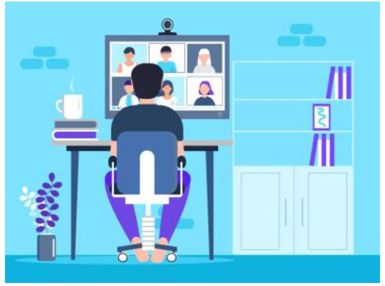
Módulo: Fundamentos de probabilidad y estadística

Agenda

Reglas del Juego

Mantener el micrófono apagado en caso de que no vayamos a hablar.

Nos encantaría verte.
Ten tu cámara encendida y conozcámonos virtualmente.



Preguntar en caso que tengan dudas.

Disfruta de este espacio.
Desconecta del resto y participa.

Por cada clase tendremos 10 min o 15 min de receso.

Modo de Evaluación

DATA ANALÍTICA

20%

Evaluación continua:

Notebooks de ejercicios, formularios de ejercicios o tareas (challenges).

30%

Test de seguimiento:

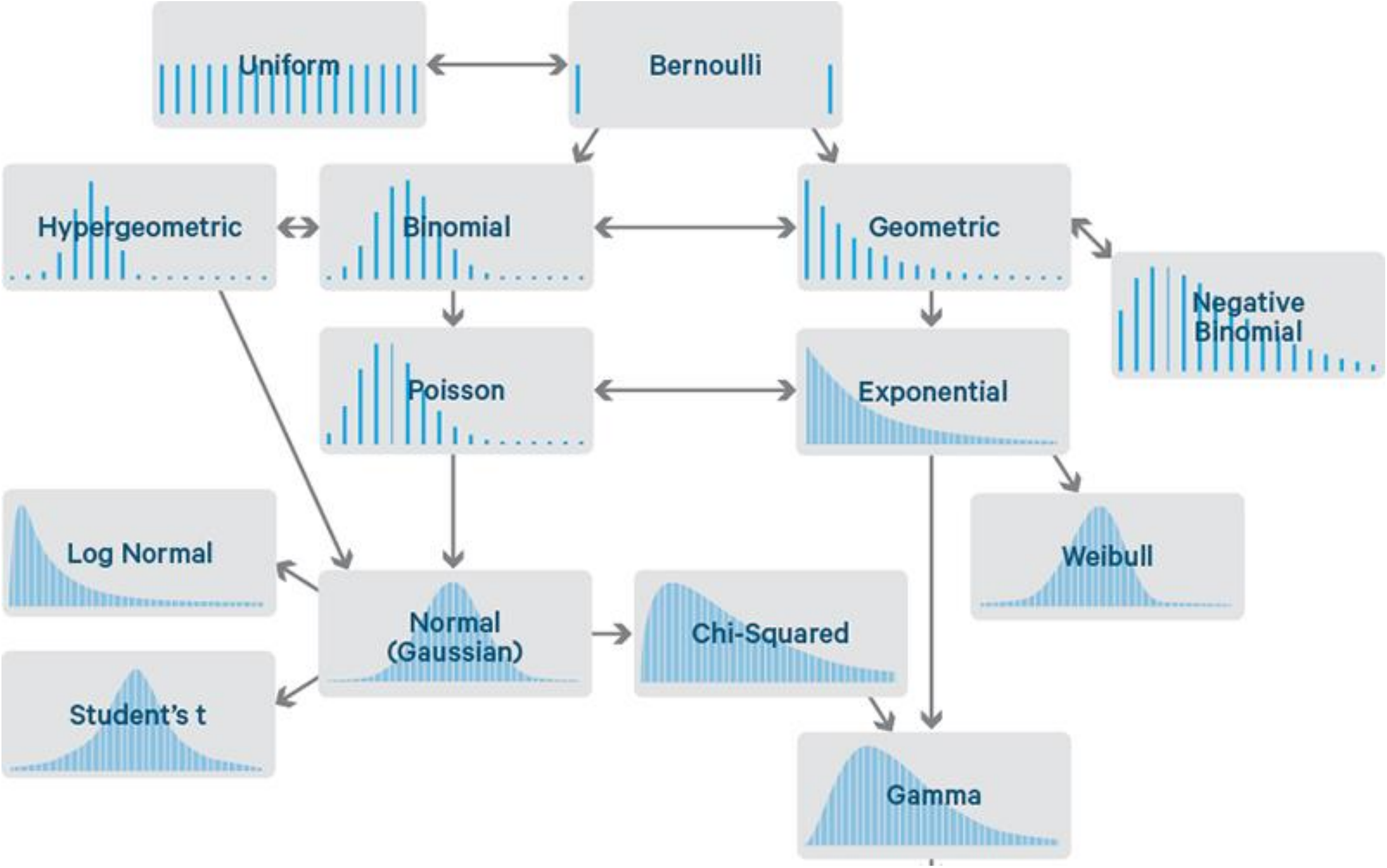
Exámenes de 5 a 10 preguntas que serán tomados al inicio de la 3ra, 5ta y 7ma sesión.

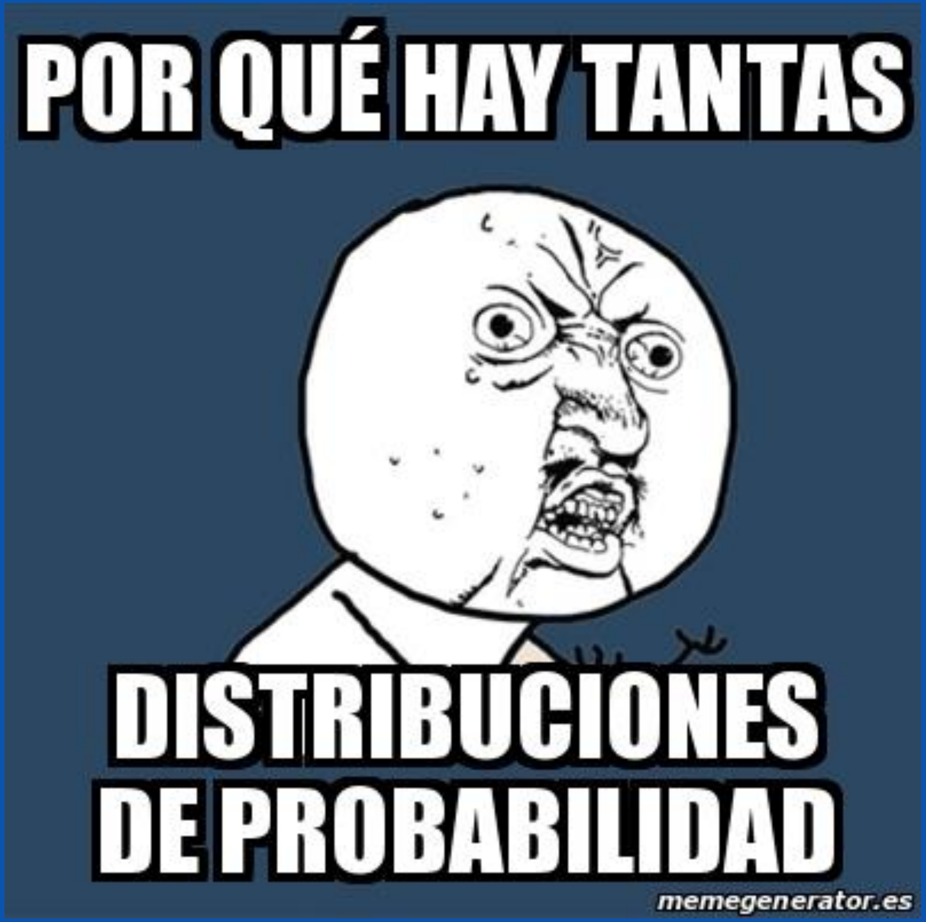
50%

Examen Final:

Examen donde evaluaremos los aprendizajes obtenidos a lo largo del curso

Para comenzar …



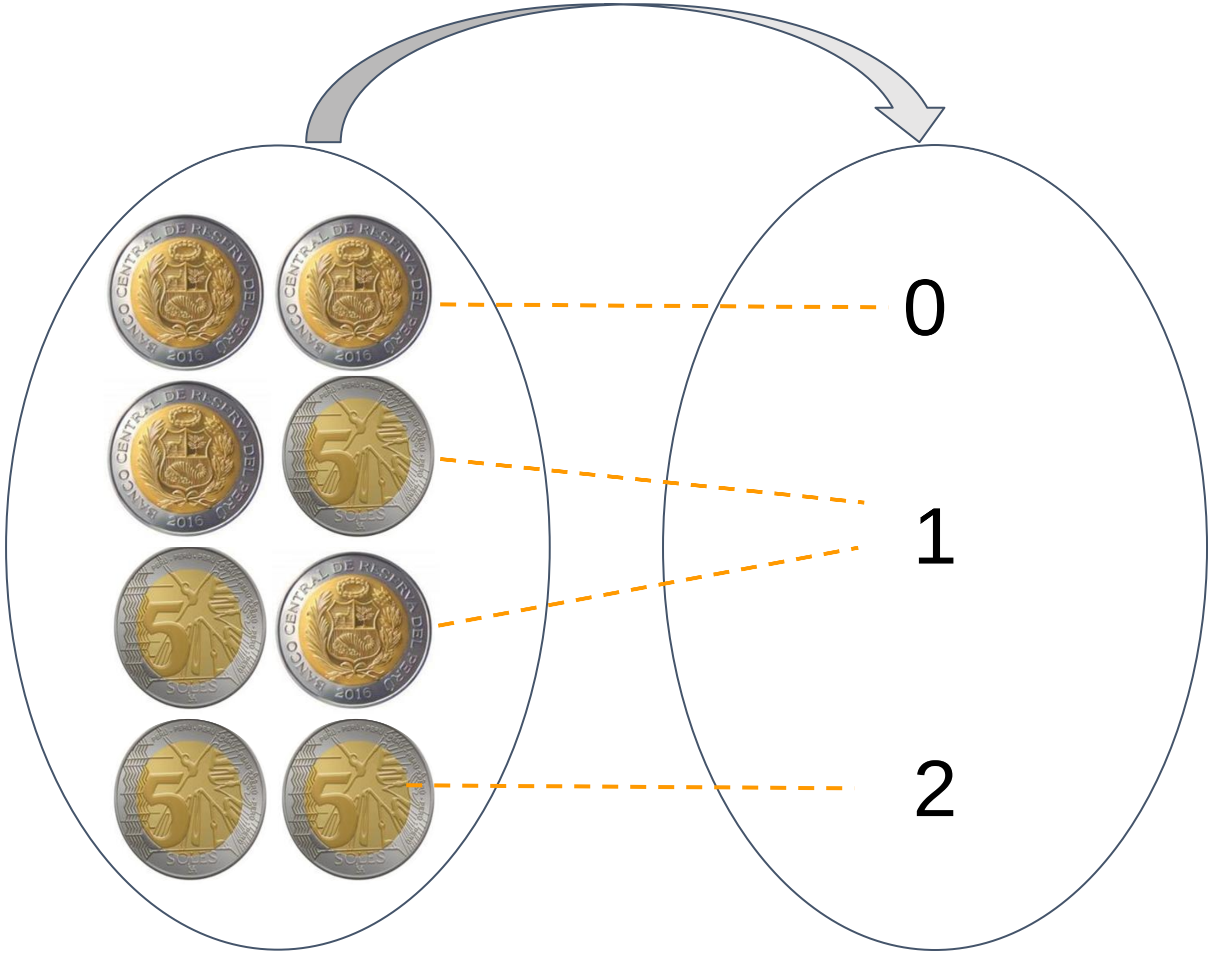

POR QUÉ HAY TANTAS
DISTRIBUCIONES DE PROBABILIDAD
memegenerator.es

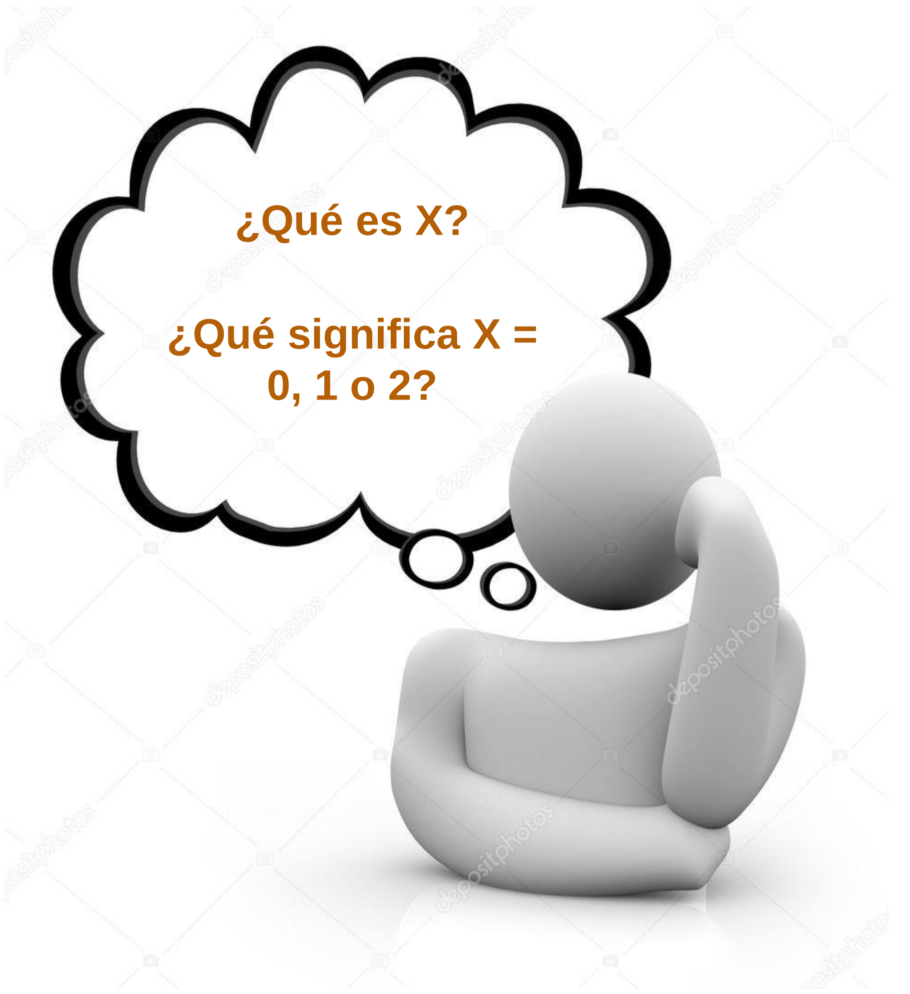
VARIABLES ALEATORIAS

Caso: Lanzamiento de dos monedas

Sea **X = Número de caras**

(fenómeno a estudiar)





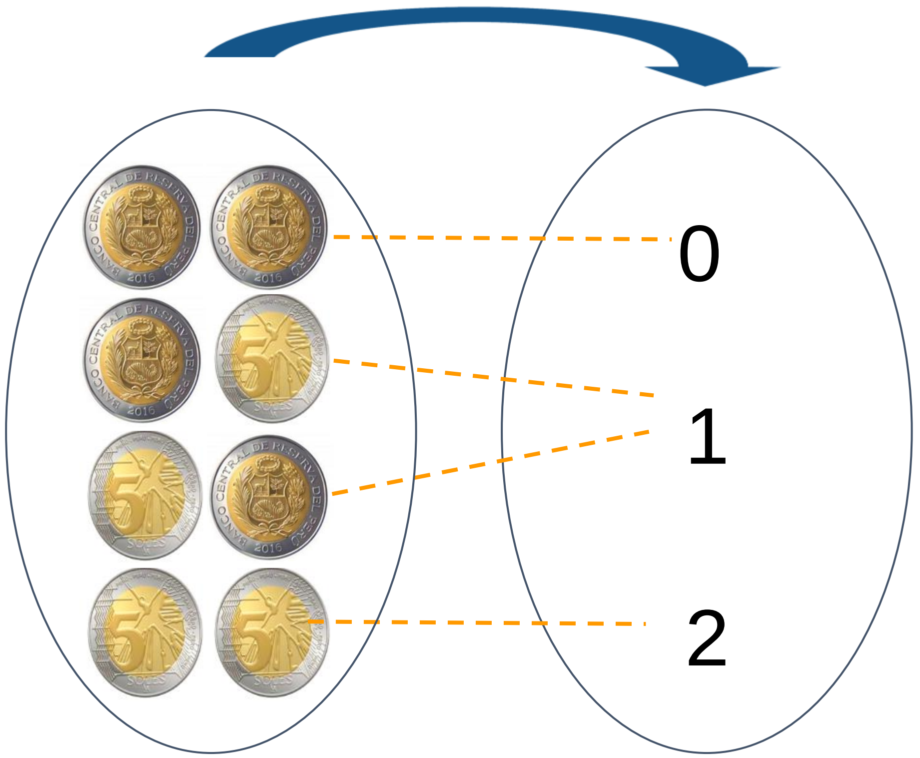
¿Qué es una variable aleatoria?

- **Definición**: Es una **función** que asigna un valor numérico, al resultado de un experimento aleatorio.

- **Representación:** Letras mayúsculas. Por ejemplo X, Y, Z, etc.

Nota: Para cada valor de X se calcula su probabilidad

Sea **X = Número de caras**



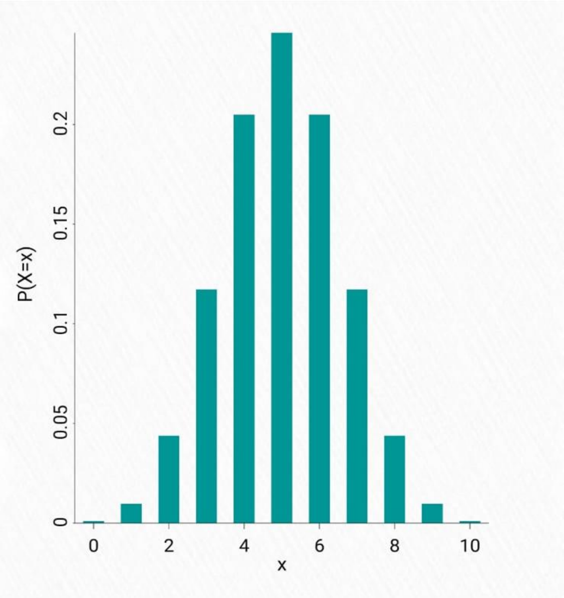
D_X = espacio muestral (S) R_X = Números reales

Tipos de variables aleatorias

DISCRETA

- Variable aleatoria que toma valores aislados o separados como 0,1,2,3,etc.

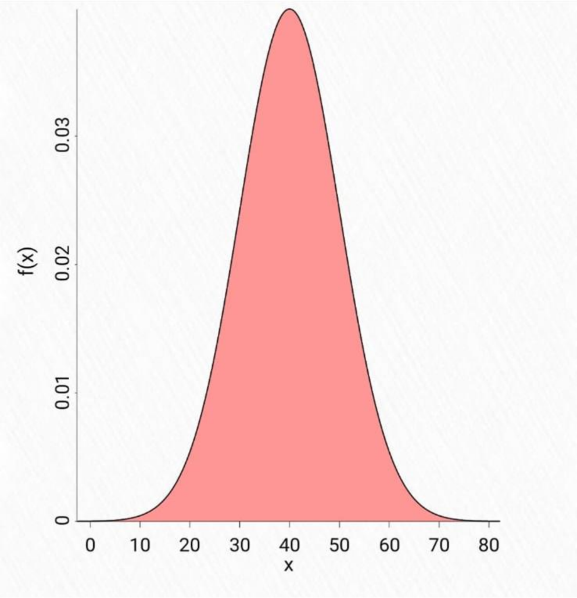
- Se relaciona con el conteo de elementos



CONTINUA

- Variable aleatoria que puede tomar valores entre dos valores dados

- Se relaciona con el tiempo, peso, longitud, precio, etc.



QUIZ TIME

Herramienta a utilizar

1er paso : Ingresar a Playstore

2do paso: Buscar la aplicación Probability Distributions

3er paso : Instalar

DISTRIBUCIONES DE PROBABILIDAD

DISCRETAS

Distribución Bernoulli

¿Cuándo utilizar la distribución bernoulli?

Para modelar **experimentos binarios**, es decir, variables que solo pueden tomar dos valores, generalmente 1 (éxito) y 0 (fracaso).

¿Cómo se define matemáticamente?

$$X \sim Ber(p)$$

p = probabilidad de éxito = P(X=1)

$$P(X = x) = p^x (1-p)^{1-x} \quad , \quad x = 0, 1$$

función masa de probabilidad (pmf)

$$E(X) = p \ \ y \ \ V(X) = 1 - p$$

esperanza (promedio) y varianza

Distribución Bernoulli - Aplicación

Regresión logística

Contexto: Supongamos que el área de riesgos desea construir un modelo de admisión de créditos hipotecarios. Para ello define su variable objetivo o target: Y = 1(cliente cae en mora), Y = 0 (cliente no cae en mora). Además se tiene una muestra de n=5000 clientes.

$$Y_1, Y_2, \cdots, Y_n$$

$$X_1^T \qquad X_2^T \qquad \cdots \qquad X_n^T$$

$$Y_i \sim Ber(\theta_i)$$

$$[0,1] \longrightarrow \theta_i = X_i^T\beta \longleftarrow \mathbb{R}$$

$$\theta_i = f(X_i^T\beta)$$

Función sigmoide

$$\theta_i = \frac{1}{1 + e^{-X_i^T\beta}}$$

$$Y_i \sim Ber\left(\frac{1}{1 + e^{-X_i^T\beta}}\right)$$

$i = 1, 2, \ldots, n$

Distribución Binomial

¿Cuándo utilizar la distribución binomial?

Para modelar la **cantidad de éxitos en una secuencia de ensayos Bernoulli** (n repeticiones). Donde la probabilidad de éxito (p) es constante.

¿Cómo se define matemáticamente?

$$X \sim Bin(n, p)$$

p = probabilidad de éxito
n = número de repeticiones Bernoulli

$$P(X = x) = \binom{n}{x} p^{x} (1-p)^{n-x} \quad , \ x \in \mathbb{N}, \ x \leq n$$

función masa de probabilidad (pmf)

$$E(X) = np \ \ y \ \ V(X) = np(1-p)$$

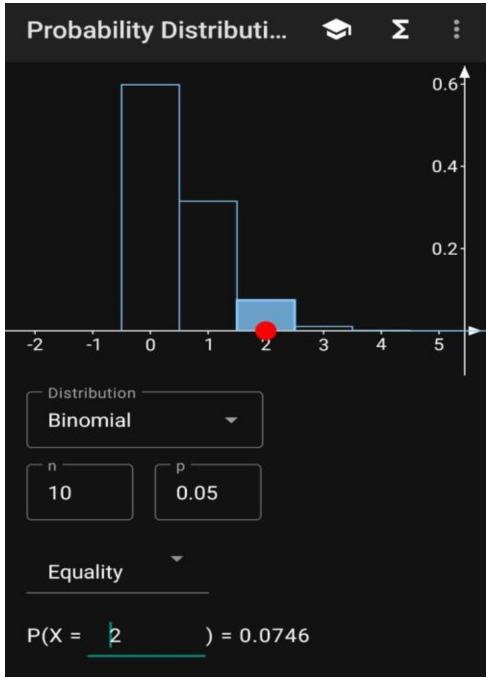
esperanza (promedio) y varianza

Distribución Binomial - Ejemplos

Supongamos que en una fábrica de bombillas se sabe que el 5% de las bombillas son defectuosas. Si se seleccionan al azar 10 bombillas, **¿cuál es la probabilidad de que exactamente 2 de ellas sean defectuosas?**

Solución:

- Éxito: bombilla defectuosa
- p = probabilidad de éxito = probabilidad defectuosas = 0.05
- n = número de ensayos bernoulli = 10
- X = número de bombillas defectuosas al seleccionar 10 bombillas.

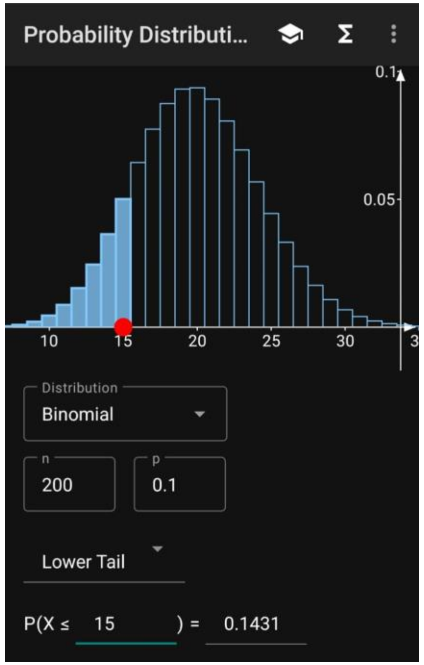


Distribución Binomial - Ejemplos

Un distribuidor de productos alimenticios sabe que el 10% de los productos que recibe están caducados. Si el distribuidor recibe un envío de 200 productos, **¿cuál es la probabilidad de que como máximo 15 de ellos estén caducados?**

Solución:

- Éxito: producto caducado
- p = probabilidad de éxito = probabilidad caducado = 0.1
- n = número de ensayos bernoulli = 200
- X = número de productos caducados al seleccionar 200 productos.



Distribución Poisson

¿Cuándo utilizar la distribución poisson?

Para modelar la **cantidad o número de eventos** que ocurren **en un intervalo de tiempo** o espacio. Donde la tasa de ocurrencia de eventos es constante.

¿Cómo se define matemáticamente?

$$X \sim Pois(\lambda)$$

λ = tasa promedio de ocurrencia en un intervalo de tiempo o espacio

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , \lambda > 0 \, , \, x = 0, 1, 2, \ldots$$

función masa de probabilidad (pmf)

$$E(X) = \lambda \;\; y \;\; V(X) = \lambda$$

esperanza (promedio) y varianza

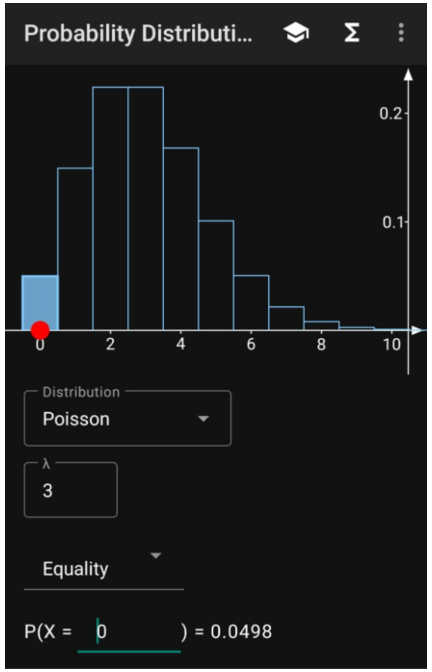
Distribución Poisson - Ejemplos

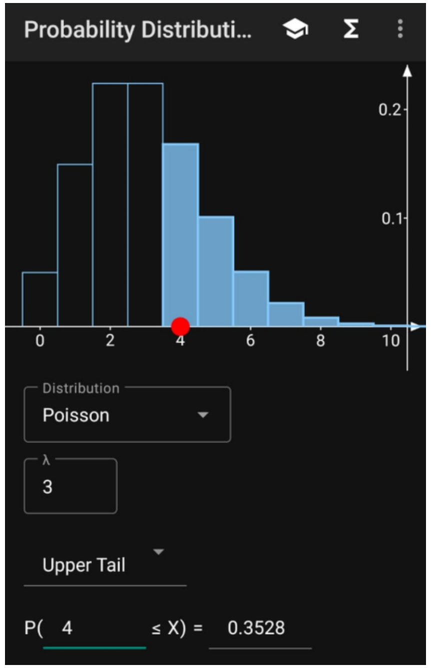
Supongamos que llegan aleatoriamente una serie de llamadas a una central telefónica con un **promedio de 3 llamadas** en intervalos de un minuto. **Calcule la probabilidad de que en un minuto cualquiera:**

a) No llegue ninguna llamada
b) Ocurran al menos 4 llamadas

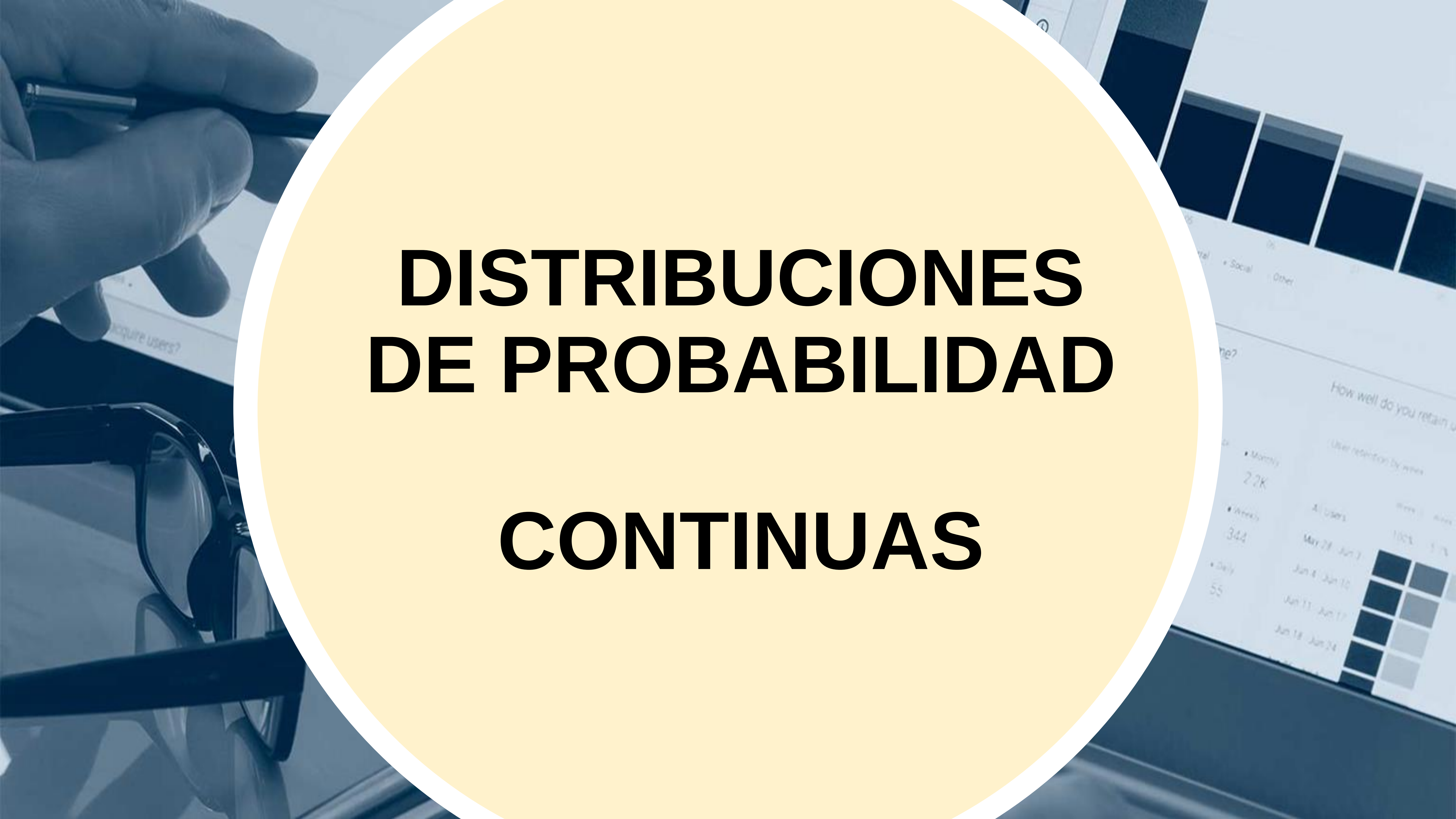
Solución:

- Unidad de tiempo: **minutos**
- λ = 3 llamadas/**1 minuto**
- X = número de llamadas que ocurren en el periodo de **un minuto**.





QUIZ TIME



DISTRIBUCIONES DE PROBABILIDAD

CONTINUAS

Distribución Exponencial

¿Cuándo utilizar la distribución exponencial?

Para modelar **tiempos de espera para la ocurrencia de cierto evento o el tiempo de vida** de dispositivos o sistemas. Donde la tasa de ocurrencia de eventos es constante

¿Cómo se define matemáticamente?

$$X \sim Exp(\lambda)$$

λ = inversa del número promedio de ocurrencias en un intervalo de tiempo o espacio. **Si λ aumenta entonces menor será el tiempo de espera.**

$$f(x) = \lambda e^{-\lambda x} \quad , \ \lambda > 0, \ x > 0$$

función de densidad de probabilidad (pdf)

$$E(X) = \frac{1}{\lambda} \ \ y \ \ V(X) = \frac{1}{\lambda^2}$$

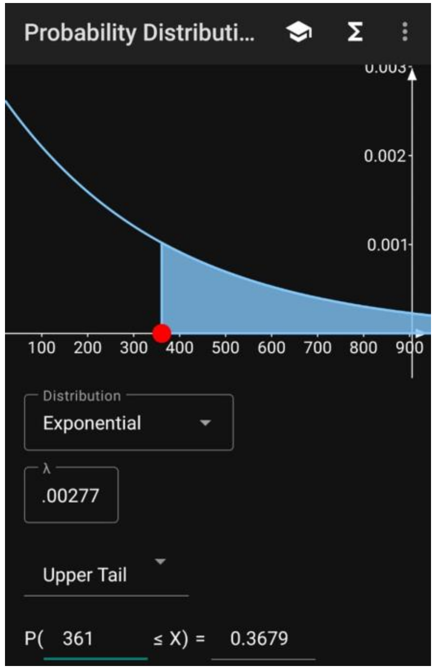
esperanza (promedio) y varianza

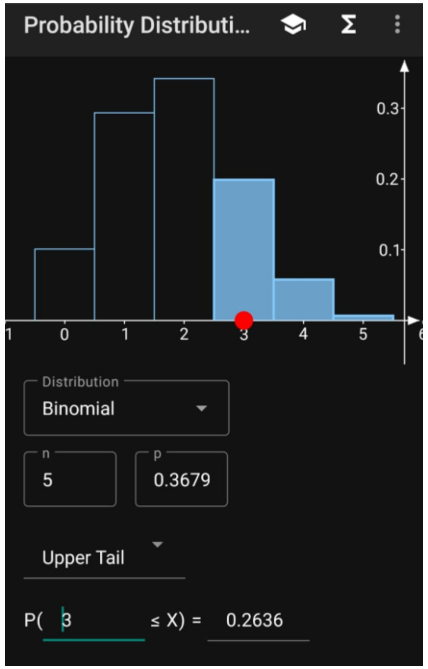
Distribución Exponencial - Ejemplos

El tiempo promedio durante el cual una batería trabaja de forma efectiva hasta que falle (tiempo de vida) es de 360 días. **¿Cuál es la probabilidad de que más de 2 baterías de 5 que adquirió un consumidor continúen trabajando después de 360 días?**

Solución:

- Éxito: batería dure más de 360 días
- p = probabilidad de éxito = P(X>360). Donde X es el tiempo de vida de la batería
- n = número de ensayos bernoulli = 5
- Y = número de baterías que duran más 360 días, de un total de 5 baterías.

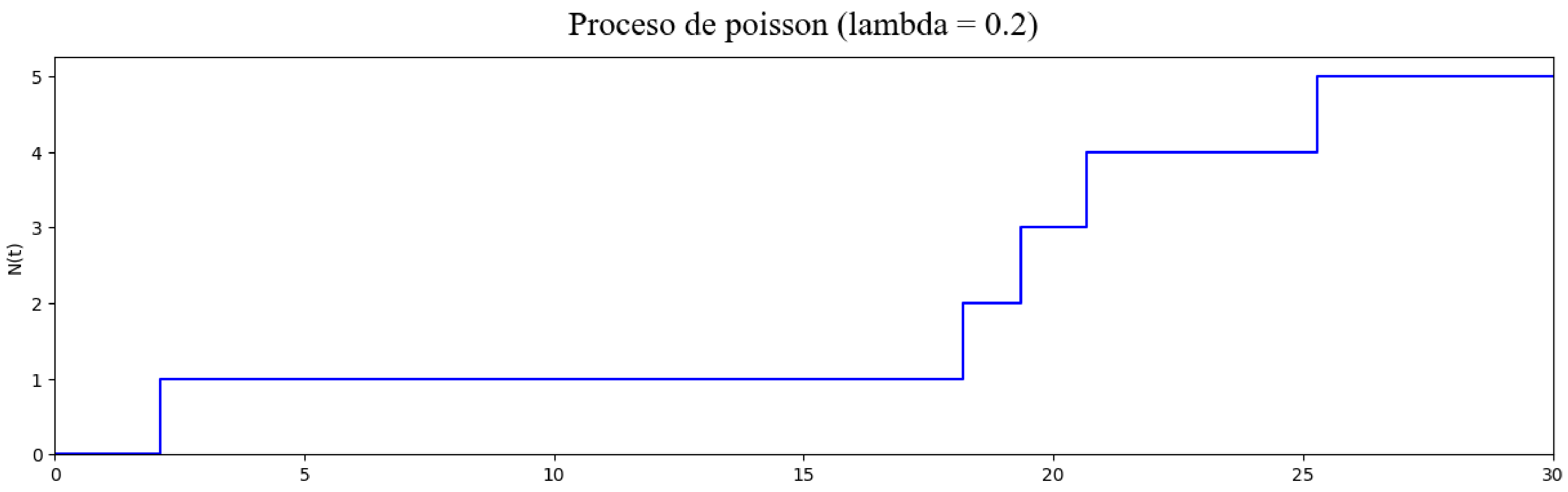




Distribución Exponencial - Aplicación

Proceso de Poisson

Contexto: Combinaremos los conceptos de la distribución exponencial y la distribución poisson para modelar **la cantidad de eventos que ocurren en un intervalo de tiempo t**



Nota: Revisar Anexo para simular procesos de poisson utilizando python

Distribución Exponencial - Aplicación

Proceso de Poisson

Definición:

Es un proceso estocástico que consiste en contar el número de eventos que han ocurrido hasta el tiempo t. Donde los tiempos de interocurrencias de los eventos son variables aleatorias que siguen una distribución Exp(λ)

***N(t)* = número de eventos que ocurren en un intervalo de tiempo t**

$$N(t) \sim Pois(\lambda t)$$

Distribución Normal o Gaussiana

¿Cuándo utilizar la distribución exponencial?

Para modelar distintos fenómenos aleatorios en el cual se desea introducir los términos de **promedio** y **varianza.**

¿Cómo se define matemáticamente?

$$X \sim N(\mu, \sigma)$$

Distribución con 2 parámetros

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \; x \in \mathbb{R}, \, \mu \in \mathbb{R}, \, \sigma > 0$$

función de densidad de probabilidad (pdf)

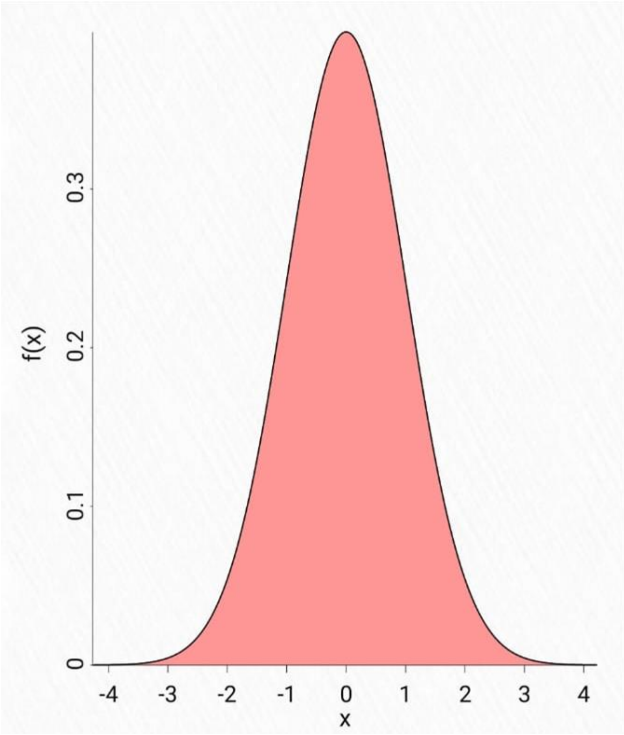
$$E(X) = \mu \;\; y \;\; V(X) = \sigma^2$$

esperanza (promedio) y varianza

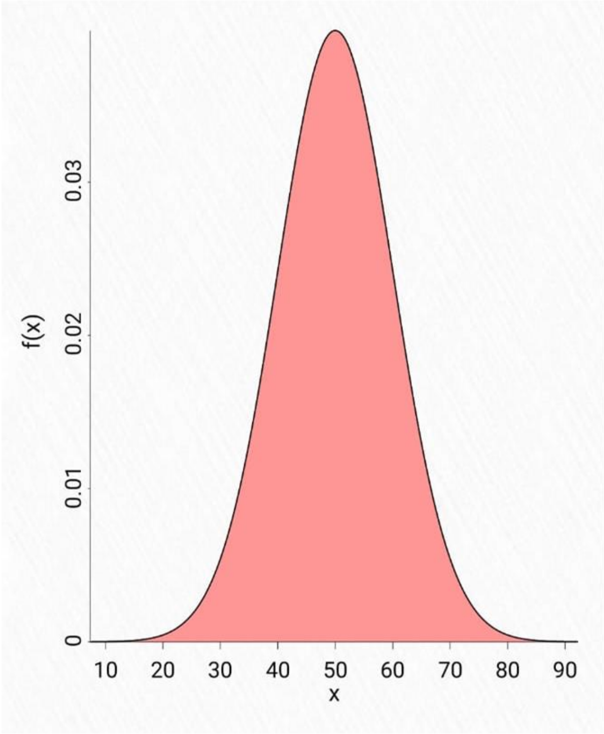
Distribución Normal o Gaussiana

¿Cómo se ve gráficamente?

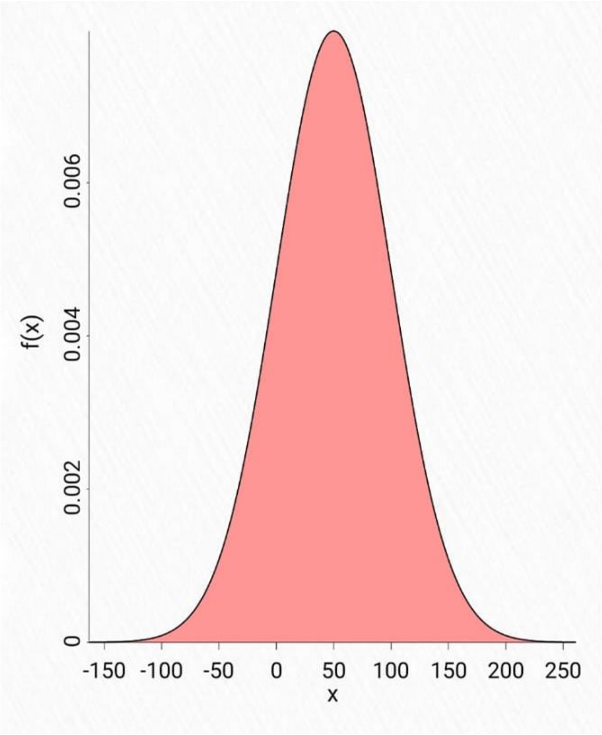
$$X \sim N(\mu = 0, \sigma = 1)$$



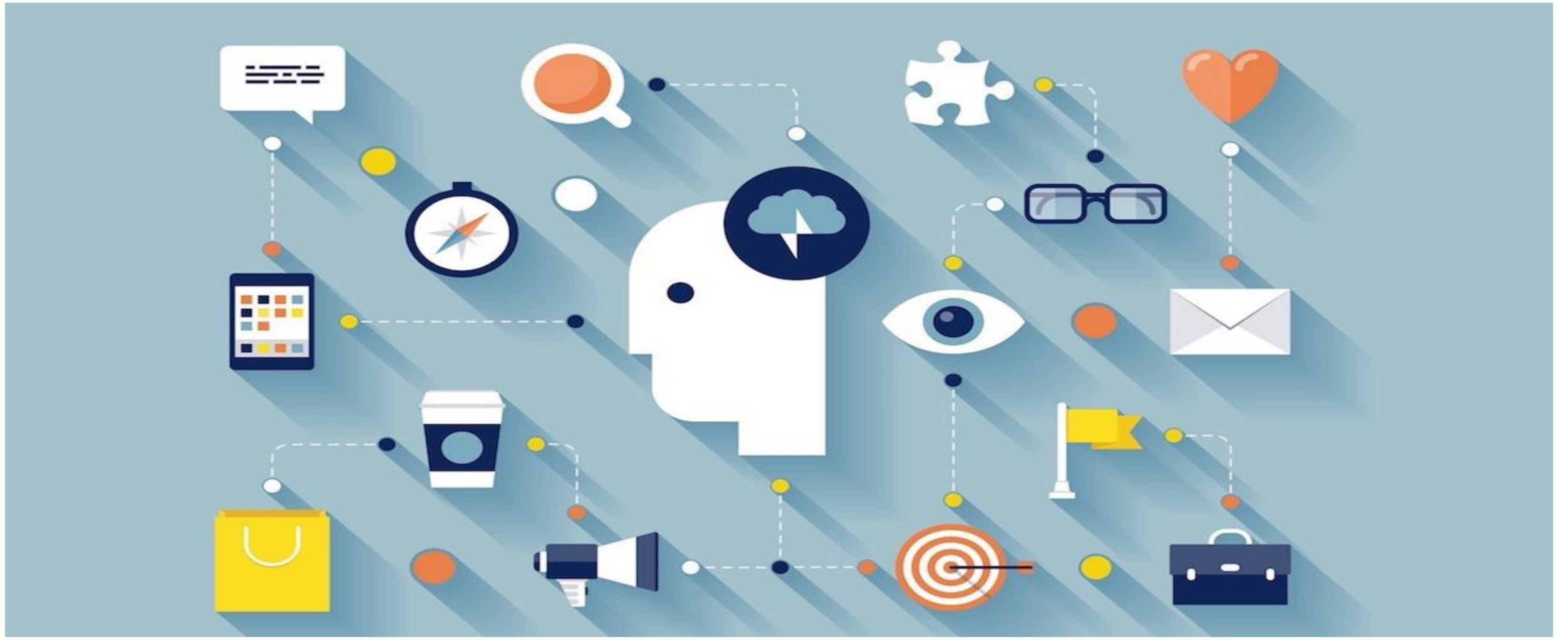
$$X \sim N(\mu = 50, \sigma = 10)$$



$$X \sim N(\mu = 50, \sigma = 50)$$



Para nuestra curiosidad…



[Distribución Normal - Ejercicios Resueltos - Nivel 1 - YouTube]

[TEORÍA DE LA DISTRIBUCIÓN NORMAL - YouTube]

Principales distribuciones discretas y continuas

DISCRETA

- Uniforme discreta
- Bernoulli
- Binomial
- Poisson
- Geométrica
- Hipergeométrica
- Binomial Negativa

CONTINUA

- Uniforme continua
- Exponencial
- Gamma
- Normal o Gaussiana
- Chi cuadrado
- Laplace o doble exponencial
- Erlang
- Weibull
- t-student
- F de Snedecor

QUIZ TIME

Aplicación en Estadística bayesiana

Estimemos la probabilidad de obtener cara

Supongamos que se ha realizado el lanzamiento de una moneda 100 veces obteniendo 40 caras y 60 sellos. ¿Cuál sería la probabilidad de obtener cara de esta moneda?

Según el Ciclo básico de un análisis estadístico:

- Parámetro (característica desconocida): Probabilidad de obtener cara (%)
- Muestra: Resultado de 100 lanzamientos (secuencia de 1s y 0s)
- Estadístico (característica conocida) = 40/100
- **Estimación del parámetro: p=0.4**

¿Podré incluir mis sospechas o creencias sobre la moneda para poder estimar el parámetro?

Bienvenido(a) a un nuevo mundo sobre estadística

- La estadística bayesiana **rompe el paradigma de la estadística clásica** de asumir el parámetro como desconocido.

- La estadística bayesiana **permite incluir nuestras creencias previas** o prioris sobre el parámetro, utilizando o **asignándole una distribución de probabilidad.**

- Bajo el enfoque bayesiano, el parámetro es tratado como una variable aleatoria.

Información de la muestra

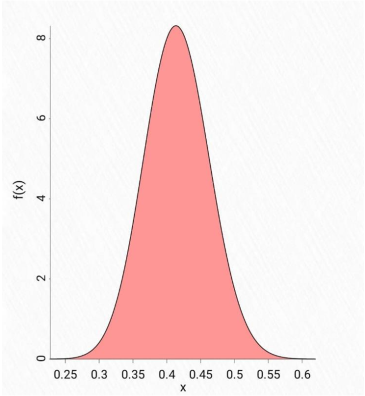
(**Distribución** muestral)

Creencias previas sobre el parámetro

(**Distribución** priori)

Creencias actualizadas sobre el parámetro

(**Distribución** posteriori)

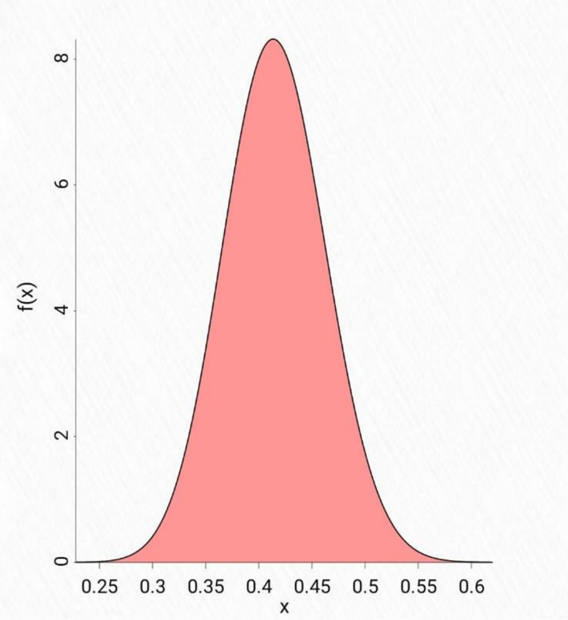


¡Bienvenido(a) a la estadística bayesiana!

Estimemos la probabilidad de obtener cara (E. Bayesiano)

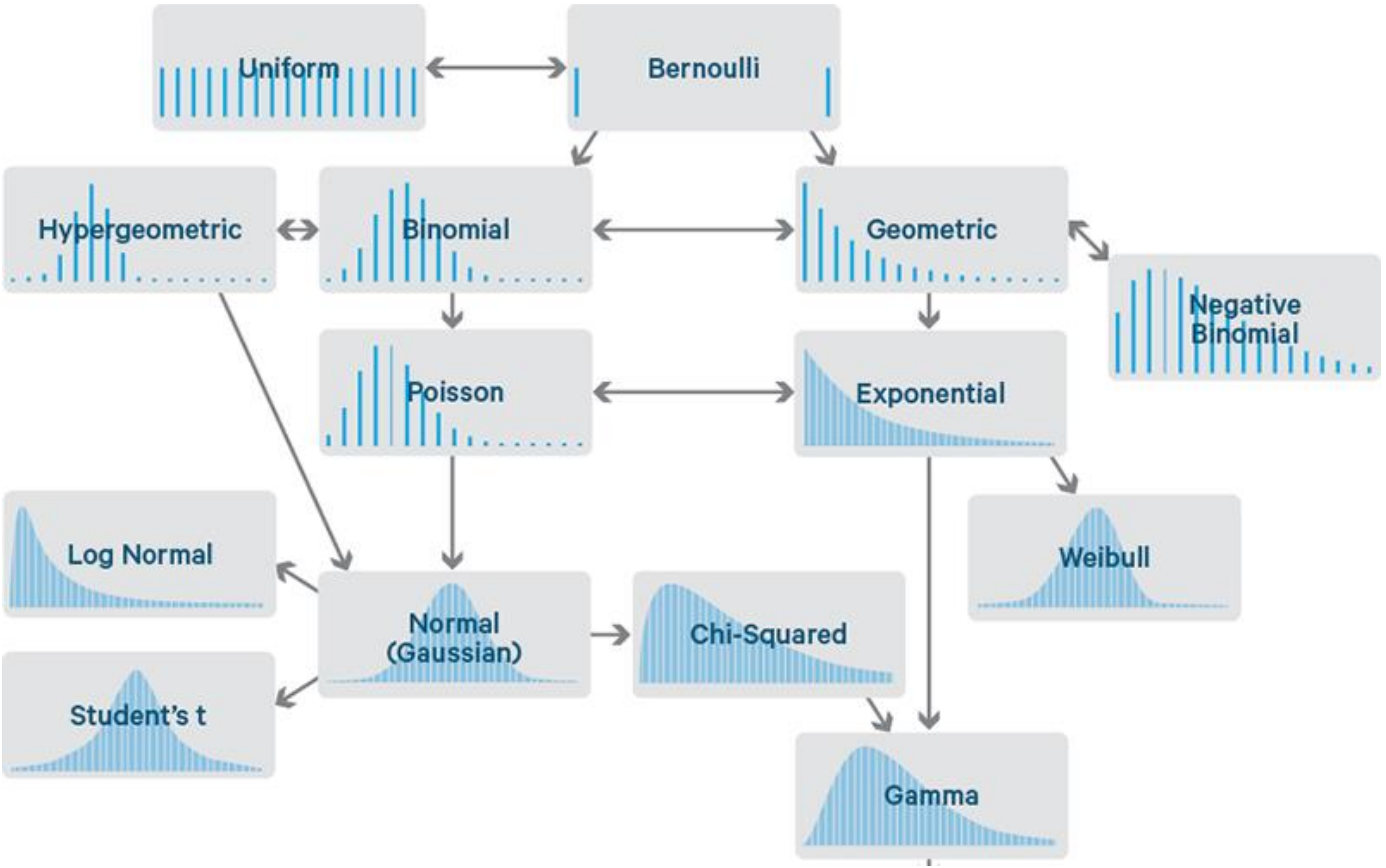
Reemplazando los valores para obtener los parámetros de la distribución Beta posteriori:

$$\theta \mid y \sim \ Beta(40{+}4{,}60{+}2)$$



Media= 0.4151

Para finalizar…



Examen Final:

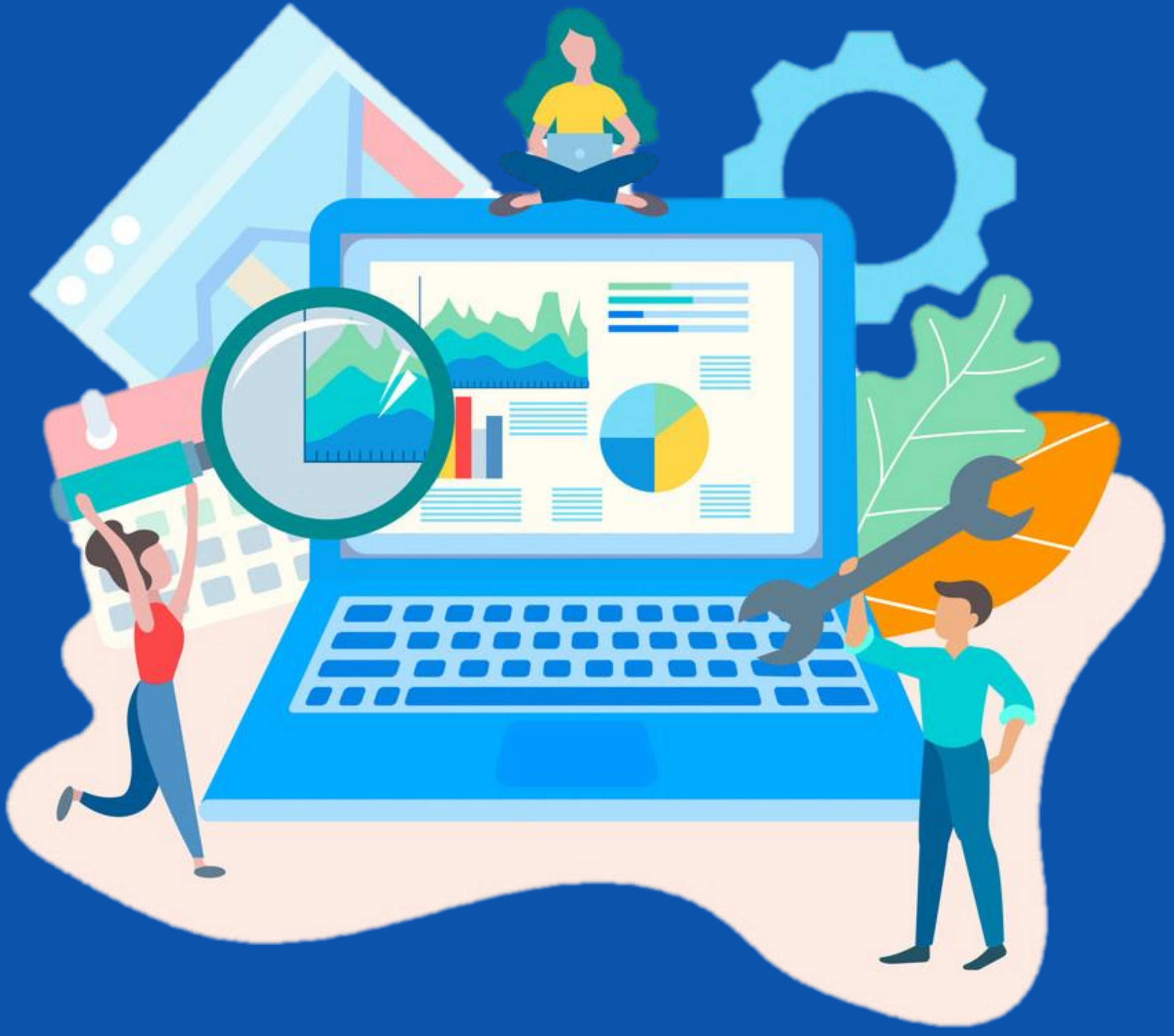
Resolver las preguntas del formulario:
Examen Final - Parte 1 - Fundamentos para la ciencia de datos y estadística

Resolver los ejercicios de Python del notebook:
Examen Final- Parte 2

Enviarlo el notebook resuelto con el asunto: Examen Final – Fundamentos – [Apellidos y nombres]

Correo: team@dataanalitica.net
Fecha máxima: 10pm | 09 de marzo
El siguiente curso empieza el viernes 17 de marzo.

GRACIAS



ANEXO - Proceso de Poisson

Simulación de un proceso de poisson

```python
import numpy as np

def poisson_process(val_lambda,val_t):
    val_0 = np.random.exponential(1/val_lambda,1)[0]
    W_t = val_0
    rng_valx = list([0,val_0])
    while W_t <= val_t:
        val = np.random.exponential(1/val_lambda,1)[0]
        W_t += val
        if W_t <= val_t: rng_valx.append(val)
    rng_ejex = list(np.cumsum(rng_valx))
    N_t = len(rng_ejex)
    rng_ejey = list(np.arange(N_t))
    rng_ejey.append(N_t-1)
    rng_ejex.append(val_t)
    print('Cantidad de realizaciones : ',N_t-1)
    print('Tiempos de interllegadas  : ',np.round(rng_ejex[1:-1],3))
    plt.step(x=rng_ejex,y=rng_ejey,where='post',color='blue')
    plt.title('Proceso de poisson (lambda = ' + str(val_lambda) + ')',
              fontdict={'fontname': 'Times New Roman', 'fontsize': 19}, y=1.03)
    plt.ylim(0)
    plt.ylabel('N(t)')
    plt.xlim([0,val_t])
    plt.show()
```

```python
poisson_process(0.7,30)
```

✓ 0.2s Python

```
Cantidad de realizaciones :  22
Tiempos de interllegadas  :  [ 0.645  2.231  2.28   4.256  4.595  5.822  7.096 11.116 13.329 15.585
 16.094 16.121 16.536 18.01  18.276 21.085 21.75  23.784 25.289 26.866
 26.952 27.856]
```