MÓDULO 2: MÉTODOS ESTADÍSTICOS PARA IA

OBJETIVO GENERAL: INTRODUCIR LOS FUNDAMENTOS ESTADÍSTICOS NECESARIOS PARA ANALIZAR DATOS Y EVALUAR MODELOS DE INTELIGENCIA ARTIFICIAL, FAVORECIENDO LA TOMA DE DECISIONES BASADAS EN EVIDENCIA.

- CLASE 1: NOCIONES DE PROBABILIDAD. ESTADÍSTICA DESCRIPTIVA.
- CLASE 2: TEST DE HIPÓTESIS.
- CLASE 3: REGRESIÓN.
- CLASE 4: ESTADÍSTICA NO PARAMÉTRICA.

MÓDULO 2: MÉTODOS ESTADÍSTICOS PARA IA

OBJETIVO GENERAL: INTRODUCIR LOS FUNDAMENTOS ESTADÍSTICOS NECESARIOS PARA ANALIZAR DATOS Y EVALUAR MODELOS DE INTELIGENCIA ARTIFICIAL, FAVORECIENDO LA TOMA DE DECISIONES BASADAS EN EVIDENCIA.

- CLASE 1: NOCIONES DE PROBABILIDAD. ESTADÍSTICA DESCRIPTIVA.
- CLASE 2: TEST DE HIPÓTESIS.
- CLASE 3: REGRESIÓN.
- CLASE 4: ESTADÍSTICA NO PARAMÉTRICA.

Clase 1 - Parte 1: Métodos Estadísticos para IA

Nociones de Probabilidad.

Objetivos de esta presentación

- Repasar algunas Probabilidades.
- Repasar algunas distribuciones de probabilidades discretas y continuas.

Resumen

Introducción

Probabilidades y distribuciones

Distribuciones de probabilidad discretas

Distribuciones de probabilidad continuas

Cuantiles

Resumen

Estadística: Introducción

- El campo de la estadística tiene que ver con la recopilación, organización, análisis y uso de datos para tomar decisiones razonables basadas en tal análisis.

- Al recoger datos relativos a las características de un grupo de individuos u objetos, suele ser imposible o poco práctico observar todo el grupo completo, en especial si es muy grande.

- En general, en lugar de analizar el conjunto entero llamado **población** o **universo**, se examina una pequeña parte del mismo, llamada **muestra**.

- En muchos problemas estadísticos es necesario utilizar una muestra de observaciones tomadas de la población de interés con objeto de obtener conclusiones sobre ella.

Estadística: Definiciones

- Una ***población*** está formada por la totalidad de las observaciones en las cuales se tiene cierto interés.
- Un ***individuo*** es cada uno de los elementos de la población.
- Una ***muestra*** es un subconjunto de observaciones seleccionada de una población.
- Una ***muestra aleatoria*** es una muestra cuyos elementos son elegidos al azar. Es decir, durante el procedimiento de selección, todos y cada uno de los elementos de la población tiene una cierta probabilidad de resultar elegidos.
- Una ***variable aleatoria*** es una función que asigna un valor numérico al resultado de un experimento aleatorio.

Estadística: Definiciones

- Si una muestra es representativa de una población, es posible **inferir** importantes conclusiones sobre la población a partir del análisis de la muestra.

- La parte de la estadística que trata sobre las condiciones bajo las cuales tal inferencia es válida se llama ***estadística inductiva*** o ***inferencia estadística***. Dado que dicha inferencia no es del todo exacta, el lenguaje de las probabilidades aparecerá al establecer nuestras conclusiones.

- La parte de la estadística que estudia la muestra sin inferir alguna conclusión sobre la población es la ***estadística descriptiva***, la cual trata sobre los métodos para recolectar, organizar y resumir datos.

Resumen

Introducción

Probabilidades y distribuciones

Distribuciones de probabilidad discretas

Distribuciones de probabilidad continuas

Cuantiles

Inferencia Estadística

- La **inferencia estadística** es el conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una **muestra**, cuál es el comportamiento de una determinada **población** con un riesgo de error medible en términos de **probabilidad**.

- Los métodos paramétricos de la inferencia estadística se pueden dividir, básicamente, en dos: **métodos de estimación de parámetros** y **métodos de contraste de hipótesis**.

- Ambos métodos se basan en el conocimiento teórico de la distribución de probabilidad del estadístico muestral que se utiliza como estimador de un parámetro.

Probabilidades y Distribuciones

- **Probabilidad:** Una definición informal puede ser que la **probabilidad** es el estudio de la aleatoriedad, es decir, cualquier situación donde podría ocurrir uno de varios casos posibles.

- Una definición un poco más formal nos diría que si tenemos un evento (conjunto) A, se puede asociar un número real c (que cumpla $0 \leq c \leq 1$), tal que $P(A) = c$.
  (P(A) se debe leer "probabilidad de A").

- Cuando estudiamos un problema con un enfoque probabilístico se debe definir una variable que se denomina **Variable Aleatoria** X (la representaremos con mayúscula).

- Una Variable Aleatoria es una **función** que asocia valores numéricos a los resultados de un experimento aleatorio, de modo que cada resultado posible del experimento tiene un valor real asignado.

Probabilidades y Distribuciones

- Las **Variables Aleatorias (VAs)** se clasifican en **discretas** o **continuas**.

- Las **VAs discretas** están caracterizadas por su **función de probabilidad puntual (fpp)**. Las denotaremos $P(X = x_i)$.

- Las **VAs continuas** están caracterizadas por su **función de densidad de probabilidad (fdp)**. Las denotaremos $f(X)$.

- En general, nos va a interesar conocer la probabilidad en un intervalo.

- Si quisiéramos calcular $P(X \leq x)$, es decir, la probabilidad que la VA X tome valores menores o iguales que el número x, entonces tendremos que utilizar la **Función de distribución acumulada (Fda)**: $P(X \leq x) = F(x)$.

- Ejemplos de VAs discretas: Binomial, Geométrica, Poisson, etc.

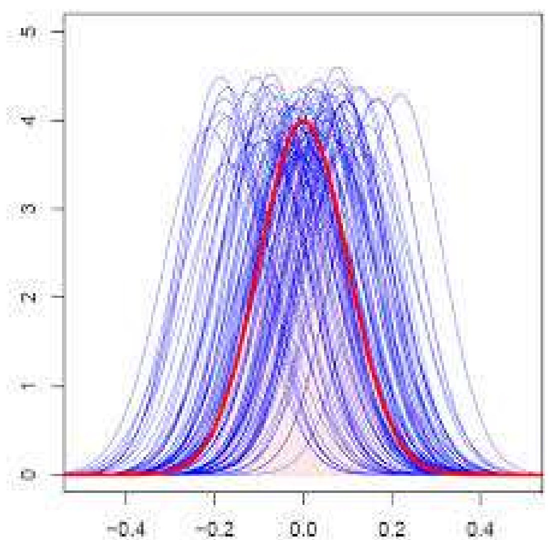
- Ejemplos de VAs continuas: Uniforme, Normal, χ-cuadrado, etc.

Esperanza y Varianza de Variables Aleatorias

- Muchas veces usamos solo dos números para resumir una distribución de probabilidad de una variable aleatoria X.

- El primero es la **esperanza (o valor esperado, o media)**, que nos dice dónde está el centro o valor típico de la distribución. Puede interpretarse intuitivamente como el valor medio de infinitas observaciones.

- El segundo es la **varianza**, que nos indica qué tan dispersos o alejados están los valores respecto de la media.



Esperanza y Varianza de Variables Aleatorias

- **Propiedades** de la **esperanza *E(X)*** y de la **varianza V(X)** de una variable aleatoria *X*:

- $E(k) = k$
- $E(k \cdot X) = k \cdot E(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(k + X) = k + E(X)$
- Si X y Y son independientes $\Rightarrow$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

- $V(k) = 0$
- $V(k \cdot X) = k^2 \cdot V(X)$
- Si X y Y son independientes $\Rightarrow$

$$V(X + Y) = V(X) + V(Y)$$

$$V(X - Y) = V(X) + V(Y)$$

- $V(k + X) = V(X)$

Resumen

Distribuciones de probabilidad discretas

- Una **distribución de probabilidad discreta** es aquella distribución que define las probabilidades de una **variable aleatoria discreta**.

- Por lo tanto, una distribución de probabilidad discreta solo puede tomar un número finito de valores (generalmente enteros).

- En una distribución de probabilidad discreta se asocia cada valor de la variable discreta que representa (x_i) a un valor de probabilidad ($f(x_i) = p_i$) que va desde 0 hasta 1. De manera que la suma de todas las probabilidades de una distribución discreta da como resultado uno.

$$(1) \quad f(x_i) \geq 0$$

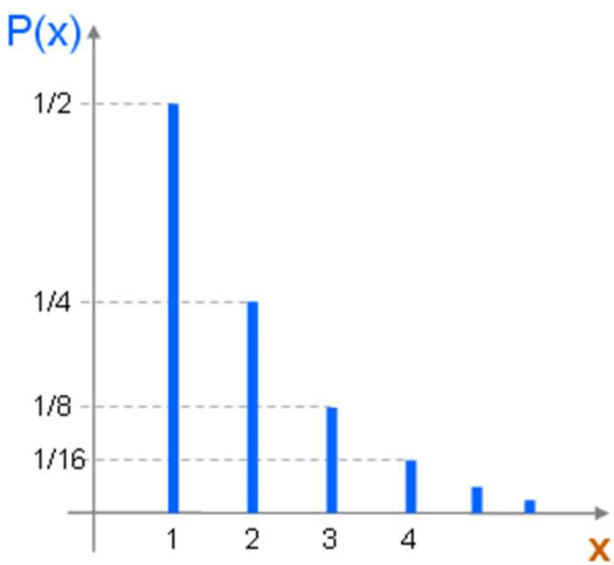
$$(2) \quad \sum_{i=1}^{n} f(x_i) = 1$$

$$(3) \quad f(x_i) = P(X = x_i)$$

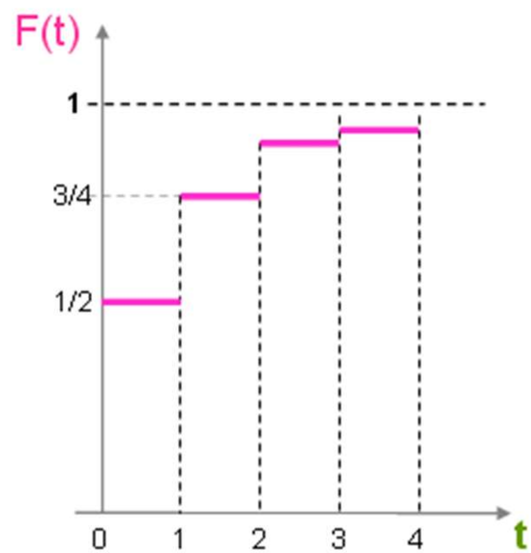
Distribuciones de probabilidad discretas

- La **Función de distribución acumulada (Fda)** de una **variable aleatoria discreta** es:

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$



Distribución de probabilidades



Función de distribución

Distribuciones de probabilidad discretas

- La **esperanza de una distribución discreta** se calcula como la suma de cada valor de la variable multiplicado por su probabilidad.

$$\mu = E(X) = \sum_{x} x f(x)$$

- La **varianza de una distribución discreta** se calcula como la suma de los cuadrados de las diferencias entre cada valor y la esperanza, ponderadas por su probabilidad.

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_{x} (x - \mu)^2 f(x) = \sum_{x} x^2 f(x) - \mu^2$$

- La **desviación estándar** de *X* es:

$$\sigma = \sqrt{\sigma^2}$$

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución Binomial

- Es una distribución de probabilidad que cuenta el número de éxitos al realizar una serie de experimentos independientes y dicotómicos (es decir, que solo puede tener dos resultados posibles: “éxito” o “fracaso”)[1], con una probabilidad de éxito constante.

- Por ejemplo, el número de veces que sale ”cara” al lanzar una moneda 25 veces es una distribución binomial.

- Una variable aleatoria que sigue una distribución binomial se escribe como:

$$X \sim \mathrm{Bin}(n, p)$$

donde *n* es el número total de experimentos realizados, y *p* es la probabilidad de éxito de cada experimento.

[1] Son conocidos como Ensayos de Bernoulli.

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución Binomial

► La **función de probabilidad puntual (fpp)** es:

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

► La **función de distribución acumulada (Fda)** es:

$$P[X \leq x] = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

► La **esperanza** y la **varianza** son:

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = np(1-p)$$

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución Geométrica

- Es una distribución de probabilidad que define el número de ensayos de Bernoulli necesarios hasta obtener el primer resultado con éxito. Es decir, modela aquellos procesos en los que se repiten experimentos de Bernoulli hasta que se consigue uno con un resultado de éxito.

- Por ejemplo, el número de autos que pasan por una ruta hasta ver un auto amarillo es una distribución geométrica.

- Una variable aleatoria que sigue una distribución geométrica se escribe como:

$$X \sim \text{Geométrica}(p)$$

  donde *p* es la probabilidad de éxito que tienen todos los experimentos realizados.

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución Geométrica

► La **función de probabilidad puntual (fpp)** es:

$$P[X = x] = (1-p)^{x-1}p$$

► La **función de distribución acumulada (Fda)** es:

$$P[X \le x] = 1 - (1-p)^{x}$$

► La **esperanza** y la **varianza** son:

$$\mu = E(X) = 1/p$$

$$\sigma^2 = V(X) = (1-p)/p^2$$

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución de Poisson

- Es una distribución de probabilidad que define la probabilidad de que ocurra un determinado número de eventos durante un período de tiempo. Es decir, sirve para modelizar variables aleatorias que describen el número de veces que se repite un fenómeno en un intervalo de tiempo.

- Por ejemplo, el número de llamadas que recibe una central telefónica por minuto es una distribución de Poisson.

- Una variable aleatoria que sigue una distribución de Poisson se escribe como:

$$X \sim \mathrm{Poisson}(\lambda)$$

donde λ indica el número de veces que se espera que ocurra el evento estudiado durante un intervalo dado.

Distribuciones de probabilidad discretas: Ejemplos

❖ Distribución de Poisson

► La **función de probabilidad puntual (fpp)** es:

$$P[X = x] = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \qquad x! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot x$$

► La **función de distribución acumulada (Fda)** es:

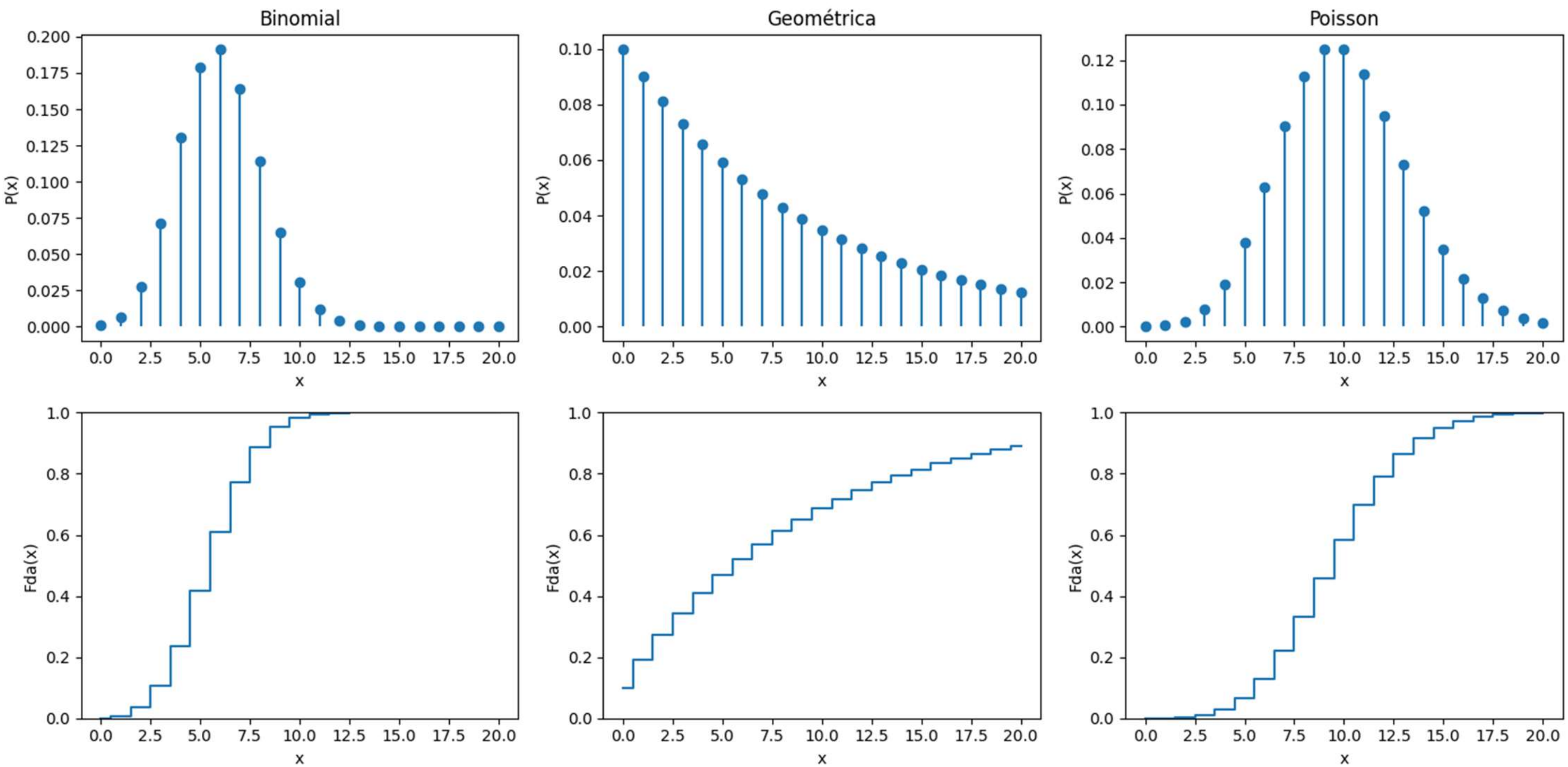
$$P(X \leq k) = \sum_{x=0}^{k} p(x)$$

► La **esperanza** y la **varianza** son:

$$\mu = E(X) = \lambda \qquad \sigma^2 = V(X) = \lambda$$

Distribuciones de probabilidad discretas: Ejemplos

- El ejemplo_1.ipynb genera los gráficos de diferentes distribuciones de Variables Aleatorias discretas:



Resumen

Introducción

Probabilidades y distribuciones

Distribuciones de probabilidad discretas

Distribuciones de probabilidad continuas

Cuantiles

Distribuciones de probabilidad continuas

- Una **distribución de probabilidad continua** es aquella distribución que define las probabilidades de una **variable aleatoria continua**.
- Por lo tanto, una distribución de probabilidad continua puede tomar infinitos valores dentro de un intervalo.
- En una distribución de probabilidad continua, a cada valor se le asocia una función de densidad de probabilidad $f(x)$ que no representa directamente la probabilidad de un punto, sino la densidad en torno a ese valor. La probabilidad de que la variable se encuentre en un intervalo $[a, b]$ se obtiene integrando la función de densidad en dicho rango. De esta forma, la integral de $f(x)$ en todo el dominio es igual a uno.

$$(1) \quad f(x) \geq 0$$

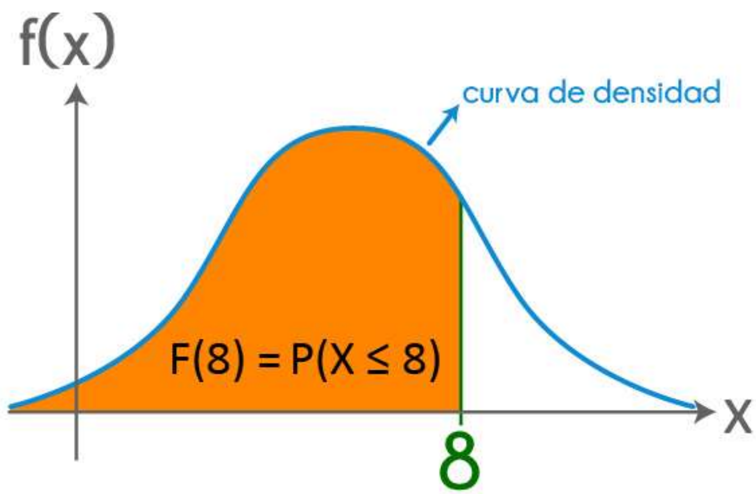
$$(2) \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

$$(3) \quad P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx$$

Distribuciones de probabilidad continuas

- La **Función de distribución acumulada (Fda)** de una **variable aleatoria continua** es:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$$



Distribuciones de probabilidad continuas

- La **esperanza de una distribución continua** se calcula como la integral de cada valor de la variable multiplicado por su probabilidad.

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

- La **varianza de una distribución continua** se calcula como la integral de los cuadrados de las diferencias entre cada valor y la esperanza, ponderadas por su probabilidad.

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

- La **desviación estándar** de *X* es:

$$\sigma = \sqrt{\sigma^2}$$

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Uniforme

- Es una distribución de probabilidad en la que todos los valores dentro de un intervalo tienen la misma probabilidad de ocurrir.

- Se utiliza para describir variables con probabilidad constante y procesos aleatorios donde ningún resultado es más probable que otro. Por ejemplo, si un experimento puede dar cualquier valor entre 5 y 9 con igual probabilidad.

- Una variable aleatoria que sigue una distribución uniforme se escribe como:

$$X \sim U(a, b)$$

donde a y b definen el intervalo de equiprobabilidad.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Uniforme

► La **función de densidad de probabilidad (fdp)** es:

$$f(x) = \frac{1}{b-a} \qquad\qquad x \in [a, b]$$

► La **función de distribución acumulada (Fda)** es:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x < b \\ 1 & \text{si } x \geq b \end{cases}$$

► La **esperanza** y la **varianza** son:

$$\mu = E(X) = \frac{(a+b)}{2} \qquad\qquad \sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal

- Es una distribución de probabilidad cuya gráfica tiene forma de campana y es simétrica respecto a su media. En Estadística, la **distribución normal (o gaussiana)** se considera la distribución más importante de todas.

- Permite modelizar un gran número de fenómenos reales de características muy diferentes. Por ejemplo, la estatura de los alumnos de un curso, el número de piezas defectuosas producidas en una fábrica durante un día.

- Además, se puede usar para aproximar otros tipos de distribuciones bajo ciertas condiciones (Teorema del Límite Central).

- Una variable aleatoria que sigue una distribución normal se escribe como:

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

donde μ y σ definen su media y su desviación estándar, respectivamente.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal

- La **función de densidad de probabilidad (fdp)** es:

$$P[X = x] = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- La **función de distribución acumulada (Fda)** es:

$$P[X \leq x] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx, \quad x \in \mathbb{R}$$

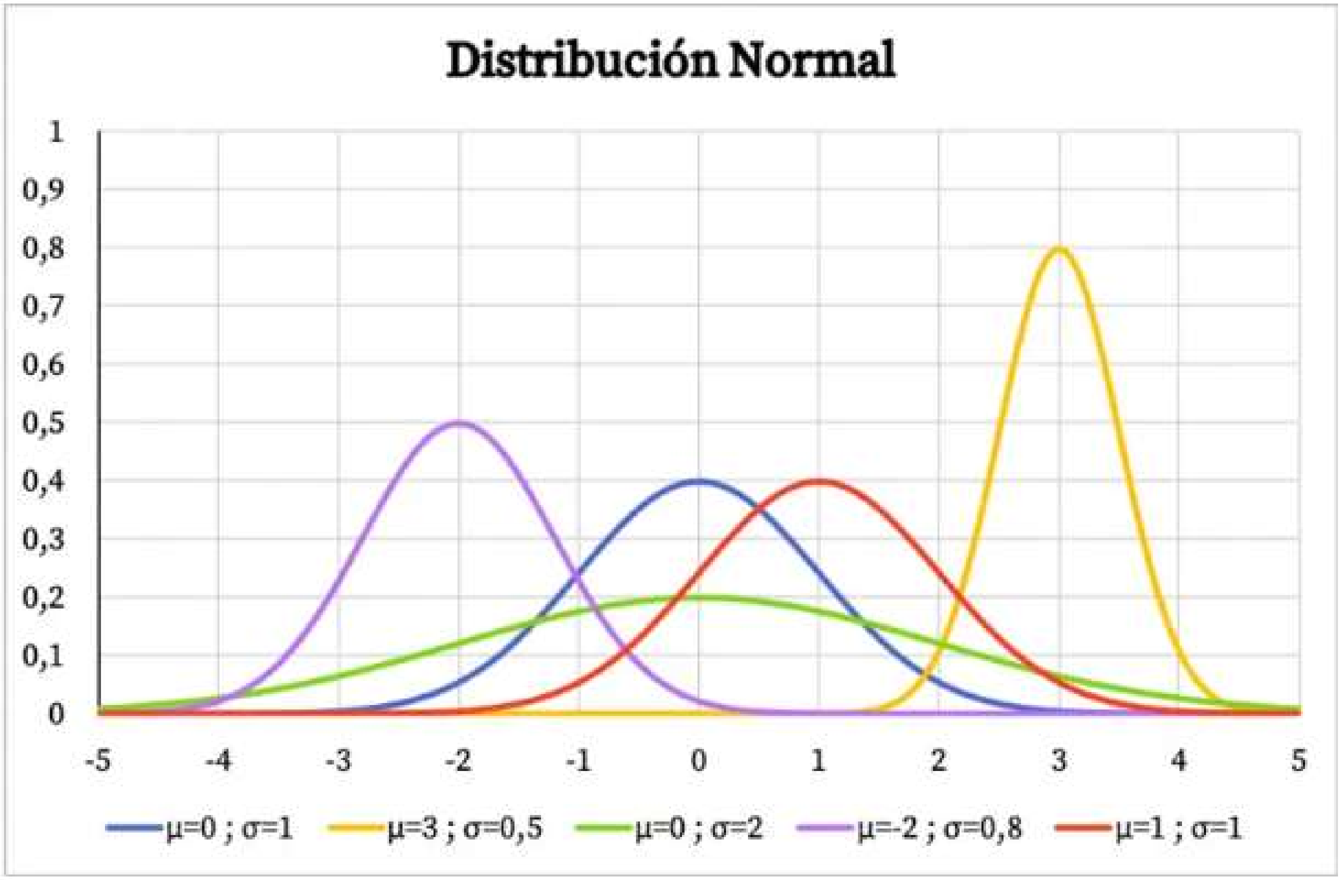
- La **esperanza** y la **varianza** son:

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

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal

► La gráfica de la **fdp de la distribución normal** es:



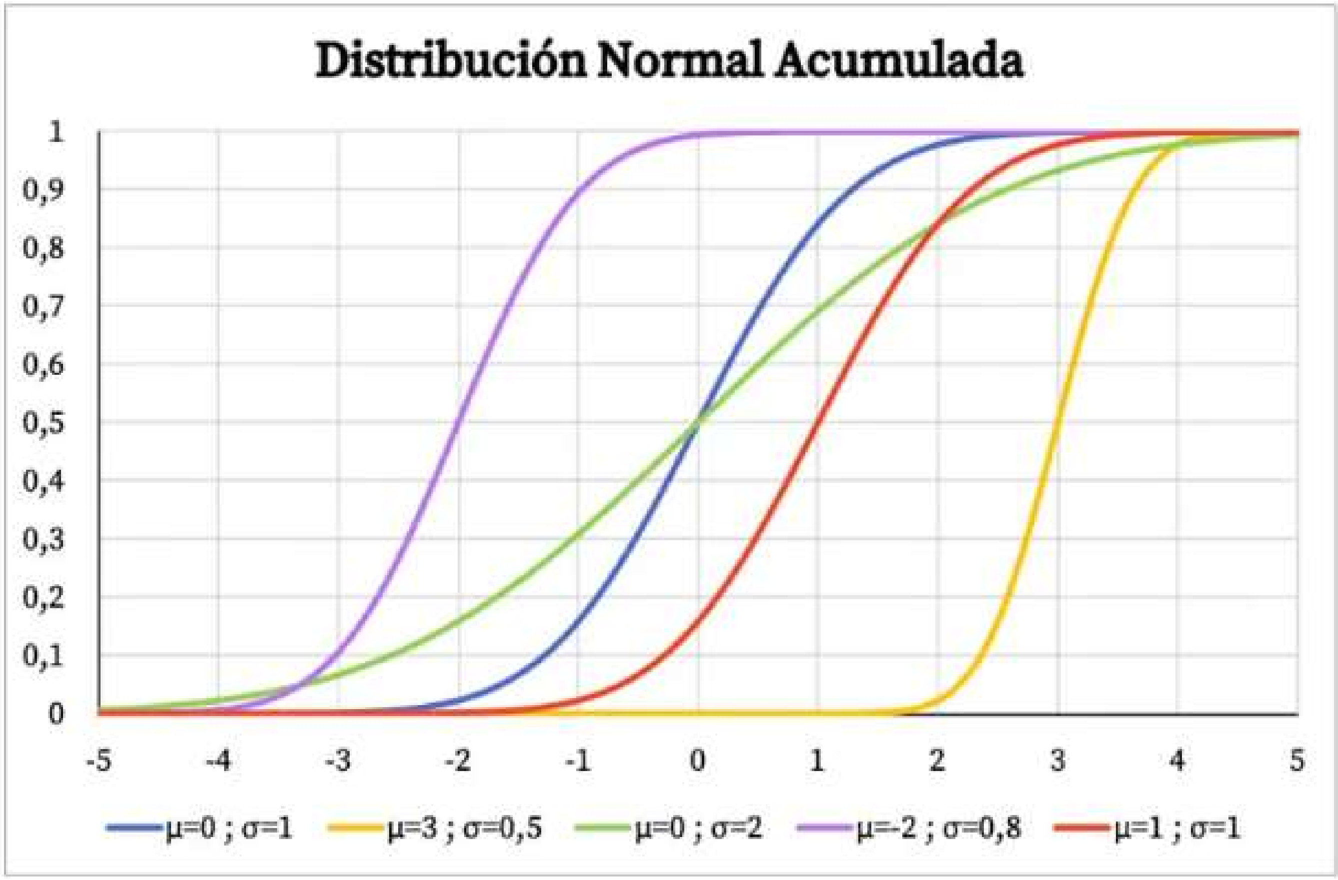
La **distribución normal** tiene forma de campana centrada en la media, donde los valores cercanos a ella son más frecuentes que los extremos.

Una mayor desviación típica genera una curva más ancha y aplanada en su representación gráfica.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal

► La gráfica de la **Fda de la distribución normal** es:



También depende de los valores de su media y de su desviación típica.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal Estándar

- La **distribución normal estándar**, también llamada **distribución normal unitaria**, es el caso más simple de una distribución normal. En concreto, la distribución normal estándar es una distribución normal con valores de media y desviación estándar iguales a 0 y 1 respectivamente.

$$N(0,1) \longrightarrow \begin{cases} \mu = 0 \\ \sigma = 1 \end{cases}$$

- Cualquier distribución normal se puede transformar en una distribución normal estándar aplicando un proceso llamado **tipificación (o normalización)** de una variable, que consiste en restar a cada uno de los valores su media aritmética y después dividir por su desviación típica.

$$\boxed{Z = \frac{X-\mu}{\sigma}}$$

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

Fda de una VA normal estándar

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z) = \Phi(z)$$

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Normal Estándar

► Tabla de la **distribución normal estándar:**

Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

z0	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	2,0
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	2,1
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	2,2
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	2,3
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	2,4
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	2,5
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	2,6
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	2,7
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	2,8
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	2,9
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990	3,0
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993	3,1
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	3,2
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	3,3
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	3,4
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	3,5
3,6	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	3,6
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	3,7
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	1,0000	1,0000	1,0000	3,8
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	3,9

probabilidadyestadistica.net

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Chi-cuadrado

- Es una distribución de probabilidad que es la suma del cuadrado de *k* variables aleatorias independientes con distribución normal. Se dice que la **distribución chi-cuadrado (o distribución de Pearson)** tiene *k* grados de libertad.

- Por lo tanto, una distribución chi-cuadrado tiene tantos grados de libertad como la suma de los cuadrados de variables con distribución normal que representa.

- Su símbolo es χ^2.

$$X \sim \chi^2_k \longrightarrow \begin{array}{l}\text{Distribución chi-cuadrado}\\ \text{con k grados de libertad}\end{array}$$

- La distribución chi-cuadrado se utiliza mucho en inferencia estadística, por ejemplo, se usa en el contraste de hipótesis e intervalos de confianza.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución Chi-cuadrado

► La **función de densidad de probabilidad (fdp)** es:

$$P[X = x] = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$ donde $\Gamma(z)$ es la **función gamma**

(https://es.wikipedia.org/wiki/Funci%C3%B3n_gamma)

► La **función de distribución acumulada (Fda)** es:

$$P[X \leq x] = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)}$$

► La **esperanza** y la **varianza** son:

$$E(X) = k \qquad V(X) = 2k$$

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución t de Student

- Es una distribución de probabilidad que se la utiliza para estimar la media de una población normal cuando el tamaño muestral es pequeño y la desviación estándar poblacional es desconocida, así como para contrastar diferencias entre dos medias **(Test de Student)**.

- Su forma depende de los grados de libertad, que se calculan restando una unidad al número de observaciones: $\nu = n - 1$.

- Una variable aleatoria que sigue una distribución t de Student se escribe como: $X \sim t_\nu$

- La distribución t de Student se utiliza mucho en inferencia estadística, por ejemplo, se usa en el contraste de hipótesis y en intervalos de confianza.

Distribuciones de probabilidad continuas: Ejemplos

❖ Distribución t de Student

► La **función de densidad de probabilidad (fdp)** es:

$$P[X=x]=\frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1+x^2/\nu)^{-(\nu+1)/2}$$

donde $\Gamma(z)$ es la **función gamma**

► La **función de distribución acumulada (Fda)** es:

$$P[X\leq x]=\frac{1}{2}+x\Gamma\left(\frac{\nu+1}{2}\right)\cdot\frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$$

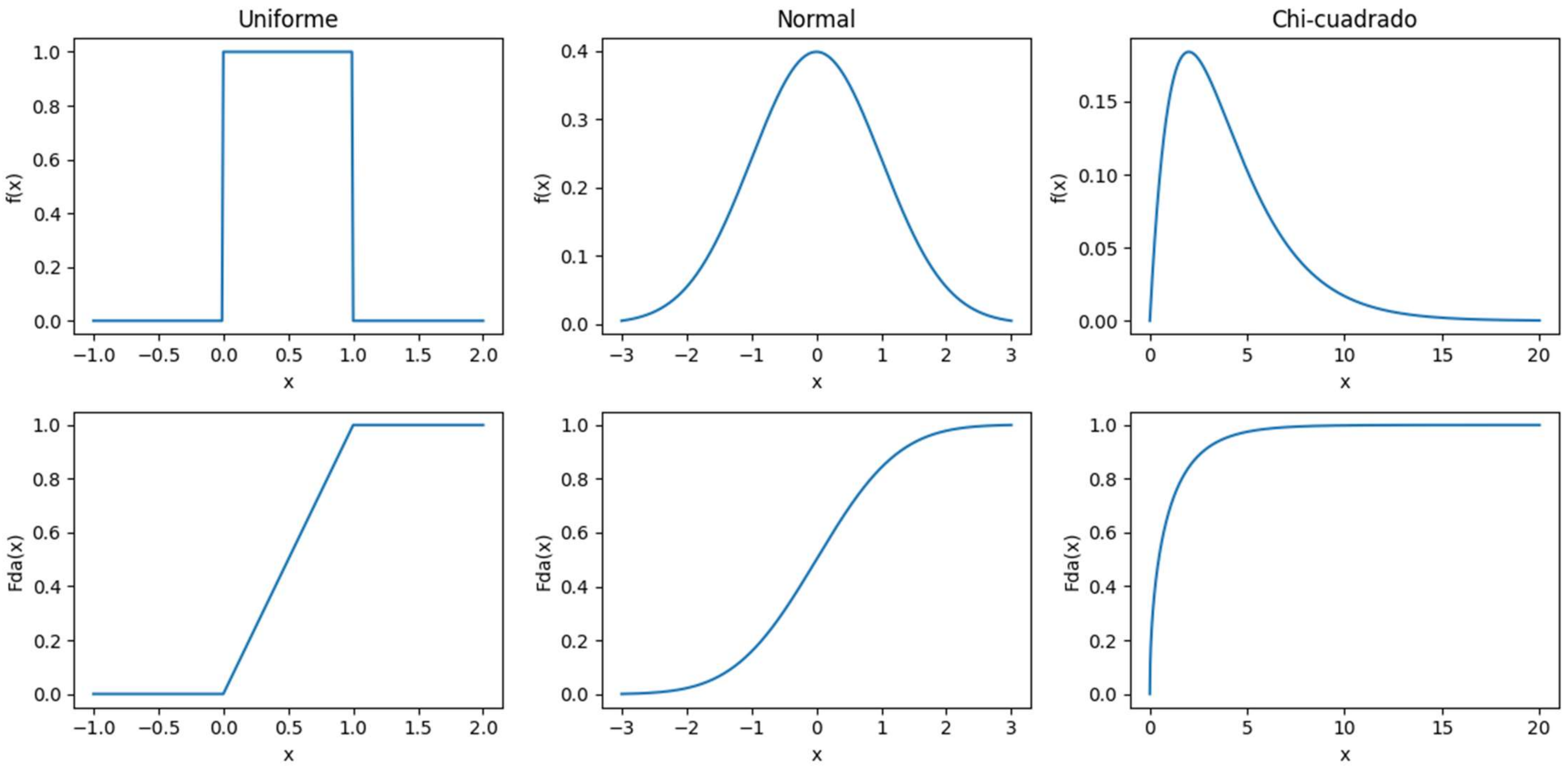
► La **esperanza** y la **varianza** son:

$$E(X)=0$$

$$V(X)=\frac{\nu}{\nu-2},\nu>2$$

Distribuciones de probabilidad continuas: Ejemplos

- El ejemplo_2.ipynb genera los gráficos de diferentes distribuciones de Variables Aleatorias continuas:



Resumen

Introducción

Probabilidades y distribuciones

Distribuciones de probabilidad discretas

Distribuciones de probabilidad continuas

Cuantiles

Cuantiles:

- Un **cuantil** es un punto que divide al conjunto de datos en dos subconjuntos: uno con valores menores o iguales que él y el otro con valores mayores que él.
- Un **cuantil** *c* (donde 0 ≤ *c* ≤ 1, es decir, es una probabilidad) será un número x_c tal que a su izquierda deje una probabilidad de *c* y a su derecha una probabilidad de 1-*c*.
- Por ejemplo, para calcular $x_{0,25}$ de una distribución Normal con media $\mu = 0$ y desviación típica $\sigma = 1$ (se utiliza la función ppf):

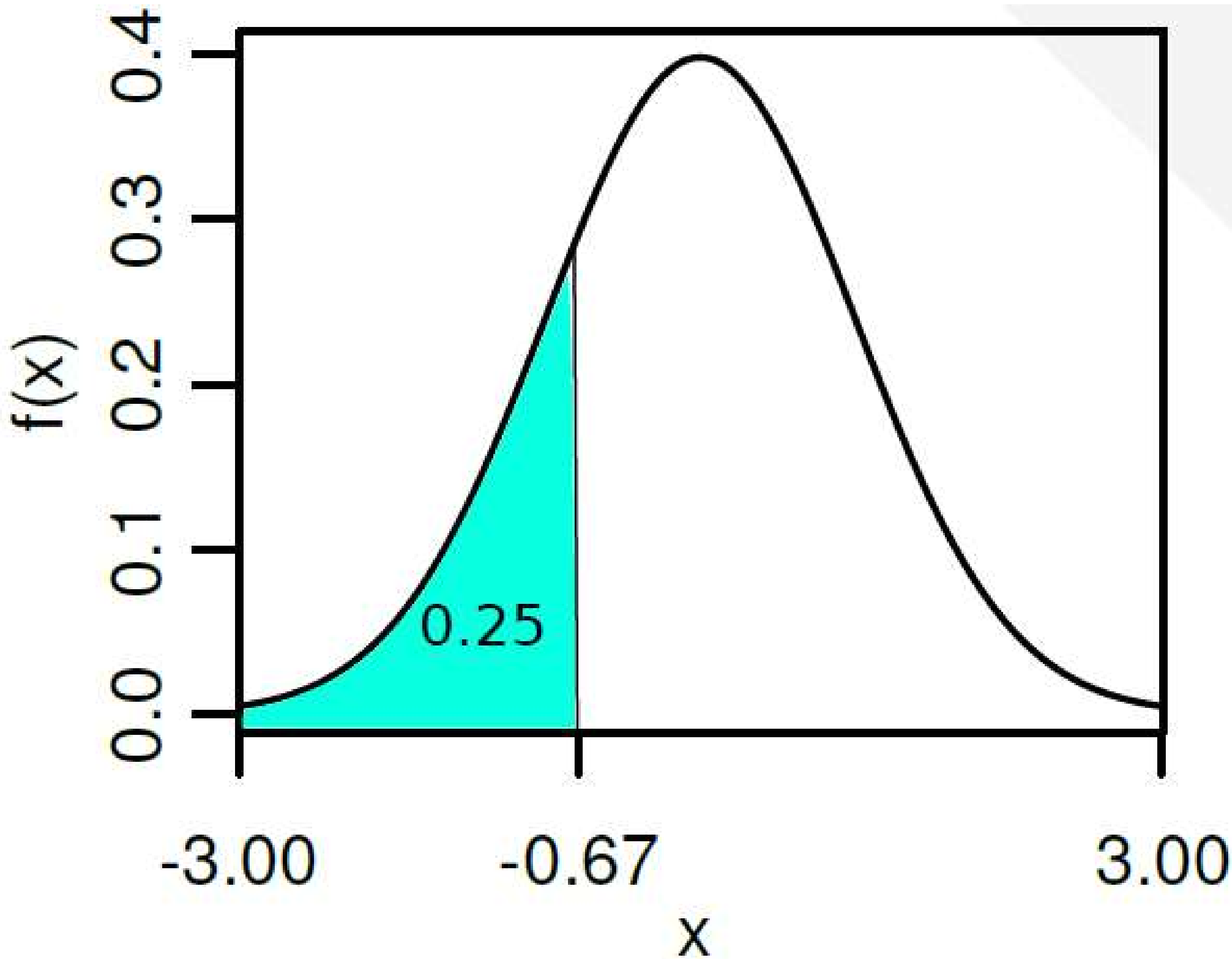
```
# Calculo el cuantil x_{0.25}
c = 0.25
xc = norm.ppf(c)  # Equivalente a qnorm(c) en R

# Probabilidad acumulada a la izquierda de xc
z = norm.cdf(xc)  # Equivalente a pnorm(xc) en R
```

Código/ejemplo_2.ipynb

da como resultado $x_{0,25} = -0{,}6744898$.

Cuantiles: Ejemplo

► Ejemplo de **cuantil** $x_{0,25}$ (primer cuartil Q1)



Números aleatorios

- Muchas veces es útil poder generar **números pseudo-aleatorios** con una distribución dada. El comando en la librería Numpy tiene la forma *np.random.distribucion*.

- Por ejemplo, para generar números con distribución Normal con media $\mu = 0$ y desviación típica $\sigma = 1$:

```python
# Setea la semilla
np.random.seed(182)

# Genera 5 números aleatorios N(0,1)
y = np.random.normal(loc=0, scale=1, size=5)
print("Números aleatorios N(0,1):", y)
```

Código/ejemplo_2.ipynb

da como resultado[2]:

```
Números aleatorios N(0,1): [ 0.26416281 -0.34912185 -0.11645481  1.05844668  1.22340598]
```

[2]¡Cuidado! Para obtener los mismos valores se debe setear la semilla.

Números aleatorios

- Es posible también generar **números pseudo-aleatorios** con una distribución dada, utilizando la librería Scipy y el módulo Stats. En este caso la forma es: *distribucion.rvs(df=k, size=n)*.

- Por ejemplo, para generar números con distribución Chi cuadrado con 4 grados de libertad:

```
# Parámetros
n = 10000
k = 4  # Grados de libertad

# Genero n números con distribución chi-cuadrado con k grados de libertad
y = chi2.rvs(df=k, size=n)
```

Código/ejemplo_2.ipynb

previamente se debe haber importado la distribución:

```
from scipy.stats import chi2
```

Resumen de funciones principals en scipy.stats

Función	Nombre	Uso principal	Ejemplo
`rvs()`	Random Variates Sampler	Generar valores aleatorios según la distribución	`norm.rvs(size=5)`
`pdf(x)`	Probability Density Function	Densidad de probabilidad en un punto (continuas)	`norm.pdf(0)`
`pmf(k)`	Probability Mass Function	Probabilidad puntual (discretas)	`binom.pmf(2, n=5, p=0.5)`
`cdf(x)`	Cumulative Distribution Function	Probabilidad acumulada $P(X \leq x)$	`norm.cdf(1.96)`
`ppf(q)`	Percent Point Function	Cuantil (inversa de la CDF)	`norm.ppf(0.975)`