Tarea 3
Probabilidad y Estadística aplicada

Sebastián Decuadro

Integrantes:
Facundo Acosta 49721889,
Nicolás Cartalla 48613239,
Rodrigo Jauregui 44493683

Fecha:
03/06/2023

Índice

Introducción

En esta tarea trabajaremos con tres distribuciones discretas, las cuales son:

- Distribución binomial
  En esta distribución se trabajara con el valor de los parámetros n = 100 y p = 0.35

- Distribución geométrica
  En esta distribución se trabajara con el valor del parámetro p = 0.08

- Distribución poisson
  En esta distribución se trabajara con el valor del parámetro λ = 30

Para cada una de ellas, el objetivo base es simular variables aleatorias discretas usando Python para responder ciertas preguntas, con la ayuda obtención de muestras para hacer estadística descriptiva y una verificación basada en la ley de los grandes números. Para esto nos ayudaremos usando una librería que nos brinda Python llamada scipy.stats.

a) Generar muestras aleatorias de tamaños 100 , 1000 , 10000 y 100000 .

b) Hacer un diagrama de cajas para cada una de las muestras generadas en la parte anterior. ¿Existen datos atípicos en las muestras?

c) Realizar un histograma de las muestras generadas.

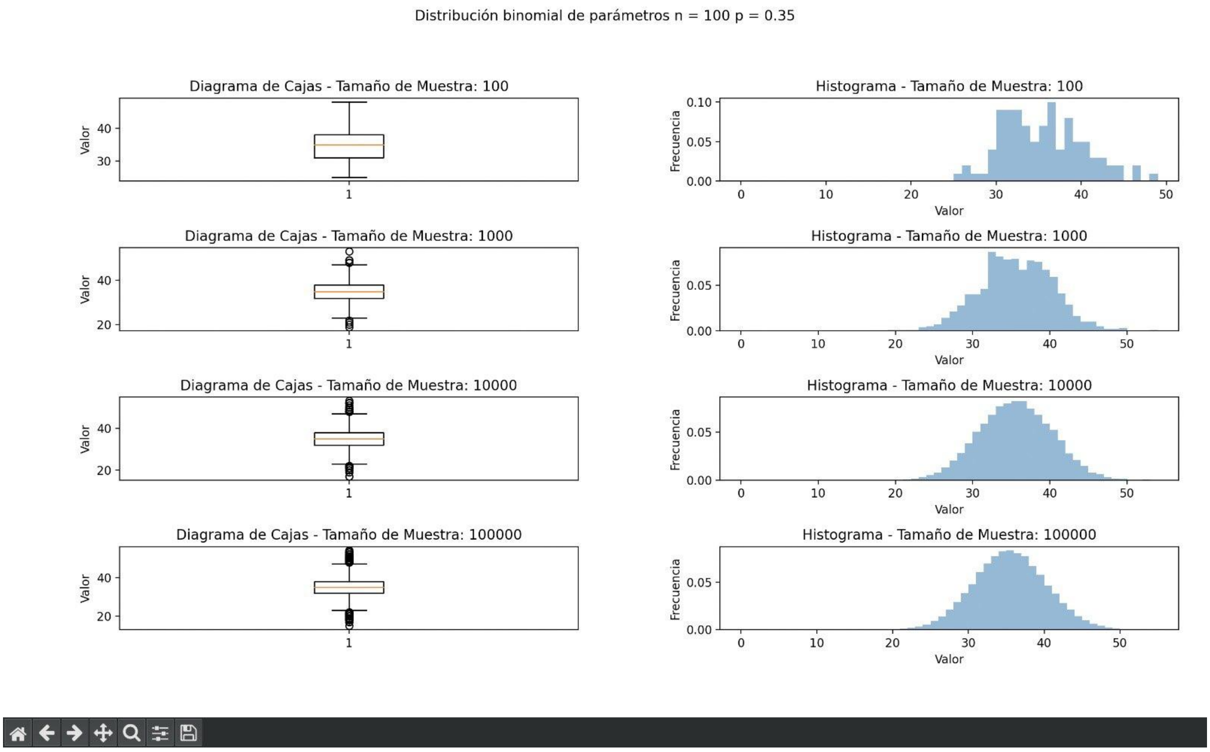
d) Hallar la mediana y la moda de cada muestra.

e) Hallar la media empíırica de cada muestra y compararla con la esperanza teórica de la distribución 3. ¿Qué se puede observar en las muestras más grandes?

f) Hallar la varianza empírica de cada muestra y compararla con la varianza teórica de la distribución 3. ¿Qué se puede observar en las muestras más grandes?

Diagrama de Cajas e Histogramas

Distribución binomial



1.

b) ¿Existen datos atípicos en las muestras?

Con un tamaño de muestra de 100, no aparece ningún dato atípico. Al aumentar el tamaño de la muestra a 1000 empiezan a aparecer algunos pocos datos atípicos. Este incremento se mantiene al incrementar el tamaño de muestra a 10000 y 100000, sugiriendo que la cantidad de datos atípicos incrementa junto al tamaño de la muestra.

d) Mediana y Moda

- Muestra 100
  - Mediana: 35
  - Moda: 32
- Muestra 1000
  - Mediana: 35
  - Moda: 34,74
- Muestra 10000
  - Mediana: 35
  - Moda: 34
- Muestra 100000
  - Mediana: 35

Moda: 35

e) Media Empírica y Esperanza Teórica

- Muestra 100
    - Media Empírica: 35,07
    - Esperanza Teórica: 35
- Muestra 1000
    - Media Empírica: 34,74
    - Esperanza Teórica: 35
- Muestra 10000
    - Media Empírica: 34,97
    - Esperanza Teórica: 35
- Muestra 100000
    - Media Empírica: 35
    - Esperanza Teórica: 35

Comparación entre la media y la esperanza:

La media y la esperanza son dos medidas de tendencia central que pueden ser utilizadas para describir una distribución de probabilidad. En una distribución binomial, la media y la esperanza son iguales y están dadas por el producto de la probabilidad de éxito (p) y el tamaño de la muestra (n). En este caso, p = 0.35 y n varía en cada muestra.

Observaciones en las muestras más grandes:

A medida que el tamaño de la muestra aumenta, la media y la esperanza se mantienen cercanas al valor teórico de 35. Esto es consistente con las propiedades de la distribución binomial y la ley de los grandes números. La ley de los grandes números establece que, a medida que el tamaño de la muestra se vuelve más grande, la media de la muestra tiende a aproximarse a la media poblacional o esperanza.

En todas las muestras analizadas, la media se encuentra muy cerca de la esperanza de 35. Esto sugiere que, en promedio, la proporción de éxitos en cada muestra se aproxima a la probabilidad de éxito teórica (p = 0,35) multiplicada por el tamaño de la muestra (n). Estas muestras más grandes proporcionan estimaciones más precisas de la media y la esperanza de la distribución binomial.

En resumen, al aumentar el tamaño de la muestra, la media y la esperanza se acercan cada vez más al valor teórico. Esto indica que las estimaciones de la proporción de éxitos en la distribución binomial se vuelven más precisas a medida que se utilizan muestras más grandes.

f) Varianza Empírica y Varianza Teórica

- Muestra 100
    - Teórica: 22,75
    - Empírica: 28,09

- Muestra 1000
    - Teórica: 22,75
    - Empírica: 22,41
- Muestra 10000
    - Teórica: 22,75
    - Empírica: 22,63
- Muestra 100000
    - Teórica: 22,75
    - Empírica: 22,79

Comparación entre la varianza empírica y la varianza teórica:

- En todas las distribuciones, la varianza teórica tiene un valor constante de 22,75, mientras que la varianza empírica varía en cada muestra.
- La varianza teórica es un parámetro calculado a partir de la fórmula de la varianza para la distribución binomial, mientras que la varianza empírica es una estimación de la variabilidad basada en los datos de la muestra.
- En general, esperamos que la varianza empírica se acerque a la varianza teórica a medida que aumenta el tamaño de la muestra, debido a la ley de los grandes números.

Observaciones en las muestras más grandes:

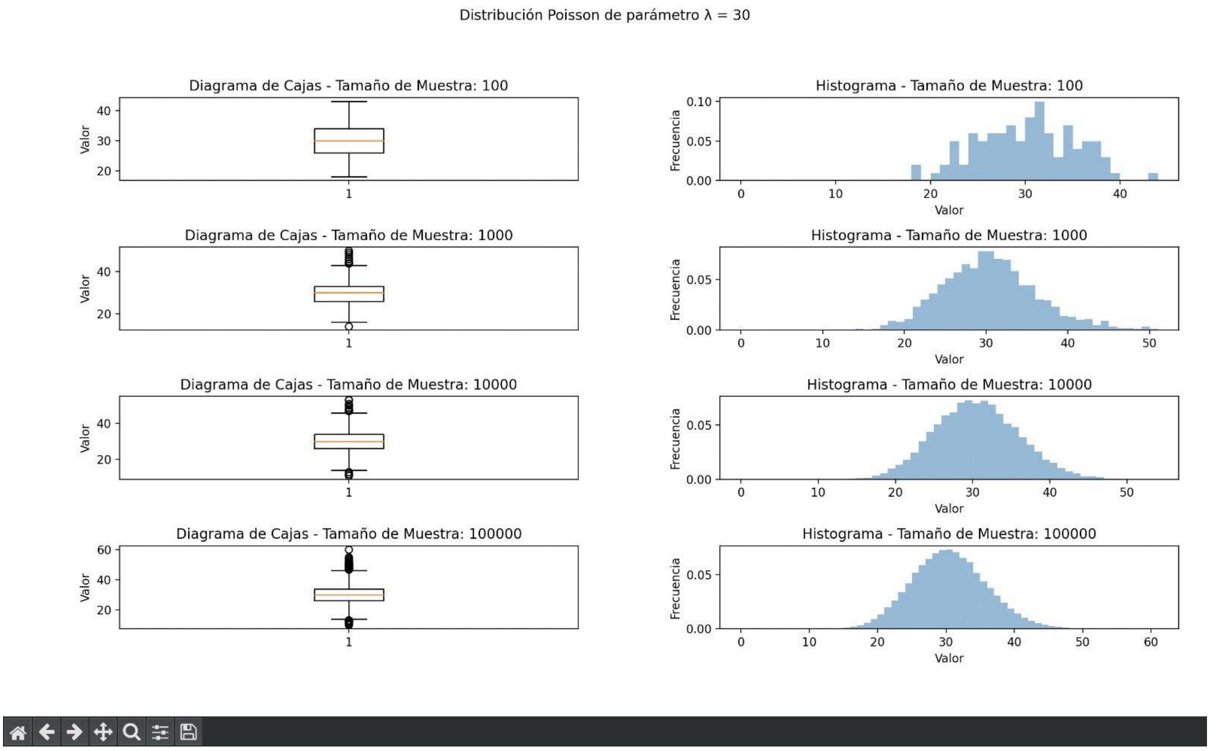
A medida que aumenta el tamaño de la muestra, la varianza empírica tiende a acercarse a la varianza teórica.

En las muestras más pequeñas (tamaño de muestra de 100), la varianza empírica puede diferir ligeramente de la varianza teórica, ya que las estimaciones basadas en una muestra pequeña pueden tener más variabilidad y ser menos precisas.
Sin embargo, a medida que el tamaño de la muestra aumenta a 1000, 10000 y 100000, la varianza empírica se aproxima cada vez más a la varianza teórica.

Esto se debe a que, con muestras más grandes, se obtiene más información y se reduce el efecto de la variabilidad de la muestra en la estimación de la varianza. La ley de los grandes números garantiza que a medida que aumenta el tamaño de la muestra, la varianza empírica converge a la varianza teórica.

En resumen, en las muestras más grandes, podemos observar una mayor precisión en la estimación de la varianza empírica, que se acerca cada vez más a la varianza teórica esperada.

<u>Distribución Geométrica</u>



2.

b) ¿Existen datos atípicos en las muestras?

La distribución geométrica parece reflejar el mismo incremento en cantidad de datos atípicos proporcional al incremento del tamaño de la muestra a como fue visto en la distribución binomial. Un tamaño de muestra de 100 no resulta en ningún dato atípico, y al incrementar el tamaño de la muestra a 1000, 10000 y 100000, empiezan a aparecer más datos atípicos.

d) Mediana y Moda

- Muestra 100
    - Mediana: 10
    - Moda: 1
- Muestra 1000
    - Mediana: 9
    - Moda: 1
- Muestra 10000
    - Mediana: 9
    - Moda: 1
- Muestra 100000
    - Mediana: 9

    Moda: 1

e) Media Empírica y Esperanza Teórica

- Muestra 100
    - Media Empírica: 12,77
    - Esperanza Teórica: 11,5
- Muestra 1000
    - Media Empírica: 12,34
    - Esperanza Teórica: 11,5
- Muestra 10000
    - Media Empírica: 12,39
    - Esperanza Teórica: 11,5
- Muestra 100000
    - Media Empírica: 35
    - Esperanza Teórica: 11,5

Comparación entre la media y la esperanza:

Podemos comparar la media y la esperanza en diferentes tamaños de muestra para la distribución geométrica con un parámetro p = 0.08. La esperanza de una distribución geométrica se calcula como E(X) = 1/p, donde p es el parámetro de probabilidad. En este caso, la esperanza es 11,5, lo cual es consistente en todos los tamaños de muestra.

Observaciones en las muestras más grandes:

A medida que el tamaño de la muestra aumenta, la media tiende a acercarse más a la esperanza. Esto se debe al efecto de la Ley de los Grandes Números, que establece que a medida que el tamaño de la muestra aumenta, la media muestral se aproxima a la media poblacional.

En este caso, se puede observar que a medida que el tamaño de la muestra aumenta de 100 a 100000, la media varía ligeramente pero se mantiene cercana a la esperanza de 11,5. Esto sugiere que la media muestral es un estimador confiable de la media poblacional en esta distribución geométrica con un parámetro p = 0,08.

f) Varianza Empírica y Varianza Teórica

- Muestra 100
    - Teórica: 143,75
    - Empírica: 126,63
- Muestra 1000
    - Teórica: 143,75
    - Empírica: 159,19
- Muestra 10000
    - Teórica: 143,75
    - Empírica: 143,17
- Muestra 100000
    - Teórica: 143,75
    - Empírica: 142,47

Comparación entre varianza Empírica y Teórica:

Podemos comparar la varianza teórica y la varianza empírica en diferentes tamaños de muestra para la distribución geométrica con un parámetro p = 0.08. La varianza teórica de una distribución geométrica se calcula como Var(X) = (1-p) / p^2, donde p es el parámetro de probabilidad. En este caso, la varianza teórica es 143.75, la cual es constante para todos los tamaños de muestra.

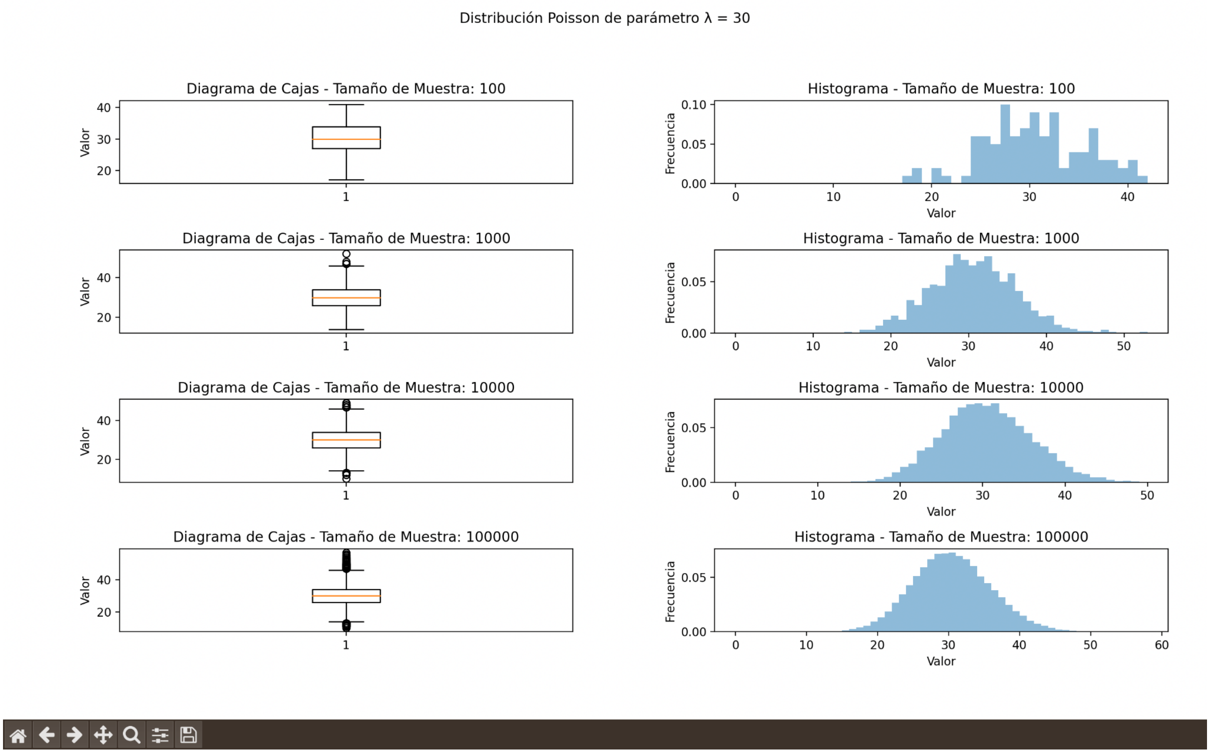
Observaciones en las muestras más grandes:

Podemos observar que la varianza empírica tiende a acercarse a la varianza teórica a medida que el tamaño de la muestra aumenta. Sin embargo, en algunos casos, como en el tamaño de muestra de 10000, la varianza empírica puede ser ligeramente mayor que la varianza teórica.

Este comportamiento se debe a las fluctuaciones naturales en los datos muestrales y a la precisión limitada de la estimación de la varianza basada en una muestra finita. A medida que el tamaño de la muestra se hace cada vez más grande, las estimaciones de la varianza empírica se vuelven más precisas y se acercan a la varianza teórica.

En resumen, con un tamaño de muestra más grande, se espera que la varianza empírica se acerque más a la varianza teórica, lo que indica una mejor estimación de la variabilidad en la distribución geométrica con un parámetro p = 0.08.

Distribución Poisson



3. **b) ¿Existen datos atípicos en las muestras?**

La misma tendencia de incremento de cantidad de datos atípicos junto al incremento del tamaño de la muestra que fue vista en la anteriores distribuciones aparece en los diagramas de caja de la distribución de Poisson. Tamaño de muestra de 100 muestra ausencia de datos atípicos, pero al pasar a un tamaño de muestra de 1000 empiezan a aparecer unos pocos. Este número incrementa al pasar a un tamaño de muestra de 10000, y vuelve a incrementar con 100000.

d) Mediana y Moda

- Muestra 100
    - Mediana: 29,5
    - Moda: 25
- Muestra 1000
    - Mediana: 30
    - Moda: 30
- Muestra 10000
    - Mediana: 30
    - Moda: 28
- Muestra 100000
    - Mediana: 30
    Moda: 29

e) Media Empírica y Esperanza Teórica

- Muestra 100
    - Media Empírica: 29,52
    - Esperanza Teórica: 30
- Muestra 1000
    - Media Empírica: 30
    - Esperanza Teórica: 30
- Muestra 10000
    - Media Empírica: 29,90
    - Esperanza Teórica: 30
- Muestra 100000
    - Media Empírica: 30
    - Esperanza Teórica: 30

Comparación entre la media y la esperanza:
En el caso específico de la distribución de Poisson con parámetro λ = 30, se observa que la media varía ligeramente en diferentes tamaños de muestra, pero la esperanza se mantiene constante en 30. A medida que el tamaño de la muestra aumenta, la media tiende a acercarse más a la esperanza, siguiendo la ley de los grandes números.

Observaciones en las muestras más grandes:
Se puede observar que la esperanza se mantiene constante en todos los casos, con un valor de 30, lo cual es consistente con el parámetro λ utilizado para generar las muestras.

Sin embargo, las medias de las muestras varían ligeramente alrededor del valor de 30. Esto se debe a la naturaleza aleatoria de las muestras y a las fluctuaciones inherentes en la distribución de Poisson. A medida que el tamaño de la muestra aumenta, las medias tienden a acercarse más al valor de 30, lo que es esperado según la ley de los grandes números.

En resumen, se puede concluir que la esperanza se mantiene constante en todas las muestras, mientras que la media tiende a converger hacia el valor de la esperanza a medida que el tamaño de la muestra aumenta. Esto sugiere que la media muestral es un estimador consistente del parámetro de la distribución y se acerca al valor teórico esperado a medida que se utilizan muestras más grandes.

f) Varianza Empírica y Varianza Teórica

- Muestra 100
    - Teórica: 30
    - Empírica: 26,95
- Muestra 1000
    - Teórica: 30
    - Empírica: 30,37
- Muestra 10000
    - Teórica: 30
    - Empírica: 29,35

- Muestra 100000
    - Teórica: 30
    - Empírica: 30,11

Observaciones en las muestras más grandes:

En la primera muestra de tamaño 100, la varianza teórica es de 30, mientras que la varianza empírica es de 26,95. La varianza empírica es ligeramente menor que la varianza teórica, lo que indica que en esta muestra en particular, la dispersión de los valores observados es un poco más baja de lo esperado teóricamente.

En la segunda muestra de tamaño 1000, la varianza teórica sigue siendo de 30, mientras que la varianza empírica aumenta a 29,35. En este caso, la varianza empírica está más cerca de la varianza teórica, pero aún muestra cierta diferencia.

En la tercera muestra de tamaño 10000, la varianza teórica y la varianza empírica son prácticamente iguales, ambas alrededor de 30. Esto sugiere que a medida que el tamaño de la muestra aumenta, la varianza empírica tiende a acercarse a la varianza teórica.

Finalmente, en la muestra más grande, con un tamaño de 100000, la varianza teórica es de 30 y la varianza empírica aumenta ligeramente a 30,11. Aunque la diferencia es pequeña, indica que la varianza empírica puede fluctuar un poco más que la varianza teórica en muestras muy grandes.

En resumen, en las muestras más pequeñas, se observa cierta diferencia entre la varianza empírica y la varianza teórica. Sin embargo, a medida que el tamaño de la muestra aumenta, la varianza empírica tiende a acercarse a la varianza teórica, indicando una mayor consistencia con el valor esperado.

Ejecutar las distribuciones en Python

Para ejecutar el programa se debe de clonar el repo https://github.com/nicocartalla/PyE, en el directorio tarea3 se encuentran las 3 distribuciones (tres scripts diferentes en python), en este directorio se encuentra también el archivo requirements.txt el cual tiene todas las dependencias necesarias,

Instalar las dependencias:
pip install -r requirements.txt

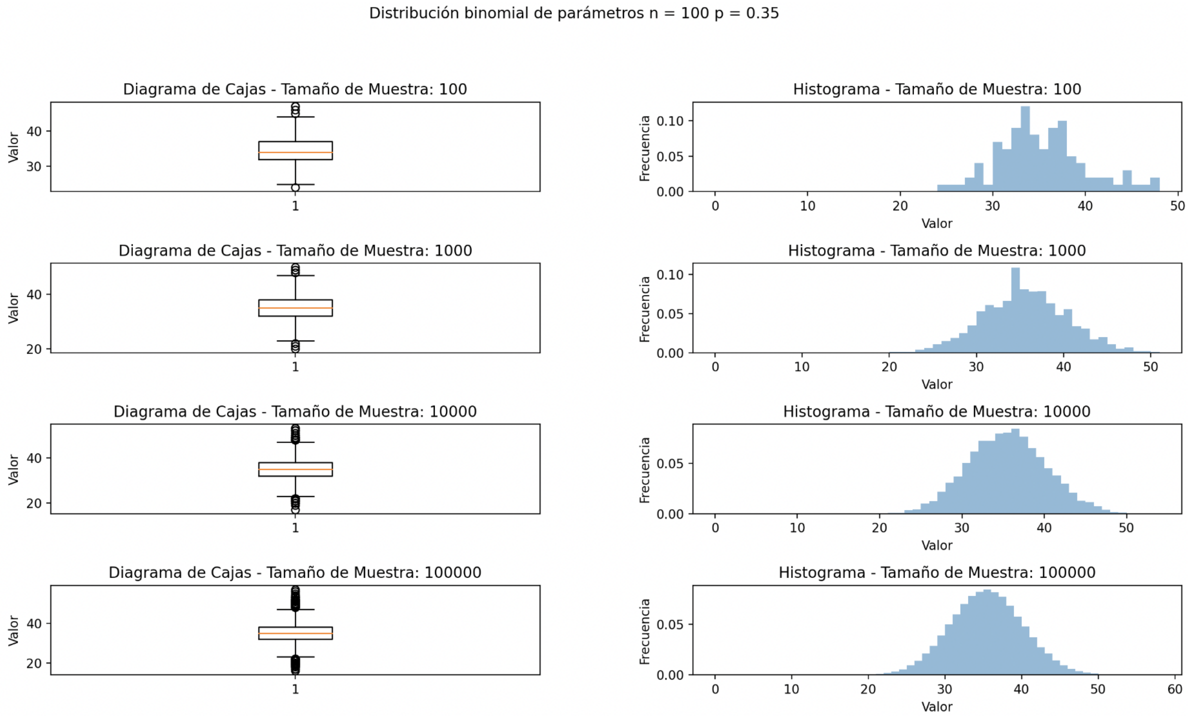
Ejecución
Para ejecutar las distribuciones, se debe ejecutar el siguiente comando:

python distribucion_<tipo-de-distribución>.py

Ejemplo con distribución binomial
python distribucione_binomial.py

Resultados

Gráficas
Se mostrará en pantalla completa (fullscreen) todas las gráficas de la distribución.



Consola

En la consola al cerrar la ventana de las gráficas se mostrarán tamaño muestra, mediana, moda, varianzas y esperanza para cada una de las muestras.

```
------- Distribución binomial de parámetros n = 100 p = 0.35 -------
Tamaño de Muestra: 100000
Mediana: 35.0
Media: 35.01417
Moda: 35
Varianza Teorica: 22.75
Varianza Empírica: 22.87
Esperanza: 35.0
-------------------------

------- Distribución binomial de parámetros n = 100 p = 0.35 -------
Tamaño de Muestra: 100000
Mediana: 35.0
Media: 35.01417
Moda: 35
Varianza Teorica: 22.75
Varianza Empírica: 22.87
Esperanza: 35.0
-------------------------

------- Distribución binomial de parámetros n = 100 p = 0.35 -------
Tamaño de Muestra: 100000
Mediana: 35.0
Media: 35.01417
Moda: 35
Varianza Teorica: 22.75
Varianza Empírica: 22.87
Esperanza: 35.0
-------------------------

------- Distribución binomial de parámetros n = 100 p = 0.35 -------
Tamaño de Muestra: 100000
Mediana: 35.0
Media: 35.01417
Moda: 35
Varianza Teorica: 22.75
Varianza Empírica: 22.87
Esperanza: 35.0
-------------------------
```