Capítulo 5

Propiedades en una muestra aleatoria

5.1. Conceptos básicos sobre muestras aleatorias

Definición 5.1.1 $X_1, \cdots, X_n$ son llamadas una *muestra aleatoria de tamaño n de una población* $f(x)$ si son variables aleatorias mutuamente independientes y la función de probabilidad o densidad marginal de cada X_i es $f(x)$. Alternativamente, $X_1, \cdots, X_n$ son llamadas variables aleatorias independientes e identicamente distribuidas con función de probabilidad o densidad $f(x)$.

Si la función de probabilidad o densidad es miembro de una familia paramétrica $f(x|\theta)$, entonces la función de probabilidad o densidad conjunta es:

$$f(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{5.1.1}$$

Ejemplo 5.1.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $\mathcal{E}(\beta)$ que corresponde al tiempo de funcionamiento (en años) de n circuitos idénticos sometidos a prueba. La función de densidad conjunta de la muestra es:

$$f(x_1, \cdots, x_n|\beta) = \prod_{i=1}^{n} f(x_i|\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-(x_1+\cdots+x_n)/\beta}$$

La probabilidad que todos los circuitos funcionen al menos dos años es:

$$
\begin{aligned}
\Pr(X_1 > 2, \cdots, X_n > 2) &= \int_2^\infty \cdots \int_2^\infty \frac{1}{\beta^n} e^{-(x_1+\cdots+x_n)/\beta} dx_1 \cdots dx_n \\
&= e^{-2/\beta} \int_2^\infty \cdots \int_2^\infty \frac{1}{\beta^{n-1}} e^{-(x_2+\cdots+x_n)/\beta} dx_2 \cdots dx_n \\
&= (e^{-2/\beta})^n = e^{-2n/\beta}
\end{aligned}
$$

Usando independencia:

$$
\begin{aligned}
\Pr(X_1 > 2, \cdots, X_n > 2) &= \Pr(X_1 > 2) \cdots \Pr(X_n > 2) \\
&= (e^{-2/\beta})^n = e^{-2n/\beta}
\end{aligned}
$$

5.2. Sumas de variables aleatorias a partir de una muestra aleatoria

Definición 5.2.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria de tamaño n de una población y sea $T(X_1, \cdots, X_n)$ una función cuyo dominio incluye el espacio muestral de $(X_1, \cdots, X_n)$, entonces la variable aleatoria $Y = T(X_1, \cdots, X_n)$ es llamada una *estadística* cuya distribución es llamada la *distribución de muestreo de* Y.

Definición 5.2.2 La *media muestral* es el promedio aritmético de los valores en la muestra aleatoria. Usualmente se denota por:

$$
\overline{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i
$$

Definición 5.2.3 La *varianza muestral* es la estadística definida por:

$$
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2
$$

La *desviación estándar muestral* es la estadística definida por $S = \sqrt{S^2}$.

Teorema 5.2.1 Sean $x_1, \cdots, x_n$ números cualesquiera y $\bar{x} = (x_1 + \cdots + x_n)/n$, entonces:

a. $\text{mín} \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2$

b. $(n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$

Teorema 5.2.2 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población con media μ y varianza $\sigma^2 < \infty$, entonces:

a. $\mathrm{E}[\overline{X}] = \mu$

b. $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$

c. $\mathrm{E}[S^2] = \sigma^2$

Teorema 5.2.3 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población con función generatriz de momentos $M_X(t)$, entonces la función generatriz de momentos de la media muestral es:

$$M_{\overline{X}}(t) = [M_X(t/n)]^n$$

Ejemplo 5.2.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $\mathcal{N}(\mu, \sigma^2)$. Hallar la distribución de $\overline{X}$.

Ejemplo 5.2.2 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $\mathcal{G}(\alpha, \beta)$. Hallar la distribución de $\overline{X}$.

5.3. Muestreo desde la distribución Normal

5.3.1. Propiedades de la media y variancia muestral

Teorema 5.3.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria de la distribución $\mathcal{N}(\mu, \sigma^2)$ y sean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Entonces:

a. $\overline{X}$ y S^2 son variables aleatorias independientes.

b. $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

c. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Lema 5.3.1 Sea χ_p^2 una variable aleatoria con distribución chi-cuadrado con p grados de libertad.

a. Si $Z \sim \mathcal{N}(0,1)$ entonces $Z^2 \sim \chi_1^2$.

b. Si $X_i \sim \chi_{p_i}^2$ son independientes, entonces $X_1 + \cdots + X_n \sim \chi_{p_1+\cdots+p_n}^2$.

5.3.2. Distribuciones derivadas: t de Student y F

Definición 5.3.1 Sea $X_1, \cdots, X_n$ es una muestra aleatoria de una distribución $\mathcal{N}(\mu, \sigma^2)$. La cantidad $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ tiene *distribución t-student con $n-1$ grados de libertad.* Equivalentemente, una variable aleatoria T tiene distribución t de student con p grados de libertad, y se denota por $T \sim t_p$, si tiene la siguiente función de densidad:

$$f_T(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(p\pi)^{1/2}(1+t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty \tag{5.3.1}$$

Si $p = 1$ entonces 5.3.1 se convierte en la distribución Cauchy, lo cual ocurre cuando el tamaño de muestra es 2. Si T_p es una variable aleatoria con distribución t_p entonces:

$$\begin{aligned} \mathrm{E}[T_p] &= 0 \qquad \text{si } p > 1 \\ \mathrm{Var}(T_p) &= \frac{p}{p-2} \quad \text{si } p > 2 \end{aligned} \tag{5.3.2}$$

Definición 5.3.2 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población $\mathcal{N}(\mu_X, \sigma_X^2)$ y sea $Y_1, \cdots Y_m$ una muestra aleatoria de una población independiente $\mathcal{N}(\mu_Y, \sigma_Y^2)$. La variable aleatoria $F = (S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$ tiene *distribución $\mathcal{F}$ con $n-1$ y $m-1$ grados de libertad.* Equivalentemente, la variable aleatoria F tiene distribución $\mathcal{F}$ con p y q grados de libertad, si su función de densidad es:

$$f_F(x) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1+(p/q)x]^{(p+q)/2}}, \quad 0 < x < \infty \tag{5.3.3}$$

Teorema 5.3.2 Usando la técnica de la transformación es posible establecer los siguientes resultados:

a. Si $X \sim \mathcal{F}_{p,q}$ entonces $1/X \sim \mathcal{F}_{q,p}$.

b. Si $X \sim t_q$ entonces $X^2 \sim \mathcal{F}_{1,q}$.

c. Si $X \sim \mathcal{F}_{p,q}$ entonces $(p/q)X/(1+(p/q)X) \sim \mathcal{BE}(p/2, q/2)$.

5.4. Estadísticas de orden

Definición 5.4.1 Las *estadísticas de orden* de una muestra $X_1, \cdots, X_n$ son los valores puestos en orden ascendente y se denotan por $X_{(1)}, \cdots, X_{(n)}$. Las estadísticas de orden son variables aleatorias que satisfacen $X_{(1)} \leq \cdots \leq X_{(n)}$, tal que:

$$\begin{aligned} X_{(1)} &= \min_{1 \leq i \leq n} X_i \\ X_{(2)} &= \text{segundo valor más pequeño}\, X_i \\ &\vdots \\ X_{(n)} &= \max_{1 \leq i \leq n} X_i \end{aligned}$$

El *rango muestral*, $R = X_{(n)} - X_{(1)}$ es la diferencia entre la observación más grande y pequeña. La *mediana muestral*, denotada por M, es el número tal que aproximadamente la mitad de las observaciones son menores que M y la otra mitad es mayor. En términos de las estadísticas de orden, M se define por:

$$M = \begin{cases} X_{((n+1)/2)} & \text{si}\, n \text{ es impar} \\ (X_{(n/2)} + X_{(n/2+1)})/2 & \text{si}\, n \text{ es par} \end{cases} \tag{5.4.1}$$

Para todo número p entre 0 y 1, el percentil muestral $100p$ es la observación tal que aproximadamente np de las observaciones son menores que el mencionado percentil y $n\,(1-p)$ de las observaciones restantes son mayores.

El *rango intercuartil* es una medida de dispersión definida como la diferencia entre el tercer y el primer *cuartil* (percentil 75 y percentil 25 respectivamente).

Como las estadísticas de orden son funciones de la muestra, sus probabilidades pueden ser calculadas en términos de $X_1, \cdots, X_n$. En el caso discreto el cálculo de las probabilidades para las estadísticas de orden es básicamente un problema de conteo.

Teorema 5.4.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria de una distribución discreta con función de probabilidad $f_X(x_i) = p_i$ donde $x_1 < x_2 < \cdots$ son los posibles valores de X en orden ascendente. Se definen:

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= p_1 + p_2 + \cdots + p_i \end{aligned}$$

Sean $X_{(1)}, \cdots, X_{(n)}$ las estadísticas de orden de la muestra, entonces:

$$\Pr(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \tag{5.4.2}$$

y

$$\Pr(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right] \tag{5.4.3}$$

Si $X_1, \cdots, X_n$ es una muestra aleatoria obtenida de una población continua se tiene que la probabilidad que dos X_j sean iguales es cero con lo cual dejamos de preocuparnos por los *empates*.

Teorema 5.4.2 Sean $X_{(1)}, \cdots, X_{(n)}$ las estadísticas de orden de una muestra aleatoria de una población continua con función de distribución acumulada $F_X(x)$ y función de densidad $f_X(x)$. La función de densidad de $X_{(j)}$ es:

$$f_{X_{(j)}}(u) = \frac{n!}{(j-1)!(n-j)!} f_X(u) [F_X(u)]^{j-1} [1 - F_X(u)]^{n-j} \tag{5.4.4}$$

Ejemplo 5.4.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria extraída desde la distribución de *Rayleigh* con parámetro θ cuya función de densidad es:

$$f(x|\theta) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\}$$

para $x > 0$. Hallar la distribución de $X_{(1)}$. ¿Cuáles son sus parámetros?

La función de probabilidad conjunta de dos o más estadísticas de orden puede usarse para determinar la distribución de algunas de las estadísticas mencionadas al inicio de esta sección. El siguiente teorema permite obtener la función de densidad conjunta de dos estadísticas de orden.

Teorema 5.4.3 Sean $X_{(1)}, \cdots, X_{(n)}$ las estadísticas de orden de una muestra aleatoria $X_1, \cdots, X_n$ de una población continua con función de distribución acumulada $F_X(x)$ y función de densidad $f_X(x)$. La función de densidad conjunta de $X_{(i)}$ y $X_{(j)}$, $1 \leq i < j \leq n$, es:

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} & f_X(u) f_X(v) [F_X(u)]^{i-1} \\ & [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j} \end{aligned} \tag{5.4.5}$$

para $-\infty < u < v < \infty$.

Ejemplo 5.4.2 Sea $X_1, \cdots, X_n$ una muestra extraída desde la distribución $\mathcal{UC}(0,1)$. El rango muestral fue definido como $R = X_{(n)} - X_{(1)}$. El *rango medio* o *semirango* es una medida de localización como la mediana o media muestral, y se define por $S = (X_{(1)} + X_{(n)})/2$. Hallar la función de densidad conjunta de R y S. ¿Cuál es la distribución de R?

Es posible obtener la función de densidad conjunta de tres o más estadísticas de orden usando argumentos similares pero más complicados. La función de densidad conjunta de todas las estadísticas de orden está dada por:

$$f_{X_{(1)},\cdots,X_{(n)}}(x_1, \cdots, x_n) = \begin{cases} n! f_X(x_1) \cdots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{de otro modo} \end{cases}$$

Ejemplo 5.4.3 Sea $X_{(1)}, X_{(2)}, X_{(3)}$ las estadísticas de orden obtenidas en una muestra de tamaño tres obtenida desde la distribución $\mathcal{E}(\beta = 1)$. Se definen $Z_1 = 5X_{(1)}$, $Z_2 = 4(X_{(2)} - X_{(1)})$ y $Z_3 = 3(X_{(3)} - X_{(2)})$. ¿Son independientes Z_1, Z_2 y Z_3?

5.5. Conceptos de convergencia

5.5.1. Convergencia en probabilidad

Definición 5.5.1 Una secuencia de variables aleatorias $X_1, X_2, \cdots$ *converge en probabilidad* hacia la variable aleatoria X, si para todo $\epsilon > 0$:

$$\lim_{n\to\infty} \Pr\left(|X_n - X| \geq \epsilon\right) = 0 \quad \text{o} \quad \lim_{n\to\infty} \Pr\left(|X_n - X| < \epsilon\right) = 1$$

Las variables aleatorias $X_1, X_2, \cdots$ en la secuencia no son necesariamente independientes e identicamente distribuidas como en una muestra aleatoria.

Frecuentemente se tiene que la secuencia de variables aleatorias corresponde a medias muestrales y que la variable aleatoria límite es constante. El resultado más famoso es el siguiente.

Teorema 5.5.1 (Ley débil de los grandes números) Sean $X_1, X_2, \cdots$ variables aleatorias independientes e identicamente distribuidas con $\mathrm{E}[X_i] = \mu$ y $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Si se define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ entonces, para todo $\epsilon > 0$:

$$\lim_{n\to\infty} \Pr\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

es decir, $\overline{X}_n$ converge en probabilidad hacia μ.

Teorema 5.5.2 Si $X_1, X_2, \cdots$ converge en probabilidad hacia la variable aleatoria X y h es una función continua, entonces $h(X_1), h(X_2), \cdots$ converge en probabilidad hacia $h(X)$.

5.5.2. Convergencia casi segura

Definición 5.5.2 Una secuencia de variables aleatorias, $X_1, X_2, \cdots$ *converge de manera casi segura* hacia la variable aleatoria X, si para todo $\epsilon > 0$:

$$\Pr\left(\lim_{n\to\infty} |X_n - X| < \epsilon\right) = 1$$

Ejemplo 5.5.1 Sea el espacio muestral $S = [0, 1]$ con distribución de probabilidad uniforme. Se definen las variables aleatorias $X_n(s) = s + s^n$ y $X(s) = s$. Para todo $s \in [0, 1)$, $s^n \to 0$ conforme $n \to \infty$ y $X_n(s) \to X(s) = s$. Sin embargo $X_n(1) = 2$ para todo n, es decir $X_n(1)$ no converge a $X(1) = 1$. La convergencia ocurre en el conjunto $[0,1)$ y $\Pr([0,1)) = 1$, luego X_n converge de forma casi segura hacia X .

Teorema 5.5.3 (Ley fuerte de los grandes números) Sean $X_1, X_2, \cdots$ variables aleatorias independientes e identicamente distribuidas con $\mathrm{E}[X_i] = \mu$ y $\mathrm{Var}(X_i) = \sigma^2 < \infty$ y se define $\overline{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Entonces, para todo $\epsilon > 0$:

$$\Pr\left(\lim_{n\to\infty}\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

es decir, $\overline{X}_n$ converge de forma casi segura hacia μ.

5.5.3. Convergencia en distribución

Definición 5.5.3 Una secuencia de variables aleatorias, $X_1, X_2, \cdots$ *converge en distribución* a la variable aleatoria X si:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

en todos los puntos x donde $F_X(x)$ es continua.

Ejemplo 5.5.2 Sean $X_1, X_2, \cdots$ variables aleatorias independientes con distribución $\mathcal{UC}(0,1)$ y sea $X_{(n)} = \max_{1\leq i\leq n}\{X_i\}$. Conforme $n \to \infty$ se espera que $X_{(n)}$ se encuentre cerca de 1, entonces para todo $\epsilon > 0$:

$$\begin{aligned}
\Pr\left(\left|X_{(n)} - 1\right| \geq \epsilon\right) &= \Pr\left(X_{(n)} \geq 1+\epsilon\right) + \Pr\left(X_{(n)} \leq 1-\epsilon\right) \\
&= \Pr\left(X_{(n)} \leq 1-\epsilon\right) \\
&= \Pr\left(X_1 \leq 1-\epsilon, \cdots, X_n \leq 1-\epsilon\right) \\
&= (1-\epsilon)^n
\end{aligned}$$

luego $X_{(n)}$ converge en probabilidad hacia 1 ya que $(1-\epsilon)^n \to 0$ cuando $n \to \infty$. Además, si se toma $\epsilon = t/n$ se tiene:

$$\Pr\left(X_{(n)} \leq 1 - t/n\right) = (1-t/n)^n \to e^{-t}$$

lo cual es equivalente a:

$$\Pr\left(n(1 - X_{(n)}) \leq t\right) \to 1 - e^{-t}$$

es decir, la variable aleatoria $n(1 - X_{(n)})$ converge en distribución a la variable aleatoria $\mathcal{E}(1)$.

Teorema 5.5.4 (Teorema central del límite) Sean $X_1, X_2, \cdots$ variables aleatorias independientes e identicamente distribuidas cuyas funciones generatrices de momentos existen. Sea $\mathrm{E}[X_i] = \mu$ y $\mathrm{Var}\,(X_i) = \sigma^2 > 0$. Se define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ y sea $G_n(x)$ la función de distribución acumulada de $\sqrt{n}(\overline{X}_n - \mu)/\sigma$. Entonces para $-\infty < x < \infty$:

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

esto es, $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ tiene distribución límite normal estándar.

Prueba: Se define $Y_i = (X_i - \mu)/\sigma$ tal que:

$$W = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

Luego:

$$M_W(t) = \left[M_Y(t/\sqrt{n})\right]^n$$

Se expande $M_Y(t/\sqrt{n})$ en una serie de potencias de Taylor alrededor de 0. Entonces:

$$M_Y(t/\sqrt{n}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!} \text{ donde } M_Y^{(k)}(0) = \frac{d^k}{dt^k} M_Y(t)\bigg|_{t=0}$$

Se sabe que $M_Y^{(0)} = 1$, $M_Y^{(1)} = 0$ y $M_Y^{(2)} = 1$, ya que por construcción la media y varianza de Y son 0 y 1 respectivamente. Entonces:

$$M_Y(t/\sqrt{n}) \;=\; 1 + \frac{(t/\sqrt{n})^2}{2} + R_Y(t/\sqrt{n})$$

donde R_Y es el residuo en la expansión de Taylor. Una aplicación del teorema 7.4.1 de Taylor muestra que, para $t \neq 0$, se tiene:

$$\lim_{n\to\infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0 \text{ entonces } \lim_{n\to\infty} nR_Y(t/\sqrt{n}) = 0$$

Luego:

$$\begin{aligned}\lim_{n\to\infty} \left[M_Y(t/\sqrt{n})\right]^n &= \lim_{n\to\infty} \left[1 + \frac{(t/\sqrt{n})^2}{2} + R_Y(t/\sqrt{n})\right]^n \\ &= \lim_{n\to\infty} \left[1 + \frac{1}{n}\left(\frac{t^2}{2} + nR_Y(t/\sqrt{n})\right)\right]^n\end{aligned}$$

y usando el lema 3.4.1 se tiene:

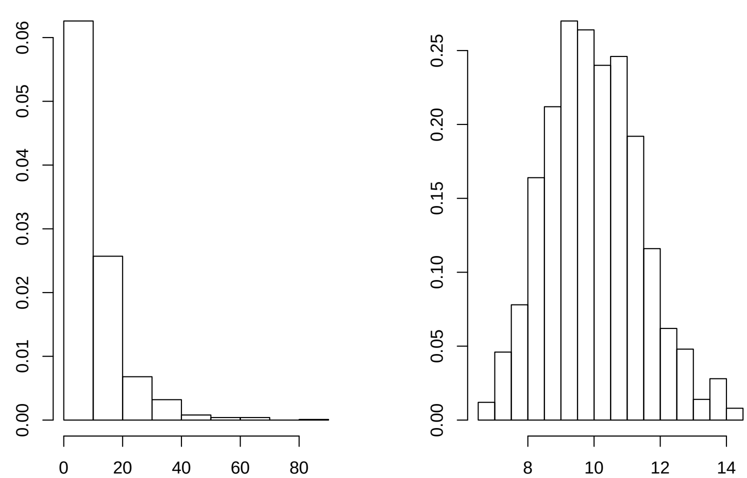
$$\lim_{n\to\infty}\left[M_Y(t/\sqrt{n})\right]^n = e^{t^2/2}$$

que es la función generatriz de momentos de la distribución normal estándar.

Usando R podemos obtener números aleatorios de la distribución exponencial y generar con ellos la distribución de la media muestral.

```
> x <- rexp(1000, rate=1/10)
> hist(x, freq=F, xlab="", ylab="", main="")
> medias <- rep(0, 1000)
> for (i in 1:1000){
+ medias[i] <- mean(rexp(50, rate=1/10))}
> hist(medias, freq=F, xlab="", ylab="", main="")
```

Figura 5.1: Teorema central del límite usando R



Ejemplo 5.5.3 El número de infracciones de estacionamiento aplicadas en una ciudad en un día cualquiera tiene distribución de Poisson con media igual a 50 infracciones. ¿Cuál es la probabilidad que el número total de estacionamiento cometidas en los próximos 40 días sea mayor a 2045? Suponga independencia de ser necesario.