

Trabajo Práctico 2

Olivia Luchetti y Martina Mariño

```
set.seed(0264)
```

Ejercicio 1 - Distribución Uniforme (0,18)

Ejercicio 1.a: Función para Generar Realizaciones

Creamos una función que devuelva un vector con R realizaciones de una variable aleatoria $X \sim U(0, 18)$.

```
X_dist <- function(R) {  
  return(runif(n = R, min = 0, max = 18))  
}
```

Ejercicio 1.b: Cálculo de Media y Varianza Muestral

Calculamos la media y la varianza muestral para diferentes tamaños de muestra R.

```
comparacion <- data.frame(  
  R = c(2, 30, 100, 10^4),  
  Media_muestral = c(  
    mean(X_dist(2)),  
    mean(X_dist(30)),  
    mean(X_dist(100)),  
    mean(X_dist(10^4))  
  ),  
  Varianza_muestral = c(  
    var(X_dist(2)),  
    var(X_dist(30)),  
    var(X_dist(100)),  
    var(X_dist(10^4))  
  )  
)  
  
print(comparacion)
```

##	R	Media_muestral	Varianza_muestral
## 1	2	12.783315	43.82488
## 2	30	8.589811	25.11464
## 3	100	9.191024	26.50414
## 4	10000	8.937333	27.00084

Ejercicio 1.c: Valores Teóricos y Comparación

El valor teórico de la esperanza $E(X)$ y la varianza $V(X)$ para una distribución uniforme $U(0, 18)$ son:

- $E(X) = (a + b)/2 = (0 + 18)/2 = 9$
- $V(X) = (b - a)^2/12 = (18)^2/12 = 27$

```

E_X = 9
Var_X = 27

R_values <- c(2, 30, 100, 10000)

X <- numeric(4)
s <- numeric(4)
abs_X_E_X <- numeric(4)
abs_s2_VarX <- numeric(4)

for (i in 1:4) {
  R <- R_values[i]
  sample_X <- runif(R, 0, 18)

  X[i] <- mean(sample_X)
  s[i] <- var(sample_X)
  abs_X_E_X[i] <- abs(X[i] - E_X)
  abs_s2_VarX[i] <- abs(s[i] - Var_X)
}

R_values <- c(R_values, "")
X <- c(X, E_X)
s <- c(s, Var_X)
abs_X_E_X <- c(abs_X_E_X, 0)
abs_s2_VarX <- c(abs_s2_VarX, 0)

tabla <- data.frame(
  R = R_values,
  X_barra = round(X, 3),
  s_2 = round(s, 3),
  abs_X_barra_E_X = round(abs_X_E_X, 3),
  abs_s_2_Var_X = round(abs_s2_VarX, 3)
)

print(tabla)

```

Comparación de Valores Teóricos y Muestrales:

##	R	X_barra	s_2	abs_X_barra_E_X	abs_s_2_Var_X
## 1	2	12.310	34.557	3.310	7.557
## 2	30	8.854	20.261	0.146	6.739
## 3	100	8.357	24.100	0.643	2.900
## 4	10000	8.968	27.320	0.032	0.320
## 5		9.000	27.000	0.000	0.000

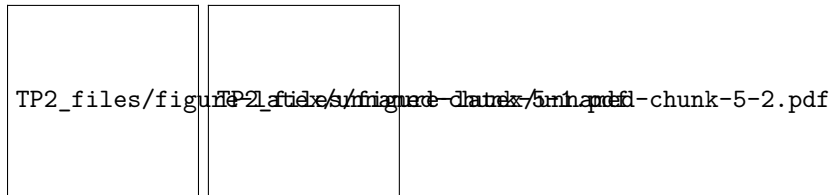
Conclusión: A medida que el tamaño de la muestra R aumenta, la media y varianza muestrales se aproximan progresivamente a los valores teóricos. Esto es consistente con la Ley de los Grandes Números, que establece que, a medida que aumenta el tamaño de la muestra, las estimaciones muestrales tienden a converger hacia los parámetros poblacionales. La Ley de los Grandes Números asegura que las estimaciones obtenidas de una muestra suficientemente grande serán cercanas a los verdaderos valores de la población de donde provienen. En el contexto de la tabla, se puede observar cómo, para tamaños de muestra pequeños (como $R = 2$), la diferencia entre las estimaciones muestrales y los valores teóricos es considerable. Sin embargo, al aumentar R , estas diferencias disminuyen, lo que indica que la media y la varianza muestrales

están cada vez más alineadas con sus valores teóricos. Este comportamiento refleja cómo el aumento en el tamaño de la muestra no solo mejora la precisión de las estimaciones, sino que también reduce la variabilidad inherente en las muestras.

Ejercicio 1.d: Histogramas de X

Creamos histogramas para $R = 100$ y $R = 10^4$ utilizando 30 bins.

Distribución esperada: Con una distribución uniforme $U(0,18)$, esperamos que el histograma muestre una distribución bastante uniforme, especialmente para tamaños de muestra grandes. Con $R=100$, el histograma debería comenzar a parecerse a la distribución uniforme. Para $R = 10^4$, la distribución debería ser aún más uniforme y parecerse a la distribución teórica.



Discusión: Al aumentar el tamaño de la muestra, el histograma se acerca más a la forma esperada de la distribución uniforme. Para muestras pequeñas, el histograma puede mostrar variabilidad significativa debido a la menor cantidad de datos. A medida que R aumenta, la variabilidad disminuye y el histograma se asemeja más a la distribución uniforme teórica.

Ejercicio 2 - Promedio de Quince Valores Independientes

Ejercicio 2.a: Función para Generar Realizaciones de Y

Creamos una función que devuelva un vector con R realizaciones de Y , donde cada Y es el promedio de 15 valores independientes de X .

```
Y_dist <- function(R) {
  # Inicializa un vector numérico de longitud R, donde se almacenarán los promedios
  # calculados.
  Y <- numeric(R)
  # Bucle que se repite R veces, una por cada realización de Y.
  for(i in 1:R) {
    # Genera 15 valores independientes de X, donde X sigue una distribución
    # uniforme entre 0 y 18.
    X_15 <- runif(n = 15, min = 0, max = 18)
    # Calcula el promedio de los 15 valores generados de X y lo almacena en la
    # i-ésima posición del vector Y.
    Y[i] <- mean(X_15)
  }
  # Devuelve el vector Y, que contiene R promedios de 15 valores independientes de
  # X.
  return(Y)
}
```

Ejercicio 2.b: Cálculo de Media y Varianza Muestral

Calculamos la media y varianza muestral para diferentes tamaños de muestra R .

```
comparacion <- data.frame(
  R = c(2, 30, 100, 10^4),
```

```

    Media_muestral = c(
      mean(Y_dist(2)),
      mean(Y_dist(30)),
      mean(Y_dist(100)),
      mean(Y_dist(10^4))
    ),
    Varianza_muestral = c(
      var(Y_dist(2)),
      var(Y_dist(30)),
      var(Y_dist(100)),
      var(Y_dist(10^4))
    )
  )
)

print(comparacion)

```

```

##           R Media_muestral Varianza_muestral
## 1         2          9.457222         0.001930543
## 2        30          8.852968         1.886629933
## 3       100          9.268562         1.521618640
## 4    10000          8.998140         1.850213107

```

Ejercicio 2.c: Comparación de Medias Empíricas

Dado que $E(Y) = E(X)$, el valor teórico de la esperanza $E(Y)$ y la varianza $V(Y)$ son:

- $E(Y) = E(X) = 9$
- $V(Y) = [(b - a)^2 / 12] / n = V(X) / n = V(X) / 15 = 27 / 15 = 1.8$

```

E_Y_teorico <- 9
V_Y_teorico <- 27 / 15

comparacion <- data.frame(
  R = c(2, 30, 100, 10^4),
  Media_muestral = c(
    mean(X_dist(2)),
    mean(X_dist(30)),
    mean(X_dist(100)),
    mean(X_dist(10^4))
  ),
  Varianza_muestral = c(
    var(X_dist(2)),
    var(X_dist(30)),
    var(X_dist(100)),
    var(X_dist(10^4))
  ),
  Media_teorica = E_Y_teorico,
  Varianza_teorica = V_Y_teorico
)

print(comparacion)

```

Media Esperada y Varianza para Y:

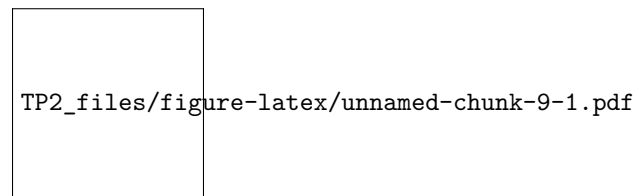
##	R	Media_muestral	Varianza_muestral	Media_teorica	Varianza_teorica
## 1	2	10.941158	4.175705	9	1.8
## 2	30	7.774725	21.974596	9	1.8
## 3	100	8.641142	27.634614	9	1.8
## 4	10000	9.011387	27.073208	9	1.8

Discusión: Las medias empíricas de Y deberían acercarse más al valor esperado $E(Y) = 9$ comparado con las medias empíricas de X, ya que Y es un promedio de varias observaciones de X. La Ley de los Grandes Números garantiza que a medida que el tamaño de la muestra aumenta, la media empírica se aproxima más al valor esperado de la distribución. Esta ley asegura que el promedio de una gran cantidad de observaciones será más preciso en reflejar la media teórica. La varianza de Y debería ser más pequeña que la varianza de X debido al promedio de múltiples observaciones.

Ejercicio 2.d: Histogramas de Y

Creamos histogramas para $R = 100$ y $R = 10^4$ utilizando 30 bins.

Distribución esperada: Esperaría ver una distribución normal con media $E(Y) = 9$ y varianza $V(Y) = 27/15$, especialmente para tamaños grandes de muestra. Con $R=100$, el histograma debería comenzar a parecerse a una distribución normal, y para $R = 10^4$, la aproximación debería ser aún mejor.



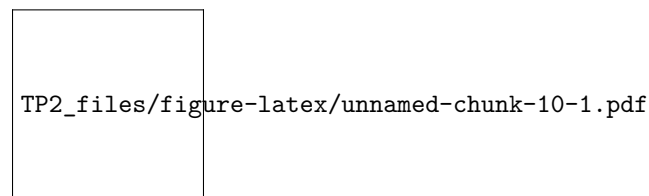
Discusión:

- Para $R=100$: Es posible que la distribución de Y aún no se parezca completamente a una distribución normal, pero debería comenzar a mostrar una forma más aproximada a una campana en comparación con una distribución uniforme.
- Para $R=10^4$: Con un tamaño de muestra grande, la distribución de Y debería aproximarse más a una normal con media 9 y varianza $27/15$. Esto es consistente con el Teorema Central del Límite, que establece que la media de muestras grandes de una variable aleatoria se distribuye aproximadamente normalmente, independientemente de la distribución original de la variable.

Efecto de variar R: A medida que aumenta R, la distribución de Y se aproxima más a una normal con la media y varianza teóricas. Con un tamaño de muestra pequeño, el histograma puede mostrar más variabilidad y desviarse más de la distribución normal teórica.

Ejercicio 3 - Teorema Central del Límite

Ejercicio 3.a: Histogramas de X



Ejercicio 3.b: Histogramas

```
X_dist <- function(R, N){
  Y <- numeric(R)
  for (i in 1:R){
    X_15 <- runif(N, 0, 18)
    mX_15 <- mean(X_15)
    Y[i] <- mX_15
  }
  return(Y)
}

color_vec <- c("#689C6E", "#73AD9C", "#6F8AA6", "#D2691E")

R_vec <- c(10^2, 10^6)
N_vec <- c(1, 2, 5, 15)

par(mfrow = c(4, 2),
    mar = c(1, 1, 1, 1),
    oma = c(1, 1, 2.5, 1),
    cex.main = 1.0,
    cex.lab = 1.2,
    cex.axis = 1.1,
    las = 1,
    mgp = c(3, 0.5, 0))

for (irow in 1:4) {
  color <- color_vec[irow]
  N <- N_vec[irow]
  for (icol in 1:2) {
    R <- R_vec[icol]
    X <- X_dist(R=R, N = N)
    hist(X, col = color,
         main = paste("R =", R), freq = FALSE)
  }
}
```

TP2_files/figure-latex/unnamed-chunk-11-1.pdf

Discusión:

- Cambios al variar n : a medida que n aumenta, la forma de la distribución de las muestras se va acercando más a una distribución normal. Esto es consistente con el Teorema Central del Límite, que establece que la suma de un número suficiente de variables aleatorias independientes e idénticamente distribuidas tenderá hacia una distribución normal, independientemente de la distribución original de las variables.
 - Para $n = 1$, la distribución sigue siendo uniforme, ya que cada muestra es una realización directa de X .
 - A medida que aumentas n , la dispersión de los datos disminuye a medida que n aumenta. Esto

se debe a que la media de un mayor número de muestras es menos variable que la media de un número menor de muestras.

- Cambios al variar R :
 - Para $R = 10^2$, los histogramas muestran más variabilidad debido a que se toman menos muestras, lo que introduce más ruido en la estimación de la densidad.
 - Para $R = 10^6$, la distribución estimada se aproxima mejor a la forma teórica de la distribución, ya que un mayor número de repeticiones reduce la variabilidad en la estimación.

En resumen:

- Al variar n : Se observa una transición de una distribución uniforme hacia una distribución normal al aumentar n . Este cambio ilustra la aplicación del Teorema Central del Límite, donde la suma (o promedio) de un número creciente de variables aleatorias idénticamente distribuidas tiende a una distribución normal.
- Al variar R : La forma general de los histogramas no cambia significativamente con R ; lo que cambia es la precisión con la que esa forma se estima. Con un R mayor, se obtiene una mejor aproximación de la forma verdadera de la distribución subyacente. R no cambia la naturaleza de la distribución, solo mejora la precisión con la que se representa.