Tarea 3

Probabilidad y Estadística Aplicada

Facultad de Ingeniería

Julián Cardozo - 4329323-0

Pedro Solomita - 4943444-6

Ana Silvera - 4230076-5

Teorema Central del Límite

El teorema central del límite lo que nos dice es nuevamente algo acerca del promedio, lo que nos dice es que:

Sean X1, X2, … Xn … una sucesión de variables aleatorias independientes e idénticamente distribuidas.

Supongamos que E(X1) = μ y Var(X1) = σ^2
Entonces para todo t real se cumple que

$$\lim_{n \to \infty} P(\sqrt{n}(\frac{\overline{Xn} - \mu}{\sigma}) \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, dx$$

Este teorema nos dice que si una muestra es lo bastante grande (generalmente cuando el tamaño muestral (n) supera los 30), sea cual sea la distribución de la media muestral, seguirá aproximadamente una distribución normal.

Es decir, dada cualquier variable aleatoria, si extraemos muestras de tamaño n (n>30) y calculamos los promedios muestrales, dichos promedios seguirán una distribución normal. Además, la media será la misma que la de la variable de interés, y la desviación estándar de la media muestral será aproximadamente el error estándar. Un caso concreto del teorema central del límite es la distribución binomial.

A partir de n=30, la distribución binomial se comporta estadísticamente como una normal, por lo que podemos aplicar los tests estadísticos apropiados para esta distribución.

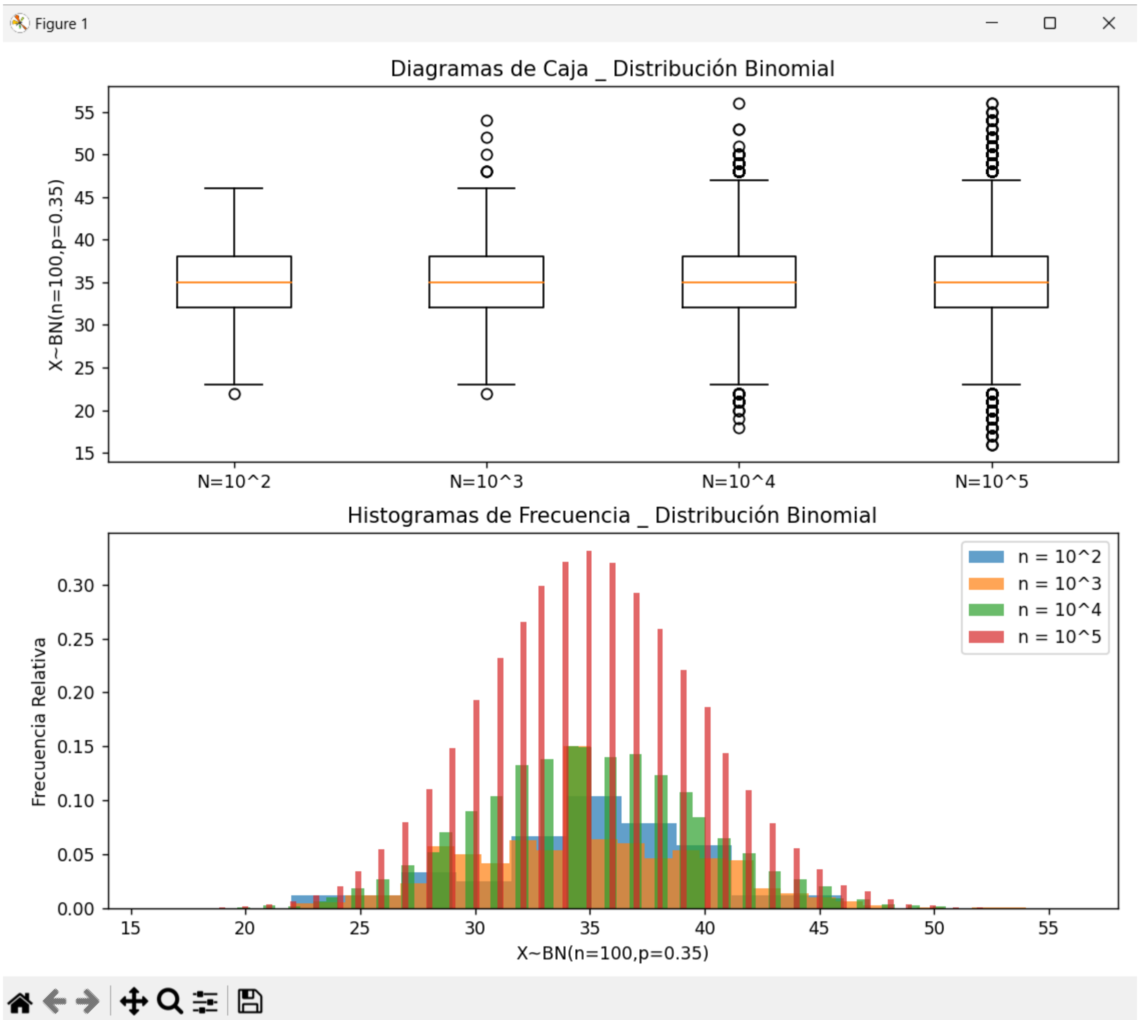
La importancia del teorema central del límite radica en que mediante un conjunto de teoremas, se desvelan las razones por las cuales, en muchos campos de aplicación, se encuentran en todo momento distribuciones normales o casi normales.

Repositorio del código de los ejercicios: https://github.com/juliancardozo/Repo-Tarea3/

Distribución 1: binomial de parámetros n = 100, p = 0,35.

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import binom
from scipy.stats import mode
#a) Muestras aleatorias de tama˜nos 10^2, 10^3, 10^4 y 10^5
r1 = binom.rvs(100,0.35,size = 10**2)
r2 = binom.rvs(100,0.35,size = 10**3)
r3 = binom.rvs(100,0.35,size = 10**4)
r4 = binom.rvs(100,0.35,size = 10**5)
#b) Diagrama de cajas para cada muestra generada
fig, (ax1, ax2) = plt.subplots(2,1,figsize=(8,8))
ax1.boxplot([r1, r2, r3, r4])
ax1.set_xticklabels(['N=10^2', 'N=10^3', 'N=10^4', 'N=10^5'])
ax1.set_ylabel('X~BN(n=100,p=0.35)')
ax1.set_title('Diagramas de Caja _ Distribución Binomial')
#c) Histograma de las muestras generadas
ax2.hist(r1, bins='auto', alpha=0.7, density=True, label='n = 10^2')
ax2.hist(r2, bins='auto', alpha=0.7, density=True, label='n = 10^3')
ax2.hist(r3, bins='auto', alpha=0.7, density=True, label='n = 10^4')
ax2.hist(r4, bins='auto', alpha=0.7, density=True, label='n = 10^5')
ax2.set_xlabel('X~BN(n=100,p=0.35)')
ax2.set_ylabel('Frecuencia Relativa')
ax2.set_title('Histogramas de Frecuencia _ Distribución Binomial')
ax2.legend()
fig.tight_layout()
plt.show()
print("Muestras :  [10^2, 10^3, 10^4, 10^5]")
#d) Mediana y moda de cada muestra
print("Medianas : ",[np.median(r1), np.median(r2), np.median(r3),
np.median(r4)])
print("Modas    : ",[mode(r1,axis=None,keepdims=True).mode[0],
                mode(r2,axis=None,keepdims=True).mode[0],
                mode(r3,axis=None,keepdims=True).mode[0],
                mode(r4,axis=None,keepdims=True).mode[0]])
#e) Medias muestrales y media poblacional
print("Medias   : ",[np.mean(r1), np.mean(r2), np.mean(r3),
np.mean(r4)])
print("µ = ",100*0.35)
#f) Varianzas muestrales y varianza teórica
v=[np.var(r1), np.var(r2), np.var(r3), np.var(r4)]
print("Varianzas: ",[np.var(r1), np.var(r2), np.var(r3),
np.var(r4)])
print("σ = ",100*0.35*0.65)
```

Los diagramas de caja mostrados en la figura a continuación indican que existen datos atípicos en todas las muestras, la cantidad se incrementa con el tamaño de la muestra y es simétrica respecto de la media muestral, tal como sigue la distribución de probabilidad cuya forma tiende a la correspondiente a la distribución Binomial



La ley de los grandes números indica que para una sucesión de v.a.i.i.d (X1, X2, …, Xn) con media finita μ, se satisface la siguiente convergencia

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

Independientemente de la función de distribución asociada a X.

El siguiente cuadro de resultados corrobora esta ley, a medida que el tamaño de la muestra crece, la media muestral efectivamente se acerca a la media poblacional μ = 35

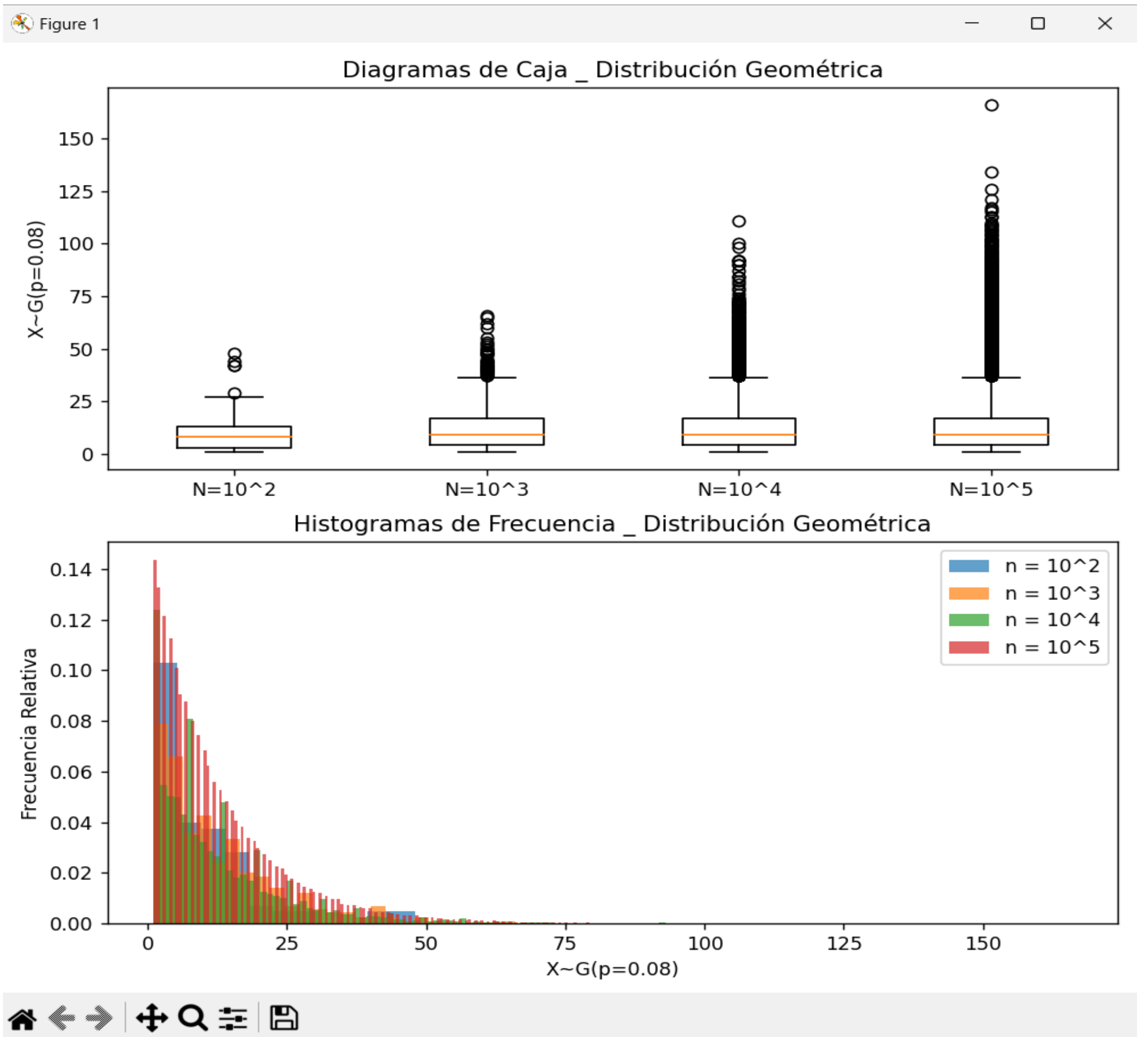
```
Muestras :  [10^2, 10^3, 10^4, 10^5]
Medianas :  [35.0, 35.0, 35.0, 35.0]
Modas    :  [35, 34, 34, 35]
Medias   :  [34.88, 34.878, 34.9803, 35.00041]
μ  =  35.0
Varianzas:  [22.965599999999995, 23.603116000000004, 22.87251191,
22.5606698319]
σ  =  22.75
```

En el caso de la varianza, no se observa una convergencia estricta para los grandes números ya que la velocidad de convergencia para la media es relativamente lenta.

Distribución 2: geométrica de parámetro p = 0,08

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import geom
from scipy.stats import mode
#a) Muestras aleatorias de tama˜nos 10^2, 10^3, 10^4 y 10^5
r1 = geom.rvs(0.08,size = 10**2)
r2 = geom.rvs(0.08,size = 10**3)
r3 = geom.rvs(0.08,size = 10**4)
r4 = geom.rvs(0.08,size = 10**5)
#b) Diagrama de cajas para cada muestra generada
fig, (ax1, ax2) = plt.subplots(2,1,figsize=(8,8))
ax1.boxplot([r1, r2, r3, r4])
ax1.set_xticklabels(['N=10^2', 'N=10^3', 'N=10^4', 'N=10^5'])
ax1.set_ylabel('X~G(p=0.08)')
ax1.set_title('Diagramas de Caja _ Distribución Geométrica')
#c) Histograma de las muestras generadas
ax2.hist(r1, bins='auto', alpha=0.7, density=True, label='n = 10^2')
ax2.hist(r2, bins='auto', alpha=0.7, density=True, label='n = 10^3')
ax2.hist(r3, bins='auto', alpha=0.7, density=True, label='n = 10^4')
ax2.hist(r4, bins='auto', alpha=0.7, density=True, label='n = 10^5')
ax2.set_xlabel('X~G(p=0.08)')
ax2.set_ylabel('Frecuencia Relativa')
ax2.set_title('Histogramas de Frecuencia _ Distribución Geométrica')
ax2.legend()
fig.tight_layout()
plt.show()
print("Muestras :  [10^2, 10^3, 10^4, 10^5]")
#d) Mediana y moda de cada muestra
print("Medianas : ",[np.median(r1), np.median(r2), np.median(r3),
np.median(r4)])
print("Modas    : ",[mode(r1,axis=None,keepdims=True).mode[0],
                mode(r2,axis=None,keepdims=True).mode[0],
                mode(r3,axis=None,keepdims=True).mode[0],
                mode(r4,axis=None,keepdims=True).mode[0]])
#e) Medias muestrales y media poblacional
print("Medias   : ",[np.mean(r1), np.mean(r2), np.mean(r3),
np.mean(r4)])
print("µ = ",1/0.08)
#f) Varianzas muestrales y varianza teórica
v=[np.var(r1), np.var(r2), np.var(r3), np.var(r4)]
print("Varianzas: ",[np.var(r1), np.var(r2), np.var(r3),
np.var(r4)])
print("σ = ",(1-0.08)/(0.08**2))
```

Los diagramas de caja mostrados en la figura a continuación indican que existen datos atípicos en todas las muestras, la cantidad se incrementa con el tamaño de la muestra y es asimétrica respecto de la media muestral, tal como sigue la distribución de probabilidad cuya forma tiende a la correspondiente a la distribución Geométrica



La ley de los grandes números indica que para una sucesión de v.a.i.i.d (X1, X2, …, Xn) con media finita μ, se satisface la siguiente convergencia

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

Independientemente de la función de distribución asociada a X.

El siguiente cuadro de resultados corrobora esta ley, a medida que el tamaño de la muestra crece, la media muestral efectivamente se acerca a la media poblacional μ = 12.5

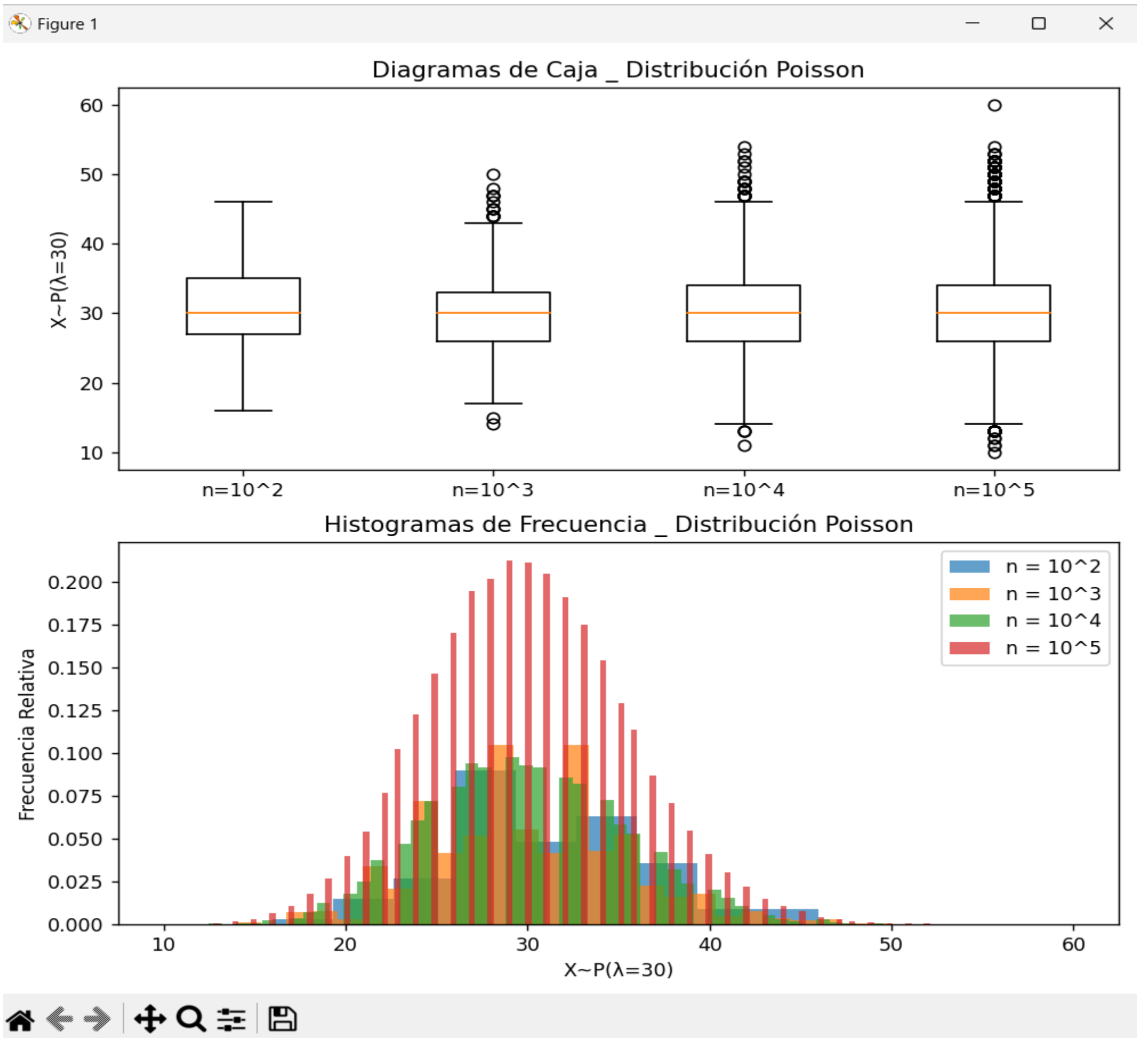
```
Muestras :  [10^2, 10^3, 10^4, 10^5]
Medianas :  [8.0, 9.0, 9.0, 9.0]
Modas    :  [1, 1, 1, 1]
Medias   :  [9.52, 12.507, 12.6526, 12.50039]
μ  =  12.5
Varianzas:  [89.18960000000001, 123.307951, 145.73051324,
144.25725984790003]
σ  =  143.75
```

En el caso de la varianza, se observa una convergencia estricta para los grandes números ya que la velocidad de convergencia para la media es relativamente rápida.

Distribución 3: poisson de parámetro λ = 30

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import poisson
from scipy.stats import mode
#a) Muestras aleatorias de tama˜nos 10^2, 10^3, 10^4 y 10^5
r1 = poisson.rvs(30,size = 10**2)
r2 = poisson.rvs(30,size = 10**3)
r3 = poisson.rvs(30,size = 10**4)
r4 = poisson.rvs(30,size = 10**5)
#b) Diagrama de cajas para cada muestra generada
fig, (ax1, ax2) = plt.subplots(2,1,figsize=(8,8))
ax1.boxplot([r1, r2, r3, r4])
ax1.set_xticklabels(['n=10^2', 'n=10^3', 'n=10^4', 'n=10^5'])
ax1.set_ylabel('X~P(λ=30)')
ax1.set_title('Diagramas de Caja _ Distribución Poisson')
#c) Histograma de las muestras generadas
ax2.hist(r1, bins='auto', alpha=0.7, density=True, label='n = 10^2')
ax2.hist(r2, bins='auto', alpha=0.7, density=True, label='n = 10^3')
ax2.hist(r3, bins='auto', alpha=0.7, density=True, label='n = 10^4')
ax2.hist(r4, bins='auto', alpha=0.7, density=True, label='n = 10^5')
ax2.set_xlabel('X~P(λ=30)')
ax2.set_ylabel('Frecuencia Relativa')
ax2.set_title('Histogramas de Frecuencia _ Distribución Poisson')
ax2.legend()
fig.tight_layout()
plt.show()
print("Muestras :  [10^2, 10^3, 10^4, 10^5]")
#d) Mediana y moda de cada muestra
print("Medianas : ",[np.median(r1), np.median(r2), np.median(r3),
np.median(r4)])
print("Modas    : ",[mode(r1,axis=None,keepdims=True).mode[0],
                mode(r2,axis=None,keepdims=True).mode[0],
                mode(r3,axis=None,keepdims=True).mode[0],
                mode(r4,axis=None,keepdims=True).mode[0]])
#e) Medias muestrales y media poblacional
print("Medias   : ",[np.mean(r1), np.mean(r2), np.mean(r3),
np.mean(r4)])
print("µ = ",30)
#f) Varianzas muestrales y varianza teórica
v=[np.var(r1), np.var(r2), np.var(r3), np.var(r4)]
print("Varianzas: ",[np.var(r1), np.var(r2), np.var(r3),
np.var(r4)])
print("σ = ",30)
```

Los diagramas de caja mostrados en la figura a continuación indican que existen datos atípicos en todas las muestras, la cantidad se incrementa con el tamaño de la muestra y es técnicamente simétrica respecto de la media muestral, tal como sigue la distribución de probabilidad cuya forma tiende a la correspondiente a la distribución Poisson



La ley de los grandes números indica que para una sucesión de v.a.i.i.d (X1, X2, …, Xn) con media finita μ, se satisface la siguiente convergencia

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu$$

Independientemente de la función de distribución asociada a X.

El siguiente cuadro de resultados corrobora esta ley, a medida que el tamaño de la muestra crece, la media muestral efectivamente se acerca a la media poblacional μ = 35

```
Muestras :  [10^2, 10^3, 10^4, 10^5]
Medianas :  [30.0, 30.0, 30.0, 30.0]
Modas    :  [27, 32, 29, 29]
Medias   :  [30.77, 29.992, 29.9565, 29.99693]
μ =  30
Varianzas:  [30.637099999999997, 28.645936000000003, 29.71560775,
30.2923005751]
σ =  30
```

En el caso de la varianza, no observa una convergencia estricta para los grandes números ya que la velocidad de convergencia para la media es relativamente lenta.