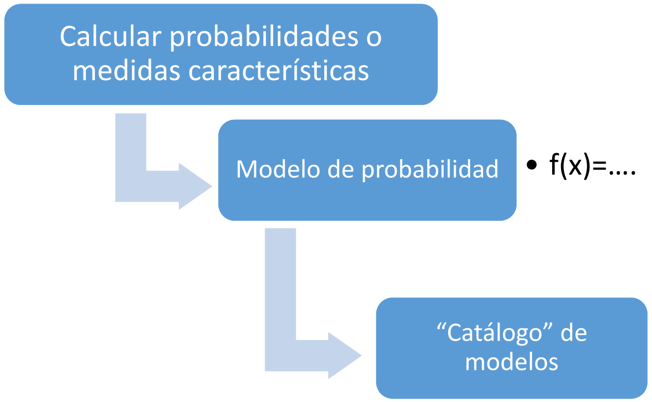
# Tema 5
# Modelos de probabilidad

Carlos Montes – uc3m



## 1. Concepto




## 1. Proceso de Bernoulli. Definición.

- Fenómeno aleatorio dicotómico
- La observación consiste en la clasificación del resultado obtenido en una de 2 categorías posibles:

  *Éxito*

  *Fracaso*

- La proporción de cada una de las categorías en la población es constante:

  *p: probabilidad de éxito*

  *q = 1-p: probabilidad de fracaso*

- Las observaciones son independientes entre sí.

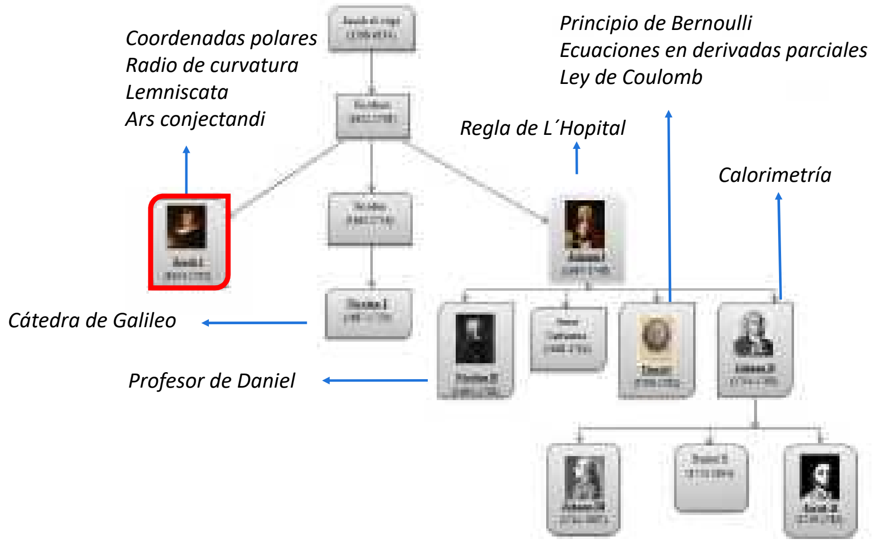
## 2.1. Variable de Bernoulli

$$x = \begin{cases} 0 \text{ si obtenemos un fracaso} \\ 1 \text{ si obtenemos un exito} \end{cases}$$

**Jakob Bernoulli**
**(1654-1715)**

Carlos Montes – uc3m

## 2.1. Variable de Bernoulli




## 1. Ley binomial (1,0) o de Bernoulli

### *Características*

*Función de probabilidad*

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

$$P(x=1) = p$$

$$P(x=0) = q$$

*Esperanza*

$$E(X) = 0 \cdot q + 1 \cdot p = p$$

## 1. Ley binomial (1,0) o de Bernoulli

*Varianza*

$$var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i) =$$

$$= (0-p)^2 \cdot q + \underbrace{(1-p)^2}_{q} \cdot p = pq\underbrace{(p+q)}_{1} = pq$$

La varianza será máxima si:

$$\frac{d(pq)}{dp} = \frac{d[p(1-p)]}{dp} = 1 - 2p = 0; \quad p = 0,5$$

## 2.3. Distribución binomial

Modeliza una serie de fenómenos dicotómicos independientes entre sí.

(nº de veces que ha aparecido el suceso S en n repeticiones independientes del experimento, con P(S)=p)



## 2.2. Distribución binomial

*¿Cuántas caras saldrán en n tiradas de la moneda?*

*Función de probabilidad, P(X=r)*

$$\underbrace{E, E, E, E\ldots E}_{r}, \ \underbrace{F, F\ldots}_{n-r}$$

$$p^r(1-p)^{n-r}$$

(si los sucesos son independientes)

## 2.2. Distribución binomial

Órdenes posibles:

$$C_n^r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

$$\boxed{P(X=r) = \binom{n}{r} p^r (1-p)^{n-r} \qquad r = 0,1,\ldots n}$$

## 2.3. Distribución binomial

*Esperanza*

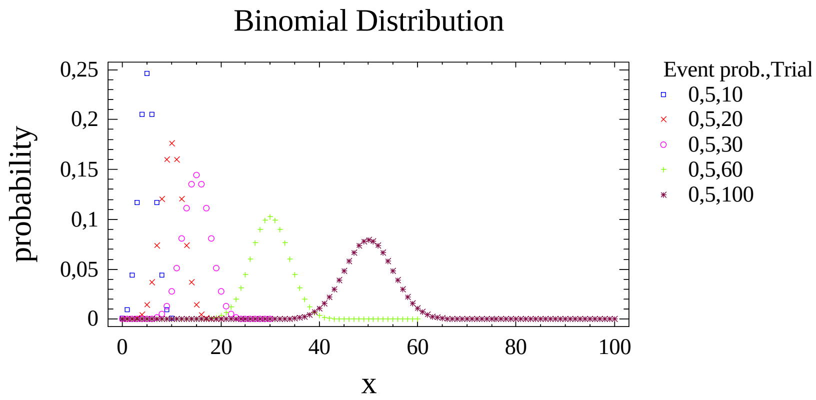
$$E(X) = E\left(\sum_{j=1}^{n} x_j\right) = \sum_{j=1}^{n} E(x_j) = np$$

*Varianza*

$$var(X) = var\left(\sum_{j=1}^{n} x_j\right) = \sum_{j=1}^{n} var(x_j) = npq$$

*Desviación típica:* $\sqrt{npq}$

## 2.3. Distribución binomial








*La probabilidad de encontrar una persona zurda es de 0,1. En una clase de 20 alumnos hay 3 pupitres para zurdos. Calcule la probabilidad de que no haya suficientes pupitres*

$X$: *número de personas zurdas en la clase* $\sim B(20, 0.10)$

$$P(X > 3) = 1 - \mathrm{P}(\mathrm{X} \leq 3) =$$
$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$P(X = 0) = \binom{20}{0} 0.10^0 \cdot 0.90^{20} = 0.12158$$

$$\binom{20}{0} = \frac{20!}{0!\,20!} = 1$$

$$P(X = 1) = \binom{20}{1} 0.10^1 \cdot 0.90^{19} = 0.27017$$

$$\binom{20}{1} = \frac{20!}{1!\,19!} = \frac{20 \cdot 19!}{1 \cdot 19!} = 20$$

$$P(X = 2) = \binom{20}{2} 0.10^2 \cdot 0.90^{18} = 0.28518$$

$$\binom{20}{2} = \frac{20!}{2!\,18!} = \frac{20 \cdot 19 \cdot 18!}{2 \cdot 18!} = 190$$

$$P(X = 3) = \binom{20}{3} 0.10^3 \cdot 0.90^{17} = 0.190120$$

$$\binom{20}{3} = \frac{20!}{3!\,17!} = \frac{20 \cdot 19 \cdot 18 \cdot 17!}{3 \cdot 2 \cdot 17!} = 1140$$

$$P(X > 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$P(X > 3) = 1 - [0.12158 \; + 0.27017 \; + 0.28518 \; + 0.190120\;] =$$

$$= 0.13295$$



**Cumulative Distribution**

Distribution: Binomial

Lower Tail Area (<)

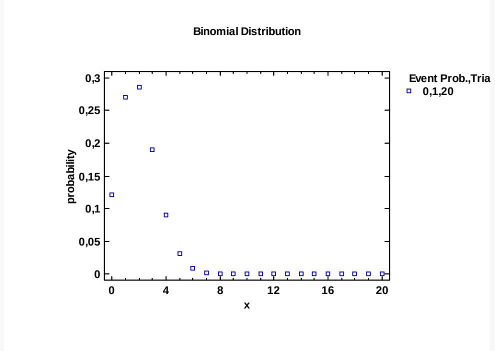
| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 3 | 0.676927 | | | | |

Probability Mass (=)

| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 3 | 0.19012 | | | | |

Upper Tail Area (>)

| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 3 | 0.132953 | | | | |




$$P(X > 3) = 0.132953$$

$$P(X > 3) = 1 - (0.676927 + 0.19012) = 0.132953$$

## 3.1. Proceso de Poisson. Definición

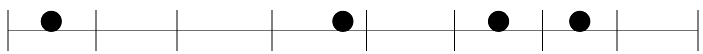
Aparición de sucesos puntuales
sobre un soporte continuo,
suponiendo que el proceso generador
de estos sucesos:

- es estable.
- produce sucesos independientes.

## 3.2. Distribución de Poisson

Modeliza la aparición de cierto número de sucesos
sobre un soporte continuo
en un intervalo de longitud fija.

*X: número de sucesos en un intervalo de longitud fija T*



***p pequeña*** ⇨ probabilidad despreciable de aparición de 2 o más sucesos en uno de los n segmentos.

Observamos si aparece o no el suceso estudiado
en cada segmento.

Es una distribución binomial en la cual:

$$n \to \infty$$

$$p \to 0$$

$$np \to \lambda$$

**Simeon Denis Poisson**
**(1781-1840)**

Carlos Montes – uc3m



***Características***

*Función de probabilidad*

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

$$P(X = r) \binom{n}{r} \left(\frac{\lambda}{n}\right)^r (1 - \frac{\lambda}{n})^{n-r} \quad r = 0, 1, 2 \ldots$$

Tomando límites: $$\boxed{P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad r = 0,1,2 \ldots}$$

(1838)



*Esperanza*

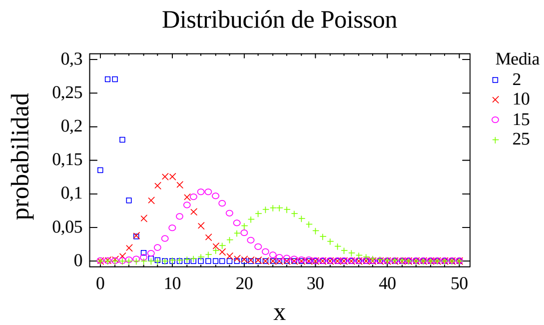
$$E(X) = \lambda$$

*Varianza*

$$var(X) = \lambda$$



Es una distribución asimétrica,
que tiende a la simetría al aumentar λ



## 3.2. Distribución de Poisson

La suma de varias variables de Poisson independientes, también es una variable de Poisson.

Sea $X_i \sim P(\lambda_i)$, i = 1, ..., k, un conjunto de variables de Poisson independientes.

$$Y = \sum_{i=1}^{k} X_i \rightarrow \mathrm{P}(\lambda^*)$$

$$\lambda^* = \sum_{i=1}^{k} \lambda_i$$



*Un servidor de una pequeña red recibe una media de 7 accesos por minuto. Suponiendo que los accesos a dicho servidor suceden de forma independiente y con ritmo medio constante, se quiere calcular la probabilidad de que reciba más de 10 accesos en un minuto, porque el servidor tendría entonces un rendimiento deficiente.*

$X$: *número de accesos en un minuto*

$X \sim P(\lambda = 7)$

$$P(X > 10) = 1 - P(X \le 10) = 1 - \sum_{r=0}^{10} \frac{7^r}{r!} e^{-7} = 0.09852$$

**Cumulative Distribution**
Distribution: Poisson

Lower Tail Area (<)

| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 10 | 0.830496 | | | | |

Probability Mass (=)

| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 10 | 0.0709833 | | | | |

Upper Tail Area (>)

| Variable | Dist. 1 | Dist. 2 | Dist. 3 | Dist. 4 | Dist. 5 |
|---|---|---|---|---|---|
| 10 | 0.0985208 | | | | |

$$P(X > 10) = 1 - P(X \le 10)$$
$$= 1 - (0.830496 + 0.0709833)$$
$$= 0.09852$$

## 3. Distribución exponencial

Modela el tiempo entre la ocurrencia de dos sucesos consecutivos, siendo estos independientes y estables.

***Características***

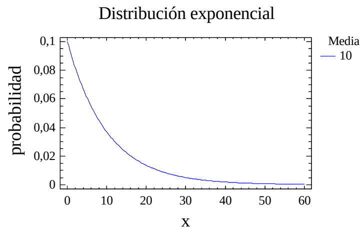
*Función de distribución*

$$F(t) = P(T \le t) = 1 - e^{-\lambda t}$$

Con λ número medio de sucesos **por unidad de tiempo**.

3.3. Distribución exponencial

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t}$$




*Esperanza*

$$E(T) = \frac{1}{\lambda}$$

*Varianza*

$$var(T) = \frac{1}{\lambda^2}$$



*Considerando la red anterior, se pide:*

*a) Calcular el tiempo medio que transcurre entre dos accesos consecutivos.*

$$\lambda = 7\ accesos/minuto$$

$$E(T) = \frac{1}{\lambda} = \frac{1}{7} = 0.143\ minutos/acceso$$

*b) Probabilidad de que entre dos accesos consecutivos transcurran más de 15 segundos.*

$$P(T > 0.25\ minutos) = 1 - P(T < 0.25) =$$

$$1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t} = e^{-7 \cdot 0.25} = 0.17$$

ej. 35

*La duración de ciertos componentes electrónicos sigue una distribución exponencial de media 100 días.*

*a) ¿Cuál es la probabilidad de que uno de los componentes anteriores dure más de 50 días?*

$$X: duración\ de\ un\ componente \quad \rightarrow \exp(\lambda = {}^{1}/_{100})$$

$$P(X > 50) = 1 - P(X \leq 50) = 1 - \left(1 - e^{-\lambda t}\right)$$

$$= 1 - \left(1 - e^{-{}^{50}/_{100}}\right) = e^{-\frac{1}{2}} = 0.607$$

*b) Un equipo electrónico está formado por 5 componentes de los anteriores, que trabajan de manera independiente, y funciona mientras funcionen correctamente al menos dos de ellos, ¿cuál es la probabilidad de que dicho equipo dure más de 50 días?*

$Y$: *Número de componentes que funcionan más de* 50 *días*

$Y \rightarrow B(5, 0.607)$

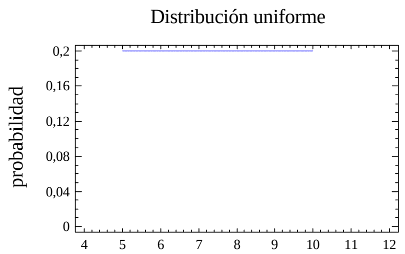
$$P(\text{equipo funcione}) = P(Y \geq 2) = 1 - P(Y=0) - P(Y=1) = 1 - \binom{5}{0} 0.607^0 \cdot 0.393^5 - \binom{5}{1} 0.607^1 \cdot 0.393^4 = 0.9182$$





***Características***

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

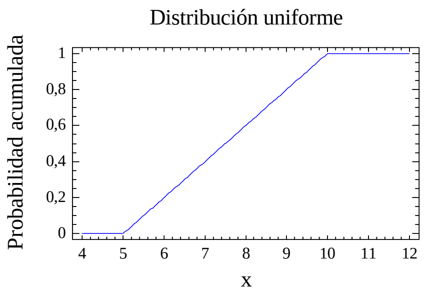





*Función de distribución*

$$F(x) = P(X \leq x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$

$$a \leq x \leq b$$






*Esperanza*

$$E(X) = \frac{a+b}{2}$$

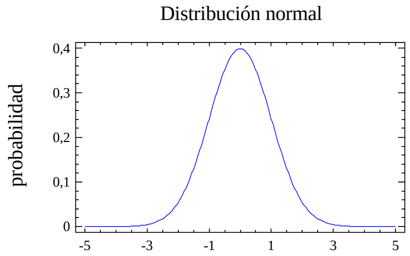
*Varianza*

$$var(X) = \frac{(b-a)^2}{12}$$

## 4. Distribución normal

Es la distribución con función de densidad:

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

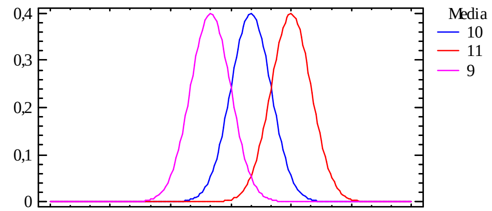


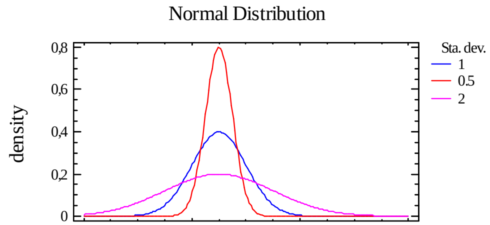



## 4. Distribución normal

La distribución normal depende de dos parámetros:

- ✓ Media μ
- ✓ Desviación típica σ



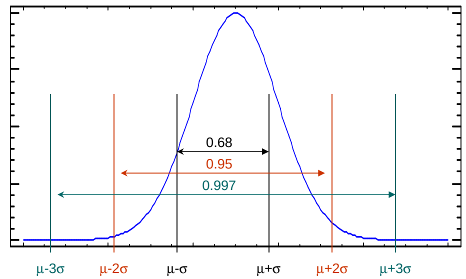




## 4. Distribución normal

- Es uno de los modelos más frecuentes para describir variables reales continuas.
- Simétrica, centrada en la media μ que es su mediana y su moda.
- Forma de campana.
- Coeficiente de apuntamiento igual a 3.

## 4. Distribución normal

*Regla empírica*



• Se ajusta a lo observado en muchos procesos de medición, si no influyen los errores sistemáticos.

• La normal de media 0 y desviación típica 1 se denomina:
- ❖ normal tipificada (Z)
- ❖ normal estándar
- ❖ normal (0,1)

y su función de distribución está tabulada.





$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

$$Z \rightarrow N(0,1)$$

*La longitud L en milímetros, de las piezas fabricadas en un proceso es una variable aleatoria que se distribuye según una N(32, 0.3²), considerándose aceptables aquellas cuya medida se encuentra dentro del intervalo (31.1, 32.6). Calcule la probabilidad de que una pieza elegida al azar sea aceptable.*

$$L \rightarrow N(32, 0.3^{2)}$$

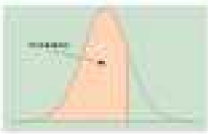
$$P(aceptable) = P(31.1 < L < 32.6)$$

$$P\left(\frac{31.1-32}{0.3} < Z < \frac{32.6-32}{0.3}\right) = P(-3 < Z < 2)$$

$$F(2) - F(-3) = F(2) - [1 - F(3)] = F(2) - 1 + F(3) =$$

$$= 0.9772 - 1 + 0.9987 = 0.9759$$

## 4. Distribución normal



$$F(2) - 1 + F(3)$$

0.9772 0.9987

0.9759

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

## 6. Distribución lognormal

Se llama *lognormal* a la variable aleatoria cuyo logaritmo neperiano es normal:

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

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2} \qquad \operatorname{var}(X) = \left(e^{\sigma^2} - 1\right) \cdot e^{2\mu + \sigma^2}$$

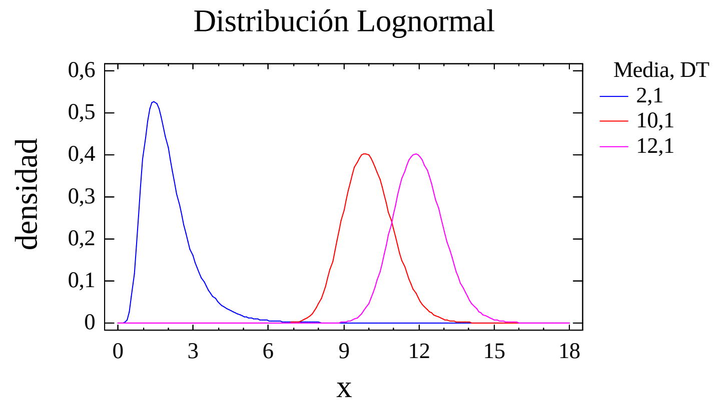
## 6. Distribución lognormal

Si la variable y puede considerarse como un producto de variables aleatorias:

$$y = x_1 \cdot x_2 \cdot x_3 \ldots x_n$$

su logaritmo neperiano
seguirá una distribución normal
(distribución lognormal o de Mac Alister).

## 6. Distribución lognormal



## 7.1. Teorema central del Límite. Definición

La suma de un conjunto
de variables aleatorias se aproxima,
al aumentar el número de variables,
a una variable aleatoria **normal**
independientemente de cual sea
la distribución de esas variables.



## 7.1. Teorema central del Límite. Definición

Sean $x_1$ , $x_2$ ,... , $x_n$ v.a. independientes
con media $\mu_i$, desviación típica $\sigma_i$
y distribución

CUALQUIERA

$$Y = x_1 + x_2 + \ldots + x_n$$

$$\frac{Y - \sum \mu_i}{\sqrt{\sum \sigma_i^2}} \rightarrow N(0,1)$$

Al crecer n:

$$Y \rightarrow N\left(\sum \mu_i, \sqrt{\sum \sigma_i^2}\right)$$

## 7.1. Teorema central del Límite. Definición

$$Y \rightarrow N\left(\sum \mu_i, \sqrt{\sum \sigma_i^2}\right)$$

La desviación típica
es la raíz cuadrada
de la suma de varianzas,
NO la suma
de las desviaciones típicas.

## 7.2. Teorema central del Límite. Aplicaciones

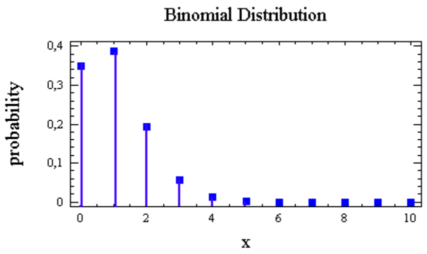
Binomial → Normal

$$E(x_i) = p \quad Var(x_i) = pq$$

Aplicando el Teorema Central del Límite:

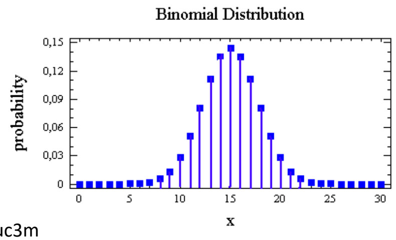
$$Y \rightarrow N\left(np, \sqrt{npq}\right)$$

$n > 30$
$npq > 5$



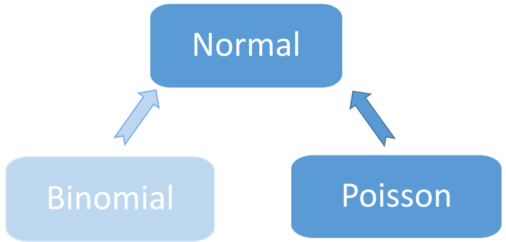

*npq=0,9*

*npq=7,5*








Sea Y(0,T) una variable de Poisson que cuenta el nº de sucesos en el intervalo (0,T)

$$Y(0,T) = x_1(0,t_1) + x_2(t_1,t_2) + \ldots + x_n(t_{n-1},T)$$

Para λ>5
podemos aproximar esta variable mediante una distribución Normal

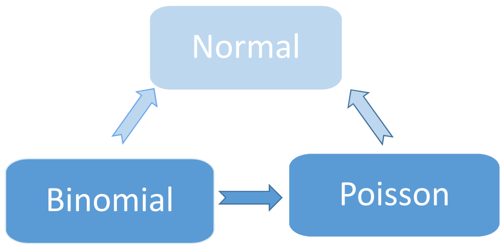


Para λ>5
podemos aproximar esta variable mediante una distribución Normal

$$x_p \to P(\lambda) \quad \lambda > 5$$

$$x_n \to N\left(\lambda, \sqrt{\lambda}\right)$$





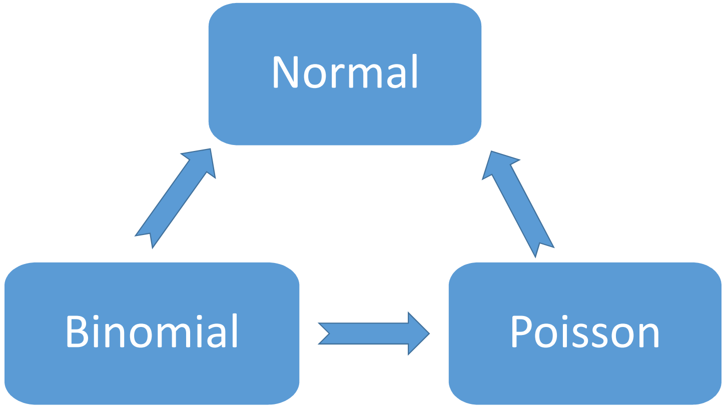

La aproximación es válida cuando:

$$np > 1$$

$$p < 0{,}1$$

$$\lambda = np$$

## 7.2. Teorema central del Límite. Aplicaciones






## 8. Modelo de regresión lineal simple

La recta de regresión de y sobre x es de la forma:

$$y_i = a + bx_i + e_i$$

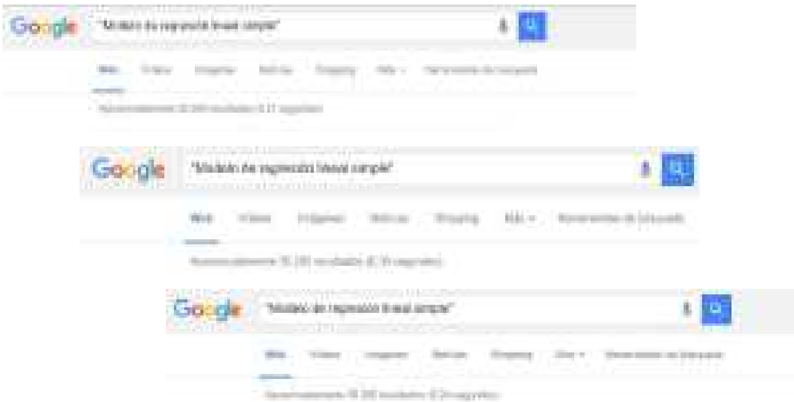
$e_i$ → Término de error

$y_i$: Valor observado de la variable *y* para el individuo i-ésimo

$x_i$: Valor observado de la variable x para el individuo i-ésimo

Si queremos definir un tipo de relación válido para toda la población, encontramos numerosos factores que no controlamos.

## 8. Modelo de regresión lineal simple



## 8. Modelo de regresión lineal simple

Si fijamos el valor de la variable X en $X = x_i$ repitiendo el experimento, observaremos valores diferentes debidos al efecto de las variables recogidas en el término *e*.

$$y_i = \underbrace{a + bx_i}_{\textit{fijo}} + \underbrace{e_i}_{\textit{aleatorio (ruido)}}$$

## 8. Modelo de regresión lineal simple

Si asumimos que todas las variables que influyen sobre Y lo hacen de forma lineal (aditiva):

$$Y = a + bX + \underbrace{\left(c_1 Z_1 + c_2 Z_2 + c_3 Z_3 + \ldots\right)}_{\rightarrow\ e}$$

Por el Teorema del Límite Central, *e* seguirá una distribución normal.



## 8. Modelo de regresión lineal simple

$$E(e) = 0$$

$$\mathrm{var}(e) = \sigma^2 \quad \text{(constante)}$$

$$e \rightarrow N(0, \sigma^2)$$

$$E\left(Y \setminus X = x_i\right) = E\left(a + bx_i + e\right) = a + bx_i + E(e) = a + bx_i$$

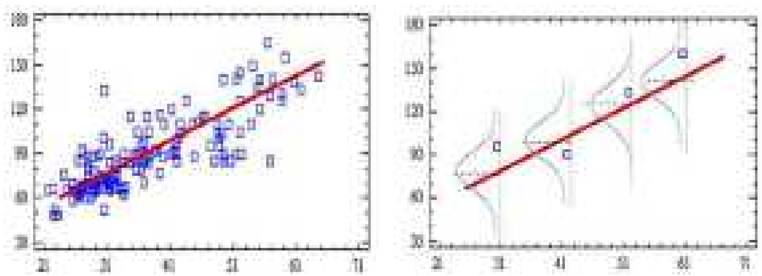
$$\mathrm{var}(Y \setminus X = x_i) = \mathrm{var}(a + bx_i + e) = \mathrm{var}(e)$$

$$Y \rightarrow N(a + bx_i, \sigma^2)$$

## 8. Modelo de regresión lineal simple

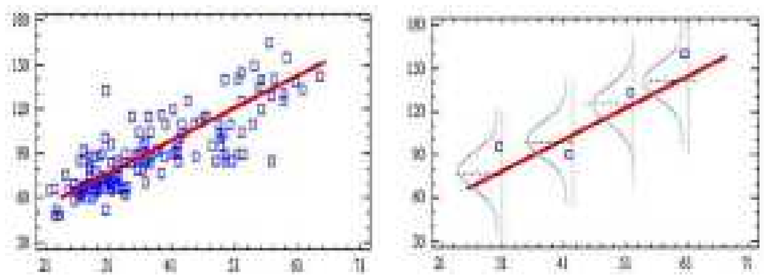
Cada punto $y_i$ que observamos se interpreta como un valor al azar de la normal.

$$y_i \rightarrow N(a + bx_i, \sigma^2)$$



## 8. Modelo de regresión lineal simple

Suponemos que el "ruido" es homogéneo a lo largo de la recta (varianza constante u *homocedasticidad*).



*Sea el modelo de regresión simple Y=50 +2X + e, siendo e =N(0, $5^2$). Calcule la probabilidad de que Y sea mayor que 160 para los casos siguientes:*

*a) X=60*

*b) X=50*



$P(Y > 160)$ *Y=50 +2X + e,*

*a) Necesitamos el modelo que sigue Y*

$Y \to N(a + bx_i; \sigma^2)$

$Y = 50 + 2X + e = 50 + 2 \cdot 60 + e = 170 + e$

$Y \to N(170; 5^2)$

$$P(Y > 160) = P\left(Z > \frac{160 - 170}{5}\right) =$$

$$= P(Z > -2) = P(Z < 2) = 0.9772$$

$P(Y > 160)$ *Y=50 +2X + e,*

*b)* $Y \to N(a + bx_i; \sigma^2)$

$Y = 50 + 2X + e = 50 + 2 \cdot 50 + e = 150 + e$

$Y \to N(150; 5^2)$

$$P(Y > 160) = P\left(Z > \frac{160 - 150}{5}\right) =$$

$$= P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$