

Primer Proyecto de Probabilidad y Estadística

Cada grupo trabajará con 2 distribuciones asignadas, una discreta y otra continua, llamadas **Distribución 1** y **Distribución 2**.

Ejercicio 1. En este ejercicio se utiliza la **Distribución 1** (que es discreta) para explorar los resultados de su simulación y compararlos con los valores teóricos respectivos. Para eso se deberán generar 4 muestras independientes de tamaño n , para $n = 10^2, 10^3, 10^4$ y 10^5 de la distribución. Se pide:

- Presentar en un mismo gráfico los 4 boxplot (diagrama de cajas) de las muestras generadas y dar una interpretación del mismo.
- Calcular los valores empíricos de la esperanza y de la varianza de cada una de las 4 muestras, y compararlos con los valores teóricos correspondientes.
- Presentar en un mismo gráfico la función de distribución teórica de la **Distribución 1** y la función de distribución empírica de la muestra para $n = 10^2$. Hacer otro gráfico similar para $n = 10^5$. Concluya sobre los resultados obtenidos.

Ejercicio 2. En este ejercicio se utiliza la **Distribución 2** (que es continua) para explorar los resultados de su simulación y compararlos con los valores teóricos respectivos. Para eso se generarán 4 muestras aleatorias simples, de tamaño n , para $n = 10^2, 10^3, 10^4$ y 10^5 de la distribución.

- Presentar en un mismo gráfico los 4 boxplot (diagrama de cajas) de las muestras generadas e interpretarlo.
- Calcular los valores empíricos de la esperanza, de la varianza de cada una de las 4 muestras, y compararlos con los valores teóricos correspondientes.
- Presentar para $n = 100$, en un mismo gráfico, el histograma de los valores simulados y la densidad teórica de la variable simulada. Hacer otro gráfico similar, pero para $n = 10^5$. Concluya sobre los resultados obtenidos.

Ejercicio 3. En este ejercicio se explorará empíricamente el *Teorema Central del Límite*, que establece que dadas X_1, X_2, \dots, X_n , variables aleatorias independientes e idénticamente distribuidas, con valor esperado μ y varianza σ^2 , se verifica

$$\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \approx N(0, 1), \text{ para valores grandes de } n.$$

Parte 1.

- Generar una muestra aleatoria simple de 1000 datos con **Distribución 2**.
- Calcular la media empírica de la muestra, que llamaremos $\hat{\mu}$.
- Hallar la media teórica, que llamaremos μ y la desviación estándar teórica, que llamaremos σ . Calcular el valor estandarizado de la media empírica, que se define como

$$\sqrt{n} \left(\frac{\hat{\mu} - \mu}{\sigma} \right)$$

Parte 2.

- a) Repetir la **Parte 1** $k = 500$ veces para obtener 500 valores del promedio estandarizado definido anteriormente.
- b) Presentar los datos de estos k valores estandarizados en un histograma, y en el mismo gráfico superpuesto presentar la función de densidad de la normal estándar. Concluir según el resultado observado.

Sobre el informe a presentar:

- La fecha límite para entregar el informe es el 10 de octubre a las 23 hs. Más adelante se creará un recurso en webasignatura para realizar la entrega online.
- El informe deberá estar en formato pdf, debe incluir los gráficos solicitados y los scripts utilizados. No es necesario entregar por separado los scripts utilizados.
- El informe deberá contener título, fecha, nombre y cédula de los integrantes del equipo. Se recomienda la estructura: Introducción, desarrollo conclusiones, bibliografía y anexos si tuviere.
- Se evaluará: prolijidad del informe, utilización correcta del idioma español, redacción, prolijidad del código presentado en los scripts, calidad de los gráficos, conclusiones.

Algunos comandos sugeridos: A continuación se dan algunos comandos que pueden ser de utilidad, no es obligatorio usarlos. Se sugiere buscar más información en internet sobre cada uno de ellos.

- `rbinom(m,size=n,prob=p)` genera una muestra aleatoria simple de tamaño m con distribución binomial de parámetros (n, p) .
- `rpois(n,lambda= λ)` genera una muestra aleatoria simple de tamaño n con distribución de Poisson de parámetro λ .
- `rgeom(n,prob=p)` genera una muestra aleatoria simple de tamaño n con distribución geométrica de parámetro p .
- `runif(n,min=a,max=b)` genera una muestra aleatoria simple de tamaño n con distribución uniforme en el intervalo $[a, b]$.
- `rexp(n,rate= λ)` genera una muestra aleatoria simple de tamaño n con distribución exponencial de parámetro λ .
- `rnorm(n,mean= μ ,sd= σ)` genera una muestra aleatoria simple de tamaño n con distribución normal de parámetros (μ, σ^2) .
- `boxplot(x)` genera un diagrama de cajas de los datos en x . Se sugiere buscar información para agregar colores, etiquetas, y otras cosas.
- `mean(vector)` calcula la media empírica (el promedio) del vector de datos "**vector**".
- `sd(vector)` calcula el desvío estándar empírico de un vector de datos "**vector**".
- `hist(vector,freq=FALSE)` calcula el histograma del vector de datos "**vector**". Para cambiar la cantidad de intervalos definidos por defecto consultar la documentación oficial de R, también consultar la documentación para agregar colores, etiquetas, y otras cosas para mejorar la calidad del histograma. Usar **FALSE** en el segundo parámetro muestra frecuencias relativas en el eje vertical, usar **TRUE** muestra frecuencias absolutas.
- `for` estructura de control iterativa.
- `table(x)` nos da las frecuencias en un vector x .

- `plot(x,y)` grafica los valores del vector `y` en el eje vertical y los valores del vector `x` en el eje horizontal. Se sugiere consultar la documentación de R para mejorar la calidad de los gráficos, agregar colores, inclusión de nombres a los ejes coordenados, y varias otras opciones.
- `lines(z,t)` agrega un gráfico similar sobre el anterior. Debe utilizarse después de la función `plot` y los puntos estarán conectados por líneas.
- `curve(ddist, from=inicio, to=fin, n=m, add = TRUE)` agrega la densidad de `dist` al gráfico o histograma anterior, desde el valor `inicio` hasta el valor `fin`, usando `m` puntos para evaluar la función a dibujar. Si usamos `FALSE` como último parámetro entonces se crea un nuevo gráfico en vez de agregarse al gráfico anterior. A continuación algunos ejemplos de posibilidades para sustituir `ddist`:

Utilizar `dnorm(x, mean= μ , sd= σ)` para dibujar la densidad de la normal de parámetros (μ, σ^2) .

Utilizar `dexp(x, mean= λ)` para dibujar la densidad de la exponencial de parámetro λ .

Utilizar `dunif(x, min= a , max= b)` para dibujar la densidad de la uniforme de parámetros (a, b) .

En todos los casos la `x` es importante y no importa que valor tenga como variable, el resultado será el mismo gráfico.

- `stepfun(...)` devuelve una función escalonada construida a partir de los parámetros, ésta puede ser graficada con el comando `plot`. Complementar con las funciones `pbinom`, `pgeom` o `ppois` respectivamente según corresponda para graficar la función de distribución acumulada teórica de una variable aleatoria discreta. Ejemplos de ejecución se mostraron en las clases.
- `cumsum(vector)` da el vector de sumas acumuladas del vector de datos "`vector`".
- `abline(h=a)` agrega una línea horizontal en un gráfico a la altura del valor `a`.
- `ecdf(vector)` devuelve la función de distribución empírica del vector de datos "`vector`".