Tarea 3 - Distribuciones

Sebastián Martínez, Juan Darracq, Diego Handalian, Francisco Cabarcos

PROBABILIDAD Y ESTADÍSTICA APLICADA

08/06/2024

Introducción

En este informe, utilizaremos el concepto de variables aleatorias discretas vistas en el curso y la realización de un script Python para obtener muestras con las que hacer estadística descriptiva y una verificación empírica de la ley de los grandes números.

En primer lugar explicaremos todo el marco teórico utilizado para realizar este análisis, luego una explicación del código formulado en Python donde mostraremos los diferentes histogramas y diagramas de cajas juntos con un análisis descriptivo para cada distribución. Asimismo, responderemos a las preguntas planteadas en la tarea, en esa parte, compararemos los valores empíricos obtenidos de las muestras con los valores teóricos esperados. Primero, calcularemos la esperanza teórica para la distribución binomial, geométrica y de poisson y la compararemos con la media obtenida para los diferentes tamaños de muestra. Luego, calcularemos la varianza teórica para la distribución binomial, geométrica y de poisson y la compararemos con la varianza empírica obtenida para los diferentes tamaños de muestra. Este análisis nos permitirá observar cómo las diferentes muestras empíricas se comportan en función de los valores teóricos a medida que aumenta el tamaño de las muestras.

Finalmente, concluiremos con una verificación empírica de la ley de los grandes números basada en todo el análisis previo.

Marco Teórico

Variables Aleatorias

Asignación numérica a cada posible resultado de un experimento aleatorio. Por ejemplo, el número de caras obtenidas al lanzar seis veces una moneda.

Distribución Geométrica

Modelado basado en repetir de forma independiente un experimento de tipo “éxito-fracaso” hasta que ocurre el primer éxito. Por ejemplo, el número de lanzamientos necesarios hasta obtener la primera “cara” en una moneda. Su fórmula es $P(X = k) = p(1 - p)^{k-1}$ donde k es la cantidad de intentos y p la probabilidad de éxito. Tiene recorrido N ∪ {0} y se dice que tiene distribución geométrica de parámetro p.

Distribución Binomial

Modelado basado en un número de éxitos sobre una cantidad fija de experimentos de tipo "éxito-fracaso" independientes, donde cada experimento cuenta con la misma probabilidad de éxito. Por ejemplo, probabilidad de conseguir 3 caras en 10 lanzamientos. Su fórmula es $P(k, n; p) = (\frac{n}{k})p^{k}(1 - p)^{n-k}$ donde p es la probabilidad de éxito, n es la cantidad de intentos y k la cantidad de éxitos. El recorrido de una distribución de Bernoulli es Rec(X) = {0,1} y de la binomial es Rec(X) = {0,1,2,...,n}.

Distribución de Poisson

Modelado basado en contar la cantidad de ocurrencias de un evento en un intervalo de tiempo. Por ejemplo, la cantidad de personas que entran a una tienda en tres horas. Su fórmula es $P(X = k) = e^{-k}(\frac{\lambda^{k}}{k!})$ donde k es la cantidad de éxitos y lambda es el promedio de ocurrencia del evento. Su recorrido es Rec(X) = N

Mediana

El valor obtenido al sumar todos los resultados y dividirlos sobre la cantidad, representa el punto de equilibrio de los datos.

Moda

Dato que ocurre con mayor frecuencia

Esperanza

La esperanza se suele llamar media teórica. Para calcularla solo se necesita información sobre la distribución de la variable aleatoria (recorrido y función de probabilidad de masa).En síntesis, es una medida de tendencia central de una distribución de probabilidad.

- Esperanza teórica Binomial
  - E(X) = $n \times p$
- Esperanza teórica Geométrica
  - E(X) = $\frac{1}{p}$
- Esperanza teórica Poisson
  - E(X) = λ

Varianza

La varianza teórica es el promedio teórico de la distancia al cuadrado entre una variable aleatoria y su centro

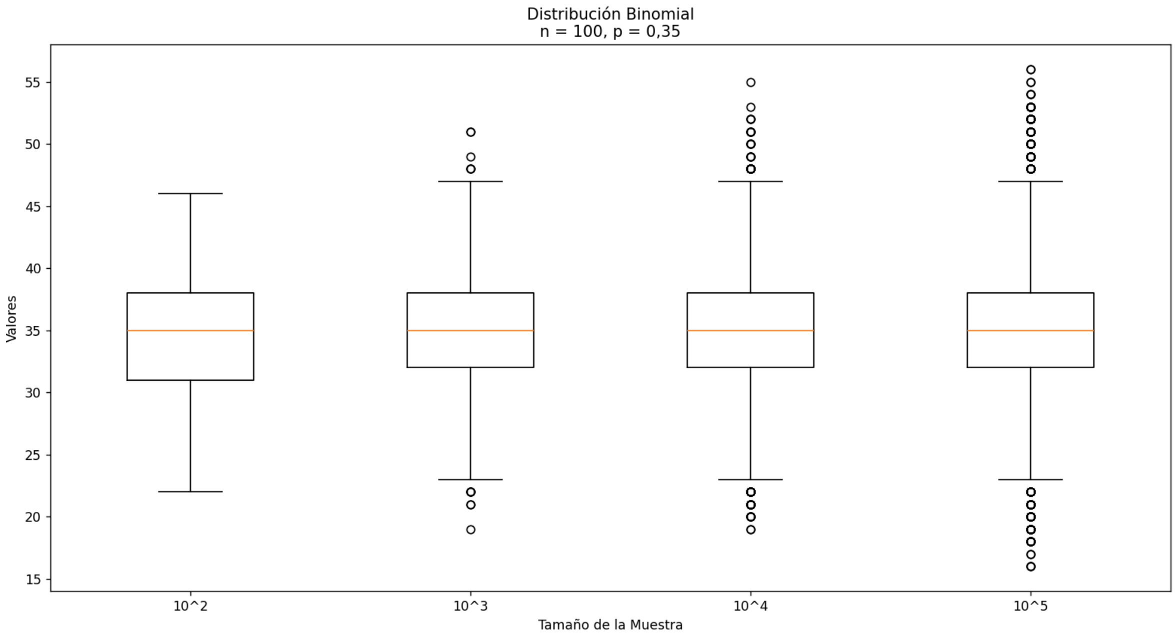
- Varianza teórica Binomial
    - Var(X) = $n \times p \times (1 - p)$
- Varianza teórica Geométrica
    - Var(X) = $\frac{1-p}{p^2}$
- Varianza teórica Poisson
    - Var(X) = λ

<u>Ley de los grandes números</u>

La ley de los grandes números es una de las leyes fundamentales de la teoría de la probabilidad. Esta ley dice que, si repetimos un experimento aleatorio muchas veces, el resultado promedio se acercará cada vez más al valor esperado teórico.

Distribución Binomial

Estadística Descriptiva



<u>Tamaño 10^2:</u>

Caja: La caja se extiende aproximadamente desde 31 hasta 38, con la mediana en alrededor de 35.

Bigotes:

- Inferior se extiende hasta aproximadamente 22

- Superior se extiende hasta aproximadamente 46.

Datos atípicos: No hay datos atípicos, ya que no se observan puntos fuera de los bigotes.

Tamaño 10^3:

Caja: La caja se extiende aproximadamente desde 32 hasta 38, con la mediana en alrededor de 35.

Bigotes:

- Inferior se extiende hasta aproximadamente 23.
- Superior se extiende hasta aproximadamente 47.

Datos atípicos: Se observan 3 puntos por debajo del bigote inferior y 3 puntos por encima del bigote superior, indicando datos atípicos bajos y datos atípicos altos, con valores aproximados a 19, 21 y 22, y con valores aproximados 48, 49 y 51 respectivamente.

Tamaño 10^4:

Caja:La caja se extiende aproximadamente desde 32 hasta 38, con la mediana en alrededor de 35.

Bigotes:

- Inferior se extiende hasta aproximadamente 23.
- Superior se extiende hasta aproximadamente 47.

Datos atípicos: Se observan 4 puntos por debajo del bigote inferior y 7 puntos por encima del bigote superior, indicando datos atípicos bajos y datos atípicos altos, con valores aproximados a 19, 20, 21 y 22, y con valores aproximados 48, 49, 50, 51, 52, 53 y 55 respectivamente.

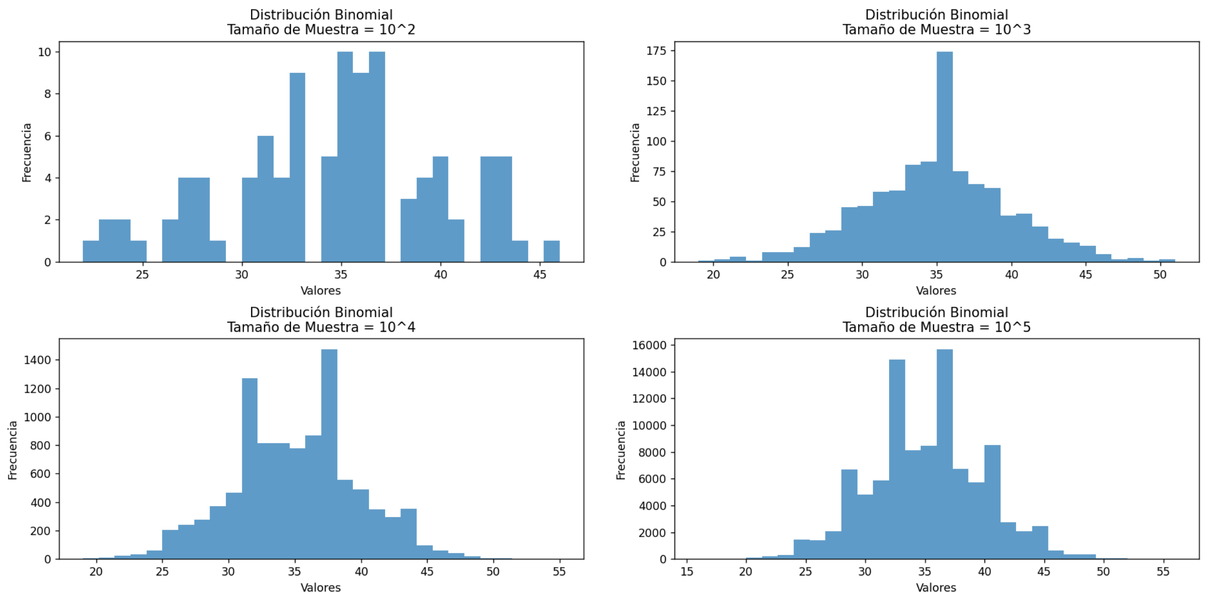
Tamaño 10^5:

Caja: La caja se extiende aproximadamente desde 32 hasta 38, con la mediana en alrededor de 35.

Bigotes:

- Inferior se extiende hasta aproximadamente 23.
- Superior se extiende hasta aproximadamente 47.

Datos atípicos: Se observan 7 puntos por debajo del bigote inferior y 9 puntos por encima del bigote superior, indicando datos atípicos bajos y datos atípicos altos, con valores aproximados a 16, 17, 18, 19, 20, 21 y 22, y con valores aproximados 48, 49, 50, 51, 52, 53, 54, 55 y 56 respectivamente.



En general, podemos observar en este histograma que para la muestras de tamaño 10^2 se puede ver una distribución uniforme.Y para las muestras de tamaño 10^3, 10^4, 10^5 una distribución de tipo acampanada .Este histograma muestra un claro agrupamiento alrededor de un valor central igual a 35. Esta concentración indica una tendencia de los valores hacia este punto específico, y esta tendencia se vuelve más pronunciada a medida que aumenta el tamaño de la muestra.

Comparación de Valores

a)

Utilizando la fórmula de esperanza teórica para la distribución binomial definida en el marco teórico, según los parámetros n=100 y p=0,35 la esperanza teórica es igual a 35. En la simulación, podemos notar que a medida que el tamaño de las muestras aumenta, las medias empíricas se acercan cada vez más a esta esperanza teórica.

Por ejemplo:

La media para el tamaño 10^2 es igual a 35,42

La media para el tamaño 10^3 es igual a 34,915

La media para el tamaño 10^4 es igual a 34,9853

La media para el tamaño 10^5 es igual a 34,96247

b)

Utilizando la varianza teórica para la distribución binomial expuesta en el marco teórico, obtenemos una varianza teórica de 22.75. A continuación, presentaremos los resultados de una simulación con las varianzas empíricas para cada muestra:

Varianzas empíricas obtenidas para los diferentes tamaños de muestra.

La varianza para el tamaño 10^2 es igual a 19,990000000000002

La varianza para el tamaño 10^3 es igual a 21,897404

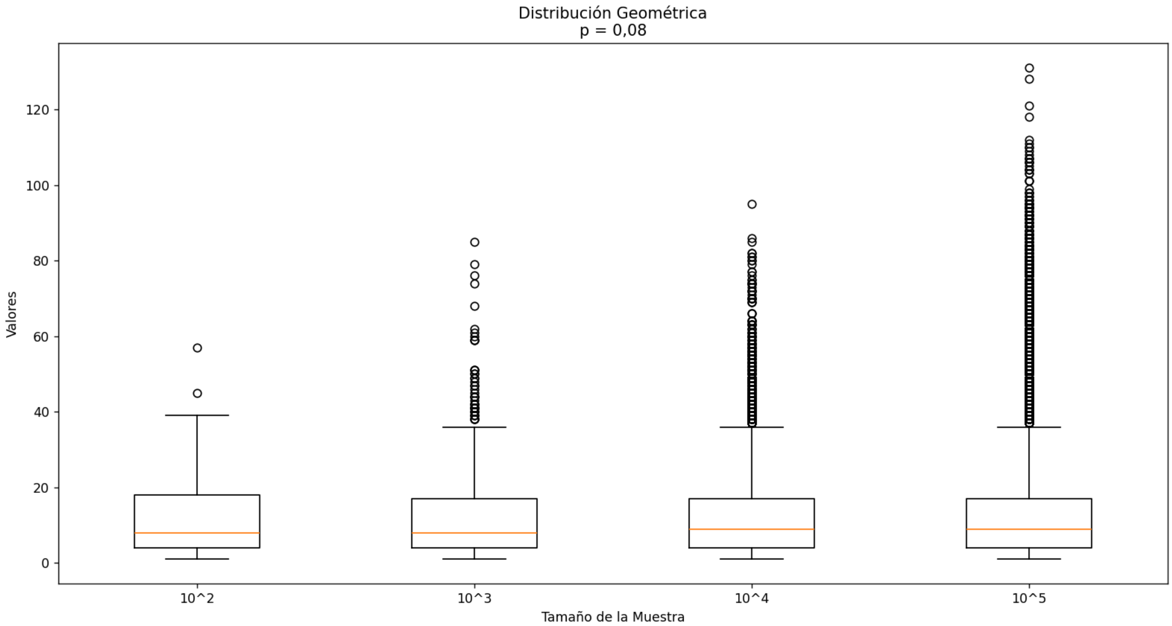
La varianza para el tamaño 10^4 es igual a 22,18479424

La varianza para el tamaño 10^5 es igual a 22,7325699039

En la simulación, podemos observar que a medida que el tamaño de las muestras aumenta, las varianzas empíricas se acercan progresivamente a esta varianza teórica.

Distribución Geométrica

Estadística Descriptiva



<u>Tamaño 10^2:</u>

Caja: La caja se extiende aproximadamente desde 4 hasta 18, con la mediana en alrededor de 8.

Bigotes:

- Inferior se extiende hasta aproximadamente 1.

- Superior se extiende hasta aproximadamente 39.

Datos atípicos: Se observan 2 puntos por encima del bigote superior, indicando datos atípicos altos, con valores aproximados a 45 y 57. No hay datos atípicos bajos ya que no hay puntos debajo del bigote inferior.

Tamaño 10^3:

Caja: La caja se extiende aproximadamente desde 4 hasta 17, con la mediana en alrededor de 8.

Bigotes:

- Inferior se extiende hasta aproximadamente 1.
- Superior se extiende hasta aproximadamente 36.

Datos atípicos: Se observan varios puntos por encima del bigote superior, indicando datos atípicos altos. No hay datos atípicos bajos ya que no hay puntos debajo del bigote inferior.

Tamaño 10^4:

Caja: La caja se extiende aproximadamente desde 4 hasta 17, con la mediana en alrededor de 9.

Bigotes:

- Inferior se extiende hasta aproximadamente 1.
- Superior se extiende hasta aproximadamente 36.

Datos atípicos: Se observan una gran cantidad de puntos por encima del bigote superior, indicando datos atípicos altos. No hay datos atípicos bajos ya que no hay puntos debajo del bigote inferior.

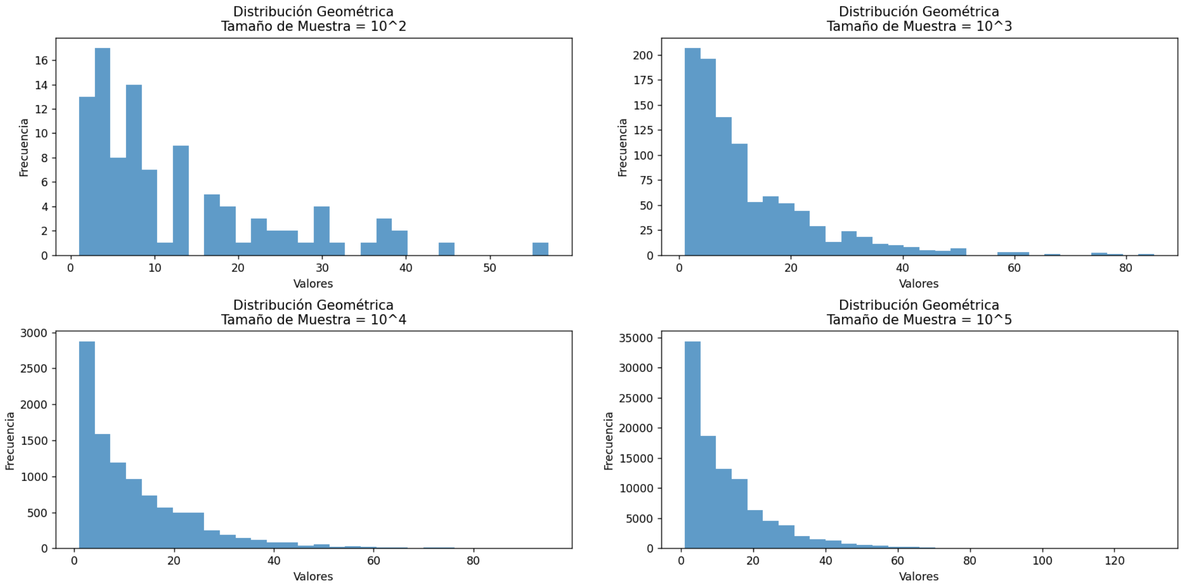
Tamaño 10^5:

Caja: La caja se extiende aproximadamente desde 4 hasta 17, con la mediana en alrededor de 9.

Bigotes:

- Inferior se extiende hasta aproximadamente 1.
- Superior se extiende hasta aproximadamente 36.

Datos atípicos: Se observan una enorme cantidad de puntos por encima del bigote superior, indicando datos atípicos altos. No hay datos atípicos bajos ya que no hay puntos debajo del bigote inferior.



En general, podemos observar en este histograma una distribución de tipo asimetría derecha en los datos. Esa asimetría se ve bien marcada en las muestras 10^3, 10^4, 10^5.

Comparación de Valores

a)

Utilizando la fórmula de esperanza teórica para la distribución geométrica definida en el marco teórico, obtenemos que, con el parámetro p=0.08, la esperanza teórica es igual a 12,5. A continuación, presentaremos los resultados de una simulación con las medias muestrales para diferentes tamaños de muestra. Las medias empíricas presentadas han sido calculadas con la función mean() de la librería numpy.

Medias empíricas obtenidas para los diferentes tamaños de muestra.

La media para el tamaño 10^2 es igual a 10.59

La media para el tamaño 10^3 es igual a 11.897

La media para el tamaño 10^4 es igual a 12.6626

La media para el tamaño 10^5 es igual a 12.52846

b)

Utilizando la varianza teórica para la distribución geométrica expuesta en el marco teórico, obtenemos una varianza teórica con el parámetro p=0.08 de 143,75 . A continuación, presentaremos los resultados de una simulación con las varianzas muestrales para cada muestra.Las varianzas empíricas presentadas han sido calculadas con la función var() de la librería numpy.

Varianzas empíricas obtenidas para los diferentes tamaños de muestra.

La varianza para el tamaño 10^2 es igual a 151.50509999999997

La varianza para el tamaño 10^3 es igual a 117.89

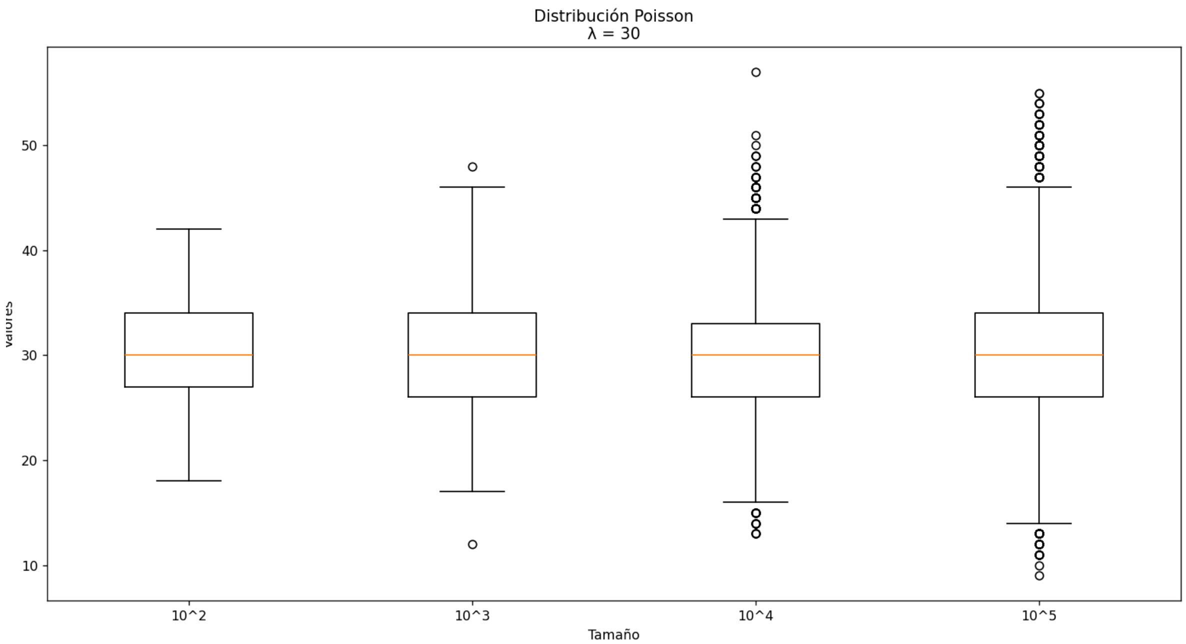
La varianza para el tamaño 10^4 es igual a 146.40854204

La varianza para el tamaño 10^5 es igual a 143.35685193749995

Observando los resultados obtenidos la esperanza empírica, podemos comprobar que a medida que el tamaño de las muestras aumenta, las varianzas empíricas se acercan progresivamente a la varianza teórica.

Distribución de Poisson

Estadística Descriptiva



<u>Tamaño 10^2:</u>

Caja: La caja se extiende aproximadamente desde 27 hasta 34, con la mediana en alrededor de 30.

Bigotes:

- Inferior se extiende hasta aproximadamente 18.
- Superior se extiende hasta aproximadamente 42.

Datos atípicos: No hay datos atípicos, ya que no se observan puntos fuera de los bigotes.

Tamaño 10^3:

Caja: La caja se extiende aproximadamente desde 26 hasta 34, con la mediana en alrededor de 30.

Bigotes:

- Inferior se extiende hasta aproximadamente 17.
- Superior se extiende hasta aproximadamente 46.

Datos atípicos: Se observan 1 punto por debajo del bigote inferior y 1 punto por encima del bigote superior, indicando un dato atípico bajo y un dato atípico alto, con valor aproximado a 12 y con valor aproximado a 48 respectivamente.

Tamaño 10^4:

Caja: La caja se extiende aproximadamente desde 26 hasta 33, con la mediana en alrededor de 30.

Bigotes:

- Inferior se extiende hasta aproximadamente 16.
- Superior se extiende hasta aproximadamente 43.

Datos atípicos: Se observan 3 puntos por debajo del bigote inferior y 9 puntos por encima del bigote superior, indicando datos atípicos bajos y datos atípicos altos, con valores aproximados a 13, 14 y 15, y con valores aproximados 44, 45, 46 , 47, 48, 49, 50, 51 y 57 respectivamente.

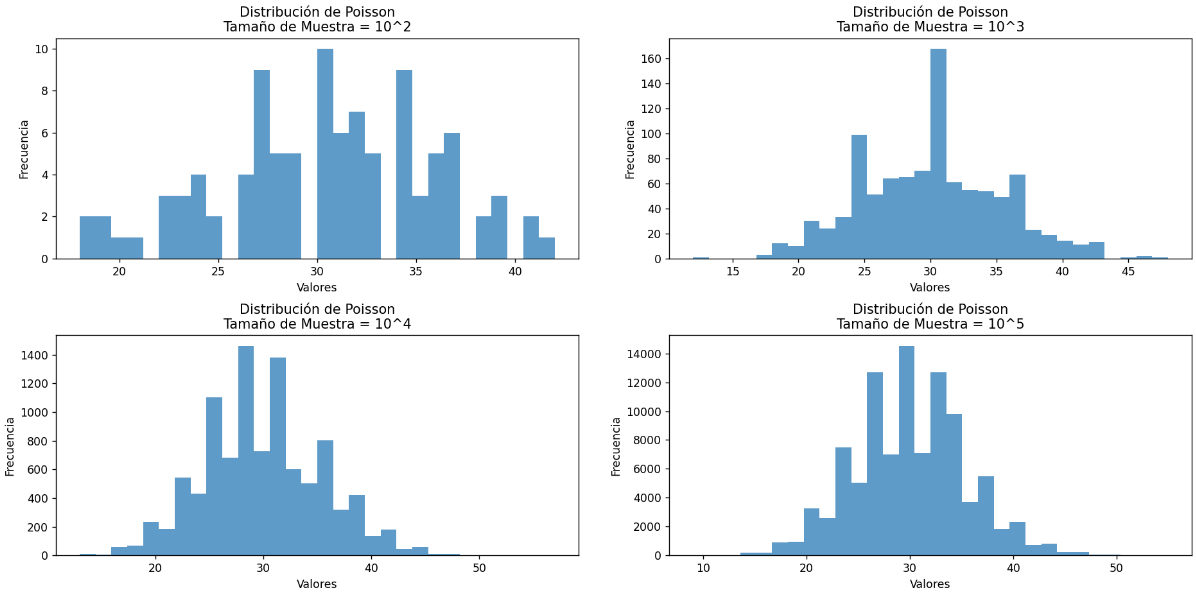
Tamaño 10^5:

Caja: La caja se extiende aproximadamente desde 26 hasta 34, con la mediana en alrededor de 30.

Bigotes:

- Inferior se extiende hasta aproximadamente 14.
- Superior se extiende hasta aproximadamente 46.

Datos atípicos: Se observan 5 puntos por debajo del bigote inferior y 9 puntos por encima del bigote superior, indicando datos atípicos bajos y datos atípicos altos, con valores aproximados a 9, 10, 11, 12, y 13, y con valores aproximados 47, 48, 49, 50, 51, 52, 53, 54 y 55 respectivamente.



En general, podemos observar en este histograma una distribución acampanada bien marcada para las muestras de tamaño 10^3, 10^4, 10^5 . Este histograma muestra un claro agrupamiento alrededor de un valor central igual a 30. Esta concentración indica una tendencia de los valores hacia este punto específico, y esta tendencia se vuelve más pronunciada a medida que aumenta el tamaño de la muestra.

Comparación de Valores

a)

Utilizando la fórmula de esperanza teórica para la distribución de poisson definida en el marco teórico, obtenemos que, con el parámetro λ=30 , la esperanza teórica es igual a 30 . A continuación, presentaremos los resultados de una simulación con las medias muestrales para diferentes tamaños de muestra. Las medias empíricas presentadas han sido calculadas con la función mean() de la librería numpy.

Medias empíricas obtenidas para los diferentes tamaños de muestra.

La media para el tamaño 10^2 es igual a 29,86

La media para el tamaño 10^3 es igual a 29,663

La media para el tamaño 10^4 es igual a 29,9756

La media para el tamaño 10^5 es igual a 29,98196

b)

Utilizando la varianza teórica para la distribución de poisson expuesta en el marco teórico, obtenemos una varianza teórica de 30. A continuación, presentaremos los resultados de una simulación con las varianzas muestrales para cada muestra.Las varianzas empíricas presentadas han sido calculadas con la función var() de la librería numpy.

Varianzas empíricas obtenidas para los diferentes tamaños de muestra.

La varianza para el tamaño 10^2 es igual a 34,360400000000006

La varianza para el tamaño 10^3 es igual a 27,095430999999998

La varianza para el tamaño 10^4 es igual a 29,81240464

La varianza para el tamaño 10^5 es igual a 29,9487545584

Observando los resultados obtenidos la esperanza empírica, podemos comprobar que a medida que el tamaño de las muestras aumenta, las varianzas empíricas se acercan progresivamente a la varianza teórica.

Conclusión

En conclusión, de acuerdo con lo expuesto y analizado a lo largo del informe, podemos destacar que, gracias a las comparaciones entre los resultados empíricos y los resultados teóricos esperados para las distribuciones binomial, geométrica y de Poisson, se ha logrado validar empíricamente la ley de los grandes números. En particular, los resultados obtenidos al comparar la media empírica con la esperanza teórica y la varianza empírica con la varianza teórica para las distintas distribuciones han sido fundamentales para esta validación. Esto proporciona una confirmación importante de que, para un número suficientemente grande de muestras, los resultados empíricos convergen hacia los valores esperados teóricamente.

Bibliografía

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_geom%C3%A9trica
https://es.wikipedia.org/wiki/Distribuci%C3%B3n_binomial
https://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_Poisson