Trabajo Práctico 1 en R

Ley de los Grandes Números y Teorema Central del Límite

Las actividades indicadas con el ícono son para resolver y entregar. En esta guía sólo deberán entregar la Actividad 1. Se deberá entregar un archivo que incluya el código con la resolución y las respuestas a las preguntas. Para ello, puede entregar un archivo R Markdown (o similar) o, en su defecto, un archivo de código prolijamente desarrollado y comentado. Les dejamos también una Actividad 2 que es optativa y NO se debe entregar.

Primera parte: Ley de los Grandes Números

Introducción

Antes de empezar con las simulaciones, recordemos que la Ley de los Grandes Números establece que

$$\overline{W}_n \xrightarrow{p} \mathbb{E}(W) \tag{1}$$

donde $(W_i)_{i\geq 1}$ son variables aleatorias i.i.d y $\mathbb{E}(W) = \mathbb{E}(W_1)$.

En particular, vimos que si las variables tienen distribución Bernoulli, obtenemos que la frecuencia relativa de éxitos converge a su probabilidad. Más generalmente, si $(X_i)_{i\geq 1}$ son variables i.i.d. y A es un conjunto de números reales, podemos definir

$$Y_i = \mathbb{I}(X_i \in A) = \begin{cases} 1 & \text{si } X_i \in A \\ 0 & \text{en caso contrario.} \end{cases}$$

Es decir, Y_i vale 1 si X_i pertenece al conjunto A y 0 en caso contrario. En tal caso, las variables Y_i tienen distribución Bernoulli, con $P(Y_i = 1) = P(X_i \in A)$, y por consiguiente, $\mathbb{E}(Y) = P(X \in A)$. Si invocamos la Ley de los Grandes Números dada en (1), pero utilizando Y_i en lugar de W_i, tenemos que

$$\overline{Y}_n \xrightarrow{p} \mathbb{E}(Y),$$

que es lo mismo que decir que la frecuencia relativa converge a la probabilidad pues

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \in A) = \frac{\#\{1 \leq i \leq n : X_i \in A\}}{n}$$

y

$$E(Y) = P(X \in A),$$

por lo tanto,

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i \in A) \xrightarrow{p} P(X \in A). \tag{2}$$

Simulación de ejemplo: lanzamiento de una moneda equilibrada

En esta sección, te proponemos simular algunos experimentos. Te vamos a guiar para que puedas hacer esta simulación y para que tengas las herramientas necesarias para hacer luego la Actividad 1 para entregar y (si querés) la Actividad 2 optativa. Contrastaremos frecuencias relativas con probabilidades teóricas y promedios (medias muestrales) con esperanzas, a modo de corroborar empíricamente la Ley de los Grandes Números.

Para simular, utilizaremos `R` para que nos provea datos que emulen un experimento aleatorio determinado. En tal caso, diremos que `R` genera o simula datos de acuerdo a cierto mecanismo que debemos explicitar.

Comenzaremos simulando el lanzamiento de una moneda equilibrada. En `R`, creamos una moneda; la lanzamos y vemos qué se obtiene. Para ello utilizaremos el comando `sample(x, size, replace = FALSE, prob = NULL)` como se ve a continuación.

```
# Creamos la moneda: vector con los posibles resultados del lanzamiento
moneda <- c("cara", "ceca")
## [1] "cara" "ceca"

# Lanzamos la moneda
sample(x = moneda, size = 1)
## [1] "ceca"

# Lanzamos 10 monedas
sample(x = moneda, size = 10, replace = TRUE)
##  [1] "cara" "ceca" "cara" "cara" "cara" "cara" "ceca" "cara" "cara" "
   cara"
```

El argumento `replace = TRUE` indica a la función `sample` que el muestreo se hace con reposición, algo razonable en este experimento.

Como sabemos, la Ley de los Grandes Números indica que la frecuencia relativa de "cara" en n ensayos independientes converge (en probabilidad) a $1/2$ cuando n tiende a infinito.

Vamos entonces a lanzar $N_{rep} = 1000$ veces la moneda y a guardar los resultados en el vector `muchas_monedas`. De alguna manera ese valor hará el papel de infinito, y trabajaremos con $n = 1, 2, \ldots, N_{rep}$ ensayos, con n creciendo a N_{rep}.

```
Nrep <- 1000
muchas_monedas <- sample(x = moneda , size = Nrep , replace = TRUE)

# Veamos los primeros 10 resultados
muchas_monedas[1:10]
##  [1] "ceca" "cara" "cara" "cara" "cara" "cara" "cara" "cara" "cara" "
   cara"
```

Podemos preguntarnos cuántas veces se obtuvo "cara" en los primeros 10 lanzamientos. Para averiguarlo, podemos verificar si cada coordenada del vector de resultados es (o no) "cara". Eso nos devuelve un vector con TRUE y FALSE, según el valor de cada coordenada sea o no "cara". Luego, sumamos el vector lógico, de forma tal que los TRUE pasan a ser

1 y los FALSE, 0.

```
muchas_monedas[1:10] == "cara"
##  [1] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE

sum(muchas_monedas[1:10] == "cara")
## [1] 9
```

Tenemos entonces que la frecuencia relativa de "cara" en los primeros 10 lanzamientos se obtiene sumando los TRUE (cada uno de ellos es un 1) y dividiendo por 10; es decir, promediando.

```
sum(muchas_monedas[1:10] == "cara")/10
## [1] 0.9

# Es lo mismo que promediar usando "mean"
mean(muchas_monedas[1:10] == "cara")
## [1] 0.9
```

Vamos a estudiar la frecuencia de "cara" en los primeros $n = 20$ lanzamientos.

```
mean(muchas_monedas[1:20] == "cara")
## [1] 0.6
```

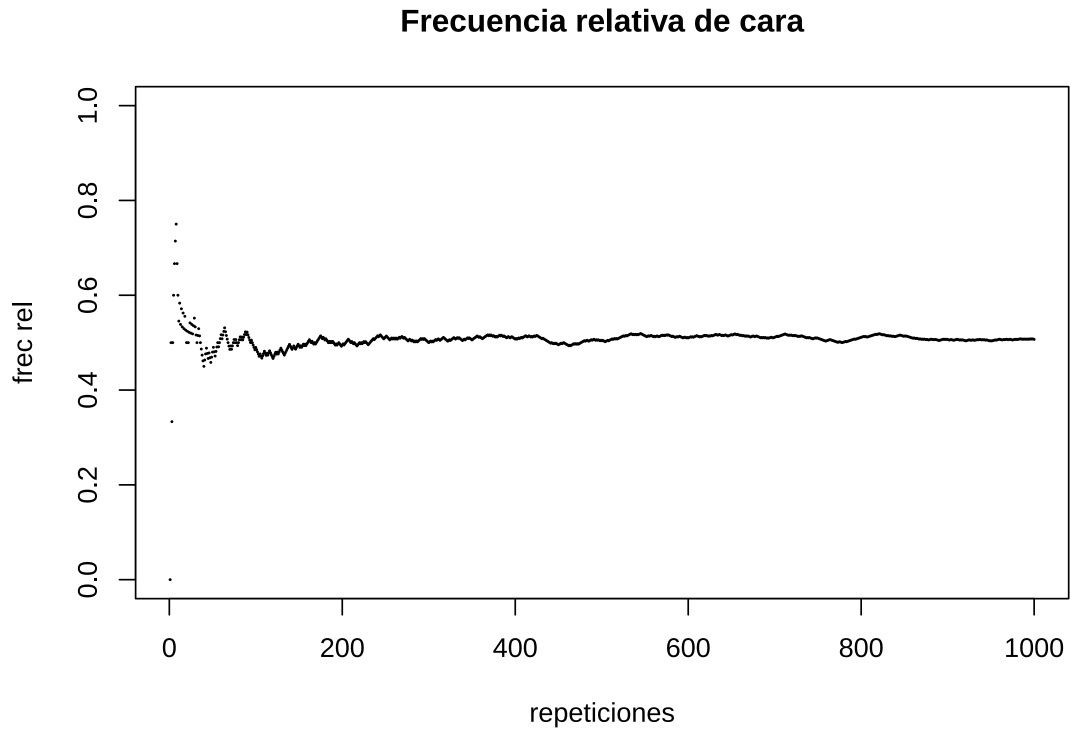
Podemos, en realidad, calcular la frecuencia de "cara" en los primeros n lanzamientos, siendo n un valor entre 1 y $N_{rep} = 1000$. Para guardar la frecuencia relativa de "cara" en los primeros n lanzamientos creamos el vector `frec_relativa_vec`.

```
# Creamos el vector de longitud Nrep y lo llenamos con NA's
frec_relativa_vec <- rep(NA, Nrep)

# Completamos con los promedios
for(n in 1:Nrep)
{
  frec_relativa_vec[n] <- mean(muchas_monedas[1:n] == "cara")
}
```

Vamos ahora a ver gráficamente cómo se comporta la frecuencia relativa a medida que la cantidad n de lanzamientos aumenta.

```
plot(seq(1:Nrep), frec_relativa_vec, cex = 0.1, ylim = c(0,1),
     xlab = "repeticiones", ylab = "frec rel",
     main = "Frecuencia relativa de cara")
# El argumento ylim = c(0,1) indica los limites en el eje y.
# El argumento cex = 0.1 permite "achicar" a un 10% el tama o de los
   puntos en el grafico.
```



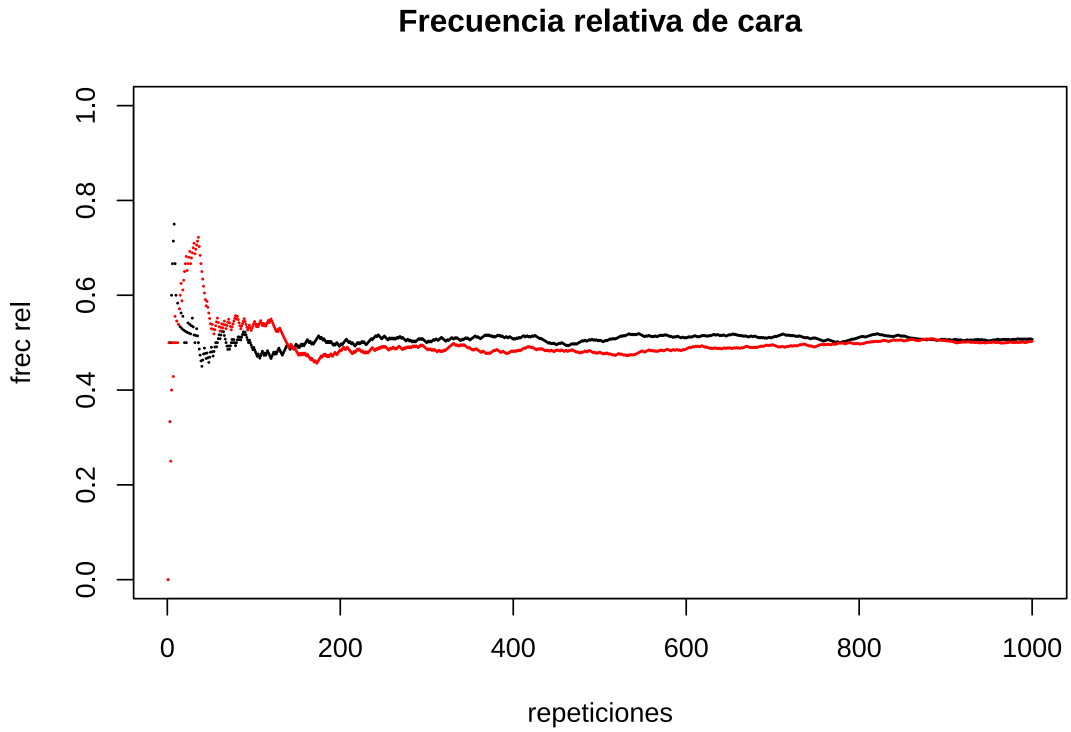
En el gráfico vemos el comportamiento **asintótico** de las frecuencias relativas de **esta** simulación, con **estos** lanzamientos. Pero, ¿siempre obtendremos estos resultados? Repitamos el experimento y agreguemos estos nuevos resultados en otro color.

```
# Nueva simulacion
muchas_monedas_2 <- muchas_monedas <- sample(x = moneda , size = Nrep ,
    replace = TRUE)
frec_relativa_vec_2 <- rep(NA, Nrep)

for(n in 1:Nrep)
{
  frec_relativa_vec_2[n] <- mean(muchas_monedas_2[1:n] == "cara")
}

# Grafico para las frel de la primera simulacion
plot(seq(1:Nrep), frec_relativa_vec, cex = 0.1, ylim = c(0,1),
     xlab = "repeticiones", ylab = "frec rel", main = "Frecuencia relativa
    de cara")

# Agregamos las de la nueva simulacion
points(seq(1:Nrep), frec_relativa_vec_2, cex = 0.1, col = "red")
```



Supongamos que en este curso somos 30 participantes. Sería interesante ver qué ocurriría si cada uno de nosotros hiciera una de estas simulaciones y, al final, viéramos gráficamente qué comportamiento tienen todas ellas.

Más aún: sería deseable no tener que hacer las 30 simulaciones una por una, sino de manera automática. Para ello, repetiremos lo que hicimos recién, pero 28 veces más utilizando un bucle. El índice del bucle lo usaremos, también, para indicar el color de los correspondientes puntos.

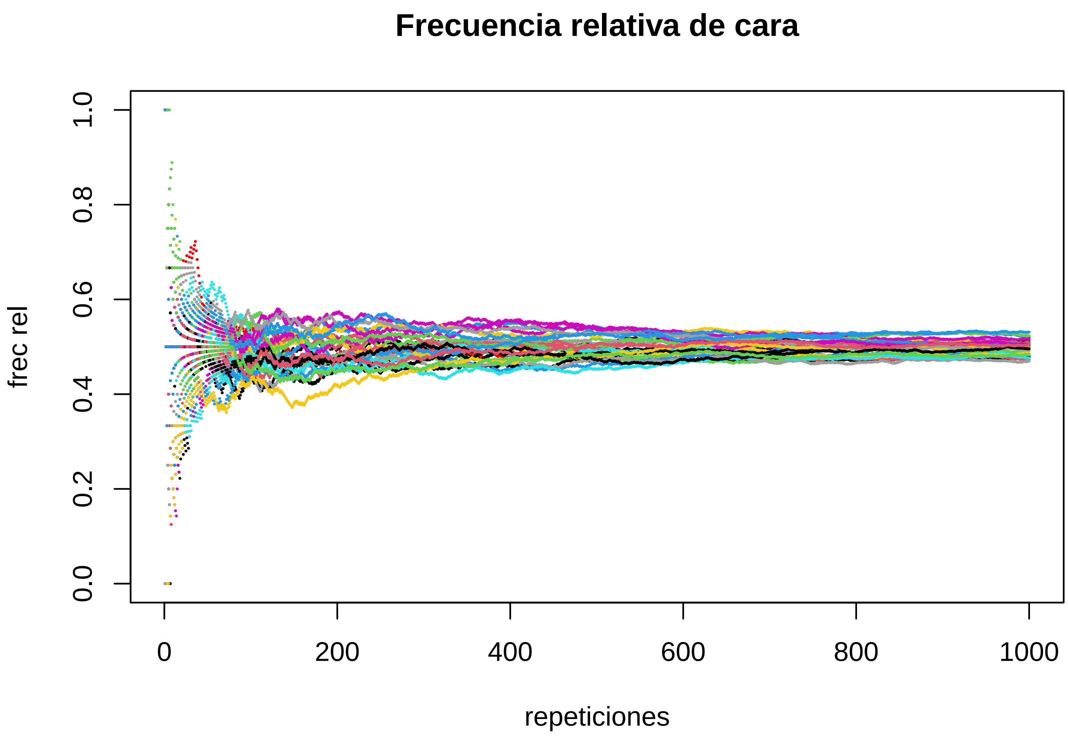
```
# Empecemos haciendo el grafico con la primera simulacion
plot(seq(1:Nrep), frec_relativa_vec, cex = 0.1, ylim = c(0,1),
     xlab = "repeticiones", ylab = "frec rel", main = "Frecuencia relativa
   de cara")

# Agregamos la segunda
points(seq(1:Nrep), frec_relativa_vec_2, cex=0.1, col="red")

# Creamos (y agregamos) las de los 28 participantes restantes
N_gente <- 28
for(j in 1:N_gente)
{
  # Agreguemos la simulacion del individuo j
  muchas_monedas <- muchas_monedas <- sample(x = moneda , size = Nrep ,
   replace = TRUE) # aca estan los lanzamientos del individuo j
  frec_relativa_vec <- rep(NA, Nrep)
  for(n in 1:Nrep)
  {
```

```
    frec_relativa_vec[n] <- mean(muchas_monedas[1:n] == "cara")
  } # aca tenemos las frecuencias relativas del individuo j

  # Agregamos al grafico los puntos del individuo j en color j
  points(seq(1:Nrep), frec_relativa_vec, cex = 0.1, col = j)
}
```



Actividad 1: Simulación del lanzamiento de un dado (PARA ENTREGAR)

En esta actividad te proponemos que estudies algo similar a lo que hicimos en la simulación de la moneda, pero con la simulación del lanzamiento de un dado.

Sea X el resultado del lanzamiento de un dado equilibrado.

1. Indicar el valor de $P(X = 5)$ y el de $\mathbb{E}(X)$.
2. Construir un vector `muchos_dados` y guardar en él los resultados correspondientes a lanzar `Nrep=1000` veces el dado.
3. Para cada valor $n = 1, \ldots, N_{rep}$, calcular la frecuencia relativa con la que el 5 aparece en los primeros n lanzamientos y guardarla en el vector `frec_relativa_dado5_vec`.
4. Graficar en el eje x los valores de n y en el eje y las correspondientes frecuencias relativas. ¿Qué observa? Indicar, si es posible, a qué valor deberían converger esas

frecuencias y por qué. ¿Se corresponde lo que observa en la práctica con lo que espera de la teoría?

5. Guardar en el vector `promedios_dado_vec` el valor de los promedios, para $n = 1, \ldots, N_{rep}$.

6. Graficar en el eje x los valores de n y en el eje y los correspondientes promedios. Indicar, si es posible, a qué valor deberían converger esos promedios y por qué. ¿Se corresponde lo que observa en la práctica con lo que espera de la teoría?

7. Repetir `Ngen = 10` veces, agregando los promedios de cada individuo en otro color. ¿Qué debería pasar? ¿Por qué? ¿Qué se observa?

Actividad 2: Simulación de diferentes distribuciones para ver la convergencia empírica del promedio y la frecuencia relativa (OPTATIVA)

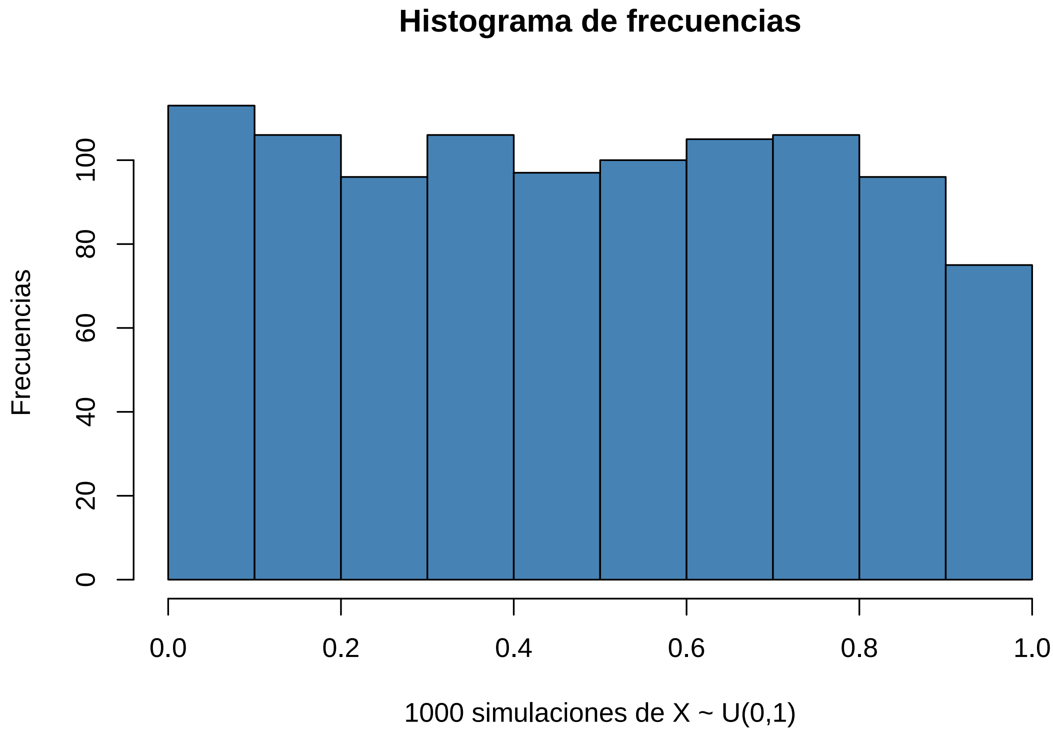
Para lo que sigue, utilizaremos X para denotar de manera genérica una variable con distribución F. Es decir, $X \sim F$. Generaremos N_{rep} datos con la distribución F para luego corroborar la convergencia de promedios a esperanzas, como postula la ecuación (1) y de frecuencias relativas a probabilidades, como postula la ecuación (2).

Generar datos de una distribución F dada no es algo tan sencillo. Sin embargo, existen comandos en `R` que pueden hacerlo para ciertas distribuciones famosas (como la uniforme y la normal, entre otras).

Por ejemplo, el comando `runif(n, min = 0, max = 1)` generará n datos provenientes de una distribución uniforme en el intervalo $(0, 1)$. Podemos simular N_{rep} de esos datos y hacer un histograma para ver cómo se distribuyen y si es consistente con lo que se espera de dicha distribución.

```r
# Generamos Nrep uniformes en (0,1)
muchas_uniformes <- runif(Nrep, min = 0, max = 1)

# Realizamos un histograma de frecuencias
hist(muchas_uniformes, xlab = "1000 simulaciones de X ~ U(0,1)", ylab = "
    Frecuencias", main = "Histograma de frecuencias", col = "steelblue")
```



En lo sucesivo, utilizarás los comandos `rbinom()`, `rnorm()` y `rexp()` para simular las distribuciones pedidas. Recordá que siempre podés invocar el `help()` para acceder a más información sobre los comandos. En particular, aquí será especialmente útil, pues es necesario saber qué argumentos requieren esas funciones para poder simular las distribuciones en cuestión.

✔ Distribución Bernoulli

Sea $X \sim \mathcal{B}e(p)$ siendo $p = 0,2$.

Promedios y esperanzas

- Indicar el valor de $\mathbb{E}(X)$.
- Guardar en el vector `muchas_bernoullies`, $N_{rep} = 1000$ datos de esta distribución utilizando el comando `rbinom()`. Prestar atención a los parámetros que utiliza esta función leyendo antes el `help(rbinom)`.
- Guardar en el vector `muchos_promedios_bernoullies` el promedio de las primeras n observaciones con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y los correspondientes promedios. Utilizar `ylim = c(0,1)`.

- Repetir los ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior los promedios de cada individuo en un color diferente.
- Indicar, según (1), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa en el gráfico?

✔ Distribución Uniforme

Sea $X \sim \mathcal{U}(a, b)$ siendo $a = 67$, $b = 73$.

Promedios y esperanzas

- Indicar el valor de $\mathbb{E}(X)$.
- Guardar en el vector `muchas_uniformes`, $N_{rep} = 1000$ datos de esta distribución utilizando el comando `runif()`. Prestar atención a los parámetros que utiliza esta función leyendo antes el `help(runif)`.
- Guardar en el vector `muchos_promedios_uniformes` el promedio de las primeras n observaciones con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y los correspondientes promedios. Utilizar `ylim = c(65,75)`.
- Repetir los ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior los promedios de cada individuo en un color diferente.
- Indicar, según (1), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa en el gráfico?

Frecuencias relativas y probabilidades

- Indicar el valor de $P(68 \leq X \leq 72)$.
- Guardar en el vector `muchos_frec_rel_uniformes` la frecuencia relativa con la cual el dato de `muchas_uniformes` aparece en el intervalo pedido $(68, 72)$ a lo largo de las primeras n observaciones, con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y las correspondientes frecuencias relativas. Utilizar `ylim = c(0,1)`.
- Repetir los dos ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior las frecuencias relativas de cada individuo en un color diferente.
- Indicar, según (2), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa en el gráfico?

✔ Distribución Normal

Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ siendo $\mu = 70$, $\sigma^2 = 1{,}2$.

Promedios y esperanzas

- Indicar el valor de $\mathbb{E}(X)$.
- Guardar en el vector `muchas_normales`, $N_{rep} = 1000$ datos de esta distribución utilizando el comando `rnorm()`. Prestar atención a los parámetros que utiliza esta función leyendo antes el `help(rnorm)`.
- Guardar en el vector `muchos_promedios_normales` el promedio de las primeras n observaciones, con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y los correspondientes promedios. Utilizar `ylim = c(60,80)`.
- Repetir los ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior los promedios de cada individuo en un color diferente.
- Indicar, según (1), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa el gráfico?

Frecuencias relativas y probabilidades

- Indicar el valor de $P(68 \leq X \leq 72)$.
- Guardar en el vector `muchos_frec_rel_normales` la frecuencia relativa con la cual el dato de `muchas_normales` aparece en el intervalo pedido $(68, 72)$ a lo largo de las primeras n observaciones, con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y las correspondientes frecuencias relativas. Utilizar `ylim = c(0,1)`.
- Repetir los dos ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior las frecuencias relativas de cada individuo en un color diferente.
- Indicar, según (2), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa en el gráfico?

✔ Distribución Exponencial

Antes de comenzar, te sugerimos revisar el ejercicio 1 de la Práctica 1. Vamos a simular datos de una distribución exponencial.

Sea $X \sim \mathcal{E}(\lambda)$ siendo $\lambda = 1/10$.

Promedios y esperanzas

- Indicar el valor de $\mathbb{E}(X)$.
- Guardar en el vector `muchas_exponenciales`, $N_{rep} = 1000$ datos de esta distribución utilizando el comando `rexp()`. Prestar atención a los parámetros que utiliza esta función leyendo antes el `help(rexp)`.
- Realizar un histograma con los datos del vector `muchas_exponenciales`. Si quiere, adicionalmente también puede graficar la densidad asociada a esos datos usando la función `dexp()` y el comando `curve(..., add = TRUE)`. En ese caso, deberá realizar un histograma de densidad (deberá indicar `probability = TRUE` en el argumento de `hist()`). Recuerde que los datos fueron simulados con $\lambda = 1/10$.
- Guardar en el vector `muchos_promedios_exponenciales` el promedio de las primeras n observaciones, con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y los correspondientes promedios. Utilizar `ylim = c(0,20)`.
- Repetir los ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior los promedios de cada individuo en un color diferente.
- Indicar, según (1), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que muestra el gráfico?

Frecuencias relativas y probabilidades

- Indicar el valor de $P(X \leq 12)$.
- Guardar en el vector `muchos_frec_rel_exponenciales` la frecuencia relativa con la cual el dato de `muchas_exponenciales` aparece en el intervalo pedido $(-\infty, 12)$ a lo largo de las primeras n observaciones, con $n = 1, \ldots, N_{rep}$.
- Graficar en el eje x los valores de n y en el eje y las correspondientes frecuencias relativas. Utilizar `ylim = c(0,1)`.
- Repetir los dos ítems anteriores $N_{gen} = 10$ veces, agregando al gráfico anterior las frecuencias relativas de cada individuo en un color diferente.
- Indicar, según (2), cuál es el valor límite de estas sucesiones. ¿Se condice con lo que observa en el gráfico?