

Capítulo 3. Distribuciones

Índice

1. Distribuciones discretas	1
1.1. Bernoulli	1
1.2. Binomial	2
1.3. Poisson	4
2. Distribuciones continuas	6
2.1. Normal	6
2.2. Exponencial	8
2.3. Gamma	11
2.4. Beta	12
3. Distribuciones en lenguaje R	15
3.1. Funciones de probabilidad y de densidad	16
3.2. Función de distribución	16
3.3. Función de distribución inversa o cuantiles	17
3.4. Simular datos	17
3.5. Simulación con múltiples muestras	17
4. Ejercicios	19

En este capítulo se describen algunas de las distribuciones más conocidas y de mayor aplicación en ciencias sociales. Habitualmente las distribuciones dependen de uno o varios *parámetros*, que son cantidades que determinan las propiedades de las distribuciones. El conjunto de posibles valores que pueden tomar los parámetros se denomina *espacio paramétrico*, y cada parámetro tiene su espacio paramétrico. Por ejemplo, según se ha visto en cursos anteriores, la distribución normal depende de la media, μ , y la desviación típica, σ ; el espacio paramétrico de μ es $-\infty < \mu < \infty$, en cambio σ no puede ser negativa y su espacio paramétrico es $\sigma > 0$.

1. Distribuciones discretas

Las distribuciones discretas son aquellas que corresponden a una variable aleatoria discreta; es decir, con un número finito o infinito numerable de posibles valores.

1.1. Bernoulli

La distribución de Bernoulli describe los experimentos aleatorios que solamente toman dos resultados, que por conveniencia se indican mediante 0 y 1 y suelen denominarse *fracaso* y *éxito*. La función de probabilidad de una variable de Bernoulli $X = (0, 1)$ es:

$$f(x) = \pi^x(1 - \pi)^{1-x},$$

donde el parámetro $\pi \in (0, 1)$ indica la probabilidad de éxito, es decir, $\pi = \text{Prob.}(X = 1)$. Se comprueba fácilmente que la probabilidad de cada resultado es:

$$\begin{aligned} f(1) &= \pi \\ f(0) &= 1 - \pi \end{aligned}$$

Momentos. El valor esperado y la varianza de una variable de Bernoulli son:

$$\begin{aligned} E(X) &= \pi \\ \text{Var}(X) &= \pi(1 - \pi) \end{aligned}$$

Muestra aleatoria simple. Supongamos que tenemos n observaciones procedentes de una distribución Bernoulli(π): x_1, \dots, x_n . Asumiendo muestreo aleatorio simple, la función de probabilidad de la muestra es

$$\begin{aligned} f(\mathbf{x}) &= \pi^{x_1}(1 - \pi)^{1-x_1} \times \dots \times \pi^{x_n}(1 - \pi)^{1-x_n} \\ &= \pi^{\sum_{i=1}^n x_i} (1 - \pi)^{\sum_{i=1}^n (1-x_i)} \\ &= \pi^y (1 - \pi)^{(n-y)} \end{aligned}$$

Donde $y = \sum_{i=1}^n x_i$ es el número de éxitos en la muestra. Vemos entonces que la función de probabilidad de la muestra depende únicamente del número de éxitos, no de qué observaciones exactamente sean éxito o fracaso. Dos muestras que tengan el mismo número de éxitos tienen también la misma probabilidad aunque dichos éxitos se correspondan con diferentes elementos de la muestra. Es decir, el número de éxitos resume toda la información contenida en la muestra acerca de la distribución de Bernoulli.

Ejemplo 1. Tomamos una muestra aleatoria simple de dos observaciones procedentes de una distribución Bernoulli (π). El espacio muestral de dicho experimento consiste en los cuatro posibles patrones de respuesta compuestos por 0 y 1, es decir:

X_1	X_2
0	0
0	1
1	0
1	1

Supongamos que los valores encontrados en la muestra se indican por el vector $\mathbf{x} = (x_1, x_2)$. Al ser la muestra aleatoria simple, la función de probabilidad de la muestra es el producto de la probabilidad de cada una de las dos observaciones:

$$\begin{aligned} f(\mathbf{x}) &= f(x_1)f(x_2) \\ &= \pi^{x_1}(1 - \pi)^{1-x_1} \times \pi^{x_2}(1 - \pi)^{1-x_2} \\ &= \pi^{x_1+x_2} (1 - \pi)^{2-x_1-x_2}. \end{aligned}$$

Por ejemplo, la probabilidad de obtener un éxito en la muestra es

$$\begin{aligned} P(Y = 1) &= f(0, 1) + f(1, 0) \\ &= 2\pi(1 - \pi) \end{aligned}$$

Vemos entonces que las muestras (0, 1) y (1, 0) tienen la misma probabilidad porque el número de éxitos es igual en ambas aunque dichos éxitos se correspondan con diferentes observaciones en cada muestra.

1.2. Binomial

Supongamos que tenemos una población dividida en dos modalidades: hombres y mujeres, personas que se curan o no curan de una enfermedad, etc. La proporción poblacional de una de las dos modalidades, denominada *éxito*, es π . Tomamos una muestra aleatoria de n observaciones y la variable aleatoria $Y = 0, 1, \dots, n$ indica

el número de éxitos en la muestra. La función de probabilidad de Y se denomina binomial (n, π) , y sus parámetros son el tamaño muestral n y la probabilidad π . La función de probabilidad de Y es:

$$\begin{aligned} f(Y = y) &= \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y} \\ &= \binom{n}{y} \pi^y (1-\pi)^{n-y} . \end{aligned}$$

donde $y!$ es el factorial de y , que consiste en el producto de y por todos los números enteros positivos menores o iguales a él:

$$y! = y(y-1)(y-2) \cdots 1.$$

Por ejemplo, el factorial de algunos números es:

$$0! = 1 \text{ (por definición)}$$

$$1! = 1$$

$$2! = 2$$

$$3! = 6$$

$$4! = 24$$

El término $\binom{n}{y}$ representa el coeficiente binomial, que indica el número de subconjuntos distintos de y elementos que pueden tomarse a partir de un conjunto de n objetos:

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}.$$

Ejemplo 2. Supongamos que tenemos una población dividida en dos modalidades, 0 y 1. A la modalidad 1 se le denomina *éxito* y su probabilidad es π . Tomamos muestras de tamaño dos, por lo que los posibles resultados de cada una de las extracciones, la probabilidad de cada resultado y el número de éxitos se corresponden con la siguiente tabla

X_1	X_2	$f(\mathbf{x})$	Y
0	0	$(1-\pi)^2$	0
0	1	$\pi(1-\pi)$	1
1	0	$\pi(1-\pi)$	1
1	1	π^2	2

Vemos que los resultados $Y = 0$ e $Y = 2$ solo pueden darse de una manera. En cambio el resultado $Y = 1$ puede darse de dos maneras distintas, con la muestra $(0, 1)$ y con $(1, 0)$. El coeficiente binomial indica justamente esto, el número de formas en que puede darse cada resultado.

La función de probabilidad de Y es la binomial $(n = 2, \pi)$. En concreto:

Y	$f(y)$
0	$(1-\pi)^2$
1	$2\pi(1-\pi)$
2	π^2

Ejemplo 3. La distribución binomial $(n = 3, \pi)$ es

Y	$f(y)$
0	$(1 - \pi)^3$
1	$3 \pi(1 - \pi)^2$
2	$3 \pi^2(1 - \pi)$
3	π^3

Por ejemplo, hay tres muestras que conducen al resultado $Y = 1$, que son $(1, 0, 0)$, $(0, 1, 0)$ y $(0, 0, 1)$. El coeficiente binomial nos dice que esto es así, el número de formas de obtener un éxito cuando la muestra es de tamaño tres es

$$\binom{3}{1} = \frac{3!}{1!(3-1)!} = \frac{6}{2} = 3.$$

Cada una de las muestras que conducen al resultado $Y = 1$ tiene un éxito y dos fracasos, por lo que su probabilidad es $\pi(1 - \pi)$. En definitiva, la probabilidad de obtener un éxito es

$$P(Y = 1) = 3\pi(1 - \pi)^2.$$

La figura 1 contiene varias distribuciones binomiales con distintos valores para n y π . Puede verse que la asimetría de la distribución depende de π y que cuando n es grande su forma se asemeja a una distribución normal.

El valor esperado y la varianza de una variable binomial son:

$$\begin{aligned} E(Y) &= n\pi \\ Var(Y) &= n\pi(1 - \pi) . \end{aligned}$$

La distribución de Bernoulli es el caso particular de la binomial cuando $n = 1$. Entonces los posibles valores del número de éxitos, Y , son simplemente 0 y 1. El coeficiente binomial toma el valor 1 y la función de probabilidad de Y se reduce a $f(y) = \pi^y(1 - \pi)^{1-y}$.

1.3. Poisson

La distribución de *Poisson* suele aplicarse cuando la variable aleatoria es una frecuencia, recuento de un número de casos, por lo que tiene gran utilidad para el análisis de tablas de contingencia y en modelos de regresión en los que se predice dicha frecuencia a partir de una variable independiente. La variable aleatoria está definida en el conjunto de los números naturales ($Y = 0, 1, 2, \dots$), por lo que no existe un máximo y se trata de un espacio muestral infinito numerable. La función de probabilidad de Y es:

$$f(Y = y) = \frac{\lambda^y}{y!} \exp(-\lambda)$$

Momentos: La distribución de distribución de Poisson tiene la peculiaridad de que su varianza es igual a su media, y ambas coinciden con el parámetro λ :

$$\begin{aligned} E(Y) &= \lambda \\ Var(Y) &= \lambda \end{aligned}$$

La figura 2 muestra varias distribuciones de Poisson.

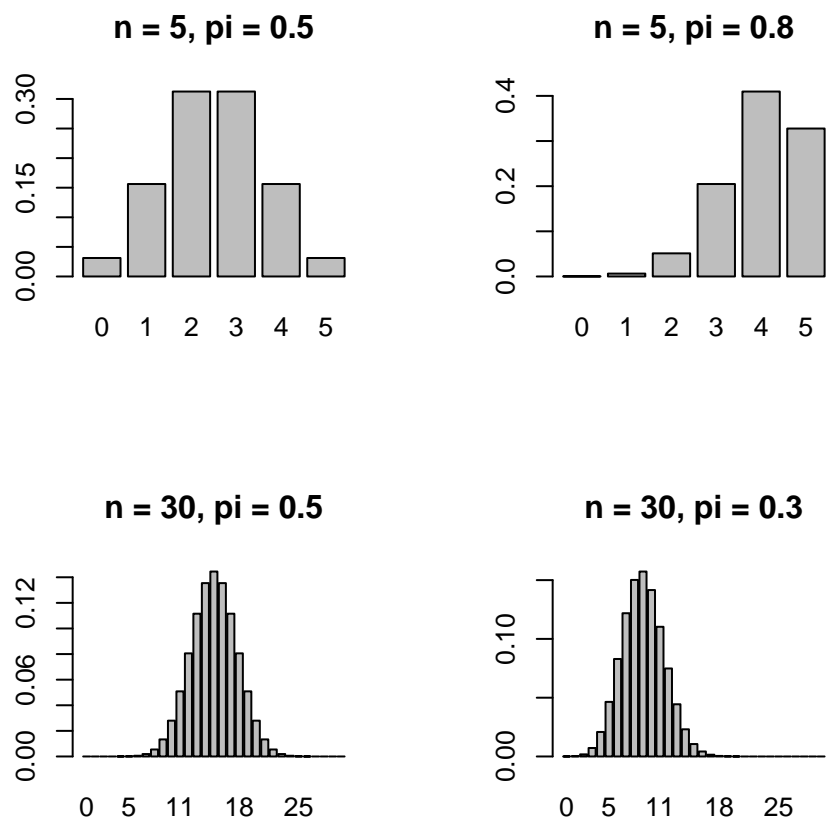


Figura 1: Distribuciones binomiales

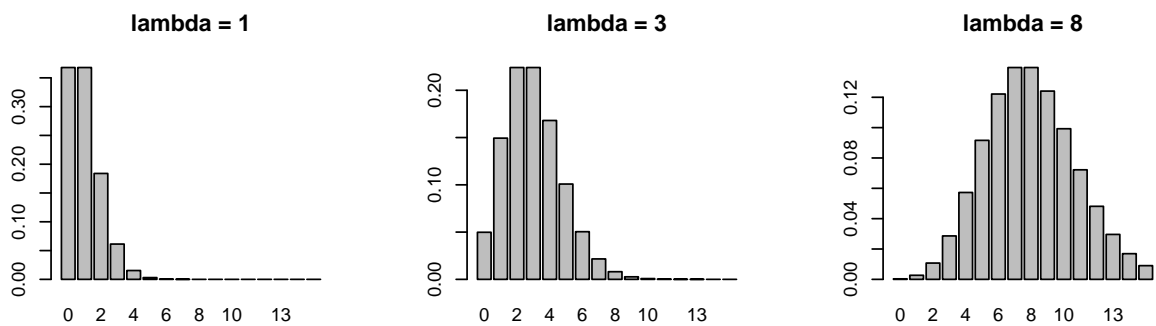


Figura 2: Distribuciones de Poisson

La distribución de Poisson tiene la propiedad de *aditividad*, que significa que si la variable X sigue la distribución de Poisson (λ) e Y es Poisson (δ) e independiente de X , entonces la suma de ambas, $Z = X + Y$, también sigue la distribución de Poisson con parámetro $\lambda + \delta$.

Muestra aleatoria simple. Supongamos que disponemos de una muestra aleatoria simple de tamaño n procedente de la distribución de Poisson, $\mathbf{y} = (y_1, \dots, y_n)$. La función de probabilidad de la muestra es

$$\begin{aligned} f(\mathbf{y}) &= \prod_{i=1}^n \frac{\lambda^{y_i}}{y_i!} \exp(-\lambda) \\ &= \frac{\lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \exp(-n\lambda) \\ &= \frac{\lambda^{n\bar{Y}}}{\prod_{i=1}^n y_i!} \exp(-n\lambda) \end{aligned}$$

Ejemplo 4. Supongamos que estamos investigando un tratamiento para el tabaquismo, Y representa el número de cigarrillos fumados al día y X es el número de sesiones de tratamiento que ha recibido un paciente. Queremos estimar la regresión de Y sobre X . Como Y es una frecuencia y no puede ser menor de cero ni tomar valores decimales, no tiene sentido asumir normalidad, como es habitual en el modelo de regresión lineal simple estudiado en cursos anteriores. Por el contrario, podemos asumir que, para un valor fijo de X , la distribución de Y es Poisson con parámetro

$$\lambda = \exp(a + bX),$$

donde a y b son el intercepto y la pendiente de la regresión. El parámetro λ indica el valor esperado de Y condicionado a X , es decir $\lambda = E(Y | X)$. El modelo resultante se conoce como regresión de Poisson. La regresión de Poisson

2. Distribuciones continuas

Las distribuciones continuas se utilizan cuando la variable aleatoria toma valores en un intervalo de números reales. El espacio muestral de una variable continua es infinito no numerable.

2.1. Normal

Es una de las distribuciones más conocidas y utilizadas en ciencias sociales. Esto se debe en parte a que es la distribución muestral de la media cuando n es grande, según afirma el *teorema del límite central*. Gracias al teorema del límite central, la distribución normal es la que se emplea en contrastes sobre medias y en otros contrastes, como los de chi-cuadrado, que son transformaciones de la distribución normal. También se denomina distribución *gaussiana* en honor de Karl Friedrich Gauss, que derivó su ecuación a partir del estudio de los errores que se cometen al realizar repetidas veces una medición en determinadas condiciones.

Una variable aleatoria distribuida según la normal(μ, σ) toma valores en el intervalo $(-\infty, \infty)$ y su función de densidad es:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right).$$

Momentos. La distribución normal tiene varias propiedades que la hacen muy conveniente:

1. El valor esperado de la distribución normal es $E(Y) = \mu$.

2. La varianza es $Var(Y) = \sigma^2$.

Además la distribución normal cumple la propiedad de aditividad. Supongamos que X es una variable normal (μ_X, σ_X) e Y es normal (μ_Y, σ_Y) . Entonces la variable $Z = X \pm Y$ es normal (μ_Z, σ_Z) , donde $\mu_Z = \mu_X \pm \mu_Y$, y la varianza es $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \pm 2Cov(X, Y)$.

Muestra aleatoria simple. Supongamos que tomamos una muestra aleatoria simple procedente de una distribución normal (μ, σ) . La función de densidad de probabilidad conjunta de la muestra es

$$\begin{aligned} f(\mathbf{y}) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_i - \mu}{\sigma}\right)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2}\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sigma^2}\right) \end{aligned}$$

El término $\sum_{i=1}^n (y_i - \mu)^2$ puede desarrollarse del modo

$$\begin{aligned} \sum_{i=1}^n (y_i - \mu)^2 &= \sum_{i=1}^n (y_i - \bar{Y} + \bar{Y} - \mu)^2 \\ &= \sum_{i=1}^n (y_i - \bar{Y})^2 + \sum_{i=1}^n (\bar{Y} - \mu)^2 - 2 \sum_{i=1}^n (y_i - \bar{Y})(\bar{Y} - \mu) \\ &= nS_Y^2 + n(\bar{Y} - \mu)^2 \end{aligned}$$

Por tanto, la función de densidad de la muestra es

$$f(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{nS_Y^2 + n(\bar{Y} - \mu)^2}{2\sigma^2}\right)$$

Cómo $f(\mathbf{y})$ sólo depende de \bar{Y} y de S_Y^2 , dos muestras que tengan la misma media y varianza tienen también la misma densidad de probabilidad. Por este motivo la media y la varianza resumen toda la información disponible en la muestra acerca de la distribución normal, y se dice que \bar{Y} y S_Y^2 son *estadísticos suficientes*.

Distribuciones derivadas de la normal. Hay varias distribuciones que se derivan de la normal. Una de ellas es la chi-cuadrado. Si dos variables, X_1 y X_2 , siguen la distribución normal $(0, 1)$ y son estadísticamente independientes, la distribución de la suma de las variables al cuadrado

$$Y = X_1^2 + X_2^2$$

es chi-cuadrado con dos grados de libertad. De modo general, si X_1, \dots, X_n son variables normales estandarizadas e independientes entre si, y definimos la suma de cuadrados

$$Y = X_1^2 + \dots + X_n^2$$

entonces $Y \sim \text{chi-cuadrado}(n)$.

Ejemplo 5. Según se ha estudiado de cursos anteriores, la distribución muestral de la media es normal $(\mu, \sigma/\sqrt{n})$. Por tanto la variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

sigue la distribución normal $(0, 1)$. En consecuencia, la variable

$$Z^2 = \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

sigue una distribución chi-cuadrado(1).

Ejemplo 6. Dos variables, X e Y , siguen una distribución normal $(10, 6)$ y normal $(15, 2)$, siendo su covarianza 10. Calculamos la suma $Z = X + Y$ cuya distribución es normal gracias a la propiedad de aditividad y tiene como parámetros

$$\begin{aligned}\mu_Z &= \mu_X + \mu_Y = 10 + 15 = 15 \\ \sigma_Z^2 &= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = 36 + 4 + 2 \times 10 = 60\end{aligned}$$

La figura 3 muestra las funciones de densidad de las tres distribuciones.

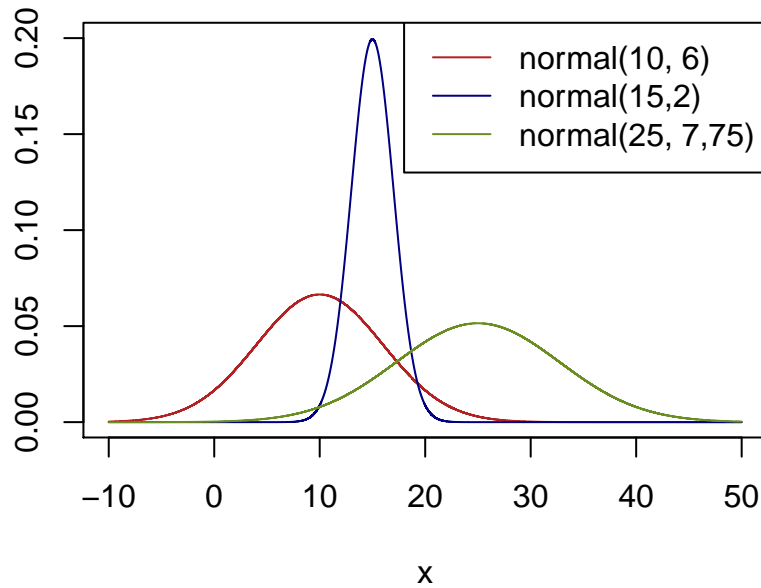


Figura 3: Distribuciones normales

2.2. Exponencial

La distribución exponencial se define en el intervalo $(0, \infty)$. Se utiliza habitualmente con variables aleatorias que indican una medida de tiempo. Por ejemplo, el tiempo que tarda un sujeto en completar una determinada tarea o el tiempo de vida útil de un producto. La función de densidad de una variable exponencial con parámetro ω , siendo $\omega > 0$, es:

$$f(y) = \omega \exp(-\omega y) .$$

Momentos. La media y varianza de una variable exponencial son:

$$E(Y) = \frac{1}{\omega}$$

$$Var(Y) = \frac{1}{\omega^2}$$

El parámetro ω puede interpretarse como la velocidad de ejecución de proceso modelizado porque es inversamente proporcional al valor esperado de Y , a mayor valor de ω menor tiempo esperado. En la figura 5 pueden verse tres funciones de densidad exponencial con diferentes valores de ω .

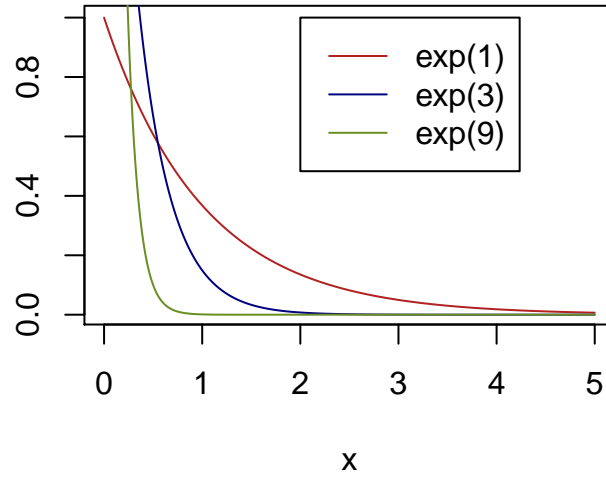


Figura 4: Distribuciones exponenciales

Muestra aleatoria simple. La función de densidad conjunta de una muestra aleatoria simple de la distribución exponencial es

$$f(\mathbf{y}) = \prod_{i=1}^n \omega \exp(-\omega y_i)$$

$$= \omega^n \exp\left(-\omega \sum_{i=1}^n y_i\right)$$

$$= \omega^n \exp(-\omega n \bar{Y})$$

Por tanto, la media muestral, \bar{Y} , es el estadístico suficiente.

En estadística es muy habitual trabajar con el logaritmo de las funciones de densidad en lugar de con la función de densidad en si misma. Esto se debe a que al tomar el logaritmo la función resultante es casi siempre más sencilla, y suele tener una forma lineal. En el caso de la distribución exponencial el logaritmo de $f(\mathbf{y})$ es

$$\log f(\mathbf{y}) = \log \omega^n \exp(-\omega n \bar{Y})$$

$$= n \log \omega - \omega n \bar{Y}$$

Ejemplo 7. Hemos tomado una muestra aleatoria simple de tres elementos procedentes de una distribución exponencial y el resultado es

$$\mathbf{y} = (1, 3, 2)$$

Cómo $\bar{Y} = 2$, la función de densidad de la muestra y su logaritmo son

$$f(\mathbf{y}) = \omega^3 \exp(-\omega 6)$$

$$l(\mathbf{y}) = 3 \log \omega - \omega 6$$

donde $l(\mathbf{y})$ es el logaritmo de la función de densidad, $l(\mathbf{y}) = \log f(\mathbf{y})$. La figura 6 muestra la representación de estas dos funciones para valores de ω entre 0 y 2. Puede advertirse que ambas funciones tienen el máximo en el mismo punto. Esto se debe a que el logaritmo es una función monótona, si $f(\mathbf{y})$ es creciente (decreciente) en ω la función $l(\mathbf{y})$ también es creciente (decreciente); por esta razón, el máximo de $f(\mathbf{y})$ y $l(\mathbf{y})$ está situado en el mismo valor de ω .

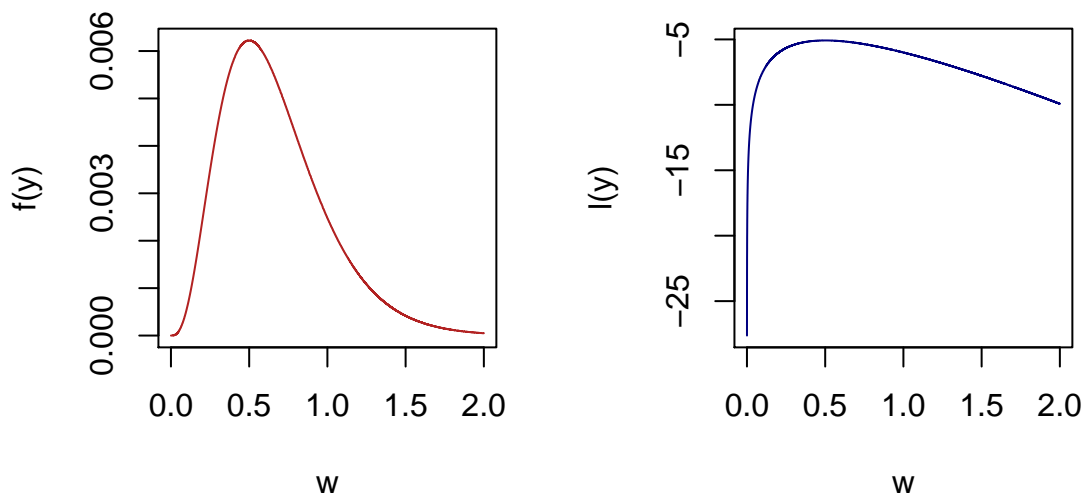


Figura 5: Función de densidad de la muestra y su logaritmo

Queremos saber cual es el valor de ω más compatible con los datos. Cómo $f(\mathbf{y})$ es la función de densidad de los datos para un valor determinado de ω , el valor de ω más compatible con nuestra muestra será aquel que maximice $f(\mathbf{y})$. El siguiente código R busca el valor de ω que maximiza $f(\mathbf{y})$ y $l(\mathbf{y})$. Puede verse que el valor de ω es el mismo en ambos casos, y resulta ser, redondeando, $\omega = 0,5$. Este resultado es bastante razonable teniendo en cuenta que $E(Y) = 1/\omega$ y que la media de esta muestra ha sido $\bar{Y} = 2$. De nuevo este ejemplo nos permite ver que R trabaja con aproximaciones numéricas, si la media es 2 el valor de omega debería ser exactamente 0,5. R da un resultado con un grado de aproximación muy bueno, pero que no es idéntico al valor real.

```
f <- function(w) w^3 * exp(-w * 6)
l <- function(w) 3*log(w) - w*6
fMax <- optimize(f, c(0, 2), maximum=TRUE)
lMax <- optimize(l, c(0, 2), maximum=TRUE)
print(fMax)
```

```
## $maximum
## [1] 0.4999931
##
## $objective
## [1] 0.006223384
```

```
print(lMax)
```

```
## $maximum
## [1] 0.4999991
##
## $objective
## [1] -5.079442
```

Podemos preguntarnos cual es la ventaja de trabajar con $l(\mathbf{y})$ en lugar de con $f(\mathbf{y})$ dado que R ha sido igualmente capaz de encontrar el máximo en ambos casos. Si tuviéramos que buscar el máximo mediante procedimientos de cálculo matemático, tomando la derivada de la función para buscar el máximo, es claro que sería más sencillo hacerlo con $l(\mathbf{y})$ dado que es una función más simple. Al ser más sencilla, esto tiene la ventaja de que el error de aproximación que cometa R será en general menor con $l(\mathbf{y})$ que con $f(\mathbf{y})$; esto puede apreciarse en el ejemplo atendiendo a los últimos decimales de ω y que será más relevante cuanto mayor sea la muestra y más compleja la función de distribución. Por otra parte, en temas posteriores veremos que algunos resultados de interés estadístico requieren necesariamente trabajar con $l(\mathbf{y})$.

2.3. Gamma

La distribución gamma está definida únicamente para valores positivos de la variable aleatoria, por lo que se utiliza en el análisis de tiempos de respuesta, por ejemplo, en el contexto de la psicología cognitiva. La función de densidad gamma es

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

donde $\alpha > 0$ y $\beta > 0$ son los parámetros de la distribución. La función gamma, Γ , da nombre a la distribución y es una generalización de la función factorial. En concreto, si x es un número entero entonces $\Gamma(x) = (x-1)!$. La función Γ es matemáticamente compleja y no profundizaremos en ella. Es fácil ver que la distribución exponencial es un caso particular de la gamma que se obtiene fijando $\alpha = 1$.

Momentos. El valor esperado y varianza de la distribución gamma son

$$E(X) = \frac{\alpha}{\beta}$$

$$Var(X) = \frac{\alpha}{\beta^2}$$

Además la moda de la distribución gamma, para $\alpha > 1$, es

$$Moda = \frac{\alpha - 1}{\beta}$$

Los valores de α y β también pueden calcularse en función de la media y la varianza de la distribución. En concreto, despejando

$$\alpha = \frac{\mu^2}{\sigma^2}$$

$$\beta = \frac{\mu}{\sigma^2}$$

La figura 7 muestra varias funciones de densidad gamma para distintos valores de α y β .

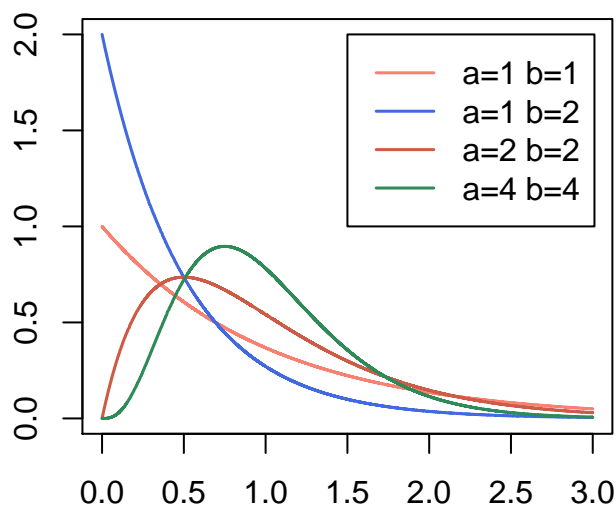


Figura 6: Distribuciones gamma

2.4. Beta

La distribución beta se utiliza con una variable aleatoria continua definida en el intervalo $(0, 1)$. Esta distribución suele aplicarse cuando la variable aleatoria es una proporción o una frecuencia. Por ejemplo, supongamos una escala de actitudes compuesta por 100 preguntas que deben responderse como acuerdo/desacuerdo; si la puntuación total del sujeto en la escala se modela mediante la proporción de preguntas en que el sujeto está de acuerdo, puede utilizarse la distribución beta para elaborar un modelo.

Una variable aleatoria $X \in (0, 1)$ sigue la distribución beta (α, β) si su función de densidad es

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

La función que aparece en el denominador, $B(\alpha, \beta)$, es la denominada *función beta* que da nombre a la distribución; en este curso no profundizaremos en la función $B(\alpha, \beta)$ por ser compleja matemáticamente, basta con saber que es la constante de normalización y que juega un papel similar al que tenía el coeficiente $\binom{n}{y}$ en la distribución binomial.

Momentos. El valor esperado, varianza y coeficiente de asimetría de la distribución beta son

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

$$Asimetría = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}}$$

donde $\alpha > 0$ y $\beta > 0$. Además la moda de la distribución es

$$Moda = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Una de las características de la distribución beta es su gran flexibilidad. Puede ser una distribución uniforme, tener forma de U, de U invertida y ser simétrica o asimétrica. La figura 8 muestra distintos ejemplos de distribuciones beta.

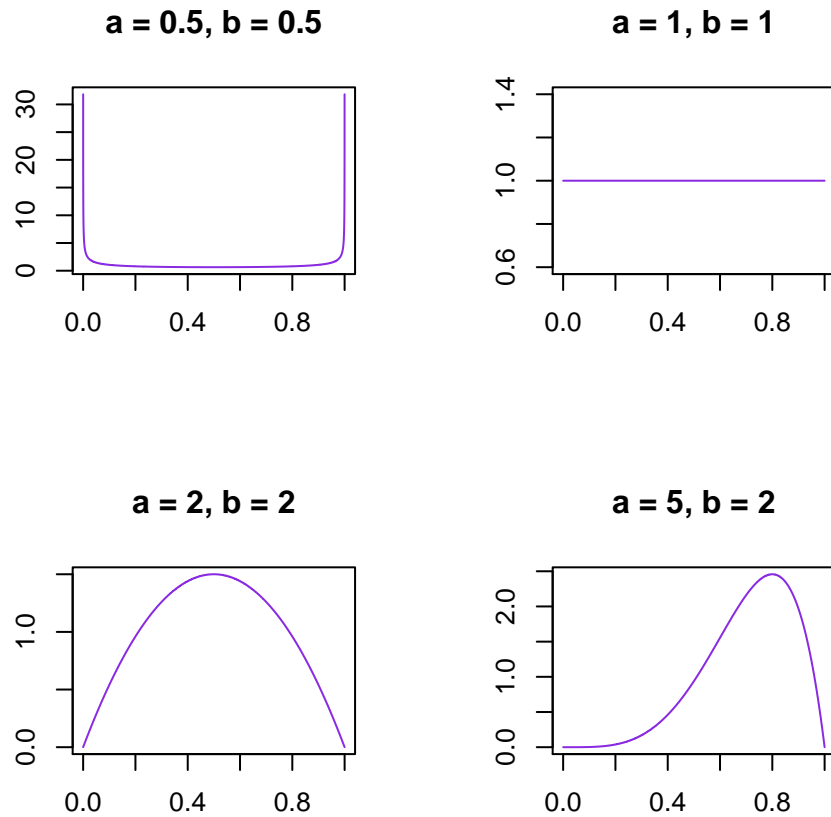


Figura 7: Distribuciones beta

Ejemplo 8. Una muestra de 300 personas responde a un examen de 50 preguntas. La figura 9 muestra la distribución de la proporción de aciertos obtenida por los sujetos junto con las funciones de densidad normal y beta ajustados al mismo. La distribución beta tiene algunas ventajas con respecto a la normal. En primer lugar, la distribución es simétrica y no capta la asimetría positiva existente en estos datos. Por otra parte,

una variable normal tiene rango $(-\infty, \infty)$ por lo que esta distribución asume que existe una probabilidad superior a cero de encontrar valores fuera del rango $(0, 1)$ en el que se define la variable objeto de estudio.

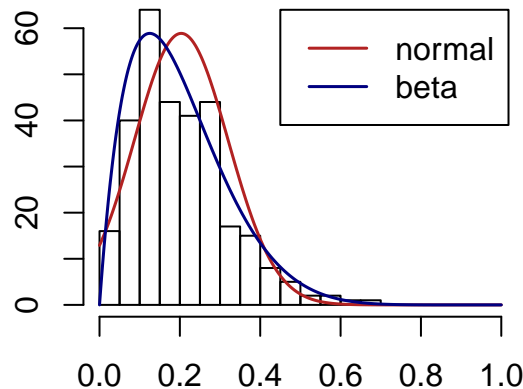


Figura 8: Distribución de la proporción de aciertos

Para ajustar la distribución beta a estos datos hemos utilizado los momentos de la distribución. Es posible despejar el valor de α y β en función de la media y varianza estimadas en la muestra, con lo que se llega a las expresiones

$$\alpha = \bar{X} \left(\frac{\bar{X}(1 - \bar{X})}{S^2} - 1 \right)$$

$$\beta = (1 - \bar{X}) \left(\frac{\bar{X}(1 - \bar{X})}{S^2} - 1 \right)$$

Podemos implementar estas fórmulas en R para obtener los valores de los parámetros:

```
alpha <- function(m,v) m *(m*(1-m)/v - 1)
beta <- function(m,v) (1-m)*(m*(1-m)/v - 1)
asimetria <- function(a,b) 2*(b-a)*sqrt(a+b+1)/((a+b+2)*sqrt(a*b))
media <- mean(muestra)
varianza <- var(muestra)
a <- alpha(media, varianza)
b <- beta(media, varianza)
cat(sprintf(
  "Los datos de la muestra son: media = %6.4f, varianza = %6.4f y asimetría = %6.4f.
  Los parámetros estimados son: alpha = %6.4f y beta = %6.4f.",
  media, varianza, asimetria(a,b), a, b))
```

```
## Los datos de la muestra son: media = 0.2034, varianza = 0.0136 y asimetría = 0.7885.
## Los parámetros estimados son: alpha = 2.2142 y beta = 8.6708.
```

Ejemplo 9. Sea X una variable aleatoria con distribución $\text{beta}(2, 3)$. Según hemos visto, el valor esperado de X es

$$E(X) = \frac{2}{2+3} = 0,4$$

Vamos a reproducir este resultado utilizando la integración numérica en R. Es decir, crearemos un programa que aplique la fórmula:

$$E(x) = \int_0^1 x \frac{x(1-x)^2}{B(2,3)} dx$$

El código es el siguiente

```
xfx <- function(x) x * dbeta(x, 2, 3)
Ex <- integrate(xfx, lower = 0, upper = 1)
print(Ex)
```

```
## 0.4 with absolute error < 4.4e-15
```

3. Distribuciones en lenguaje R

El lenguaje R asocia cuatro funciones a cada distribución de probabilidad. Estas funciones llevan por nombre una letra (*d*, *p*, *q* o *r*) seguida del nombre de la distribución. Por ejemplo, en el caso de la distribución normal las funciones son

Comando de R	Función
dnorm(x, mean = 0, sd = 1, log = FALSE)	Función de densidad, $f(x)$
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)	Función de distribución, $F(x)$
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)	Función de distribución inversa
rnorm(n, mean = 0, sd = 1)	Simular datos

La lista completa de funciones de densidad incluidas en R es la siguiente.

Palabra clave en R	Distribución
<i>beta</i>	beta
<i>binom</i>	binomial
<i>cauchy</i>	Cauchy
<i>chisq</i>	chi-cuadrado
<i>exp</i>	exponencial
<i>f</i>	F
<i>gamma</i>	gamma
<i>geom</i>	geométrica
<i>hyper</i>	hipergeométrica
<i>lnorm</i>	log-normal
<i>multinom</i>	multinomial
<i>nbinom</i>	binomial negativa
<i>norm</i>	normal
<i>pois</i>	Poisson
<i>t</i>	<i>t</i> de Student
<i>unif</i>	uniforme
<i>weibull</i>	Weibull

Con estas distribuciones podemos realizar las mismas operaciones que hemos visto en el caso de la distribución normal añadiendo la primera al nombre de la palabra clave. Para obtener una explicación adicional sobre el sentido de cada función podemos utilizar el comando `?` de R. Por ejemplo, ejecutando `?pchisq` obtendremos una explicación de las funciones relativas a la distribución chi-cuadrado. A continuación veremos algunos ejemplos.

3.1. Funciones de probabilidad y de densidad

El comando `dnorm` resulta útil para realizar la representación gráfica de una función de probabilidad o densidad. Para ello, hay que definir en primer lugar el rango de valores de la variable X , después calculamos el valor de $f(x)$ para cada valor de x y finalmente representamos ambas cantidades. Por ejemplo, el código de la figura 10 es:

```
par(mfrow=c(1,2))
y <- 0:10
f1 <- dpois(y,4)
barplot(f1, names=as.character(y), main="Poisson(4)")
x <- seq(0, 10, by=0.001)
f2 <- dchisq(x,4)
plot(x, f2, type="l", lwd=1.5, col="darkblue", main="chi-cuadrado(4)", xlab="", ylab="")
```

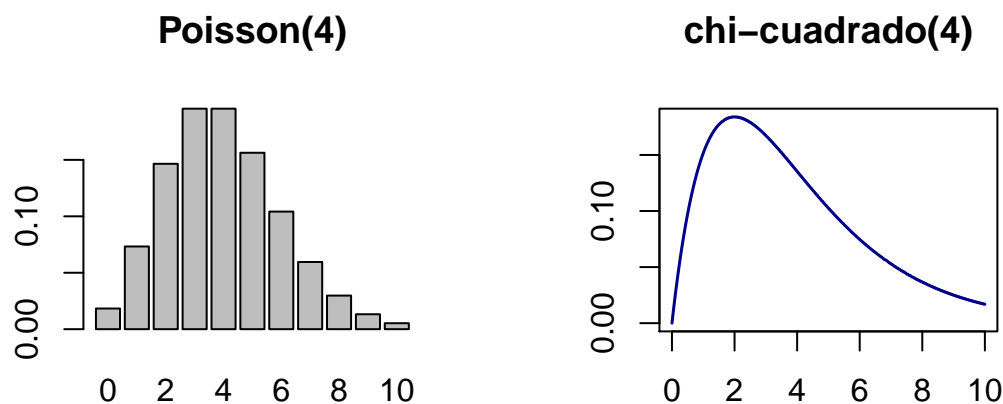


Figura 9: Distribuciones discreta y continua con igual media

3.2. Función de distribución

El comando `pnorm` proporciona el valor de la función de distribución normal $F(x)$, que indica la probabilidad de que X sea menor o igual de un valor concreto: $F(x) = P(X \leq x)$. Podemos utilizar `pnorm` tanto para representar gráficamente una distribución normal como para conocer la probabilidad asociada a una variable. Por ejemplo, para conocer cual es la probabilidad de que X , distribuido según sea menor o igual de 80, es decir $F(-1, 68)$, ejecutamos el código `pnorm(80)`.

El argumento `lower.tail` de `pnorm` indica si debe obtenerse la probabilidad a la izquierda de x , cuando `lower.tail` es `FALSE` el comando `pnorm` proporciona la probabilidad del lado derecho. Por ejemplo, si quisiéramos conocer $P(X \geq 6)$ siendo $X \sim \text{binomial}(n = 10, \pi = 0,5)$ ejecutamos el código


```
pbinom(5, 10, 0.5, lower.tail=FALSE)
```

```
## [1] 0.3769531
```

3.3. Función de distribución inversa o cuantiles

El comando *qnorm* proporciona los cuantiles de una distribución normal, es decir el valor de la función de distribución inversa $F^{-1}(p)$. Dada una probabilidad p , *qnorm* el valor de la variable aleatoria al que le corresponde dicha probabilidad. Por ejemplo, supongamos que necesitamos conocer cual es el valor de Z deja a su izquierda la probabilidad $p = 0,05$ en una distribución normal estándar. Esto podemos calcularlo con el código *qnorm(0,05)*. Si quisiéramos saber cual es el valor de Z que deja a su derecha la probabilidad 0,05 podríamos hacerlo de dos maneras, mediante *qnorm(0.05, lower.tail=FALSE)* o también *qnorm(0.95)*.

3.4. Simular datos

El comando *rnorm* permite simular una muestra de datos de tamaño n de una distribución normal. Por ejemplo, con el comando *hist(rnorm(100))* obtendríamos el histograma de frecuencias de 100 valores tomados al azar de la distribución normal estándar.

Vamos a realizar un pequeño estudio de simulación. Tomaremos una muestra de tamaño nueve de una distribución normal(100, 15); después calculamos \bar{X} y S^2 .

```
muestra <- rpois(100, 15)
cat(sprintf("La media es %5.2f y la varianza %5.2f", mean(muestra), sd(muestra)))
```

```
## La media es 15.66 y la varianza 4.05
```

Podemos realizar con sencillez simulaciones más sofisticadas. Por ejemplo, simularemos 1000 muestras de tamaño 50 de una distribución beta(0,5, 0,5). A continuación calculamos la media de las muestras, obtenemos el histograma de frecuencias de las medias, la estimación del valor esperado de la distribución muestral de la media y la estimación de la varianza de las medias. El resultado aparece en la figura 11.

```
n <- 50
muestras <- 1000
dat <- rbeta(muestras*n, 0.5, 0.5)
muestra <- matrix(dat, ncol=n)
medias <- colMeans(muestra)
hist(medias, ylab="Frecuencia", main=
  sprintf("Media = %4.2f, Sd = %4.2f", mean(medias), sd(medias)))
```

La figura 9 ilustra uno de los resultados más importantes de la estadística. La distribución muestral de la media es aproximadamente normal en muestras grandes con independencia de cual sea la distribución original de la variable aleatoria. En este ejemplo la variable es beta(0,5, 0,5), que es altamente divergente de la distribución normal, aún así la distribución muestral de las medias de dicha variable tiende a la normal al aumentar n .

3.5. Simulación con múltiples muestras

Un estudio de simulación real no se basa sólo en una muestra, es necesario tomar un número elevado de muestras de la distribución bajo estudio y calcular las cantidades de interés en cada una de las muestras. El número de muestras simuladas depende de características de la simulación cómo la complejidad de los cálculos a realizar en cada muestra, y oscila entre las decenas y los miles de muestras.

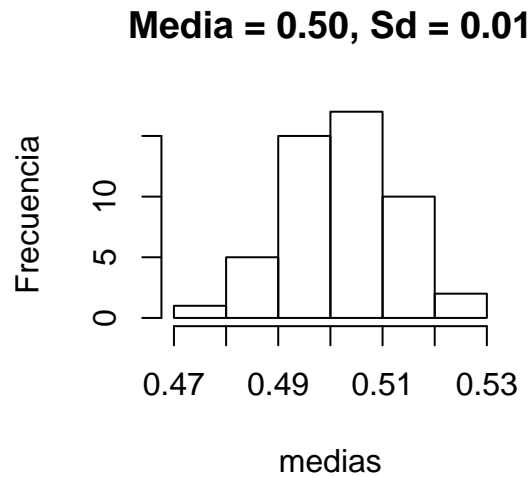


Figura 10: Distribución muestral de la media de una distribución beta

El lenguaje R proporciona varias formas para realizar una simulación con más de una muestra. A continuación veremos dos de ellas, el bucle *for* y la función *replicate*. Aplicaremos ambos procedimientos a un problema muy sencillo, obtener la media y la varianza de la media muestral cuando los datos proceden de una distribución chi-cuadrado(10) y el tamaño muestral es 50. Además tomaremos 1000 muestras simuladas de dicha distribución. Un resultado conocido de la estadística matemática es que una distribución chi-cuadrado(gl) tiene $E(\bar{X}) = gl$ y $Var(\bar{X}) = 2gl/n$, vamos a comprobarlo mediante simulación.

El bucle *for* recibe como entrada un conjunto de valores y el nombre de una variable que tiene que iterar sobre ellos. En el ejemplo, los valores son los números enteros del 1 al 1000 y a la variable le llamaremos *muestra*. Además vamos a definir un vector que contenga el valor de las 1000 muestras. El código es

```
muestras <- 1000
vectorMedias <- rep(0, 1000)
for(muestra in 1:muestras){
  vectorMedias[muestra] <- mean(rchisq(50, 10))
}

cat(sprintf("La media es %5.2f y la varianza %5.2f", mean(vectorMedias), var(vectorMedias)))
```

```
## La media es 10.02 y la varianza 0.40
```

La función *replicate* recibe como argumentos de entrada el número de repeticiones y el código que debe repetirse. Como argumento de salida proporciona un vector con el resultado del cálculo de dicho código. La simulación se podría programar del siguiente modo

```
muestras <- 1000
vectorMedias <- replicate(muestras, mean(rchisq(50, 10)))

cat(sprintf("La media es %5.2f y la varianza %5.2f", mean(vectorMedias), var(vectorMedias)))
```

```
## La media es 10.00 y la varianza 0.42
```

4. Ejercicios

Ejercicio 1. Demuestre las expresiones de $E(X)$ y $Var(X)$ siendo X una variable de Bernoulli.

Ejercicio 2. Sea X una variable aleatoria con rango $(1, 3)$ que sigue una función de densidad uniforme:

$$f(x) = \frac{1}{2}$$

Obtenga el valor esperado de X , su varianza y la función de probabilidad, $F(x)$.

Ejercicio 3. Obtenga la función de distribución acumulada exponencial $F(y)$.

Ejercicio 4. Sea X una variable normal (μ, σ) . Obtenga los puntos de inflexión de $f(x)$.

Ejercicio 5. Demuestre que en la distribución exponencial $E(X) = 1/\omega$.

Ejercicio 6. Sea $\mathbf{x} = (2, 1, 3)$ una muestra aleatoria simple procedente de una distribución de Poisson. Utilizando R, obtenga el valor de λ que maximiza $f(\mathbf{x})$.

Ejercicio 7. Sea X una variable distribuida según la beta $(3, 2)$. Utilice las funciones de cálculo numérico de R para obtener:

1. El valor esperado, μ .
2. La varianza, σ^2 .
3. La moda.
4. El índice de asimetría de Pearson: $A = (\mu - \text{moda})/\sigma$.