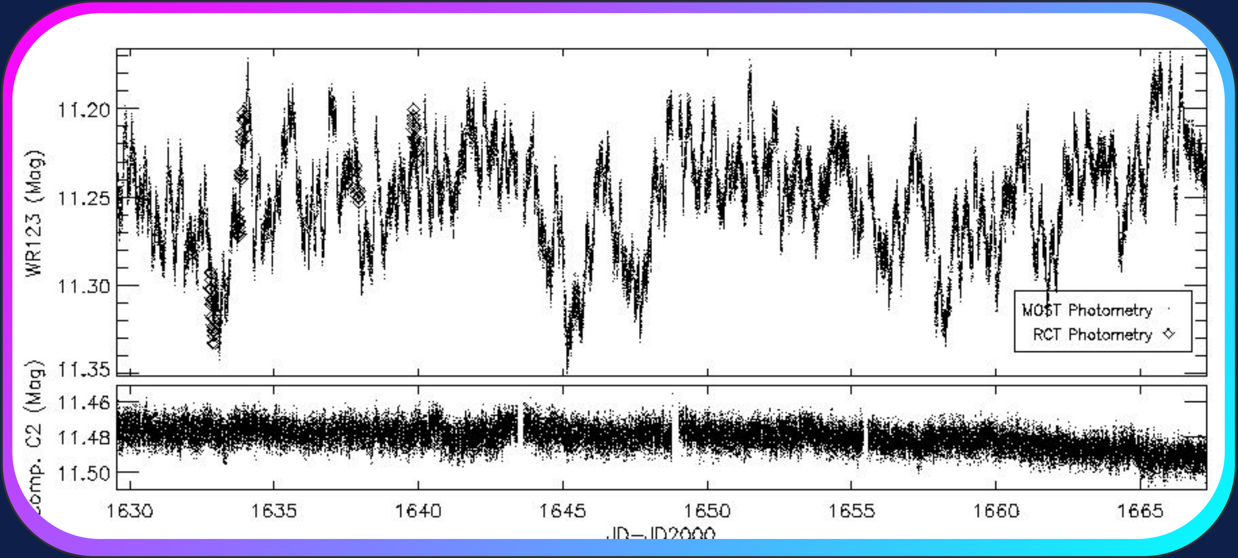
ASTRONOMÍA Y CIENCIA DE DATOS: DE LAS ESTRELLAS A LOS NÚMEROS

Clase 2: Introducción y Herramientas Fundamentales II: Estadística

ESTADÍSTICA EN EL CONTEXTO ASTRONÓMICO

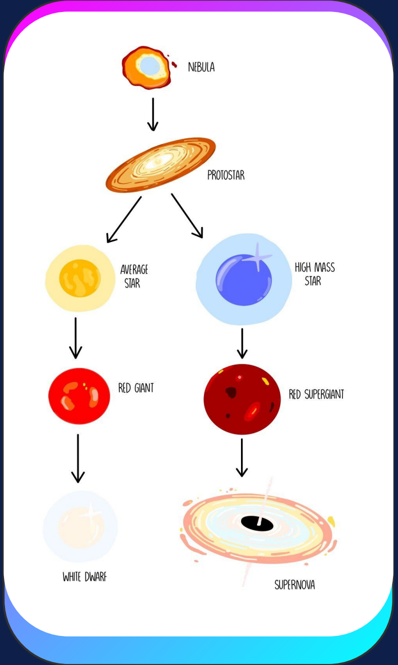


¿QUÉ VEMOS EN ESTAS IMÁGENES?

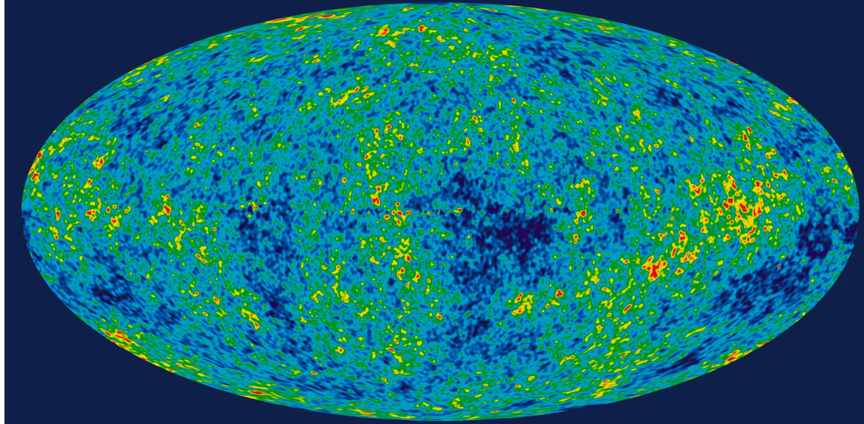
¿Ruido o señal?

¿Cómo podemos distinguirlos?

Modelo de Evolución Estelar



Modelo de Evolución del Universo



Modelo de Evolución de Galaxias

¿CUÁNTO PODEMOS **CONFIAR** EN ESTOS MODELOS?

LA ESTADÍSTICA ES NUESTRA SALVACIÓN <3

CONCEPTOS BÁSICOS DE ESTADÍSTICA EN UN CONTEXTO ASTRONÓMICO

¿QUÉ ES UNA VARIABLE ALEATORIA?

Definición

Una variable aleatoria (v.a.) es una función que, dado un evento aleatorio de un espacio muestral, asigna un valor numérico.

$$X : E \longrightarrow \mathbb{R}$$

¿QUÉ ES UNA VARIABLE ALEATORIA?

X (⚀) = 1

X (⚁) = 2

X (⚂) = 3

X (⚃) = 4

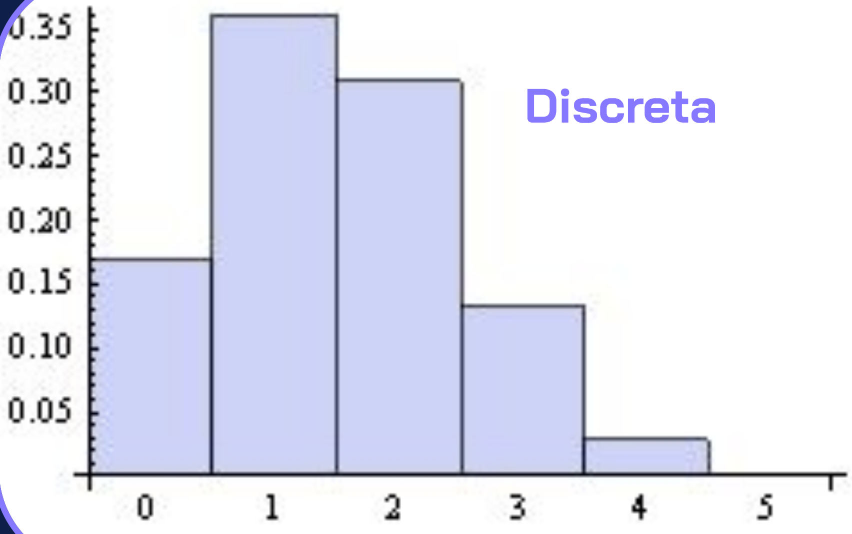
X (⚄) = 5

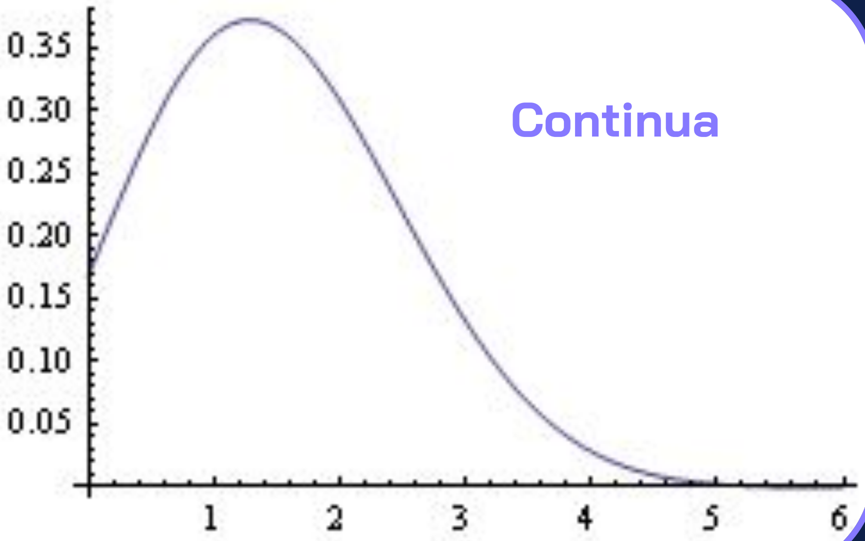
X (⚅) = 6

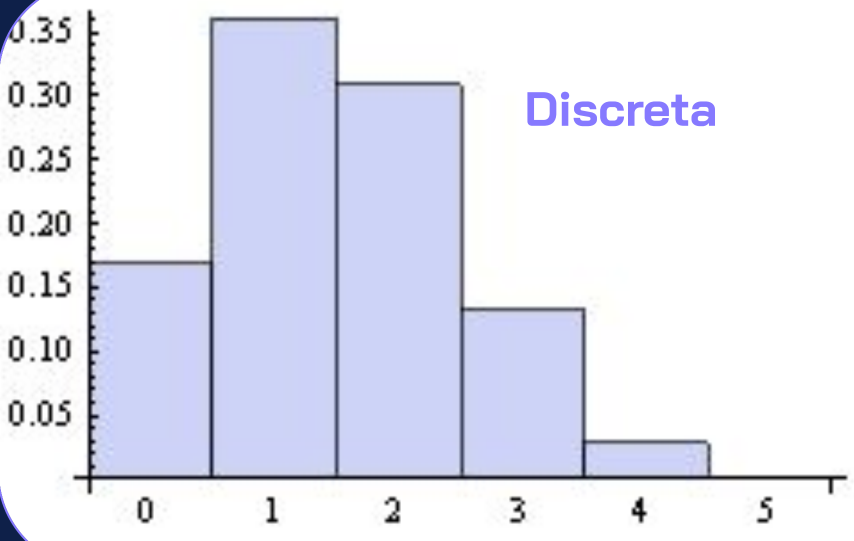
¿QUÉ ES UNA DISTRIBUCIÓN DE PROBABILIDAD?

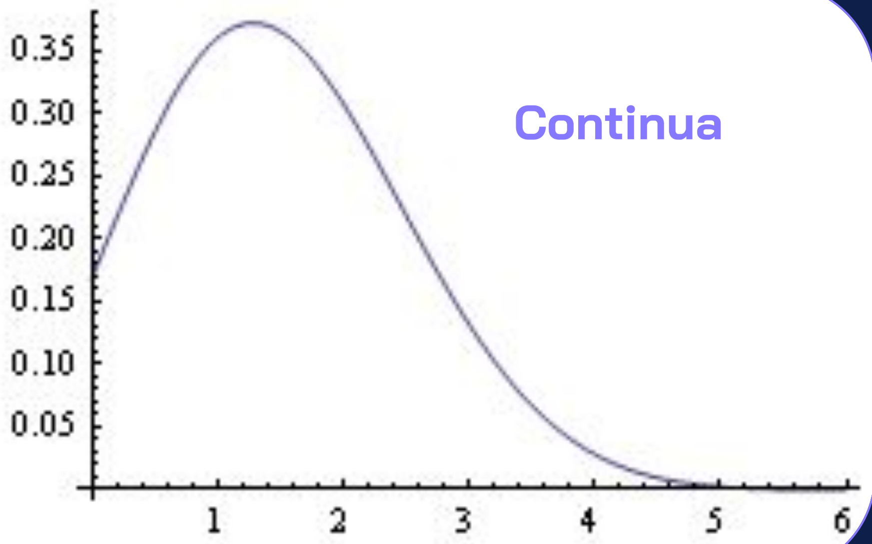
Función que muestra las probabilidades asociadas a cada posible resultado.

Hay dos tipos:









Los resultados se pueden contar y enumerar porque son específicos y finitos.

Ejemplo: Lanzar dado.

Los resultados se pueden medir pero no contar.

Ejemplo: Altura de una población.

Funciones de probabilidad

Discreta

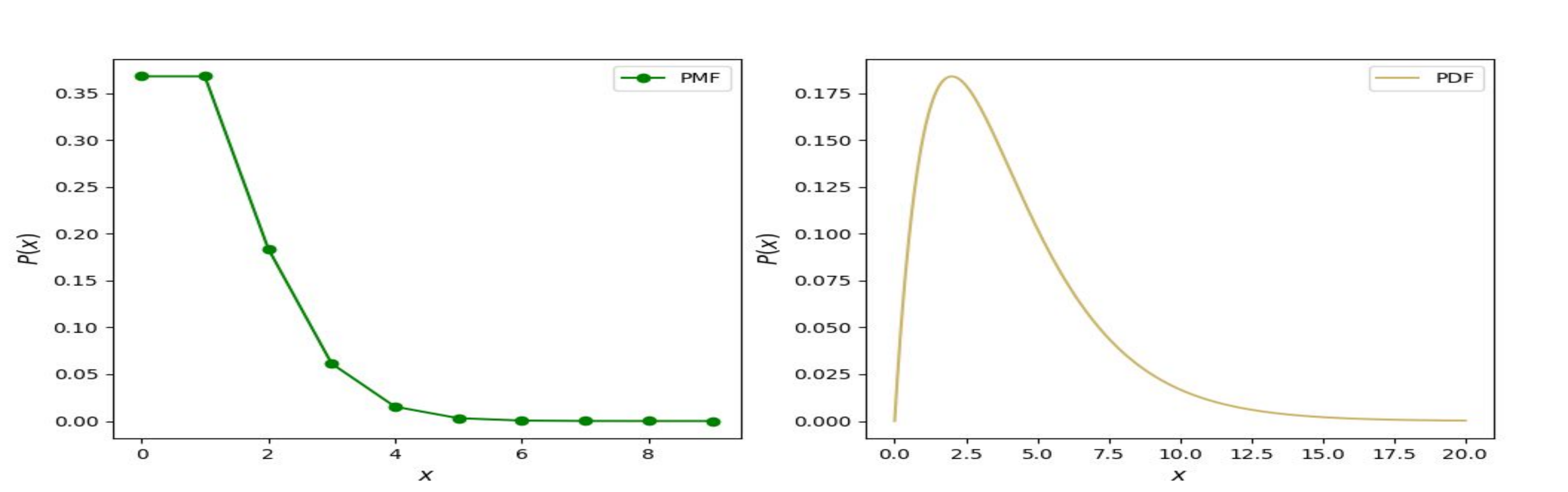
Función de probabilidad (PMF) $p(x)$

$$P(X = x) = p(x)$$

Continua

Densidad de probabilidad (PDF) $f(x)$

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



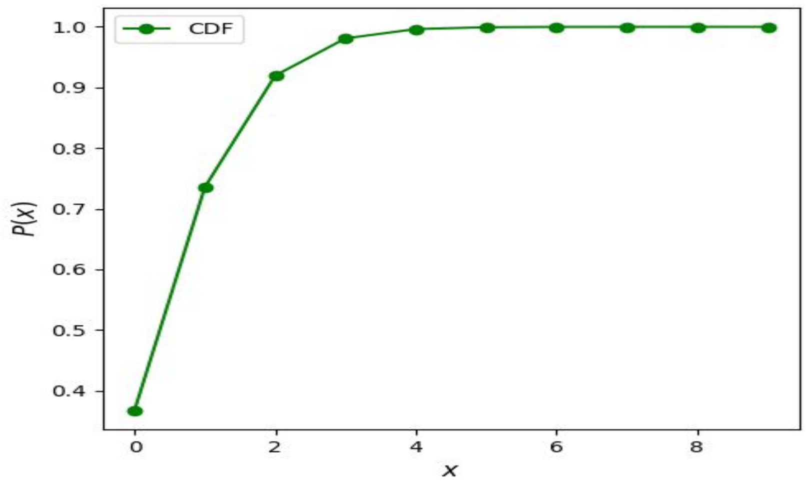
Probabilidad acumulada (CDF)

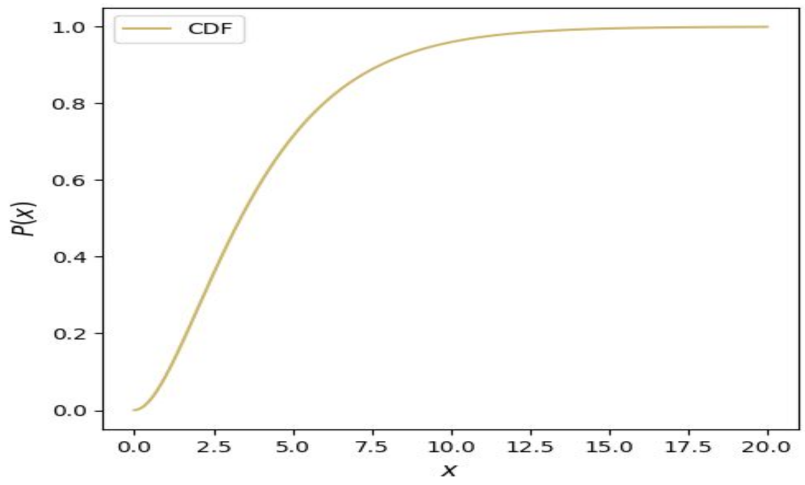
Discreta

$$P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} p(x)$$

Continua

$$P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$



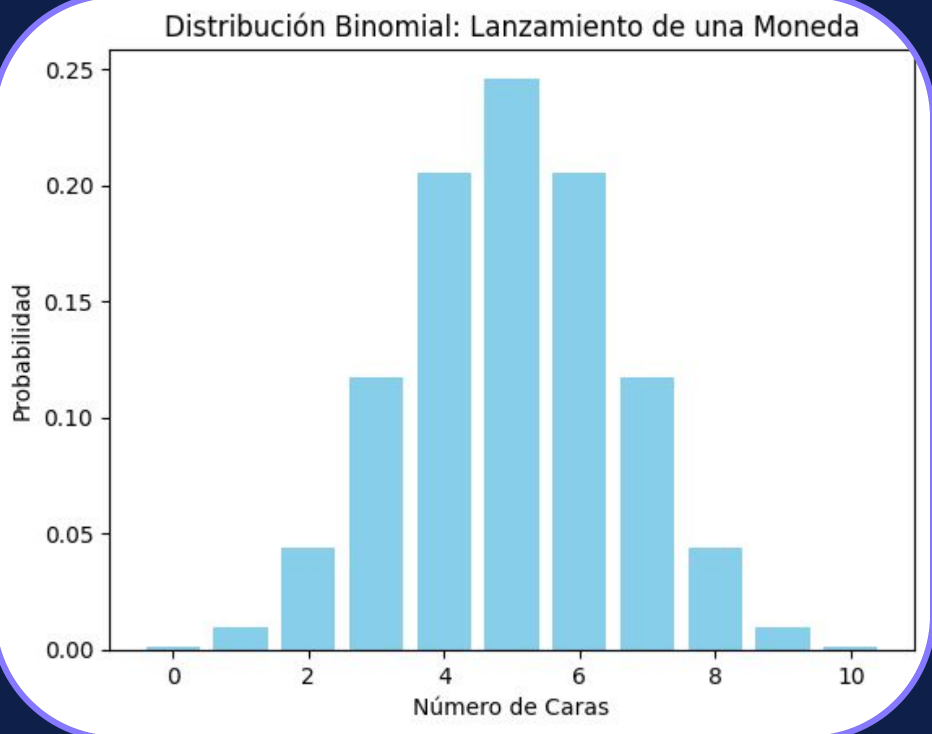


DISTRIBUCIÓN BINOMIAL

Solo hay dos posibles resultados a y b (éxito o fracaso).

Cuenta el número de veces que sale a (o b) y le asocia una probabilidad.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

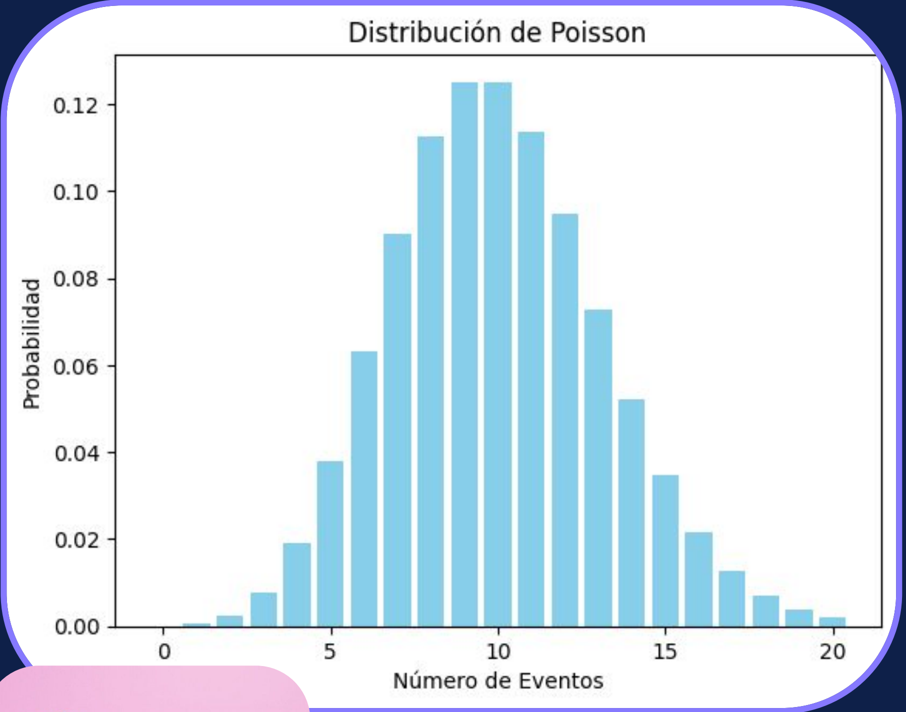


DISTRIBUCIÓN DE POISSON

Describe la probabilidad de que ocurra un número específico de eventos en un ΔT, dado un promedio conocido de eventos por intervalo.

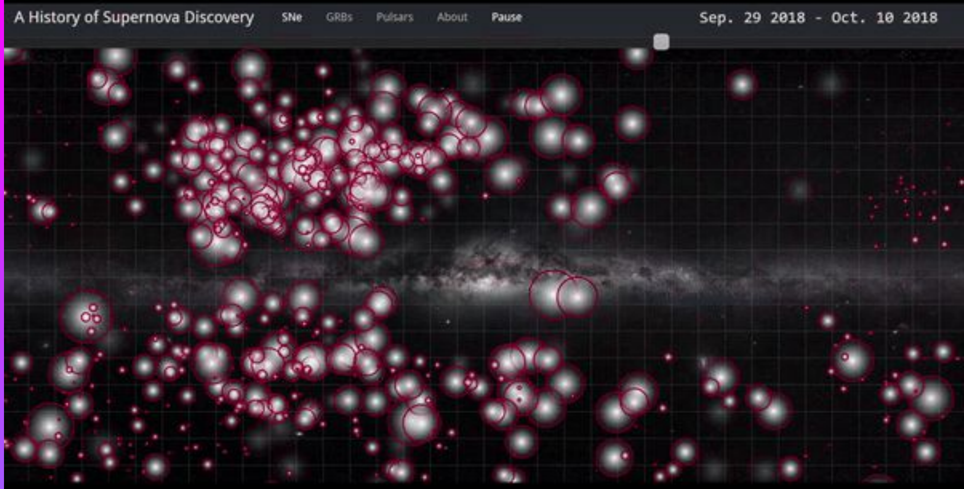
Ejemplo: La probabilidad de vender X donas en 1h dado que por hora se venden 10 donas en promedio

$$\Pr(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$



POISSON EN ASTRONOMÍA

FRECUENCIA DE ESTALLIDOS DE SUPERNOVAS



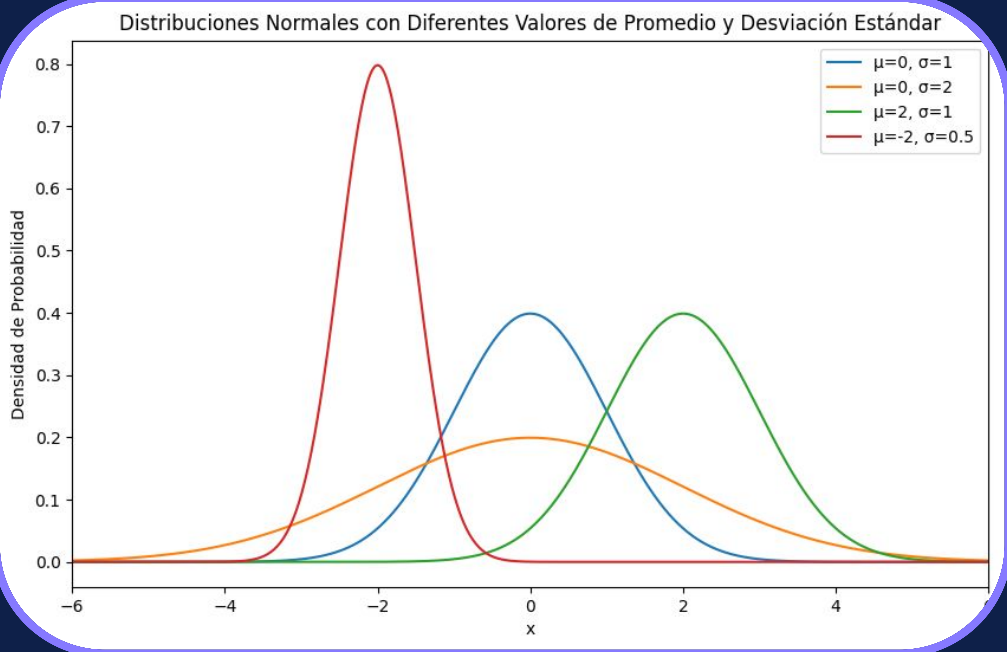
FRECUENCIA DE LLEGADA DE FOTONES A DETECTORES DE TELESCOPIOS

DISTRIBUCIÓN NORMAL

Representa datos que se agrupan alrededor de un valor central con variaciones simétricas.

Común en las ciencias pues muchos fenómenos naturales tienden a seguir este patrón.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

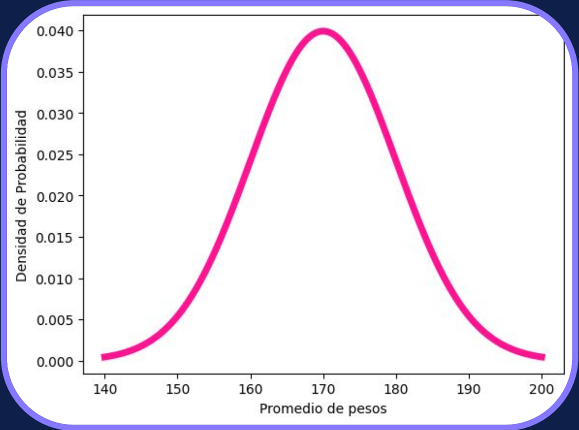


TEOREMA DEL LÍMITE CENTRAL (TLC)

Al azar **muchos** dulces

Calcular promedio de los pesos

Si se repite el procedimiento **muchas** veces...
El promedio distribuye como una Gaussiana

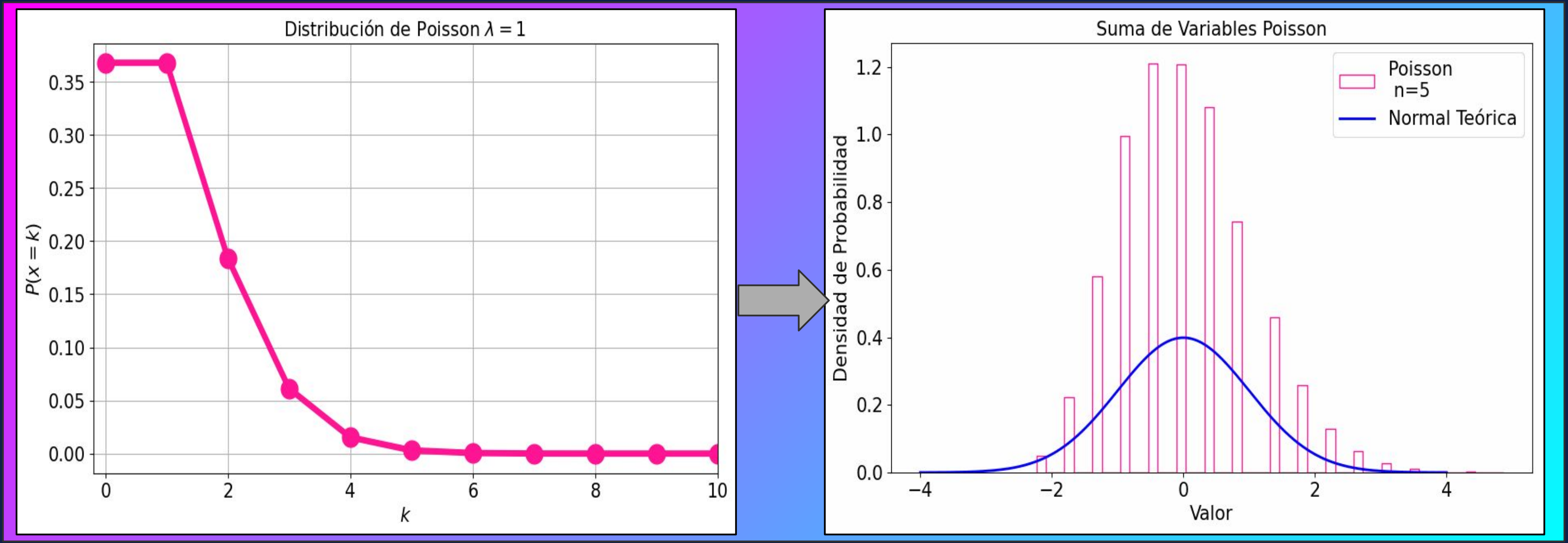


TEOREMA DEL LÍMITE CENTRAL (TLC)

Sea $X_1, X_2, \ldots, X_n$ una muestra de tamaño n de una población con media μ y desviación σ . Entonces el promedio muestral $\overline{X} = (X_1 + X_2 + \ldots + X_n)/n$ tiene una distribución normal de media μ y desviación estándar $\sigma/\sqrt{n}$.

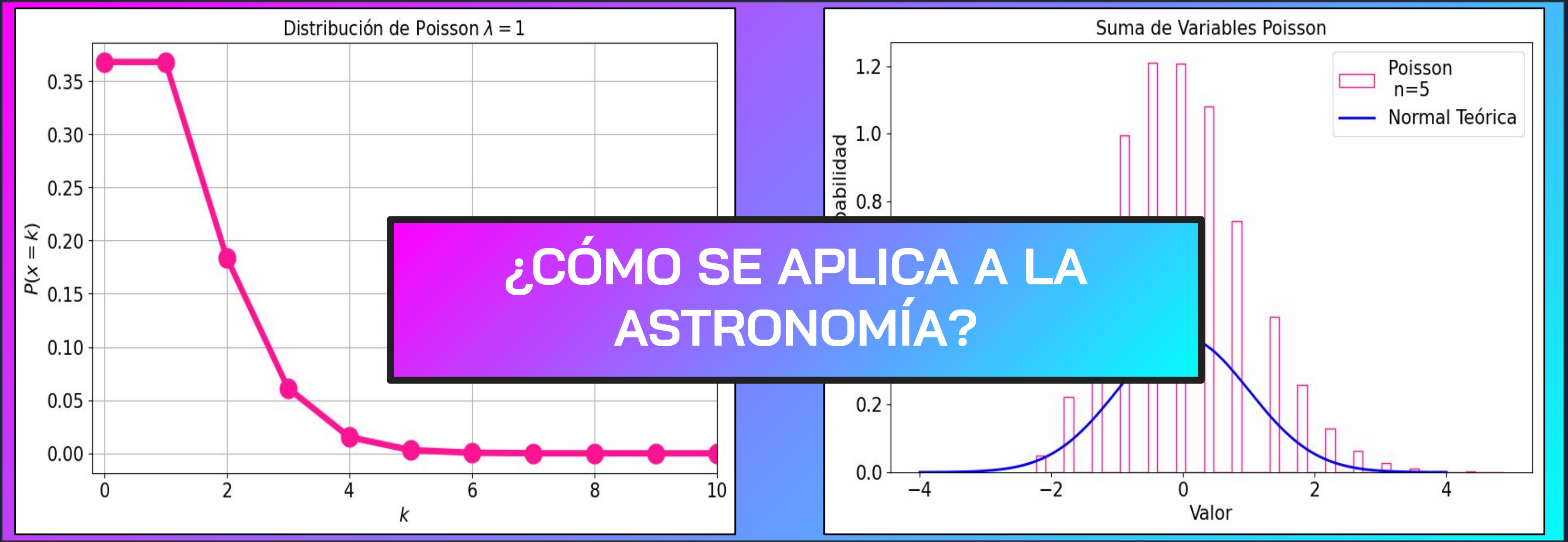
Notamos que tanto $X_1, X_2, \ldots, X_n$
como $\overline{X}$ son variable aleatorias.

TEOREMA DEL LÍMITE CENTRAL (TLC)



Cualquiera sea la distribución madre, **los promedios de las muestras de una variable aleatoria se acercan a una distribución normal cuando se aumenta el tamaño de las muestras.**

TEOREMA DEL LÍMITE CENTRAL (TLC)



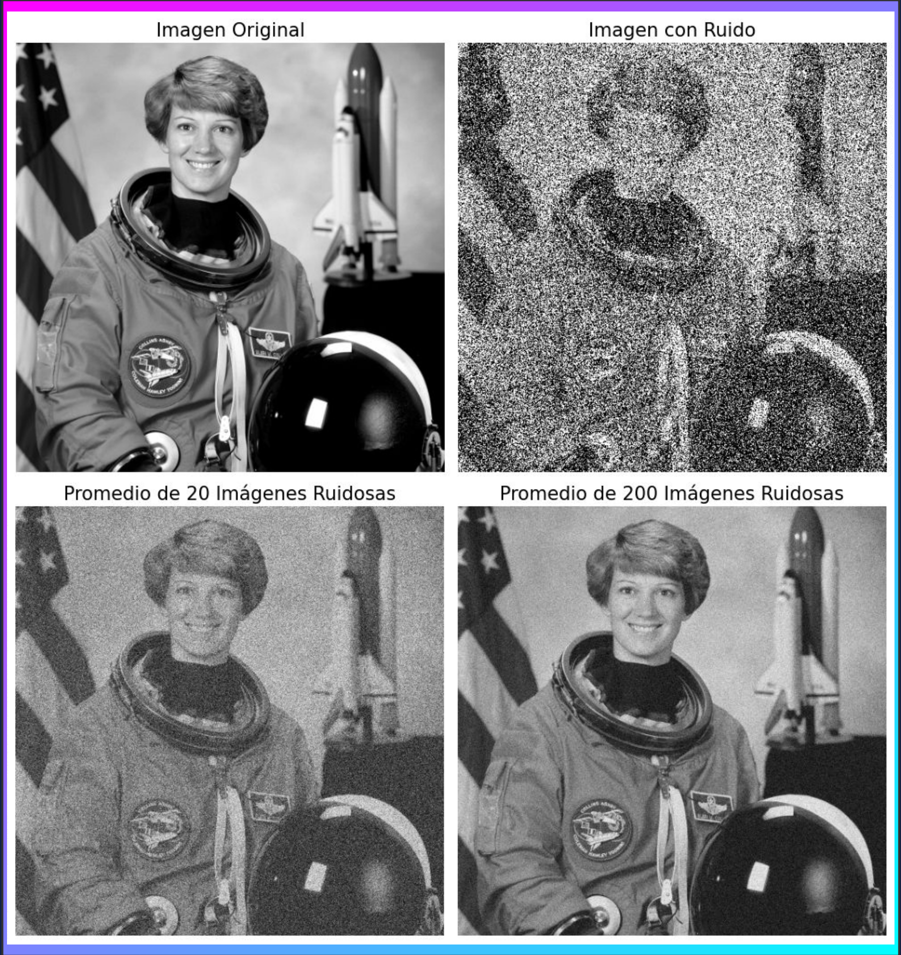
¿CÓMO SE APLICA A LA ASTRONOMÍA?

Cualquiera sea la distribución madre, **los promedios de las muestras de una variable aleatoria se acercan a una distribución normal cuando se aumenta el tamaño de las muestras.**

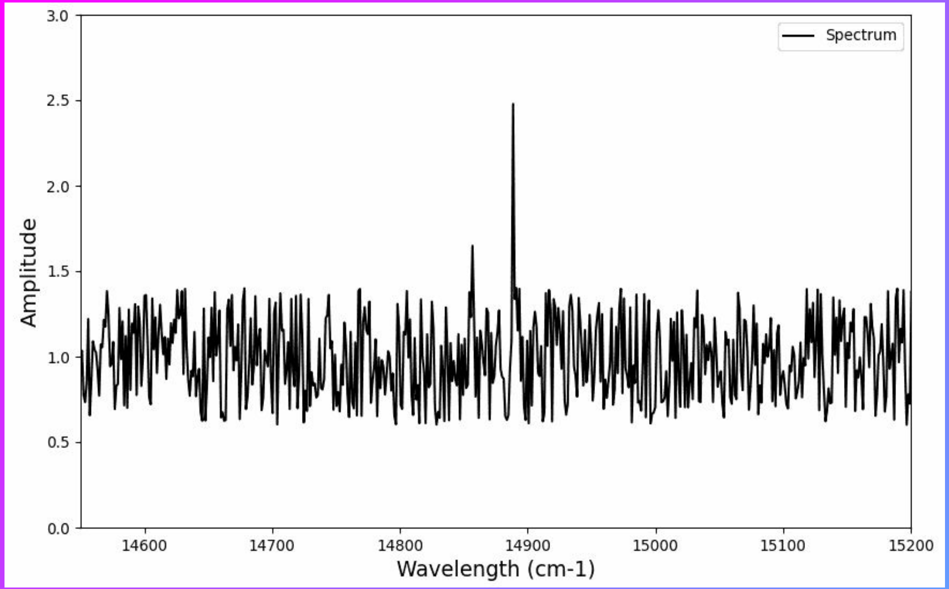
DISTRIBUCIÓN NORMAL EN ASTRONOMÍA Y TLC

DISMINUCIÓN DEL RUIDO EN LOS DATOS

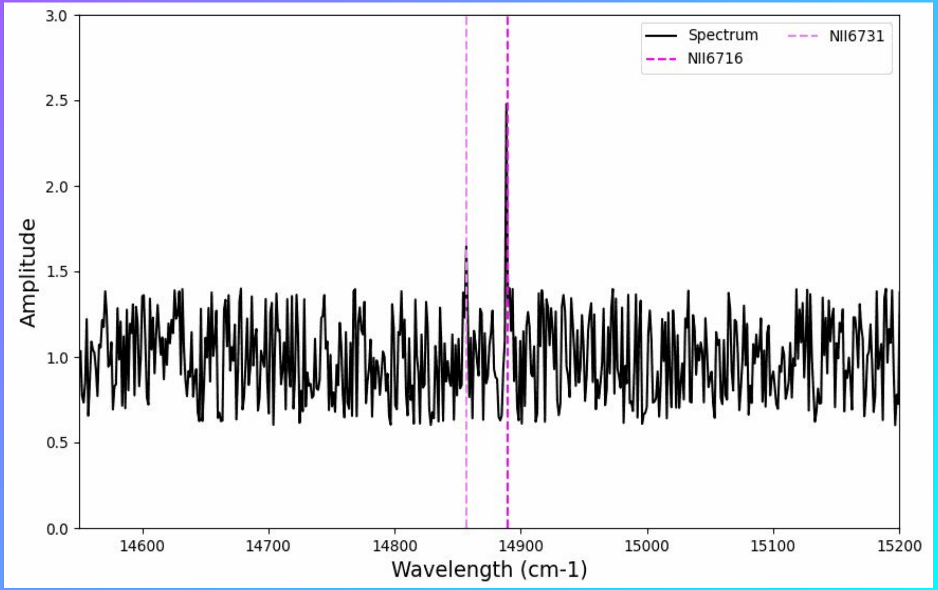
- Múltiples mediciones de una cantidad.
- Cada medición está afectada por ruido y errores.
- Al promediar múltiples observaciones el ruido tiende a cancelarse (TLC).
- Mientras más mediciones, al promediarlas se distinguirá mejor la señal c/r al ruido.



OBSERVACIONES/ESPECTROS INDIVIDUALES



PROMEDIO DE ESPECTROS



“Señal a ruido” $\stackrel{\text{def}}{=}$ P_S/P_N
aumenta al promediar observaciones individuales

DISTRIBUCIÓN NORMAL EN ASTRONOMÍA Y TLC

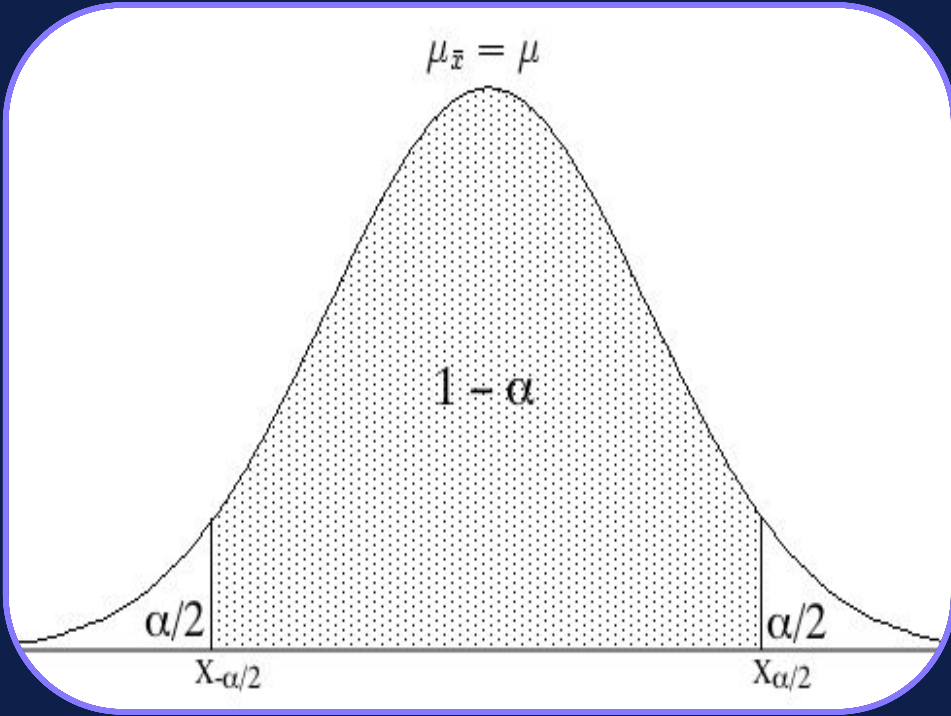
ESTIMACIÓN DE PARÁMETROS

Medición de luminosidad de estrellas

Caracterización de poblaciones estelares (edad, masa, metalicidad)

Intervalos de confianza

DISTRIBUCIÓN NORMAL EN ASTRONOMÍA Y TLC



Intervalos de confianza

Intervalo de confianza del $(1-\alpha) * 100\%$:

$$P[z_1 \leq Z \leq z_2] = 1 - \alpha$$

Estandarización de distribución de parámetro:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

MÉTRICAS: Caracterización y comparación de distribuciones

Son esenciales para entender la forma, la dispersión de datos y **<u>tomar decisiones informadas</u>**

PROMEDIO:

$$\mu = E[X]$$

VARIANZA:

$$\sigma = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

SKEWNESS (sesgo):

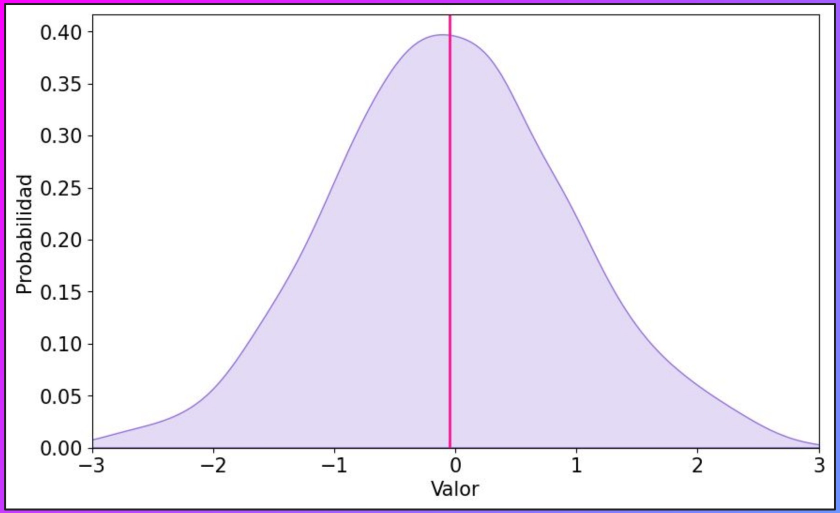
$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

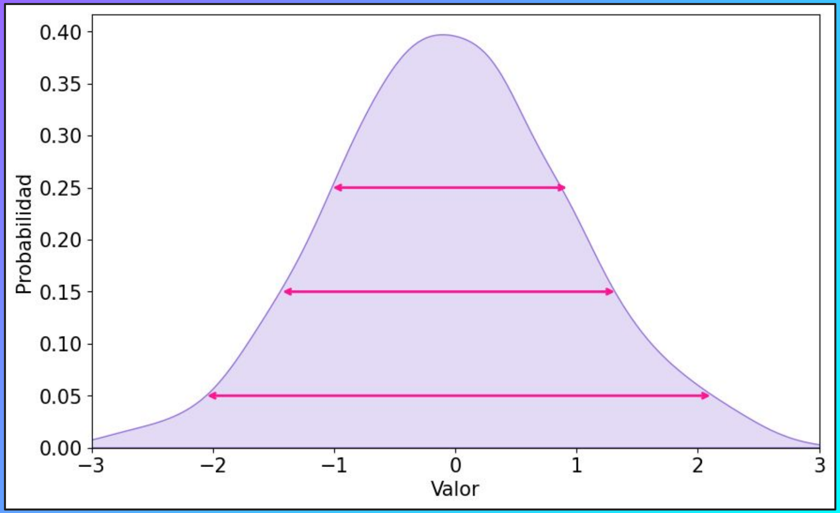
KURTOSIS (curtosis):

$$\mu_4 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

MÉTRICAS: Caracterización y comparación de distribuciones

Son esenciales para entender la forma, la dispersión de datos y **tomar decisiones informadas**



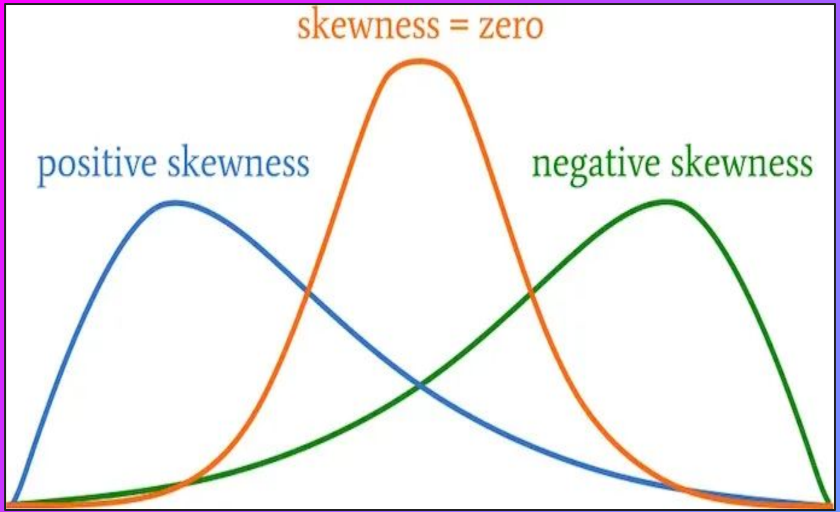


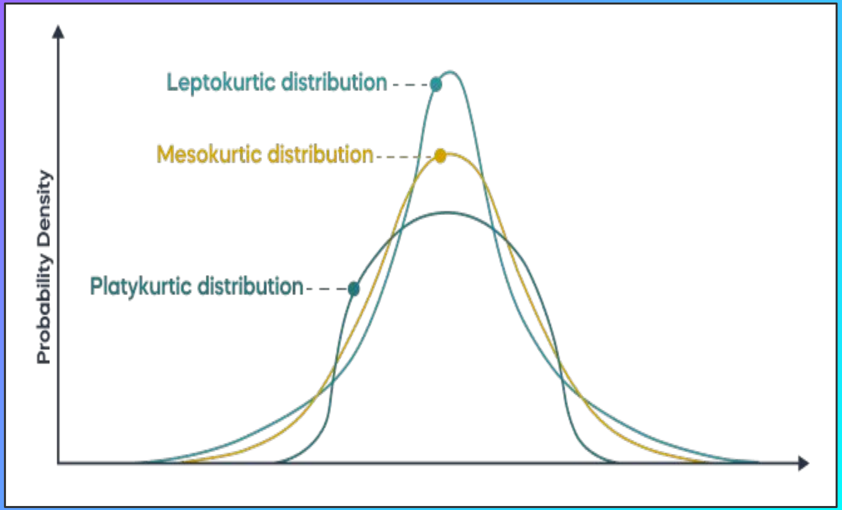
PROMEDIO:
Es el valor esperado

VARIANZA:
Dispersión de calores c/r al promedio

MÉTRICAS: Caracterización y comparación de distribuciones

Son esenciales para entender la forma, la dispersión de datos y **tomar decisiones informadas**





SKEWNESS (sesgo):
Mide asimetría de la distribución

KURTOSIS (curtosis):
Describe forma c/r a la normal