0, 1, \dots, n.$$

En este caso, se escribe $X \sim \mathcal{H}(N, R, n)$.

Teorema 24

Si $X \sim \mathcal{H}(N, R, n)$ entonces

$$E[X] = \frac{nR}{N}$$

$$V[X] = n \frac{R}{N} \left(1 - \frac{R}{N}\right) \frac{N - n}{N - 1}$$

Demostración Demasiado complicado.

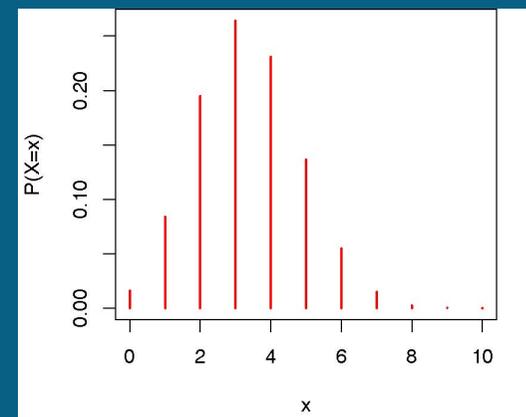
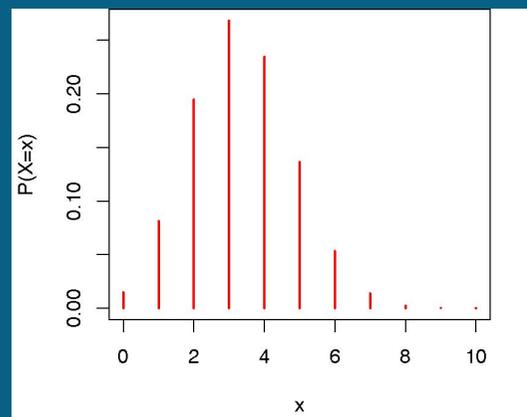
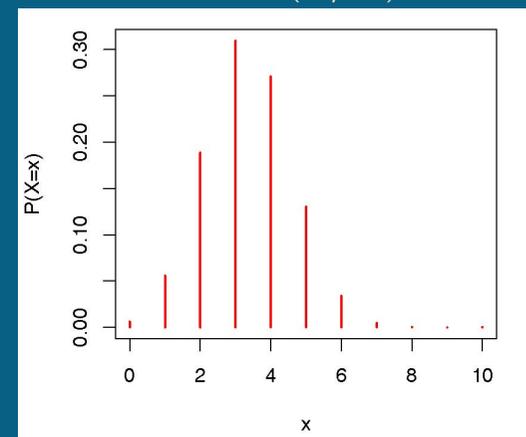
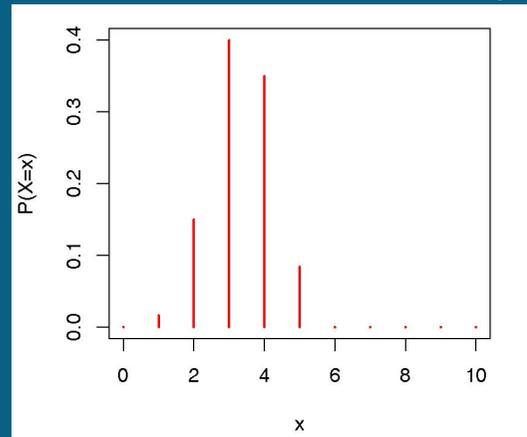


Ejemplo

Entre las 20 celdas solares que se presentan en una expresión comercial, 12 son celdas planas y las otras son celdas de concentración. Si una persona que visita la exposición selecciona al azar 6 de las salas solares para revisarlas. ¿Cuál es la probabilidad de que 3 de estas sean planas? Calcular la esperanza y la varianza del número de celdas planas en 6 celdas seleccionadas.

La función de probabilidad

Los gráficos muestran la función de probabilidad hipergeométrica con $n = 10$ y $(N, R) = (15, 5), (30, 10), (150, 50)$ y $(300, 100)$ respectivamente. Observamos que la proporción inicial de bolas rojas es la misma ($1/3$) en cada caso.



Aproximación de probabilidades hipergeométricas

En los ejemplos anteriores, parece que la distribución de X se acerque a un límite.

Obviamente, si N e R son grandes en comparación con n , la probabilidad de sacar una bola roja en la primera repartición ($p = R/N$) es casi la misma como la probabilidad de sacar una roja en la segunda dado que la primera fuese roja ($(R - 1)/(N - 1)$) o dado que la primera fuese azul ($R/(N - 1)$) etc.

Podemos formalizar este idea recordando la fórmula alternativa para la probabilidad hipergeométrica:

$$\begin{aligned}
P(X = x) &= \frac{\binom{R}{x} \binom{N-R}{n-x}}{\binom{N}{n}} = \frac{\binom{n}{x} \binom{N-n}{R-x}}{\binom{N}{R}} \\
&= \binom{n}{x} \frac{(N-n)!}{(R-x)!((N-n)-(R-x))!} \frac{R!(N-R)!}{N!} \\
&= \binom{n}{x} \frac{(N-n)!}{N!} \frac{R!}{(R-x)!} \frac{((N-R)!}{((N-n)-(R-x))!} \\
&= \binom{n}{x} \frac{R \times \cdots \times (R-x+1) \times (N-R) \times \cdots \times ((N-R)-(n-x)+1)}{N \times (N-1) \times \cdots \times (N-n+1)} \\
&= \binom{n}{x} \frac{p \left(p - \frac{1}{N}\right) \cdots \left(p - \frac{x-1}{N}\right) q \left(q - \frac{1}{N}\right) \cdots \left(q - \frac{n-x-1}{N}\right)}{\left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n-1}{N}\right)}
\end{aligned}$$

donde $q = 1 - p$.

Dejando $N, R \rightarrow \infty$ tal que $\frac{R}{N} = p$ en la fórmula anterior, todos los términos $\frac{c}{N} \rightarrow 0$ y tenemos el siguiente teorema.

Teorema 25

Sea $X \sim \mathcal{H}(N, R, n)$ donde $\frac{R}{N} \rightarrow p$ cuando $N \rightarrow \infty$, y n es un número fijo. Luego,

$$P(X = x) \rightarrow \binom{n}{x} p^x (1 - p)^{n-x}$$

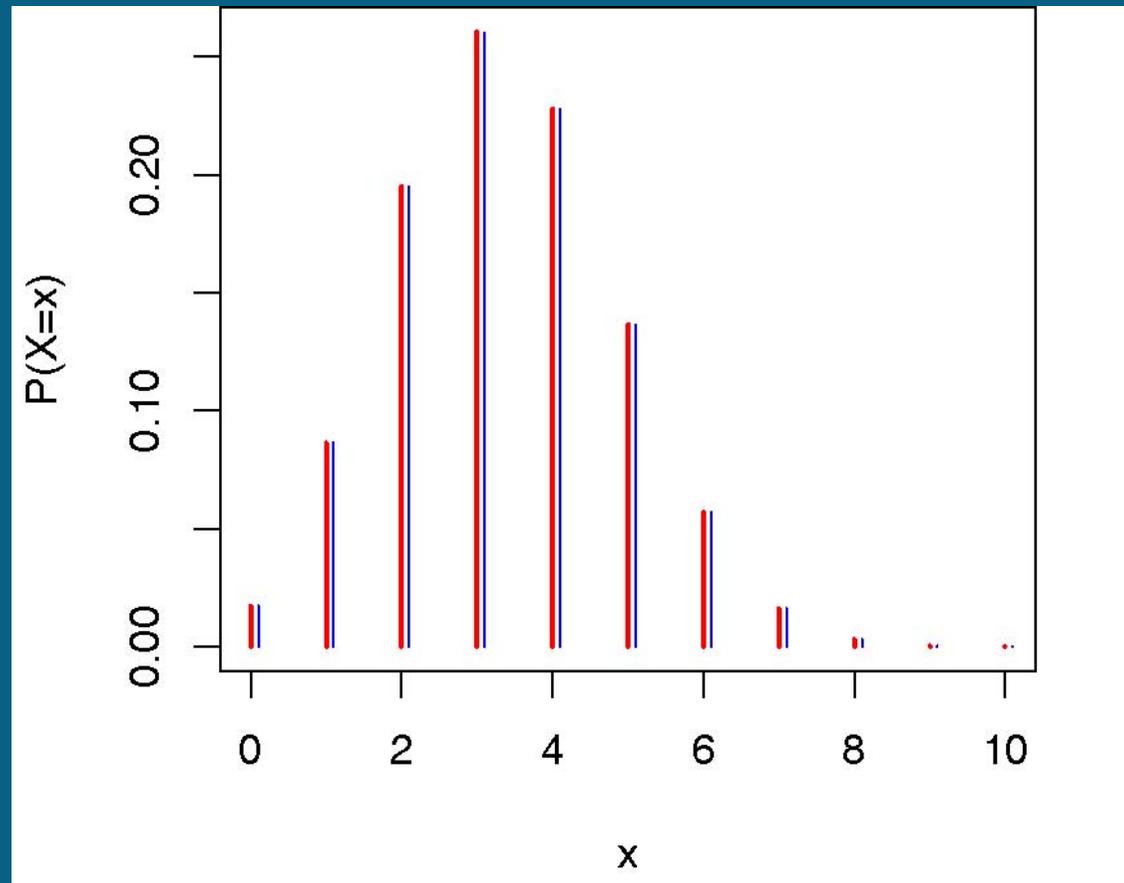
y $X \approx \mathcal{BI}(n, p)$.

El teorema demuestra que en poblaciones grandes, es casi lo mismo muestrear sin reemplazamiento que con reemplazamiento.

Una *regla empírica* es que la aproximación binomial será bastante razonable si

$$\frac{n}{N} \leq 0.1 \quad \text{y} \quad N \geq 50.$$

El siguiente gráfico muestra la distribución hipergeométrica $\mathcal{H}(3000, 1000, 10)$ en rojo y la distribución binomial $\mathcal{BI}(10, 1/3)$. Las probabilidades son casi iguales.



Ejemplo

Un embarque de 200 alarmas contra robo contiene 10 piezas defectuosas. Se selecciona al azar 5 alarmas contra robo para enviarlas a un cliente.

- a. Use la distribución hipergeométrica para encontrar la probabilidad de que el cliente reciba exactamente una alarma contra robo defectuosa.*
- b. Use la aproximación binomial para la distribución hipergeométrica para obtener la probabilidad de que el cliente reciba exactamente una alarma contra robo defectuosa.*
- c. Encuentre el error de la aproximación.*