Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

MODELADO PROBABILÍSTICO

Distribuciones de Probabilidad

BORJA CALVO • borja.calvo@ehu.es

DOMINIO DE DEFINICIÓN

- Categórica, ordinal o continua
- Acotada o no acotada

MEDIA (1ER MOMENTO)

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

ASPECTO DE LA FUNCIÓN

- Depende de los parámetros
- Caracterización por medio de los *momentos*
- Los más habituales, *media y varianza*

MOMENTOS EN TORNO AL CERO

$$\mu_k' = E[X^k] = \int_{-\infty}^{\infty} x^k f(x)\, dx$$

MOMENTOS EN TORNO A LA MEDIA

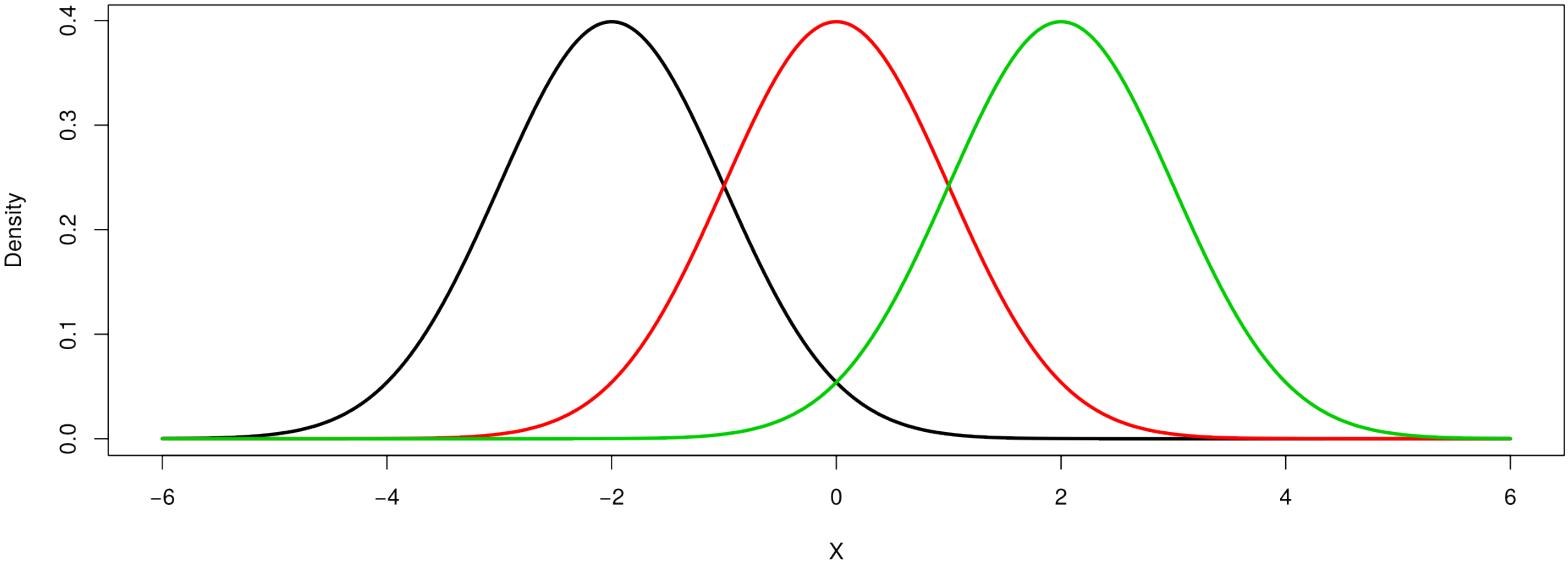
$$\mu_k = E[(X - \mu_1')^k] = \int_{-\infty}^{\infty} (x - \mu_1')^k f(x)\, dx$$

MOMENTOS MUESTRALES

$$m_k' = \frac{1}{n} \sum_{i=1}^{n} x_i^k \qquad m_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_1')^k$$

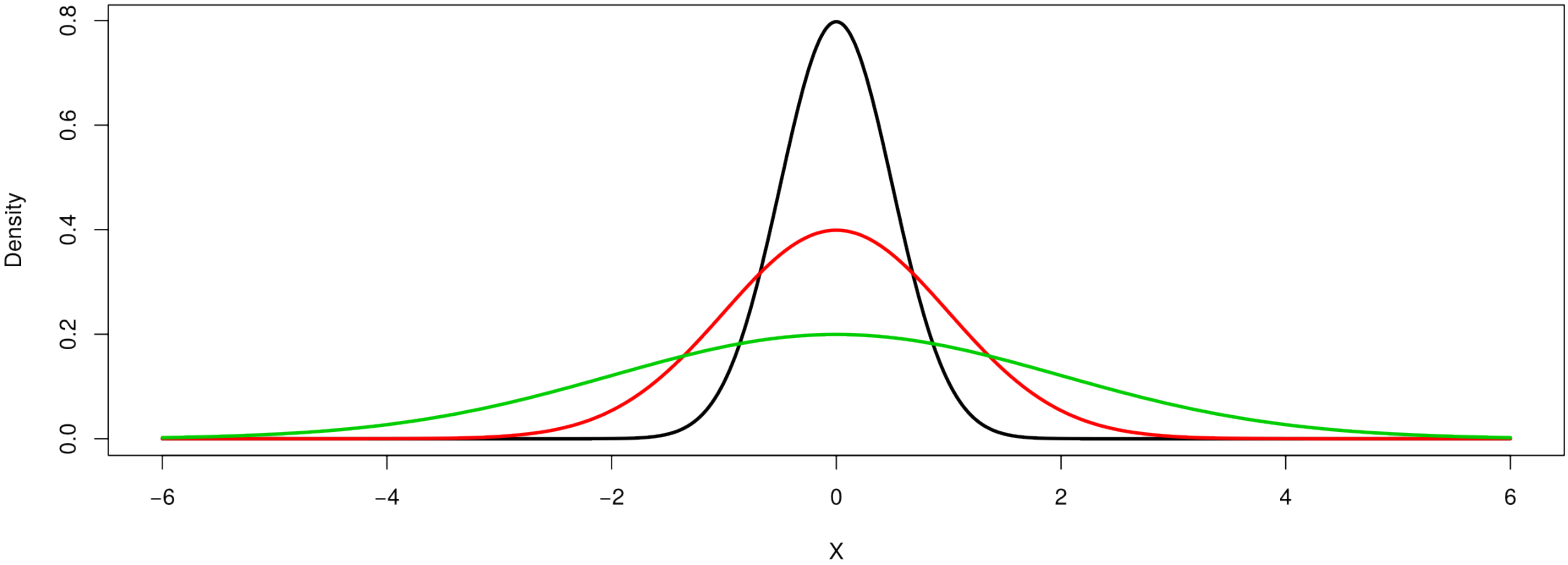
1ER MOMENTO: LA MEDIA

El primer momento nos da idea de la *localización*



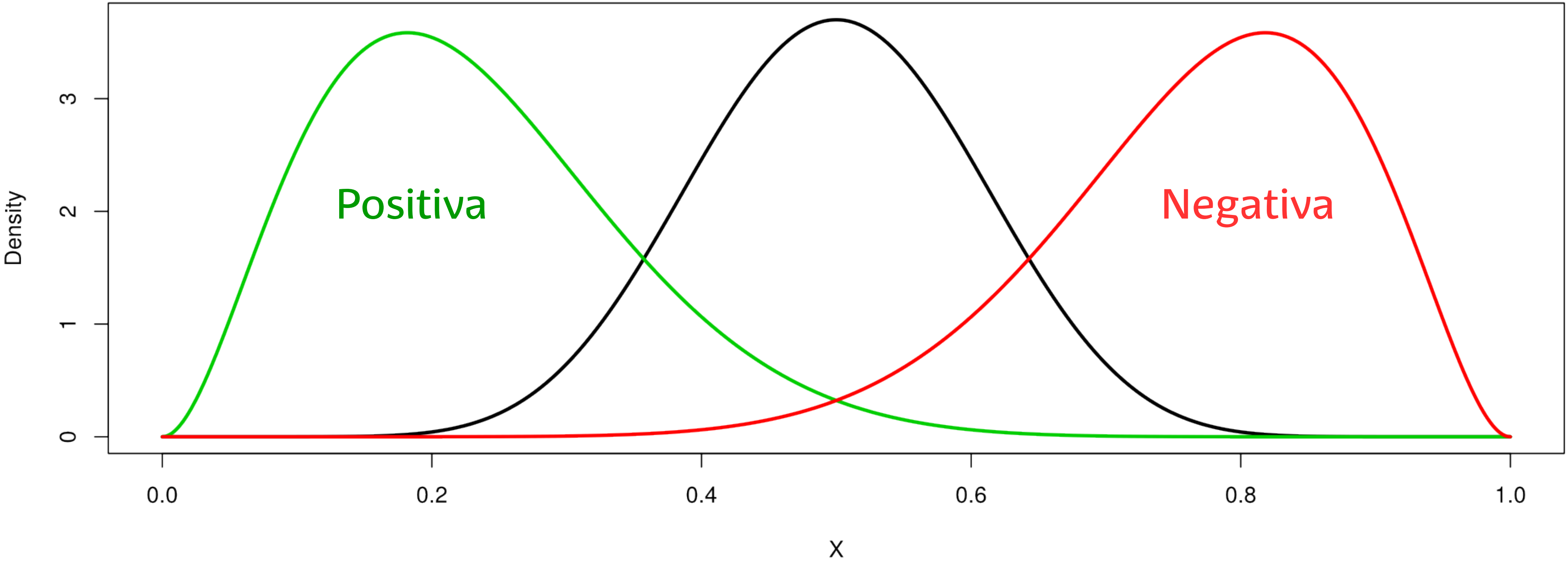
2º MOMENTO: LA VARIANZA

El segundo momento nos da idea de la *dispersión*



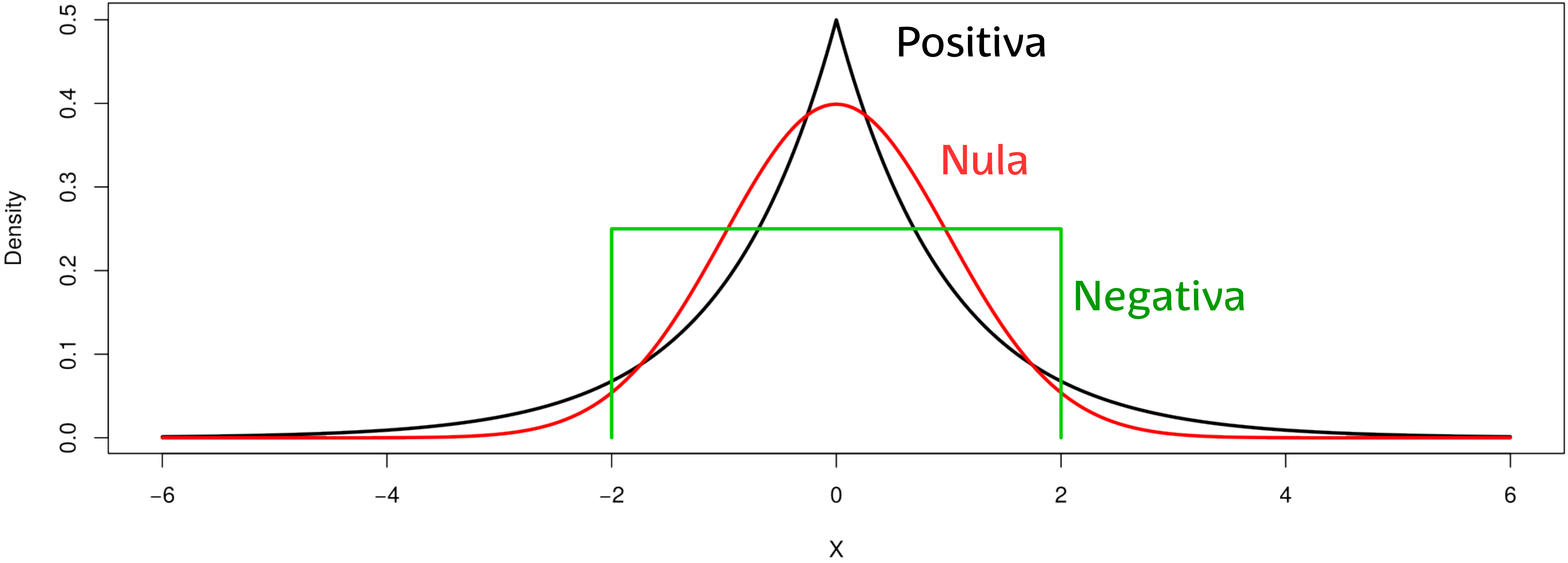
3º MOMENTO: LA ASIMETRÍA ESTADÍSTICA (SKEWNESS)

El tercer momento nos da idea de la *sesgadez*



4º MOMENTO: LA CURTOSIS (KURTOSIS)

El cuarto momento nos da idea de como de *picuda* es la distribución



Distribuciones discretas

Distribución de probabilidad asociada a variables categóricas (no existe un orden entre los valores que puede tomar).

DOMINIO DE DEFINICIÓN

$$X \in \{1, 2, \ldots, r\}$$

PARAMETROS

$$0 \leqslant p_1, \ldots, p_r \leqslant 1 \,;\, \sum_{i=1}^{n} p_i = 1$$

FUNCION DE PROBABILIDAD

$$P(X = i) = p_i$$

CASOS PARTICULARES

- Distribución discreta uniforme
- Distribución de Bernouilli (r=2)

En un muestreo de una distribución de Bernouilli, cuenta el número de resultados de un cierto tipo.

DOMINIO DE DEFINICIÓN $X \in \{0, 1, \ldots, n\}$

PARAMETROS $n > 0, 0 \leqslant p \leqslant 1$

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

MOMENTOS $\mu_1' = np \,;\, \mu_2 = np(1-p)$

En un conjunto de *N* elementos hay *M* “defectuosos”. Si tomamos una muestra de tamaño *n* con reemplazamiento, esta variable representa el número de elementos defectuosos.

DOMINIO DE DEFINICIÓN

$$max\{0, M-N+n\} < X < min\{n, M\}$$

PARAMETROS

$$N>0, 0 \leqslant M \leqslant N, 0 \leqslant n \leqslant N$$

FUNCION DE PROBABILIDAD

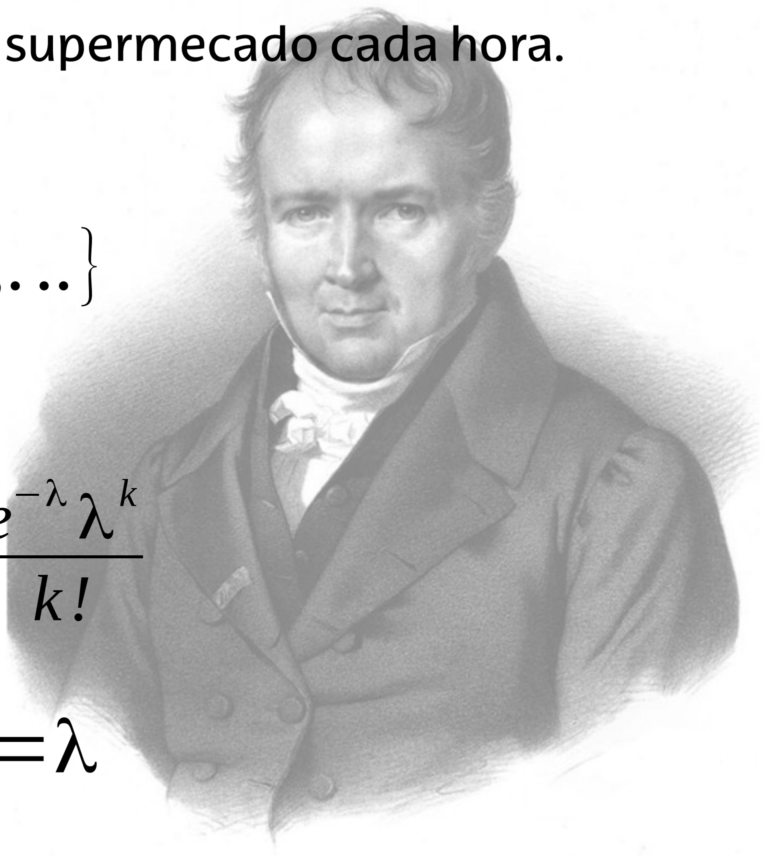
$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

MOMENTOS

$$\mu_1^{'}=\frac{nM}{N};\mu_2=\frac{nM}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Distribución de Poisson

Representa el número de eventos que ocurren en un intervalo de tiempo, bajo ciertas suposiciones. Por ejemplo, el número de clientes en el supermecado cada hora.



DOMINIO DE DEFINICIÓN $X \in \{0,1,2,\ldots\}$

PARAMETROS $\lambda > 0$

FUNCION DE PROBABILIDAD $P(X=k) = \dfrac{e^{-\lambda}\lambda^{k}}{k!}$

MOMENTOS $\mu_1^{'} = \lambda \,; \mu_2 = \lambda$

Distribuciones continuas

Distribución de Normal o Gausiana

Se usa habitualmente para modelar datos de poblaciones (la edad de los estudiantes de la Facultad, por ejemplo) o repeticiones de medidas experimentales (el contenido real de, por ejemplo, una lata de refresco).

DOMINIO DE DEFINICIÓN $X \in \mathbb{R}$

PARAMETROS $\mu \in \mathbb{R}, \sigma > 0$

FUNCION DE PROBABILIDAD $P(X=k) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

MOMENTOS $\mu_1' = \mu ; \mu_2 = \sigma^2$

Distribución Gamma

Supongamos un proceso de Poisson. La distribución gamma representa el tiempo de expera hasta que ocurre el evento a-esimo.

DOMINIO DE DEFINICIÓN	$X > 0$
PARAMETROS	$a > 0, b > 0$
FUNCION DE PROBABILIDAD	$f(x) = \frac{1}{\Gamma(a) b^a} e^{-\frac{x}{b}} x^{a-1}$
MOMENTOS	$\mu_1' = ab\,; \mu_2 = ab^2$

Representa el conciente entre dos distribuciones gamma. Se usa habitualmente para representar variables que varian entre 0 y 1, tales como ratios.

DOMINIO DE DEFINICIÓN

$$0 \leqslant X \leqslant 1$$

PARAMETROS

$$a>0, b>0$$

FUNCION DE PROBABILIDAD

$$f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$$

MOMENTOS

$$\mu_1'=\frac{a}{a+b};\mu_2=\frac{ab}{(a+b)^2(a+b+1)}$$

Distribución Chi Cuadrado

Representa la suma del cuadrado de n variables aleatorias que siguen una distribución normal de media 0 y desviación 1. Se usa para representar la varianza de una muestra y en ciertos test estadísticos.

DOMINIO DE DEFINICIÓN $X > 0$

PARAMETROS $n > 0$

FUNCION DE PROBABILIDAD $f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2} x^{n/2-1}$

MOMENTOS $\mu_1' = n ; \mu_2 = 2n$