EDEM

Máster en Data Analytics

Estadística con Python II

Miguel Rua del Barrio

Conceptos básicos

EDEM
Escuela de Empresarios

Función de Probabilidad, P, es una función que a cada suceso A del espacio muestral Ω le hace corresponder un número P (A), y que verifica:

1. O ≤ P(A) ≤ 1, ∀ suceso A de Ω
2. P(Ω) = 1
3. Si $A_1, A_2, \ldots, A_n$ son sucesos disjuntos, entonces $P(\bigcup_{i=1}^{n} A_n) = \sum_{i=1}^{n} P(A_n)$

Definición:

EDEM
Escuela de Empresarios

- Se llama **función de distribución** de la variable aleatoria X a la función $F : R \to [0, 1]$ dada por $F(x) = P(X \leq x)$.

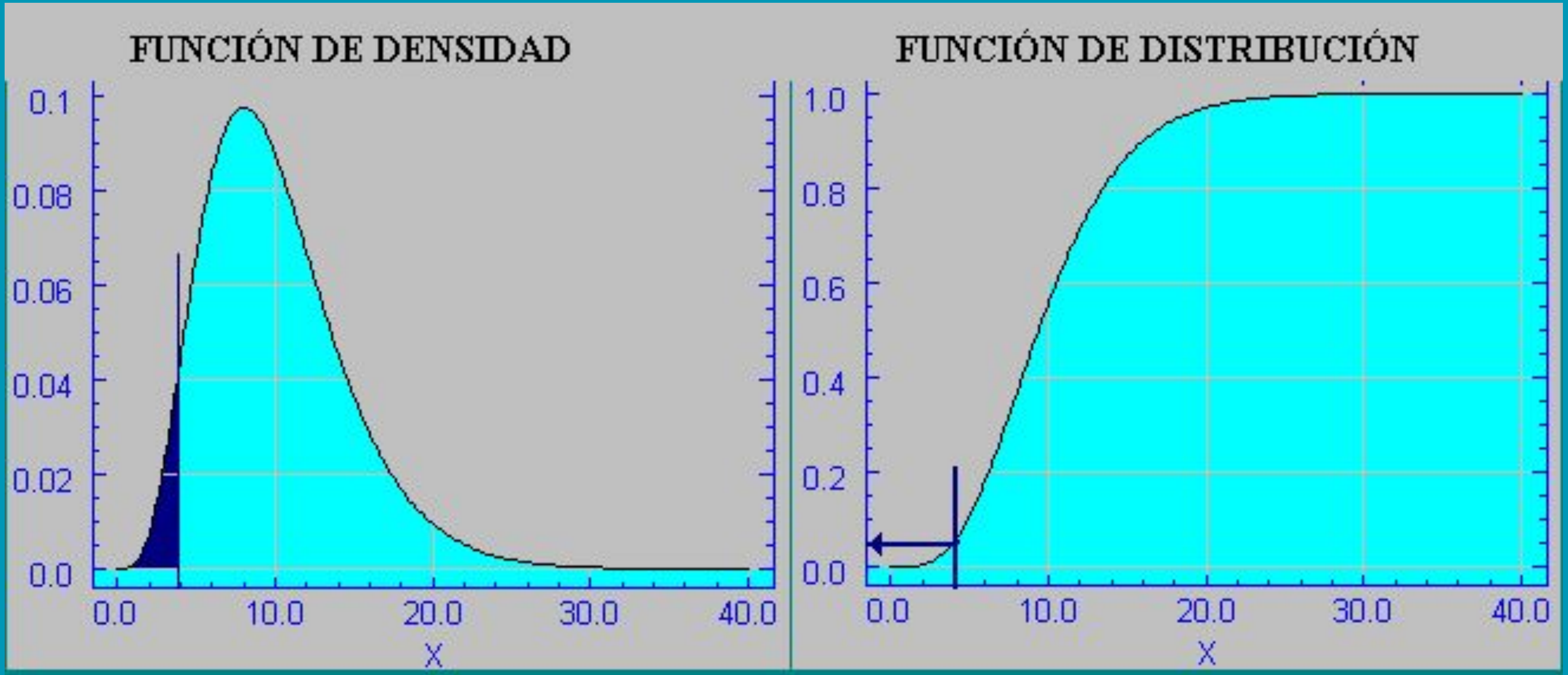
Si X es una v.a.continua con función de densidad $f(x)$, entonces

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

- La **función de densidad** f(x) describe la probabilidad de que la variable tome un determinado valor.

- La **función de distribución** F(x) describe la probabilidad acumulada de la variable hasta ese valor.

Distribuciones de probabilidad:



$P(X \leq 4)$

Distribuciones de probabilidad:

EDEM
Escuela de Empresarios

1. **DISTRIBUCIONES DISCRETAS**
2. **DISTRIBUCIONES CONTINUAS**
3. **RELACIÓN ENTRE LAS PRINCIPALES DISTRIBUCIONES**

Funciones de probabilidad:

EDEM
Escuela de Empresarios

Distribuciones discretas:

- Distribución Uniforme Discreta
- Distribución de Bernouilli
- Distribución Binomial
- Distribución de Poisson

Distribuciones continuas:

- Distribución normal
- Distribución chi-cuadrado de Pearson
- Distribución t de Student
- Distribución F de Fisher-Snedecor

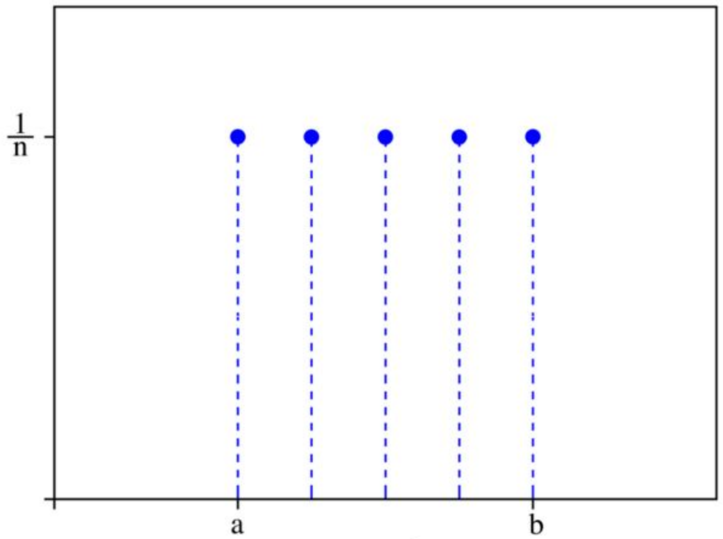
Distribución uniforme discreta

EDEM
Escuela de Empresarios

$$P(X = x_i) = \begin{cases} \frac{1}{n} & si\ i = 1,\ 2,\ 3,\ \ldots,\ n \\ 0 & resto \end{cases}$$

Esperanza: $E[X] = \frac{n+1}{2}$

Varianza: $V[X] = \frac{n^2-1}{12}$



Distribución uniforme discreta.

Distribución uniforme discreta

EDEM Escuela de Empresarios

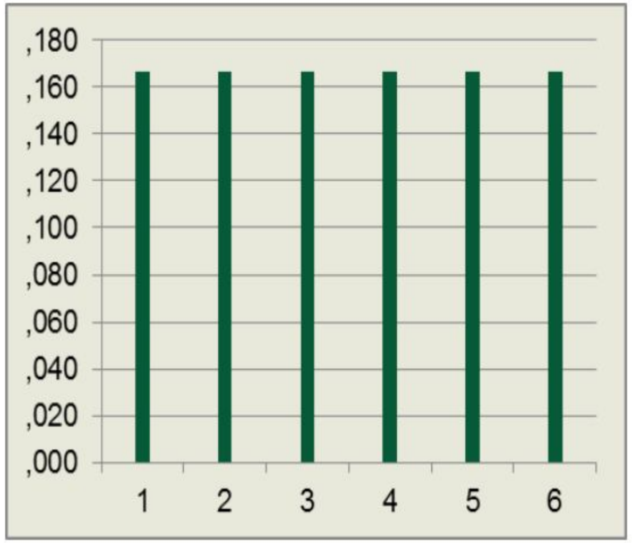
Ejemplo: Puntuación obtenida al lanzar un dado

Se distribuye uniformemente sobre 6 puntos:

$\boldsymbol{x_1}$=1, $\boldsymbol{x_2 = 2}$, $\boldsymbol{x_3 = 3}$, $\boldsymbol{x_4}$=4, $\boldsymbol{x_5 = 5}$, $\boldsymbol{x_6 = 6}$.

Su función de probabilidad o masa es:

$$P(X= x_i)=\begin{cases} \frac{1}{6} & si\ i = 1,\ 2,\ 3,\ 4,5,6 \\ 0 & resto \end{cases}$$



$$E[X] = \frac{n+1}{2} = \frac{6+1}{2} = 3.5$$

$$V[X] = \frac{n^2-1}{12} = \frac{6^2-1}{12} = 2.92$$

Distribución de Bernouilli

EDEM
Escuela de Empresarios

La Función de Probabilidad o Masa de una variable X con Distribución de Bernoulli viene dada por

P(X=x) = $\boldsymbol{p^x(1-p)^{1-x}}$ **para x = 0, 1.**

La esperanza de una variable aleatoria de Bernoulli es E[X] = p.

La varianza viene dada por V[X] = p(1-p).

Ejemplo:
Lanzar una moneda al aire y considerar la variable aleatoria X=(Número de caras obtenidas), en cuyo caso X=0 si: 1-p=1/2, y X=1 si p=1/2

Distribución Binomial

EDEM
Escuela de Empresarios

Distribución de la variable aleatoria

X = "Nº de éxitos obtenidos en n pruebas de Bernoulli

P(E) = p en cada una de ellas"

X = $\mathbf{X_1} + \mathbf{X_2} + \mathbf{X_3} + \ldots + \mathbf{X_n}$ Donde cada una de las Xi es una distribución de Bernouilli

Función de probabilidad:

P(X=x) = $\binom{n}{x}\boldsymbol{p^x(1-p)^{n-x}}$ para cada $\boldsymbol{x = 0, 1}$,…,n

donde $\binom{n}{x} = \frac{n!}{(n-x)!x!}$ con 0!=1.

Esperanza: E[X] = np;

Varianza: V[X] = np(1-p).

Distribución Binomial

EDEM
Escuela de Empresarios

Ejemplos de variables que siguen una distribución binomial:

- **Número de caras cuando lanzamos una moneda 100 veces,**
- **Número de aprobados entre los matriculados una asignatura,**
- **Número de empresas españolas que han superado el millón de euros de beneficios,**
- **Número de vehículos que han superado los 140 km/h de los 187 que han pasado por un determinado punto.**

Distribución Binomial

Ejemplos de variables que siguen una distribución binomial:

Por estudios anteriores, se sabe que el 20% de los alumnos matriculados en un determinado curso no asistirán al mismo. Si el aula en el que se va a impartir el curso sólo dispone de 30 puestos, ¿Cuál es la probabilidad de que todas las personas que asistan al curso tengan un puesto disponible, si se han inscrito 35? ¿Cuál es el número esperado de asistentes?

Distribución de Poisson

Distribución de los sucesos "raros"

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

$\lambda = E(x)$

K: es el número de éxito cuya probabilidad se está calculando

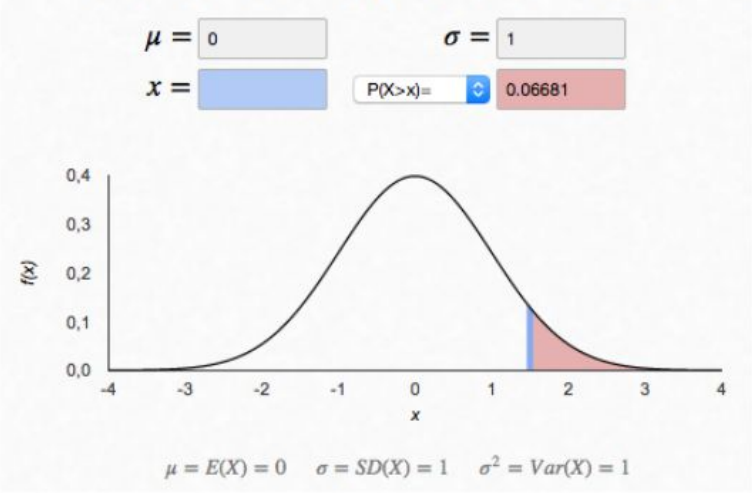
Distribución Normal

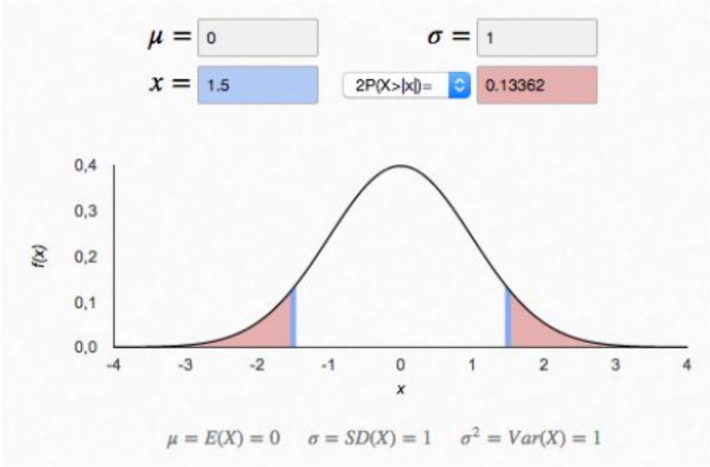
EDEM
Escuela de Empresarios

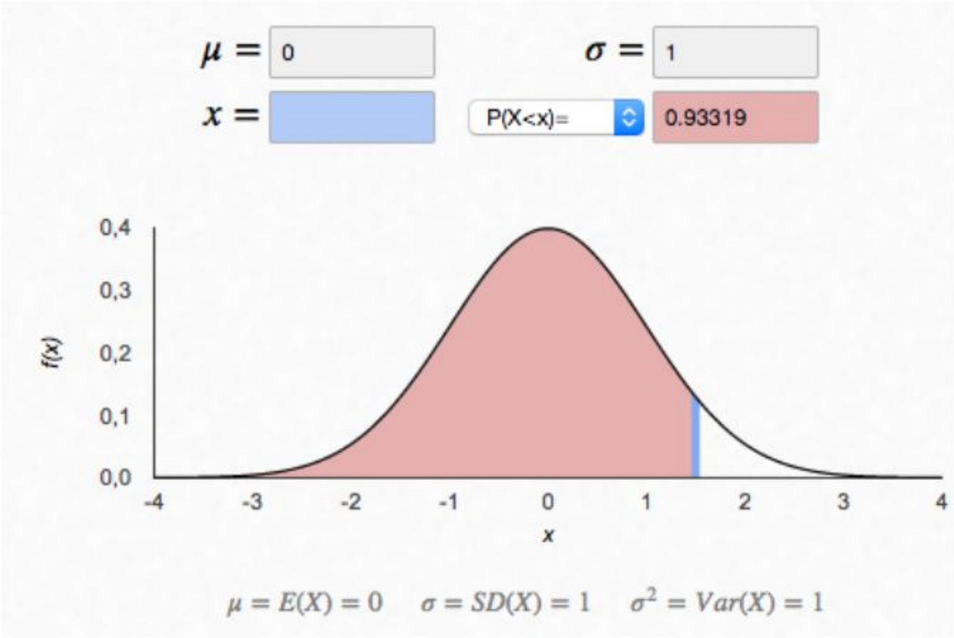
$$f(x)= \frac{1}{\sigma\sqrt{2\pi}}\, e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]},\ \forall\, x \in \mathcal{R}.$$

La esperanza de una variable aleatoria X con distribución normal viene dada por $E[X] = \mu$ y su varianza por $V[X] = \sigma^2$.

Funciones de probabilidad: Distribución Normal

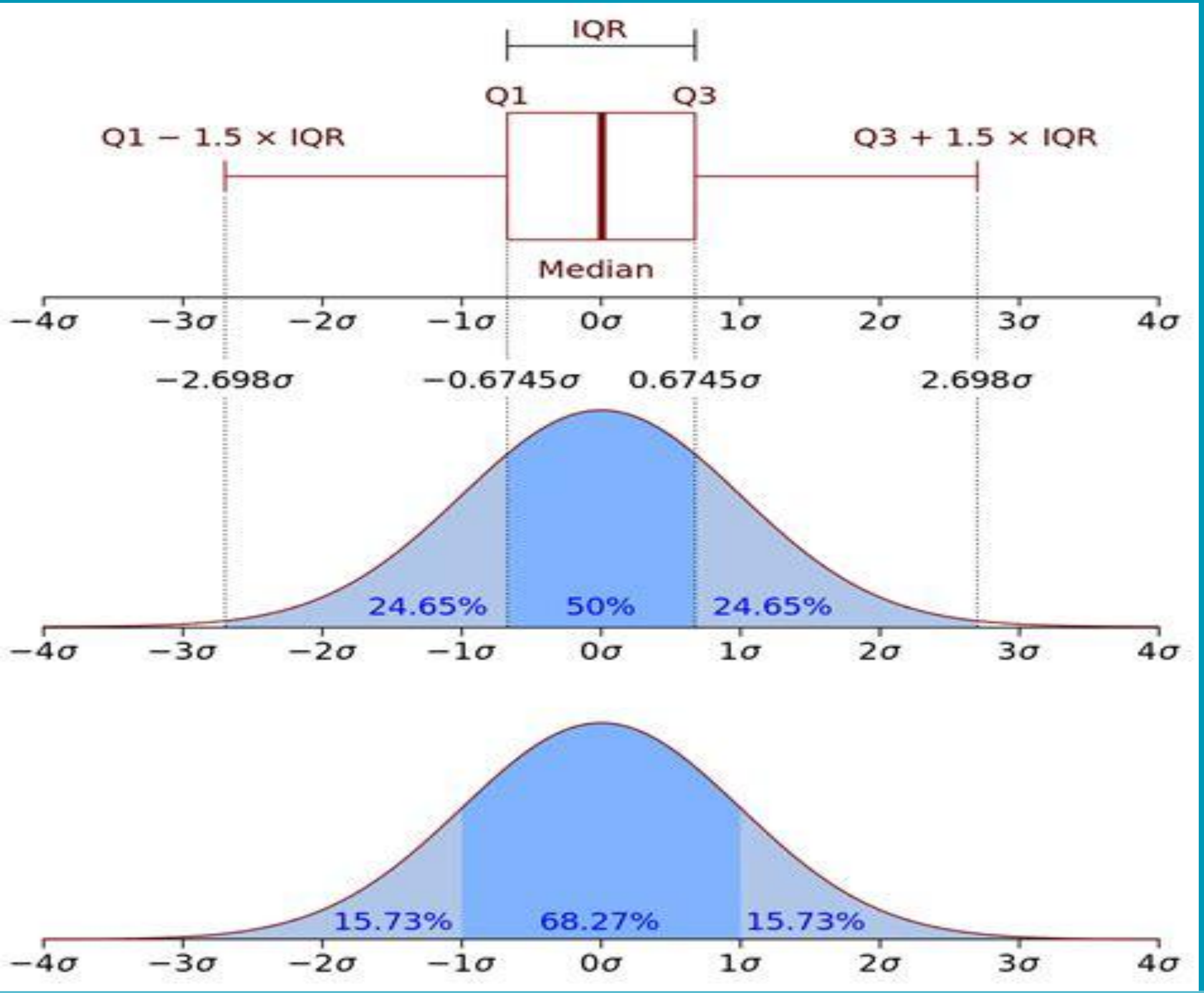






Distribuciones de probabilidad:

EDEM
Escuela de Empresarios



Teorema del Límite Central

EDEM
Escuela de Empresarios

Funciones de probabilidad: Teorema del límite central afirma que dadas X1 ,…,Xn variables aleatorias independientes e idénticamente distribuidas de una población X de tamaño n, con media μ y desviación típica σ, se verifica que:

Si X~N(μ, σ) entonces

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ para cualquier n}$$

Si X no es Normal,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ para n grande}$$

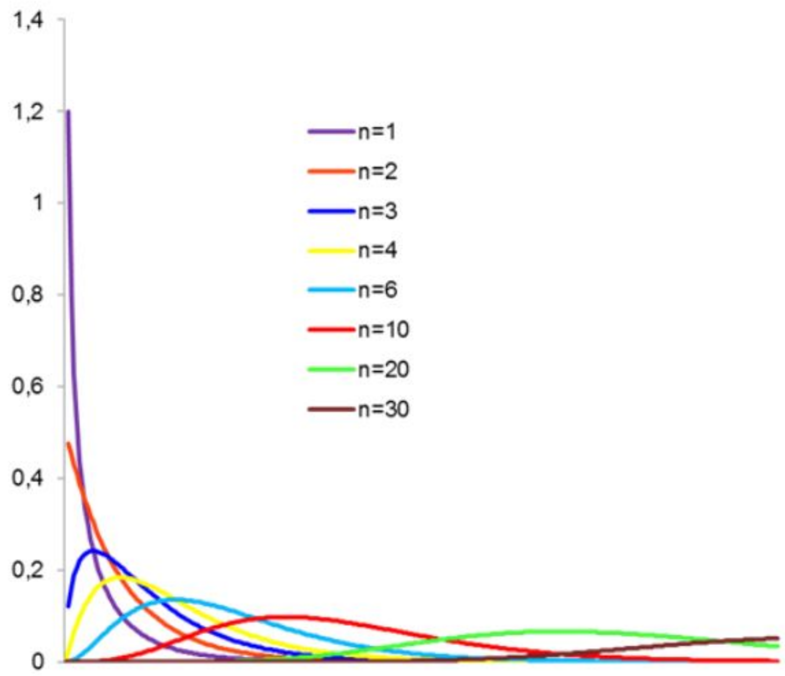
Es decir, cuanto más asimétrica sea la distribución, mayor tiene que ser el valor de n para que sea aceptable la aproximación.

Distribución χ2 de Pearson

EDEM
Escuela de Empresarios

Sean Z1 , Z2 , …,Zn v. a. independientes con distribución N(0,1). La Distribución χ 2 de Pearson con n grados de libertad, χ 2 n , es la distribución de la variable aleatoria:

$$\sum_{i=1}^{n} Z_i^2$$



E(x)= n

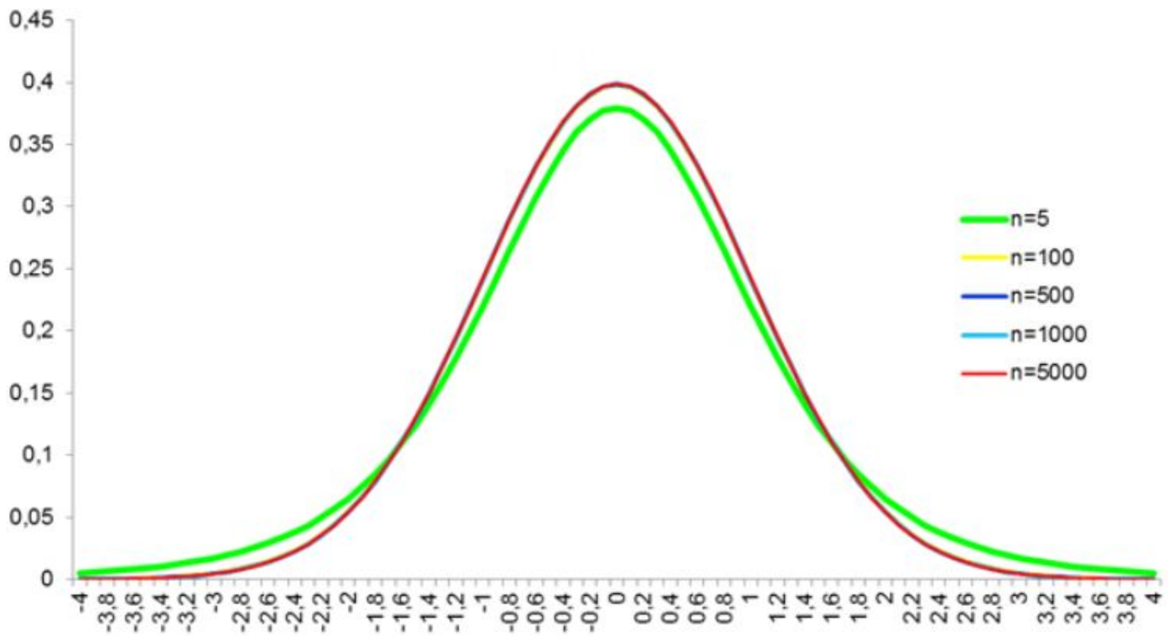
V(x)= 2n

Distribución t de Student

EDEM
Escuela de Empresarios

Sean Z1 , Z2 , …,Zn v. a. independientes con distribución N(0,1). La Distribución t de Student con n grados de libertad, tn , es la distribución de la variable aleatoria:

$$\frac{Z}{\sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n}}} = \frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}$$

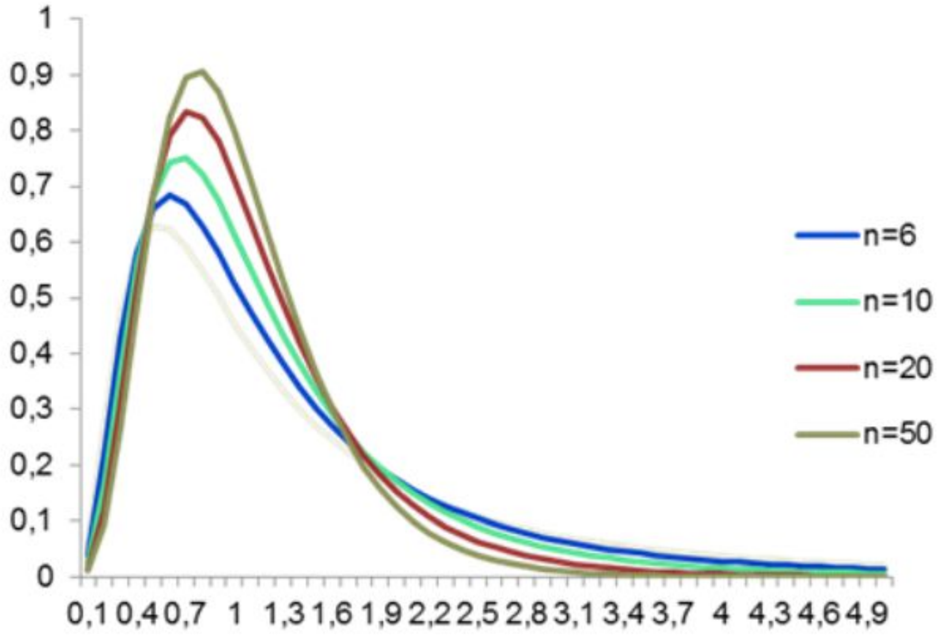


Su esperanza es E$[t_n]$=0 si $n > 1$, y su varianza V$[t_n]$=$\frac{n}{n-2}$ $para\, n > 2$

Distribución F de Fisher-Snedecor

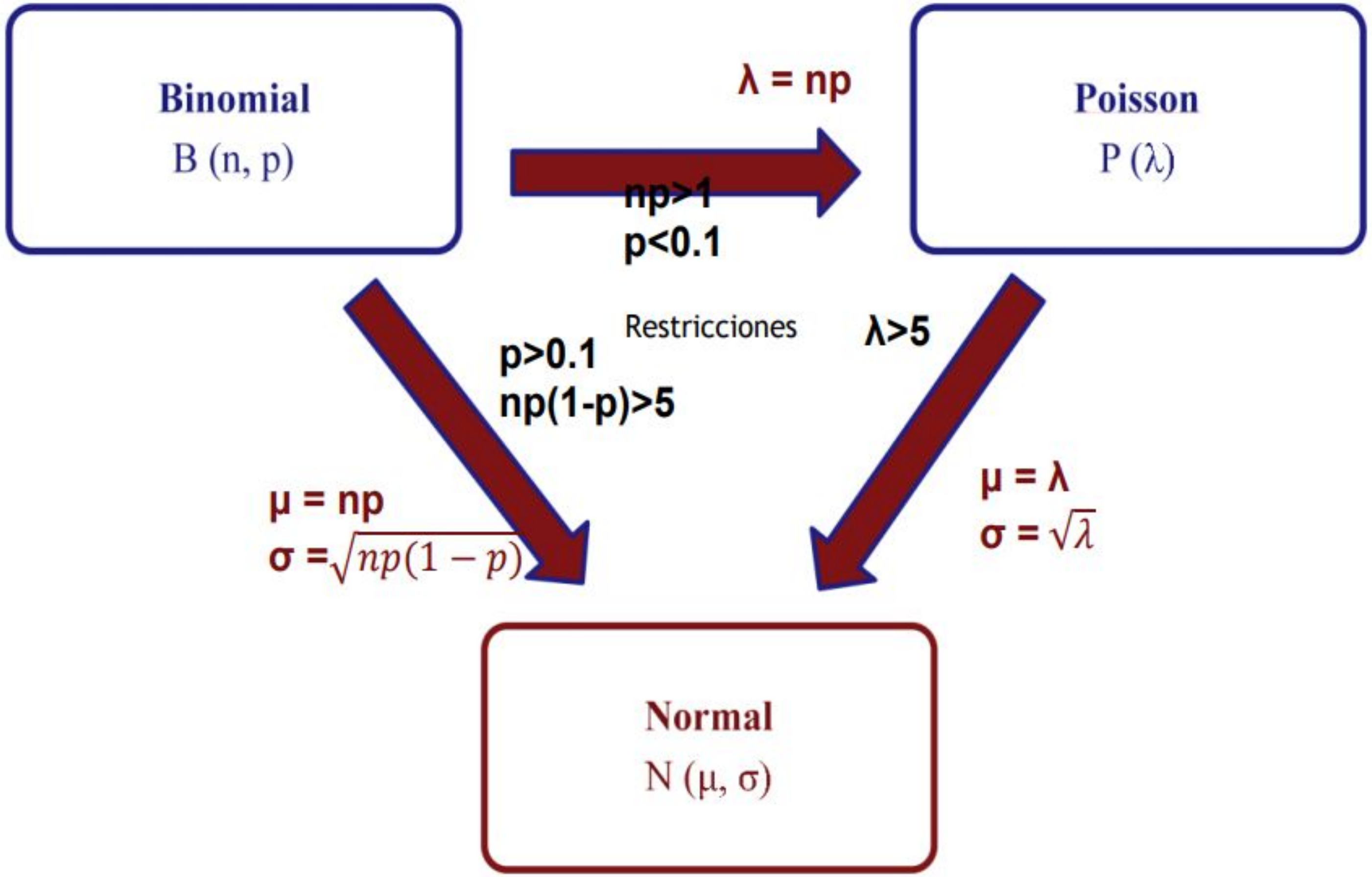
Sean X1 , …, Xm, Y1 , …,Yn variables aleatorias independientes todas ellas N(0,1). La Distribución F de Fisher-Snedecor con m y n grados de libertad, Fm:n , es la distribución de la variable aleatoria:

$$F = \frac{\frac{\sum_{i=1}^{m} X_i^2}{m}}{\frac{\sum_{i=1}^{n} Y_i^2}{n}} = \frac{\frac{\chi^2{}_m}{m}}{\frac{\chi^2{}_n}{n}}$$



Su esperanza viene dada por $E[F_{m:n}] = \frac{m}{n-2}$ $\boldsymbol{para\ n > 2.}$

Relación entre distribuciones



Práctica con Python

EDEM
Escuela de Empresarios

Parte práctica de la sesión:

1. Distribuciones de probabilidad con Python

Resumen

Durante esta unidad hemos visto:

- **Definición de función de probabilidad, distribución y densidad.**
- **Definición de las distribuciones más conocidas**
- **Trabajar con dichas distribuciones en Python con simulación**
- **Hemos hecho hincapié en la distribución normal**
- **Realización de informes estadísticos con Google Colab.**

Siguientes objetivos:

- **Contrastes de hipótesis.**
- **ANOVA.**

EDEM
Escuela de Empresarios

¡GRACIAS POR VUESTRA ATENCIÓN!

miguel.ruadelbarrio@ams-europe.com

linkedin.com/in/miguel-rua-del-barrio-5214661b5