Trabajo Práctico 2: Ley de los Grandes Números y Teorema Central del Límite

Juan Merhle, Juan Martin Goyeneche, Luciano Pozzoli

2025-08-26

Ejercicio 1

Sea $X \sim \mathrm{U}(0,18)$ una variable aleatoria con distribución uniforme entre 0 y 18.

Cuando queremos calcular el valor teorico, tanto de la esperanza como de la varianza en una variable continua, debemos hacer uso de la integral. En las variables aleatorias discretas tenemos un conjunto finito de valores de los cuales podemos hacer una suma ponderada, pero este no es el caso para las variables aleatorias continuas. Entonces, recurrimos al uso de la integral. En esta usamos los infinitos valores que puede tomar la variable aleatoria, representados con x y hacemos la ponderación multiplicandolos por su función de densidad de probabilidad $f_X(x)$.

Entonces, la esperanza se calcula con: $\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$

Pero, nuestra funcion de probabilidad es distinta de cero solo cuando x esta entre 0 y 18. Ademas, sabemos como esta definida la funcion de probabilidad de la uniforme. Entonces, la integral para esta funcion especifica nos quedaria de la siguiente forma:

$$\mathbb{E}(X) = \int_{a}^{b} x \cdot f_{X}(x) \, dx = \int_{0}^{18} x \cdot \frac{1}{18} \, dx$$

Resolvemos la integral:

$$\int_0^{18} x \, dx = \left. \frac{x^2}{2} \right|_0^{18} = \frac{18^2}{2} - \frac{0^2}{2} = \frac{324}{2} = 162$$

Entonces:

$$\mathbb{E}(X) = \frac{1}{18} \cdot 162 = 9$$

Por lo tanto, la esperanza teórica es:

$$\boxed{\mathbb{E}(X) = 9}$$

En cuanto a la varianza, sabemos que el valor teorico tiene la siguiente forma:

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Afortunadamente ya contamos con el valor de la esperanza y podemos elevarlo al cuadrado sin problema. Lo que si nos resta hacer es el calculo del primer termino de la resta. Para esto tenemos que volver a la integral con la cual resolvemos la esperanza pero utilizar el cuadrado de x en lugar de x. Entonces:

$$\mathbb{E}(X^2) = \int_0^{18} x^2 \cdot f_X(x) \, dx = \int_0^{18} x^2 \cdot \frac{1}{18} \, dx$$

Sacamos la constante $\frac{1}{18}$ fuera de la integral:

$$\mathbb{E}(X^2) = \frac{1}{18} \int_0^{18} x^2 \, dx$$

Calculamos la integral:

$$\int_0^{18} x^2 dx = \frac{x^3}{3} \Big|_0^{18} = \frac{18^3}{3} - \frac{0^3}{3} = \frac{5832}{3} = 1944$$

Entonces:

$$\mathbb{E}(X^2) = \frac{1}{18} \cdot 1944 = 108$$

$$\mathbb{E}(X^2) = 108$$

Teniendo ya todos los valores pertinentes mostramos que:

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 108 - 81 = 27$$

Por lo tanto, la varianza teórica es:

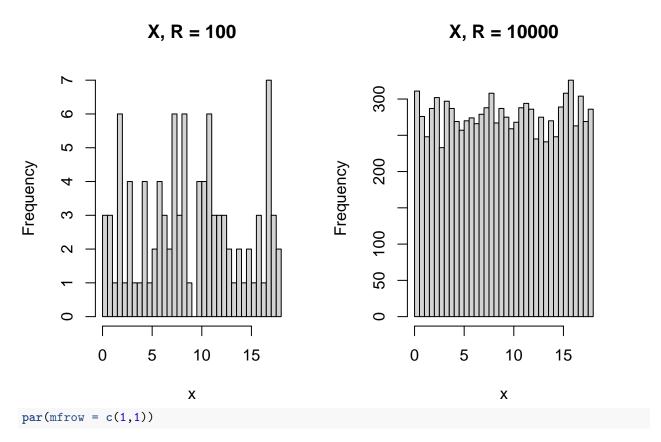
$$Var(X) = 27$$

Inciso B

```
set.seed(13)
X_dist <- function(R){
  return(runif(n= R, min= 0, max=18))
}</pre>
```

(1.c) Histogramas con R = 100 y $R = 10^4$

```
R100 <- X_dist(100)
R1e4 <- X_dist(10^4)
summary(R100)
     Min. 1st Qu. Median
                             Mean 3rd Qu.
  0.1968 5.2054 8.3281 8.9050 12.3849 17.9886
summary(R1e4)
##
       Min.
             1st Qu.
                         Median
                                     Mean
                                            3rd Qu.
   0.000525 4.480285
                       8.992434 9.023786 13.611984 17.997011
par(mfrow = c(1,2))
hist(R100, breaks = 30, main = "X, R = 100", xlab = "x")
hist(R1e4, breaks = 30, main = "X, R = 10000", xlab = "x")
```



La Ley de los Grandes Números nos dice que a medida que el tamaño de la muestra (n) aumente, la diferencia entre la media muestral y la media poblacional disminuye. En otras palabras, con un mayor n en la muestra deberiamos observar que la media muestral se acerca a la esperanza.

En nuestro caso la cantidad de valores que generan las muestras aleatorias estan definidas por el parametro R. Y queremos observar que a mayor R la media muestral se aproxima al valor teorico de la esperanza que calculamos anteriormente. Los resultados obtenidos fueron:

$$\overline{X}_{100} = 8.905031$$
 $\overline{X}_{10000} = 9.0237862$

Vemos que en estas observaciones se cumple que al aumentar n estamos aproximandonos al valor teorico de $\mathbb{E}(X) = 9$. La diferencia de los promedios muestrales con la esperanza teorica es casi 10 veces menor cuando aumentamos de R=100 a R=10000. Si vieramos los graficos y juzgaramos por intuicion, tambien coincidiriamos que R=10000 es mas parecido a una distribución Uniforme.

De la misma forma, esta tendencia tambien puede observarse en la varianza. Veamos si a medida que el tamaño de la muestra aumenta, se estabiliza la varianza muestral. Los datos que observamos de las varianzas muestrales fueron:

$$Var(X_{100}) = 26.1728695 \quad Var(X_{10000}) = 27.346619$$

En este caso la convergencia de las varianzas muestrales hacia su valor teorico de 27 es muy evidente. A medida que aumentamos el tamaño de la muestra la dispersion se reduce y se asemeja a la varianza de la distribución.

Cabe destacar que estas muestras de datos se realizan con una sola iteración para cada tamaño, y los datos generados son aleatorios.

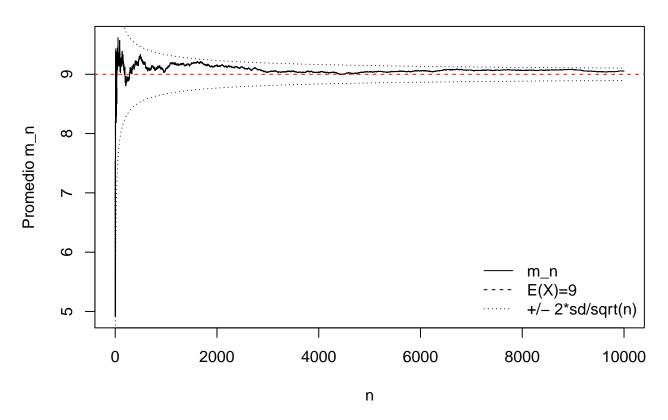
(1.d) Promedios acumulados y Ley de los Grandes Números

Construimos los promedios acumulados:

$$m_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad n = 1, 2, \dots, 10^4$$

y los comparamos con el valor teórico E(X)=9.

Promedios acumulados (X ~ U(0,18))



(2.a) Esperanza de $Y \sim \text{Exp}(\lambda)$ con $\lambda = \frac{1}{9}$

La densidad es $f_Y(y) = \lambda e^{-\lambda y}$ para $y \ge 0$. Entonces,

$$\mathbb{E}(Y) = \int_0^\infty y \, \lambda e^{-\lambda y} \, dy$$

Para $\lambda = \frac{1}{9}$,

$$\mathbb{E}(Y) = \frac{1}{9} \int_0^\infty y \, e^{-\frac{1}{9}y} \, dy$$

Integración por partes. Tomamos $u = y \Rightarrow du = dy$ y $dv = e^{-\frac{1}{9}y}dy \Rightarrow v = -9 e^{-\frac{1}{9}y}$.

$$\int_0^\infty y \, e^{-\frac{1}{9}y} \, dy = \left[-9y \, e^{-\frac{1}{9}y} \right]_0^\infty + \int_0^\infty 9 \, e^{-\frac{1}{9}y} \, dy$$

El primer término es 0 (porque $ye^{-\frac{1}{9}y} \to 0$ al $+\infty$ y en 0 vale 0). Para la integral restante:

$$\int_0^\infty 9 e^{-\frac{1}{9}y} dy = 9 \left[-9 e^{-\frac{1}{9}y} \right]_0^\infty = 9 \left(0 - (-9) \right) = 81$$

Por lo tanto,

$$\mathbb{E}(Y) = \frac{1}{9} \cdot 81 = 9.$$

$$\boxed{\mathbb{E}(Y) = 9}$$

Observación: en general, si $Y \sim \text{Exp}(\lambda)$, entonces $\mathbb{E}(Y) = \frac{1}{\lambda}$ y $\text{Var}(Y) = \frac{1}{\lambda^2}$. Para $\lambda = \frac{1}{9}$, Var(Y) = 81.

Teniendo el valor teorico de la esperanza, recordamos que la formula de la varianza es:

$$Var(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$$

El calculo anterior nos sirve dado que nos resuelve uno de los terminos ya que hacer el cuadrado de la esperanza es una cuenta fácil, pero si necesitamos volver a hacer la integral con el cuadrado de y para poder resolver la primera parte de la resta.

La esperanza de Y^2 se calcula como:

$$\mathbb{E}(Y^{2}) = \int_{0}^{\infty} y^{2} \cdot \frac{1}{9} e^{-\frac{1}{9}y} \, dy$$

Sacamos la constante fuera de la integral:

$$\mathbb{E}(Y^2) = \frac{1}{9} \int_0^\infty y^2 \cdot e^{-\frac{1}{9}y} \, dy$$

Ahora resolvemos la integral por partes dos veces.

Primera integración por partes: Sea

$$u = v^2$$
 \Rightarrow $du = 2u \, du \, dv = e^{-\frac{1}{9}y} \, du$ \Rightarrow $v = -9e^{-\frac{1}{9}y}$

Aplicamos:

$$\int y^2 \cdot e^{-\frac{1}{9}y} dy = -9y^2 e^{-\frac{1}{9}y} + \int 18y \cdot e^{-\frac{1}{9}y} dy$$

Ahora resolvemos la segunda integral:

$$\int 18y \cdot e^{-\frac{1}{9}y} dy$$

Segunda integración por partes:

$$u = y \quad \Rightarrow \quad du = dydv = e^{-\frac{1}{9}y}dy \quad \Rightarrow \quad v = -9e^{-\frac{1}{9}y}$$

Entonces:

$$\int 18y \cdot e^{-\frac{1}{9}y} dy = 18 \left(-9y e^{-\frac{1}{9}y} + \int 9e^{-\frac{1}{9}y} dy \right)$$

Resolvemos la última integral:

$$\int 9e^{-\frac{1}{9}y}dy = -81e^{-\frac{1}{9}y}$$

Sumamos todo:

$$\int y^2 e^{-\frac{1}{9}y} dy = -9y^2 e^{-\frac{1}{9}y} + 18\left(-9y e^{-\frac{1}{9}y} - 81e^{-\frac{1}{9}y}\right)$$

Multiplicamos y reordenamos:

$$= -9y^2e^{-\frac{1}{9}y} - 162ye^{-\frac{1}{9}y} - 1458e^{-\frac{1}{9}y}$$

Evaluamos en los límites 0 a ∞ . Todos los términos tienden a 0 en $y \to \infty$, y en y = 0 queda:

$$-0 - 0 - 1458 \cdot 1 = -1458$$

Entonces la integral completa da 1458, y por lo tanto:

$$\mathbb{E}(Y^2) = \frac{1}{9} \cdot 1458 = 162$$

Resultado final:

$$\mathbb{E}(Y^2) = 162$$

Teniendo ya todos los terminos necesarios, procedemos a hacer el calculo teorico de la varianza.

$$Var(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = 162 - 81 = 81$$

$$\boxed{Var(Y) = 81}$$

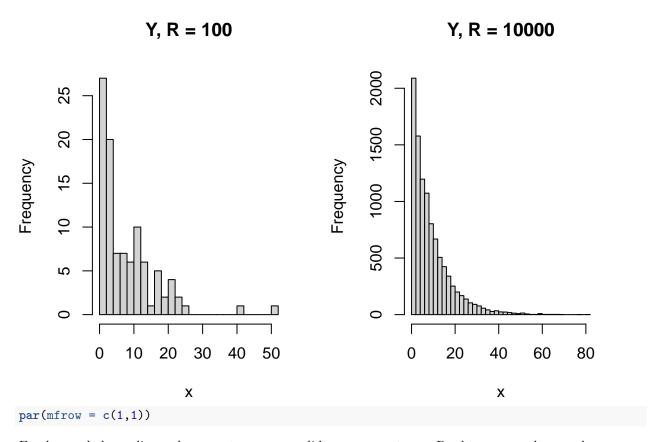
(2.b) Función generadora de muestras

Definimos la función generadora de muestras Y dist(R) que devuelve R realizaciones de Y. Usamos nuestra semilla de grupo (13) para asegurar reproducibilidad.

```
set.seed(13)
Y_dist <- function(R) {</pre>
  rexp(n = R, rate = 1/9)
```

(2.c) Histogramas con R = 100 y $R = 10^4$

```
YR100 <- Y_dist(100)
YR1e4 <- Y_dist(10^4)
summary(YR100)
      Min. 1st Qu. Median
                             Mean 3rd Qu.
   0.1331 1.9403 4.8549 8.0517 11.8086 51.6450
summary(YR1e4)
##
       Min.
            1st Qu.
                      Median
                                 Mean 3rd Qu.
    0.00018 2.46954
                     6.16815 8.94091 12.32194 81.27150
par(mfrow = c(1,2))
hist(YR100, breaks = 30, main = "Y, R = 100", xlab = "x")
hist(YR1e4, breaks = 30, main = "Y, R = 10000", xlab = "x")
```



En el caso de la media, podemos notar que a medida que aumentamos R, obtenemos valores cada vez mas aproximados a la esperanza teorica de valor 9. Esta observacion se sustenta con lo que dice la Ley de lo Grandes Numeros. Los valores obtenidos de las medias muestrales son:

$$\overline{Y}_{100} = 8.0516985 \quad \overline{Y}_{10000} = 8.9409069$$

En este caso podemos ver que la convergencia se cumple y queda en evidencia que la diferencia entre E(Y)=9 y las medias muestrales disminuye a medida que aumenta el tamaño de la muestra.

Lo mismo podemos observar con los valores de la varianza. Al aumentar el n notamos un acercamiento muy grande de la varianza muestral al valor de la varianza poblacional. Recordando que Var(Y) = 81:

$$Var(Y_{100}) = 72.9703857 \quad Var(Y_{10000}) = 82.0615821$$

Punto 3: Promedios de variables aleatorias

Sea $X \sim U(0, 18)$. Definimos:

$$X_{15} = \frac{1}{15} \sum_{i=1}^{15} X_i$$

donde los X_i son 15 variables independientes con la misma distribución que X.

(3.a) Esperanza y varianza de X_{15}

Usamos propiedades lineales:

$$E(X_{15}) = \frac{1}{15} \sum_{i=1}^{15} E(X_i) = \frac{1}{15} \cdot 15 \cdot E(X) = E(X) = 9$$

$$Var(X_{15}) = \frac{1}{15^2} \sum_{i=1}^{15} Var(X_i) = \frac{1}{225} \cdot 15 \cdot Var(X) = \frac{Var(X)}{15}$$

Como Var(X) = 27:

$$Var(X_{15}) = \frac{27}{15} = 1.8$$

Comentarios:

- La esperanza sigue siendo 9.
- La varianza disminuye al dividir por n, pasa de 27 a 1.8.
- En general, promediar reduce la dispersión de los valores.

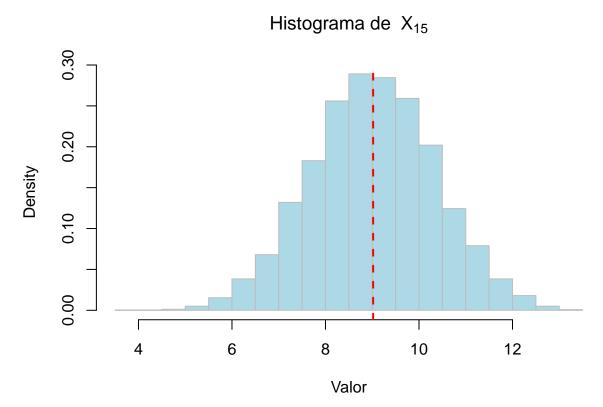
(3.b) Función Xn_dist(n, R)

Definimos una función que genere R realizaciones del promedio de n variables X:

```
Xn_dist <- function(n, R) {
  replicate(R, mean(runif(n, min = 0, max = 18)))
}</pre>
```

(3.c) Histograma de X15

```
R <- 10^4
X15 <- Xn_dist(15, R)
hist(X15, breaks = 30, probability = TRUE,
    main = expression("Histograma de " ~ X[15]),
    xlab = "Valor", col = "lightblue", border = "gray")
abline(v = mean(X15), col = "red", lwd = 2, lty = 2)</pre>
```



Comentarios:

- -El histograma de X15 ya no es uniforme como el de X. Su forma es más acampanada y concentrada alrededor de la media (9).
- -Esto se debe a que al promediar 15 valores independientes, se reduce la variabilidad y la distribución se aproxima a una normal.

(4.a) Histogramas de X, X_{40}, Y, Y_{40} con curvas teóricas.

En este inciso aplicamos el Teorema Central del Límite a $X \sim U(0, 18)$ y $Y \sim \text{Exp}(1/9)$. Primero mostramos la forma de las distribuciones originales y luego la de los promedios de 40 observaciones. Superponemos las curvas teóricas para ver cómo aparecen las aproximaciones normales.

```
set.seed(13)

R_big <- 10^6
n <- 40

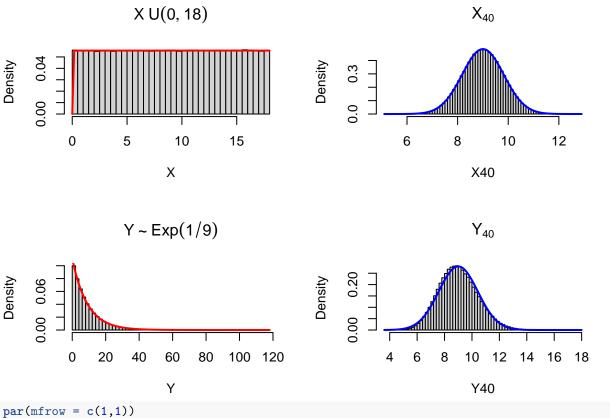
#Usamos funciones definidas antes
X <- X_dist(R_big)
Y <- Y_dist(R_big)

#Promedios de n=40

X40 <- replicate(R_big, mean(X_dist(n)))
Y40 <- replicate(R_big, mean(Y_dist(n)))

#Parámetros teóricos
mu_X <- 9
var_X <- 27
mu_Y <- 9</pre>
```

```
var_Y <- 81
#Gráficos
par(mfrow = c(2,2))
hist(X, breaks = 60, freq = FALSE, main = X ~ U(0,18))
curve(dunif(x,0,18), add = TRUE, col="red", lwd=2)
hist(X40, breaks = 60, freq = FALSE, main = expression(X[40]))
curve(dnorm(x, mean=mu_X, sd=sqrt(var_X/n)), add = TRUE, col="blue", lwd=2)
hist(Y, breaks = 60, freq = FALSE, main = expression(Y %~% Exp(1/9)))
curve(dexp(x, rate=1/9), add = TRUE, col="red", lwd=2)
hist(Y40, breaks = 60, freq = FALSE, main = expression(Y[40]))
curve(dnorm(x, mean=mu_Y, sd=sqrt(var_Y/n)), add = TRUE, col="blue", lwd=2)</pre>
```



4(b) — TCL con X_n : efecto de n y R

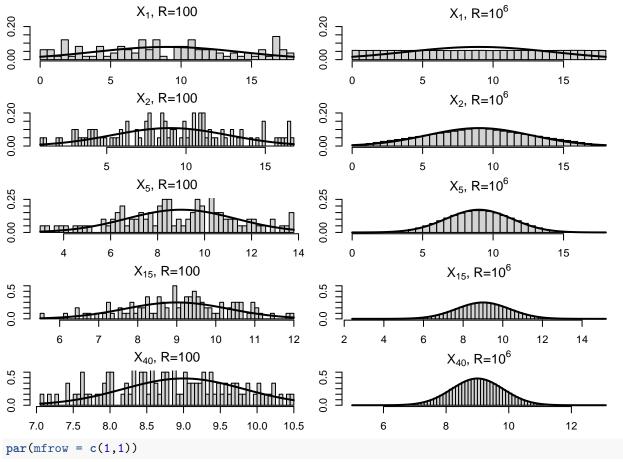
Ahora estudiamos cómo varía la distribución de X_n (el promedio de n uniformes) cuando cambiamos n y el número de simulaciones R.

El panel 5×2 nos deja comparar la forma de los histogramas y el ajuste a la normal según el tamaño de la muestra y el tamaño de la simulación.

```
set.seed(13)

Xn_dist <- function(n, R) replicate(R, mean(X_dist(n)))</pre>
```

```
par(
  mfrow = c(5, 2),
  mar = c(2, 2, 2, 1))
n_{vals} \leftarrow c(1, 2, 5, 15, 40)
R_{vals} \leftarrow c(1e2, 1e6)
Xn <- replicate(10^2, mean(X_dist(1)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[1] * ", R=100"), xlab = "", ylab = "",
     ylim = c(0, 0.2))
curve(dnorm(x, mean = 9, sd = sqrt(27/1)), add = TRUE, lwd = 2)
Xn <- replicate(10^6, mean(X_dist(1)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[1] * ", R=10"^6), xlab = "", ylab = "",
     ylim = c(0, 0.2))
curve(dnorm(x, mean = 9, sd = sqrt(27/1)), add = TRUE, lwd = 2)
Xn <- replicate(10^2, mean(X_dist(2)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[2] * ", R=100"), xlab = "", ylab = "",
     ylim = c(0, 0.2))
curve(dnorm(x, mean = 9, sd = sqrt(27/2)), add = TRUE, lwd = 2)
Xn <- replicate(10^6, mean(X_dist(2)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[2] * ", R=10"^6), xlab = "", ylab = "",
     ylim = c(0, 0.2)
curve(dnorm(x, mean = 9, sd = sqrt(27/2)), add = TRUE, 1wd = 2)
Xn <- replicate(10^2, mean(X_dist(5)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[5] * ", R=100"), xlab = "", ylab = "",
     ylim = c(0, 0.25))
curve(dnorm(x, mean = 9, sd = sqrt(27/5)), add = TRUE, lwd = 2)
Xn <- replicate(10^6, mean(X_dist(5)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[5] * ", R=10"^6), xlab = "", ylab = "",
     ylim = c(0, 0.25))
curve(dnorm(x, mean = 9, sd = sqrt(27/5)), add = TRUE, 1wd = 2)
Xn <- replicate(10^2, mean(X_dist(15)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[15] * ", R=100"), xlab = "", ylab = "",
     ylim = c(0, 0.6))
curve(dnorm(x, mean = 9, sd = sqrt(27/15)), add = TRUE, lwd = 2)
Xn <- replicate(10^6, mean(X_dist(15)))</pre>
hist(Xn, breaks = 60, probability = TRUE,
     main = expression(X[15] * ", R=10"^6), xlab = "", ylab = "",
```



Comentarios (Xn):

- n = 1: uniforme en [0,18], media = 9. Con R grande se ve perfectamente rectangular.
- n = 2: distribución triangular, ya se "redondea" en el centro.
- n = 5: comienza a verse simétrica y con forma de campana; la normal se ajusta mejor.
- n = 15: claramente normal, muy concentrada alrededor de 9; varianza mucho menor.

- n = 40: se ajusta muy bien a la normal teórica $\mathcal{N}(9, 27/40)$, extremos casi nulos.
- se puede concluir: como dice el TCL, a medida que la muestra sea lo suficientemente grande, la media de la muestra tiene distribucion aproximadamente normal

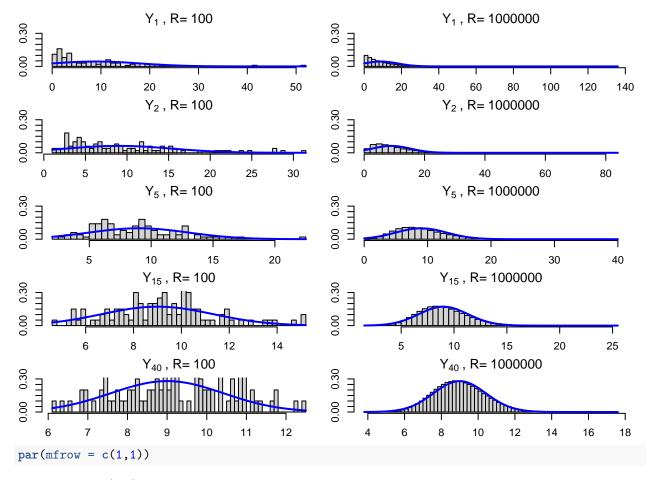
Efecto de R: más simulaciones \rightarrow histograma más suave.

4(c) — TCL con Y_n : efecto de n y R

Finalmente repetimos el análisis para Y_n (promedios de exponenciales).

Así podemos discutir cómo la convergencia hacia la normal es más lenta en una distribución asimétrica: hacen falta valores de n más grandes para que la forma se parezca a una campana.

```
set.seed(13)
\# función general: devuelve R realizaciones del promedio de n muestras de Y
Yn_dist <- function(n, R) {</pre>
 replicate(R, mean(Y_dist(n)))
n_{vals} < c(1, 2, 5, 15, 40)
R_{vals} \leftarrow c(10^2, 10^6)
par(mfrow = c(5,2), mar = c(2, 2, 2, 1))
for (n in n_vals) {
 for (R in R_vals) {
    Yn <- Yn_dist(n, R)
    hist(Yn, breaks = 50, freq = FALSE,
         main = bquote(Y[.(n)] ~ ", R=" ~ .(format(R, scientific=FALSE))),
         xlab = "", ylab = "",
         ylim = c(0, 0.3))
    #Superponemos la normal del TCL
    mu <- 9
    sd <- sqrt(81 / n)
                        #varianza de Y es 81
    curve(dnorm(x, mean = mu, sd = sd), add = TRUE, col="blue", lwd=2)
  }
}
```



Comentarios (Yn):

- $\mathbf{n} = \mathbf{1}$: distribución exponencial pura, muy asimétrica a la derecha, media = 9.
- n = 2: sigue siendo claramente asimétrica, pero empieza a concentrarse alrededor de 9.
- n = 5: la media se mantiene en 9, la varianza ya es menor (81/5), pero todavía hay cola a la derecha.
- n = 15: la forma empieza a parecer normal; la asimetría se atenúa bastante.
- $\mathbf{n} = 40$: se observa un buen ajuste a la normal $\mathcal{N}(9, 81/40)$, la cola derecha es mucho menos notoria.

Comparación con Xn:

- En ambos casos la media se mantiene en 9 y la varianza decrece como 1/n.
- La diferencia está en la **velocidad de convergencia**: para la uniforme, con n = 5 ya se veía muy normal; en la exponencial, recién con n = 15 o más se aproxima bien a la campana.

Conclusion Ejercicio 4 Tanto para la uniforme como para la exponencial, los promedios mantienen la misma media que la distribución original y su varianza disminuye proporcionalmente a 1/n.

En el caso de X_n (uniforme), la convergencia hacia la normal es rápida: con valores de n moderados ya se observa un buen ajuste.

En el caso de Y_n (exponencial), la asimetría inicial hace que la convergencia sea más lenta, necesitando n mayores para que la forma se aproxime bien a una normal.

En conjunto, los resultados ilustran con claridad cómo el TCL opera de manera general, aunque la velocidad del ajuste depende de la distribución original.