Apuntes de 2°GITI

María Ballesteros

APUNTES ESTADISTICA

# Estadística Descriptiva

## 1 TIPOS DE DATOS

En estadística clasificaremos los datos en:

Cuantitativos
- Continuos: consumo del coche, potencia, aceleración, peso –
- Discretos: nº cilindros – nº

Cualitativos
- Ordinales: categoría
- No ordinales: país, gasolina / gasoil

## 2 MEDIDAS DE CENTRO

Para una distribución: $x_1, x_2, \ldots x_n$ tenemos:

Media aritmética $\longrightarrow$ $\boxed{\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$

Media geométrica (si $x_i > 0 \ \forall i$) $\longrightarrow$ $\overline{x}_G = \sqrt[n]{x_1 x_2 \ldots x_n}$

Media armónica (si $x_i > 0 \ \forall i$) $\longrightarrow$ $\overline{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

NOTA!!

$$\boxed{\overline{x}_H \leq x_G \leq \overline{x}}$$

## 3 MEDIDAS DE DISPERSIÓN

Desviación típica $\quad S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}}$

La desviación típica informa de la media de distancias que tienen los datos respecto a su media aritmética, es decir, representa la variación esperada con respecto a la media.

Varianza: $s^2$ (el cuadrado de la desviación típica)

# 4 MEDIANA Y CUARTILES

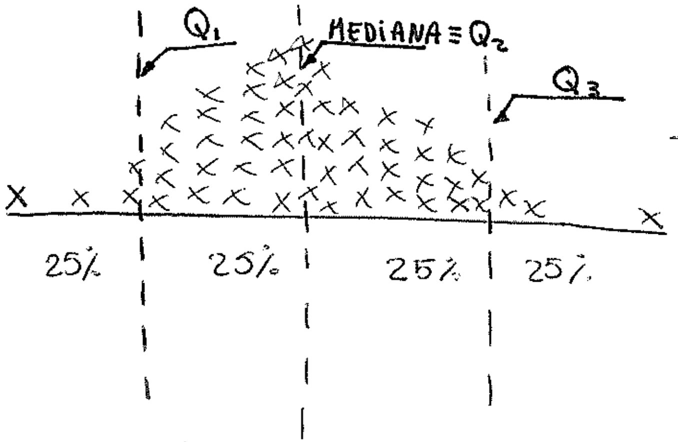
Sea la distribución $x_1, x_2$ ___ $x_n$ de datos ordenados tal que

$x_{(1)} \le x_{(2)} \le$ ___ $\le x_{(n)}$

$$\text{Mediana} \begin{cases} X_{(p)} & \text{con } \boxed{p = \frac{n+1}{2}} \text{ si } \underline{n \text{ es impar}} \\ \frac{X_{(p)} + X_{(p+1)}}{2} & \text{con } \boxed{p = \frac{n}{2}} \text{ si } \underline{n \text{ es par}} \end{cases}$$

Parte entera!!

$$\text{Cuartiles} \begin{cases} Q_1 = X_{(r)} & \text{con } \boxed{r = \left[\frac{p+1}{2}\right]} \\ Q_3 = X_{(s)} & \text{con } \boxed{s = n - r + 1} \end{cases}$$



$\Longrightarrow$

$fr\,(X_i \le Q_1) = 0{,}25$

$fr\,(X_i \le Q_3) = 0{,}75$

$fr\,(X_i \le Med) = 0{,}5$

Ejemplo

$x_1, x_2,$ ___ $x_n \equiv 7, \overset{Q_1}{⑨}, 15, \underline{18}, 23, \overset{Q_2}{㊺}, 67$

$n = 7 \longrightarrow$ Mediana $= X_{(p)}$ con $p = \frac{n+1}{2} = 4 \longrightarrow \boxed{\text{Mediana} = 18}$

$Q_1 = X_{(r)}$ con $r = \left[\frac{p+1}{2}\right] = \left[\frac{5}{2}\right] = 2$ $\boxed{Q_1 = X_2 = 9}$

$Q_3 = X_{(s)}$ con $s = n - r + 1 = 7 - 2 + 1 = 6$ $\boxed{Q_3 = X_6 = 45}$

NOTA!! $Q_1$ está a la misma distancia de la mediana que $Q_2$

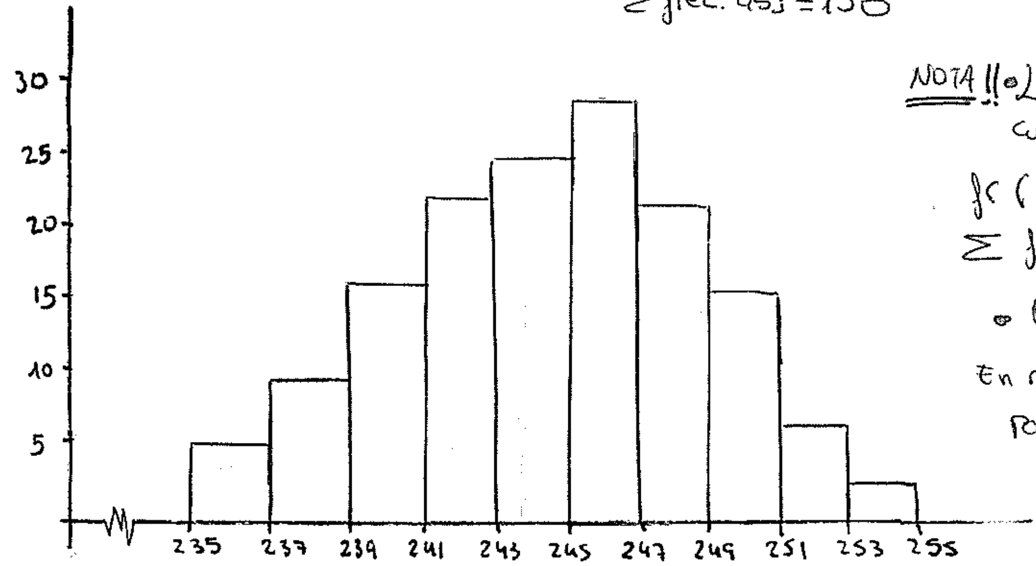
# 5 DIAGRAMAS

## • El Histograma

⊕ $\text{frec. relativa}_i = \dfrac{\text{frec. abs}_i}{\Sigma \text{ frec. abs}}$

| Nº clase | Clase | Frecuencia absoluta | Frec. relativa |
|---|---|---|---|
| 1 | 235 - 237 | 5 | $5/150 = 0{,}0\overline{3}$ |
| 2 | 237 - 239 | 9 | $9/150 = 0{,}06$ |
| 3 | 239 - 241 | 16 | $16/150 = 0{,}10\overline{6}$ |
| 4 | 241 - 243 | 23 | $23/150 = 0{,}15\overline{3}$ |
| 5 | 243 - 245 | 24 | $24/150 = 0{,}16$ |
| 6 | 245 - 247 | 29 | $29/150 = 0{,}19\overline{3}$ |
| 7 | 247 - 249 | 22 | $22/150 = 0{,}14\overline{6}$ |
| 8 | 249 - 251 | 15 | $15/150 = 0{,}1$ |
| 9 | 251 - 253 | 5 | $5/150 = 0{,}0\overline{3}$ |
| 10 | 253 - 255 | 2 | $2/150 = 0{,}01\overline{3}$ |

$\Sigma$ frec. abs = 150



NOTA !! • La frec. relativa cumple

$fr\,(a < x_i \leq b) \geq 0$

$\Sigma\, fr(a_j < x_i < b_j) = 1$

• Rango = max - min

En nuestro caso:

Rango = 255 - 235 = $\underline{\underline{20}}$

## • Diagrama de puntos

Sean dos grupos ejemplo

Cuando tenemos pocos datos elaboramos diagramas de puntos en lugar de histogramas

| Ⓐ | Ⓑ |
|---|---|
| 179 | 174 |
| 169 | 168 |
| 170 | 174 |
| 171 | 177 |
| 160 | 173 |
| 182 | 175 |
| 177 | 174 |
| 171 | 171 |
| 182 | 173 |
| 166 | 170 |

Diagrama de Ⓐ

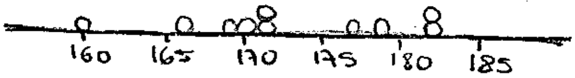
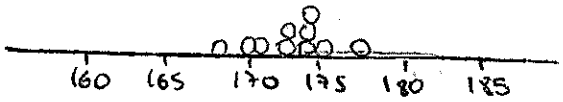


Diagrama de Ⓑ



# Diagrama de caja

Densidad de la tierra 5,5 5,47 5,55 5,75 5,29 5,27
5,57 4,88 5,34 5,29 5,34 5,85
5,42 5,62 5,3 5,1 5,26 5,65
5,61 5,63 5,36 5,86 5,44 5,39
5,53 4,07 5,79 5,58 5,46

Necesitamos { Mediana $= Q_2$ Rango intercuartílico ≡ $RI = Q_3 - Q_1$
$Q_1$ Lím. inf ≡ $LI = Q_1 - 1{,}5\,RI$
$Q_3$ Lím sup ≡ $LS = Q_3 + 1{,}5\,RI$

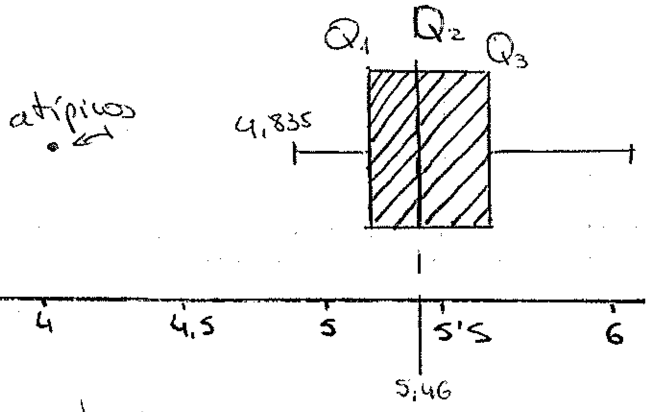
Ordenamos datos: 4,07 4,88 5,1 5,26 5,27 5,29 5,29
5,3 5,34 5,34 5,36 5,39 5,42 5,44
$\boxed{5{,}46}$ 5,47 5,5 5,53 5,55 5,57 5,58
5,61 5,62 5,63 5,65 5,75 5,79 5,85
5,86

$n = 29$ impar ⟶ Mediana $= X_{(p)}$ con $p = \frac{n+1}{2} = \frac{30}{2} = 15$

Med $= 5{,}46$

$Q_1 = X_{(r)}$ con $r = \left[\frac{p+1}{2}\right] = \frac{16}{2} = 8 \rightarrow Q_1 = 5{,}3$

$Q_3 = X_{(s)}$ con $s = n - r + 1 = 29 - 8 + 1 = 22 \rightarrow Q_3 = 5{,}61$



$RI = Q_3 - Q_1 = 5{,}61 - 5{,}3$

$RI = 0{,}31$

$LI = Q_1 - 1{,}5\,RI$

$LI = 4{,}835$

$LS = Q_3 + 1{,}5\,RI$

$LS = 6{,}075$

# Diagrama de tallos y hojas

Usaremos los datos anteriores tomando $\underbrace{5,}_{\text{TALLO}}\underbrace{47}_{\text{HOJA}}$

Bajos 4,07

| | | |
|---|---|---|
| 48 | 8 | |
| 49 | | |
| 50 | | |
| 51 | 0 | |
| 52 | 6 7 9 9 | |
| 53 | $\boxed{0}$ 4 4 6 9 | $Q_1$ |
| 54 | 2 4 $\boxed{6}$ 7 | $Q_2$ o Mediana |
| 55 | 0 3 5 7 8 | |
| 56 | $\boxed{1}$ 2 3 5 | $Q_3$ |
| 57 | 5 9 | |
| 58 | 5 6 | |

# 6 DESIGUALDAD DE CHEBYCHEV

$$\boxed{\text{frel}\left(|X_i-\overline{X}|\leq Ks\right) > 1-\frac{1}{K^2}}$$

**Demo:** Sean dos subconjuntos: $\begin{cases} A_1=\{x_i / |x_i-\overline{x}| > K\cdot s\} \\ A_2=\{x_i / |x_i-\overline{x}| \leq K\cdot s\} \end{cases}$

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\frac{\sum_{A_1}(x_i-\overline{x})^2}{n}+\frac{\sum_{A_2}(x_i-\overline{x})^2}{n}\geq\frac{\sum_{A_1}(x_i-\overline{x})^2}{n}>$$

$$>\frac{\sum_{A_1}K^2s^2}{n}=\frac{n_{A_1}K^2s^2}{n}=K^2s^2\,\text{frel}\left(|x_i-\overline{x}|>Ks\right)\Longrightarrow$$

$$S^2>K^2s^2\,\text{frel}\left(|x_i-\overline{x}|>Ks\right)$$

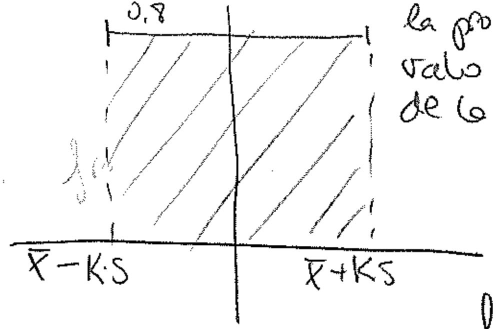
$$\frac{1}{K^2}>\text{frel}\left(|x_i-\overline{x}|>Ks\right);\quad \frac{1}{K^2}>1-\text{frel}\left(|x_i-\overline{x}|\leq Ks\right)$$

Por tanto: $\text{frel}\left(|x_i-\overline{x}|\leq Ks\right)>1-\frac{1}{K^2}$

⇨ **Ejercicio práctico:** Un asesor bursátil le dice a su cliente que el precio de una acción determinada será 55 € con una desviación típica de 1,5 €. El asesor no conoce la distribución de la probabilidad del precio de la acción. Proporcione un intervalo centrado en el valor esperado que contenga el precio de la acción con una probabilidad igual o superior a 0,8

$|s=1,5|$

$55=\overline{x}$



El intervalo pedido será: $[\overline{x}-KS,\ \overline{x}+KS]$

Por desigualdad de Chebychev

$$\text{frel}\left(|x_i-\overline{x}|\leq Ks\right)\geq 1-\frac{1}{K^2}=0,8$$

$$1-\frac{1}{K^2}=0,8;\quad \frac{1}{K^2}=0,2;\quad K=2,236$$

NOTA!! Relacionar siempre probabilidad con frec. rel.

Intervalo $\begin{cases}\text{Lim. inf} \longrightarrow 55-2,236\cdot 1,5=51,65\\ \text{Lim. sup} \longrightarrow 55+2,236\cdot 1,5=58,35\end{cases} \longrightarrow$ $\boxed{\text{Intervalo pedido } [51,65,\ 58.35]}$

# 7 MOMENTOS DE LOS DATOS. Medidas características de forma → MODELO IDEAL

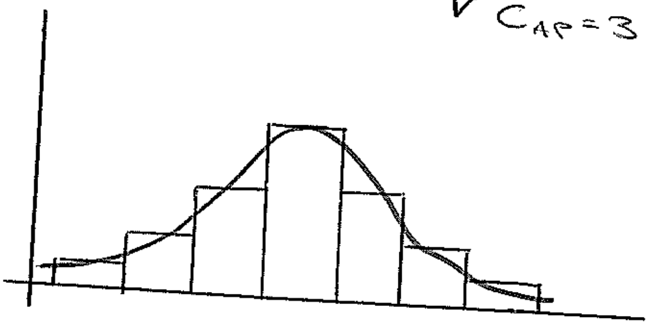
Momento respecto al origen $\boxed{a_k = \frac{\sum_{i=1}^{n} x_i^k}{n}}$

Momento respecto a la media $\boxed{m_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^k}{n}}$

Definimos dos medidas características de forma

- Coeficiente de asimetría $\boxed{C_{AS} = \frac{m_3}{S^3}}$
- Coeficiente de curtosis o apuntamiento $\boxed{C_{AP} = \frac{m_4}{S^4}}$

EL MODELO IDEAL ⇒ $C_{AS} = 0$, $C_{AP} = 3$



Los momentos de orden impar respecto a la media valen cero en el modelo ideal

$C_{AS} > 0$ ⟶ 

$C_{AS} < 0$ 

⇨ Ejercicio aplicación → cuartiles, diagramas cox-box

En la siguiente tabla se muestran los resultados de la prueba de longitud de las Olimpiadas de 1988. Construye el diagrama de cajas:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7,72 | 7,45 | 7,45 | 7,45 | 7,44 | 7,43 | 7,43 | 7,38 | 7,37 |
| 7,36 | 7,34 | 7,29 | 7,28 | 7,23 | 7,19 | 7,12 | 7,09 | 7,08 |
| 7,08 | 7,07 | 7,05 | 7,04 | 7,02 | 7,01 | 7,00 | 6,98 | 6,97 |
| 6,95 | 6,90 | 6,83 | 6,75 | 6,43 | 6,22 | 5,83 | 6 | |

$n=34 \longrightarrow \text{Mediana} = \frac{X_p + X_{p+1}}{2}$ con $p = \frac{n}{2} = 17 \rightarrow \text{Med} = \frac{7{,}08 + 7{,}09}{2} = 7{,}085$

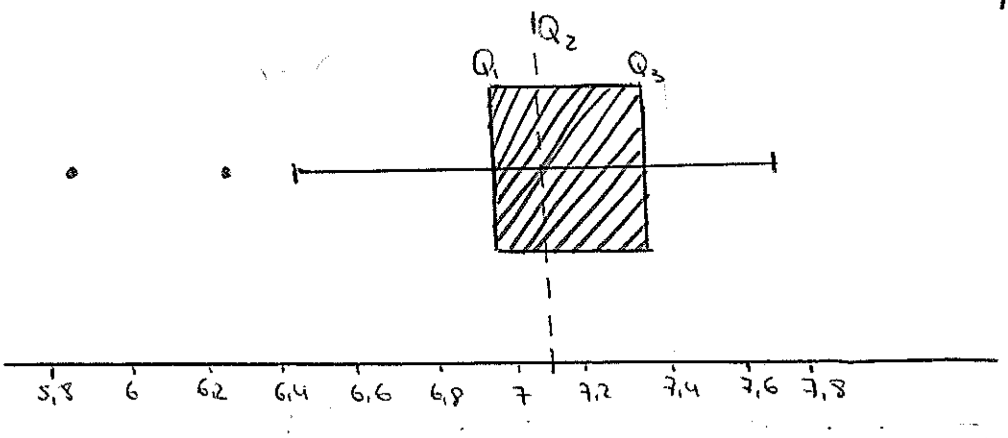
$Q_1 = X_r$ con $r = \left[\frac{p+1}{2}\right] = \left[\frac{18}{2}\right] = 9 \longrightarrow Q_1 = 6{,}98$

$Q_2 = X_s$ con $S = n - r + 1 = 26 \longrightarrow Q_3 = 7{,}37$

$RI = Q_3 - Q_1 = 0{,}39$

$LI = Q_1 - 1{,}5 RI = 6{,}395 \rightarrow 6{,}43$

$LS = Q_3 + 1{,}5 RI = 7{,}955 \rightarrow 7{,}72$



# 8 TRANSFORMACIONES

- **Lineales** → $y_i = a + bx_i$

No cambian la forma (asimetría y curtosis) de la distribución pero

$$\bar{y} = a + b\bar{x}$$

$$S_y = |b|\, S_x$$

Demo:

$$\bar{y} = \frac{\hat{\sum} y_i}{n} = \frac{\sum(a+bx_i)}{n} = \frac{na}{n} + b\overset{\bar{x}}{\boxed{\frac{\sum x_i}{n}}} =$$

$$= a + b\bar{x}$$

Demo

$$S_y^2 = \frac{\sum(y_i-\bar{y})^2}{n} = \frac{\sum(\cancel{a}+bx_i-\cancel{a}-b\bar{x})^2}{n} = b^2 \overset{S_x^2}{\boxed{\frac{\sum(x_i-\bar{x})^2}{n}}} = b^2 S_x^2 \longrightarrow S_y = |b|\, S_x$$

- **No lineales** → $y_i = h(x_i)$

Cambian la forma ($C_{As}$ y $C_{Ap}$). La media y varianza de las nuevas distribuciones se calcula por Taylor.

Para $C_{As} > 0$

usaremos transformaciones como $\begin{cases} y_i = \log x_i \\ y_i = \dfrac{1}{x_i} \\ y_i = \sqrt{x_i} \end{cases}$

Para $C_{As} < 0$

usaremos transformaciones como $\left\{ y_i = x_i^2 \right.$

**Aproximaciones:**

$$h(x) = h(a) + h'(a)(x-a) + \frac{h''(a)}{2}(x-a)^2 + \dots$$

$$\Rightarrow h(x) = h(\bar{x}) + h'(\bar{x})(x-\bar{x}) + \frac{h''(\bar{x})}{2}(x-\bar{x})^2$$

Como $y_i = h(x_i) \longrightarrow \boxed{\bar{y} \simeq h(\bar{x}) + \frac{h''(\bar{x})}{2} S_x^2}$; $\boxed{S_y^2 = |h'(\bar{x})|^2 S_x^2}$

Ej.: $y_i = \log(x_i) \longrightarrow \bar{y} = \frac{\sum \log x_i}{n} \simeq \log(\bar{x}) - \frac{1}{2}\frac{1}{\bar{x}^2} S_x^2$

$$S_y^2 \simeq \left|\frac{1}{100}\right|^2 S_x^2$$

**Estandarización**

$x_1, x_2, x_3 \text{ ——— } x_n \left\{ \begin{array}{l} \bar{x} \\ S_x \end{array} \right.$; la transformación $z_i = \dfrac{x_i - \bar{x}}{S_x} \Rightarrow \left\{ \begin{array}{l} \bar{z} = 0 \\ S_z = 1 \end{array} \right.$

**Box-Cox**

Transformación de la forma $\begin{cases} y_i = \dfrac{x_i^p - 1}{p} & \forall p \neq 0 \\ y_i = \log x_i & p = 0 \end{cases}$

# 9 ESTADÍSTICA DESCRIPTIVA MULTIVARIANTE

- Vector de medias $\bar{X} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_n \end{pmatrix}$

→ Covarianza (tiene unidades)

$$\boxed{S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}}$$

- Matriz de varianzas

Esta matriz es cuadrada, simétrica y semidefinida positiva.

$$\begin{pmatrix} S_{x1}^2 & S_{x1x2} & \cdots & S_{x1xk} \\ S_{x2x1} & S_{x2}^2 & \cdots & S_{x2xk} \\ \vdots & & \ddots & \vdots \\ S_{xkx1} & S_{xkx2} & \cdots & S_{xk}^2 \end{pmatrix} = S^2$$

A su determinante se le denomina → varianza generalizada

→ Coeficiente de correlación

$$\boxed{r_{xy} = \frac{S_{xy}}{S_x S_y}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Es adimensional
Está acotado $[-1, 1]$
Ver demo. ejercicio 1.10

- Matriz de correlaciones → $\begin{pmatrix} 1 & r_{xy} \\ r_{xy} & 1 \end{pmatrix}$

# 10 TRANSFORMACIONES LINEALES MULTIVARIANTES

Sea $y_i = a_1 X_{1i} + a_2 X_{2i} + \text{\_\_\_\_} + a_k X_{ki} \Rightarrow y_i = \overbrace{(a_1\ a_2 \cdots a_k)}^{\text{Vector de la transformación}} \begin{pmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{ki} \end{pmatrix} = a^T X_i$

Como $y_i = a^T X_i \longrightarrow \boxed{\begin{array}{l} \bar{y} = a^T \bar{X} \\ S_y^2 = a^T S_x^2 a \end{array}}$ Vectorial

Sea

$$\begin{array}{l} y_{1i} = a_{11} X_{1i} + a_{12} X_{2i} + \text{\_\_\_\_} + a_{1k} X_{ki} \\ y_{2i} = a_{21} X_{1i} + a_{22} X_{2i} + \text{\_\_\_\_} + a_{2k} X_{ki} \\ \vdots \\ y_{m1} = a_{m1} X_{1i} + a_{m2} X_{2i} + \text{\_\_\_\_} + a_{nk} X_{ki} \end{array} \Rightarrow \begin{pmatrix} y_{1i} \\ y_{2i} \\ y \vdots \\ y_{mi} \end{pmatrix} = \overbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{1k} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{pmatrix}}^{\text{Matriz de la transformación}} \begin{pmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{ki} \end{pmatrix}$$

Como $y_i = A X_i$

$$\boxed{\begin{array}{l} \bar{y} = A \bar{X} \\ S_y^2 = A S_x^2 A^T \end{array}}$$

## 11 OTRAS FÓRMULAS DE INTERÉS

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (x_i^2 + \bar{x}^2 - 2x_i\bar{x})}{n} = \frac{\sum x_i^2}{n} + \frac{\cancel{n}\,\bar{x}^2}{\cancel{n}} - 2\bar{x}\overset{\bar{x}}{\left(\frac{\sum x_i}{n}\right)} =$$

$$= \boxed{\frac{\sum x_i^2}{n} - \bar{x}^2 = S_x^2}$$

$$\operatorname{cov}(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum (x_i y_i + \bar{x}\bar{y} - \bar{y}x_i - \bar{x}y_i)}{n} =$$

$$= \frac{\sum x_i y_i}{n} + \frac{\cancel{n}\,\bar{x}\bar{y}}{\cancel{n}} - \bar{y}\overset{\bar{x}}{\left(\frac{\sum x_i}{n}\right)} - \bar{x}\overset{\bar{y}}{\left(\frac{\sum y_i}{n}\right)} = \boxed{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y} = \operatorname{cov}(x,y)}$$

## 12 EJERCICIOS

➯ Sea la transformación lineal $y_i = x_{1i} + x_{2i}$ y sabiendo los valores de $\bar{x}_1$ y $\bar{x}_2$; así como $S_x^2 = \begin{pmatrix} S_{x_1}^2 & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2}^2 \end{pmatrix}$. Hallar $\bar{y}$, $S_y^2$ en función de los datos.

Se trata de una transformación lineal que escribiremos de la forma:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (1 \quad 1) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \longrightarrow$$ Expreión vectorial de la transformación

$y_i = a^T X_i$ , con $a = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\bar{y} = a^T \bar{X} = (1\ 1)\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} = \bar{X}_1 + \bar{X}_2 \longrightarrow \boxed{\bar{y} = \bar{X}_1 + \bar{X}_2}$$

$$S_y^2 = a^T S_x^2 a = (1\ 1)\begin{pmatrix} S_{x_1}^2 & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2}^2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} S_{x_1}^2 + S_{x_1x_2} & S_{x_1x_2} + S_{x_2}^2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} =$$

$$= \boxed{S_{x_1}^2 + 2 S_{x_1x_2} + S_{x_2}^2 = S_y^2}$$

¿Y si la transformación fuera $y_i = x_{1i} - x_{2i}$?

$$y_i = \underbrace{(1\ -1)}_{} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \bar{y} = a^T \bar{x} = (1\ -1)\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} = \boxed{\bar{X}_1 - \bar{X}_2 = \bar{y}}$$

$$y_i = a^T \cdot X_i$$

$$S_y^2 = a^T S_x^2 a = (1\ -1)\begin{pmatrix} S_{x_1}^2 & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2}^2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \boxed{S_{x_1}^2 - 2 S_{x_1x_2} + S_{x_2}^2 = S_y^2}$$

Por ultimo, usar la transformación $y_i = 4x_{1i} + 3x_{2i}$

$y_i = (4\ \ 3)\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix}$ $\quad \bar{y} = a^T\bar{X} = (4\ \ 3)\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} = \boxed{4\bar{X}_1 + 3\bar{X}_2 = \bar{y}}$

$y_i = a^T x_i$

$\boxed{S_y^2} = a^T X_i a = (4\ \ 3)\begin{pmatrix} S_{x_1}^2 & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2}^2 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \left(4S_{x_1}^2 + 3S_{x_1x_2} \quad 4S_{x_1x_2} + 3S_{x_2}^2\right)\begin{pmatrix} 4 \\ 3 \end{pmatrix} =$

$= 4^2 S_{x_1}^2 + 4\cdot 3 S_{x_1x_2} + 4\cdot 3 S_{x_1x_2} + 3^2 S_{x_2}^2 = \boxed{4^2 S_{x_1}^2 + 2\cdot 3\cdot 4\, S_{x_1x_2} + 3^2 S_{x_2}^2}$

⇨ Siendo $X_1$ y $X_2$ tal que $\bar{X} = \begin{pmatrix} 10 \\ 12 \end{pmatrix}$ y $S_x^2 = \begin{pmatrix} 36 & 4 \\ 4 & 36 \end{pmatrix}$

Siendo $y_1 = x_1 + x_2$ e $y_2 = x_1 - x_2$, calcula: $\bar{y}$, $S_y^2$

$y_i = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ; $\bar{X} = \begin{pmatrix} \bar{X}_1 = 10 \\ \bar{X}_2 = 12 \end{pmatrix}$ ; $S_x^2: \begin{pmatrix} S_{x_1}^2 = 36 & S_{x_1x_2} = 4 \\ S_{x_1x_2} = 4 & S_{x_2}^2 = 36 \end{pmatrix}$

$y = A\, X$

$\bar{y} = A\bar{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 10 \\ 12 \end{pmatrix} = \begin{pmatrix} 22 \\ -2 \end{pmatrix} \longrightarrow \boxed{\bar{y} = \begin{pmatrix} \bar{y}_1 = 22 \\ \bar{y}_2 = -2 \end{pmatrix}}$

$S_y^2 = A S_x^2 A^T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 36 & 4 \\ 4 & 36 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 40 & 40 \\ 32 & -32 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} S_{y_1}^2 = 80 & S_{y_1y_2} = 0 \\ S_{y_1y_2} = 0 & S_{y_2}^2 = 64 \end{pmatrix}$

**1.1** Datos: 22, 25, 28, 31, 32, 35, 35, 37, 38, 40, 41, 42, 44, 53, 58, 61

a) Media, mediana y desviación típica directamente

$n = 16$ ; $\bar{X} = \frac{\sum_{i=1}^{16} x_i}{16} = \boxed{38{,}875 \equiv \text{Media}}$

$S_x = \sqrt{\frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{16}} = \boxed{10{,}71 \equiv \text{Desviación típica}}$

Mediana $= \frac{x_p + x_{p+1}}{2}$ con $p = \frac{n}{2} = 8$

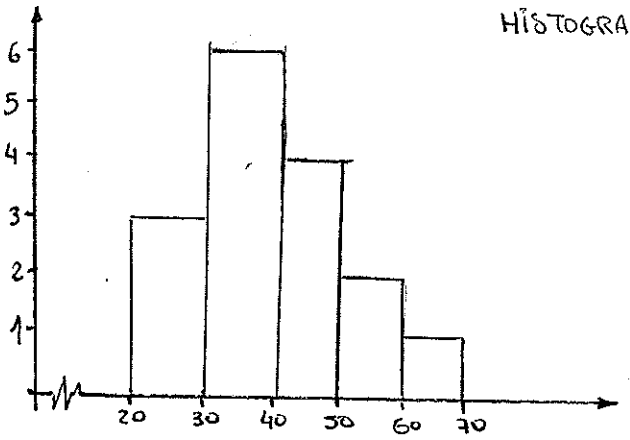
Med $= \frac{37 + 38}{2} = \boxed{37{,}5}$

b) Lo mismo pero agrupando los datos en 5 clases de long. 10.

| Clase | Frec. absoluta | F. rel |
|---|---|---|
| [20-30) | 3 | 3/16 |
| [30-40) | 6 | 6/16 |
| [40-50) | 4 | 4/16 |
| [50-60) | 2 | 2/16 |
| [60-70) | 1 | 1/16 |
| | 16 | |

No entra!

c) Histograma y diagrama de tallo y hojas


HISTOGRAMA

DIAG. TALLOS-HOJAS

| | | |
|---|---|---|
| 3 | 2 | 2 5 8 |
| 9 | 3 | 1 2 5 5 7 8 |
| (4) | 4 | 0 1 2 4 |
| 3 | 5 | 3 8 |
| 1 | 6 | 1 |

**1.2** 4 profesores dan clase a grupos de 10, 18, 22 y 150 alumnos. Preguntamos a los profesores. Calcular la media y la desviación típica

$$\bar{X}_p = \frac{10+18+22+150}{4} = \boxed{50} \qquad S_p = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{4}} = \sqrt{\frac{40^2+32^2+28^2+100^2}{4}} = \boxed{57,9}$$

¿Y preguntando a los alumnos?

$10 \longrightarrow f_{rel} = 10/200$

$18 \longrightarrow f_{rel} = \frac{18}{200}$

$22 \longrightarrow f_{rel} = \frac{22}{200}$

$150 \longrightarrow f_{rel} = \frac{150}{200}$

$\overline{200}$

$$\bar{X}_a = \frac{10^2+18^2+22^2+150^2}{200} = \boxed{117,04}$$

$$S_p = \sqrt{f_{rel}\,(x_i'-\bar{x})^2} = \sqrt{\frac{10}{200}(10-117,04)^2+\frac{18}{200}(18-117,04)^2+\frac{22}{200}(22-117,04)^2+\frac{150}{200}(150-117,04)^2} = \boxed{57,13}$$

<u>NOTA!!</u> Los resultados son tan diferentes debido a la heterogeneidad de los datos

**1.3** Si $y_i = \frac{1}{x_i}$, obtener $\bar{y}$ e $S_y^2$ en función de $\bar{x}$ y $S_x^2$

Se trata de una transformación no lineal

Consideraremos la aproximación $h(x) = h(\bar{x}) + h'(\bar{x})(x-\bar{x}) + \frac{h''(\bar{x})}{2}(x-\bar{x})^2$

$$h(x) = \frac{1}{x}\ ;\ h'(x) = \frac{-1}{x^2},\ h''(x) = \frac{2x}{x^4} = \frac{2}{x^3}$$

Por tanto $y = \frac{1}{x} = \frac{1}{\bar{x}} - \frac{1}{x^2}(x-\bar{x}) + \frac{1}{x^3}(x-\bar{x})^2$

$$\bar{y} \approx h(\bar{x}) + \frac{1}{2}h''(\bar{x})S_x^2 = \frac{1}{\bar{x}} + \frac{1}{2}\frac{2}{\bar{x}^3}S_x^2 = \boxed{\frac{1}{\bar{x}} + \frac{S_x^2}{\bar{x}^3}}$$

$$S_y^2 \approx [h'(\bar{x})]^2 S_x^2 = \left(\frac{-1}{\bar{x}^2}\right)^2 S_x^2 = \frac{1}{\bar{x}^4} S_x^2 = \boxed{\frac{S_x^2}{\bar{x}^4}}$$

**1.7** $S_x^2 = 4$; $S_y^2 = 9$; $z = x+y$. ¿Es posible que $S_z^2 = 2$?

$z_i = \overbrace{(1,1)}^{a^T} \overbrace{\begin{pmatrix} x \\ y \end{pmatrix}}^{M}$; $z = a^T \cdot M$; $\underline{S_z^2 = a^T S^2 a}$, con $S^2 = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix}$

$$S_z^2 = (1\ 1)\begin{pmatrix} 4 & S_{xy} \\ S_{xy} & 9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = (4+S_{xy} \quad S_{xy}+9)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4 + S_{xy} + S_{xy} + 9 = 13 + 2S_{xy}$$

Para que $S_z^2 = 2 \longrightarrow S_{xy} = -\frac{11}{2}$; $S^2 = \begin{vmatrix} 4 & -\frac{11}{2} \\ -\frac{11}{2} & 9 \end{vmatrix} = 36 - \frac{11^2}{4} = 30 - 30{,}25 > 0 \Rightarrow \boxed{\text{Es posible}}$

**1.8** Sea la transformación $W_i = K_1 x_i$, $\Omega_i = K_2 y_i$
$K_1$ y $K_2 \to$ mismo signo

Demo $r_{w\Omega} = r_{xy}$ $\bar{w} = K_1 \bar{x}$, $\bar{\Omega} = K_2 \bar{y}$

$r_{xy} = \frac{S_{xy}}{S_x S_y}$; $S_w^2 = \frac{\sum(w_i - \bar{w})^2}{n} = \frac{\sum(K_1 x_i - K_1 \bar{x})^2}{n} = \frac{K_1^2 \sum(x_i - \bar{x})^2}{n} = K_1^2 S_x^2$

$S_w = |K_1| S_x$

$r_{w\Omega} = \frac{S_{w\Omega}}{S_w S_\Omega}$

$S_\Omega^2 = K_2^2 S_y^2$; $S_\Omega = |K_2| S_y$

$$S_{w\Omega} = \frac{\sum(w_i - \bar{w})(\Omega_i - \bar{\Omega})}{n} = \frac{\sum(K_1 x_i - K_1 \bar{x})(K_2 y_i - K_2 \bar{y})}{n} =$$

$$= K_1 K_2 \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n} = K_1 K_2 S_{xy}$$

Por tanto

$$r_{w\Omega} = \frac{K_1 K_2}{|K_1||K_2|}\underbrace{\frac{S_{xy}}{S_x S_y}}_{r_{xy}} = \begin{cases} \boxed{K_1 \text{ y } K_2 \text{ mismo signo} \Longrightarrow r_{w\Omega} = r_{xy}} \\ K_1 \text{ y } K_2 \text{ distinto signo} \Longrightarrow r_{w\Omega} = -r_{xy} \end{cases}$$

**1.9** $y = a + bx$; con $b > 0 \longrightarrow$ Demo $r_{xy} = 1$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{n}\sum(y_i - \bar{y})(x_i - \bar{x})}{S_x \cdot S_y} \overset{\uparrow}{=} \frac{\frac{1}{n}\sum(a + bx_i - a - b\bar{x})(x_i - \bar{x})}{|b| \cdot S_x^2} \underset{b>0}{=}$$

$\bar{y} = a + b\bar{x}$, $S_y = |b| S_x$

$$= \frac{\cancel{b}\underbrace{\frac{1}{n}\sum(x_i - \bar{x})^2}_{S_x^2}}{\cancel{b} \cdot S_x^2} = \frac{S_x^2}{S_x^2} = 1$$

Como queríamos demostrar!

**1.10** Demo $|r| \leq 1$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} \; ; \; S_{xy} = r_{xy} S_x S_y$$

Matriz de varianzas es semidef. positiva

Sea $S = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix}$ ; $|S| = S_x^2 S_y^2 - S_{xy}^2 \geq 0$

$$S_x^2 S_y^2 - r_{xy}^2 S_x^2 S_y^2 \geq 0 \; ; \; S_x^2 S_y^2 (1 - r_{xy}^2) \geq 0$$

$\Longrightarrow (1 - r_{xy}^2) \geq 0$ ; $r_{xy}^2 \leq 1$ ; $\boxed{|r_{xy}| \leq 1}$ como queríamos demostrar

**1.11** $S^2 = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 2 \end{pmatrix}$ ¿Hay error de cálculo?

La matriz es simétrica. $|S| = 16 + 6 + 6 - 4 - 18 - 8 = -2 < 0 \longrightarrow$ La matriz ha de ser definida positiva por tanto no es posible $|S| < 0$

<u>Hay error</u> $\longleftarrow$

**1.12** $S^2 = \begin{pmatrix} 20 & 8 & 7 \\ 8 & 10 & 4 \\ 7 & 4 & 15 \end{pmatrix}$ en cm. Calcular var. generalizada en mm.

$$|S^2_{(mm)}| = 10^3 \cdot |S^2_{(cm)}| = 10^3 \begin{vmatrix} 20 & 8 & 7 \\ 8 & 10 & 4 \\ 7 & 4 & 15 \end{vmatrix} = 10^3 |3000 + 224 + 224 - 490 - 960 - 320| =$$

$$= \boxed{1678}$$

**1.14** Representar diagrama de cajas partiendo de el de tallos y hojas. La rama $-6 \mid 91$ representa $-0{,}69$ y $-0{,}61$

| | | |
|---|---|---|
| 2 | -6 | 91 |
| 2 | -5 | |
| 4 | -4 | 00 |
| 10 | -3 | 766320 |
| 18 | -2 | 98754310 |
| 29 | -1 | 98⑥5432 1100 |
| (16) | -0 | 9977666555 44$\boxed{3}$3211 |
| 36 | 0 | 015566677 |
| 27 | 1 | 233347⑧ |
| 20 | 2 | 134789 |
| 14 | 3 | 23455699 |
| 6 | 4 | 011355 |

$n = 29 + 36 + 16 = 81$

$\text{Med} = X_p$ con $p = \frac{81+1}{2} = 41 \rightarrow \text{Med} = -0{,}03$

$Q_1 = X_r$ con $r = \left[\frac{p+1}{2}\right] = 21 \longrightarrow Q_1 = -0{,}16$

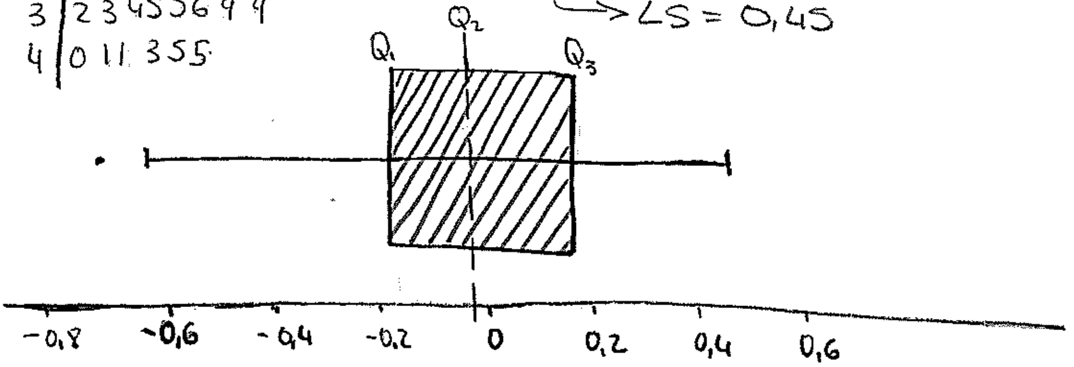
$Q_3 = X_s$ con $s = n - r + 1 = 81 - 21 + 1 = 61 \rightarrow Q_3 = 0{,}18$

$RI = Q_3 - Q_1 = 0{,}18 - (-0{,}16) = 0{,}34$

$LI = Q_1 - 1{,}5 \cdot 0{,}34 = -0{,}16 - 0{,}51 = -0{,}67 \rightarrow LI = -0{,}61$

$LS = Q_3 + 1{,}5 \cdot 0{,}34 = 0{,}18 + 0{,}51 = 0{,}69$

$\hookrightarrow LS = 0{,}45$



1.4 $y = \log X$. Obtener $\bar{y}$ y $S_y^2$ en función $\bar{X}$ y $S_x^2$

Consideraremos las aproximaciones $\begin{cases} \bar{y} = h(\bar{x}) + \frac{1}{2} h''(\bar{x}) \cdot S_x^2 \\ S_y^2 = |h'(\bar{x})|^2 S_x^2 \end{cases}$ con $y = h(x) = \log X$

$h'(x) = \frac{1}{X}$ ; $h''(x) = \frac{-1}{X^2}$

$$\bar{y} = \log(\bar{X}) + \frac{1}{2}\left(\frac{-1}{\bar{X}^2}\right) \cdot S_x^2 = \boxed{\log(\bar{x}) - \frac{1}{2\bar{X}^2} S_x^2}$$

$$\boxed{S_y^2 = \left(\frac{1}{\bar{X}}\right)^2 S_x^2 = \boxed{\frac{S_x^2}{\bar{X}^2}}}$$

1.13 $\bar{X} = 100$ ; $y = \log_{10}(x)$ ; ¿ $\bar{y} = 2.5$ es posible?

Consideramos la aproximación $\bar{y} = h(\bar{x}) + \frac{1}{2} h''(\bar{x}) S_x^2$ con $y = h(x) = \log_{10}(x)$

$h'(x) = \frac{1}{X} \cdot \frac{1}{\ln 10}$

$h''(x) = \frac{-\ln 10}{X^2 (\ln 10)^2}$

$$\bar{y} = \log_{10}(100) - \frac{1}{2} \frac{\ln 10}{100^2 (\ln(10))^2} \cdot S_x^2$$

$$\bar{y} = 2 - \underbrace{\frac{\ln 10}{2 \cdot 100^2 (\ln(10))^2}}_{>0} \cdot S_x^2 < 2 \Longrightarrow \boxed{\bar{y} \neq 2.5}$$ No es posible

Libro 1 Mirar datos en libro

a) Media, desviación típica, cuartiles y RI

Se define la media aritmética como $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ , $\boxed{\bar{X} = 383 \text{ ptas}}$

Se define la desviación típica como $S_x = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} = \boxed{246{,}807 \text{ ptas}}$

Interpretación: la desviación esperada a la media es de 246,807 ptas → alta

$Q_1$ es un nº tal que el 25% de los valores de la muestra son menores o iguales que él.

$Q_2$ o mediana es un nº tal que el 50% de los val de la muestra son $\leq$ que él

$Q_3$ es un nº tal que el 75% de los datos son $\leq$ que él.

Como los datos están ordenados $\longrightarrow$ $Q_2 = \begin{cases} X_{(p)} & \text{con } p = \frac{n+1}{2} \text{ si } n \text{ es impar} \\ \frac{X_{(p)} + X_{(p+1)}}{2} & \text{con } p = \frac{n}{2} \text{ } n \text{ par} \end{cases}$

$n = 70 \rightarrow p = 35 \longrightarrow Q_2 = \frac{X_{(35)} + X_{(36)}}{2} = \frac{300 + 325}{2} = \boxed{312{,}5}$

$Q_1 = X_{(r)}$ con $r = \left[\frac{p+1}{2}\right] = \frac{35+1}{2} = 18$ , $Q_1 = X_{(18)} = \boxed{200}$

$Q_3 = X_{(s)}$ con $s = n - r + 1 = 70 - 18 + 1 = 53$ ; $Q_3 = X_{(53)} = \boxed{520}$

$$\boxed{RI = 520 - 200 = 320}$$

Interpretación de los datos: El 25% de los estudiantes gastan una cantidad inferior o igual a 200 ptas. La mitad de los estudiantes gastan menos de 312,5 ptas y sólo un 25% gasta una cantidad superior o igual a 520 ptas.

b) Diagrama de tallo y hojas

Se tomará en tallo la cifra de las decenas y las hojas la cifra de las decenas de forma que el diagrama queda:

| | |
|---|---|
| 0 | 15 |
| 1 | 01111224455678 |
| 2 | 00111112224557888 |
| 3 | 0025678 |
| 4 | 0113578 |
| 5 | 00011222557 |
| 6 | 005 |
| 7 | 0036 |
| 8 | 0 |
| 9 | 0 |
| 10 | 00 |
| 11 | |
| 12 | 0 |

Interpretación:

- Asimetría positiva → Diagrama desplazado hacia arriba

  Muestra que es mayor el nº de alumnos con gastos bajos que con altos

- Heterogeneidad en los datos: Vemos dos picos en el diagrama

  → Una modalidad en la clase 200 ptas y otra en 500 ptas

- Un valor separado del resto (1200 ptas) que está en consonancia con la asimetría del diagrama

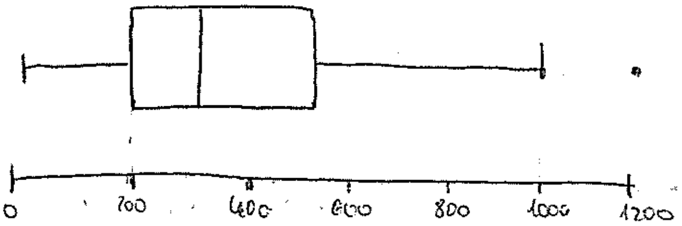
c) Diagrama de cajas

Necesitamos Q₁, Q₂, Q₃, RI
LI, LS

$Q_1 = 200$
$Q_2 = 312{,}5$
$Q_3 = 520$
$RI = 320$

$LS = Q_3 + 1{,}5\,RI = 520 + 1{,}5 \cdot 320 = 1000$

$LI = Q_1 - 1{,}5\,RI = 200 - 1{,}5 \cdot 320 = -280$



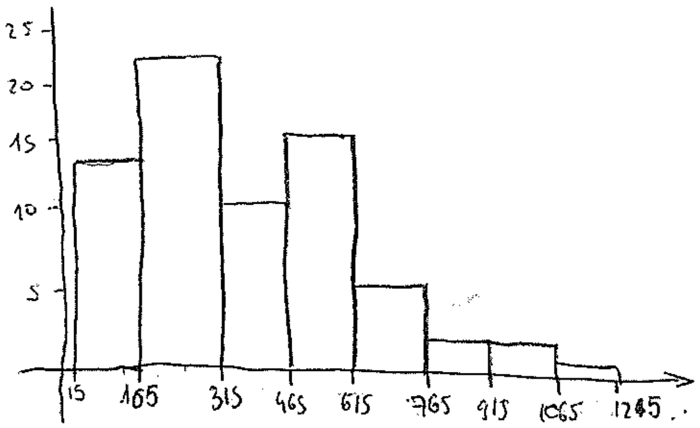
De nuevo vemos el punto atípico en 1200 ptas

Una asimetría positiva por estar la cola del diagrama a la derecha

Mismas conclusiones que en ap. anterior

d) Histograma

nº clases = $\sqrt{\text{nº elementos}} = \sqrt{70} = 8{,}37 \simeq 8$ clases; Ancho clase: $(1200-15)/8 = 148 \simeq 150$; 1ª clase [15,165)



| Clase | F absoluta |
|---|---|
| [15,165) | 13 |
| [165,315) | 22 |
| [315,465) | 10 |
| [465,615) | 15 |
| [615,765) | 5 |
| [765,915) | 2 |
| [915,1065) | 2 |
| [1065,1215) | 1 |

Interpretación

- Asimetría positiva
- Bimodalidad → Hay dos poblaciones (2 picos)

1.2 Una pieza pasa por la maquina 1 y la 2. X e y son los tiempos de cada máquina

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 5 | 15 | 10 | 10 | 10 | 5 | 10 | 10 | 15 | 10 |
| y | 10 | 15 | 15 | 10 | 10 | 15 | 20 | 15 | 20 | 20 |
| | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | | ↑ |

a) Distribución de frecuencias de los tiempos de ambas máquinas

| | X=5 | X=10 | X=15 |
|---|---|---|---|
| y=10 | $1/10$ | $2/10$ | 0 |
| y=15 | $1/10$ | $2/10$ | $1/10$ |
| y=20 | 0 | $2/10$ | $1/10$ |

Distribución conjunta de (X, y)

b) Distribución de la maquina 1, $\overline{X}$, $S_x^2$

| | X=5 | X=10 | X=15 |
|---|---|---|---|
| Frel | $2/10$ | $6/10$ | $2/10$ |

$$\overline{X} = \frac{\sum_{i=1}^{10} x_i}{10} = \boxed{10}$$

$$S_x^2 = \frac{\sum_{i=1}^{10} (x_i - \overline{x})}{10} = \boxed{10}$$

c) Covarianza + interpretación , $\overline{y} = 15$

$$Cov(x,y) = \frac{\sum_{i=1}^{10} (x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{25+0+0+0+0+0+0+0+25}{10} = \boxed{5}$$

La covarianza es positiva por lo que ambas variables crecen o decrecen simultaneamente. De forma que cuando una pieza invierte más tiempo en la máquina 1, también se espera que invierta un tiempo grande en la 2.

d) Coste $= 0.8X + 0.6Y \rightarrow$ Distribución de frecuencias de los costes, media y varianza.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 5 | 15 | 10 | 10 | 10 | 5 | 10 | 10 | 15 | 10 |
| y | 10 | 15 | 15 | 10 | 10 | 15 | 20 | 15 | 20 | 20 |
| coste | 10 | 21 | 17 | 14 | 14 | 13 | 20 | 17 | 24 | 20 |

Distribución de frecuencias

| Coste | 10 | 13 | 14 | 17 | 20 | 21 | 24 |
|---|---|---|---|---|---|---|---|
| Frel | $1/10$ | $1/10$ | $2/10$ | $2/10$ | $2/10$ | $1/10$ | $1/10$ |

$$\boxed{\overline{C} = 17}$$

$$\boxed{S_c^2 = 16,6}$$

# PROBABILIDAD Y VARIABLE ALEATORIA

## 1 CONCEPTOS BÁSICOS

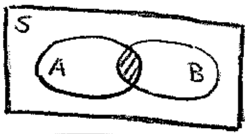
- **Experimento aleatorio:** Proceso cuyo resultado no es conocido de antemano
- **Espacio muestral:** Conjunto formado por todos los resultados posibles de un experimento aleatorio
- **Suceso:** Cualquier subconjunto del espacio muestral. $\emptyset \equiv$ suceso imposible

Las operaciones con sucesos se resumen en:

Sean A y B dos subconjuntos de S:

Unión: $\boxed{A \cup B} = \{x : (x \in A) \text{ o } (x \in B)\}$

Intersección: $\boxed{A \cap B} = \{x : (x \in A) \text{ y } (x \in B)\}$



Complementario: $\boxed{\bar{A}} = \{x : x \notin A\}$

Sus propiedades son:

Conmutativa $\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \end{cases}$ ; Asociativa $\begin{cases} A \cup (B \cup C) = (A \cup B) \cup C \\ A \cap (B \cap C) = (A \cap B) \cap C \end{cases}$

Distributiva $\begin{cases} A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \end{cases}$

Leyes de Morgan $\begin{cases} \overline{A \cup B} = \bar{A} \cap \bar{B} \\ \overline{A \cap B} = \bar{A} \cup \bar{B} \end{cases}$

## 2 AXIOMAS DE PROBABILIDAD Y PROPIEDADES

Dado un espacio muestral S, una función de probabilidad asigna valores P(A) a cada suceso $A \subset S$ y satisface:

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- Si una secuencia de sucesos $A_1, A_2, \ldots A_n$ cumple $A_i \cap A_j = \emptyset \;\; \forall i \neq j \longrightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

Propiedades

$P(\emptyset) = 0 \longleftrightarrow P(S \cup \emptyset) = P(S) + P(\emptyset) = 1 + 0 = 1$

$P(\bar{A}) = 1 - P(A)$

Si $A \subset B$ entonces $P(A) \leq P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Para n sucesos $A_1, A_2, \ldots A_n \subset S$

$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \underset{\text{de orden 2}}{\text{intersecciones}} + \underset{\text{de orden 3}}{\text{intersecciones}} + \ldots$

# 3 PROBABILIDAD CLÁSICA

Sea un experimento con un número finito N de resultados excluyentes y equiprobables, la probabilidad del suceso A es:

$$p(A) = \frac{N(A)}{N}$$

N: nº resultados posibles

N(A): nº resultados favorables al suceso A

Ejemplos:

a) Lanzar moneda → Salir cara

$S: \{c, x\}$ $\quad p(c) = \frac{1}{2}$

b) Lanzar dado → Salir par

$S: \{1,2,3,4,5,6\}$ $\quad p(par) = \frac{3}{6} = \frac{1}{2}$

c) Lanzar dos dados → "suma 7"

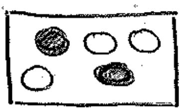
$N = 6 \times 6 = 36$

$N(suma\ 7) = 6$

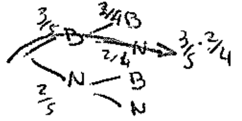
$p(suma\ 7) = \frac{6}{36} = \frac{1}{6}$

| | DADO 1 | | | | | |
|---|---|---|---|---|---|---|
| DADO 2 | (1,1) | (2,1) | (3,1) | (4,1) | (5,1) | (6,1) |
| | (1,2) | (2,2) | (3,2) | (4,2) | (5,2) | (6,2) |
| | (1,3) | (2,3) | (3,3) | (4,3) | (5,3) | (6,3) |
| | (1,4) | (2,4) | (3,4) | (4,4) | (5,4) | (6,4) |
| | (1,5) | (2,5) | (3,5) | (4,5) | (5,5) | (6,5) |
| | (1,6) | (2,6) | (3,6) | (4,6) | (5,6) | (6,6) |

d) Urna → Sacar 1ª blanca y 2ª negra sin reponer



$$p(B\ y\ N) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} = \frac{3}{10}$$



$N = 25 - 5 = 20$

$N(B\ y\ N) = 6$

| | BOLA 1 | | | | |
|---|---|---|---|---|---|
| BOLA 2 | ~~B1 B1~~ | B2 B1 | B3 B1 | N1 B1 | N2 B1 |
| | B1 B2 | ~~B2 B2~~ | B3 B2 | N1 B2 | N2 B2 |
| | B1 B3 | B2 B3 | ~~B3 B3~~ | N1 B3 | N2 B3 |
| | B1 N1 | B2 N1 | B3 N1 | ~~N1 N1~~ | N2 N1 |
| | B1 N2 | B2 N2 | B3 N2 | N1 N2 | ~~N2 N2~~ |

y reponiendo $N = 25$

$N(B\ y\ N) = 6 \longrightarrow p(B\ y\ N) = \frac{6}{25}$

# 4 COMBINATORIA

Número posible de reordenaciones de n objetos tomados de r en r.

| | SIN REENPLAZAMIENTO | CON REEMPLAZAMIENTO |
|---|---|---|
| IMPORTA EL ORDEN | $\frac{n!}{(n-r)!}$ | $n^r$ |
| NO IMPORTA EL ORDEN | $\binom{n}{r}$ | $\binom{n+r-1}{r}$ |

- Primitiva: Se eligen 6 números distintos del 1 al 49

P(acertar los 6)

No hay remplazamiento pues los núm. son distintos
NO importa el orden $\Big\} \Longrightarrow$

$$\Rightarrow N = \binom{49}{6} = \frac{49!}{6!(49-6)!} = \frac{49 \cdot \cancel{48} \cdot 47 \cdot 46 \cdot \overset{3}{\cancel{45}} \cdot 44}{\cancel{6} \cdot \cancel{8} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2}} = 13983816$$

$$p(acertar\ 6) = \frac{1}{13983816} = 0,000000072$$

p(acertar 5)

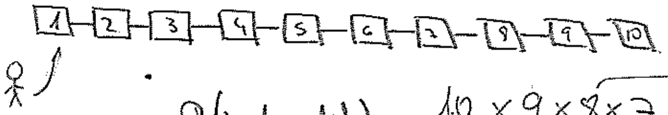
$$p(acertar\ 5) = \frac{\binom{6}{5}\binom{43}{1}}{13983816} = \frac{6 \cdot 43}{13983816} = 0,000018$$

$$p(\text{acertar }4)=\frac{\binom{6}{4}\binom{43}{2}}{13983816}=\frac{\frac{6!}{4!\,2!}\,\frac{43\cdot 42}{2!}}{13983816}=\frac{6\cdot 5\cdot 43\cdot 42}{4\cdot 13983816}=0{,}00097$$

$$p(\text{acertar }0)=\frac{\binom{6}{0}\binom{43}{6}}{13983816}=\frac{43\cdot\overset{7}{\cancel{42}}\cdot 41\cdot\cancel{40}\cdot\overset{13}{\cancel{39}}\cdot 38}{\cancel{6}\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot 13983816}=0{,}44$$

$$p(\text{salir }1)=\frac{\binom{1}{1}\binom{47}{5}}{13983816}=\frac{48\cdot 47\cdot 46\cdot 45\cdot 44}{5\cdot 4\cdot 3\cdot 2\cdot 13983816}=\frac{720}{49}\cdot\frac{1}{5!}=\frac{6}{49}=0{,}1224$$

En una estación de metro hay 5 pasajeros esperando a un tren con 10 vagones. Cada pasajero elige el vagón ¿p(todos en vagón diferente)?



Casos posibles $N$: $10^5$ → Importa el orden, hay remplazamiento, $n=10$, $r=5$

$$p(\text{todos dif})=\frac{10\times 9\times 8\times 7\times 6}{10^5}=0{,}3024$$

→ el 1º puede meterse en 10 vagones. El sig. sólo en 9, El sig. en 8,...

De un lote con 100 piezas se toman al azar 10. Si todas son buenas se acepta el lote y si no, se rechaza. $\underbrace{p(\text{aceptar lote con 10 piezas malas})}_{P(A)}$

$$N=\binom{100}{10}\;;\quad N(A)=\binom{10}{10}\binom{90}{10}\qquad P(A)=\frac{\binom{10}{10}\binom{90}{10}}{\binom{100}{10}}=\frac{1\cdot\frac{90!}{10!\,80!}}{\frac{100!}{10!\,90!}}=\frac{10!\,(90!)^2}{10!\,80!\,100!}=$$

$$=\frac{\cancel{90!}\;90\cdot 89\cdot 88\cdot 87\cdot 86\cdot 85\cdot 84\cdot 83\cdot 82\cdot 81\;\cancel{80!}}{\cancel{80!}\;100\;99\;98\;97\;96\;95\;94\;93\;92\;91\;\cancel{90!}}=0{,}33$$

Prob. de que en un grupo de 25 personas haya al menos 2 con el mismo cumpleaños

igual que los trenes

$$p(\text{no coincidencias})=\frac{365\cdot(365-1)(365-2)\dots(365-24)}{365^{25}}=0{,}432$$

$$p(\text{hay una coincidencia o más})=1-p(\text{no coinc.})=1-0{,}43=0{,}578$$

## 5 PROBABILIDAD CONDICIONADA

Sea B un suceso con probabilidad distinta de cero, se define la probabilidad del suceso A condicionado por B como:

$$p(A|B)=\frac{P(A\cap B)}{p(B)}$$

| | Mujeres | Hombres | |
|---|---|---|---|
| Fumadores | 0,12 | 0,18 | → 0,30 |
| No fumadores | 0,39 | 0,31 | → 0,70 |
| | 0,51 | 0,49 | |

- ¿Probabilidad de que, si es hombre, sea fumador?

$$p(F|H)=\frac{0{,}18}{0{,}49}=0{,}37$$

- ¿Prob de que, si es mujer, sea fumadora?

$$p(F|M)=\frac{0{,}12}{0{,}51}=0{,}24$$

- ¿Prob. de que, si es mujer, no sea fumadora? ¿y si es hombre?

$$p(NF|M)=\frac{0{,}39}{0{,}51}=0{,}76$$

$$p(NF|H)=\frac{0{,}31}{0{,}49}=0{,}63$$

1

1

¿Prob de que, si es fumador, sea mujer? ¿y hombre?

$$p(M|F)=\frac{0{,}12}{0{,}3}=0{,}4\,,\quad p(H|F)=\frac{0{,}18}{0{,}3}=0{,}6$$

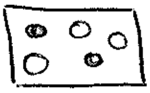
1

Mediante la probabilidad condicionada podemos calcular:

$$p(A|I) = \frac{P(A\cap I)}{p(I)} \longrightarrow \text{Intersección de sucesos} \quad p(A\cap I) = p(A|I)\,p(I)$$

$\hookrightarrow$ Probabilidad de A $\quad p(A) = p[(A\cap I)\cup(A\cap \bar{I})] =$

$= p(A|I)\,p(I) + p(A|\bar{I})\cdot p(\bar{I})$

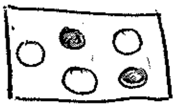
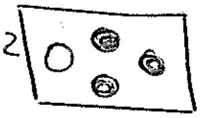
➜ Urna  a) Sin remplazamiento

$p(1^a \text{ blanco } 2^a \text{ negra})$ $\quad P(B1\cap N2) = p(N2|B1)\,p(B1) = \frac{2}{4}\cdot\frac{3}{5} = \frac{6}{20} = \frac{3}{10}$ //

b) Con remplazamiento

$P(B1\cap N2) = p(N2|B1)\,p(B1) = \frac{2}{5}\;\frac{3}{5} = \frac{6}{25}$ //

➜ Prob. de que se extraiga una bola blanco eligiendo una Urna al azar.

U1  U2 

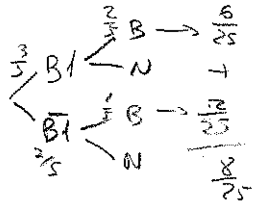
B: Saca blanca $\quad$ U2: Elige Urna 2

U1: Elige Urna 1

$$P(B) = p(B|U1) + p(B|U_2) = p(B\cap U1)\,p(U1) + p(B\cap U2)\cdot p(U2) =$$

$$= \frac{3}{5}\cdot\frac{1}{2} + \frac{1}{4}\;\frac{1}{2} = \frac{3}{10} + \frac{1}{8} = \frac{12+5}{40} = \boxed{\frac{17}{40}}$$

Se toma una bola de U1 y se mete en U2. ¿P(B)?



B: Saca blanca

B1: Saca blanca en 1ª extracción

$$P(B) = p(B\cap B1) + p(B\cap\overline{B1}) =$$

$$= p(B|B1)\,p(B1) + p(B|\overline{B1})\cdot p(\overline{B1}) = \frac{2}{5}\;\frac{3}{5} + \frac{1}{5}\;\frac{2}{5} = \boxed{\frac{8}{25}}$$

## 6 INDEPENDENCIA

Si en el conocimiento de la ocurrencia de B cambia la probabilidad de que ocurra otro suceso A se dice que A y B son dependientes $\rightarrow$ $p(A|B) \neq p(A)$

Si A es i.dependiente de B $\quad p(A|B) = p(A)$ y $p(B|A) = p(B)$

y como $\quad p(A|B) = \frac{P(A\cap B)}{P(B)} \longrightarrow \boxed{p(A\cap B) = p(A)\cdot p(B)}$

Con 3 o más sucesos, para que sean independientes deberá cumplirse que:

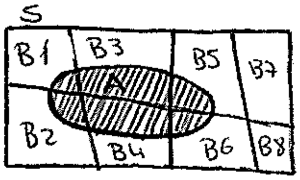
$p(A\cap B\cap C) = p(A)\cdot p(B)\,p(C)$

$p(A\cap B) = p(A)\cdot p(B)$

$P(A\cap C) = p(A)\cdot p(C)$

$p(B\cap C) = p(B)\cdot p(C)$

# 7 PROBABILIDAD TOTAL Y TEOREMA DE BAYES



$$P(A)=P(A\cap S)=P[A\cap[B1\cup B2\cup\ldots\cup BN]]=$$
$$=P[(A\cap B1)\cup(A\cap B2)\cup\ldots(A\cap BN)]=$$
$$=P(A\cap B1)+p(A\cap B2)+\ldots+p(A\cap BN)$$

$$\boxed{p(A)=p(A|B1)\cdot p(B1)+p(A|B2)\cdot p(B2)+\ldots+p(A|BN)\cdot p(BN)}$$

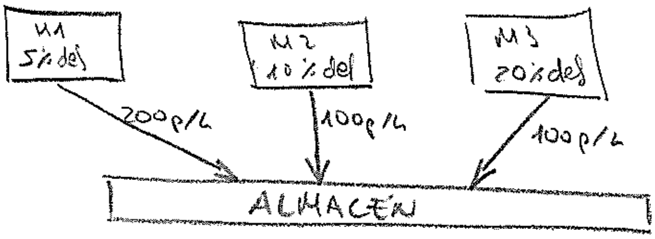
Esto será equivalente a decir $p(A)=p(A|B)\,p(B)+p(A|\bar{B})\cdot p(\bar{B})$

## BAYES

Sea la partición de S $B_1, B_2, \ldots B_n$ con $p(B_i)>0$ $i=1,2,3,\ldots n$
Sea A cualquier suceso con $p(A)>0$. Entonces

$$\boxed{P(B_i|A)=\frac{P(B_i\cap A)}{P(A)}=\frac{P(A|B_i)\cdot P(B_i)}{\sum_{j=1}^{n}P(A|B_j)\,p(B_j)}}$$

➔ Porcentaje de piezas defectuosas fabricadas por tres máquinas es 5%, 10% y 20%
Maquina 1 → 200 piezas/h, Maquina 3 y 2 → 100 piezas/h. Se toma una pieza del almacén y es defectuosa ¿p(procede de M1)? ¿p(procede M2)? ¿p(procede M3)?



D: pieza defectuosa
M1: " maq. 1
M2: " " 2
M3: " " 3

$$P(M1|D)=\frac{P(M1\cap D)}{P(D)}\overset{\uparrow}{=}\frac{p(D|M1)\,p(M1)}{0,1}=\frac{0,05\cdot\frac{1}{2}}{0,1}=\boxed{0,25}$$

$$p(D)=p(M1\cap D)+p(M2\cap D)+p(M3\cap D)=p(D|M1)\,p(M1)+p(D|M2)\,p(M2)+$$
$$+p(D|M3)\,p(M3)=0,05\cdot\frac{1}{2}+0,1\,\frac{1}{4}+0,2\cdot\frac{1}{4}=0,1$$

$$P(M2|D)=\frac{P(M2\cap D)}{P(D)}=\frac{P(D|M2)\cdot p(M2)}{0,1}=\frac{0,1\cdot 1/4}{0,1}=\boxed{0,25}$$

$$p(M3|D)=\frac{P(M3\cap D)}{P(D)}=\frac{P(D|M3)\cdot p(M3)}{P(D)}=\frac{0,2\cdot\frac{1}{4}}{0,1}=\boxed{0,5}$$

Si una persona porta el virus A, el analisis lo detecta el 99% veces. El analisis proporciona falsos positivos indicando presencia del virus en el 3% de los sanos. Solo 5 de cada 1000 tienen virus ¿p(tener el virus si el analisis da positivo?

V. tener virus
P: analisis positivo

Se pide $P(V|P)$

Datos $P(P|V) = 0{,}99$
$P(P|\bar{V}) = 0{,}03$
$P(V) = \frac{5}{1000} = 0{,}005$

$$p(V|P) = \frac{P(V \cap P)}{P(P)} \underset{\uparrow}{=} \frac{P(P|V) \cdot p(V)}{P(P)} = \frac{0{,}99 \cdot 0{,}005}{0{,}0348} = \boxed{0{,}142}$$

$$p(P) = p(P|V) \cdot p(V) + p(P|\bar{V}) \cdot p(\bar{V}) =$$
$$= 0{,}99 \; 0{,}005 + 0{,}03 \cdot 0{,}995 =$$
$$= 0{,}0348$$

Estudio 1.000.000 →

| | Sanos ↓ | Enfermos ↓ |
|---|---|---|
| Dan ⊕ → | 29850 | 4950 |
| Dan ⊖ → | 965150 | 50 |

Hay 5000 enfermos
$V \cap P = 5000 \cdot 0{,}99 = 4950$
$V \cap \bar{P} = 5000 - 4950 = 50$

Hay 995000 sanos →
$\bar{V} \cap P = 995000 \cdot 0'3 = 29850$
$\bar{V} \cap \bar{P} = 995000 - 29850 = 965150$

$P = 29850 +$
$+ 4950 =$
$= 34800$

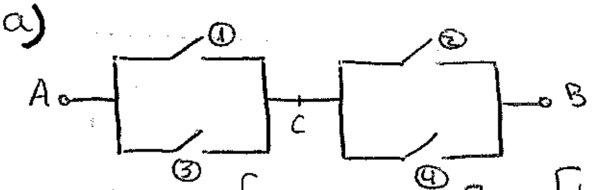
Por tanto, de los 34.800 positivos, sólo 4.950 tenían el virus

$$P(V|P) = \frac{4950}{34800} = 0{,}142$$

# 8 EJERCICIOS DE PROBABILIDAD

(2.1) Probabilidad de que el interruptor esté cerrado es p. Calcular la probabilidad de que pase corriente de A a B.

a)



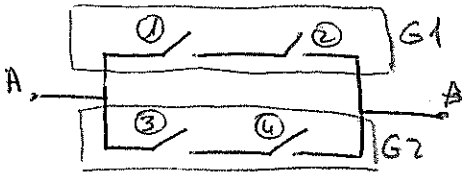
AB: pasa corriente de AB
AC: " " " AC

$$P(AB) = P[(AC) \cap (CB)] \underset{\uparrow}{=} [1-(p-1)^2]^2$$

$$p(AB) = [1-(1-p)^2]^2$$

$$p(AC) = p(CB) = 1 - p(\overline{AC}) = 1 - \underset{\substack{\uparrow \\ 1\,\text{abierto}}}{(p-1)}\underset{\substack{\nwarrow \\ 3\,\text{abierto}}}{(p-1)}$$

b)



G1: pasa corriente por G1
G2 " " " G2

$$p(AB) = 1 - p(\overline{AB}) = 1 - p(\overline{G1})\,p(\overline{G2}) = \boxed{1-(1-p^2)^2}$$

$$p(\overline{G1}) = 1 - p(G1) = 1 + p^2$$

**2.2** Para cualquier par de sucesos $A_1, A_2$ se cumple que:

$$p(A_1^{\cup}A_2) \geq P(A_1) + p(A_2) - 1 \longleftarrow \text{DESIGUALDAD DE BONFERRONI}$$

Demostrar que para n sucesos se cumple $P(A_1 A_2 \dots A_n) = \sum_{i=1}^{n} P(A_i) - (n-1)$

n=2 $p(A_1^{\cup}A_2) \geq p(A_1) + p(A_2) - 1$

n=3 $P(\underbrace{A_1^{\cup}A_2}_{B}{}^{\cup}A_3) = p(BA_3) \geq p(B) + p(A_3) - 1 = p(A_1 \cup A_2) + p(A_3) - 1 \geq p(A_1) + p(A_2) + p(A_3) - 2$

n-1) $p(\underbrace{A_1 \cup A_2 \cup \dots}_{B} \cup A_{n-1}) = p(B A_{n-1}) \geq p(B) + p(A_{n-1}) - 1 \Rightarrow p(A_1 \cup A_2 \cup \dots \cup A_{n-2}) + p(A_{n-1}) - 1 \geq$

$\geq p(A_1) + p(A_2) + \dots + p(A_{n-1}) - (n-2)$

Generalizando para n

$$P(A_1 A_2 \dots A_n) \geq p(A_1) + p(A_2) + \dots + p(A_n) - (n-1) \quad \text{CQD}$$

**2.3** P(componente se averíe antes de 100h) = 0,01. Hay 50 componentes.

P(averíe la maquina antes de 100h)

a) La maquina se avería si lo hace uno o más componentes.

$p(\text{avería}) = 1 - p(\text{no avería}) = 1 - [p(\text{funcionen todos})] = 1 - [p(\text{funcione } 1)]^{50} =$

$= 1 - (1-0,01)^{50} = 1 - 0,99^{50} = \boxed{0,395}$

b) La maquina se avería si fallan dos o más componentes

$p(\text{avería}) = 1 - p(\text{no avería}) = 1 - [p(\text{funcionen todos}) + p(\text{funcionen } 49)] =$

$= 1 - 0,99^{50} - \binom{50}{49} \cdot 0,99^{49} \cdot 0,01 = \boxed{0,089}$

↳ $0,99^{49} \cdot 0,01 + 0,99 \cdot 0,01 + \dots$

¿Cuántas formas podemos ordenar 49 el. de 50? $\binom{50}{49}$

c) La máquina se avería si lo hacen todos los componentes

$p(\text{avería}) = p(\text{no funcione elemento})^{50} = \boxed{0,01^{50}}$

**2.4** ¿Cual es mayor? $\underbrace{P(AB|A)}_{P1}$ o $\underbrace{P(AB|A\cup B)}_{P2}$

$P1 = P(AB|A) = \frac{P(AB \cap A)}{P(A)} = \frac{P(AB)}{P(A)}$

$P2 = P(AB|A\cup B) = \frac{P(AB \cap (A\cup B))}{P(A\cup B)} = \frac{P(AB)}{P(A\cup B)}$

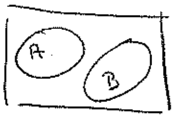
Como $p(A\cup B) \geq p(A)$

$\rightarrow \frac{p(AB)}{p(A)} \geq \frac{p(AB)}{p(A\cup B)}$

$\boxed{p(1) \geq p(2)}$

**2.5** A, B. $p(A)>0$ y $p(B)>0$. Verdadero o falso

a) Si A y B son mutuamente excluyentes entonces no pueden ser independientes



Condición de independencia

$$p(A\cap B) = p(A)\cdot p(B)$$

Excluyente $\longrightarrow p(A\cap B)=0 \Longrightarrow p(A\cap B) \neq p(A)\cdot p(B)$ pues $p(A)>0$ y $p(B)>0$

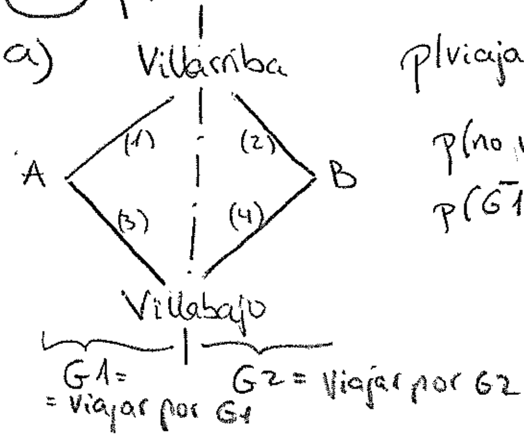
$\boxed{\text{Verdadero}}$

b) Si A y B son independientes entonces no pueden ser mutuamente excluyentes

$p(A\cap B) = p(A)\,p(B)$ $\boxed{\text{Verdadero}}$

**2.6** p(tramo transitable) = 0,8. ¿p(viajar Villarriba a Villabajo?

a)

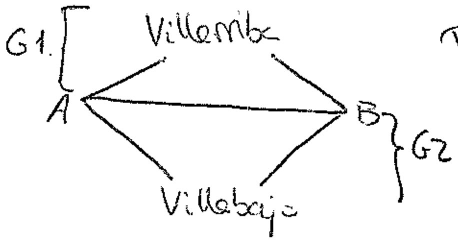


$p(\text{viajar}) = 1 - p(\text{no viajar}) = 1-(1-0,8^2)^2 = \boxed{0,8704}$

$p(\text{no viajar}) = p(\bar{G1})\cdot p(\bar{G2})$

$p(\bar{G1}) = 1 - p(G1) = 1-(0,8)^2$

b) Se construye el tramo de carretera AB con p(viajar AB) = 0,8.
¿p(viajar Villarriba a Villabajo)?



$p(\text{viajar}) = \underbrace{p(\text{viajar}|\bar{AB})}_{0,8704}\cdot\underbrace{p(\bar{AB})}_{0,2} + p(\text{viajar}|AB)\cdot\underbrace{p(AB)}_{0,8} \longrightarrow$

$\Longrightarrow p(\text{viajar}|AB) = p(G1)\cdot p(G2) = (1-(1-0,8)^2)^2 = 0,9216$

$p(G1) = 1 - p(\bar{G1}) = 1-(1-0,8)^2$

$p(\text{viajar}) = 0,8704\cdot 0,2 + 0,9216\cdot 0,8 = \boxed{0,91136}$

**2.7** MONTY HALL: Concursante elige entre 3 puertas. Una tiene premio. Elige y el presentador le muestra que detrás de una de las otras dos no hay premio y le ofrece cambiar. ¿Qué debe hacer?

$A_1$: elegir premio

$$p(\text{acertar cambiando}) = p(\text{acert}|A_1)\,p(A_1) + p(\text{acierto}|\bar{A}_1)\,p(\bar{A}_1) =$$
$$= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \boxed{\frac{2}{3}}$$

$$p(\text{acertar sin cambio}) = p(\text{aciert}|A_1)\cdot p(A_1) + p(\text{aciert}|\bar{A}_1)\cdot p(\bar{A}_1) =$$
$$= 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \boxed{\frac{1}{3}}$$

**2.8** Demo: Si A,B son independientes, también lo son $A$ y $\bar{B}$, y $\bar{A}$ y $\bar{B}$

A y B independientes ⟶ $p(A\cap B) = p(A)\cdot p(B)$; $\begin{cases} p(A|B) = \frac{p(A\cap B)}{p(B)} = p(A) \\ p(B|A) = \frac{p(B\cap A)}{p(A)} = p(B) \end{cases}$

$A$ y $\bar{B}$)

$$p(A) = \underbrace{p(A|B)}_{p(A)} \cdot \underbrace{p(B)}_{p(B)} + p(A|\bar{B})\,p(\bar{B}) = p(A)p(B) + p(A|\bar{B})\,p(\bar{B});$$

$$p(A) - p(A)p(B) = p(A|\bar{B})\cdot p(\bar{B}); \quad p(A)\,(\underbrace{1-p(B)}_{p(\bar{B})}) = p(A|\bar{B})\,p(\bar{B})$$

$$p(A)\cancel{p(\bar{B})} = p(A|\bar{B})\cdot\cancel{p(\bar{B})}; \quad p(A) = p(A|\bar{B}) \longrightarrow$$ A es independiente de $\bar{B}$ ⟹

⟹ $\boxed{A \text{ y } \bar{B} \text{ son independientes}}$ ⟶ $\begin{cases} p(\bar{A}|B) = p(\bar{A}) \\ p(\bar{B}|A) = p(\bar{B}) \end{cases}$; $p(\bar{A}\cap B) = p(\bar{A})\,p(B)$

$$p(A|\bar{B}) = \frac{p(A\cap\bar{B})}{p(\bar{B})} = p(A) \longrightarrow \underline{p(A\cap\bar{B}) = p(A)\cdot p(\bar{B})}$$

$\bar{A}$ y $\bar{B}$)
$$p(\bar{A}) = \underbrace{p(\bar{A}|B)}_{p(\bar{A})}\,p(B) + p(\bar{A}|\bar{B})\,p(\bar{B});$$

$$p(\bar{A}) - p(\bar{A})\,p(B) = p(\bar{A}|\bar{B})\,p(\bar{B}); \quad p(\bar{A})\,(\underbrace{1-p(B)}_{p(\bar{B})}) = p(\bar{A}|\bar{B})\,\cancel{p(\bar{B})}$$

$$\underbrace{p(\bar{A}) = p(\bar{A}|\bar{B})}\ \longrightarrow \boxed{\bar{A} \text{ y } \bar{B} \text{ son independientes}}$$

$$p(\bar{A}|\bar{B}) = \frac{p(\bar{A}\cap\bar{B})}{p(\bar{B})} = p(\bar{A}); \quad p(\bar{A}\cap\bar{B}) = p(\bar{A})\cdot p(\bar{B}) \longrightarrow$$

**2.9** Estudiantes A y B. A asiste 80% clases y B 60%. Uno está en clase, cual es la probabilidad de que sea A?

X: hay uno en clase $\rightarrow p(X) = p(A\bar{B}) + p(\bar{A}B) = 0,8\cdot 0,4 + 0,2\cdot 0,6$

$$p(A|X) = \frac{p(A\cap X)}{p(X)} = \frac{p(A\bar{B})}{p(A\bar{B})+p(\bar{A}B)} = \frac{0,8\cdot 0,4}{0,8\cdot 0,4+0,2\cdot 0,6} = \boxed{0,72}$$

**2.10** p(componente se averíe) = 0,01. Se comprueba su estado con un ensayo que falla con una p = 0,05 y si el componente está averiado el ensayo no se equivoca. El ensayo indica que se ha averiado el componente. ¿p(realmente este averiado)?

$A_e$: ensayo detecta fallo ; A: averiado realmente

$$P(A|A_e) \underset{\substack{\uparrow\\ \text{Bayes}}}{=} \frac{P(A_e|A)\cdot p(A)}{p(A_e|A)\cdot p(A) + p(A_e|\bar{A})\,p(\bar{A})} =$$

$$= \frac{1\cdot 0,01}{1\cdot 0,01 + 0,05\cdot(1-0,01)} = \frac{0,01}{0,01+0,05\cdot 0,99} = \boxed{0,168}$$

**2.11** 4 fichas marcadas con las letras A, B, C, ABC. Se toma 1 al azar ¿Son independientes los sucesos: $\underbrace{\text{presencia de A}}_{A}$, $\underbrace{\text{presencia de B}}_{B}$ y $\underbrace{\text{presencia de C}}_{C}$?

A B C ABC

Condición de independencia

$$\begin{cases} p(ABC) = p(A)\,p(B)\,p(C) \\ p(AB) = p(A)\,p(B) \\ p(AC) = p(A)\,p(C) \\ p(BC) = p(B)\,p(C) \end{cases}$$

$p(AB) = \frac{1}{4}$ ; $p(A)p(B) = \frac{2}{4}\cdot\frac{2}{4} = \frac{1}{4}$ ✓

$p(AC) = \frac{1}{4}$ , $p(A)p(C) = \frac{2}{4}\cdot\frac{2}{4} = \frac{1}{4}$ ✓

$p(BC) = \frac{1}{4}$ ; $p(B)\cdot p(C) = \frac{2}{4}\cdot\frac{2}{4} = \frac{1}{4}$ ✓

$\underbrace{p(ABC) = \frac{1}{4} ; \; p(A)p(B)p(C) = \frac{2}{4}\cdot\frac{2}{4}\cdot\frac{2}{4}}_{\hookrightarrow\, p(ABC) \neq p(A)p(B)p(C)\; \Longrightarrow} = \frac{1}{8}$

$\Longrightarrow$ Los sucesos no son independientes!

**2.12** 3 personas: A, B C. De las llamadas $\frac{2}{5}$ son para A, $\frac{2}{5}$ para B y $\frac{1}{5}$ para C. A sale 50% tiempo, B y C el 25%. Probabilidad de que:

a) No esté ninguno para responder el teléfono.

$L_A$: Llaman a A ; A: A está en la oficina
$L_B$: " " B ; B: B " " "
$L_C$: " " C ; C: C " " "

$$p(\bar{A}\cap\bar{B}\cap\bar{C}) = p(\bar{A})\,p(\bar{B})\,p(\bar{C}) = \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{32} = \boxed{0{,}03}$$

b) Esté la persona a la que llaman

$$p(L_A A + L_B B + L_C C) = \frac{2}{5}\cdot\frac{1}{2} + \frac{2}{5}\cdot\frac{3}{4} + \frac{1}{5}\cdot\frac{3}{4} = \frac{1}{5} + \frac{3}{10} + \frac{3}{20} = \frac{4+6+3}{20} = \frac{13}{20} =$$
$$= \boxed{0{,}65}$$

c) Haya 3 llamadas para una misma persona

$$p(L_AL_AL_A + L_BL_BL_B + L_CL_CL_C) = [p(L_A)]^3 + (p(L_B))^3 + [p(L_C)]^3 = \left(\frac{2}{5}\right)^3 + \left(\frac{2}{5}\right)^3 + \left(\frac{1}{5}\right)^3 = \frac{17}{125} =$$
$$= \boxed{0{,}136}$$

d) Haya 3 llamadas seguidas para 3 personas diferentes.

$$p(L_AL_BL_C + L_AL_CL_B + L_BL_AL_C + L_BL_CL_A + L_CL_AL_B + L_CL_BL_A) =$$
$$= \left(\frac{2}{5}\cdot\frac{2}{5}\cdot\frac{1}{5}\right)\cdot 6 = \frac{24}{125} = \boxed{0{,}192}$$

**2.13** Campeonato de tenis. Dos secuencias A-B-A o B-A-B donde A y B son los oponentes. Hay que ganar 2 partidos seguidos. A es mejor que B, ¿qué secuencia eliges? SECUENCIA A-B-A

$$P(\text{ganar}) = p(ABA) + p(AB\bar{A}) + p(\bar{A}BA) = aba + ab(1-a) + (1-a)ba =$$
$$= ab(a+1-a+1-a) = ab(2-a)$$

SECUENCIA B-A-B $(2-b)ab$

$$p(\text{ganar}) : p(BAB) + p(BA\bar{B}) + p(\bar{B}AB) = bab + ba(1-b) + (1-b)ab =$$
$$= ab(b+1-b+1-b) = ab(2-b)$$

Como $p(a) < p(b) \longrightarrow ab(2-b) < ab(2-a)$

$\boxed{\text{Mejor secuencia A-B-A}}$

**2.14** Un jurado de 3 miembros decide por mayoría. Dos miembros tienen una probabilidad $p$ de acertar y el tercero lanza una moneda. Si un juez tiene una probabilidad, $p$, de acertar ¿cuál de los dos sistemas es mejor?

A: acierta el 1^er juez $\longrightarrow p(A)=p\ ;\ p(\bar{A})=1-p$
B: " " 2º juez $\longrightarrow p(B)=p\ ;\ p(\bar{B})=1-p$
C: " " 3^er juez $\longrightarrow p(C)=\frac{1}{2}\ ;\ p(\bar{C})=\frac{1}{2}$

$$p(\text{acertar sist. 3 jueces})=p(AB\bar{C}+A\bar{B}C+\bar{A}BC+ABC)=p(AB\bar{C})+p(A\bar{B}C)+p(\bar{A}BC)+p(ABC)=$$
$$=p\cdot p\cdot\frac{1}{2}+p(1-p)\frac{1}{2}+(1-p)p\frac{1}{2}+pp\frac{1}{2}=$$
$$=\cancel{p^2/2}+\frac{p-\cancel{p^2}}{2}+\frac{p-\cancel{p^2}}{2}+\cancel{\frac{p^2}{2}}=p$$

$p(\text{acertar sist 1 juez})=p$

Los dos sistemas tienen igual probabilidad

**2.15** Una comunidad de vecinos tiene una alarma. Si roban se pone en funcionamiento siempre. $p(\text{robo})=0,001$. $p(\text{falsa alarma})=0,01$. Si se declara una señal de alarma ¿$p(\text{sea falsa})$?

A: suena la alarma , R: Roban

$$p(\bar{R}|A)=\frac{p(A|\bar{R})\cdot p(\bar{R})}{p(A|\bar{R})p(R)+p(A|R)p(R)}=\frac{0,01\cdot(1-0,001)}{0,01\cdot(1-0,001)+1\cdot 0,001}=$$

↑ Bayes

$$=\frac{0,00999}{0,01099}=\boxed{0,909}$$

● Una pareja tiene dos hijos. Cual es la probabilidad de que si al menos uno de ellos es varón, lo sean los dos

Casos posibles: $V_1V_2\quad V_1M_2\quad M_1V_2\quad M_1M_2$

$$p(V_1V_2\,|\,(V_1M_2)\cup(M_1V_2)\cup(V_1V_2))=\frac{1/4}{3/4}=\boxed{\frac{1}{3}}$$

● Federer y Nadal están empatados 40-40 y saca Nadal. $p(\text{Nadal gane el punto con su saque})=0,6$. Calcula la $p$ de que gane el juego Nadal

N: gana Nadal 2 puntos $\longrightarrow p(N)=0,6^2$
E: Empatan, mete un punto uno y luego el otro ↳ $p(E)=(0,6\cdot 0,4)2$

N
EN
EEN
EEEN
⋮

$$p(\text{gana Nadal})=p(N)+p(EN)+p(EEN)+p(EEEN)+\ldots=$$
$$=(0,6^2+2\cdot 0,6\cdot 0,4\cdot 0,6^2+(2\cdot 0,6\cdot 0,4)^2\cdot 0,6^2+(2\cdot 0,6\cdot 0,4)^3\,0,6^2=$$

↑ Prog geom $\frac{a_1}{1-\text{razon}}$

$$=\frac{0,6^2}{1-2\cdot 0,6\cdot 0,4}=\boxed{0,6923}$$

# 9 VARIABLE ALEATORIA DISCRETA. Distribución de probabilidad y función de distribución

- Una variable aleatoria es una función que asigna un número real a cada uno de los resultados de un experimento aleatorio

  Cuando los valores que toma una variable aleatoria son finitos o infinitos numerables se dice que es discreta. Ej.: Resultados al lanzar un dado $\{1, 2, 3, 4, 5, 6\}$

  Nº veces que lanzar moneda hasta obtener cara $\{1, 2, 3, 4, \dots\}$

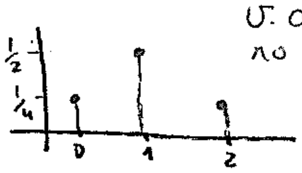
- Sea $\{X_1, X_2, \dots X_n\}$ valores posibles de una variable aleatoria $X$. Se denomina distribución de la probabilidad de la v. aleatoria a $P(X=x)$ que cumple:

$$P(X=x_i) \geq 0$$

$$\sum_{i=1} P(X=x_i) = 1$$

Ej.: Nº caras al lanzar dos monedas

| $X$ | $P(X=x)$ |
|---|---|
| 0 | 1/4 |
| 1 | 1/2 |
| 2 | 1/4 |
| | 1 |



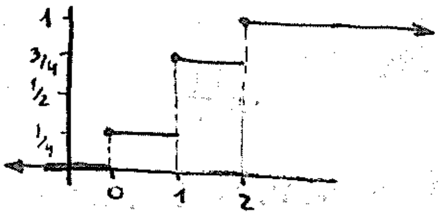
V. aleatoria no uniforme

- La función de distribución $F_x(x)$ de una v. aleatoria $X$ se define para todo número real $x$ como:

$$\boxed{F_x(x) = P(X \leq x)}$$

Ej.: Nº caras al lanzar dos monedas

| $X$ | $F_x(x)$ |
|---|---|
| $(-\infty, 0)$ | 0 |
| $[0, 1)$ | 1/4 |
| $[1, 2)$ | 3/4 |
| $[2, +\infty)$ | 1 |



Propiedades

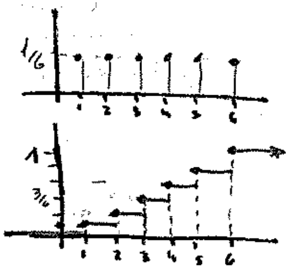
$$\lim_{x \to -\infty} F_x(x) = 0$$

$$\lim_{x \to +\infty} F_x(x) = 1$$

$F_x(x)$ es una función no decreciente

$F_x(x)$ es continua por la derecha

Ej. v.a. discreta uniforme
Lanzamiento de 1 dado



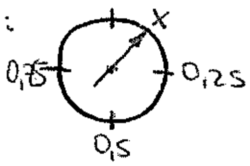
| $x$ | $P(X=x)$ |
|---|---|
| 1 | 1/6 |
| 2 | 1/6 |
| 3 | 1/6 |
| 4 | 1/6 |
| 5 | 1/6 |
| 6 | 1/6 |

$x$ . $F_x(x)$

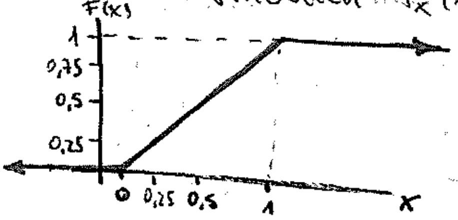
# 10 VARIABLE ALEATORIA CONTINUA. Función de densidad

- Una variable aleatoria $X$ es continua si su función de distribución $F_x(X)$ es continua

Ej.:



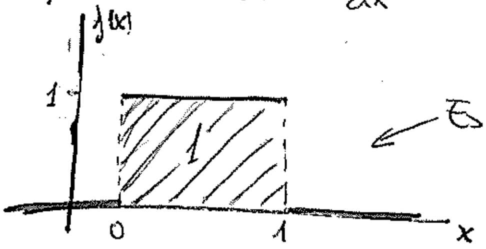
$$F_x(x) = \begin{cases} X & x \in [0,1) \\ 1 & x \geq 1 \\ 0 & x < 0 \end{cases}$$



- La función de densidad de probabilidad $f_x(x)$ de una variable aleatoria continua $X$ es la que verifica $F_x(x) = \int_{-\infty}^{x} f_x(t)\,dt$ $\forall x$ y si $F(x)$ es derivable:

$$\boxed{f_x(x) = \frac{d}{dx} F_x(x)}$$

Ej anterior: $f_x(x) = \frac{d}{dx}(x) = 1 \quad \forall x \in [0,1]$
$0 \quad \forall x \in (-\infty, 0) \cup (1, \infty)$

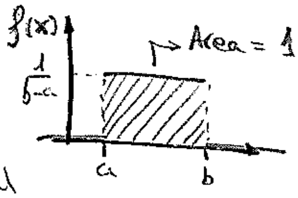


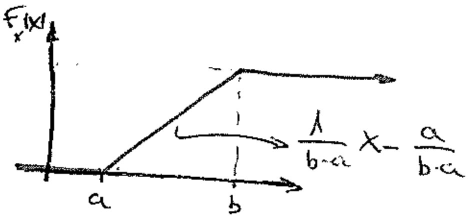
Es uniforme!

Toda variable aleatoria uniforme continua en [a,b) cumple:

$$f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & x\notin[a,b]\end{cases} \qquad F_x(x)=\int_a^x f_x(x)\,dx=\int_a^x\frac{1}{b-a}\,dx=\frac{x-a}{b-a}=\frac{1}{b-a}x-\frac{a}{b-a}$$

Por tanto



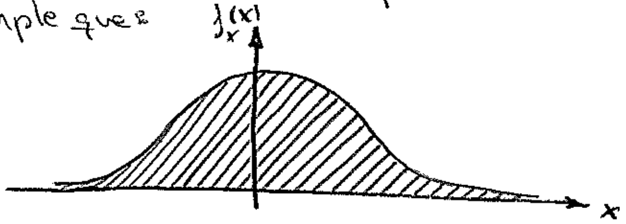


- Propiedad general

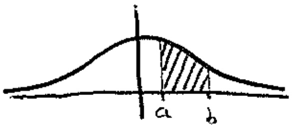
Una función $f_x(x)$ es función de densidad de probabilidad de una variable aleatoria X si y sólo si cumple que:

a) $f_x(x)\geq 0 \quad \forall x$

b) $\int_{-\infty}^{\infty} f_x(x)\,dx=1$



- Cálculo de probabilidad



$$P(a\leq x\leq b)=F_x(b)-F_x(a)=\int_a^b f_x(x)\,dx$$

$$P(X>b)=1-F_x(b)=1-P(X\leq b)=\int_b^{\infty} f_x(x)\,dx$$

## 11 ESPERANZA Y VARIANZA

- Se define esperanza o media de una variable aleatoria discreta X y se representa por E[X] al valor:

$$\boxed{E[X]=\sum_{i=1}^{n} x_i P(X=x_i)}$$

Ej.: Lanzamiento de un dado: $E[X]=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\underline{\underline{3,5}}$

- Se define esperanza o media de una variable aleatoria continua X con función de densidad $f_x(x)$ y se representa por E[X] al valor:

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

Ej.: Distrib. unif. $f_x(x)=1 \quad 0\leq x\leq 1$

$E[X]=\int_0^1 x\cdot 1\,dx=\frac{x^2}{2}\Big|_0^1=\frac{1}{2}$

→ Propiedad: Sea la transformación lineal $Y=aX+b$ con a, b constantes

$$E[Y]=E[aX+b]=aE[X]+b$$

Demo: X v. aleatoria $Y=aX+b$

$$E[Y]=E[aX+b]=\int_{-\infty}^{\infty}(aX+b)f_x(x)\,dx=\int_{-\infty}^{\infty}aXf_x(x)\,dx+\int_{-\infty}^{\infty}b\,f_x(x)\,dx=a\underbrace{\int Xf_x(x)\,dx}_{E[X]}+$$

$$+b\underbrace{\int_{-\infty}^{\infty}f_x(x)\,dx}_{1}=aE[X]+b \qquad \text{CQD.}$$

- Sea X una variable aleatoria con media $\mu$, se denomina varianza a:

$$\boxed{Var(X) = E[(X-\mu)^2]}$$

V. aleatoria discreta

$$Var[X] = \sum_{X=-\infty}^{\infty} (X-\mu)^2 P(X=x)$$

V. aleatorias continuas

$$Var[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x)\, dx$$

→ Propiedades

i) Sea la transformación lineal $Y = aX+b$ con a y b constantes

$$Var[Y] = E[(Y-\mu_y)^2] = E[(aX+\cancel{b}-a\mu_x-\cancel{b})^2] = E[a^2(X-\mu_x)^2] =$$

$$= a^2 E[(X-\mu_x)^2] = a^2 Var[X]$$

ii) $\boxed{Var(X) = E[(X-\mu)^2] = E[X^2]-\mu^2}$

Ejemplos:-Lanzamiento del dado

$$Var[X] = E[X^2]-\mu^2 = \left(1^2 \cdot \frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6}\right) - (3,5)^2 =$$

$$= \frac{35}{12}$$

Otra forma

$$Var[X] = E[(X-\mu)^2] = \sum (X-\mu)^2 \cdot P(X=x) =$$

$$= (1-3,5)^2\frac{1}{6} + (2-3,5)^2\frac{1}{6} + (3-3,5)^2\frac{1}{6} + (4-3,5)^2\frac{1}{6} + (5-3,5)^2\frac{1}{6} + (6-3,5)^2\frac{1}{6} =$$

$$= \frac{35}{12}$$

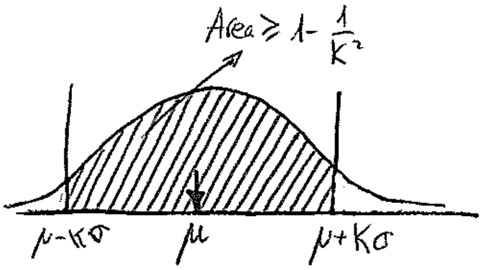
- Distribución uniforme $\longrightarrow f_x(x) = 1 \quad \mu = 0,5$

$$Var(X) = E[(X-\mu)^2] = \int_0^1 (x-\mu)^2 \cdot 1\, dx = \frac{X^3}{3}\Big|_0^1 + 0,5^2 X\Big|_0^1 - 2\frac{X^2}{2}0,5\Big|_0^1 =$$

$$= \frac{1}{3} + 0,5^2 - 0,5 = \frac{1}{12}$$

otra forma: $Var[X] = E[X^2] - \mu_x^2 = \int_0^1 x^2 \cdot 1 dx - 0,5^2 = \frac{X^3}{3}\Big|_0^1 - 0,5^2 = \frac{1}{3} - 0,5^2 = \frac{1}{12}$

# 12 DESIGUALDAD DE TCHEBYCHEV



Para cualquier variable aleatoria

$\mu = E[X]$ y $\sigma^2 = Var[X]$

$$P(|X-\mu| \leq K\sigma) > 1-\frac{1}{K^2}$$

# 13 MEDIANA, MOMENTOS Y $C_{AS}$, $C_{AP}$

- Mediana en variables discretas: Ha de ser un valor de X que cumple que $\rightarrow$ $P_X(X<m) < 0.5$

$P_X(X \leq m) \geq 0.5$

Ejemplo dados

| $X$ | $F_X(x) = P(X \leq x)$ |
|---|---|
| $(-\infty, 0]$ | 0 |
| $[0,1)$ | 0 |
| $[1,2)$ | 1/6 |
| $[2,3)$ | 2/6 |
| $[3,4)$ | 3/6 |
| $[4,5)$ | 4/6 |
| $[5,6)$ | 5/6 |
| $[6,\infty)$ | 1 |

$m=3$?

$P(X<3) = \frac{2}{6} < 0.5$ ✓ $\longrightarrow$ $\underline{\underline{m=3}}$

$P(X \leq 3) = \frac{3}{6} = 0.5 \geq 0.5$ ✓

- Momentos respecto al origen

$\mu_1 = E[X] = \mu$

$\mu_2 = E[X^2]$

$\vdots$

$\mu_p = E[X^p]$

- Momentos respecto a la media

$\alpha_1 = E[(X-\mu)] = 0$

$\alpha_2 = E[(X-\mu)^2] = \sigma^2$

$\vdots$

$\alpha_p = E[(X-\mu)^p]$

Demo $\alpha_1 = 0$ $\alpha_1 = E[(X-\mu)] = 0$

1ª forma $\alpha_1 = E[X-\mu] = E[X] - \mu = \mu - \mu = 0$ CQD

2ª forma $\alpha_1 = E[X-\mu] = \int_{-\infty}^{\infty} (X-\mu) f_x(x)dx = \int_{-\infty}^{\infty} x f_x(x)dx - \int_{-\infty}^{\infty} \mu f_x(x)dx =$

$$= E[X] - \mu \underbrace{\int_{-\infty}^{\infty} f_x(x)dx}_{1} = \mu - \mu = 0 \quad \text{CQD}$$

- $C_{AS} = \frac{\alpha^3}{\sigma_x^3} = \frac{E[(X-\mu_x)^3]}{\sigma_x^3}$ $C_{AP} = \frac{\alpha^4}{\sigma_x^4} = \frac{E[(X-\mu_x)^4]}{\sigma_x^4}$

# 14 TRANSFORMACIONES NO LINEALES

Son transformaciones de la forma $y = h(X)$

- Aproximaciones: $Y \approx h(\mu_x) + h'(\mu_x)(X-\mu_x) + \frac{1}{2} h''(\mu_x)(X-\mu_x)^2$

$$E[y] \approx h(\mu_x) + h'(\mu_x) \cancel{E[(X-\mu_x)]}^{0 \text{ (momento de orden 1)}} + \frac{1}{2} h''(\mu_x) \underbrace{E[(x-\mu_x)^2]}_{\sigma_x^2} ;$$

$$\boxed{E[y] = h(\mu_x) + \frac{1}{2} h''(\mu_x) \sigma_x^2}$$

Aproximamos ahora $y \approx h(\mu_x) + h'(\mu_x)(x-\mu_x)$

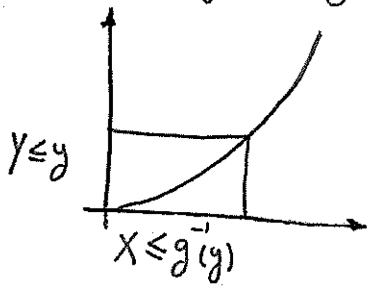
$\mu_y = E[y] = h(\mu_x)$

$$Var[y] \approx E[(y-\mu_y)^2] = E[(\cancel{h(\mu_x)} + h'(\mu_x)(x-\mu_x) - \cancel{h(\mu_x)})^2] = h'(\mu_x)^2 \underbrace{E[(x-\mu_x)^2]}_{\sigma_x^2}$$

$$\boxed{Var[y] = (h'(\mu_x))^2 \sigma_x^2}$$

- Dada una variable aleatoria $X$ con función de densidad $f_x(x)$ veamos como se obtiene $f_y(y)$ de la variable aleatoria $\boxed{Y = g(X)}$

Si la función $g$ es creciente



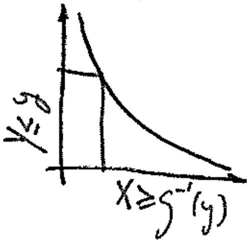
$$F_y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_x(g^{-1}(y))$$

$$\boxed{F_y(y) = F_x(g^{-1}(y))}$$

$$f_y(y) = \frac{d F_y(y)}{dy} = \frac{d F_x(g^{-1}(y))}{d g^{-1}(y)} \cdot \frac{d g^{-1}(y)}{dy} = f_x(g^{-1}(y)) \cdot \frac{d g^{-1}(y)}{dy}$$

$$\boxed{f_y(y) = f_x(g^{-1}(y)) \cdot \frac{d g^{-1}(y)}{dy}}$$

Si la función $g$ es decreciente



$$F_y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - P(X \leq g^{-1}(y)) = 1 - F_x(g^{-1}(y))$$

$$f_y(y) = \frac{d F_y(y)}{dy} = \frac{d}{dy}(1 - F_x(g^{-1}(y))) = \frac{d}{d g^{-1}(y)}(1 - F_x(g^{-1}(y))) \frac{d g^{-1}(y)}{dy} =$$

$$= \boxed{-f_x(g^{-1}(y)) \frac{d g^{-1}(y)}{dy} = f_y(y)}$$

En general diremos que si g es una función monótona

$$\boxed{f_Y(y) = \left|\frac{d g^{-1}(y)}{dy}\right| \cdot f_X(g^{-1}(y))}$$

El radio de una esfera es una variable aleatoria con $f_x(x) = 3x^2 \quad 0 \le x \le 1$ ¿Cuál es la función de densidad del volumen?

$$Y = \frac{4}{3}\pi X^3 = g(x); \quad X = \left(\frac{3y}{4\pi}\right)^{1/3} = g^{-1}(y)$$

$$f_y(y) = \left|\frac{d}{dy}(g^{-1}(y))\right| \cdot f_x(g^{-1}(y)) = \frac{d}{dy}\left[\left(\frac{3y}{4\pi}\right)^{1/3}\right] \cdot 3\left(\left(\frac{3y}{4\pi}\right)^{1/3}\right)^2 =$$

$$= \frac{1}{\cancel{3}}\cancel{\left(\frac{3y}{4\pi}\right)^{-2/3}} \cdot \frac{\cancel{3}}{4\pi} \cdot 3\cancel{\left(\frac{3y}{4\pi}\right)^{2/3}} = \frac{3}{4\pi} \quad \text{con } 0 \le y \le \frac{4\pi}{3}$$

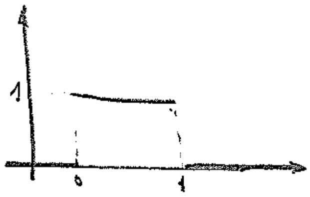
# 15 EJERCICIOS VARIABLE ALEATORIA

(2.16) X v.a. uniforme en (0,1). Calcular $P(Y > 0{,}8)$ si $Y = e^{-x^2}$

$X \longrightarrow f_X(x) = 1$

$\downarrow$

$F_X(x) = \int_0^x 1\,dx = X$



Se pide

$$P(Y>0{,}8) = P(e^{-x^2} > 0{,}8) = P(-X^2 > \log 0{,}8) =$$

$$= P(X^2 \le -\log 0{,}8) = P(X \le \sqrt{-\log 0{,}8}) =$$

$$= F_X(\sqrt{-\log 0{,}8}) = \sqrt{-\log 0{,}8} = \boxed{0{,}472}$$

(2.18) Si X es una variable aleatoria con media $\mu$. Demostrar que cuando $m = \mu$, $E[(x-m)^2]$ es mínima.

$Var[X] = E(x^2) - \mu^2; \quad \sigma^2 + \mu^2 = E[x^2]$

Calculamos $E[(X-m)^2] = E[X^2] + E[m^2] - 2mE[X] =$

$$\sigma^2 + \mu^2 + m^2 - 2m\mu = g(m)$$

$$g'(m) = 0 + 0 + 2m - 2\mu$$

$$g'(m) = 0 \longrightarrow 2m = 2\mu; \quad m = \mu$$

$$g''(m) = 2 > 0 \longrightarrow \boxed{m = \mu \text{ es mínimo}}$$

$\boxed{2.19}$ La función de densidad de v.a. X es:

$$f(x) = \begin{cases} 1/Kx & \text{si } 25 \le x \le 50 \\ 0 & \text{en el resto} \end{cases}$$

Obtener K, la media y la varianza de X.

a) K

$$\int_{-\infty}^{\infty} f(x) = 1 \longrightarrow \int_{-\infty}^{25} \underbrace{\cancel{0 \cdot dx}}_{0} + \int_{25}^{50} \frac{1}{Kx}\,dx + \int_{50}^{\infty} \underbrace{\cancel{0\,dx}}_{0} = 1 ; \quad \frac{1}{K} \ln|X| \Big|_{25}^{50} = 1$$

$$\ln \frac{50}{25} = K ; \quad \boxed{K = \ln 2}$$

b) $$E[X] = \int_{-\infty}^{\infty} x f_x(x) = \int_{25}^{50} x \frac{1}{\ln 2 \cdot x}\,dx = \frac{x}{\ln 2}\Big|_{25}^{50} = \boxed{\frac{25}{\ln 2}} \leftarrow \mu_x$$

$$Var[X] = E[X^2] - \mu_x^2 = \int_{25}^{50} x^2 \frac{1}{\ln 2 \cdot x}\,dx - \left(\frac{25}{\ln 2}\right)^2 = \frac{x^2}{2 \ln 2}\Big|_{25}^{50} - \left(\frac{25}{\ln 2}\right)^2 =$$

$$= \frac{x^2}{2 \ln 2}\Big|_{25}^{50} - \frac{25^2}{(\ln 2)^2} = \frac{50^2 - 25^2}{2 \ln 2} - \frac{25^2}{(\ln 2)^2} = \boxed{51,67} \leftarrow \sigma_x^2$$

Calcular la función de distribución $F_x(x)$

$$F_x(x) = P(X \le x) = \int_{-\infty}^{x} f_x(x)\,dx = \int_{25}^{x} \frac{1}{\ln 2 \cdot x}\,dx = \frac{1}{\ln 2} \ln|x| \Big|_{25}^{x} = \frac{1}{\ln 2}\left(\ln\left(\frac{x}{25}\right)\right) =$$

$$= \begin{cases} \dfrac{\ln\left(\frac{x}{25}\right)}{\ln 2} & \forall x \in [25, 50] \\ 0 & \forall x \in (-\infty, 25) \\ 1 & \forall x \in (50, \infty) \end{cases}$$

Calcular $P(X \ge 30)$

$$P(X \ge 30) = 1 - \overbrace{P(X \le 30)}^{F_x(30)} = 1 - \frac{\ln\left(\frac{30}{25}\right)}{\ln 2} = \boxed{0,737}$$

Otra forma

$$P(X \ge 30) = \int_{30}^{\infty} f_x(x)\,dx = \int_{30}^{50} \frac{1}{\ln 2 \cdot x}\,dx = \frac{1}{\ln 2} \ln\left|\frac{50}{30}\right| = 0,\underline{737}$$ ✓

**2.20** La velocidad $V$ de una molécula de masa $m$ de un gas a temperatura $T$ es una v. aleatoria con $f(v) = \frac{4}{\alpha^3\sqrt{\pi}} v^2 e^{-v^2/\alpha^2}$, $v \geq 0$ donde $\alpha = \sqrt{2KT/m}$. Además $E(V) = 2\alpha/\sqrt{\pi}$ y $Var[V] = (3/2 - 4/\pi)\alpha^2$

a) Calcular el valor medio de la energía cinética $mV^2/2$ ¿A misma $T$, que gas tiene mayor valor medio de energía, uno ligero o uno pesado?

Se pide $E[\frac{mV^2}{2}] = \frac{m}{2}E[V^2] \underset{\uparrow}{=} \frac{m}{2}\frac{3}{2}\alpha^2 = \frac{3}{4}m\alpha^2 = \frac{3}{4}\cancel{m}\frac{2KT}{\cancel{m}} = \boxed{\frac{3}{2}KT}$

$$Var[V] = E(V^2) - (E[V])^2 = \left(\frac{3}{2} - \frac{4}{\pi}\right)\alpha^2$$

$$\rightarrow E[V^2] = \left(\frac{3}{2} - \frac{4}{\pi}\right)\alpha^2 + \left(\frac{2\alpha}{\sqrt{\pi}}\right)^2 = \alpha^2\left(\frac{3}{2} - \cancel{\frac{4}{\pi}} + \cancel{\frac{4}{\pi}}\right) = \frac{3}{2}\alpha^2$$

El valor medio de la energía es independiente de la masa del gas

b) Obtener la función de densidad de la energía cinética de una molécula. Indicar si depende de la masa molecular.

$Y = \frac{mV^2}{2}$, $y = g(V)$

$V = \sqrt{\frac{2y}{m}} = g^{-1}(y)$

$$f_y(y) = \left|\frac{d g^{-1}(y)}{dy}\right| f_v(g^{-1}(y)) = \frac{\frac{2}{m}}{2\sqrt{\frac{2y}{m}}} \cdot \frac{4}{\alpha^3\sqrt{\pi}} \frac{2y}{m} e^{-\frac{2y}{m}/\alpha^2} =$$

$$= \sqrt{\frac{1/m^2}{2y/m}} \cdot \frac{4}{\alpha^3\sqrt{\pi}}\frac{2y}{m} e^{-\frac{2y}{m\alpha^2}} = \sqrt{\frac{1}{2my}} \frac{4m^{3/2}}{(2KT)^{3/2}\pi^{1/2}m} 2y e^{-\frac{2ym}{m2KT}} =$$

$$= \frac{\cancel{2^2}\cancel{m^{3/2}}\,2y}{\cancel{2^2}\cancel{m^{1/2}}y^{1/2}(KT)^{3/2}\pi^{1/2}\cancel{m}} e^{-y/KT} = \frac{2}{(KT)^{3/2}}\sqrt{\frac{y}{\pi}} e^{-y/KT}$$

$$\boxed{f_y(y) = \frac{2}{(KT)^{3/2}}\sqrt{\frac{y}{\pi}} e^{-y/KT} \qquad y \geq 0}$$

**2.21** La función de distribución v.a. $X$ es $F_X(x)$. Obtener la función de densidad de la v. a. $Y = F(X) \rightarrow X = F^{-1}(y)$

$$F_y(y) = P(Y \leq y) = P(F(x) \leq y) = P(F^{-1}(F_x(x)) \leq F_x^{-1}(y)) = P(X \leq F_x^{-1}(y)) =$$

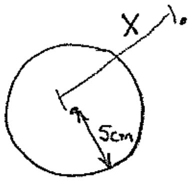
$$= F_X(F_x^{-1}(y)) = y \Longrightarrow F_y(y) = y$$

$$f_y(y) = \frac{dF_y(y)}{dy} \underset{\frac{d}{dy}y = 1}{\overset{\uparrow}{=}} \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

NOTA!! La transformación $Y = F(X)$ convierte la v.a. $X$ en uniforme!!

(2.22) $f(x) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right]$, $x \geq 0$ ; $\sigma \geq 0$

X es la v.a de la distancia del punto de impacto del proyectil al centro del blanco y σ el parámetro que mide la precisión. Si para una distancia determinada de disparo $\sigma = 10$ cm ¿p( lanzar 10 proyectiles y que ninguno impacte a una distancia menor de 5 cm del centro del blanco?



$p(X \leq 5) = p(\text{impactar dentro})$

$p(\text{impactar fuera}) = p(X > 5) = 1 - p(X \leq 5)$

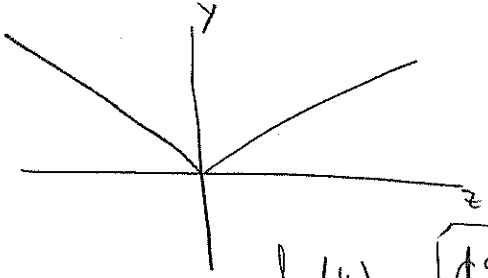
$$p(X \leq 5) = F_X(5) = \int_0^5 \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \underset{\substack{\uparrow \\ \sigma = 10}}{=} \int_0^5 \frac{x}{10^2} e^{-\frac{x^2}{2 10^2}} = -e^{-\frac{x^2}{200}} \Big|_0^5 =$$

$$= e^0 - e^{-25/200} = 1 - 0{,}88 = 0{,}1175$$

$$p(\text{impactar todos fuera}) = (1 - p(X \leq 5))^{10} = (1 - 0{,}1175)^{10} = \boxed{0{,}2865}$$

(2.24) Z v.a $f(z) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{z^2}{2\sigma^2}\right)$; $-\infty < z < \infty$ ; $\sigma > 0$

Obtener la función de densidad de $Y = |Z|$ y su media. $y = g(z)$



$$y = \begin{cases} z & \text{si } z \geq 0 \longrightarrow g^{-1}(y) = z = y \\ -z & \text{si } z < 0 \longrightarrow g^{-1}(y) = z = -y \end{cases}$$

$$f_y(y) = \overbrace{\left|\frac{dg^{-1}(y)}{dy}\right| f_z(g^{-1}(y))}^{\to z \geq 0} + \overbrace{\left|\frac{dg^{-1}(y)}{dy}\right| f_z(g^{-1}(y))}^{\to z < 0} =$$

$$= \underbrace{\left|\frac{d}{dy}(y)\right|}_{1} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} + \underbrace{\left|\frac{d}{dy}(-y)\right|}_{1} \cdot \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(-y)^2}{2\sigma^2}} = \frac{2}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} \qquad y \geq 0$$

$$\boxed{f_y(y) = \frac{2}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} \; ; \; y \geq 0}$$

$$E[y] = \int_0^\infty y\, f_y(y)\, dy = \frac{2}{\sqrt{2\pi\cancel{\sigma}}} (-\sigma^{\cancel{2}}) \int_0^\infty y\, e^{\frac{-y^2}{2\sigma^2}} dy = \frac{-2\sigma}{\sqrt{2\pi}} e^{\frac{-y^2}{2\sigma^2}} \Big|_0^\infty = \boxed{\frac{+2\sigma}{\sqrt{2\pi}}}$$

(2.25) V. a discreta $X$. $P(X=x) = KX$ con $x = 1, 2, 3, \dots 20$.

Se pide $K$ y $E(X \mid X > 10)$

P. aritmetica $= \dfrac{1^{er}\text{ termino} + \text{Ultimo}}{2} \cdot n^{\circ}$ terminos

$$\sum_{i=1}^{20} (P(X=x_i)) = 1 \Rightarrow \sum_{i=1}^{20} K X_i = K \sum_{i=1}^{20} X_i \overset{\downarrow}{=} K \cdot \frac{1+20}{2} \cdot 20 = K \cdot 210 \Longrightarrow$$

$$\Rightarrow K\,210 = 1 \;, \boxed{K = \frac{1}{210}}$$

$$E\left[X \mid X > 10\right) = \sum_{11}^{20} x\, P(x \mid x > 10) = \sum_{11}^{20} X \cdot \frac{X}{155} = \frac{1}{155} \sum_{11}^{20} x^2 = \boxed{16{,}0323}$$

$$P(X \mid x > 10) = \frac{P(x \cdot x > 10)}{P(x > 10)} = \frac{P(X=x)}{p(X > 10)} = \frac{\cancel{K} x}{\cancel{K} \sum_{11}^{20} x} = \frac{x}{\frac{11+20}{2} \cdot 10} = \frac{x}{155}$$

(2.26) V. a. $X$ con $f(x) = \begin{cases} K(1-x^2) & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$

Obtener $K$, la media y la varianza de $Y = 3X - 1$

$$F_x(X) = \int_{-\infty}^{\infty} f(x) = 1 \longrightarrow \int_{-\infty}^{0} \cancelto{0}{0 \cdot dx} + \int_0^1 K(1-x^2)\,dx + \int_1^{\infty} \cancelto{0}{0 \cdot dx} = 1 \longrightarrow$$

$$\longrightarrow K \int_0^1 (1-x^2)\,dx = K\left[x - \frac{x^3}{3}\right]\Big|_0^1 = K\left(1 - \frac{1}{3}\right) = \frac{2}{3}K \longrightarrow \frac{2}{3}K = 1 \;, \boxed{K = \frac{3}{2}}$$

$$E[X] = \int_0^1 X f_x(x)\,dx = \int_0^1 X \frac{3}{2}(1-x^2)\,dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_0^1 = \frac{3}{2} \cdot \frac{1}{4} = \frac{3}{8}$$

$$E[Y] = 3\,E(X) - 1 = 3 \cdot \frac{3}{8} - 1 = \frac{9}{8} - 1 = \frac{9-8}{8} = \boxed{\frac{1}{8}}$$

$$Var[X] = \int_0^1 (X - \mu_x)^2 f_x(x)\,dx = \int_0^1 \left(X - \frac{3}{8}\right)^2 \cdot \frac{3}{2}(1-x^2)\,dx = \frac{3}{2}\int_0^1 \left(X^2 + \frac{9}{64} - \frac{3}{4}X\right)(1-x^2)\,dx =$$

$$= \frac{3}{2}\int_0^1 \left(X^2 + \frac{9}{64} - \frac{3}{4}x - X^4 - \frac{9}{64}x^2 + \frac{3}{4}x^3\right)dx = \frac{3}{2}\left(\frac{X^3}{3} + \frac{9X}{64} - \frac{3X^2}{8} - \frac{X^5}{5} - \frac{9X^3}{64 \cdot 3} + \frac{3}{4}\frac{X^4}{4}\right)\Big|_0^1$$

$$= \frac{3}{2}\left(\frac{1}{3} + \frac{9}{64} - \frac{3}{8} - \frac{1}{5} - \frac{3}{64} + \frac{3}{16}\right) = \frac{3}{2}\,\frac{320 + 135 - 360 - 192 - 45 + 180}{960} = \frac{3 \cdot 38}{2 \cdot 960} =$$

$$= \frac{19}{320}$$

$$Var[y] = 3^2 \cdot Var[X] = 9 \cdot \frac{19}{320} = \boxed{\frac{171}{320}}$$

**2.27** Diametro D de bolas de acero sigue distribución con media $\mu$ y desviación típica $\sigma$. Obtener de forma aproximada media y varianza de los volumenes.

$$V = \frac{4}{3}\pi\left(\frac{D}{2}\right)^3 ;$$

$$\Downarrow$$

$$V = h(D)$$

Consideramos la aproximación

$$\mu_V = h(\mu) + \frac{1}{2}h''(\mu)\sigma^2 \quad y \quad \sigma_V = \left(h'(\mu)\right)^2\sigma^2$$

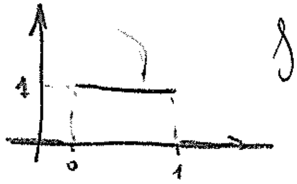
$$h(\mu) = \frac{4}{3}\pi\left(\frac{\mu}{2}\right)^3 ; \quad h'(D) = \frac{4}{\cancel{3}}\pi\frac{\cancel{3}D^2}{2^3} = \frac{\pi D^2}{2} \longrightarrow h'(\mu) = \frac{\pi\mu^2}{2}$$

$$h''(D) = \pi D \longrightarrow h''(\mu) = \pi \cdot \mu$$

$$\boxed{\mu_V = \frac{\pi}{6}\mu^3 + \frac{1}{2}\pi\mu\sigma^2} \qquad \boxed{\sigma_V = \left(\frac{\pi\mu^2}{2}\right)^2\sigma^2}$$

Sea X una v.a. <u>uniforme</u> continua en $[0,1]$. $y = -\log(x) = g(x)$

Calcular $f_y(y)$



$$f_x(x) = 1 \quad 0 \leq X \leq 1$$

$$x = g^{-1}(y)$$

$$y = g(x) = -\log(x) ; \quad X^{-1} = e^{y} ; \quad X = \frac{1}{e^{y}} \quad y \geq 0$$

$$f_y^{(y)} = \left|\frac{d}{dy}g^{-1}(y)\right| f_x(g^{-1}(y)) = \left|\frac{d}{dy}(e^{-y})\right| \cdot 1 = \left|-e^{-y}\right| = e^{-y} ; \quad \boxed{f_y(y) = e^{-y} \text{ con } y \geq 0}$$

Dada X con función de densidad $f_x(x)$ $x \in \mathbb{R}$, calcular $f_y(y)$ si

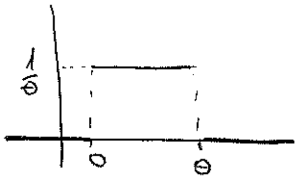
$$Y = X^2 = g(x) ; \quad X = +\sqrt{y} = g^{-1}(y)$$

$$\boxed{f_y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| \cdot f_x(g^{-1}(y)) = \left|\frac{d}{dy}(\sqrt{y})\right| f_x(\sqrt{y}) = \left|\frac{1}{2\sqrt{y}}\right| f_x(\sqrt{y}) ; \quad y \geq 0}$$

Si X fuera continua uniforme en $[a,b]$ → $f_x(x) = \frac{1}{b-a}$

$$\boxed{f_y(y) = \frac{1}{2\sqrt{y}}\left(\frac{1}{b-a}\right) , \quad y \geq 0}$$

(2.28) Si X es una v.a. uniforme entre 0 y θ, obtener la función de densidad de la v.a. $Y=+\sqrt{X}=g(x)$; $X=y^2=g^{-1}(y)$



$$f_X(x)=\frac{1}{\theta} \quad x\in[0,\theta]$$

$$f_y(y)=\left|\frac{d}{dy}(g^{-1}(y))\right|\cdot f_x(g^{-1}(y))=\frac{d}{dy}(y^2)\cdot f_x(y^2)=$$

$$=2y\frac{1}{\theta}=\frac{2y}{\theta}\Rightarrow \boxed{f_y(y)=\frac{2y}{\theta}\ \text{con}\ y\in[0,+\sqrt{\theta}]}$$

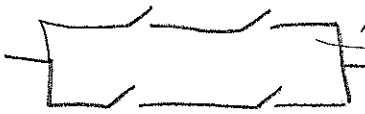
## 16 1er PARCIAL

➔ p(pase corriente por interruptor) = p. Calcular la probabilidad de que pase corriente de A a B

C: pase por tramo central

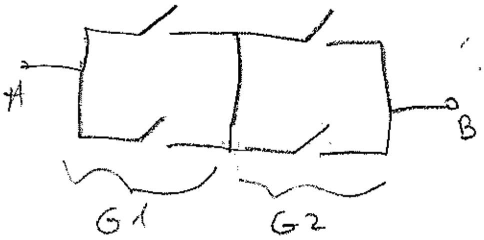
$$p(\text{pasa corriente})=p(AB|C)\cdot p(C)+p(AB|\bar{C})p(\bar{C})$$

Supongamos circuito $\longrightarrow p(AB|\bar{C})=1-(1-p^2)^2$



$$\longrightarrow \underset{\text{pasa por el Grupo}}{p^2}\longrightarrow \underset{\text{no pase por grupo}}{1-p^2}\longrightarrow \underset{\text{no pase por ningún grupo}}{(1-p^2)^2}$$

Ahora, con C cerrado

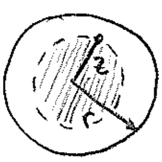


$$p(G1)=1-p(\bar{G1})=1-(1-p)^2=p(G2)$$

$$p(AB|C)=p(G1)\,p(G2)=(1-(1-p)^2)^2=p(AB|C)$$

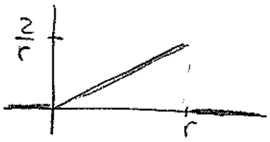
$$\longrightarrow \boxed{p(\text{pasa corriente})=(1-(1-p)^2)^2p+(1-p^2)(1-p)}$$

(2.17) Se elige punto interior a $x^2+y^2=r^2$. Z es v.a. que define la distancia al centro. Calcular $F_z$ y $f_z$



$$F_z(z)=P(Z\le z)=\overset{\text{área del círculo/área total}}{\frac{\pi z^2}{\pi r^2}}=\frac{z^2}{r^2}\longrightarrow \boxed{F_z(z)=\frac{z^2}{r^2}\quad 0\le z\le r}$$

$$f_z(z)=\frac{dF_z}{dz}=\frac{2z}{r^2}\ ;\ \boxed{f_z(z)=\frac{2z}{r^2}\ \text{con}\ 0\le z\le r}$$

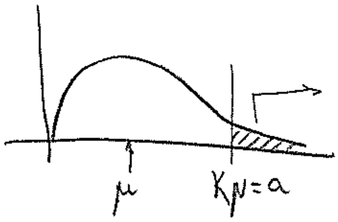


$\boxed{2.23}$ Chebychev: $P(|X-\mu| \leq K\sigma) > 1 - \frac{1}{K^2}$

Demostrar a partir de esta desigualdad la desigualdad de Markov

$$P(X > a) \leq \frac{1}{a} E[X] \quad \text{con} \quad P(X>0) = 1$$

$$\mu = E(X)$$



$$\leq \frac{1}{K} = \frac{1}{a}\mu \longrightarrow p(X > K\mu) \leq \frac{1}{K}$$

$$E[X] = \int_0^\infty x f_x(x)\,dx = \int_0^a x f_x(x)\,dx + \int_a^\infty x\, f_x(x)\,dx \Rightarrow$$

$$\Rightarrow E[X] \geqslant \int_a^\infty x\, f_x(x)\,dx \geqslant \int_a^\infty a\, f_x(x)\,dx \Rightarrow E[X] \geq a P(X \geqslant a) \longrightarrow$$

$$\longrightarrow p(X \geqslant a) \leq \frac{1}{a} E[X] \quad \text{CQD}$$

**Libro:** **2.1** Razonar, si A, B son sucesos con probabilidad no nula y son complementarios, ¿pueden ser independientes?

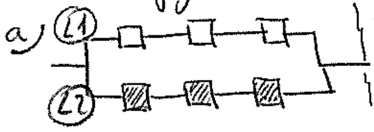
Sucesos complementarios $\longrightarrow \begin{cases} \text{Son excluyentes } A \cap B = \emptyset \\ p(A+B) = 1 \end{cases}$

$p(A \cap B) = 0 \longrightarrow$ Si fueran independientes $p(A \cap B) = p(A)p(B) > 0 \longrightarrow$ No pueden ser independientes

**Libro** **2.4** p(fallo componente blanco) = 0,01 ; p( fallo componente oscuro) = 0,05

¿Cuál configuración es mejor?

a)



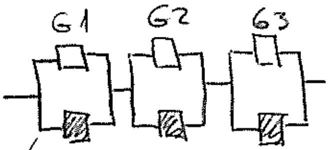
p(funciona linea 1) = p(funcionan los 3) = $(1-0{,}01)^3 = 0{,}970299$

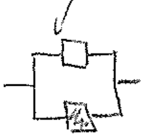
p(funcione Linea 2) = p(funcionan los 3) = $(1-0{,}05)^3 = 0{,}857375$

p(funcione maquina) = 1 − p(no funcione) = $1 - p(\overline{L1})\, p(\overline{L2}) = 1 - (1-0{,}970299)(1-0{,}857375) =$

$= 0{,}996 \longrightarrow$ p(averia) = $4{,}236 \cdot 10^{-3}$

b)





p(funcione G1) = 1 − p(no funcione) = 1 − (0,01)(0,05)

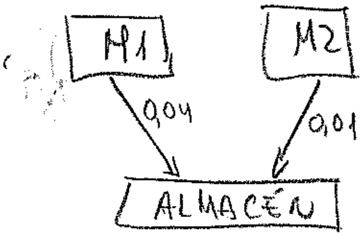
p(funcione maquina) = $p(G1 \cdot G2 \cdot G3) = (p(G1))^3 = (1-(0{,}01)(0{,}05))^3 = 0{,}9985$

$\longrightarrow$ p(avería) = $1{,}499 \cdot 10^{-3}$

$\Rightarrow$ Preferible configuración B

**Libro** **2.5** M1 → 0,04 defectuosas M2 → 0,01 defectuosas

Se toman 2 piezas al azar y resulten aceptables. Calcular p(las dos piezas sean de la maquina 1)



$$p(\text{coger defectuosa}) = \underbrace{p(D|M1)}_{0,04}\underbrace{p(M1)}_{\frac{1}{2}} + \underbrace{p(D|M2)}_{0,01}\cdot\underbrace{p(M2)}_{1/2} = 0,025$$

$$p(\text{aceptable}) = 0,975$$

$$p(M1|\text{aceptable}) = \frac{p(\text{aceptable}|M1)\cdot p(M1)}{p(\text{acept.}|M1)\cdot p(M1) + p(\text{acep}|M2)\,p(M2)} =$$

$$= \frac{(1-0,04)\cdot\frac{1}{2}}{(1-0,04)\frac{1}{2} + (1-0,01)\frac{1}{2}} = 0,4923 \longrightarrow$$ Si sacamos dos piezas

$$p(\text{las dos } M1) = 0,4923^2 = \boxed{0,2423}$$

**Libro** **2.7** Empresa legal chips defectuosos 5% y vende en paquetes de 5 unidades

Empresa ilegal " " 50% y vende en " " 5 "

10% chips vendidos son ilegales. ¿ p(un paquete con 2 chips defectuosos sea ilegal?

$$p(\text{2 chips mal}|EI) = \binom{5}{2}0,5^2\,0,5^3 = 10\cdot 0,5^5,$$

$$p(EI) = 0,1$$

$$p(\text{2 chip mal}|EL) = \binom{5}{2}0,05^2\,0,95^3$$

$$p(EL) = 0,9$$

En paquete ilegal

$\overline{C_1}\,\overline{C_2}\,C_3\,C_4\,C_5 \to 0,5^2\cdot 0,5^3$

$\overline{C_1}\,C_2\,\overline{C_3}\,C_4\,C_5$

$\overline{C_1}\,C_2\,C_3\,\overline{C_4}\,C_5$

$\overline{C_1}\,C_2\,C_3\,C_4\,\overline{C_5}$

$C_1\,\overline{C_2}\,\overline{C_3}\,C_4\,C_5$

$C_1\,\overline{C_2}\,C_3\,\overline{C_4}\,C_5$

$C_1\,\overline{C_2}\,C_3\,C_4\,\overline{C_5}$

$C_1\,C_2\,\overline{C_3}\,\overline{C_4}\,C_5$

$C_1\,C_2\,\overline{C_3}\,C_4\,\overline{C_5}$

$C_1\,C_2\,C_3\,\overline{C_4}\,\overline{C_5}$

$\Big\}\binom{5}{2}$

$$p\left(EI\middle|\substack{\text{2 chips}\\ \text{mal}}\right) = \frac{p\left(\substack{\text{2 chips}\\ \text{mal}}\middle|EI\right)\cdot p(EI)}{p\left(\substack{\text{2 chips}\\ \text{mal}}\middle|EI\right)p(EI) + p\left(\substack{\text{2 chips}\\ \text{mal}}\middle|EL\right)p(EL)}$$

$$p\left(EI\middle|\substack{\text{2 chips}\\ \text{mal}}\right) = \frac{10\cdot 0,5^5\cdot 0,1}{10\,0,5^5\cdot 0,1 + 10\cdot 0,5^2\cdot 0,95^3\cdot 0,9} = \boxed{0,618}$$

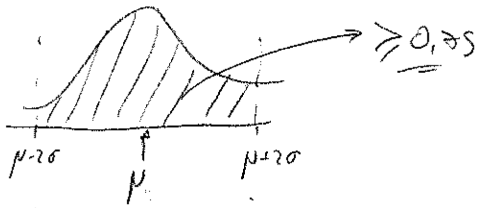
**Libro** 2.8 Entre media y dos desviaciones típicos se encuentra el 75% de la prob. total.
Justificar

Por Tchebychev sabemos $P(|X-\mu| \leq K\sigma) \geq 1-\frac{1}{K^2}$; $1-\frac{1}{K^2} = 0,75$, $K = \sqrt{\frac{1}{0,25}} = 2$

$$K=2 \longrightarrow P(|X-\mu| \leq 2\sigma) \geq 0,75$$

$$\longrightarrow |X-\mu| \leq 2\sigma, \quad \begin{matrix} X-\mu \leq 2\sigma \to X \leq \mu+2\sigma \\ X-\mu \geq -2\sigma \quad X \geq \mu-2\sigma \end{matrix}$$

Por tanto $P(\mu-2\sigma \leq X \leq \mu+2\sigma) \geq 0,75$



**Libro** 2.12 Máquina 1 $\rightarrow \mu_1$ , $\sigma_1$ 60% total componentes
Máquina 2 $\rightarrow \mu_2$ , $\sigma_2$ 40% total componentes } Calcular la $\mu_{TOTAL}$ y

TOTAL = 0,6 · piezas M1 + 0.4 piezas M2 $\longrightarrow$ Transf. lineal

$$E[TOTAL] = 0,6\, E[M1] + 0.4\, E[M2] = \boxed{0,6\,\mu_1 + 0,4\,\mu_2}$$

$$(M1\ M2)\begin{pmatrix}0,6\\0,4\end{pmatrix} = 0,6M1 + 0,4M2 = Tot.$$

Tenemos $F_{TOTAL}(x) = \underbrace{P(X=x)}_{F_{M1}}\, p(M1) + \underbrace{P(X=x)}_{F_{M2}}\, p(M2)$

$$F_{TOTAL}(x) = 0,6\, F_{M_1}(x) + 0,4\, F_{M_2}(x)$$

$$f_{TOT}(x) = 0,6\, f_{M_1}(x) + 0,4\, f_{M_2}(x)$$

$$E[X_{TOT}] = \int_{-\infty}^{\infty} x\, f_{TOT}(x)\,dx = \underbrace{\int_{-\infty}^{\infty} 0,6\,x\, f_{M_1}(x)\,dx}_{0,6\,E_{M_1}[X]} + \underbrace{\int_{-\infty}^{\infty} 0,4x\, f_{M_2}(x)\,dx}_{0,4\,E_{M_2}[X]} =$$

$$= \boxed{0,6\,\mu_1 + 0,4\,\mu_2}$$

$$Var[X_{TOT}] = E[X_{TOT}^2] - (E[X_{TOT}])^2$$

$$E[X_{TOT}^2] = \int_{-\infty}^{\infty} x^2 f_{TOT}(x)\,dx = \int_{-\infty}^{\infty} x^2\, 0,6\, f_{M_1}(x)\,dx + \int_{-\infty}^{\infty} x^2\, 0,4\, f_{M_2}(x)\,dx =$$

$$= 0,6 \int_{-\infty}^{\infty} x^2 f_{M_1}(x)\,dx + 0,4\int x^2 f_{M_2}(x)\,dx = 0,6\, E_{M_1}[x^2] + 0,4\, E_{M_2}[x^2] =$$

Como $E_{M_1}[X^2] = Var_{M_1}[X] + \mu_1^2 = \sigma_1^2 + \mu_1^2$ como $E_{M_2}[X^2] = Var_{M_2}[X] + \mu_2^2$ $= 0,6(\sigma_1^2 + \mu_1^2) + 0,4(\sigma_2^2 + \mu_2^2)$

$Var_{M_1}[X] = E(x^2) - \mu_1^2$

Sustituimos $Var[X_{TOT}] = 0,6(\sigma_1^2 + \cancel{\mu_1^2}) + 0,4(\sigma_2^2 + \cancel{\mu_2^2}) - (0,6\,\mu_1)^2 - (0,4\,\mu_2)^2 - 0,48\,\mu_1\mu_2$

$$= 0,6\,\sigma_1^2 + 0,4\,\sigma_2^2 + \mu_1^2(0,6 - 0,6^2) + \mu_2^2(0,4 - 0,4^2) - 0,48\,\mu_1\cdot\mu_2 =$$

$$= 0,6\,\sigma_1^2 + 0,4\,\sigma_2^2 + 0,24\,\mu_1^2 + 0,24\,\mu_2^2 - 0,48\,\mu_1\mu_2 = 0,6\,\sigma_1^2 + 0,4\,\sigma_2^2 + 0,24(\mu_1^2 + \mu_2^2 - 2\mu_1\mu_2) =$$

$$= \boxed{0,6\,\sigma_1^2 + 0,4\,\sigma_2^2 + 0,24\,(\mu_1 - \mu_2)^2}$$

**Libro 2.15** v.a. $X$ $f(x)=\begin{cases} 12X^2(1-X) & 0<x<1 \\ 0 & \text{en el resto}\end{cases}$ Calcular $CV=\frac{\sigma}{\mu}$

$$\mu=\int_{-\infty}^{\infty} x\, f_x(x)=\int_0^1 x\cdot 12x^2(1-x)\,dx=\int_0^1 (12x^3-12x^4)dx=\left(12\frac{x^4}{4}-12\frac{x^5}{5}\right)\Big|_0^1=$$

$$=3-2{,}4=\boxed{0{,}6}$$

$$\sigma^2=\int_{-\infty}^{\infty}(X-\mu)^2 f_x(x)\,dx=\int_0^1 (X-0{,}6)^2\,12X^2(1-X)\,dx=\boxed{\frac{1}{25}}$$

**Libro 2.17** $P(X\leq 0{,}29)=0{,}75$ $X$ v.a. continua $(0,1)$; $Y=1-X$

$P(Y\leq K)=0{,}25$ ¿K?

$P(X\leq x)=F_x(x)$

$$P(Y\leq K)=P(1-X\leq K)=P(-X\leq K-1)=p(X\geq 1-K)=1-\underbrace{p(X\leq 1-K)}_{F_x(1-K)}$$

$$1-\underbrace{p(X\leq 1-K)}_{F_x(1-K)}=0{,}25,\quad F_x(1-K)=0{,}75 \longrightarrow$$

Dato → $P(X\leq 0{,}29)=F_x(0{,}29)=0{,}75$

$$1-K=0{,}29;\quad \boxed{K=0{,}71}$$

**Libro 2.20** $X$ v.a. $f_x(X)=\begin{cases} kX & 0<x<6 \\ 0 & \text{en el resto}\end{cases}$

a) Calcular $f_y(y)$ si $Y=1/X$ $\begin{cases} y=\frac{1}{x}=g(x) \\ x=\frac{1}{y}=g^{-1}(y)\end{cases}$ $y\in\left(\frac{1}{6},\infty\right)$

$$f_y(y)=\left|\frac{d}{dy}(g^{-1}(y))\right| f_x(g^{-1}(y))=\left|\frac{-1}{y^2}\right| K/y=\frac{+K}{y^3}=\frac{1}{18y^3}$$

$$\int_0^6 f_x(x)dx=1;\quad \int_0^6 kx\,dx=k\frac{x^2}{2}\Big|_0^6=K\,18=1;\quad K=\frac{1}{18}$$

$$f_y(y)=\begin{cases}\frac{1}{18y^3} & y\in\left(\frac{1}{6},\infty\right)\\ 0 & \text{en el resto}\end{cases}$$

b)
$$E[Y]=\int_{-\infty}^{\infty} y\, f_y(y)\,dy=\int_{1/6}^{\infty} y\frac{1}{18y^3}dy=\int_{1/6}^{\infty}\frac{1}{18y^2}dy=\frac{y^{-1}}{-18}\Big|_{1/6}^{\infty}=$$

$$=\lim_{y\to\infty}\underbrace{\cancel{\left(\frac{-1}{18y}\right)}}_{0}-\left(\frac{1}{-18\,{}^1/_6}\right)=\boxed{\frac{1}{3}}$$

**Libro 2.21** X v.a. $f_X(X)=\begin{cases} \frac{2x}{\Theta^2} & \text{si } 0<x<\Theta \\ 0 & \text{en el resto} \end{cases}$ Obtener la función de densidad y la mediana de $Y=4X+5$

$Y=g(X)=4X+5$ ; $g^{-1}(y)=\frac{y-5}{4}$ $\quad 5<Y<4\Theta+5$

$$f_Y(y)=\left|\frac{d}{dy}(g^{-1}(y))\right| f_X(g^{-1}(y))=\frac{d}{dy}\left(\frac{y-5}{4}\right)\cdot\frac{2}{\Theta^2}\left(\frac{y-5}{4}\right)=\frac{1}{4}\,\frac{2}{\Theta^2}\,\frac{y-5}{4}=\frac{1}{8\Theta^2}(y-5)$$

$$f_Y(y)=\begin{cases} \frac{1}{8\Theta^2}(y-5) & 5<y<4\Theta+5 \\ 0 & \text{en el resto} \end{cases}$$

Mediana Y es el valor que verifica $\underbrace{P(Y\leq m_y)}_{F_Y(m_y)}=\frac{1}{2}$

$$F_Y(y)=\int_5^y\frac{1}{8\Theta^2}(y-5)\,dy=\frac{1}{8\Theta^2}\left(\frac{y^2}{2}-5y\right)\Big|_5^y=\frac{1}{8\Theta^2}\left[\frac{y^2}{2}-5y-\frac{5^2}{2}+25\right]$$

$$F_Y(m)=\frac{1}{2}\longrightarrow\ \frac{1}{{}_4\cancel{8}\Theta^2}\left(\frac{m^2}{2}-5m+12{,}5\right)=\frac{1}{2}$$

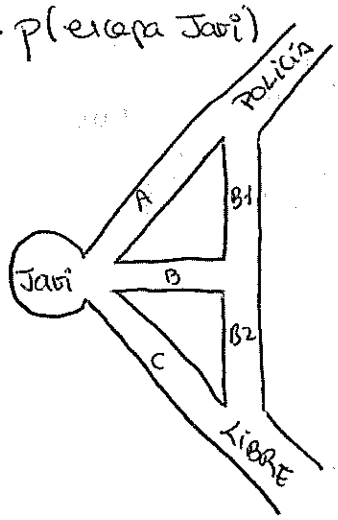
$$\frac{m^2}{2}-5m+12{,}5=4\Theta^2,\quad m^2-10m+25-8\Theta^2=0$$

$$m=\frac{10\pm\sqrt{100-4(25-8\Theta^2)}}{2}=5\pm\sqrt{25-25+8\Theta^2}\overset{m\in(5,4\Theta+5)}{=}5+2\sqrt{2}\,\Theta\longrightarrow\boxed{M_Y=5+2\sqrt{2}\,\Theta}$$

~~**Ayrée 1** 800 habitaciones. 15% reservan cancelados $\longrightarrow$ p(cancelación) = 0,15~~
~~p(overbooking) = 0,99 ¿Cuantas reservas debe fijar el hotel?~~
~~X v.a. nº de clientes que acuden al hotel con reserva~~
~~N reservan~~
~~$p(X\leq 500)=0{,}99$ } se pide N~~
~~$X=N\cdot 0'85$~~

Muy difíciles!

⇨ p(escapa Javi)



$$p(\text{escapar})=p(C)+p(B2|B)\,p(B)=\frac{1}{3}+\frac{1}{2}\,\frac{1}{3}=\frac{2+1}{6}=\boxed{\frac{1}{2}}$$

$$p(\text{escapar}|B)=p(\text{escapar}|B1)\,p(B1)+p(\text{escapar}|B2)\,p(B2)=$$
$$=0\cdot\frac{1}{2}+1\cdot\frac{1}{2}=\boxed{\frac{1}{2}}$$

$p(\text{escapar}|C)=\boxed{1}$

$p(\text{escapar}|A)=\boxed{0}$

# FÓRMULAS RESUMEN: $1^{er}$ Parcial

## TEMA 1

## UNIVARIANTE

- Media $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

- Varianza $S_x^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ ; • Desviación típica $S_x = \sqrt{S_x^2} = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$

- Cuartiles → $Q_2 = \text{Mediana} = \begin{cases} X_{(p)} & \text{si } n \text{ es impar } p = \frac{n+1}{2} \\ \dfrac{X_{(p)} + X_{(p+1)}}{2} & \text{si } n \text{ es par } p = \frac{n}{2} \end{cases}$

  $RI = Q_3 - Q_1$

  → $Q_1 = X_{(r)}$ con $r = \left[\dfrac{p+1}{2}\right]$

  → $Q_3 = X_{(s)}$ con $s = n - r + 1$

- Límites → $LI = Q_1 - 1{,}5\,RI$

  → $LS = Q_3 + 1{,}5\,RI$

- Momentos → Respecto al origen $a_k = \dfrac{\sum x_i^k}{n}$

  → Respecto a la media $m_k = \dfrac{\sum (x_i - \bar{x})^k}{n}$

- Coeficientes → $C_{AS} = \dfrac{m_3}{S^3}$ En modelo ideal $\begin{cases} C_{AS} = 0 \\ C_{AP} = 3 \end{cases}$

  → $C_{AP} = \dfrac{m_4}{S^4}$

- Tchebychev

$$f_{rel}\left(|x_i - \bar{x}| \leq KS\right) \geq 1 - \frac{1}{K^2}$$

## MULTIVARIANTE

- Covarianza $S_{xy} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$ • Mat. varianzas $S^2 = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{yx} & S_y^2 \end{pmatrix}$

- Correlación $r_{xy} = \dfrac{S_{xy}}{S_x S_y}$ $[-1, 1]$ • Mat. correlaciones $\begin{pmatrix} 1 & r_{xy} \\ r_{xy} & 1 \end{pmatrix}$

  Coef. correlación $= r_{xy}^2$ $\geq 0{,}5$ → alto

# TRANSFORMACIONES UNIVARIANTES

- **Lineales** $y = aX + b \longrightarrow \bar{y} = a\bar{x} + b$
  $\searrow S_y^2 = a^2 S_x^2$

- **No lineales**
  - Para $C_{AS} > 0 \longrightarrow y_i = \log x_i \,,\; y_i = \frac{1}{x_i} \,,\; y_i = \sqrt{x_i}$
  - Para $C_{AS} < 0 \longrightarrow y_i = x_i^2$
  - Aproximaciones
    $$\bar{y} = h(\bar{y}) + \frac{1}{2} h''(\bar{y}) \cdot S_x^2$$
    $$S_y^2 = \left(h'(\bar{y})\right)^2 S_x^2$$
  - Estandarización
    $$z_i = \frac{x_i - \bar{x}}{S_x} \quad \begin{cases} \bar{z} = 0 \\ S_z = 1 \end{cases}$$
  - Box-Cox
    $$y_i = \frac{x_i^p - 1}{p} \quad \forall p \neq 0$$
    $$y_i = \log x_i \quad p = 0$$

# TRANSFORMACIONES MULTIVARIANTES → Solo lineales

- **Vectorial**

$$y = a_1 X_{1i} + a_2 X_{2i} + \_\_\_\_ + a_K x_{iK} \longrightarrow y_i = (a_1 a_2 \dots a_k) \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Ki} \end{pmatrix} = a^T X_i$$

$$y_i = a^T X_i \qquad \bar{y} = a^T \bar{X}$$
$$S_y^2 = a^T S_x^2 a$$

- **Matricial**

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{mi} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mK} \end{pmatrix} \begin{pmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{Ki} \end{pmatrix} \Rightarrow y = AX_i$$

$$\bar{y} = A\bar{X}$$
$$S_y^2 = A S_x^2 A^T$$

# OTRAS FÓRMULAS

- $S_x^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$

- $\mathrm{Cov}(x,y) = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y}$

# Modelos Univariantes de Probabilidad

## 1 PROCESO DE BERNOUILLI

Se habla de procesos de Bernouilli cuando el resultado del experimento realizado admite dos categorías: "Aceptable" y "Defectuoso".

- Se repite el experimento n veces
- Siendo p la probabilidad de defectuoso, p es constante
- Los experimentos son independientes

La variable de Bernouilli es discreta. Ej.: Lanzamiento de un dado n veces clasificando el resultado como 6 o distinto de 6.

## 2 DISTRIBUCIÓN BINOMIAL

Sea X una variable aleatoria con distribución binomial:

X = nº de elementos defectuosos al observar n"

El experimento habrá de hacerse con reposición o con un $N \gg n$ $\left(\frac{N}{n} > 10\right)$

$\boxed{X \rightarrow B(n, p)}$ con n: nº elementos observados
p: prob. de defectuoso

| X | P(X=x) |
|---|---|
| 0 | $\binom{n}{0} p^0 (p-1)^n$ |
| 1 | $\binom{n}{1} p^1 (p-1)^{n-1}$ |
| 2 | $\binom{n}{2} p^2 (p-1)^{n-2}$ |
| ⋮ | ⋮ |
| n | $\binom{n}{n} p^n$ |

$$\Longrightarrow \boxed{p(X=K) = \binom{n}{K} p^K (1-p)^{n-K}} \qquad K = 0, 1, 2, \ldots n$$

$$\text{Obviamente} \begin{cases} p(X=K) \geq 0 \\ \sum_{K=0}^{n} \binom{n}{K} p^K (1-p)^{n-K} = 1 \end{cases}$$

Propiedades $E[X] = \sum_{K=0}^{n} K \cdot p(X=K) = np$ , $\boxed{E[X] = np}$

$Var[X] = E[X^2] - E[X]^2 = np(1-p)$ $\boxed{Var[X] = np(1-p)}$

Cuando p = 0,5 la distribución binomial es simétrica

## Ejemplos

① N° caras al lanzar 100 monedas $B(n=100,\ p=0{,}5)$

¿Probabilidad de que salgan 10 caras?

$$p(X=10)=\binom{100}{10}p^{10}(1-p)^{100-10}\ \underset{\substack{\uparrow\\ p=\frac{1}{2}}}{=}\ \binom{100}{10}0{,}5^{10}(1-0{,}5)^{90}$$

¿Probabilidad de que salgan al menos 60 caras?

$$p(X\geq 60)=\sum_{x=60}^{100}\binom{100}{x}p^{x}(1-p)^{100-x}\ \underset{\substack{\uparrow\\ p=\frac{1}{2}}}{=}\ \sum_{x=60}^{100}\binom{100}{x}0{,}5^{x}(1-0{,}5)^{100-x}$$

¿Probabilidad de que salgan máximo 40 caras?

$$p(X\leq 40)=\sum_{x=0}^{40}\binom{100}{x}p^{x}(1-p)^{100-x}\ \underset{\substack{\uparrow\\ p=\frac{1}{2}}}{=}\ \sum_{x=0}^{40}0{,}5^{x}(1-0{,}5)^{100-x}$$

② Un contrato estipula la compra de componentes en lotes grandes que deben contener un máximo de 10 % de piezas con algún defecto. Para comprobar la calidad se toman 11 unidades y se acepta el lote si como máximo hay dos piezas defectuosas. ¿Es un buen procedimiento de control?

$X$ v.a n° piezas defectuosas $X\to B(n=11,\ p)$

$$p(\text{aceptar lote})=p(X\leq 2)=\binom{11}{0}p^{0}(1-p)^{11}+\binom{11}{1}p^{1}(1-p)^{10}+\binom{11}{2}p^{2}(1-p)^{9}$$

| $p$: | 5% | 10% | 15% | 20% | 25% |
|---|---|---|---|---|---|
| p(aceptar) | 0,985 | 0,910 | 0,779 | 0,617 | 0,455 |

## 3 DISTRIBUCIÓN GEOMÉTRICA

Sea $Y$ una v.a. de distribución de probabilidad geométrica

$Y$: n° elementos hasta que aparezca 1 defectuoso

| | $Y$ | $P(Y=y)$ |
|---|---|---|
| ● | $1 \longrightarrow$ | $p$ |
| ○● | $2 \longrightarrow$ | $(1-p)p$ |
| ○○● | $3 \longrightarrow$ | $(1-p)^2p$ |
| ⋮ | ⋮ | ⋮ |
| ○○○…● | $K \longrightarrow$ | $(1-p)^{K-1}p$ |

$$\boxed{p(Y=K)=(1-p)^{K-1}p}$$

Propiedades $\boxed{E[Y]=\frac{1}{p}}$ $\boxed{Var[Y]=\frac{1-p}{p^{2}}}$

# 4 DISTRIBUCIÓN DE POISSON

Definiremos esta distribución con un ejemplo. Sea la fabricación continua de conductor de cobre y $\lambda$ el nº medio de defectos cada 100 m.

$X$ = nº defectos en un tramo de 100 m

Por tanto, una distribución de poisson es una binomial con $n \to \infty$

$$P_x(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad p = \frac{\lambda}{n}$$

$$P_x(x) = \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!}\lim_{n\to\infty}\underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{x}}_{\to 1} =$$

$$= \frac{\lambda^x}{x!}e^{-\lambda} \quad ; \quad x = 0,1,2,\ldots$$

$$\boxed{P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda} \quad ; \quad x=0,1,2,\ldots} \qquad \boxed{E[X]=\lambda \quad Var[X]=\lambda}$$

⊛ Para $\lambda > 5$ la distribución de Poisson tiende a ser simétrica.

⇒ Ejemplo: Una fuente radiactiva emite partículas según la distribución de poisson de media 10 partículas/minuto. Se desea calcular:

- Probabilidad de 5 partículas en 1 minuto.

$\lambda = 10$ part/min

$X$ = nº part. en 1 min

$$P(X=5) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{10^5}{5!}e^{-10} = \boxed{0{,}0378}$$

- Probabilidad de 0 partículas en 1 minuto.

$$P(X=0) = \frac{10^0}{0!}e^{-10} = \boxed{4{,}54 \cdot 10^{-5}}$$

- Probabilidad de más de 5 partículas en 1 minuto.

$$P(X>5) = 1 - P(X\leq 5) = 1 - \sum_{x=0}^{5}\frac{10^x}{x!}e^{-10} = 1 - e^{-10}\left(\frac{10^0}{0!}+\frac{10^1}{1!}+\frac{10^2}{2!}+\frac{10^3}{3!}+\frac{10^4}{4!}+\frac{10^5}{5!}\right) =$$

$$= \boxed{0{,}933}$$

- Probabilidad de 30 o menos partículas en 5 minutos

$Y$ = nº partículas en 5 mins $\longrightarrow \lambda' = 5\lambda = 50$ part/5 mins

$$P(Y\leq 30) = \sum_{y=0}^{30}\frac{50^y}{y!}e^{-50} = e^{-50}\sum_{0}^{30}\frac{50^y}{y!} = \boxed{0{,}0016}$$

## 5 DISTRIBUCIÓN EXPONENCIAL

Tomando el mismo ejemplo, de la fabricación continua de conductor de cobre con $\lambda \equiv$ nº medio de defectos cada 100 m y

$T \equiv$ distancia entre dos defectos consecutivos → v.a. exponencial

$$p(T \geq t) = \overset{p(0 \text{ defectos en el intervalo } [0,t])}{p(X=0 \mid X \to \text{Poisson}(\lambda t))} = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

$$\boxed{\begin{array}{ll} p(T \geqslant t) = e^{-\lambda t} & t \geqslant 0 \\ F_T(t) = p(T \leqslant t) = 1 - e^{-\lambda t} & t \geqslant 0 \\ f_T(t) = \lambda e^{-\lambda t}; & t \geqslant 0 \end{array}}$$

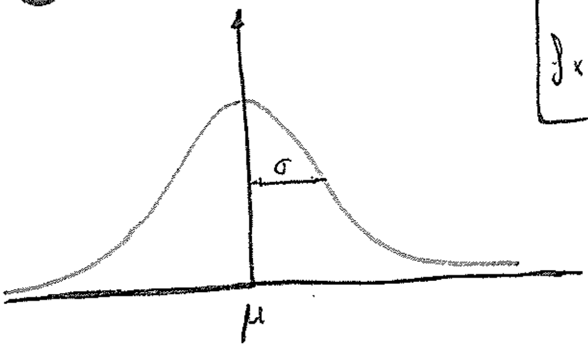
Propiedades $E[T] = \int_{-\infty}^{\infty} t\, f_T(t)\, dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \lambda \left[ \cancelto{0}{t \frac{e^{-\lambda t}}{-\lambda}} \Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda t}}{\lambda} dt \right] =$

$$= \lambda \frac{1}{\lambda} e^{-\lambda t} \frac{1}{-\lambda} \Big|_0^{\infty} = 0 - \left(\frac{1}{-\lambda}\right) = \frac{1}{\lambda} \to \boxed{E[T] = \frac{1}{\lambda}}$$

$$Var[T] = E[T^2] - (E[T])^2 = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = \lambda \left( \cancelto{0}{t^2 \frac{e^{-\lambda t}}{-\lambda}} \Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda t}}{\lambda} 2t\, dt \right) -$$

$$- \frac{1}{\lambda^2} = \underbrace{\lambda \left( \frac{2}{\lambda} \int_0^{\infty} e^{-\lambda t}\, t\, dt \right)}_{E[T]} - \frac{1}{\lambda^2} = \frac{2}{\lambda} \cdot \frac{1}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \qquad \boxed{Var[T] = \frac{1}{\lambda^2}}$$

## 6 DISTRIBUCIÓN NORMAL. Campana de Gauss



$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \; ; \quad x \in \mathbb{R}$$

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

Propiedades $E[X] = \mu$

$E[(X-\mu)^3] = 0$ (todos los momentos de orden impar son nulos)

$CA_s = \frac{E[(X-\mu)^3]}{\sigma^3} = \underline{\underline{0}}$

$Var[X] = \sigma^2$

$E[(X-\mu)^4] = 3\sigma^4$

$CA_p = \frac{E[(X-\mu)^4]}{\sigma^4} = \underline{\underline{3}}$

Los valores de una distribución normal podremos mirarlos en tablas estandarizando la variable

NORMAL ESTANDAR: $Z \to N(0,1)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \ \Big\{ \text{Valor de tablas}$$

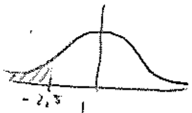
ESTANDARIZACIÓN

$$X \to N(\mu,\sigma) \ ; \quad Z = \frac{X-\mu}{\sigma} \to N(0,1)$$

$$p(X<a \mid X \to N(\mu,\sigma)) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = P\left(Z \le \frac{a-\mu}{\sigma}\right) = \phi\left(\frac{a-\mu}{\sigma}\right)$$

⇨ Ejemplo: La longitud X de ciertos tornillos es una v.a. con distribución normal de media 30mm y desv. típica 0,2 mm. Se aceptan como válidos los tornillos que cumplen 29,5 < X < 30,4

- Proporción de tornillos no aceptables por cortos.

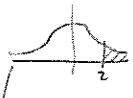


$$P(X \le 29,5) = p\left(\frac{X-30}{0,2} \le \frac{29,5-30}{0,2}\right) = p(Z \le -2,5) \overset{\downarrow}{=} p(Z \ge 2,5) =$$

$$= 1 - \underbrace{p(Z \le 2,5)}_{\phi(2,5)} = 1 - 0,9938 = 6,2 \cdot 10^{-3}$$

$\boxed{\text{0,62 \% de los tornillos no serán aceptables por cortos}}$

- Proporción de tornillos no aceptables por largos



$$P(X \ge 30,4) = p\left(\frac{X-30}{0,2} \ge \frac{30,4-30}{0,2}\right) = p(Z \ge 2) \overset{\downarrow}{=} 1 - \Phi(2) = 1 - 0,9772 =$$

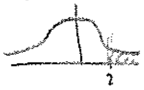
$$= 0,0228$$

$\boxed{\text{2,28 \% de los tornillos no serán aceptables por largos}}$

- Proporción de tornillos válidos.

$$P(29,5 < X < 30,4) = p\left(\frac{29,5-30}{0,2} < \frac{X-30}{0,2} < \frac{30,4-30}{0,2}\right) = p(-2,5 < Z < 2) \overset{\downarrow}{=}$$

$$= p(Z > -2,5) - p(Z > 2) = p(Z < 2,5) - (1 - p(Z<2)) = \Phi(2,5) - 1 + \Phi(2) =$$



$$= 0,9938 - 1 + 0,9772 = 0,971$$

$\boxed{\text{97,1 \% de los tornillos serán válidos}}$

# 7 EJERCICIOS

③.1 Llamadas a una centralita siguen distribución Poisson $\lambda = 3 \text{ llam}/5 \text{ min}$

Calcular la probabilidad de:

a) Seis llamadas en 5 mins.

X v.a. llamadas a la centralita cada 5 minutos. $\Big\{ X \to P(\lambda = 3 \text{ llamadas}/5 \text{ min})$

$$P(X=6) = \frac{\lambda^x}{X!} e^{-\lambda} = \frac{3^6}{6!} e^{-3} = \boxed{0{,}05}$$

b) Tres llamadas en 10 mins.

Y v.a. llamadas a la centralita cada 10 min $\Big\{ Y \to P(\lambda' = 2\lambda = 6 \text{ llam}/10 \text{ min})$

$$P(Y=3) = \frac{6^3}{3!} e^{-6} = \boxed{0{,}089}$$

c) Más de 15 en un cuarto de hora

W v.a. llamadas a la centralita cada 15 mins $\Big\{ W \to P(\lambda'' = 3\lambda = 9 \text{ llam}/15 \text{ mins})$

$$P(W>15) = 1 - P(W \leq 15) = 1 - \sum_{K=0}^{15} e^{-9} \frac{9^K}{K!} =$$

$$= \boxed{1 - e^{-9}\left(1 + 9 + \frac{9^2}{2!} + \frac{9^3}{3!} + \cdots + \frac{9^{15}}{15!}\right)}$$

d) Dos en 1 minuto

V v.a. llamadas a la centralita cada minuto $\Big\{ V \to P(\lambda''' = \frac{\lambda}{5} = \frac{3}{5} \text{ llam}/\text{min})$

$$P(V=2) = \frac{\left(\frac{3}{5}\right)^2}{2!} e^{-3/5} = \boxed{0{,}099}$$

③.2 X distribución exponencial con media 1. Obtener la función de distribución y la de densidad de $W = aX^{1/b}$ $a>0$, $b>0$.

$f_x(X) = e^{-x}$

$F_x(X) = 1 - e^{-x}$

$$F_W(w) = P(W \leq w) = P(aX^{1/b} \leq w) = P\left(X \leq \left(\frac{w}{a}\right)^b\right) = F_x\left(X = \left(\frac{w}{a}\right)^b\right) =$$

$$= 1 - e^{-\left(\frac{w}{a}\right)^b}$$

$$\boxed{F_W(w) = 1 - e^{-\left(\frac{w}{a}\right)^b}} \to \begin{matrix} w \geq 0 \\ a>0 \quad b>0 \end{matrix}$$

$$\boxed{f_w(w) = \frac{d\,F_w(w)}{dw} = -e^{-\left(\frac{w}{a}\right)^b} \cdot \left(-\left(\frac{w}{a}\right)^{b-1} b \cdot \left(\frac{1}{a}\right)\right) = e^{-\left(\frac{w}{a}\right)^b} \frac{b}{a^b}\, w^{b-1}}$$

$w \geq 0$
$a > 0$
$b > 0$

**NOTA!!** Se define la tasa de fallo como $\boxed{\lambda_w = \frac{f_w(w)}{1-F_w(w)}}$

En nuestro caso → $$\boxed{\lambda(w) = \frac{\cancel{e^{-\left(\frac{w}{a}\right)^b}}\ \frac{b}{a^b}\, w^{b-1}}{\cancel{1}-\cancel{1}+\cancel{e^{-\left(\frac{w}{a}\right)^b}}} = \frac{b}{a^b}\, w^{b-1}}$$

③③ El nº averías diarias de una máquina sigue una distribución de Poisson de media 0,4 averías. Calcular la probabilidad de que halla 3 días consecutivos sin avería.

X v.a. nº averías por día    $X \to P(\lambda = 0{,}4$ averías/día$)$

Y v.a. nº averías 3 días    $Y \to P(\lambda' = 3\lambda = 1{,}2$ averías/3 días$)$

$$P(Y=0) = e^{-1,2}\,\frac{1{,}2^0}{0!} = \boxed{e^{-1,2}}$$

Otra forma $$\left(P(X=0 \mid X \to \text{Poisson}(\lambda = 0{,}4\ \text{av/día}))\right)^3 = \left(e^{-0,4}\,\frac{0{,}4^0}{0!}\right)^3 = \boxed{e^{-1,2}}$$

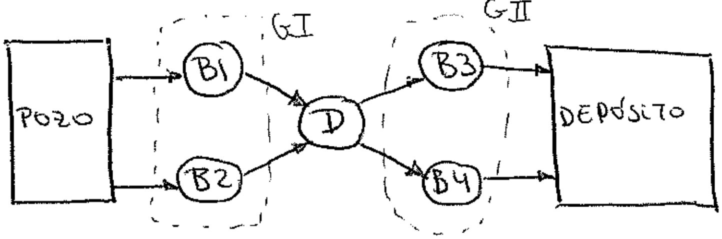
③.④ A un puerto llegan de media 10 clientes/h de manera indep. Calcular la prob. de que lleguen 8 clientes en la prox. media hora, sabiendo que en la última hora llegaron 14 clientes y que la v. a. numero de clientes que llegan en 1h sigue distribución de Poisson.

X: v.a nº clientes que llegan en 1h :    $X \to \text{Poisson}(\lambda = 10$ clientes/h$)$

Y: v.a. nº clientes " " " $\frac{1}{2}$ h.    $Y \to \text{Poisson}\left(\lambda' = \frac{1}{2}\lambda = 5 \text{ clientes}/\tfrac{1}{2}\text{h}\right)$

$$P(Y=8 \mid X=14) \underset{\substack{\uparrow \\ \text{por indep}}}{=} P(Y=8) = \boxed{e^{-5}\,\frac{5^8}{8!}}$$

3.5 Los tiempos de vida de la depuradora y las bombas son variables aleatorias independientes con distribución exponencial, siendo 20000 h la vida media de la depuradora y 30000 h la de las bombas.



a) Calcular la probabilidad de que llegue agua al depósito después de 20000 h de funcionamiento.

$$E[T_B] = 30000 = \frac{1}{\lambda_B} \longrightarrow f_{T_B}(T_B) = \frac{1}{30000} e^{-\frac{t}{30000}}, \quad T_B \geq 0$$

$$E[T_D] = 20000 = \frac{1}{\lambda_D} \longrightarrow f_{T_D}(T_D) = \frac{1}{20000} e^{-\frac{t}{20000}} \quad T_D \geq 0$$

$T_B \equiv$ v.a exponencial tiempo de vida de una bomba

$T_D \equiv$ v.a " " " " " la depuradora

$$p(\text{llegue agua al depósito tras } 20000h) = p(GI) \cdot p(D) \, p(GII)$$

$$p(GI) = 1 - p(\overline{GI}) = 1 - p(\overline{B1}) p(\overline{B2}) = p(GII) \Longrightarrow$$

$$\left\{\begin{array}{l} f_{T_B} = \frac{1}{30000} e^{-\frac{t}{30000}} \longrightarrow F_{T_B}(T_B < t) = \int_0^t \frac{1}{30000} e^{-\frac{t}{30000}} = 1 - e^{-t/30000} \\ p(\overline{B1}) = p(T_B < 20000) = F_{T_B}(20000) = 1 - e^{-\frac{20000}{30000}} = 1 - e^{-2/3} = p(\overline{B2}) \end{array}\right\} \Longrightarrow$$

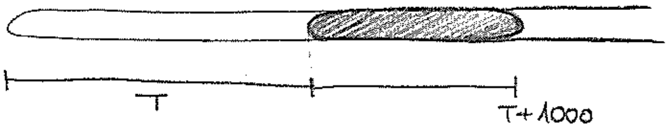
$$\Longrightarrow p(G1) = 1 - p(\overline{GI}) = 1 - (1 - e^{-2/3})^2 \, ; \text{ por tanto}$$

$$p(\text{llegar agua al dep tras } 20000h) = \left(1 - (1 - e^{-2/3})^2\right)^2 \cdot p(D) = \left(1 - (1 - e^{-2/3})^2\right)^2 e^{-1} =$$

$$f_{T_D} = \frac{1}{20000} e^{-\frac{t}{20000}}, \quad F_{T_D} = 1 - e^{-\frac{t}{20000}} \qquad = \boxed{0,214}$$

$$p(D) = p(T_D > 20000) = 1 - F_{T_D}(20000) = 1 - (1 - e^{-1}) = e^{-1}$$

b) Calcular la probabilidad de que una depuradora que ha trabajado T horas falle antes de las 1000 h siguientes. ¿Es razonable que, para evitar fallos de la depuradora se renueve esta cada 20000 h? ¿por qué?



$$p(T_D < T+1000 \mid T_D > T) = \frac{p(T_D < T+1000 \cdot T_D > T)}{p(T_D > T)} = \frac{p(T < T_D < T+1000)}{1 - F_{TD}(T)} =$$

$$= \frac{F_{TD}(T+1000) - F_{TD}(T)}{1 - F_{TD}(T)} = \frac{1 - e^{-\left(\frac{T+1000}{20000}\right)} - \left(1 - e^{-\frac{T}{20000}}\right)}{1 - \left(1 - e^{-\frac{T}{20000}}\right)} =$$

$$= \frac{e^{-T/20000}\left(1 - e^{-1/20}\right)}{e^{-T/20000}} = 1 - e^{-\frac{1000}{20000}} = F_{TD}(1000) = \underbrace{P(T_D < 1000)}_{\text{prob. de que la depuradora dure menos de 1000h}} \Rightarrow$$

$\Rightarrow$ No tiene sentido cambiar la depuradora transcurridas 20000h pues hay la misma probabilidad de que esta se estropee independientemente de cuantas horas lleve funcionando

NOTA// Las variables aleatorias con distribución exponencial no tienen memoria (ver 3.13)

3.6 1 de cada 100 personas porta un virus. Se realiza un análisis combinado que consiste en: Se toman las muestras de sangre de 50 en 50, se mezclan y se analiza la mezcla. Si el resultado del análisis es negativo, se concluye que los 50 individuos están sanos. Si el análisis es positivo se repite a cada persona individualmente. El análisis es infalible.

a) N° esperado de análisis que se realizarán siguiendo esta estrategia

$Y$ = v.a. n° análisis realizados $\longrightarrow$ se busca $E[Y]$

| $y$ | $P(Y=y)$ |
|---|---|
| 1 | p(mezcla negativa) |
| 51 | p(mezcla positiva) |

$X$ = n° individuos enfermos de un total de 50 $\equiv X \to B(n=50, p=0{,}01)$

$$p(Y=1) = p(X=0 \mid X \sim B(n=50, p=0{,}01)) = \binom{50}{0} p^0 (1-p)^{50} = 0{,}99^{50}$$

$$p(Y=51) = 1 - 0{,}99^{50}$$

$$E[Y] = 1 \cdot p(Y=1) + 51\, p(Y=51) = 1 \cdot 0{,}99^{50} + 51 \cdot (1 - 0{,}99^{50}) = 20{,}74 \simeq 21$$

El n° esperado de análisis es 21

b) ¿Cuál es la probabilidad de que un individuo sea portador del virus si el análisis realizado a su grupo de 50 ha resultado positivo?

X: individuo es portador ; ⊕: Análisis positivo

$$p(X|\oplus)=\frac{P(X\cdot\oplus)}{P(\oplus)}=\frac{P(\oplus|X)\cdot P(X)}{P(y=51)}=\frac{1\cdot\frac{1}{100}}{0,395}=\boxed{0,025}$$

③.7 Lote con proporción de defectuosas p se extraen piezas hasta observar la K-ésima defectuosa. Se pide la distribución de prob de v. a X: nº tot. de piezas observadas.

| $X$ | $P(X=x)$ |
|---|---|
| $K$ | $p^K$ |
| $K+1$ | $p^K(1-p)$ |
| $K+2$ | $p^K(1-p)^2$ |
| $\vdots$ | $\vdots$ |

$$\underset{\text{Observamos } n}{P}(X=n)=\binom{n-1}{K-1}p^{(K-1)}\,p^{(n-1)-(K-1)}\;p^K=\binom{n-1}{K-1}p^K\,p^{n-K}$$

$$\underbrace{\underbrace{\circ\circ\circ\ldots}_{(n-K)}\quad\bullet\ \text{Késima}}_{n}\Rightarrow \text{Pero! La Késima es la última pieza extraída}\quad \underbrace{\underbrace{\circ\circ\circ\circ\ldots}_{(n-1)\to(K-1)\text{ defectuosas}}\quad\overset{\text{Késima}}{\bullet}}_{n}$$

$$\boxed{P(X=n)=\binom{n-1}{K-1}p^K\,p^{n-K}}$$

③.8 La función de densidad de una v.a. X es:

$$f(x)=\begin{cases}x/8, & \text{si } 0\le x\le 4\\ 0, & \text{en el resto}\end{cases}$$

Se generan secuencialmente los valores de esta variable. ¿Cuantos valores de X habrá que generar por término medio para obtener un valor mayor que 3? v.a. geometrica (p)

Y: v.a. valores generados de X hasta observar uno mayor que 3

Se pide $E[Y]=\frac{1}{p}$ ; $P(X>3)=1-P(X\le 3)=1-F_x(3)=1-\frac{9}{16}=0,4375=\frac{7}{16}$

$$F_x(x)=\int_0^x\frac{x}{8}\,dx=\frac{x^2}{16}$$

Por ser Y una v.a. geométrica $E[Y]=\frac{1}{p}=\frac{16}{7}=2,29\rightarrow$ $\boxed{\text{Entre 2 y 3 valores}}$

3.9 Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcular la probabilidad de que tengan más de 4 hijos.

$p(\text{niño}) = p(\text{niña}) = 0{,}5$

| $X$ | $p(X=x)$ |
|---|---|
| 1 | $0{,}5$ |
| 2 | $(1-0{,}5)0{,}5$ |
| 3 | $(1-0{,}5)^2 0{,}5$ |
| 4 | $(1-0{,}5)^3 0{,}5$ |
| ⋮ | |
| $K$ | $(1-0{,}5)^{K-1} 0{,}5$ |

$X$ v.a. geométrica nº de hijos hasta nacimiento de una niña.

$X \to$ geométrica $(p=0{,}5)$

$$p(X>4) = 1 - p(X\leq 4) = 1 - \sum_{K=1}^{4} (1-p)^{K-1} \overset{p=0{,}5}{p} =$$

$$= 1 - \sum_{K=1}^{4} 0{,}5^K = 1 - (0{,}5 + 0{,}5^2 + 0{,}5^3 + 0{,}5^4) = \frac{1}{16} = \boxed{0{,}0625}$$

3.10 La distancia D entre dos vehículos sigue una distribución exponencial con media 200 metros. ¿Cuál es la probabilidad de que en un tramo de 1 Km haya exactamente 5 vehículos?

D: v. aleatoria exponencial distancia entre 2 vehículos

$E[D] = 200\text{ m} = \frac{1}{\lambda}$, $F_D(D) = 1 - e^{-\lambda d}$, $f_D(D) = \lambda e^{-\lambda d}$ con $\lambda = \frac{1}{200}$

$$\Rightarrow \begin{cases} F_D(D) = 1 - e^{-\frac{1}{200}d} \\ f_D(D) = \frac{1}{200} e^{-\frac{1}{200}d} \end{cases}$$

$X$: v. aleatoria nº coches en 1 Km ; $X \to$ Poisson $(\lambda = 5 \text{ coches}/1\text{ Km})$

Se pide $p(X=5) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{5^5}{5!} e^{-5} = \boxed{0{,}175}$

3.13 Obtener tasa de fallo de T, siendo T una v.a. exponencial de media 1000 h. Interpolar resultado.

Tasa de fallo $\left\{ \lambda(t) = \frac{f_T(T)}{1-F_T(T)} \right|$, Por ser T v.a. exponencial $E[T] = 1000 = \frac{1}{\lambda}$ $\left| \begin{array}{l} F_T(T) = 1 - e^{-\frac{1}{1000}t} \\ f_T(T) = \frac{1}{1000} e^{-\frac{1}{1000}t} \end{array} \right.$

$$\lambda(t) = \frac{\frac{1}{1000} e^{-\frac{1}{1000}t}}{1-(1-e^{-\frac{1}{1000}t})} = \boxed{\frac{1}{1000} = \lambda = \text{cte}} \Longrightarrow$$

Por tanto, las variables aleatorias con distribución exponencial no tienen memoria

**3.14** Un examen consiste en 25 cuestiones. En cada cuestión el alumno elige entre 5 soluciones propuestas sólo 1. El número mínimo de respuestas correctas para aprobar es a. Se fija "a" con el siguiente criterio: Prob de aprobar para un alumno que conteste al azar sea menor que 0,05. (Una cuestión es respondida al azar si cada uno de los resultados propuestos tiene la misma prob. de ser escogido).

$Y$ v.a. nº respuestas correctas respondiendo al azar de un total de 25

$$y \to B\left(n=25, p=\frac{1}{5}\right)$$

$$p(\text{aprobar contestando al azar}) = p\left(y \geq a \mid y \to B\left(25, \frac{1}{5}\right)\right) < 0{,}05 =$$

$$\underbrace{p(y \geq a) = 1 - p(y<a) = 1 - \sum_{k=0}^{k=a} \binom{n}{k} p^k (1-p)^{n-k} < 0{,}05}$$

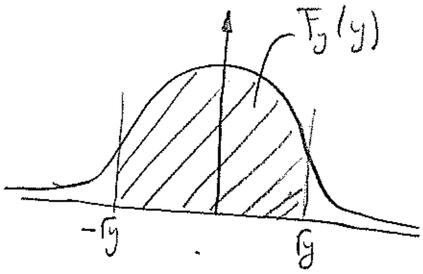
Se resolvería por aproximación a la normal que se verá en el tema siguiente!!

**3.15** Obtener la función de densidad de una v. a. $X^2$ si $X \to N(0,1)$

$Y = X^2$ $\quad X \to N(0,1)$, $F_x(x) = \Phi(x)$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) =$$

$$= \Phi(\overset{u}{\sqrt{y}}) - \Phi(-\sqrt{y})$$



$$f_y(y) = \frac{dF_y(y)}{dy} = \frac{d\Phi(u)}{du} \cdot \frac{1}{2\sqrt{y}} - \frac{d\Phi(-u)}{du} \cdot \left(-\frac{1}{2\sqrt{y}}\right) \overset{u=\sqrt{y}}{=}$$

$$= f_x(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_x(-\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} =$$

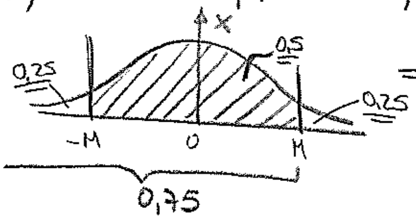
$$\left\{ \text{Como } X \to N(0,1) \to f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right\}$$

$$= \cancel{2} \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{\cancel{2}\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \longrightarrow \boxed{f_y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \quad y \geq 0}$$

**3.16** Dada una v.a $X \to N(0,\sigma^2)$, calcular la mediana de la variable $y = |X|$

$F_y(M) = 0{,}5 \longrightarrow p(y \leq M) = 0{,}5 \longrightarrow p(|X| \leq M) = 0{,}5$

$p(-M \leq X \leq M) = 0{,}5$



$\Longrightarrow p(X \leq M) = 0{,}75$

Estandarizando $p\left(\frac{X-0}{\sigma} \leq \underbrace{\frac{M-0}{\sigma}}_{z_0}\right) = 0{,}75$; $\phi(z_0) = 0{,}75 \xrightarrow{\text{Tablas}} z_0 = 0{,}68$

$\frac{M-0}{\sigma} = 0{,}68$; $\boxed{M = 0{,}68\,\sigma}$

**3.17** La longitud $L$ [mm] de longitud de piezas es una v.a $N(32, 0{,}3)$

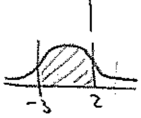
Se consideran aceptables las medidas entre $(31.1, 32.6)$

a) Calcular la probabilidad de que una pieza elegida al azar sea aceptable.

$L$ v.a $N(32, 0.3)$

$p(\text{aceptable}) = p(31{,}1 < L < 32{,}6) \overset{\text{estandarizamos}}{=} p\left(\frac{31{,}1-32}{0{,}3} < \overbrace{\frac{L-31{,}1}{0{,}3}}^{z \to N(0,1)} < \frac{32{,}6-32}{0{,}3}\right) =$

$= p(-3 < z < 2) = p(z<2) - p(z<-3) = p(z<2) - p(z>-3) =$



$= p(z<2) - (1 - p(z<3)) = \phi(2) - (1-\phi(3)) =$

$= 0{,}9772 - (1 - 0{,}9987) = \boxed{0{,}9759}$

b) Se toman al azar 3 piezas. ¿Cual es la prob. de que la 1ª y la 3ª sean aceptables y la segunda no lo sea?

$P(ADA) \underset{\uparrow \text{ independencia}}{=} p(\text{aceptable})^2 \cdot p(\text{defectuosa}) = p(\text{aceptable})^2 (1 - p(\text{aceptable})) = 0{,}9759^2 (1 - 0{,}9759) =$

$= \boxed{0{,}0229}$

c) En una muestra se toman 3 ¿prob. de que almenos 1 sea aceptable?

$X \to$ v.a nº piezas defectuosas de un total de 3 . $X \to B(n=3, p = p(\text{defectuosa}) = \underbrace{1 - 0{,}9759}_{0{,}0241})$

$P(\text{al menos 1 aceptable}) = 1 - p(\text{todos defectuosas}) = 1 - p(X=3) = 1 - \binom{n}{3} p^n (1-p)^{n-3} =$

$= 1 - \binom{3}{3}(1-0{,}9759)^3 = \boxed{0{,}999986}$

d) Las piezas se embalan en lotes de 500. Calcular la prob. de que un lote tenga máximo de 15 piezas defectuosas

$Y$: v.a nº piezas defectuosas en un total de 500 $\equiv y \rightarrow B(n=500, p=0,0241)$

$$p(y > 15 \mid y \rightarrow B(n=500, p=0,0241)) = 1 - p(y<15) = 1 - \underbrace{\sum_{y=0}^{15} \left( \binom{500}{y} (0,024)^{y} (1-0,024)^{500-y} \right)}$$

Este sumatorio es complicado de hacer por lo que aproximaríamos la v.a y a una normal. Se verá en el tema siguiente!

**3.25** Federer y Nadal están empatados, 40-40. Saca Nadal. p(Nadal gana punto) = 0,6

¿Juego lo gana Nadal?

$NN \rightarrow$ Gana Nadal (G) $0,6^2$
$NF$, $FN$ } Empate (E) $2 \cdot 0,6 \cdot 0,4$
$FF \rightarrow$ Gana Federer $0,4^2$

Gana Nadal (casos)
$G \longrightarrow 0,6^2$
$E\ G \longrightarrow (2 \cdot 0,6 \cdot 0,4) 0,6^2$
$E\ E\ G \longrightarrow (2\ 0,6 \cdot 0,4)^2 0,6^2$
$E\ E\ E\ G \longrightarrow (2\ 0,6 \cdot 0,4)^3\ 0,6^2$
$\vdots$

$$p(\text{gana Nadal}) = 0,6^2 + (2 \cdot 0,6 \cdot 0,4)0,6^2 + (2 \cdot 0,6 \cdot 0,4)^2 0,6^2 + \dots =$$

$$= 0,6^2 \underbrace{(1 + 0,6 \cdot 0,4 \cdot 2 + (0,6 \cdot 0,4 \cdot 2)^2 + \dots)}_{\text{prog geométrica razón} = 0,6 \cdot 0,4 \cdot 2} = 0,6^2 \frac{1}{1 - 0,6 \cdot 0,4 \cdot 2} = \frac{9}{13} = \boxed{0,692}$$

Suma de una PG = $\dfrac{\text{Primer término}}{1 - \text{razón}}$

Sea $X \rightarrow N(\mu, \sigma)$ Hallar $x_0$ para que $p(X \leq x_0) = 0,75$

$$p(X \leq x_0) = p(\underbrace{z \leq \frac{x_0 - \mu}{\sigma}}) = 0,75$$

$$z_0 = 0,675 \longrightarrow \frac{x_0 - \mu}{\sigma} = 0,675 ; \quad \boxed{x_0 = 0,675\sigma + \mu}$$

EJERCICIOS APROXIMACIONES (ver tema 4)

**3.14** Examen 25 cuestiones. Cada cuestión 5 soluciones y solo 1 correcta. Nº mín. de respuestas correctas para aprobar es a. Se fija a con el criterio: prob. de aprobar al azar < 0,05. Se pide a.

$Y$ v.a nº respuestas correctas de un total de 25 $\rightarrow B(n=25, p=\frac{1}{5})$

$$p(\text{aprobar}) = p(Y \geq a \mid Y \rightarrow B(25, \tfrac{1}{5})) \simeq p(Y \geq a - 0,5 \mid Y \sim N(\mu = 5, \sigma = \sqrt{5 \cdot \tfrac{4}{5}} = 2)) =$$

$$= p\left(z \geq \frac{a - 0,5 - 5}{2}\right) = 1 - p\left(z \leq \frac{a - 5,5}{2}\right) = 0,05 ; \xrightarrow{\text{tablas}} \frac{a - 5,5}{2} = 1,64$$

$$a = 2 \cdot 1,64 + 5,5 = 8,78 \longrightarrow \boxed{a = 9 \text{ respuestas correctas}}$$

(3.17) d) Piezas se embalan en lotes de 500 con prob. de defectuoso $p = 0{,}024$. Se pide la probabilidad de que un lote tenga más de 15 defectuosas.

X v.a nº piezas def. de un total de 500 → Binomial ($n = 500$, $p = 0{,}024$)

$$p(X > 15 \mid X \to B(500, 0{,}024)) \simeq p\Big(Y \geq 15{,}5 \mid X \sim N\big(\mu = \underbrace{500 \cdot 0{,}024}_{12}, \sigma = \underbrace{\sqrt{500 \cdot 0{,}024 \cdot 0{,}976}}_{3{,}42}\big)\Big) =$$

$$= p\left(Z \geq \frac{15{,}5 - 12}{3{,}42}\right) = p(Z \geq 1{,}02) = 1 - \phi(1{,}02) = 1 - 0{,}8461 = \boxed{0{,}1539}$$

(3.18) Concesionario recibe pedidos según un proceso Poisson de media 2 vehículos/semana. Los pedidos se realizan con una antelación mínima de 1 mes. ¿Cuántos automóviles ha de tener a principios de 1 mes para satisfacer con una prob. igual o mayor a 0,95 la demanda mensual? (1 mes = 4 semanas).

X v.a nº pedidos 1 semana → $\lambda = 2$/sem

Y v.a nº pedidos 1 mes → $\lambda = 8$/mes

$p(Y \leq c) \geq 0{,}95$ → Se pide c

$$p(Y \leq c \mid Y \to \text{Poisson}(\lambda = 8)) \simeq p(Y \leq c + 0{,}5 \mid Y \sim N(8, \sqrt{8})) =$$

$$= p\Big(Z \leq \underbrace{\frac{c + 0{,}5 - 8}{\sqrt{8}}}_{z_0}\Big) \geq 0{,}95$$

$z_0 = 1{,}65$

$$\frac{c + 0{,}5 - 8}{\sqrt{8}} \geq 1{,}65;$$

Con la igualdad $c = 12{,}16$ → $\boxed{c > 13}$

(3.19) p(disparo impacte en diana) = 0,0001. Se pide

P(impactar 4 o más veces en 50 000 disparos)

X v.a. nº disparos que impactan de un total de 50000 → B(50000, 0,0001)

$$p(X \geq 4 \mid X \to B(n = 50000, p = 0{,}0001)) = 1 - \sum_{0}^{3} \binom{50\,000}{x} (0{,}0001)^x (1 - 0{,}0001)^{50000 - x} \simeq$$

$\left\{\begin{array}{l} n \to \infty \\ p \to 0 \end{array}\right\}$ Aproximo por la poisson ($\lambda = np$) $\simeq p(X \geq 4 \mid X \sim \text{Poisson}(\lambda = 50000 \cdot 0{,}0001)) =$

$$= p(X \geq 4 \mid X \sim \text{Poisson}(\lambda = 5)) = 1 - \sum_{x=0}^{x=3} e^{-5} \frac{5^x}{x!} = 1 - e^{-5}\left(1 + 5 + \frac{25}{2!} + \frac{5^3}{3!}\right) =$$

$$= \boxed{0{,}735}$$

(3.22) Control calidad lotes 30.000: muestra 300 piezas. Si hay 15 o más defectuosas se rechaza el lote. Cada mes aplica el control a 200 lotes. ¿nº esperado de lotes rechazados si cada lote tiene 4 % piezas defectuosas?

$X$ v.a. nº lotes rechazados en un mes de un total de 200

$X \rightarrow B(n=200,\ p=p(\text{rechazar lote}))$

$Y$ v.a. nº piezas defectuosas en una muestra de 300

$Y \rightarrow B(n=300,\ p=0{,}04)$

$p(\text{rechazar lote}) = p(Y \geq 15 \mid Y \rightarrow B(n=300,\ p=0{,}04)) \simeq$ sin corrección por cont.

$\simeq p(Y \geq 15 \mid Y \rightsquigarrow N(\mu = \underbrace{300 \cdot 0{,}04}_{12},\ \sigma = \underbrace{\sqrt{300 \cdot 0{,}04 \cdot 0{,}96}}_{\sqrt{11{,}52}})) =$

$= p\left(Z \geq \frac{15-12}{\sqrt{11{,}52}}\right) = p(Z \geq 0{,}88) = 1 - \Phi(0{,}88) = 1 - 0{,}8106 = 0{,}19$

Con corrección sería $p(\text{rechazar lote}) \simeq p\left(Z \geq \frac{14{,}5-12}{\sqrt{11{,}52}}\right) = 1 - \Phi(0{,}74) = 1 - 0{,}7704 =$

$= 0{,}2296$

Tomaremos $p(\text{rechazar lote}) = 0{,}19$

$E[X] = n \cdot p = 200 \cdot 0{,}19 = 38$ lotes se espera rechazar

Tomando $p(\text{rechazar lote}) = 0{,}2296$

$E(X) = n \cdot p = 200 \cdot 0{,}2296 = 45{,}92 \simeq 46$ lotes se espera rechazar.

(3.24) 500 personas van a un congreso. Se regalan corbatas a hombres y pañuelos a las mujeres. Calculo el nº min. de corbatas y pañuelos que deben estar disponibles para que todos los asistentes tengan regalo con una prop. 0,99. La prob. de que el asistente sea hombre o mujer es 0,5. Son independientes

$X$ v.a nº mujeres asistentes de un total de 500 $\rightarrow B(500, 0{,}5)$

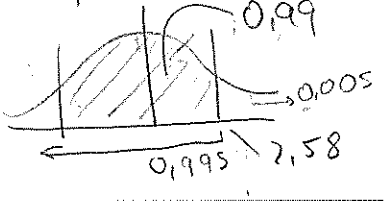
$Y$ v.a nº hombres asistentes " " " 500 $\rightarrow B(500, 0{,}5)$

$a$ = nº pañuelos = nº corbatas se pide $a$ tal que $p(X \leq a,\ Y \leq a) = 0{,}99 =$

Como $Y = 500 - X$ $\curvearrowright$ $= p(X \leq a,\ 500 - X \leq a) = p(500 - a \leq X \leq a) = 0{,}99$

Aproximando por la normal:

$p(500 - a \leq X \leq a \mid X \rightsquigarrow N(250, \sqrt{125})) = p\left(\frac{500-a-250}{\sqrt{125}} \leq X \leq \frac{a-250}{\sqrt{125}}\right) = 0{,}99$



$\frac{a-250}{\sqrt{125}} = 2{,}58$ ; $\boxed{a \simeq 279 \text{ pañuelos y } 279 \text{ corbatas}}$

**3.26** Empresa organiza viajes en autobus para asistir a actuaciones. 300 plazas que salen de Madrid y Sevilla. Precio Sevilla $= 60€$, Precio Mad. $= 50€$. Se cubren todas las plazas. $p(\text{salir Sevilla}) = \frac{1}{3}$ $p(\text{salir de Madrid}) = \frac{2}{3}$. Independencia en las reservas. Calcular $E[\text{ingresos}]$ y $Var[\text{ingresos}]$

$X$ v.a nº personas que salen de Sevilla de un tot. 300 $\rightarrow B(300, \frac{1}{3})$

$Y$ " " " " " " Madrid " " " " $\rightarrow B(300, \frac{2}{3})$

$I$ v.a. ingresos $\longrightarrow$ $I = 60X + 50Y$

$$E[I] = 60E[X] + 50E[Y] \underset{\uparrow}{=} 60 \cdot 300 \frac{1}{3} + 50 \cdot 300 \cdot \frac{2}{3} = 6000 + 10000 =$$

$E(X) = 300 \frac{1}{3}$

$E(Y) = 300 \cdot \frac{2}{3}$

$$= \boxed{16000 €}$$

$$Var[I] = Var[60X + 50Y] \underset{\substack{\uparrow \\ X+Y=300}}{=} Var[60X + 50(300 - X)] =$$

$$= Var[10X + 1500] = 10^2 \, Var[X] = 10^2 \cdot 300 \cdot \frac{1}{3} \cdot \frac{2}{3} = \boxed{6666,67}$$

**3.23** Servicio telefónico recibe 10 llam/1min a) ¿$p$(recibir más de 550 llam/1h?

$X$ v.a. llamadas en 1h $\longrightarrow$ Poisson$(\lambda = 60 \cdot 10$ llam/h$)$

$$p(X > 550 \mid X \sim \text{Poisson}(\lambda = 600)) \simeq p(X > 550,5 \mid X \sim N(600, \sqrt{600})) =$$

$$= p\left(Z > \frac{550,5 - 600}{\sqrt{600}}\right) = p(Z > -2,02) = p(Z < 2,02) = \boxed{0,9783}$$

b) Se instala un call center con cap. 650 llam/h ¿Nº esperado de h al año con nº llamadas sup. a su capacidad?

$$P(X > 650 \mid X \sim N(600, \sqrt{600})) = p\left(Z > \frac{650 - 600}{\sqrt{600}}\right) = p(Z > 2,04) =$$

$$= 1 - \Phi(2,04) = 1 - 0,9793 = 0,0207$$

$Y$ v.a. nº horas que se satura de un tot. de 8760h $\longrightarrow$ Binomial$(\overset{n}{8760}, \overset{p}{0,0207})$

$$E[Y] = n \cdot p = \boxed{181,332 \text{ h al año}}$$

3.21 Lotes de 10000 chips con porcentaje de defectuosos 10%.

Control de calidad I: Tomar 100 unidades y rechazar el lote si existen más de 15 defectuosos

Control de calidad II: Tomar 100 unidades, dividirlas en 10 grupos y si en algún grupo hay más de una pieza defectuosa rechazar el lote.

a) ¿ p(rechazar lote con 10% chips defectuosos) para cada método?

Control I

X v.a nº unidades def. de un total de 100 → $B(n=100, p=0,1)$

$$p(\text{rechazo}) = p(X \geq 15 \mid X \to B(100, 0,1)) \simeq p(X \geq 14,5 \mid X \sim N(10, \sqrt{9})) =$$

$$= p\left(Z \geq \frac{14,5-10}{3}\right) = p(Z \geq 1,5) = 1 - p(Z < 1,5) = 1 - 0,9332 = \boxed{0,0668}$$

Control II Y v.a nº unidades def. de un lot. de 10

$$p(\text{rechazo}) = [p(Y > 1 \mid Y \to B(10, 0,1))]^{10} = [1 - p(Y \leq 1)]^2 =$$

$$= \left(1 - \sum_{y=0}^{1} \binom{10}{y} 0,1^y \, 0,9^{10-y}\right)^{10} = \left(1 - 0,9^{10} - 10 \cdot 0,1 \cdot 0,9^9\right)^{10} = \boxed{1,63\ 10^{-6}}$$

# FÓRMULAS → MODELOS UNIVARIANTES

- **Distribución Binomial**

X nº elementos defectuosos al observar n

$$\boxed{\begin{array}{l} X \rightarrow B(n,p) \\ P(X=K) = \binom{n}{K} p^K (1-p)^{n-K} \end{array}} \qquad \boxed{\begin{array}{l} E[X] = np \\ Var[X] = np(1-p) \end{array}}$$

- **Distribución Geométrica**

Y nº elementos hasta aparecer uno defectuoso.

$$\boxed{\begin{array}{l} Y \rightarrow G(p) \\ P(Y=K) = (1-p)^{K-1} p \end{array}} \qquad \boxed{\begin{array}{l} E[y] = \dfrac{1}{p} \\ Var[y] = \dfrac{1-p}{p^2} \end{array}}$$

- **Distribución de Poisson**

X nº defectos en un tramo continuo con λ nº medio de defectos.

$$\boxed{\begin{array}{l} X \rightarrow Poisson(\lambda) \\ P(X=x) = e^{-\lambda} \dfrac{\lambda^x}{x!} \end{array}} \qquad \boxed{E[X] = Var[X] = \lambda}$$

$x = 0, 1, 2, \dots$

- **Distribución Exponencial** $\quad T \rightarrow Exponencial(\lambda)$

T → distancia entre dos defectos consecutivos con λ nº medio de defectos en un tramo continuo.

$$\boxed{\begin{array}{ll} P(T \geq t) = e^{-\lambda t} & ; \ t \geq 0 \\ F_T(T) = P(T \leq t) = 1 - e^{-\lambda t} & ; \ t \geq 0 \\ f_T(T) = \lambda e^{-\lambda t} & ; \ t \geq 0 \end{array}}$$

$$\boxed{E[T] = \frac{1}{\lambda} \ ; \quad Var[T] = \frac{1}{\lambda^2}}$$

- **Distribución Normal** $\quad X \rightarrow N(\mu, \sigma)$

$$\boxed{f_X(X) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad X \in \mathbb{R}}$$

$$C_{AS} = 0 \ ; \quad C_{AP} = 3$$

Estandarización $\quad X \rightarrow N(\mu, \sigma) \equiv z = \dfrac{X-\mu}{\sigma}, \ z \rightarrow N(0,1)$

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \ ; \quad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightarrow \text{Valor de tablas!}$$

⊛ **TASA DE FALLO**

$$\boxed{\lambda_w = \frac{f_w(w)}{1 - F_w(w)}}$$

TEMA 4

# Modelos Multivariantes

## 1 VARIABLES DISCRETAS

La distribución conjunta de probabilidad de dos variables aleatorias X, Y

$$P(X=x, Y=y) \begin{cases} P(X=x, Y=y) \geq 0 \quad \forall x, y \\ \sum_{x=-\infty}^{x=\infty} \sum_{y=-\infty}^{y=\infty} P(X=x, Y=y) = 1 \end{cases}$$

Función de distribución conjunta $F_{xy}(x,y) = P(X \leq x, Y \leq y)$

Distribuciones marginales

$$\begin{cases} P(X=x) = \sum_{y=-\infty}^{y=\infty} P(X=x, Y=y) & \text{Distribución marginal de la v.a. } X \\ P(Y=y) = \sum_{x=-\infty}^{x=\infty} P(X=x, Y=y) & \text{Distribución marginal de la v.a. } Y \end{cases}$$

Distribuciones condicionadas $P(Y=y \mid X=x_0) = \dfrac{P(Y=y, X=x_0)}{P(X=x_0)}$

Independencia: Dos variables aleatorias X e Y son independientes si y sólo si:

$P(X=i, Y=j) = P(X=i)\,P(Y=j) \Rightarrow$ Conocer una no aporta información sobre la otra

## 2 VARIABLES ALEATORIAS CONTINUAS

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_{c}^{d}\int_{a}^{b} f_{xy}(x,y)\,dx\,dy$$

(con $y$ bajo la primera integral y $x$ bajo la segunda)

Siendo $f_{xy}(x,y)$ la función de densidad conjunta, que cumple

$$\begin{cases} f_{xy}(x,y) \geq 0, \ \forall x, y \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\,dx\,dy = 1 \end{cases}$$

Función de distribución $\int_{-\infty}^{y}\int_{-\infty}^{x} f_{xy}(u,v)\,du\,dv = F_{XY}(x,y)$

Funciones de densidad marginales

$$\begin{cases} \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy = f_x(x) \\ \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx = f_y(y) \end{cases}$$

Funciones de densidad condicionadas

$$\begin{cases} f_{X|Y}(x|y) = \dfrac{f_{xy}(x,y)}{f_y(y)}, & \text{cuando } f_y(y) > 0 \\ f_{Y|X}(y|x) = \dfrac{f_{xy}(x,y)}{f_x(x)}, & \text{cuando } f_x(x) > 0 \end{cases}$$

## Independencia:

Las variables aleatorias X e Y son independientes si y sólo si:

$$f_{xy}(x,y) = f_x(x)\, f_y(y)$$

Por tanto, también podemos decir que las v.a. X e y son independientes si y solo si:

$$\left\{\begin{array}{l} f_{x|y}(x|y) = f_x(x) \\ f_{y|x}(y|x) = f_y(y) \end{array}\right.$$

Ejemplo 1: Las v.a. X e Y tienen como función de densidad conjunta

$$f_{xy}(x,y) = 6xy^2, \quad 0<x<1, \quad 0<y<1$$

1) Función de distribución conjunta:

$$F_{xy}(x,y) = P(X\le x,\ Y\le y) = \underset{X}{\int_0^x}\underset{Y}{\int_0^y} f(u,v)\,dv\,du = \int_0^x\int_0^y 6uv^2\,dv\,du =$$

$$= \int_0^x 6u\frac{v^3}{3}\Big|_0^y du = 2\int_0^x u y^3\,du = x^2y^3, \quad \boxed{F_{xy}(x,y) = x^2y^3}$$

2) Probabilidad de que $X+y\le 1$

$$P(X+y\le 1) = \iint_{x+y\le 1} 6xy^2\,dx\,dy = \underset{y}{\int_0^1}\underset{X}{\int_0^{1-y}} 6xy^2\,dx\,dy = \underset{y}{\int_0^1} 6y^2\frac{x^2}{2}\Big|_0^{1-y} dy =$$

$$= \int_0^1 3y^2(1-y)^2\,dy = 3\int_0^1 (y^2-2y^3+y^4)\,dy = 3\left(\frac{y^3}{3} - 2\frac{y^4}{4} + \frac{y^5}{5}\right)\Big|_0^1 =$$

$$= 3\left(\frac{1}{3} - \frac{1}{2} + \frac{1}{5}\right) = 3\,\frac{10-15+6}{30} = \boxed{\frac{1}{10}}$$

3) Marginales.

$$f_x(x) = \int_{y=0}^{y=1} 6xy^2\,dy = 6x\frac{y^3}{3}\Big|_0^1 = 6x\frac{1}{3} = 2x, \quad 0<x<1$$

$$f_y(y) = \int_{x=0}^{x=1} 6xy^2\,dx = 6y^2\frac{x^2}{2}\Big|_0^1 = 3y^2, \quad 0<x<1$$

4) Son independientes

$f_{xy}(x,y) = f_x(x)\cdot f_y(y)$ ??

$$f_{xy}(x,y) = 6xy^2 \longrightarrow \boxed{\text{Son independientes}}$$

$$f_x(x)\cdot f_y(y) = 2x\cdot 3y^2 = 6xy^2$$

5) Si fuera $f_{xy}(x,y)=\frac{1}{x}$ $\quad 0\le y\le x\le 1$ ¿Serían independientes?

$$\begin{cases} f_x(x)=\int_0^x \frac{1}{x}\,dy=\frac{1}{x}\,y\Big|_0^x=\frac{1}{x}(x-0)=1, & 0<x<1\\ f_y(y)=\int_y^1 \frac{1}{x}\,dx=\ln|x|\Big|_y^1=-\ln(y), & 0<y<1\end{cases}$$

Como $f_{xy}(x,y)\neq f_x(x)\cdot f_y(y)\longrightarrow \boxed{\text{No son independientes}}$

6) $f_{x|y}(x|y)$ de la 1ª función de densidad $f_{xy}(x,y)=6xy^2$; $0<x<1$ $\&y<1$

$$f_{x|y}(x|y)=\frac{f_{xy}(x,y)}{f_y(y)}=\frac{6xy^2}{3y^2}=2x=f_x(x)\longrightarrow \text{Independencia}$$

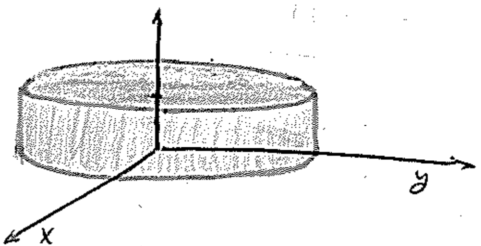
$0<x<1$

7) $f_{x|y}(x|y)$ de la 2ª función de densidad $f_{xy}(x,y)=\frac{1}{x}$ $\quad 0\le y\le x\le 1$

$$f_{x|y}(x|y)=\frac{f_{xy}(x,y)}{f_y(x)}=\frac{1/x}{-\ln(y)}=\frac{-1}{x\ln(y)} \qquad y\le x\le 1$$

$$f_{y|x}(y|x)=\frac{f_{xy}(x,y)}{f_x(x)}=\frac{1/x}{1}=\frac{1}{x} \qquad 0\le y\le x$$

Ejemplo 2 Sea $f_{xy}(x,y)=\begin{cases} c, & x^2+y^2\le r^2\\ 0, & x^2+y^2>r^2\end{cases}$



① Hallar c

Por tratarse de un cilindro

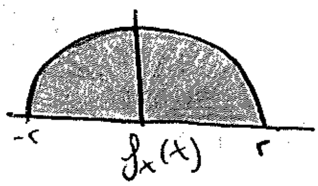
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c\,dx\,dy \neq \pi r^2 c=1 \longrightarrow \boxed{c=\frac{1}{\pi r^2}}$$

x

② $f_x(x)$

$$f_x(x)=\int_{-\infty}^{\infty} f_{xy}(x,y)\,dy=\frac{1}{\pi r^2}\int_{y=-\sqrt{r^2-x^2}}^{y=+\sqrt{r^2-x^2}} dy=\frac{1}{\pi r^2}\left(\sqrt{r^2-x^2}+\sqrt{r^2-x^2}\right)=\frac{2}{\pi r^2}\sqrt{r^2-x^2}$$

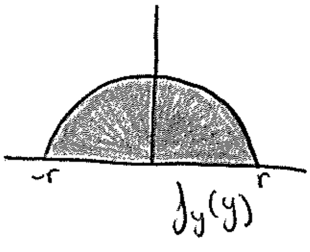
$$\boxed{f_x(x)=\frac{2}{\pi r^2}\sqrt{r^2-x^2} \qquad -r\le x\le r}$$



③ $f_x(y)$

$$f_y(y)=\int_{-\infty}^{\infty} f_{xy}(x,y)\,dx=\frac{1}{\pi r^2}\int_{-\sqrt{r^2-y^2}}^{+\sqrt{r^2-y^2}} dy=\frac{1}{\pi r^2}\left(\sqrt{r^2-y^2}+\sqrt{r^2-y^2}\right)=\frac{2}{\pi r^2}\sqrt{r^2-y^2}$$

$$\boxed{f_y(y)=\frac{2}{\pi r^2}\sqrt{r^2-y^2} \qquad -r\le y\le r}$$



④ $f_{x|y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)} = \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2}\sqrt{r^2-y^2}} = \boxed{\frac{1}{2\sqrt{r^2-y^2}}}$ ; $-\sqrt{r^2-y^2} \le x \le \sqrt{r^2-y^2}$

⑤ Son independientes ¿X y ¿Y?

Lo serán si $f_{xy}(x,y) = f_x(x)\, f_y(y)$ ← Condición de independencia

Como $f_{xy}(x,y) = \begin{cases} \frac{1}{\pi r^2}, & x^2+y^2 \le r^2 \\ 0; & x^2+y^2 > r \end{cases}$ y $\begin{cases} f_x(x) = \frac{2}{\pi r^2}\sqrt{r^2-x^2}, & -r \le x \le r \\ f_y(y) = \frac{2}{\pi r^2}\sqrt{r^2-y^2}, & -r \le y \le r \end{cases}$

No se cumple la condición de independencia → $\boxed{\text{No son independientes}}$

## 3 ESPERANZA DE g(X, Y)

Si X, Y son variables aleatorias discretas, se define su esperanza como:

$$\boxed{E[g(x,y)] = \sum_{x=-\infty}^{x=\infty} \sum_{y=-\infty}^{y=\infty} g(x,y)\, P(X=x,\, Y=y)}$$

Si X, Y son variables aleatorias continuas, se define su esperanza como:

$$\boxed{E[g(x,y)] = \underset{y}{\int_{-\infty}^{\infty}} \underset{x}{\int_{-\infty}^{\infty}} g(x,y)\, f_{xy}(x,y)\, dx\, dy}$$

Propiedades:

Si $g(X,Y) = g_1(x) + g_2(y)$ se cumple $\boxed{E[g(x,y)] = E[g_1(x)] + E[g_2(y)]}$

Demo: $E[g(x,y)] = E[g_1(x) + g_2(y)] = \underset{y}{\int_{-\infty}^{\infty}} \underset{x}{\int_{-\infty}^{\infty}} (g_1(x) + g_2(y))\, f_{xy}(x,y)\, dx\, dy =$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) f_{xy}(x,y)\, dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(y)\, f_{xy}(x,y)\, dx\, dy =$$

$$= \underset{x}{\int_{-\infty}^{\infty}} g_1(x) \left( \underbrace{\underset{y}{\int_{-\infty}^{\infty}} f_{xy}(x,y)\, dy}_{f_x(x)} \right) dx + \underset{y}{\int_{-\infty}^{\infty}} g_2(y) \left( \underbrace{\underset{x}{\int_{-\infty}^{\infty}} f_{xy}(x,y)\, dx}_{f_x(y)} \right) dy =$$

$$= \int_{-\infty}^{\infty} g_1(x)\, f_x(x)\, dx + \int_{-\infty}^{\infty} g_2(y)\, f_y(y)\, dy = E[g_1(x)] + E[g_2(y)]$$

# 4 COVARIANZA

La covarianza de dos variables aleatorias X e Y se denota por cov(X, Y) y se define como (en v.a. continuas):

$$Cov(X,Y)= E[(X-\mu_x)(Y-\mu_y)] = \int_{\underset{y}{-\infty}}^{\infty}\int_{\underset{x}{-\infty}}^{\infty}(x-\mu_x)(y-\mu_y)\, f(x,y)\,dx\,dy$$

donde $\mu_x = E[X]$ y $\mu_y = E[Y]$

Si las variables son discretas:

$$Cov(X,Y)= E[(X-\mu_x)(Y-\mu_y)] = \sum_{x=i}\sum_{y=j}(x_i-\mu_x)(y_i-\mu_y)\,P(X=x_i, Y=y_i)$$

Propiedades:

- La covarianza es una medida de la dependencia lineal entre las dos variables. Si las variables son independientes → $\boxed{cov(X,Y)=0}$

Demo para v.a. continuas

indep → $f(x,y)=f_x(x)\,f_y(y)$

$$Cov(X,Y)=\int_{\underset{y}{-\infty}}^{\infty}\int_{\underset{x}{-\infty}}^{\infty}(x-\mu_x)(y-\mu_y)\,f(x,y)\,dx\,dy \overset{\downarrow}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\,f_x(x)\,f_y(y)\,dx\,dy =$$

$$= \underbrace{\int_{-\infty}^{\infty}(x-\mu_x)\,f_x(x)\,dx}_{\int_{-\infty}^{\infty} x\, f_x(x)dx - \int_{-\infty}^{\infty}\mu_x f_x(x)\,dx = E[X]-\mu_x = 0}\ \underbrace{\int_{-\infty}^{\infty}(y-\mu_y)\,f_y(y)\,dy}_{\int_{-\infty}^{\infty} y\, f_y(y)dy - \int_{-\infty}^{\infty}\mu_y f_y(y)\,dy = E[Y]-\mu_y = 0} = 0$$

Demo para v.a. discretas

indep → $P_{xy}(X=x, Y=y) = P_x(X=x)\,P_y(Y=y)$

$$Cov(X,Y)=\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}(x-\mu_x)(y-\mu_x)\,P_{xy}(X=x,Y=y) \overset{\downarrow}{=} \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\,P_x(X=x)\,P_y(Y=y) =$$

$$= \underbrace{\sum_{x=-\infty}^{\infty}(x-\mu_x)P_x(X=x)}_{E[(X-\mu_x)] = E[X]-\mu_x = 0}\ \underbrace{\sum_{y=-\infty}^{\infty}(y-\mu_y)P_y(Y=y)}_{E[(Y-\mu_y)] = E[Y]-\mu_y = 0} = 0$$

- $\boxed{Cov(X,Y) = E[XY] - E[X]\,E[Y]}$

Demo: $Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] = E[XY - X\mu_y - Y\mu_x + \mu_x\mu_y] =$

$$= E[X\cdot Y] - \mu_y E[X] - \underbrace{\cancel{\mu_x E[Y]}}_{\mu_x\mu_y} + \cancel{\mu_x\mu_y} = E[X\cdot Y] - E[X]\,E[Y]$$

Si X, Y son independientes $E[X\cdot Y] = E[X]\,E[Y] \to Cov(X,Y) = \underbrace{E[X\cdot Y]}_{E[X]E[Y]} - E[X]E[Y] = 0$

Demo de $E[X\cdot Y]=E[X]E[Y]$ si y sólo si $X$ e $Y$ son independientes

$$E[X\cdot Y]=\int_{\underset{Y}{-\infty}}^{\infty}\int_{\underset{X}{-\infty}}^{\infty} x y \, f_{XY}(x,y)\,dx\,dy \underset{\substack{\uparrow\\ \text{independencia}}}{=} \int_{\underset{Y}{-\infty}}^{\infty}\int_{\underset{X}{-\infty}}^{\infty} x y \, f_X(x)\, f_Y(y)\,dx\,dy =$$

$$= \int_{\underset{X}{-\infty}}^{\infty} x\, f_X(x)\,dx \int_{\underset{Y}{-\infty}}^{\infty} y\, f_y(y) = E[X]E[Y]$$

- $$\boxed{\begin{aligned} Var(X+Y) &= Var(X)+Var(Y)+2\,cov(X,Y)\\ Var(X-Y) &= Var(X)+Var(Y)-2\,cov(X,y)\end{aligned}}$$

Demo: $Var(X+y) = E[((X+y)-E[X+y])^2] = E[(X+y-E[X]-E[Y])^2] =$

$$= E[((X-E[X])+(Y-E[Y]))^2] = E[(X-E[X])^2 + (y-E[Y])^2 + 2(X-E[X])(y-E[Y])] =$$

$$= E[(X-E[X])^2] + E[(y-E[y])^2] + 2E[(X-E[X])(y-E[y])] =$$

$$= Var(X) + Var(y) + 2\,cov(X,y)$$

## 5 CORRELACIÓN

Se define el coeficiente de correlación $\rho$ entre dos variables aleatorias $X, y$ como:

$$\boxed{\rho(X,y) = \frac{Cov(X,y)}{\sqrt{Var(X)}\sqrt{Var(y)}}}$$

Propiedades:

- $-1 \leq \rho(X,y) \leq +1$
- $X, y$ son independientes $\longrightarrow \rho(X,y)=0$
- $\rho(X,y)=0 \longrightarrow$ Ausencia de linealidad y sólo en algunos casos independencia
- $Y = a + bX \Longleftrightarrow \underbrace{\rho(X,y)=1 \ (b>0)}_{\text{Relación lineal perfecta positiva}} \ \text{o} \ \underbrace{\rho(X,y)=-1 \ (b<0)}_{\text{Relación lineal perfecta negativa}}$

# 6 TEORÍA GENERAL PARA n VARIABLES ALEATORIAS

Para hacer el cálculo de probabilidades de un suceso en el que intervengan n variables aleatorias $X_1, X_2, \ldots, X_n$ es preciso conocer la distribución de probabilidad conjunta

- Si las variables son continuas se emplea la función de densidad conjunta

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad f_{\mathbf{X}}(X_1, X_2, \ldots X_n)$$

o la función de distribución conjunta. $F_{\mathbf{X}}(X_1, X_2, \ldots X_n)$

$$F_{\mathbf{X}}(X_1, X_2, \ldots X_n) = \int_{-\infty}^{X_n}\int_{-\infty}^{X_{n-1}} \cdots \int_{-\infty}^{X_1} f_X(t_1, t_2, \ldots t_n)\,dt_1\,dt_2 \ldots dt_n$$

- Distribuciones marginales

$$f_{X_1}(X_1) = \underset{X_2}{\int_{-\infty}^{\infty}}\ \underset{X_3}{\int_{-\infty}^{\infty}} \cdots \underset{X_n}{\int_{-\infty}^{\infty}} f_X(X_1, X_2, \ldots X_n)\,dX_n \ldots dX_3\,dX_2 \equiv$$

Función de densidad marginal de $X_1$

$\longrightarrow$ Todas menos $X_1$

$$f_{X_1X_2}(X_1, X_2) = \underset{X_3}{\int_{-\infty}^{\infty}}\ \underset{X_4}{\int_{-\infty}^{\infty}} \cdots \underset{X_n}{\int_{-\infty}^{\infty}} f_{\bar{X}}(X_1, X_2, \ldots X_n)\,dX_n \ldots dX_4\,dX_3 \equiv$$

Función de densidad marginal de $(X_1, X_2)$

$\longrightarrow$ Todas menos $X_1, X_2$

- Esperanza

Sea $g(\mathbf{X}) = g(X_1, X_2, \ldots X_n)$

$$E[g(\mathbf{X})] = \underset{X_n}{\int_{-\infty}^{\infty}}\ \underset{X_{n-1}}{\int_{-\infty}^{\infty}} \cdots \underset{X_1}{\int_{-\infty}^{\infty}} g(X_1, X_2, \ldots X_n)\, f_{\mathbf{X}}(X_1, X_2, \ldots X_n)\,dX_1\,dX_2 \ldots dX_n$$

$$E[X_i] = \int_{-\infty}^{\infty} X_i\, f_{X_i}(X_i)\,dX_i$$

$$E[g(X_1, X_2)] = \underset{X_2}{\int_{-\infty}^{\infty}}\ \underset{X_1}{\int_{-\infty}^{\infty}} g(X_1, X_2)\, f_{X_1X_2}(X_1, X_2)\,dX_1\,dX_2$$

- Vector de medias y matriz de varianzas

$$\bar{X} = (X_1, X_2, \ldots X_n)^T \longrightarrow E[\mathbf{X}] = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix}$$

$$\hookrightarrow Var[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \quad \text{con} \quad \begin{matrix} Var(X_i) = \sigma_i^2 \\ Cov(X_i, X_j) = \sigma_{ij};\ i \neq j \end{matrix}$$

- Independencia.

Las variables $X_1, X_2, \ldots X_n$ son independientes si y sólo si:

$$f(X_1, X_2 \ldots X_n) = f_{X_1}(X_1)\, f_{X_2}(X_2) \ldots f_{X_n}(X_n)$$

## 7 TRANSFORMACIONES LINEALES VECTORIALES

$$\underbrace{X = (X_1, X_2, \ldots X_n)^T}_{\text{vector de n v.a.}}, \quad \underbrace{a = (a_1, a_2 \ldots a_n)^T}_{\text{vector de n ctes}} \longrightarrow Y = a^T X$$

$$\boxed{E[Y] = a^T E[X]} \qquad \boxed{Var[Y] = a^T Var[X]\, a}$$

⊛

## 8 TRANSFORMACIONES LINEALES MATRICIALES

$$\underbrace{X = (X_1, X_2, \ldots X_n)^T}_{\text{Vector de n v.a.}} \qquad A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \rightarrow \text{Matriz de mxn ctes}$$

$$Y = AX \longrightarrow \boxed{E[Y] = A\,E[X]} \qquad \boxed{Var[Y] = A\, Var[X]\, A^T}$$

⊛ Caso particular. Si $X = (X_1, X_2, \ldots X_n)^T \rightarrow$ Vector de n v.a <u>independientes</u>

$Y = aX$ con $a$ = vector de constantes $\longrightarrow Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$

$$Var[Y] = a^T Var[X]\, a \overset{\text{por independencia}}{=} (a_1, a_2 \ldots a_n) \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2$$

⊛ Ejemplo: Calcular la media y la varianza de la suma de 12 variables aleatorias independientes con distribución uniforme en (0,1)

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{12} \end{pmatrix} \qquad U_i \rightarrow \text{uniforme } (0,1) \longrightarrow f_{U_i}(U_i) = 1 \quad 0 < U_i < 1,$$

$U_i, U_j$ indep $\longrightarrow Cov(U_i, U_j) = 0$

$$E[U_i] = \int_0^1 U_i\, 1\, dU_i = \frac{U_i^2}{2}\Big|_0^1 = \frac{1}{2}$$

$$Var[U_i] = E[U_i^2] - (E[U_i])^2 = \int_0^1 U_i^2\, 1\, dU_i - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{4-3}{12} = \frac{1}{12}$$

$$\mho = U_1 + U_2 + \ldots + U_{12}, \quad \mho = \underbrace{(1, 1 \ldots 1)}_{a^T} U$$

$$E[\mho] = a^T E[U] = (1, 1 \ldots 1) \begin{pmatrix} 1/2 \\ 1/2 \\ \vdots \\ 1/2 \end{pmatrix} = 12 \cdot \frac{1}{2} = 6$$

$$\boxed{E[\mho] = 6}$$

$$Var[\mho] = a^T Var[U]\, a = (1, 1, \ldots 1) \begin{pmatrix} \frac{1}{12} & 0 & \ldots & 0 \\ 0 & \frac{1}{12} & \ldots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \ldots & \frac{1}{12} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = 12 \cdot \frac{1}{12} = 1 \; ; \quad \boxed{Var[\mho] = 1}$$

Ejemplo: Se dispone de n sobres con sus correspondientes cartas. Se extraen las cartas de los sobres, se sortean y se vuelven a introducir de forma aleatoria cada una en un sobre. ¿Cuál es el nº esperado de cartas que coinciden con su sobre inicial?

$X =$ v.a nº de coincidencias $\quad X = X_1 + X_2 + \dots + X_n$

$$X_i : \begin{cases} 1 & \text{si la carta } i \text{ coincide con su sobre} \\ 0 & \text{si la carta } i \text{ no coincide con su sobre} \end{cases} \qquad E[X_i] = \frac{1}{n}$$

Se espera que de n cartas coincida 1

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n] = \frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n} = \underline{1} \text{ carta}$$

Un proceso fabrica una proporción p de tornillos defectuosos. Se define X como la v.a. "número de tornillos extraídos del proceso hasta aparecer r defectuosos". Se pide E(X) y Var(X]

$X_1 =$ Nº tornillos hasta el 1º defectuoso

$X_2 =$ Nº tornillos nuevos hasta el 2º defectuoso

$\vdots$

$X_i =$ Nº tornillos nuevos hasta el i-ésimo defectuoso

$X_i \to$ v.a. geométrica (p)

$$E[X_i] = \frac{1}{p}$$

$$Var[X_i] = \frac{(1-p)}{p^2}$$

$$X = X_1 + X_2 + \dots + X_r \longrightarrow \boxed{E[X] = E[X_1] + E[X_2] + \dots + E[X_r] = \frac{r}{p}}$$

$$\boxed{Var[X] \overset{indep.}{=} Var[X_1] + Var[X_2] + \dots + Var[X_r] = \frac{r(1-p)}{p^2}}$$

Sea $Y = X_1 + 3X_2$; $M_X = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 \end{pmatrix}$; $E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \end{pmatrix}$. Calcular $E[Y]$ y $Var[Y]$

$$Y = (1;3)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = a^T X$$

$$\boxed{E[Y]} = a^T E[X] = (1,3)\begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix} = \boxed{E[X_1] + 3E[X_2]}$$

$$\boxed{Var[Y]} = a^T M_X a = (1,3)\begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = (\sigma_{x_1}^2 + 3\sigma_{x_1x_2} \quad \sigma_{x_1x_2} + 3\sigma_{x_2}^2)\begin{pmatrix} 1 \\ 3 \end{pmatrix} =$$

$$= \sigma_{x_1}^2 + 3\sigma_{x_1x_2} + 3\sigma_{x_1x_2} + 9\sigma_{x_2}^2 = \boxed{\sigma_{x_1}^2 + 9\sigma_{x_2}^2 + 6\sigma_{x_1x_2}}$$

Sean X e Y variables aleatorias continuas tal que $f_{xy}(x,y) = \begin{cases} c & 0<x<1 \\ & 0<y<1 \\ 0 & \text{en el resto} \end{cases}$

Se pide: c, $f_x(x)$, $f_y(y)$, $F_x(x)$, $F_y(y)$, $F_{xy}(x,y)$

NOTA!! Se trata de la función bidimensional continua uniforme

$$\int_{\substack{0\\x}}^{1}\int_{\substack{0\\y}}^{1} c\,dy\,dx = \int_{\substack{0\\x}}^{1} cy\Big|_0^1 dx = \int_0^1 c\,dx = c = 1 \longrightarrow \boxed{c=1}$$

$$f_x(x) = \int_{\substack{0\\y}}^{1} 1\,dy = 1 \quad \boxed{\begin{matrix} f_x(x)=1 \\ x\in(0,1)\end{matrix}} \;; \quad f_y(y) = \int_{\substack{0\\x}}^{1} 1\,dx = 1 \quad \boxed{\begin{matrix} f_y(y)=1 \\ y\in(0,1)\end{matrix}}$$

$$F_{xy}(x,y) = \int_{\substack{0\\x}}^{x}\int_{\substack{0\\y}}^{y} 1\,dy\,dx = xy \qquad \boxed{F_{xy}(x,y) = xy}$$

$$F_x(x) = \int_0^x 1\,dx = x \,, \quad F_y(y) = \int_0^y 1\,dy = y \longrightarrow \boxed{\begin{matrix} F_x(x) = x \\ F_y(y) = y \end{matrix}}$$

## 9 TEOREMA CENTRAL DEL LÍMITE

Sea $\mathbf{X} = (X_1, X_2, \ldots X_n)^T \rightarrow$ vector de n v.a. independientes.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad \begin{cases} E[\bar{X}] = \dfrac{E[X_1] + E[X_2] + \ldots + E[X_n]}{n} \\[2ex] Var[\bar{X}] = \dfrac{Var[X_1] + Var[X_2] + \ldots + Var[X_n]}{n^2} \end{cases}$$

Si las variables tienen misma media y varianza: $\mu = E[X_i] \; \forall i$ $\Longrightarrow$
$\sigma^2 = Var[X_i] \; \forall i$

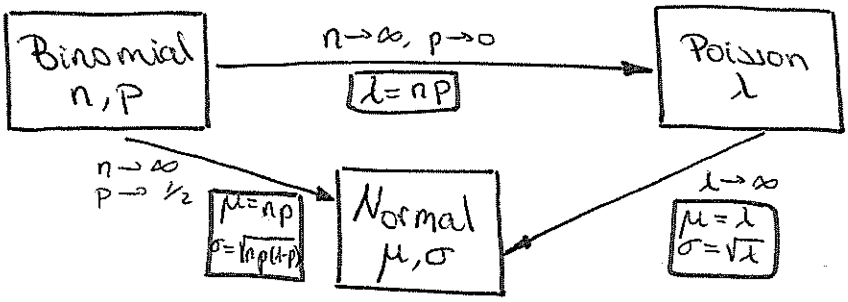
$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad \begin{cases} E[\bar{X}] = \mu \\ Var[\bar{X}] = \dfrac{\sigma^2}{n} \end{cases}$$

Teorema: Sean $X_1, X_2 \ldots X_n$ variables aleatorias independientes con la misma distribución de probabilidad de media $\mu$ y varianza $\sigma^2$

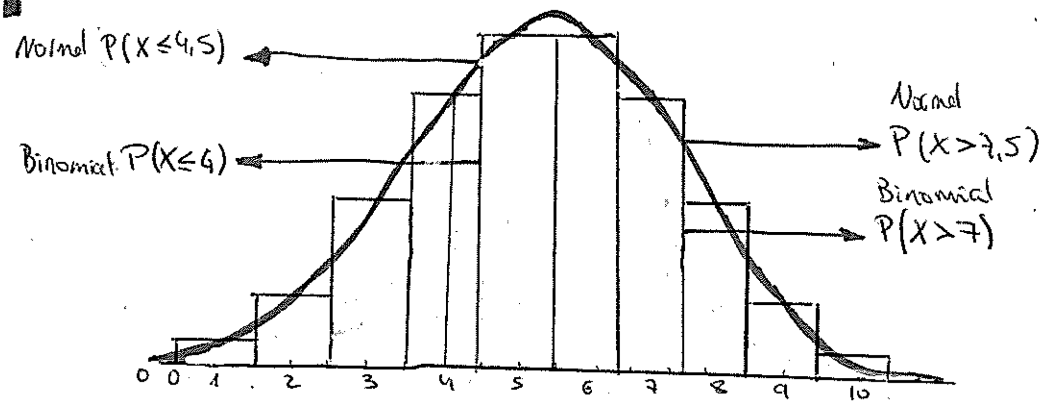
Entonces $\boxed{\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)}$

## 10 BINOMIAL- POISSON- NORMAL

En ocasiones será conveniente aproximar unas distribuciones de probabilidad a otras.



## 11 CORRECCIÓN POR CONTINUIDAD

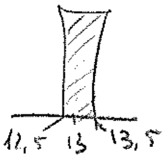


➔ Muestra 45 piezas de un proceso que fabrica un promedio de 25% piezas fuera de especificación.

① ¿Probabilidad de que en la muestra haya 13 defectuosos?

X : v.a nº elementos defectuosos de 45

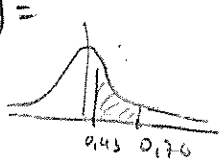
$$P(X=13 \mid X\to B(n=45,\ p=0{,}25)) = \binom{45}{13} 0{,}25^{13}\, 0{,}75^{32} = \boxed{0{,}109}$$



Por aprox. a la normal

$$p(X=13 \mid X\to B(n=45,\ p=0{,}25)) \approx p(12{,}5<X<13{,}5 \mid X \rightsquigarrow N(11{,}25,\ 2{,}9)) =$$

$$= p\left(\frac{12{,}5-11{,}25}{2{,}9} < Z < \frac{13{,}5-11{,}25}{2{,}9}\right) = p(0{,}43 < Z < 0{,}76) =$$



$$= p(X<0{,}76) - p(X<0{,}43) = \phi(0{,}76) - \phi(0{,}43) = 0{,}7764 - 0{,}6664 =$$

$$= \boxed{0{,}11}$$

② ¿Prob de que la muestra contenga 13 o más piezas defectuosas?

$$P(X \geq 13 \mid X\to B(n=45,\ p=0{,}25)) = 1-p(X<13) = 1 - \sum_{0}^{12}\binom{45}{x} 0{,}25^{x}\, 0{,}75^{45-x} \rightarrow \text{Aproximamos}$$

$$p(X \geq 12{,}5 \mid X \rightsquigarrow N(11{,}25,\ 2{,}9)) = p\left(Z \geq \frac{12{,}5-11{,}25}{2{,}9}\right) = p(Z \geq 0{,}43) =$$

$$= 1 - \phi(0{,}43) = 1 - 0{,}6664 = \boxed{0{,}3336}$$

# 12 DISTRIBUCIÓN NORMAL MULTIVARIANTE

Sean:

$X = (x_1, x_2, \dots x_n)^T \in \mathbb{R}^n$

$\mu = (\mu_1, \mu_2, \dots \mu_n)^T \in \mathbb{R}^n$

M: Matriz de varianzas $n \times n$ $\longrightarrow$ $M = Var(X) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}$

$$\boxed{f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |M|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T M^{-1} (x-\mu)\right\}}$$

Si la distribución es bivariante:

$$f(x_1, x_2) = \frac{1}{2\pi |M|^{1/2}} \exp\left\{-\frac{1}{2} (x-\mu_1 \quad x-\mu_2) M^{-1} \begin{pmatrix} x-\mu_1 \\ x-\mu_2 \end{pmatrix}\right\}$$

$x = (x_1, x_2)^T$

$\mu = (\mu_1, \mu_2)^T$

$$M = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \overset{\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}}{=} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}; \quad |M| = \sigma_1^2 \sigma_2^2 (1-\rho^2)$$

$$M^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{(1-\rho^4)} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & 1/\sigma_2^2 \end{pmatrix}$$

$$\boxed{f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}}$$

Propiedades:

- Las distribuciones marginales son normales $N(\mu_i, \sigma_i)$
- Las distribuciones condicionadas son normales
- $\rho = 0 \Longleftrightarrow$ Variables independientes
- Tranf. lineales $Y = AX \Longrightarrow X \rightarrow N(\mu, M)$; $Y \rightarrow N(A\mu, AMA^T)$

Sea $(X_1, X_2, X_3)$ una normal tridimensional de media $(10, 20, 30)$ y

$M = \begin{pmatrix}1&0&0\\0&1&0\\0&0&4\end{pmatrix}$ ¿ $P(X_1 + X_2 \geqslant X_3 + 1)$?

Definimos $Y = X_1 + X_2 - X_3 \longrightarrow P(X_1+X_2 \geqslant X_3+1) = p(Y \geqslant 1)$

Y tiene distribución $N(\mu_y, \sigma_y^2)$

$$Y = \underbrace{(1\ 1\ {-1})}_{a^t}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}, \quad \mu_y = a^t\bar{X} = (1\ 1\ {-1})\begin{pmatrix}10\\20\\30\end{pmatrix} = 0$$

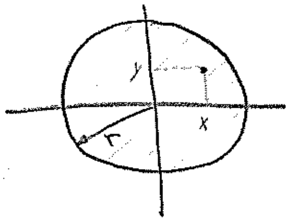
$$\sigma_y^2 = a^t M a = (1\ 1\ {-1})\begin{pmatrix}1&0&0\\0&1&0\\0&0&4\end{pmatrix}\begin{pmatrix}1\\1\\-1\end{pmatrix} = (1\ 1\ {-4})\begin{pmatrix}1\\1\\-1\end{pmatrix} = 6$$

$$p(X_1 + X_2 \geqslant X_3 + 1) = p(X_1 + X_2 - X_3 \geqslant 1) = p(Y \geqslant 1 \mid Y \to N(0, \sqrt{6})) =$$

$$= p\left(Z \geqslant \frac{1-0}{\sqrt{6}}\right) = p\left(Z \geqslant \frac{1}{\sqrt{6}}\right) = \boxed{1 - \Phi\left(\frac{1}{\sqrt{6}}\right)}$$

## 13 EJERCICIOS

Diana circular con centro en el origen de coordenadas y radio r. X, Y coordenadas de un punto elegido al azar. Cualquier otro punto de la diana tiene la misma probabilidad de ser elegido. Se pide $f_{xy}(x,y)$ y $f_x(x)$.



$$f_{xy}(x,y) = \begin{cases} C & x^2+y^2 \leq r^2 \\ 0 & \text{en el resto} \end{cases} \longrightarrow \boxed{f_{xy}(x,y) = \begin{cases} 1/\pi r^2 & x^2+y^2 \leq r \\ 0 & \text{en el resto} \end{cases}}$$

$$\iint f_{xy}(x,y)\,dx\,dy = 1 \longrightarrow \pi r^2 C = 1; \quad C = \frac{1}{\pi r^2}$$

$$\boxed{f_x(x) = \int_{-\sqrt{r^2-x^2}}^{+\sqrt{r^2-x^2}} \frac{1}{\pi r^2}\,dy = \frac{2\sqrt{r^2-x^2}}{\pi r^2} \qquad -r \leq x \leq r}$$

② Almacén guarda cajas que contienen piezas de distinto tipo. Proporción p de piezas de tipo A es una v.a. con $f(p) = Kp(1-p) \quad 0 \leq p \leq 1$

a) Calcular K, la media y la varianza de p.

$$\int_0^1 f(p)\,dp = 1 \rightarrow \int_0^1 Kp(1-p)\,dp = K\left(p^2/_2 - p^3/_3\right)\Big|_0^1 = K\left(\frac{1}{2} - \frac{1}{3}\right) = 1$$

$$K\frac{3-2}{6} = 1, \quad \boxed{K = 6}$$

$$E[p] = \int_0^1 p\, f_p(p)\,dp = \int_0^1 6p^2(1-p)\,dp = \left(6p^3/_3 - 6\,p^4/_4\right)\Big|_0^1 = 6\left(\frac{1}{3} - \frac{1}{4}\right) =$$

$$= 6\,\frac{4-3}{12} = \frac{1}{2} \longrightarrow \boxed{E[p] = 1/_2}$$

$$Var[p] = E[p^2] - (E[p])^2 = \int_0^1 p^2 f_p(p)\,dp - \left(\frac{1}{2}\right)^2 = \int_0^1 p^2\, 6p(1-p)\,dp - \frac{1}{4} =$$

$$= 6\left(\frac{p^4}{4} - \frac{p^5}{5}\right)\Big|_0^1 - \frac{1}{4} = \frac{6}{4} - \frac{6}{5} - \frac{1}{4} = \frac{25-24}{20} = \frac{1}{20}, \quad \boxed{Var[p] = \frac{1}{20}}$$

b) Se toman 10 cajas al azar ¿Cual es la probabilidad de que ninguna de ellas contenga una prop. de piezas de tipo A igual o sup. al 75%?

X v.a nº cajas con $p_A \geq 0{,}75$ de un total de 10

$X \rightarrow B(n=10, p')$

$p' = p(\text{en una caja halla } 75\% \overset{\text{o mas}}{\text{piezas}} \text{ tipo } A) = P(P \geq 0{,}75) =$

$$= 1 - P(p < 0{,}75) = 1 - \int_0^{0{,}75} f_P(p)\,dp = 1 - \int_0^{0{,}75} 6p(1-p)\,dp =$$

$$= 1 - 6\left(\frac{p^2}{2} - \frac{p^3}{3}\right)\Big|_0^{0{,}75} = 1 - 6\left(\frac{0{,}75^2}{2} - \frac{0{,}75^3}{2}\right) = \frac{5}{32}$$

$$X \rightarrow B\left(n=10, p' = \frac{5}{32}\right)$$

$$P\left(X=0 \mid X \rightarrow B\left(n=10, p'=\frac{5}{32}\right)\right) = \binom{10}{0}\left(\frac{5}{32}\right)^0\left(1 - \frac{5}{32}\right)^{10-0} = \boxed{0{,}18}$$

(4.3) X, Y v.a. independientes con la misma distribución F. Calcular la función de densidad $U = \max(X, Y)$

$$F_x(x) = F_y(y) = F \longrightarrow f_x(x) = f_y(y) = f$$

$$F' = f$$

$$F_u(u) = p(U \leq u) = P(\max(X,y) \leq u) = P(x \leq u, y \leq u) \overset{\text{indep}}{=}$$

$$= p(X \leq u)\, p(y \leq u) = F(u)\, F(u) = F^2(u)$$

$$f_u(u) = \frac{dF^2(u)}{du} = 2F(u)F'(u) \underset{\substack{\uparrow \\ F' = f}}{=} 2F(u)f(u)$$

$$\boxed{f_u(u) = 2F(u)f(u)}$$

(4.4) Obtener la distribución de probabilidad del máximo, el mínimo y la media de los resultados obtenidos al lanzar dos dados. Los resultados de los dados son v.a. independientes.

X v.a. resultado del dado 1º $\{1, 2, 3, 4, 5, 6\}$ > Los resultados son equiprobables → las v.a. son uniformes discretas

Y v.a. resultado del dado 2º $\{1, 2, 3, 4, 5, 6\}$

Calculemos $F_x(X)$, $F_y(y)$

| X | $P(X \leq x)$ |
|---|---|
| 1 | 1/6 |
| 2 | 2/6 |
| 3 | 3/6 |
| 4 | 4/6 |
| 5 | 5/6 |
| 6 | 1 |

$$P(X \leq x) = \frac{X}{6} \longrightarrow \begin{cases} F_x(X) = \dfrac{X}{6};\ X = 1,2,3,4,5,6 \\ F_y(y) = \dfrac{y}{6};\ y = 1,2,3,4,5,6 \end{cases}$$

$U = \max(X, y)$, se pide $p(U = u) = p(U \leq u) - p(U \leq u-1) = F_u(u) - F_u(u-1)$

Calculemos $F_u(u) = p(U \leq u) = p(\max(X,y) \leq u) = p(X \leq u, y \leq u) \underset{\substack{\uparrow \\ \text{indep}}}{=} p(X \leq u)\, p(y \leq u) =$

$$= F_x(u) F_y(u) = \left(\frac{u}{6}\right)^2 \longrightarrow F_u(u) = \left(\frac{u}{6}\right)^2;\ u = 1,2,3,4,5,6$$

$$P(U = u) = F_u(u) - F_u(u-1) = \left(\frac{u}{6}\right)^2 - \left(\frac{u-1}{6}\right)^2 = \frac{u^2 - u^2 - 1 + 2u}{36} = \frac{2u-1}{36}$$

$$\boxed{P(U = u) = \frac{2u-1}{36},\quad u = 1,2,3,4,5,6}$$

$$V = \min(X,y) \longrightarrow F_v(v) = p(V \leq v) = p(\min(X,y) \leq v) = 1 - p(\min(X,y) > v) =$$

$$= 1 - p(X > v, y > v) \underset{\substack{\uparrow \\ \text{indep}}}{=} 1 - p(X > v)\, p(y > v) = 1 - (1 - p(X \leq v))(1 - p(y \leq v)) =$$

$$= 1 - (1 - F_x(v))(1 - F_y(v)) = 1 - \left(1 - \frac{v}{6}\right)^2 \Longrightarrow F_v(v) = 1 - \left(1 - \frac{v}{6}\right)^2 \quad v = 1,2,3,4,5,6$$

$$P(V=v)=P(V\leq v)-P(V\leq v-1)=F_V(v)-F_V(v-1)=$$

$$=\left(\cancel{1}-\left(1-\frac{v}{6}\right)^2\right)-\left(\cancel{1}-\left(1-\frac{v-1}{6}\right)^2\right)=\left(\frac{6-v+1}{6}\right)^2-\left(\frac{6-v}{6}\right)^2=\frac{49+\cancel{v^2}-14v-36-\cancel{v^2}+12v}{36}=$$

$$=\frac{13-2v}{36}\longrightarrow\boxed{P(V=v)=\frac{13-2v}{36},\ v=1,2,3,4,5,6}$$

$$M=\frac{1}{2}(x+y)$$

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1 | 3/2 | 2 | 5/2 | 3 | 7/2 |
| 2 | 3/2 | 2 | 5/2 | 3 | 7/2 | 4 |
| 3 | 2 | 5/2 | 3 | 7/2 | 4 | 9/2 |
| 4 | 5/2 | 3 | 7/2 | 4 | 9/2 | 5 |
| 5 | 3 | 7/2 | 4 | 9/2 | 5 | 11/2 |
| 6 | 7/2 | 4 | 9/2 | 5 | 11/2 | 6 |

→ Valores de la media

De la tabla deducimos

$$\begin{cases} P(M=m)=\frac{1}{36};\ m=1,6 \\ P(M=m)=\frac{2}{36},\ m=\frac{3}{2},\frac{11}{2} \\ P(M=m)=\frac{3}{36};\ m=2,5 \\ P(M=m)=\frac{4}{36};\ m=\frac{5}{2},\frac{9}{2} \\ P(M=m)=\frac{5}{36};\ m=3,4 \\ P(M=m)=\frac{6}{36};\ m=\frac{7}{2} \end{cases}$$

**4.5** Empresa fabrica papel en planchas de 1×1 $m^2$. Se mide la limpieza en nº impurezas/$m^2$. Linea 1: 5 impurezas/$m^2$. Linea 2: 3 impurezas/$m^2$
El nº impurezas por plancha es v.a. Poisson. Se empaquetan las planchas en packs de 2000 unidades, y se clasifican segun la linea de procedencia. Si se encuentra un pack sin clasificar se toma una muestra aleatoria de 10 planchas y se determina el nº impurezas. Se le asigna el grupo de la linea 1 si el nº impurezas medio es mayor que cuatro, y la linea 2 en caso contrario.

a) Calcular la probabilidad de clasificar erroneamente un pack

L1 v.a nº impurezas en planchas de L1 (1$m^2$) ⟶ $\lambda_I=5$ imp/$m^2$

L2 v.a nº impurezas en planchas de L2 (1$m^2$) ⟶ $\lambda_{II}=3$ imp/$m^2$

X nº impurezas en 10 planchas
- → Si planchas $\in I$ ⟶ $\lambda_I'=50$ imp/10 planch. $\quad p(I)=1/2$
- → Si planchas $\in II$ ⟶ $\lambda_{II}'=30$ imp/10 planch. $\quad p(II)=1/2$

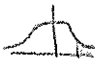
$$p(\text{clasificar erroneamente})=p(\text{clas. }I\mid II)\cdot p(II)+p(\text{clas. }II\mid I)\,p(I)$$

$$p(\text{clasificar en }I)=p(X>4\cdot 10\text{ planchas});\qquad p(I)=\tfrac{1}{2}$$

$$p(\text{clasificar en }II)=p(X\leq 4\cdot 10\text{ planchas}),\qquad p(II)=\tfrac{1}{2}$$

$p(\text{clasificar I} \mid \text{II}) = p(X > 40 \mid X \rightarrow \text{Poisson}(\lambda'_{II} = 30) \simeq p(X \geqslant 41 \mid X \rightsquigarrow N(30, \sqrt{30})) =$

$= p\left(Z \geqslant \frac{41-30}{\sqrt{30}}\right) = 1 - p(Z < 2{,}00) = 1 - \Phi(2) = 1 - 0{,}9772 = 22{,}8\ 10^{-3}$



$p(\text{clas II} \mid \text{I}) = p(X \leq 40 \mid X \rightarrow \text{Poisson}(\lambda'_{I} = 50) \simeq p(X \leq 40 \mid X \rightsquigarrow N(50, \sqrt{50})) =$

$= p\left(Z \leq \frac{40-50}{\sqrt{50}}\right) = p(Z \leq -1{,}41) = p(Z \geqslant 1{,}41) = 1 - p(Z \leq 1{,}41) = 1 - \phi(1{,}41) =$

$= 1 - 0{,}9207 = 0{,}0793$

$\boxed{p(\text{clasificar erroneamente}) = 22{,}8 \cdot 10^{-3} \cdot \frac{1}{2} + 0{,}0793 \cdot \frac{1}{2} = 0{,}05105}$

**4.6** La función de densidad de una v.a. bidimensional viene dada por la expresión:

$$f_{xy}(x,y) = \begin{cases} x \cdot y + ce^{x}, & \text{si } 0 < x < 1 \text{ y } 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

¿Son independientes las variables X e y?

X e y serán independientes si se cumple la condición $f_{xy}(x,y) = f_x(x)\, f_y(y)$

$$f_x(x) = \int_0^1 (xy + ce^{x})dy = x\frac{y^2}{2} + ce^{x}y \Big|_0^1 = \frac{x}{2} + ce^{x}, \quad 0 < x < 1$$

$$f_y(y) = \int_0^1 (xy + ce^{x})dx = \frac{x^2}{2}y + ce^{x}\Big|_0^1 = \frac{y}{2} + ce^{e} - c = \frac{y}{2} + c(e-1), \quad 0 < y < 1$$

Como $f_{xy}(x,y) \neq f_x(x)\, f_y(y) \longrightarrow$ $\boxed{\text{X e Y no son v.a. independientes}}$

b) Caso concreto n° impurezas en planchas ha sido: 1, 2, 3, 3, 4, 4, 5, 5, 6, 8. Calcular la prob. de que pertenezca a cada una de las líneas.

$n^{\circ}$ medio de impurezas $= \frac{1+2+3+3+4+4+5+5+6+8}{10} = 4{,}1$ impurezas/plancha $\rightarrow$ 41 imp/10 planchas

$$\underbrace{p(\text{I} \mid X = 41)}_{p(\text{pertenecer a I})} = \frac{p(X=41 \mid \text{I})\, p(\text{I})}{p(X=41 \mid \text{I})\, p(\text{I}) + p(X=41 \mid \text{II})\, p(\text{II})}$$

$p(X = 41 \mid X \rightarrow \text{Poisson}(\lambda'_I = 50\,\text{imp}/10) = \frac{50^{41}}{41!} e^{-50} = \underline{0{,}026}$

$p(X = 41 \mid X \rightarrow \text{Poisson}(\lambda'_{II} = 30\,\text{im}/10) = \frac{30^{41}}{41!} e^{-30} = 0{,}\underline{01}$

$\boxed{p(\text{pertenecer a I})} = \frac{0{,}026 \cdot \frac{1}{2}}{0{,}026 \cdot \frac{1}{2} + 0{,}01 \frac{1}{2}} = \boxed{0{,}7198}$

$\boxed{p(\text{pertenecer a II})} = \frac{p(X=41 \mid \text{II})\, p(\text{II})}{p(X=41 \mid \text{II})\, p(\text{II}) + p(X=41 \mid \text{I})\, p(\text{I})} = \frac{0{,}01 \cdot \frac{1}{2}}{0{,}01 \frac{1}{2} + 0{,}026 \frac{1}{2}} =$

$= \boxed{0{,}2802}$

(4.7) Billetes se fabrican en pliegos. Dos máquinas iguales, una imprime el anverso y la otra el reverso. Sean X e Y el nº defectos en el anverso y reverso respectivamente. Ambas variables son indep. con distribución de Poisson $(\lambda_1, \lambda_2)$

a) Demostrar que el nº total de defectos en un pliego $Z = X + Y$ tiene distribución de Poisson.

Nota: $P(Z=n) = \sum_{K=0}^{n} P(X=K)P(Y=n-K)$ y desarrollo del Binomio para $(\lambda_1+\lambda_2)^n$.

$$P(Z=n) = \sum_{K=0}^{n} P(X=K)P(Y=n-K) \overset{\substack{X e Y \to \text{Poisson / Independientes} \\ \downarrow}}{=} \sum_{K=0}^{n} \frac{\lambda_1^K}{K!}e^{-\lambda_1} \cdot \frac{\lambda_2^{n-K}}{(n-K)!}e^{-\lambda_2} = e^{-\lambda_1}e^{-\lambda_2}\sum_{K=0}^{n} \frac{\lambda_1^K \lambda_2^{n-K}}{K!(n-K)!} =$$

$$= \frac{e^{-\lambda_1}e^{-\lambda_2}}{n!} \underbrace{\sum_{K=0}^{n} \underbrace{\frac{n!}{K!(n-K)!}}_{\binom{n}{K}} \lambda_1^K \lambda_2^{n-K}}_{(\lambda_1+\lambda_2)^n} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n \longrightarrow \text{Es una Poisson con } \lambda = \lambda_1+\lambda_2$$

NOTA!! Cuando sumamos v.a. de Poisson independientes obtenemos otra v.a. de Poisson que tiene como $\lambda$ la suma de los $\lambda$ de las v.a. Poisson iniciales.

b) Si el nº total de defectos del pliego es $Z = n$ ¿Cuál es la prob. de que halla exactamente $X = K$ defectos en el anverso? (Resultado en función de: $\lambda_1$, $\lambda_2$ n K
¿De que distribución de probabilidad se trata?

$$P(X=K \mid Z=n) = \frac{P(X=K.\ Z=n)}{P(Z=n)} = \frac{P(X=K, Y=n-K)}{P(Z=n)} \overset{\text{independencia}}{=}$$

$$= \frac{\cancel{e^{-\lambda_1}}\frac{\lambda_1^K}{K!} \cdot \cancel{e^{-\lambda_2}}\frac{\lambda_2^{n-K}}{(n-K)!}}{\frac{\cancel{e^{-(\lambda_1+\lambda_2)}}}{n!}(\lambda_1+\lambda_2)^n} = \frac{n!}{(n-K)!\,K!}\ \frac{\lambda_1^K \lambda_2^{n-K}}{(\lambda_1+\lambda_2)^n} = \underbrace{\binom{n}{K}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^K \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-K}}_{\binom{n}{K}p^K(1-p)^{n-K}}$$

Se trata de una Binomial de parámetros $\boxed{n,\ p = \frac{\lambda_1}{\lambda_1+\lambda_2}}$

(4.8) La cantidad en miligramos de dos componentes contenidos en un producto es una v.a. bidimensional cuya función de densidad es:

$$f_{XY}(x,y) = \begin{cases} 4xy, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{en el resto} \end{cases}$$

Calcular la probabilidad de que la cantidad del 1er componente sea menor que 0,3 mg cuando la del 2º es 0,8 mg.

Nos piden $P(X < 0{,}3 \mid Y = 0{,}8)$

Necesitamos la función de densidad condicionada $f_{X|Y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)}$

Calculamos $f_y(y) = \int_0^1 4xy\,dx = 4\frac{x^2}{2}\Big|_0^1 y = 4y\frac{1}{2} = 2y \rightarrow f_{X|Y}(x|y) = \frac{4xy}{2y} = 2x$

que coincide con la marginal de $X \rightarrow \left(f_x(x) = \int_0^1 4xy\,dy = 4x\frac{1}{2} = 2x\right) \rightarrow$ Ambas v.a son independientes

Por tanto $f_{x|y}(X|Y) = f(x) = 2x$ y además $P(X<0{,}3 \mid Y=0{,}8) = P(X<0$

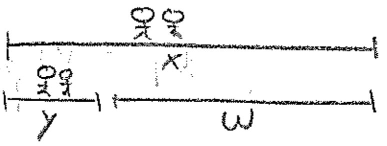
$$= \int_0^{0{,}3} 2x\,dx = 2\frac{(0{,}3)^2}{2} = \boxed{0{,}09}$$

(4.9) La llegada de clientes a un banco se considera un proceso de Poisson de parámetro $\lambda$. En la última hora llegaron 2 clientes ¿p( los dos entraran en los primeros 15 min)?

$$e^{-\lambda}\frac{\lambda^x}{x!}$$

X v.a nº de clientes que llegan a un banco en 1h → Poisson($\lambda$)

Y v.a " " " " " " " " " 15 mins → Poisson($\lambda/4$)



W v.a. nº clientes que llegan en 45 mins → Poisson($\frac{3}{4}\lambda$)

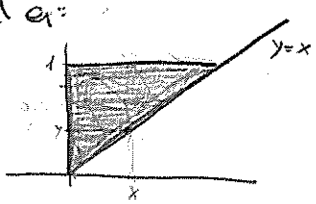
→ W, Y independientes

$$P(Y=2 \mid X=2) = \frac{P(Y=2, X=2)}{P_0(X=2)} = \frac{P(Y=2, W=0)}{P(X=2)} = \frac{P(Y=2)\,P(W=0)}{P(X=2)} =$$

$$= \frac{e^{-\frac{\lambda}{4}}\frac{\left(\frac{\lambda}{4}\right)^2}{2!}\cdot e^{-\frac{3\lambda}{4}}\frac{\left(\frac{3}{4}\lambda\right)^0}{0!}}{e^{-\lambda}\frac{\lambda^2}{2!}} = \frac{\cancel{e^{-\lambda}}\left(\frac{\cancel{\lambda}}{4}\right)^2\cancel{/2!}}{\cancel{e^{-\lambda}}\frac{\cancel{\lambda^2}}{\cancel{2!}}} = \left(\frac{1}{4}\right)^2 ; \boxed{P(Y=2 \mid X=2) = \left(\frac{1}{4}\right)^2}$$

(4.10) La función densidad de una v.a bidimensional es:

$$f_{xy}(x,y) = \begin{cases} Kxy, & 0<x<y<1 \\ 0, & \text{en el resto} \end{cases}$$



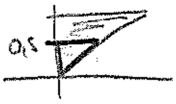
a) Calcular el valor de K.

$$\int_0^1\int_0^y Kxy\,dx\,dy = \int_0^1 Ky\frac{y^2}{2}\,dy = K\frac{y^4}{8}\Big|_0^1 = \frac{K}{8} = 1 ;\ K = 8$$

b) Calcular $P(X<0.5 \mid Y=0{,}5)$

Necesitamos $f_{x|y}(X|Y) = \frac{f_{xy}(x,y)}{f_y(y)} = \frac{8xy}{f_y(y)} = \frac{8xy}{4y^3} = \frac{2x}{y^2}$

$$f_y(y) = \int_0^y 8xy\,dx = 8\frac{y^2}{2}y = 4y^3 \quad 0<y<1$$



Para el caso $y=0{,}5$ → $f(X \mid y=0{,}5) = \frac{2x}{(0{,}5)^2} = 8x \quad 0<x<0{,}5$

Por tanto $p(X<0{,}5 \mid Y=0{,}5) = \int_0^{0.5} 8x\,dx = \frac{8\cdot 0{,}5^2}{2} = \boxed{1}$

Completamente razonable ya que si $Y=0{,}5$ es seguro que $X<0{,}5$

c) ¿Son independientes X e Y?

Serán independientes si y solo si $f_{xy}(x,y) = f_x(x) f_y(y)$

$$f_{xy}(x,y) = 8xy \quad 0<x<y<1$$

$$f_y(y) = 4y^3 \quad 0<y<1$$

$$f_x(x) = \int_x^1 8xy\,dy = 8x\left(\frac{y^2}{2}\right)_x^1 = 8x\left(\frac{1}{2}-\frac{x^2}{2}\right)$$

} No se cumple la condición de independencia

(4.14) La función de densidad conjunta de las v.a. $Y_1$, $Y_2$ es:

| | | $Y_1$ | | |
|---|---|---|---|---|
| | | -1 | 0 | 1 |
| $Y_2$ | -1 | 1/16 | 3/16 | 1/16 |
| | 0 | 3/16 | 0 | 3/16 |
| | 1 | 1/16 | 3/16 | 1/16 |

Calcular su coeficiente de correlación e indicar si son independientes.

| $Y_1$ | $P(Y_1=y)$ |
|---|---|
| -1 | 5/16 |
| 0 | 6/16 |
| 1 | 5/16 |

| $Y_2$ | $P(Y_2=y)$ |
|---|---|
| -1 | 5/16 |
| 0 | 6/16 |
| 1 | 5/16 |

Si son independientes $p(Y_1=k, Y_2=m) = P(Y_1=k)P(Y_2=m)$

$$p(Y_1=0, Y_2=0) = 0$$

$$p(Y_1=0)\,p(Y_2=0) = \frac{6}{16}\cdot\frac{6}{16} = \left(\frac{6}{16}\right)^2$$

→ $\boxed{\text{No son independientes}}$

Calulo de $\rho_{Y_1Y_2} = \dfrac{cov(y_1,y_2)}{\sigma_{y_1}\,\sigma_{y_2}}$

$$cov(y_1,y_2) = E[y_1y_2] - E[Y_1]E[Y_2]$$

$$E[Y_1] = \sum_{y=-1}^{1} Y_1 P(Y_1=y) = -1\,\frac{5}{16} + 0\cdot\frac{6}{16} + 1\,\frac{5}{16} = 0$$

$$E[Y_2] = 0$$

$$E[Y_1Y_2] = ??$$

$Y_1 \cdot Y_2$

$-1 \longrightarrow p(Y_1=-1, y_2=1) + p(Y_1=1, Y_2=-1) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16}$

$0 \longrightarrow p(Y_1=-1, y_2=0) + p(Y_1=0, Y_2=0) + p(Y_1=1, y_2=0) + p(Y_1=0, y_2=-1) + p(Y_1=0, Y_2=1) = \frac{3}{16} + 0 + \frac{3}{16} + \frac{3}{16} + \frac{3}{16} = \frac{12}{16}$

$1 \longrightarrow p(Y_1=-1, y_2=-1) + p(Y_1=1, Y_2=1) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16}$

$$E[Y_1Y_2] = \sum_{y=-1}^{1} Y_1Y_2\,(p(Y_1Y_2)=y) = -1\,\frac{2}{16} + 0\cdot\frac{12}{16} + 1\,\frac{2}{16} = 0$$

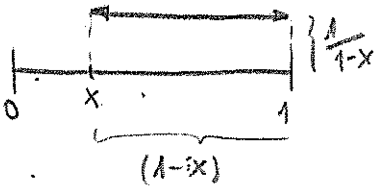
$$cov(y_1,y_2) = 0 \longrightarrow \boxed{\rho_{Y_1Y_2} = 0}$$

**NOTA!!** Hemos demostrado que, aunque el coeficiente de correlación sea cero no tienen porque ser independientes las variables.

(4.17) Sea X un valor elegido al azar de una distribución uniforme en [0,1]. A continuación se toma otro valor Y de la distribución uniforme [X, 1]. Calcular $f_Y(y)$

X v.a – uniforme [0,1] $\longrightarrow f_X(x) = \begin{cases} 1 & 0\le x\le 1 \\ 0 & \text{en el resto}\end{cases}$

Y uniforme en [X, 1]



$$f_{Y|X}(Y|X) = \begin{cases} \frac{1}{1-X} \; ; & X\le y\le 1 \\ 0 \; , & \text{en el resto}\end{cases}$$

$$f_{Y|X}(Y|X) = \frac{f_{XY}(X,Y)}{f_X(x)} \; ; \quad f_{XY}(X,Y) = f_{Y|X}(Y|X)\, f_X(x) = \frac{1}{1-X}\cdot 1$$

$$f_{XY}(X,Y) = \frac{1}{1-X} \qquad 0\le X\le y\le 1$$

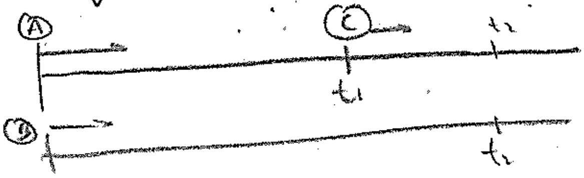
$$f_Y(y) = \int_0^y \frac{1}{1-x}\,dx = -\ln(1-X)\Big|_0^y = -\ln(1-y)+\ln(1) = -\ln(1-y)$$

$$\boxed{f_Y(y) = -\ln(1-y) \qquad 0\le y\le 1}$$

(4.19) Oficina de correos con 2 ventanillas. ABC llegan en el mismo instante y las ventanillas están desocupadas. Los tiempos de servicio requeridos por las 3 personas son v.a. exponenciales de parámetro λ. A y B comienzan de inmediato y C espera. ¿p(C no sea el último en salir)?

$T_A$ v.a. tiempo de A en la ventanilla } exponenciales (λ)
$T_B$ v.a " " B " " "
$T_C$ v.a " " C " " "

Supongamos que A es el primero en salir → B o C serán los último



$$p(\text{B sea el último}) = p(T_B > t_2 \mid T_B > t_1) = \frac{p(T_B > t_2, T_B > t_1)}{p(T_B > t_1)} = \frac{p(T_B > t_2)}{p(T_B > t_1)} = \frac{e^{-\lambda t_2}}{e^{-\lambda t_1}} = e^{-\lambda(t_2-t_1)}$$

$$p(\text{C sea el último}) = p(T_c + t_1 > t_2) = p(T_C > t_2 - t_1) = e^{-\lambda(t_2-t_1)}$$

→ Es igualmente probable que C o B sean últimos habiendo salido A el primero

Secuencias ABC equiprobables
$\boxed{ACB}$
De igual forma $\boxed{BCA}$ serán equiprobables
BAC
} 4 casos posibles equiprobables

$$p(\text{C no sea el último}) = \frac{2}{4} = \frac{1}{2} = \boxed{0{,}5}$$

Obvio, ya que la exponencial no tiene memoria

**4.20** La estatura de los ciudadanos varones sigue una distribución normal $N(\mu = 175\,cm, \sigma = 5\,cm)$. Se seleccionan al azar 100 ciudadanos ¿cuál es la probabilidad de que al menos 30 superen los 180 cm?

$X$ v.a nº ciudadanos que superan los 180 cm de un total de 100

$$\left.\begin{array}{l} \rightarrow B(n=100, p) \\ E \rightarrow N(175, 5) \end{array}\right\} \rightarrow p = p(E > 180 \mid E \rightarrow N(175,5)) = p\left(Z > \frac{180-175}{5}\right) =$$

$$= p(Z > 1) = 1 - \Phi(1) = 1 - 0,8413 = 0,1587$$

$$p(X \geq 30 \mid X \rightarrow B(n=100, p=0,1587)) \underset{\uparrow}{\simeq} p\left(X \geq 30 \mid X \rightsquigarrow N\left(15,87, \underbrace{\sqrt{15,87(0,8413)}}_{3,65}\right)\right) =$$

aproximamos a la normal sin corrección por cont.

$$= p\left(Z \geq \frac{30 - 15,87}{3,65}\right) = p(Z \geq 3,87) = 1 - \Phi(3,87) \simeq \underline{\underline{0}}$$

EXTRA en pág. 25

**4.11** $X$ e $Y$ son v.a con coeficiente de correlación $\rho = -1$. Si las varianzas son iguales, calcular la varianza de $Z = X + Y - 1$

$$Var[Z] = Var[X + Y - 1] = Var[X] + Var[Y] + 2\,cov(X, Y) =$$

$$cov(X,Y) = \rho \cdot \sigma_X \sigma_Y \underset{\uparrow}{=} \rho \sigma^2 = -\sigma^2$$

$\sigma_X = \sigma_Y = \sigma = \sqrt{Var[X]} = \sqrt{Var[Y]}$

$$\Rightarrow Var[Z] = \sigma^2 + \sigma^2 + 2(-\sigma^2) = \underline{\underline{0}}$$

Otra forma: $\rho_{XY} = -1 \rightarrow$ Linealidad perfecta negativa $\Rightarrow Y = a + bX$, $b < 0$

$$Var(Y) = b^2 Var(X) \xrightarrow{Var(Y) = Var(X)} b^2 = 1 \underset{\substack{\uparrow \\ b<0}}{\longrightarrow} b = -1 \longrightarrow Y = a - X, \quad X + Y = a$$

$$Z = X + Y - 1 = a - 1 \longrightarrow Var(Z) = Var(a-1) = \underline{\underline{0}}$$

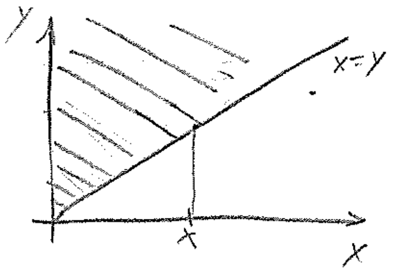
**4.12** Equipo de radio con dos partes: receptor y amplificador. La duración del receptor es una v.a. exponencial de media 500 h y la duración del amplificador es una v.a. exponencial de media 1000 h ¿Cuál es la probabilidad de que el fallo del equipo sea debido al receptor? (Las variables son independientes)

$X$ v.a tiempo del receptor $\quad f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0$; $E[X] = \frac{1}{\lambda} = 500\,h$

$Y$ v.a tiempo del ampli. $\quad f_Y(y) = \mu e^{-\mu y} \quad y \geq 0$; $E[Y] = \frac{1}{\mu} = 1000\,h$

Por ser $X$ e $Y$ independientes $\quad f_{XY}(x,y) = \lambda \mu e^{-\lambda x - \mu y}$, $x \geq 0 \; y \geq 0$

Se pide $p(Y \geq X) = p(Y-X \geq 0) = \iint\limits_{Y-X\geq 0} f_{x,y}(x,y)\,dx\,dy =$



$$= \int_{\underset{y}{0}}^{\infty} \int_{\underset{x}{0}}^{y} \lambda\mu e^{-\lambda x-\mu y}\,dx\,dy = \int_0^\infty \lambda\mu e^{-\mu y} \int_0^y e^{-\lambda x}\,dx\,dy =$$

$$= \int_0^\infty \lambda\mu e^{-\mu y}\left(\frac{e^{-\lambda x}}{-\lambda}\Big|_0^y\right) dy = \int_0^\infty -\mu e^{-\mu y}(e^{-\lambda y}-1)\,dy =$$

$$= \int_0^\infty (-\mu e^{-\mu y-\lambda y} + \mu e^{-\mu y})\,dy = \left(-\mu \frac{e^{-\mu y-\lambda y}}{-\mu-\lambda} - e^{-\mu y}\right)\Bigg|_0^\infty = 1 - \frac{\mu}{\mu+\lambda} \underset{\substack{\mu=\frac{1}{1000}\\ \lambda=\frac{1}{500}}}{\uparrow}=$$

$$= 1 - \frac{\frac{1}{1000}}{\frac{1}{1000}+\frac{1}{500}} = 1 - \frac{\frac{1}{1000}}{\frac{3}{1000}} = \boxed{\frac{2}{3}}$$

**4.13** Sea $X_1$ v.a $N(10,1)$, $X_2 \to N(20,1)$, $X_3 \to N(3,4)$ Se define

$Z_1 = X_1 + X_2 - X_3$ ; $Z_2 = X_1 + X_2 + X_3$ ; $Z_3 = X_1 - X_2 - X_3$

Calcular la matriz de varianzas de $(Z_1\ Z_2\ Z_3)$ si $X_1, X_2, X_3$ son indep.

$$\underbrace{\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}}_{Z} = \underbrace{\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}}_{X} ; \quad Var[Z] = A \overset{\text{por indep}}{M_x} A^T = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & 1 & -4 \\ 1 & 1 & 4 \\ 1 & -1 & -4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 6 & -2 & 4 \\ -2 & 6 & -4 \\ 4 & -4 & 6 \end{pmatrix} ; \quad \boxed{Var[Z] = \begin{pmatrix} 6 & -2 & 4 \\ -2 & 6 & -4 \\ 4 & -4 & 6 \end{pmatrix}}$$

**4.15** La función de densidad conjunta $f_{xy}(x,y) = xy$ $\quad \begin{matrix} 0<x<1 \\ 0<y<2 \end{matrix}$

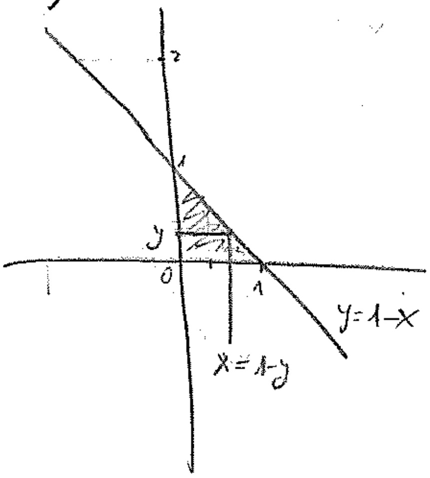
a) $f_x(x)$, $f_y(y)$ ¿Son X e Y independientes?

$f_x(x) = \int_0^2 xy\,dy = x\left(\frac{y^2}{2}\right)\Big|_0^2 = 2x \longrightarrow$ $\boxed{f_x(x) = 2x,\ 0<x<1}$

$f_y(y) = \int_0^1 xy\,dx = y\left(\frac{x^2}{2}\right)\Big|_0^1 = \frac{y}{2} \longrightarrow$ $\boxed{f_y(y) = \frac{y}{2},\ 0<y<2}$

Como $f_{x,y}(x,y) = xy = f_x(x)\,f_y(y) = 2x\,\frac{y}{2} = xy \longrightarrow$ $\boxed{\text{X e Y son independientes}}$

b) Calcular $p(X+y<1) = \iint\limits_{x+y<1} xy\,dx\,dy = \underset{Y}{\int_0^1}\underset{X}{\int_0^{1-y}} xy\,dx\,dy = \underset{Y}{\int_0^1} y\left(\frac{x^2}{2}\right)\Big|_0^{1-y} dy =$



$$= \int_0^1 y\,\frac{(1-y)^2}{2}\,dy = \frac{1}{2}\int_0^1 (y + y^3 - 2y^2)\,dy =$$

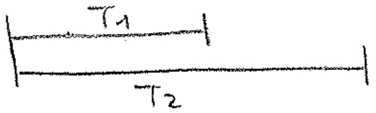
$$= \frac{1}{2}\left(\frac{y^2}{2}+\frac{y^4}{4}-2\frac{y^3}{3}\right)\Big|_0^1 = \frac{1}{2}\left(\frac{1}{2}+\frac{1}{4}-\frac{2}{3}\right) = \frac{1}{2}\,\frac{6+3-8}{12}$$

$$= \frac{1}{24} \;;\; \boxed{p(X+y<1) = \frac{1}{24}}$$

④.16 Un ordenador tarda un total de $T_2$ segundos en procesar un mensaje de correo electrónico. Esta cantidad incluye al tiempo $T_1$ durante el cual el mensaje está en la cola esperando a ser procesado $T_2 \geq T_1$.

$$f_{T_1,T_2}(T_1,T_2) = e^{-t_2} \quad 0 \leq t_1 \leq t_2 < \infty$$

Calcular la probabilidad de que un mensaje haya estado menos de 1 segundo en la cola si el tiempo total de su procesamiento ha sido mayor que 2 segundos.



$$p(T_1 \leq 1 \mid T_2 \geq 2) = \frac{p(T_1 \leq 1, T_2 \geq 2)}{p(T_2 \geq 2)}$$

Necesitamos $f_{T_2}(T_2) = \int_0^{t_2} e^{-t_2}\,dt_1 = t_2 e^{-t_2}$

$$p(T_2 \geq 2) = 1 - p(T_2 < 2) = 1 - \int_0^2 t_2 e^{-t_2}\,dt_2 = 1 - \left(t_2\frac{e^{-t_2}}{-1}\Big|_0^2 + \int_0^2 e^{-t_2}\,dt_2\right) =$$

$$= 1 - \left(-2e^{-2} + \left(-e^{-t_2}\right)\Big|_0^2\right) = 1 + 2e^{-2} + e^{-2} - 1 = \underline{\underline{3e^{-2}}}$$

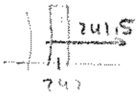
$$p(T_1 \leq 1, T_2 \geq 2) = \underset{T_2}{\int_2^\infty}\underset{T_1}{\int_0^1} e^{-t_2}\,dt_1\,dt_2 = \underset{T_2}{\int_2^\infty} e^{-t_2}\,dt_2 = \frac{e^{-t_2}}{-1}\Big|_2^\infty =$$

$$= \underline{\underline{e^{-2}}}$$

$$\Longrightarrow p(T_1 \leq 1 \mid T_2 \geq 2) = \frac{e^{-2}}{3e^{-2}} = \boxed{\frac{1}{3}}$$

**4.19** En uno fabrica 96% piezas resultan con longitudes admisibles, 3% defectuosas cortas y 1% defectuosas largas. Se pide:

a) Probabilidad de que en lote de 250 piezas sean admisibles 242 o más



X n° piezas admisibles de un total de 250: $X \to B(n=250, p=0{,}96)$

$$p(X \geq 242 \mid X \to B(n=250, p=0{,}96)) \underset{(*)}{\cong} p(X \geq 241{,}5 \mid X \rightsquigarrow N(\mu = \underbrace{250 \cdot 0{,}96}_{240}, \sigma = \underbrace{\sqrt{250 \cdot 0{,}96 \cdot 0{,}04}}_{3{,}098})) =$$

$$= p\left(Z \geq \frac{241{,}5 - 240}{3{,}098}\right) = p(Z \geq 0{,}48) = 1 - p(Z < 0{,}48) = 1 - 0{,}6844 = \boxed{0{,}3156}$$

(*) Con corrección por continuidad!!

b) En un lote de 500 sean cortas 10 o menos.

Y n° piezas cortas de un total de 500: $Y \to B(n=500, p=0{,}03)$ * Con correc. por cont.

$$p(Y \leq 10 \mid Y \to B(500, 0{,}03)) \underset{*}{\cong} p(Y \leq 10{,}5 \rightsquigarrow N(\mu = \underbrace{500 \cdot 0{,}03}_{15}, \sigma = \underbrace{\sqrt{500 \cdot 0{,}03 \cdot 0{,}97}}_{3{,}814})) =$$

$$= p\left(Z \leq \frac{10{,}5 - 15}{3{,}814}\right) = p(Z \leq -1{,}18) = 1 - p(Z < 1{,}18) = 1 - 0{,}8810 =$$

$$= \boxed{0{,}119}$$

c) En 1000 piezas haya entre 6 y 12 largas.

L n° piezas largas de un total de 100, $L \to B(n=1000, p=0{,}01)$

$$p(6 \leq L \leq 12 \mid L \to B(1000, 0{,}01)) \simeq p(5{,}5 \leq L \leq 12{,}5 \mid L \rightsquigarrow N(\mu = 10, \sigma = \sqrt{10 \cdot 0{,}99}) =$$

$$= p\left(\frac{5{,}5 - 10}{\sqrt{9{,}9}} \leq Z \leq \frac{12{,}5 - 10}{\sqrt{9{,}9}}\right) = p(-1{,}43 \leq Z \leq 0{,}79) =$$

$$= p(Z \leq 0{,}79) - p(Z > 1{,}43) = \phi(0{,}79) - (1 - \phi(1{,}43)) = 0{,}7852 - (1 - 0{,}9236) =$$

$$= \boxed{0{,}7088}$$

(*) EXTRA ejercicio ⑳: ¿Cuál es la probabilidad de que la estatura media de las 100 personas de la muestra sea mayor que 176?

Por teorema central del límite $\overline{E} \to N(\mu = 175, \sigma = \frac{5}{\sqrt{100}})$

$$p(\overline{E} > 176 \mid \overline{E} \to N(\mu = 175, \sigma = \frac{5}{\sqrt{100}})) = p\left(Z > \frac{176 - 175}{5/\sqrt{100}}\right) = p(Z > 2) =$$

$$= 1 - \phi(2) = \boxed{0{,}0228}$$

NOTA!! $\overline{E}$ tiene mucho menos dispersión que $E$

$\sigma_{\overline{E}} = 0{,}5 \quad \sigma_E = 5$

(4.21) Proceso de fabricación. Por término medio hay 1 defecto cada 20 m. Si la distribución de defectos es Poisson, calcular la probabilidad de 6 defectos en 200 m.

a) Directamente

X v.a. nº defectos en 20 m ⟶ $\lambda_x = 1 \text{ def}/20\text{m}$

Y v.a nº defectos en 200m ⟶ $\lambda_y = 10 \text{ def}/200\text{m}$

$$P(Y=6 \mid Y \to \text{Poisson}(\lambda_y = 10 \text{ def}/200\text{m})) = e^{-10}\frac{10^6}{6!} = \boxed{0{,}063}$$

b) Por aprox. a la normal

$$p(5{,}5 \le Y \le 6{,}5 \mid Y \sim N(\mu = 10, \sigma = \sqrt{10}) = p\left(\frac{5{,}5-10}{\sqrt{10}} \le Z \le \frac{6{,}5-10}{\sqrt{10}}\right) =$$

$$= p(-1{,}42 \le z \le -1{,}11) = p(z \le 1{,}42) - p(z \le 1{,}1) =$$

$$= 0{,}9222 - 0{,}8665 = \boxed{0{,}0557}$$

(4.21 bis) Para controlar la recepción de lotes de 10000 unidades de CDs se toma una muestra al azar de $n = 200$ discos clasificándolos en aceptables o no. Si el nº discos defectuosos es igual o inferior a 15 se acepta el lote. En caso contrario se rechaza.
$c=$

a) ¿Cuál es la probabilidad de aceptar un lote con el 12% discos defectuosos?

X v.a nº discos defectuosos de un total de 200

$X \to B(n=200, p=0{,}12)$

$$p(\text{aceptar lote}) = p(X \le 15 \mid X \to B(n=200, p=0{,}12)) \simeq$$

$$\simeq p(X \le 15{,}5 \mid X \sim N(\mu = 24, \sigma = \underbrace{\sqrt{12 \cdot 0{,}88}}_{\sqrt{21{,}12}})) = p\left(Z \le \frac{15{,}5-24}{\sqrt{21{,}12}}\right) =$$

$$= p(z \le -1{,}85) = 1 - \Phi(1{,}85) = 1 - 0{,}9678 = \boxed{0{,}032}$$

b) ¿Cuál es la prob. de rechazar un lote con el 5% defectuosos?

$Y \to B(n=200, p=0{,}05) \sim N(10, \sqrt{9{,}5})$

$$P(Y > 15 \mid Y \to B(100, 0{,}05)) = p(Y \ge 15{,}5 \mid Y \sim N(10, \sqrt{9{,}5}))$$

$$= p\left(Z \ge \frac{15{,}5-10}{\sqrt{9{,}5}}\right) = p(z \ge 1{,}78) = 1 - \Phi(1{,}78) = 1 - 0{,}9625 =$$

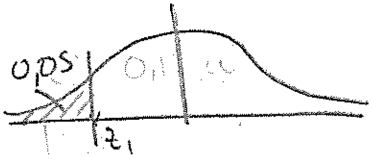
$$= \boxed{0{,}0375}$$

Muerteeeeee

c) ¿Que valores de n y c deben utilizarse si se desea que las prob. anteriores sean iguales a 0,05?

$$p(\underset{\text{con }12\%}{\text{aceptar}}) = p\left(X \leq c \mid X \to B(n, p=0{,}12)\right) \simeq p\left(X \leq c \mid X \to N\left(\mu = n\cdot 0{,}12, \sigma = \sqrt{n\cdot 0{,}12\cdot 0{,}88}\right)\right) =$$

$$= p\left(Z \leq \frac{c - n0{,}12}{\sqrt{0{,}1056\,n}}\right) = 0{,}05 \,;$$



$$z_1 = \frac{c - 0{,}12\,n}{\sqrt{0{,}1056\,n}}$$

$$p(\text{rechazar con }5\%) = p\left(Y > c \mid Y \to B(n, p=0{,}05)\right) \simeq p\left(Y > c \mid Y \sim N\left(\mu = n\cdot 0{,}05, \sigma = \sqrt{n\cdot 0{,}05\cdot 0{,}95}\right)\right) =$$

$$= p\Bigg(Z > \underbrace{\frac{c - 0{,}05\,n}{\sqrt{0{,}0475\,n}}}_{z_2}\Bigg) = 0{,}05$$

$$z_2 = \frac{c - 0{,}05\,n}{\sqrt{0{,}0475\,n}}$$

Por simetría de la gráfica de la normal y valores de tablas $\begin{cases} z_1 = -1{,}64 \\ z_2 = 1{,}64 \end{cases}$

$$\begin{cases} \dfrac{c - 0{,}12\,n}{\sqrt{66n}/25} = -1{,}64 & (1) \\ \dfrac{c - 0{,}05\,n}{\sqrt{19n}/20} = +1{,}64 & (2) \end{cases}$$

$$(1)+(2) \to \frac{c - 0{,}12\,n}{\sqrt{66n}/25} = \frac{0{,}05n - c}{\sqrt{19n}/20}$$

$$\frac{25}{\sqrt{66}}(c - 0{,}12n) = \frac{20}{\sqrt{19}}(0{,}05n - c)$$

$$\frac{25}{\sqrt{66}}c + \frac{20}{\sqrt{19}}c = \frac{20}{\sqrt{19}}\cdot 0{,}05n + \frac{25}{\sqrt{66}}\,0{,}12n$$

$$c = \frac{0{,}5987\,n}{7{,}665} = 0{,}078n \ (3)$$

$$\to \frac{0{,}078n - 0{,}12n}{\sqrt{66n}/25} = -1{,}64\,, \qquad \frac{0{,}0419\,\cancel{n}}{\frac{\sqrt{66}}{25}\sqrt{\cancel{n}}}\cdot\frac{\sqrt{n}}{\sqrt{n}} = 1{,}64\,,$$

$$\sqrt{n} = \frac{1{,}64\cdot\frac{\sqrt{66}}{25}}{0{,}0419}\,; \quad n = 161{,}84 \simeq 162$$

$$c = 0{,}078\cdot 162 = 12{,}62 \simeq 13$$

Las muestras deberían de ser de $n = 162$ CDs y el nº de discos con los que se acepta el lote debería ser igual o inferior a $c = 13$

**⓸.22** Sean X, Y, U y V v.a. Demostrar que si $Y = U + V \Rightarrow \operatorname{cov}(X,Y) = \operatorname{cov}(X,U) + \operatorname{cov}(X,V)$

$$\operatorname{cov}(X,Y) = E[(X-\mu_x)(Y-\mu_y)] \overset{\text{como } Y=U+V,\ \mu_y = \mu_u + \mu_v}{=} E[(X-\mu_x)(U+V-\mu_u-\mu_v)] =$$

$$= E[(X-\mu_x)(U-\mu_u) + (X-\mu_x)(V-\mu_v)] = E[(X-\mu_x)(U-\mu_u)] - E[(X-\mu_x)(V-\mu_v)] =$$

$$= \operatorname{cov}(X,U) + \operatorname{cov}(X,V) \qquad \text{CQD}$$

**4.23** $f_{xy}(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$

es la función de densidad de una normal bidimensional.

Si $\rho = 0,3$, se pide $p(Y \leq X+1)$

Sabemos que la función de densidad de una normal bidimensional es:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{1/2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right)\right\}$$

Identificando coeficientes

$X \longrightarrow N(0,1)$ $\qquad \rho = 0,3 \quad , \rho = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y} = 0,3$

$Y \longrightarrow N(0,1)$

$p(Y \leq X+1) = p(Y - X \leq 1)$, Llamaremos $W = Y - X \Longrightarrow$

$W = (-1,1)\begin{pmatrix} X \\ Y \end{pmatrix} = a^T \begin{pmatrix} X \\ Y \end{pmatrix}$

$\Longrightarrow E[W] = a^T E[X,Y] = (-1,1)\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$

$Var[W] = a^T \begin{pmatrix} 1 & 0,3 \\ 0,3 & 1 \end{pmatrix} a = (-1\ 1)\begin{pmatrix} 1 & 0,3 \\ 0,3 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = (-0,7\ \ 0,7)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1,4$

Tambiem $Var[W] = Var(x) + Var(y) - 2\,\text{cov}(x,y) = 1 + 1 + 2 \cdot 0,3 = 1,4$

$p(Y \leq X+1) = p(Y - X \leq 1) = p(W \leq 1 \mid N \longrightarrow (\mu = 0, \sigma = \sqrt{1,4}) =$

$= p\left(Z \leq \frac{1-0}{\sqrt{1,4}}\right) = p(Z \leq 0,85) = \boxed{0,8023}$

(4.25) Compuesto sometido durante 1h a una atmosf. de oxígeno a 200 °C
Medida de corrosión = ganancia de peso. Para un determinado compuesto, esta ganancia $X_1$, en 1h, se distribuye $\rightarrow N(100,5)$

a) Se realizan ensayos secuenciales ¿cuántos se tendrán que hacer por término medio hasta encontrar una probeta con una ganancia mayor que 105?

Y v.a. nº ensayos hasta encontrar probeta con ganancia $\geq 105$

$Y \rightarrow$ Geometrica $(p)$

$$p = p(X_1 > 105 \mid X_1 \rightarrow N(100,5)) = p\left(Z > \frac{105-100}{5}\right) = p(Z>1) = 1-\phi(1) =$$

$$= 1 - 0{,}8413 = 0{,}1587$$

Se pide $E[Y] = \frac{1}{p} = \frac{1}{0{,}1587} = 6{,}3 \longrightarrow$ Se esperarán hacer entre 6 y 7 ensayos

b) Sometemos la probeta al ensayo durante 2h. La ganancia en la 2ª h, $X_2$, es una normal $N(\mu=60, \sigma=5)$. $\rho_{x_1 x_2} = -0{,}28$.

Calcular la probabilidad de que una probeta tenga mayor ganancia en la 2ª h.

Se pide $p(X_2 > X_1) = p(\underbrace{X_2 - X_1}_{W} > 0)$

$W = X_2 - X_1$, $W \rightarrow N(\mu_W, \sigma_W)$ $\quad cov(X_1,X_2) = \rho \cdot \sigma_{x_1}\sigma_{x_2}$

$$\mu_W = \mu_{X_2} - \mu_{X_1} = 60 - 100 = -40$$

$$Var(W) = \sigma_{X_1}^2 + \sigma_{X_2}^2 - 2cov(X_1,X_2) = 5^2 + 5^2 - 2\cdot(-0{,}28)5\cdot 5 =$$

$$= 64$$

$W(-40,8)$

$$p(W>0) = p\left(Z > \frac{0-(-40)}{8}\right) = p(Z>5) \simeq \boxed{0}$$

c) Se forman índices de oxidación $Z_1 = X_1 + X_2$ y $Z_2 = X_1 - X_2$. Calcular $f_{Z_1,Z_2}(z_1,z_2)$ ¿Son independientes?

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \mu_x = \begin{pmatrix} 100 \\ 60 \end{pmatrix}, \quad S_x^2 = \begin{pmatrix} 25 & -0{,}28\cdot 5\cdot 5 \\ -0{,}28\cdot 5\cdot 5 & 25 \end{pmatrix}$$

$Z_1 \longrightarrow N(\mu_1, \sigma_1)$

$Z_2 \longrightarrow N(\mu_2, \sigma_2)$

$$\mu_Z = A\mu_X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 100 \\ 60 \end{pmatrix} = \begin{pmatrix} 160 \\ 40 \end{pmatrix} \begin{matrix} \rightarrow \mu_1 = 160 \\ \rightarrow \mu_2 = 40. \end{matrix}$$

$\sigma_1^2 = 36$
$\sigma_2^2 = 64$ $\rho = 0$

$$S_Z^2 = A S_x^2 A^T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 25 & -7 \\ -7 & 25 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 18 & 18 \\ 32 & -32 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 36 & 0 \\ 0 & 64 \end{pmatrix}$$

$$f(z_1,z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left(\left(\frac{z_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{z_2-\mu_2}{\sigma_2}\right)^2 - 2\left(\frac{z_1-\mu_1}{\sigma_1}\right)\left(\frac{z_2-\mu_2}{\sigma_2}\right)\right)\right)$$

$$f(z_1,z_2) = \frac{1}{2\pi 6\cdot 8} \exp\left(\frac{-1}{2}\left(\left(\frac{z_1-160}{6}\right)^2 + \left(\frac{z_2-40}{8}\right)^2 - 2\left(\frac{z_1-160}{6}\right)\left(\frac{z_1-40}{8}\right)\right)\right)$$

$$\boxed{\begin{array}{l} f_{z_1 z_2}(z_1, z_2) = \dfrac{1}{2\pi \cdot 48} \exp\left\{ \dfrac{-1}{2} \left( \left(\dfrac{z_1-160}{6}\right)^2 + \left(\dfrac{z_2-40}{8}\right)^2 - 2\left(\dfrac{z_1-160}{6}\right)\left(\dfrac{z_2-40}{8}\right) \right) \right\} \\ \text{Como } \rho = 0 \Longrightarrow \text{Son v. a. independientes.} \end{array}}$$

d) Si la garancia durante los dos primeras horas ha sido 170, ¿cuál es el valor medio de la garancia en la 1ª hora?

"No hacer, dificultad elevada"

J. Juan.

**4.26** Abastecimiento energético de una comarca depende de 3 centrales. $P_{nuclear} = 500$ MW, $P_{térmica} = 300$ MW, $P_{hidra.} = 200$ MW. Las centrales pueden estar disponibles (toda pot.) o averiadas ($P=0$)

En un día p(avería nuclear) = 0,10, p(av. térmica) = 0,12, p(av. hidr.) = 0,05

En un día la central no cambia de estado.

Averías indep. y a lo largo de un día la central no cambia de estado.

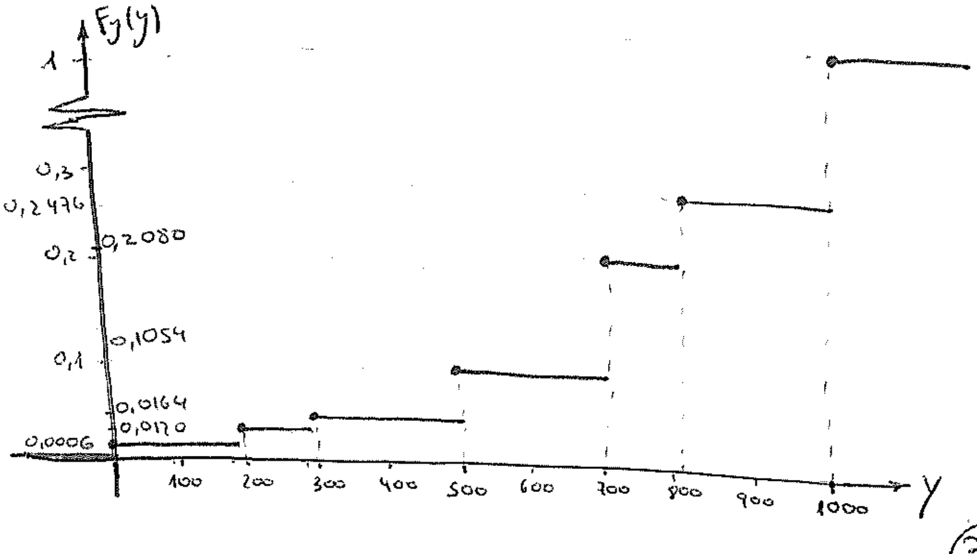
1) Calcular $F_Y$ (Y = potencia disponible)

$$N = \begin{cases} 500 \text{ MW} & p = 0,9 \\ 0 & p = 0,1 \end{cases} ; \quad T = \begin{cases} 300 \text{ MW} & p = 0,88 \\ 0 & p = 0,12 \end{cases} ; \quad H = \begin{cases} 200 \text{ MW} & p = 0,95 \\ 0 & p = 0,05 \end{cases}$$

Y v.a pot. disponible en la comarca → $Y = N + T + H$

| N T H | y | P (Y = y) |
|---|---|---|
| 0 0 0 | 0 | 0,1·0,12·0,05 = 0,0006 |
| 0 0 200 | 200 | 0,1·0,12·0,95 = 0,0114 |
| 0 300 0 | 300 | 0,1·0,88·0,05 = 0,0044 |
| 0 300 200 | 500 | 0,1·0,88·0,95 = 0,0836 |
| 500 0 0 | 500 | 0,9·0,12·0,05 = 0,0054 |
| 500 0 200 | 700 | 0,9·0,12·0,95 = 0,1026 |
| 500 300 0 | 800 | 0,9·0,88·0,05 = 0,0396 |
| 500 300 200 | 1000 | 0,9·0,88·0,95 = 0,7524 |

| y | $F_y(y) = P(Y \le y)$ |
|---|---|
| 0 | 0,0006 |
| 200 | 0,0170 |
| 300 | 0,0164 |
| 500 | 0,1054 |
| 700 | 0,2080 |
| 800 | 0,2476 |
| 1000 | 1 |



2) Un país tiene 30 centrales de cada tipo del ap. anterior. Calcular la prob. de que en un día la potencia disponible sea menor que 24000 MW (aprox. normal).

W v.a. pot disponible

$$W = N_1 + N_2 + \dots + N_{30} + T_1 + T_2 + \dots + T_{30} + H_1 + H_2 + \dots + H_{30}$$

$$W = \sum_{i=1}^{30} N_i + \sum_{i=1}^{30} T_i + \sum_{i=1}^{30} H_i \quad \text{con } W \sim N \text{, se pide } p(W < 24000 \mid W \sim N(\mu_w, \sigma_w))$$

$$E[W] = E\left[\sum_{i=1}^{30} N_i + \sum_{i=1}^{30} T_i + \sum_{i=1}^{30} H_i\right] = \sum_{i=1}^{30} E[N_i] + \sum_{i=1}^{30} E[T_i] + \sum_{i=1}^{30} E[H_i] \Rightarrow$$

$E[N_i] = 500 \cdot 0{,}9 + 0 \cdot 0{,}1 = 450$ MW

$E[T_i] = 300 \cdot 0{,}88 + 0 \cdot 0{,}12 = 264$ MW

$E[H_i] = 200 \cdot 0{,}95 + 0 \cdot 0{,}05 = 190$ MW

$E[W] = 30 \cdot 450 + 30 \cdot 264 + 30 \cdot 190 = 27120$ MW

$$Var[W] = \sum_{i=1}^{30} Var[N_i] + \sum_{i=1}^{30} Var[T_i] + \sum_{i=1}^{30} Var[H_i] =$$

$Var[N_i] = E[N_i^2] - (E[N_i])^2 = 500^2 \cdot 0{,}9 + 0^2 \cdot 0{,}1 - 450^2 = 22500$

$Var[T_i] = 300^2 \cdot 0{,}88 + 0^2 \cdot 0{,}12 - 264^2 = 9504$

$Var[H_i] = 200^2 \cdot 0{,}95 + 0^2 \cdot 0{,}05 - 190^2 = 1900$

$Var[W] = 30(22500 + 9504 + 1900) = 1017120$

$$p(W < 24000 \mid W \sim N(27120, \sqrt{1017720})) = p\left(Z < \frac{24000 - 27120}{\sqrt{101720}}\right) =$$

$$= p(Z < -3{,}09) = 1 - p(Z < 3{,}09) = 1 - 0{,}9990 = \boxed{0{,}001}$$

3) La potencia máxima diaria demandada es una v.a. $\rightarrow N(\overset{MW}{23000}, \overset{MW}{1000})$

¿p(demanda sea superior a la potencia disponible)?

D v.a. pot max diaria $\rightarrow N(23000, 1000)$

W v.a. pot disponible

$$p(D > W) = p(\underbrace{D - W}_{X} > 0) \; ; \quad X = D - W$$

$E[X] = E(D) - E[W] = 23000 - 27120 = -4120$

$Var[X] \underset{(\text{indep. } D, W)}{=} Var[D] + Var[W] = 1000^2 + 1017120 = 2017120$

$\downarrow$

$$p(X > 0 \mid X \rightarrow N(-4120, \sqrt{2017120})) = p\left(Z > \frac{4120}{\sqrt{2017120}}\right) =$$

$$= p(Z > 2{,}90) = 1 - \Phi(2{,}90) = 1 - 0{,}9981 = \boxed{0{,}0019}$$

**4.24** Inversor con dinero en 2 compañías. La rentabilidad es v.a.

$X$ v.a. rentabilidad compañía 1 $\rightarrow N(10, 2.5)$ $\quad \rho_{xy} = -0,5$

$Y$ " " " 2 $\rightarrow N(10, 1)$

¿Que proporción debe invertir en cada una para que el riesgo sea mínimo?

NOTA: Min. riesgo $\equiv$ min. varianza.

$W$ v.a. rentabilidad total — $W = aX + bY$ con $a+b=1$ (a y b en % 1)

$$Var[W] = (a \quad 1-a)\begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\begin{pmatrix} a \\ 1-a \end{pmatrix} = \left(a\sigma_x^2 + (1-a)\rho\sigma_x\sigma_y \,\middle|\, a\rho\sigma_x\sigma_y + (1-a)\sigma_y^2\right)\begin{pmatrix} a \\ 1-a \end{pmatrix}$$

$$= a^2\sigma_x^2 + 2a(1-a)\rho\sigma_x\sigma_y + (1-a)^2\sigma_y^2$$

$$\frac{d\,Var(W)}{da} = 0 \rightarrow \underline{2a\sigma_x^2} + 2\rho\sigma_x\sigma_y - 4a\rho\sigma_x\sigma_y - \underline{2(1-a)\sigma_y^2} = 0$$

$\sigma_x^2 = 2,5^2 = 6,25$

$\sigma_y^2 = 1^2 = 1$

$\rho\sigma_x\sigma_y = -0,5 \cdot 2,5 \cdot 1 = -1,25$

$$\underline{12,5a} - 1,25(\underline{2} - \underline{4a}) - 2(\underline{1} - \underline{a}) = 0$$

$$a(12,5 + 5 + 2) = 2,5 + 2$$

$$\left.\begin{aligned} a &= \frac{4,5}{19,5} = 0,23 \\ b &= 0,77 \end{aligned}\right\}$$

Debe invertir el 23% en la 1ª compañía y el 77% en la 2ª

**4.28** Plan de muestreo para controlar lotes de 10000 unidades. La cap. de inspección es de 200 piezas. Determinar c "nº max. de piezas defectuosas en

a) la muestra que debe tener un lote aceptado" si se desea que la prob. de rechazar un lote con el 10% piezas = 0,02.

$X$ v.a nº piezas defectuosas de un total de 200 $\rightarrow B(n=200, p = p(\text{deject.}))$

$$p(\text{rechazar}) = p(X \geq c) = 0,02 \quad \text{si } p(\text{def}) = 0,1$$

$$p(X > c \mid X \rightarrow B(200, 0,1)) \simeq p\left(X > c \mid X \sim N(20, \sqrt{\underbrace{20 \cdot 0,9}_{18}})\right) = 0,02$$

$$p(X > c) = p\left(z > \frac{c-20}{\sqrt{18}}\right) = 1 - p\left(z \leq \frac{c-20}{\sqrt{18}}\right) = 0,02$$

$$p\left(z \leq \frac{c-20}{\sqrt{18}}\right) = 1 - 0,02 = 0,98 \longrightarrow \frac{c-20}{\sqrt{18}} = 2,05, \quad c = 28,7$$

El lote debera contener un máximo de 28 piezas defectuosas de un total de 200 para ser aceptado

b) Un proveedor estima que sus lotes tienen un 5% de defectuosas.
¿qué prob. tiene de que un lote suyo sea aceptado?

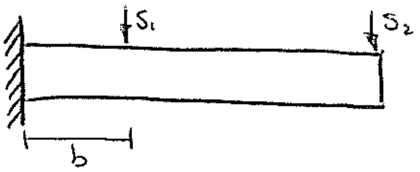
$p(\text{defec}) = 0{,}05$  Y v.a. nº piezas def. de un lot. de 200,
$n = 200$

$$p(\text{aceptar}) = p\left(Y \leq 28 \mid Y \to B(200, 0{,}05)\right) \simeq p\left(Y \leq 28 \mid Y \sim N(10, 3{,}08)\right) =$$

$$= p\left(Z \leq \frac{28-10}{3{,}08}\right) = p\left(Z \leq 5{,}84\right) \simeq \boxed{1}$$

Una viga en voladizo está sometida a dos independientes cargas $S_1$ y $S_2$ que son v.a. con mismo media y desviación típica. Obtener la correlación entre el esfuerzo cortante Q y el momento M en el empotramiento sabiendo que

$$\begin{cases} Q = S_1 + S_2 \\ M = bS_1 + 2bS_2 \end{cases}$$



$\sigma_{S_1}^2 = \sigma_{S_2}^2 = \sigma^2$
$\text{cov}(S_1, S_2) = 0$

$$\rho_{Q,M} = \frac{\text{cov}(Q,M)}{\sigma_Q \, \sigma_M} \quad , \quad M_{QM} = \begin{pmatrix} \sigma_Q^2 & \text{cov}(Q,M) \\ \text{cov}(Q,M) & \sigma_M^2 \end{pmatrix}$$

Necesitamos calcular los valores de $\underline{M_{Q,M}}$

$$\begin{pmatrix} Q \\ M \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ b & 2b \end{pmatrix}}_{A} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} , \quad M_{QM} = A \, M_{S_1S_2} \, A^T = \begin{pmatrix} 1 & 1 \\ b & 2b \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 & b \\ 1 & 2b \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma^2 & \sigma^2 \\ b\sigma^2 & 2b\sigma^2 \end{pmatrix} \begin{pmatrix} 1 & b \\ 1 & 2b \end{pmatrix} = \begin{pmatrix} 2\sigma^2 & 3b\sigma^2 \\ 3b\sigma^2 & 5b^2\sigma^2 \end{pmatrix} \longrightarrow \begin{aligned} \sigma_Q^2 &= 2\sigma^2 \\ \sigma_M^2 &= 5b^2\sigma^2 \\ \text{cov}(Q,M) &= 3b\sigma^2 \end{aligned} \Longrightarrow$$

$$\Longrightarrow \rho_{QM} = \frac{3\cancel{b}\cancel{\sigma}^2}{\cancel{\sigma}\sqrt{2} \cdot \cancel{b}\cancel{\sigma}\sqrt{5}} = \boxed{\frac{3}{\sqrt{10}}}$$

# FÓRMULAS → MODELOS MULTIVARIANTES

- Variables discretas:

Distribución conjunta de prob. $\boxed{P(X=x, Y=y)}$

Función de distribución conjunta $\boxed{F_{xy}(x,y) = P(X \leq x, Y \leq y)}$

Distribuciones marginales
$$\left\{ \boxed{\begin{array}{l} P(X=x) = \sum_{y=-\infty}^{y=\infty} P(X=x, Y=y) \\ P(Y=y) = \sum_{x=-\infty}^{x=\infty} P(X=x, Y=y) \end{array}} \right.$$

Distribuciones condicionadas
$$\boxed{P(Y=y \mid X=x_0) = \frac{P(Y=y, X=x_0)}{P(X=x_0)}}$$

Independencia → $\boxed{P(X=i, Y=j) = P(X=i)\, P(Y=j)}$

- Variables continuas

Función de densidad conjunta $f_{xy}(x,y)$

Función de distribución conjunta
$$\boxed{F_{xy}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{xy}(x,y)\, dx\, dy}$$

Funciones de densidad marginales
$$\left\{ \boxed{\begin{array}{l} \int_{-\infty}^{\infty} f_{x,y}(x,y)\, dy = f_x(x) \\ \int_{-\infty}^{\infty} f_{xy}(x,y)\, dx = f_y(y) \end{array}} \right.$$

Funciones de densidad condicionadas
$$\left\{ \boxed{\begin{array}{l} f_{X|Y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)} \quad \text{con } f_y(y) > 0 \\ f_{Y|X}(y|x) = \frac{f_{xy}(x,y)}{f_x(x)} \quad \text{con } f_x(x) > 0 \end{array}} \right.$$

Independencia $\boxed{f_{x,y}(x,y) = f_x(x)\, f_y(y)}$

- Esperanza de $g(X,Y)$

$$\boxed{\begin{array}{l} E[g(x,y)] = \sum\limits_{x=-\infty}^{x=\infty} \sum\limits_{y=-\infty}^{y=\infty} g(x,y) \cdot P(X=x,\ Y=y) \\ E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{x,y}(x,y)\, dx\, dy \end{array}}$$

en discretas;

en continuas.

Si $g(x,y) = g_1(x) + g_2(y) \longrightarrow \boxed{E[g(x,y)] = E[g_1(x)] + E[g_2(x)]}$

- Covarianza y Varianza

$$\boxed{Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]} \qquad \boxed{Cov(X,Y) = E[X \cdot Y] - E[X]E[Y]}$$

Si $X, Y$ son indep. $\to Cov(X,Y) = 0$

$\hookrightarrow E[X \cdot Y] = E[X]E[Y]$

$$\boxed{Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2ab\, cov(X,Y)}$$

$$\boxed{Var[X] = E[X^2] - (E[X])^2}$$

- Correlación $\boxed{\rho(x,y) = \dfrac{cov(x,y)}{\sigma_x\ \sigma_y}}$

- Transformaciones lineales

$Y = a^T X \longrightarrow \begin{cases} E[Y] = a^T E[X] \\ Var[Y] = a^T\, Var[X]\, a \end{cases}$

$\bar{Y} = A \bar{X} \longrightarrow \begin{cases} E[Y] = A\, E[X] \\ Var[Y] = A\, Var[X]\, A^T \end{cases}$

Mat. varianzas

- Teorema central del límite $\boxed{\bar{X} \longrightarrow N\left(\mu; \dfrac{\sigma}{\sqrt{n}}\right)} \to$ $x_1, x_2, \ldots x_n$ v.a. indep. con misma distribución de media $\mu$ y var. $\sigma^2$

- Aproximaciones

$$\begin{tikzcd} \text{Binomial}(n,p) \arrow[rr, "n\to\infty\ p\to 0"', "\boxed{\lambda = np}"] \arrow[dr, "\substack{n\to\infty \\ p\to\frac{1}{2}}"', "\boxed{\substack{\mu = np \\ \sigma = \sqrt{np(1-p)}}}"] & & \text{Poisson } \lambda \arrow[dl, "\boxed{\substack{\mu = \lambda \\ \sigma = \sqrt{\lambda}}}"', "\lambda\to\infty"] \\ & \text{Normal } \mu\ \sigma & \end{tikzcd}$$

- Distribución normal multivariante $\boxed{f(x_1, x_2 \ldots x_n) = \dfrac{1}{(2\pi)^{\frac{n}{2}} |M|^{1/2}} \exp\left\{ -\frac{1}{2}(X-\mu)^T M^{-1} (X-\mu) \right\}}$

Bivariante $\to \boxed{f(x_1, x_2) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \dfrac{-1}{2(1-\rho^2)} \left[ \left(\dfrac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\dfrac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho \left(\dfrac{x_1-\mu_1}{\sigma_1}\right)\left(\dfrac{x_2-\mu_2}{\sigma_2}\right) \right] \right\}}$

# Estimación puntual

## 1 MÉTODOS DE ESTIMACIÓN. Método de los momentos

Sea $X \rightarrow f_X(x, \theta_1, \theta_2, \ldots \theta_r)$ con $f_X$ conocida pero $\theta_1, \theta_2 \ldots \theta_r$ desconocidos.

Tomamos una muestra aleatoria simple de X: $X_1, X_2, X_3 \ldots X_n$

El método de los momentos consiste en:

- Hallar los momentos de grado 1 hasta r de la m.a.s.
- Igualamos los momentos de la muestra a los de la población.

Ejemplo: Sea la normal $N(\mu, \sigma)$, estimar $\mu$ y $\sigma$ por el método de los momentos.

Como tenemos 2 parámetros calcularemos los momentos de grado 1 y 2:

m.a.s. pob. real

$$a_1 = \frac{\sum_{i=1}^{n} x_i}{n} \longrightarrow a_1 = E[X] = \mu$$

$$a_2 = \frac{\sum_{i=1}^{n} x_i^2}{n} \longrightarrow a_2 = E[X^2] = \sigma^2 + \mu^2$$

Por tanto, igualando obtenemos la estimación pedida

$$\boxed{\begin{array}{l} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{X}^2 = S^2 \end{array}}$$

Por tanto, los estimadores por momentos de una normal $N \rightarrow (\mu, \sigma)$ son $\boxed{\hat{\mu} = \bar{X} \text{ y } \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{X}}$

(5.2) Los taxis en servicio de una ciudad están numerados del 1 al N. Se observa una muestra de 10 taxis apuntándose sus números. Obtener un estimador de N por el método de los momentos.

Taxi: 1, 2, 3, .. N = X

M.a.s: n = 10 taxis

$X_1$
$X_2$
⋮
$X_{10}$

| taxi X | prob (X=x) |
|---|---|
| 1 | 1/N |
| 2 | 1/N |
| 3 | 1/N |
| ⋮ | |
| N | 1/N |

Discreta uniforme

La media de la m.a.s es $\bar{X} = \frac{X_1 + X_2 + \ldots + X_{10}}{10}$

Calculamos $E[X] = \sum_{i=1}^{n} X\, \underbrace{p(X=x)}_{1/n} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\left(\frac{1+N}{2} \cdot N\right) = \frac{1+N}{2}$

!!

Igualemos [1]

$$\bar{X} = \frac{1+\hat{N}}{2}$$

$$\boxed{\hat{N} = 2\bar{X} - 1}$$

➡ En una urna hay N bolas numeradas de 1 hasta N. Se obtiene una muestra aleatoria simple con remplazamiento de 10 bolas resultando: 703, 785, 363, 454, 1050, 986, 798, 1338, 693, 646. Estimar N por el método de los momentos.

m.a.s: $X_1$
$X_2$
$\vdots$
$X_{10}$

población real $\underbrace{X, \Theta_1, \Theta_2, \ldots \Theta_N}_{N}$ : $P(X=x) = \frac{1}{N}$

Calculamos $\bar{X}$ (media de la m.a.s.) $= \frac{\sum_{i=1}^{10} x_i}{10} = \frac{7816}{10} = 781,6$

Calculamos $E[X]$ (pob real) $= \sum_{i=1}^{N} X_i \, p(X=x) = \frac{1}{N} \sum_{i=1}^{N} X_i = \frac{1}{N}\left(\frac{1+N}{2} \cdot N\right)$

Igualamos: $781,6 = \frac{1+\hat{N}}{2}$; $\boxed{\hat{N} = 1562,2 \approx 1562 \text{ bolas}}$

➡ El coseno de un ángulo con el que se emiten ciertas partículas de un proceso radiactivo es una v. aleatoria con función de densidad: $f_X(x) = \frac{1+\Theta x}{2}$ donde $x \in [-1, 1]$. Estimar $\Theta$ por el método de los momentos.

Tomamos una m.a.s $X_1, X_2, \ldots X_n \longrightarrow \bar{X}$

Calculamos $E[X] = \int_{-\infty}^{\infty} x f_X(x) = \int_{-1}^{1} x \frac{1+\Theta x}{2} = \frac{1}{2}\left(\frac{x^2}{2} + \frac{x^3\Theta}{3}\right)\Big|_{-1}^{1} =$ $\Longrightarrow$

$$= \frac{1}{2}\left(\frac{1}{2} + \frac{\Theta}{3} - \frac{1}{2} + \frac{\Theta}{3}\right) = \frac{\Theta}{2}$$

Igualamos $\bar{X} = \frac{\hat{\Theta}}{3} \longrightarrow \boxed{\hat{\Theta} = 3\bar{X}}$

## 2 MÉTODOS DE ESTIMACIÓN. Máxima Verosimilitud

Para aplicar este método haremos:

1) Calcular la función de verosimilitud $\ell(\Theta_1, \Theta_2, \ldots \Theta_n) = \prod_{i=1}^{n} f_X(x_i)$

2) Calculamos la función soporte $L(\Theta_1, \Theta_2, \ldots \Theta_n) = \log \ell(\Theta_1, \Theta_2, \ldots \Theta_n)$

3) Derivamos e igualamos a cero tantas veces como parámetros haya que estimar.

$$\frac{\partial L(\Theta_1, \Theta_2 \ldots \Theta_n)}{\partial \Theta_i} = 0 \longrightarrow \hat{\Theta}_i$$

//

➔ Una fuente radiactiva emite partículas según un proceso de Poisson con media λ desconocida. Durante 10 minutos se han contado el nº de partículas emitidas

se pide estimar λ

12, 6, 11, 3, 8, 5, 3, 9, 7, 5.

① Calculo $\ell(\lambda) = p(X_1=12, X_2=6, X_3=11, \dots X_{10}=5) \overset{\downarrow}{=}$

Por tratarse de una m.a.s, son independientes.
Además, todos son Poissón, con función de densidad
$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$

$$= e^{-\lambda}\frac{\lambda^{12}}{12!}\cdot e^{-\lambda}\frac{\lambda^{6}}{6!}e^{-\lambda}\frac{\lambda^{11}}{11!}\cdots e^{-\lambda}\frac{\lambda^{5}}{5!} =$$

$$= e^{-10\lambda}\frac{\lambda^{12+6+11+\dots+5}}{12!6!11!\dots5!} = e^{-10\lambda}\frac{\lambda^{69}}{12!6!\dots5!}$$

② Calculo $L(\lambda) = \log(\ell(\lambda)) = \log\left(e^{-10\lambda}\frac{\lambda^{69}}{12!6!\dots5!}\right)$

③ $$\frac{d}{d\lambda}(L(\lambda)) = \frac{-10\cancel{e^{-10\lambda}}\frac{\lambda^{69}}{\cancel{12!6!\dots5!}} + \cancel{e^{-10\lambda}}\frac{69\lambda^{68}}{\cancel{12!6!\dots5!}}}{\cancel{e^{-10\lambda}}\frac{\lambda^{69}}{\cancel{12!6!\dots5!}}} = \frac{-10\lambda^{69}+69\lambda^{68}}{\lambda^{69}} = \frac{69-10\lambda}{\lambda}$$

$$\frac{69-10\hat{\lambda}}{\hat{\lambda}} = 0 \longrightarrow \boxed{\hat{\lambda} = 6{,}9}$$

## Máx. verosimilitud. Poisson

$p(X=x) = e^{-\lambda}\frac{\lambda^x}{x!}$ ; queremos estimar λ

Muestra aleatoria simple (m.a.s): $X_1, X_2, \dots X_n$

$$\ell(\lambda) = p(X_1, X_2, \dots X_n) = e^{-\lambda}\frac{\lambda^{X_1}}{X_1!}e^{-\lambda}\frac{\lambda^{X_2}}{X_2!}\cdots e^{-\lambda}\frac{\lambda^{X_n}}{X_n!} = e^{-\lambda n}\frac{\lambda^{\sum_{i=1}^{n} X_i}}{X_1!X_2!\dots X_n!}$$

$$L(\lambda) = -\lambda n + \sum_{i=1}^{n} X_i \log\lambda - \sum_{i=1}^{n}\log X_i!$$

$$\frac{dL(\lambda)}{d\lambda} = -n + \sum_{i=1}^{n} X_i\frac{1}{\lambda} + 0; \quad -n + \sum_{i=1}^{n} X_i\frac{1}{\hat{\lambda}} = 0; \quad \boxed{\hat{\lambda} = \overline{X}}$$

## Máx. verosimilitud: $\boxed{\text{Normal}}$

$$f_x(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$ queremos estimar $\mu$ y $\sigma$

①

$$\ell(\mu,\sigma)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2}\cdots\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}=\frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2}$$

②

$$L(\mu,\sigma)=\log\ell(\mu,\sigma)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

③

$$\begin{cases}\dfrac{\partial L}{\partial\mu}=-\dfrac{1}{2\sigma^2}2\displaystyle\sum_{i=1}^{\infty}(x_i-\mu)\cdot(-1)\longrightarrow\dfrac{1}{2\sigma^2}\sum_{i=1}^{\infty}(x_i-\hat{\mu})=0;\ \boxed{\hat{\mu}=\overline{x}}\\[2ex]\dfrac{\partial L}{\partial\sigma^2}=-\dfrac{n}{2}\dfrac{1}{\sigma^2}+\dfrac{2}{4\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2;\ \dfrac{+n}{2}\dfrac{1}{\hat{\sigma}^2}-\dfrac{1}{2\hat{\sigma}^4}\sum_{i=1}^{\infty}(x_i-\hat{\mu})^2=0\end{cases}$$

$$\frac{n}{\cancel{2}}-\frac{1}{\cancel{2}\hat{\sigma}^2}\sum_{i=1}^{\infty}(x_i-\hat{\mu})^2=0;\quad\hat{\sigma}^2=\frac{\Sigma(x_i-\hat{\mu})^2}{n}\longrightarrow\boxed{\hat{\sigma}^2=s^2}$$

## Máx. verosimilitud: $\boxed{\text{Binomial}}$

$$p(X=x)=p^r(1-p)^{n-r}$$

① $\ell(p)=p(x_1,x_2,\dots x_n)=p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}=p^r(1-p)^{n-r}$ con $r=\Sigma x_i$ es el nº de defectuosas

② $L(p)=\log(\ell(p))=r\log p+(n-r)\log(1-p)$

③ $$\frac{dL}{dp}=r\frac{1}{p}+(n-r)\frac{-1}{1-p};\quad\frac{r}{\hat{p}}-\frac{(n-r)}{1-\hat{p}}=0;\quad\frac{r(1-\hat{p})-(n-r)\hat{p}}{\hat{p}(1-\hat{p})}=0$$

$$r-\cancel{r\hat{p}}-\hat{p}n+\cancel{r\hat{p}}=0;\quad\boxed{\hat{p}=\frac{r}{n}}$$

NOTA!! Cuadro resumen estimadores

| | MOMENTOS | MÁX. VEROSIMILITUD |
|---|---|---|
| NORMAL | $\hat{\mu}_m=\overline{x}$<br>$\hat{\sigma}^2_m=S^2$ | $\hat{\mu}_{mv}=\overline{x}$<br>$\hat{\sigma}^2_{mv}=S^2$ |
| POISSON | $\hat{\lambda}_m=\overline{x}$ | $\hat{\lambda}_{mv}=\overline{x}$ |
| BERNOULLI | $\hat{p}_m=\frac{r}{n}$ | $\hat{p}_{mv}=\frac{r}{n}$ |

**5.3** Una variable aleatoria discreta puede tomar los valores 0, 1, 2 con probabilidades $\frac{1.5}{\theta}$, $\frac{2.5}{\theta}$, $\frac{\theta-4}{\theta}$ respectivamente. Se toma una muestra de tamaño 25 con los resultados siguientes:

| X | 0 | 1 | 2 |
|---|---|---|---|
| Oi | 17 | 5 | 3 |

Estimar $\theta$ por máxima verosimilitud

① $$l(\theta)=\left(p(X=0)\right)^{17}\left(p(X=1)\right)^{5}\left(p(X=2)\right)^{3}=\left(\frac{1.5}{\theta}\right)^{17}\cdot\left(\frac{2.5}{\theta}\right)^{5}\left(\frac{\theta-4}{\theta}\right)^{3}\equiv \text{Fun. verosimilitud}$$

② $$L(\theta)=17\log(1,5)-17\log(\theta)+5\log(2,5)-5\log(\theta)+3\log(\theta-4)-3\log(\theta);$$

③ $$\frac{dL(\theta)}{d\theta}=-17\frac{1}{\theta}-5\frac{1}{\theta}+3\frac{1}{\theta-4}-3\frac{1}{\theta}=-25\frac{1}{\theta}+3\frac{1}{\theta-4}=\frac{-25\theta+100+3\theta}{\theta(\theta-4)}$$

$$\frac{dL}{d\theta}=0\longrightarrow -25\hat{\theta}+3\hat{\theta}+100=0;\quad \boxed{\hat{\theta}_{MV}=\frac{100}{22}}$$

**5.4** Se ha tomado una muestra de tamaño 10 del tiempo, en minutos, entre el paso de dos autobuses T en una parada con los siguientes resultados:

9, 10, 6, 4, 15, 6, 1, 5, 4, 10. $F(t)=1-\exp(-\alpha t)$. Calcular $p(T>10)$

T v.a tiempo entre dos autobuses

$$p(T>10)=1-p(T<10)=1-F(10)=e^{-\alpha 10}$$

Como no sabemos el valor de $\alpha$ lo tenemos que estimar.

Usamos el método de máxima verosimilitud:

① $$l(\alpha)=p(9,10,6\ldots 10)=f(9)\,f(10)\ldots f(10)=\alpha e^{-\alpha 9}\alpha e^{-\alpha 10}\ldots\alpha e^{-\alpha 10}=$$

$$f(t)=\frac{d}{dt}(1-e^{-\alpha t})=\alpha e^{-\alpha t}$$

$$=\alpha^{10}e^{-\alpha\sum_{i=1}^{10}t_i}=\alpha^{10}e^{-\alpha 70}$$

② $$L(\alpha)=\log(\alpha e^{-\alpha 70})=10\log\alpha-\alpha 70$$

③ $$\frac{dL}{d\alpha}=\frac{10}{\alpha}-70,\quad \frac{dL}{d\alpha}=0\longrightarrow \frac{10}{\hat{\alpha}}-70=0;\quad \frac{10}{\hat{\alpha}}=70;\quad \hat{\alpha}=\frac{1}{7}$$

Por tanto $p(T>10)=e^{-\frac{1}{7}10}=\boxed{0,24}$

**(5.5)** La función de distribución de una variable aleatoria es: $F(x)\begin{cases} 0 & x<0 \\ (x/\beta)^{\alpha} & 0\le x\le\beta \\ 1 & x>\beta \end{cases}$

con $\alpha,\beta>0$. Estimar $\alpha$ y $\beta$ por máx. verosimilitud

$$\frac{dF(x)}{dx}=\frac{\alpha}{\beta^{\alpha}}(x)^{\alpha-1} \qquad 0\le x\le\beta$$

① $$l(\alpha,\beta)=\frac{\alpha}{\beta^{\alpha}}(x_1)^{\alpha-1}\cdot\frac{\alpha}{\beta^{\alpha}}(x_2)^{\alpha-1}\cdots\frac{\alpha}{\beta^{\alpha}}(x_n)^{\alpha-1}=\left(\frac{\alpha}{\beta^{\alpha}}\right)^{n}(x_1x_2\ldots x_n)^{\alpha-1}$$

② $$L(\alpha,\beta)=\log(l(\alpha,\beta))=n\log\alpha-\alpha n\log\beta+(\alpha-1)\sum_{i=1}^{n}\log x_i$$

③ $$\begin{cases} \dfrac{\partial L}{\partial\alpha}=n\dfrac{1}{\hat\alpha}-n\log\hat\beta+\displaystyle\sum_{i=1}^{n}\log x_i=0\,;\quad \hat\alpha=\dfrac{n}{n\log\hat\beta-\sum\log x_i} \\ \dfrac{\partial L}{\partial\beta}=-\dfrac{\hat\alpha n}{\hat\beta}=0 \longrightarrow \text{Imposible obtener } \hat\beta \end{cases}$$

**NOTA!!** Siempre que un parámetro condicione el dominio no podremos obtenerlo por estimación directa

$\beta\ge x_i$ (condición de dominio) $\Longrightarrow \boxed{\hat\beta=\max\{x_1,x_2,\ldots x_n\}}$

$$\boxed{\hat\alpha=\frac{n}{n\log\hat\beta-\sum\log x_i}}$$

**(5.9)** El tiempo de duración de ciertos componentes electrónicos es una variable aleatoria con distribución exponencial. Se ha realizado un ensayo con 10 componentes cuyos tiempos de duración han sido: 37, 45, 92, 104, 109, 200, 295. Después de 400 h, tres componentes seguían funcionando. Estimar por máx. verosimilitud el parámetro de la distribución exponencial.

$t_1$

$t_2$

$\vdots$

$t_7$

Datos (censurados) $\begin{cases} t_8>400 \\ t_9>400 \\ t_{10}>400 \end{cases}$

**NOTA!!** Al trabajar con datos censurados haremos la función de verosimilitud como producto de las $f(x_i)$ de los datos corrientes junto con el producto de $p(T>t_i)$ de los datos censurados.

Por ser exponencial $f(t)=\lambda e^{-\lambda t}$; $F(t)=1-e^{-\lambda t}$

① $$l(\lambda)=f(t_1)f(t_2)\ldots f(t_7)p(T_8>400)p(T_9>400)p(T_{10}>400)=$$
$$=\lambda e^{-\lambda 37}\lambda e^{-\lambda 45}\lambda e^{-\lambda 92}\ldots\lambda e^{-\lambda 295}\left(e^{-\lambda 400}\right)^{3}=\lambda^{7}e^{-\lambda\left(\sum_{i=1}^{7}t_i+1200\right)}=\lambda^{7}e^{-\lambda 2080}$$

② $L(\lambda)=7\log\lambda-\lambda 2080$ ; ③ $\dfrac{dL}{d\lambda}=\dfrac{7}{\lambda}-2080=0$, $\boxed{\hat\lambda=\dfrac{7}{2080}=3{,}36\;10^{-3}}$

Extra → Vida media estimada $\widehat{E[T]}=\dfrac{1}{\hat\lambda}=297{,}4\,h$

➡ La velocidad del viento "y" en un emplazamiento es una variable aleatoria cuya distribución de probabilidad es:

$$p(Y \leq y) = 1 - e^{-(y/a)^2} \qquad y \geq 0$$

Con un anemómetro se ha medido la velocidad efectiva del viento en 8h distintas espaciados en el tiempo de forma que pueden considerarse independientes, obteniéndose:

$v_1 = 8,2$ $v_4 = 6,8$ $v_7 = 9,3$
$v_2 = 5,5$ $v_5 = 2,5$ $v_8 = 3,9$
$v_3 = 2,7$ $v_6 = 7,7$

Estimar "a" por máxima verosimilitud.

$$f_y(y) = \frac{2y}{a} e^{-(y/a)^2}$$

① $$l(a) = \frac{2v_1}{a} e^{-(v_1/a)^2} \cdots \frac{2v_8}{a} e^{-(v_8/a)^2} = \left(\frac{2}{a}\right)^8 (v_1 v_2 \ldots v_8)\, e^{-\frac{1}{a^2}\sum_{i=1}^{8} v_i^2}$$

$$l(a) = \frac{148001962,3}{a^8} e^{-\frac{1}{a^2} 318,26}$$

② $$L(a) = 18,81 - 8 \log a - \frac{318,26}{a^2}$$

③ $$\frac{dL}{da} = -8\frac{1}{\hat{a}} + \frac{318,26 \, 2\hat{a}}{\hat{a}^4} = 0 \, ; \quad \frac{-8\hat{a}^2 + 318,262}{\hat{a}^3} = 0$$

$$\boxed{\hat{a} = 39,78}$$

➡ La distribución $\Gamma$: $f_y(y) = \dfrac{1}{\alpha\,\Gamma(\beta)} y^{\beta-1} e^{y/\alpha}$ ; $y > 0$ es utilizada con mucha frecuencia para describir los niveles de precipitación en un lugar, donde $\alpha$ y $\beta$ son parámetros a estimar. Se dispone de las precipitaciones máximas en un día medidos en pulgadas para un total de 36 tormentas observadas en un sistema montañoso peninsular. Siendo la media 7,29 y la varianza 32,45, estimar $\alpha$ y $\beta$ por el método de los momentos, teniendo en cuenta que $E[y] = \alpha\beta$ ; $E[y^2] = \alpha^2\beta(\beta+1)$

$$\bar{x} = 7,29 \longrightarrow E[y] = \alpha\beta \longrightarrow \hat{\alpha}\hat{\beta} = 7,29 \quad (1)$$

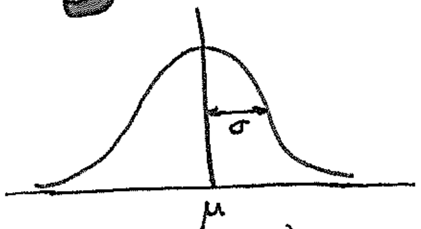
$$s^2 = 32,45 \longrightarrow E[y^2] = \sigma^2 + (E[y])^2 = \alpha^2\beta(\beta+1) \longrightarrow$$

$$\longrightarrow 32,45 + 7,29^2 = \hat{\alpha}^2\hat{\beta}(\hat{\beta}+1)$$

$$\left\{\begin{array}{l} \hat{\alpha}\hat{\beta} = 7,29 \longrightarrow \hat{\alpha} = \frac{7,29}{\hat{\beta}} \\ \hat{\alpha}^2\hat{\beta}(\hat{\beta}+1) = 32,45 + 7,29^2 \longrightarrow \frac{7,29^2}{\hat{\beta}^2}\hat{\beta}(\hat{\beta}+1) = 85,5941 \end{array}\right.$$

$$53,1441\hat{\beta} + 53,1441 = 85,5941\hat{\beta} \, , \quad 32,45\hat{\beta} = 53,1441 \, , \quad \boxed{\begin{array}{l} \hat{\beta} = 1,638 \\ \hat{\alpha} = 4,45 \end{array}}$$

## 3 DISTRIBUCIÓN DE MEDIA NORMAL: $\bar{X}$



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

$X_1, X_2, \dots X_n \equiv$ muestra aleatoria simple de una distribución normal $(\mu,\sigma)$

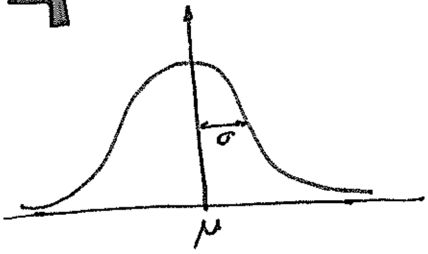
$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E[\bar{X}] = \frac{E[X_1] + E[X_2] + \dots + E[X_n]}{n} = \frac{\mu + \mu + \dots + \mu}{n} = \mu$$

$$Var[\bar{X}] = \frac{Var[X_1] + Var[X_2] + \dots + Var[X_n]}{n^2} = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Por tanto: $\boxed{\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$

## 4 DISTRIBUCIÓN DE VARIANZA NORMAL $S^2$ y VARIANZA CORREGIDA $\hat{S}^2$



$X_1, X_2, \dots X_n \equiv$ m.a.s. de distribución normal $(\mu,\sigma)$

$$S^2 = \frac{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \dots + (X_n-\bar{X})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$\frac{1}{\cancel{n}}\cancel{n}(\sigma^2+\mu^2) - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2$$

$$E[S^2] = \frac{1}{n}\underbrace{\sum E[X_i^2]}_{} - \underbrace{E[\bar{X}^2]}_{} = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right) = \frac{n-1}{n}\sigma^2$$

$Var[X_i] = E[X_i^2] - (E[X_i])^2$

$E[X_i^2] = \sigma^2 + \mu^2$

$Var[\bar{X}] = E[\bar{X}^2] - (E[\bar{X}])^2$

$E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$

Como $\boxed{E[S^2] = \frac{n-1}{n}\sigma^2}$ se esta subestimando el valor de la varianza.

Proponemos otro estimador $\hat{S}^2$ (varianza corregida)

$$\hat{S}^2 = \frac{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \dots + (X_n-\bar{X})^2}{n-1} \Longrightarrow \boxed{(n-1)\hat{S}^2 = nS^2}$$

$$E[\hat{S}^2] = E\left[\frac{nS^2}{n-1}\right] = \frac{n}{n-1}E[S^2] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2 ; \quad \boxed{E[\hat{S}^2] = \sigma^2}$$

Ahora la esperanza coincide con el valor verdadero por lo que ya no estoy subestimando.

# 5 DISTRIBUCIÓN $\chi^2$

$z_1, z_2, \dots z_n \longrightarrow N(0,1)$ independientes

$z_1^2 + z_2^2 + \dots + z_n^2 \longrightarrow \chi^2_n$ Distribución CHI$^2$ con n grados de libertad

Propiedades

$$\boxed{\begin{array}{l} E[\chi^2_n] = n \\ Var[\chi^2_n] = 2n \\ \chi^2_n + \chi^2_m = \chi^2_{n+m} \text{ con } \chi^2_n \text{ y } \chi^2_m \text{ independientes} \end{array}}$$

$$\boxed{\frac{nS^2}{\sigma^2} = \sum_{i=1}^{n} \left( \frac{X_i - \overline{X}}{\sigma} \right)^2 \longrightarrow \chi^2_{n-1}}$$

$Var[S^2] \Longrightarrow Var\left[\frac{nS^2}{\sigma^2}\right] = Var[\chi^2_{n-1}] = 2(n-1)$

$|||$

$\frac{n^2}{\sigma^4} Var[S^2] = 2(n-1)$

$\Longrightarrow \boxed{Var[S^2] = \frac{2(n-1)\sigma^4}{n^2}}$

$$\frac{(n-1)\hat{S}^2}{\sigma^2} = \sum_{i=1}^{n} \left( \frac{X_i - \overline{X}}{\sigma} \right)^2 \longrightarrow \chi^2_{n-1}$$

$Var[\hat{S}^2] \Longrightarrow Var\left[\frac{(n-1)\hat{S}^2}{\sigma^2}\right] = 2(n-1)$

$\frac{(n-1)^2}{\sigma^4} Var[\hat{S}^2] = 2(n-1)$

$\boxed{Var[\hat{S}^2] = \frac{2\sigma^4}{n-1}}$

# 6 APROXIMACIONES A LA NORMAL

- Binomial $(n, p)$: $X_1, X_2, \dots X_n$ $\begin{cases} E[X_i] = p \\ Var[X_i] = p(1-p) \end{cases}$

$$\boxed{\hat{p} = \frac{X_1 + X_2 + \_\_\_ + X_n}{n}} \longrightarrow \begin{cases} E[\hat{p}] = p \\ Var[\hat{p}] = \frac{p(1-p)}{n} \end{cases}$$

$$\boxed{\hat{p} \xrightarrow[n\to\infty]{} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)}$$

- Poisson $(\lambda)$: $X_1, X_2, \dots X_n$ $\begin{cases} E(X_i) = \lambda \\ Var(X_i) = \lambda \end{cases}$

$$\boxed{\hat{\lambda} = \frac{X_1 + X_2 + \_\_\_ + X_n}{n}} \longrightarrow \begin{cases} E[\hat{\lambda}] = \lambda \\ Var[\hat{\lambda}] = \frac{\lambda}{n} \end{cases}$$

$$\boxed{\hat{\lambda} \xrightarrow[n\to\infty]{} N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)}$$

# 7 PROPIEDADES DE LOS ESTIMADORES

- Estimador centrado o insesgado $E[\hat{\theta}]=\theta$

Ejemplos: Normal $(\mu,\sigma)$

$\hat{\mu}=\bar{x}$, $\bar{x}\to N(\mu,\frac{\sigma^2}{n})$, $E[\bar{x}]=E[\hat{\mu}]=\mu \Rightarrow$ $\hat{\mu}$ es un estimador centrado

$s^2=\sigma^2$, $E[s^2]=\frac{n-1}{n}\sigma^2\neq\sigma^2 \Rightarrow$ $s^2$ no es centrado

$\hat{S}^2=\sigma^2$; $E[\hat{S}^2]=\sigma^2 \Rightarrow$ $\hat{S}^2$ es centrado

$$\text{sesgo}(S^2)=E[S^2]-\sigma^2=$$
$$=\frac{n-1}{n}\sigma^2-\sigma^2=\sigma^2\left(\frac{n-1-n}{n}\right)=$$
$$=\frac{-1}{n}\sigma^2$$

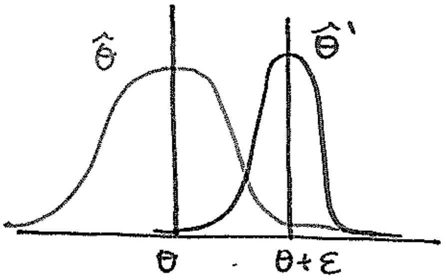
Un estimador centrado es insesgado, esto es:

$$\boxed{\text{Sesgo}[\hat{\theta}]=E[\hat{\theta}]-\theta=0}$$

- Varianza mínima

Los estimadores de máxima verosimilitud son los que tienen varianza mínima, esto es, para un mismo parámetro:

$$Var(\hat{\theta}_{MV})\leq Var(\hat{\theta})$$



- Error cuadrático medio mínimo

Cuando tenemos dos estimadores tales que:

$\hat{\theta}$ es centrado pero $\hat{\theta}'$ no lo es
$Var(\hat{\theta})\geq Var(\hat{\theta}')$

Definimos el $\boxed{ECM=E[(\hat{\theta}-\theta)^2]=\text{sesgo}^2(\hat{\theta})+Var(\hat{\theta})}$

Será mejor el estimador con menor ECM

Demo: $E[(\hat{\theta}-\theta)^2]=E[(\hat{\theta}-E[\hat{\theta}]+E[\hat{\theta}]-\theta)^2]=$

$=E[(\hat{\theta}-E[\hat{\theta}])^2+(E[\hat{\theta}]-\theta)^2+2(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta)]=$

$=\underbrace{E[(\hat{\theta}-E[\hat{\theta}])^2]}_{}+\underbrace{E[(E[\hat{\theta}]-\theta)^2]}_{}+2\underbrace{E[(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta)]}_{}=$

$=Var(\hat{\theta})+\text{sesgo}^2(\hat{\theta})$

$\underbrace{E[\hat{\theta}-E[\hat{\theta}]]}_{}E[E[\hat{\theta}]-\theta]$

$E(X-\mu_1)=0$ (momento de orden 1)

0

- Consistentes

$\boxed{\lim_{n\to\infty} E[\hat{\Theta}]=\Theta}$, es decir, que sean asintóticamente centrados

En general todos los estimadores de máximo verosimilitud lo son.

$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \longrightarrow E[\hat{p}]=p \Longrightarrow \hat{p}$ es asintóticamente centrado

$\hat{\lambda} \sim N\left(\lambda, \frac{\lambda}{n}\right) \longrightarrow E[\hat{\lambda}]=\lambda \Longrightarrow \hat{\lambda}$ es asintóticamente centrado

$S^2 \longrightarrow E[S^2]=\frac{n-1}{n}\sigma^2 \equiv \lim_{n\to\infty} E[S^2]=\lim_{n\to\infty}\frac{n-1}{n}\sigma^2=\sigma^2 \Rightarrow$ $S^2$ es asintóticamente centrado

Además $\boxed{\lim_{n\to\infty} Var[\hat{\Theta}]=0}$, es decir, han de tener varianza asintóticamente tendiendo a cero

➡ Ejemplo: Sea el estimador $S^2$ de una m.a.s. $X_1, X_2, \dots X_n$ $N(\mu,\sigma)$, estimador de $\sigma^2$. Definir sus propiedades:

$E[S^2]=\frac{n-1}{n}\sigma^2 \neq \sigma \longrightarrow$ No es centrado. $Sesgo[S^2]=E[S^2]-\sigma^2=\frac{n-1}{n}\sigma^2-\sigma^2$

$$Sesgo[S^2]=-\frac{\sigma^2}{n}$$

$Var[S^2]$; Como $Var\left[\frac{nS^2}{\sigma^2}\right]=Var(\chi^2_{n-1})=2(n-1) \Rightarrow Var[S^2]=\frac{2(n-1)\sigma^4}{n^2}$

$\lim_{n\to\infty} E[S^2]=\odot$ y $\lim_{n\to\infty} Var[S^2]=0 \Longrightarrow$ Es consistente

$ECM=sesgo^2(S^2)+var(S^2)=\left(\frac{-\sigma^2}{n}\right)^2+\frac{2(n-1)}{n^2}\sigma^4=\sigma^4\left(\frac{2(n-1)}{n^2}+\frac{1}{n}\right)$

$ECM=\sigma^4\frac{3n-2}{n^2}$

(5.10) Sea $X_1, X_2, \dots X_n$ una m.a.s. de una función de densidad: $f(x)=2(\Theta-x)/\Theta^2$, $0\leq x\leq\Theta$ Obtener por el método de los momentos un estimador insesgado de $\Theta$ y calcular su varianza

Calculemos el momento de orden 1 de la población real, esto es, $E[X]$

$E[X]=\int_0^\Theta x\,\frac{2(\Theta-x)}{\Theta^2}=\left(\frac{x^2\Theta}{\Theta^2}-\frac{2x^3}{3\Theta^2}\right)\Big|_0^\Theta=\Theta-\frac{2}{3}\Theta=\frac{\Theta}{3}$

Lo igualamos a $\bar{X}$ (media de la m.a.s) $\longrightarrow \frac{\hat{\Theta}}{3}=\bar{X}$; $\boxed{\hat{\Theta}=3\bar{X}}$

Como $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ $\hat{\Theta}$ tiene una distribución asintóticamente normal

$E[\hat{\Theta}]=E[3\bar{X}]=3\underbrace{E[\bar{X}]}=3E[X]=3\frac{\Theta}{3}=\Theta \Longrightarrow$ $\boxed{\text{Como } E[\hat{\Theta}]=\Theta,\ \hat{\Theta} \text{ es centrado}}$

↳ La esperanza de la media muestral coincide con la de la pob. real

$\Rightarrow \boxed{Sesgo[\hat{\Theta}]=0}$

↳ $\boxed{\hat{\Theta} \text{ es por tanto insesgado}}$

$$Var[\hat{\Theta}] = Var[3\bar{X}] = 9\underbrace{Var[\bar{X}]} = 9\frac{Var[X]}{n} = 9\frac{\Theta^2}{18n} = \frac{\Theta^2}{2n}$$

La varianza de la media muestral es la varianza de la población real dividida entre el nº de sumandos de la muestra

$$Var[X] = E[X^2] - (E[X])^2 =$$

$$= \int_0^0 x^2 \frac{2(\Theta - x)}{\Theta^2} - \left(\frac{\Theta}{3}\right)^2 = \frac{2\Theta^2}{3} - \frac{\Theta^2}{2} - \frac{\Theta^2}{9} =$$

$$= \frac{12\Theta^2 - 9\Theta^2 - 2\Theta^2}{18} = \frac{\Theta^2}{18}$$

Por tanto $\boxed{\hat{\Theta} \rightsquigarrow N\left(\Theta, \frac{\Theta^2}{2n}\right)}$

(5.12) $X_1, X_2, \ldots X_n$ es una m.a.s. de una $N(\mu, \sigma)$ con $\mu$ y $\sigma$ desconocidos. Para estimar la varianza se propone el estimador:

$$S^2 = K\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(X_i - X_j)^2 ;$$ Determinar K para que sea centrado.

Calculemos $E[S^2] = K\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[(X_i - X_j)^2] = K\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[(X_i^2 + X_j^2 - 2X_iX_j)] =$

$$= K\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big(\underbrace{E[X_i^2]}_{Var[X_i^2] + (E(X_i))^2 = \sigma^2 + \mu^2} + \underbrace{E(X_j^2]} - \underbrace{2E[X_iX_j]}_{2E[X_i]E[X_j] = 2\mu^2}\Big) = K\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 2\mu^2) =$$

$$= K2\sigma^2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 1 = K2\sigma^2 \underbrace{\sum_{i=1}^{n-1}(n-i)}_{(n-1)+(n-2)+(n-3)+\ldots+\underbrace{(n-n-1)}_{1}} = \left(\frac{n-1+1}{2}\cdot n-1\right)K2\sigma^2 = K\sigma^2 n(n-1)$$

Queremos que sea centrado, esto es $\longrightarrow E[S^2] = \sigma^2$

$$\longrightarrow K\sigma^2 n(n-1) = \sigma^2, \quad \boxed{K = \frac{1}{n(n-1)}}$$

➡ Sea $X_1, X_2, \ldots X_n$ una m.a.s de una variable con distribución de Poisson de media $\lambda = E[X]$. Se propone un estimador para $\lambda$: $T = \frac{X_1 + 2X_2 + 2X_3 + \ldots + 2X_{n-1} + X_n}{2(n-1)}$

¿Es un estimador insesgado de $\lambda$? Si es sesgado, calcule su sesgo. Calcule su varianza. Propon un estimador centrado con menor varianza que T.

$$E[T] = E\left(\frac{X_1 + 2X_2 + \_\_ + 2X_{n-1} + X_n}{2(n-1)}\right] = \frac{1}{2(n-1)}\left(E[X_1] + 2E[X_2] + \_\_ + 2E[X_{n-1}] + E[X_n]\right] \overset{E[X_i] = \lambda}{=}$$

$$= \frac{1}{2(n-1)}(\lambda + 2\lambda(n-2) + \lambda) = \frac{1}{n-1}(\lambda + \lambda(n-2)) = \frac{\lambda(n-1)}{n-1} = \lambda \longrightarrow$$ $\boxed{\text{T es insesgado o centrado, sesgo(T) = 0}}$

$$Var(T) = Var\left(\frac{X_1+2X_2+\_\_+2X_{n-1}+X_n}{2(n-1)}\right) = \left(\frac{1}{2(n-1)}\right)^2\left[Var(X_1)+4Var(X_2)+\_\_+4Var(X_{n-1})+Var(X\right.$$

$$\underset{\uparrow}{=} \left(\frac{1}{2(n-1)}\right)^2(\lambda+4\lambda+\_\_+4\lambda+\lambda) = \left(\frac{1}{2(n-1)}\right)^2(2\lambda+(n-2)4\lambda) = \frac{\lambda}{2(n-1)^2}(1+2n-4) = \frac{\lambda(2n-3)}{2(n-1)^2}$$

$Var(X_i)=\lambda$

$$\boxed{Var(T) = \frac{\lambda(2n-3)}{2(n-1)^2}}$$

Calcularemos un estimador $\hat{\lambda}$ por maxima verosimilitud

$E[\hat{\lambda}] = \lambda$, $Var[\hat{\lambda}] = \frac{\lambda}{n}$; $\boxed{\hat{\lambda} = \frac{\sum x_i}{n}}$

$Var(T) \geq Var(\hat{\lambda})$ porque $\hat{\lambda}$ es un estimador por máx. verosimilitud, por tanto tiene varianza minima.

$$\frac{\lambda(2n-3)}{2(n-1)^2} \geq \frac{\lambda}{n}; \quad \frac{2n-3}{2(n-1)^2} - \frac{1}{n} \geq 0;$$

$$\frac{2n^2-3n-2n^2-2+4n}{2n(n-1)^2} \geq 0; \quad \frac{n-2}{2n(n-1)^2} \geq 0 \begin{cases} \text{Si } n=1 \longrightarrow \text{Indeterminado} \\ n=2 \longrightarrow 0 \rightarrow Var(T) = Var(\hat{\lambda}) \\ n>2 \longrightarrow >0 \rightarrow Var(T) > Var(\hat{\lambda}) \end{cases}$$

$$\Longrightarrow \boxed{Var(T) \geq Var(\hat{\lambda})}$$

(5.1) $X \rightarrow B(n,p)$. mas $\{16, 18, 22, 25, 27\}$. Estimar $n$ y $p$ por el método de los momentos

Calculamos $E[X]$ de la pob. real que por ser $X$ una binomial

$E[X] = n \cdot p$

Ahora el momento de orden 1 de la muestra es:

$a_1 = \frac{\sum x_i}{n} = 21,6$ $\Big\}$ Igualando $\hat{n}\hat{p} = 21,6$

Hacemos lo mismo con $Var(X) = np(1-p)$

$$a_2 = \frac{\sum x_i^2}{n} = \frac{2418}{5} = 483,6 \longrightarrow \text{Igualando } \hat{n}\hat{p}(1-\hat{p}+\hat{n}\hat{p}) = 483,6$$

$$E[X^2] = Var(X) + (E[X])^2 = np(1-p) + (np)^2$$

$$\begin{cases} \hat{n}\hat{p} = 21,6 \rightarrow \hat{n} = \frac{21,6}{\hat{p}} \\ \hat{n}\hat{p}(1-\hat{p}+\hat{n}\hat{p}) = 483,6 \longrightarrow 21,6(1-\hat{p}+21,6) = 483,6 \end{cases}$$

$$\hat{p} = 22,6 - \frac{483,6}{21,6} = 0,21, \quad \hat{n} = \frac{21,6}{0,21} = 102,35 \simeq 103 \quad \boxed{\begin{matrix} \hat{p} = 0,21 \\ n = 103 \end{matrix}}$$

(5.6) La distancia X, del punto de impacto de un proyectil al centro de 6 di

es una variable aleatoria de $f(x)=\frac{2x}{\theta^2}\exp\left(-\frac{x^2}{\theta^2}\right)$; $x\geq 0$, $\theta\geq 0$

M.A.S: 2,1 | 3,2 | 6,3 | 5,4 | 2,2 | 6,9 | 7,1 | 6,6 | 2,5 | 9,1 | >11 | >11 | >11

Estimar θ   Estimaremos θ por max. v

$$\ell(\theta)=f(x_1)\,f(x_2)\ldots f(X_{10})\overbrace{p(X_{11}>11)\,p(X_{12}>11)\,p(X_{13}>11)}^{(1-F(11))^3}=$$

$$=\frac{2\cdot X_1}{\theta^2}e^{-\frac{x_1^2}{\theta^2}}\cdot\frac{2X_2}{\theta^2}e^{-\frac{x_2^2}{\theta^2}}\ldots\frac{2X_{10}}{\theta^2}e^{-\frac{X_{10}^2}{\theta^2}}\underbrace{\left(e^{-\frac{11^2}{\theta^2}}\right)^3}=$$

$$F(x)=\int_0^x\frac{2x}{\theta^2}e^{-\frac{x^2}{\theta^2}}=-e^{-\frac{x^2}{\theta^2}}\Big|_0^x=1-e^{-\frac{x^2}{\theta^2}}$$

$$=\left(\frac{2}{\theta^2}\right)^{10}(X_1\cdot X_2\cdot\ldots\cdot X_{10})\,e^{\frac{-X_1^2-X_2^2\ldots-X_{10}^2}{\theta^2}}\left(e^{-\frac{11^2}{\theta^2}}\right)^3$$

$$L(\theta)=\ln(\ell(\theta))=10\ln\left(\frac{2}{\theta^2}\right)+\ln(X_1\cdot X_2\cdots X_{10})+\frac{-X_1^2-X_2^2\ldots-X_{10}^2}{\theta^2}-\frac{11^2\cdot 3}{\theta^2}$$

$$\frac{dL}{d\theta}=10\frac{\frac{-4\theta}{\theta^4}}{\frac{2}{\theta^2}}+\sum_{i=1}^{10}(-X_i^2)\cdot\frac{-2\theta}{\theta^4}-11^2\cdot 3\,\frac{-2\theta}{\theta^4}=-\frac{20}{\theta}+\frac{1}{\theta^3}\left(\Sigma X_i^2(-2)+(-2)\cdot 11^2\cdot 3\right)=$$

$$=-\frac{20}{\theta}+\frac{1}{\theta^3}(637{,}96+726)$$

$$\frac{dL}{d\theta}=0\longrightarrow -20+\frac{1}{\hat\theta^2}(1363{,}96)=0;\quad \frac{1}{\hat\theta^2}=\frac{20}{1363{,}96},\quad \hat\theta=\sqrt{\frac{1363{,}96}{20}}$$

$$\boxed{\hat\theta=8{,}258\text{ cm}}$$

(5.7) Una compañía ha encuestado a personas elegidas al azar hasta encontrar 20 que utilicen su producto. Estimar por m.v. la prop. de consumidores si el nº total de entrevistados ha sido 115.

Muestra $X_1, X_2\ldots X_{115}\longrightarrow B(n,p)$

$$\ell(p)=K\,p^{20}(1-p)^{115-20};\quad L(p)=\ln K+20\ln(p)+95\ln(1-p)$$

$$\frac{dL}{dp}=20\frac{1}{p}+95\frac{-1}{1-p}=\frac{20-20p-95p}{p(1-p)}=\frac{20-115p}{p(1-p)};\quad\frac{dL}{dp}=0\rightarrow$$

$$\rightarrow 20-115\hat p=0,\quad\boxed{\hat p=\frac{20}{115}=0{,}174\rightarrow 17{,}4\%}$$

5.8

Se supone que el tiempo que tarda un paciente en recuperarse al aplicar el tratamiento A es una v.a. exponencial. Al observar $\boxed{20}$ pacientes al cabo de 20 días $\boxed{15}$ no se han recuperado y los otros $\boxed{5}$ tardaron 17, 15, 19, 18 y 17 días. Estimar por máx. verosimilitud el tiempo medio hasta la recuperación.

$$X \to \text{exponencial}(\lambda) \equiv f(t) = \lambda e^{-\lambda t},\quad F(t) = 1 - e^{-\lambda t}$$

$$\ell(\lambda) = \underbrace{(p(t>20))}_{1-F(20)}^{15} \cdot f(t_{16}) \cdot f(t_{17}) \cdots f(t_{20}) =$$

$$= (e^{-\lambda 20})^{15} \cdot \lambda e^{-\lambda 17} \lambda e^{-\lambda 15} \cdots \lambda e^{-\lambda 17} = \lambda^5 e^{-\lambda(15\cdot 20 + \underbrace{\sum_{i=16}^{20} t_i}_{86})}$$

$$L(\lambda) = \log(\ell(\lambda)) = 5\log(\lambda) - \lambda(15\cdot 20 + 86)$$

$$\frac{dL}{d\lambda} = 5\frac{1}{\lambda} - 386 = \frac{5-386\lambda}{\lambda};\ \text{si } \frac{dL}{d\lambda} = 0 \longrightarrow 5 - 386\hat{\lambda} = 0$$

$$\hat{\lambda} = \frac{5}{386}$$

$$E[\hat{\lambda}] = \frac{1}{\hat{\lambda}} = \frac{386}{5} = 77{,}2 \longrightarrow \boxed{E[\hat{\lambda}] = 77{,}2 \text{ días}}$$

## 5.11

$\bar{X}$ es la media de una m.a.s. $\to N(\mu,\sigma)$. Se define $\hat{X} = c\bar{X}$ como un estimador para $\mu$. Determinar $c(\mu,\sigma)$ para que $\hat{X}$ tenga ECM mínimo. Calcular $c$ si $\frac{\sigma}{\mu} = 2$

$$ECM = \text{sesgo}^2(\hat{X}) + Var(\hat{X})$$

- $\text{sesgo}(\hat{X}) = E[\hat{X}] - \mu = E[c\bar{X}] - \mu = c\underbrace{E[\bar{X}]}_{\mu} - \mu = c\mu - \mu$ ya que $\bar{X} \to N(\mu, \frac{\sigma}{\sqrt{n}})$
- $Var(\hat{X}) = Var(c\bar{X}) = c^2 Var(\bar{X}) = c^2\frac{\sigma^2}{n}$

$\Longrightarrow$

$$ECM = (c\mu-\mu)^2 + c^2\frac{\sigma^2}{n} = c^2\mu^2 - 2c\mu^2 + \mu^2 + \frac{c^2\sigma^2}{n}$$

$$\frac{dECM}{dc} = 2\mu^2 c - 2\mu^2 + \frac{2c\sigma^2}{n} = 0,\quad c\left(2\mu^2 + \frac{2\sigma^2}{n}\right) = 2\mu^2$$

$$c = \frac{2\mu^2 n}{2\mu^2 n + 2\sigma^2};\quad \boxed{c = \frac{\mu^2}{\mu^2 + \frac{\sigma^2}{n}}}$$

$$\text{Si } \frac{\sigma}{\mu} = 2 \longrightarrow c = \frac{\left(\frac{\mu}{\sigma}\right)^2}{\left(\frac{\mu}{\sigma}\right)^2 + \left(\frac{\sigma^2}{\sigma^2 n}\right)} = \frac{\left(\frac{1}{2}\right)^2}{\left(\frac{1}{2}\right)^2 + \frac{1}{n}} = \frac{\frac{1}{4}}{\frac{n+4}{4n}} = \frac{n}{n+4} \to \boxed{c = \frac{n}{n+4}}$$

(5.15) Un sistema de lectura emplea un mensaje de 128-bit. La probabilidad de un bit cambie dando error, p, es constante y los cambios son independientes. Estimar p si en los últimos 10000 lecturas (todas de 128-bit), 340 eran erroneas.

$p(\text{caracter mal}) = p$

$p(1\,\text{caracter funcione}) = 1-p$

$p(1\,\text{lectura OK}) = (1-p)^{128}$

$p(\text{lectura no OK}) = 1-(1-p)^{128}$

$$X_1, X_2, \dots X_n \longrightarrow B(n,p)$$

$$p(X=x) = \binom{10000}{x} \underbrace{p}_{\text{p. lectura no OK}}{}^{x} \underbrace{(1-p)}_{\text{p lectura OK}}{}^{10000-x}$$

$$l(p) = K\left(1-(1-p)^{128}\right)^{340}\left(\left(1-p\right)^{128}\right)^{10000-340}$$

$$L(p) = \log K + 340 \log\left(1-(1-p)^{128}\right) + 9660 \cdot \log (1-p)^{128}$$

$$\frac{dL}{dp} = 340 \frac{+128(1-p)^{127}}{1-(1-p)^{128}} + 9660 \cdot 128 \frac{-(1-p)^{127}}{(1-p)^{128}} =$$

$$= \frac{43520(1-p)^{127}(1-p)^{128} - 1236480(1-p)^{127} + (1-p)^{128}(1-p)^{127}1236480}{\left(1-(1-p)^{128}\right)(1-p)^{128}}$$

$$\frac{dL}{dp} = 0 \longrightarrow \cancel{(1-\hat{p})^{127}}(1-\hat{p})^{128}(43520+1236480) = 1236480\cancel{(1-\hat{p})^{127}}$$

$$(1-\hat{p})^{128} = \underbrace{\frac{1236480}{1280000}}_{0{,}966}; \quad 1-\hat{p} = \sqrt[128]{0{,}966} = 0{,}99972$$

$$\hat{p} = 0{,}00027$$

(5.13) Para estimar $\sigma^2$ de una población normal se utiliza $\hat{\sigma}^2 = K\hat{S}^2$ ($\hat{S}^2$ var. muestral corregida) Calcular el valor de K que minimiza el ECM

$$FCM = \text{sesgo}^2(\hat{\sigma}^2) + Var(\hat{\sigma}^2)$$

$$E[\hat{\sigma}^2] = E[K\hat{S}^2] = K\,E[\hat{S}^2] = K\sigma^2, \rightarrow \text{Sesgo}(\hat{\sigma}^2) = K\sigma^2 - \sigma^2 = \sigma^2(K-1)$$

$$Var(\hat{\sigma}^2) = Var(K\hat{S}^2) = K^2 Var\Big(\underbrace{\frac{\hat{S}^2(n-1)}{\sigma^2}}_{\chi^2_{n-1}} \cdot \frac{\sigma^2}{n-1}\Big) = \frac{K^2\sigma^4}{(n-1)^2}2(n-1) = \frac{2K^2\sigma^4}{n-1}$$

$$\Rightarrow ECM = \sigma^4(K-1)^2 + \frac{2K^2\sigma_4}{n-1}$$

$$\frac{d\,ECM}{dK} = \sigma^4\left(2(K-1) + \frac{4K}{n-1}\right) = 0, \quad 2(K-1)(n-1) + 4K = 0,$$

$$\cancel{2}Kn - \cancel{2}K - \cancel{2}n + \cancel{2} + \underset{2}{\cancel{4}}K = 0, \quad K(n-1+2) = n-1, \quad \boxed{K = \frac{n-1}{n+1}}$$

**(5.14)** Los tiempos de funcionamiento de dos componentes siguen distribuciones exponenciales con esperanzas $\mu$ y $2\mu$. Se obtienen los tiempos de fallo de una m.a. de cada tipo de componente de tamaño $n$. Obtener el estimador por máx. verosimilitud de $\mu$, calcular su media y su varianza.

$X$ v.a tiempo de funcionamiento del componente 1 $\longrightarrow f_X(X) = \frac{1}{\mu} e^{-X/\mu}$

$Y$ v.a " " " " " 2 $\longrightarrow f_Y(Y) = \frac{1}{2\mu} e^{-Y/2\mu}$

$$\ell(\lambda) = f_X(X_1) \dots f_X(X_n) f_Y(Y_1) \dots f_Y(y_n) = \left(\frac{1}{\mu}\right)^n e^{-\frac{\Sigma x_i}{\mu}} \left(\frac{1}{2\mu}\right)^n e^{-\frac{\Sigma y_i}{2\mu}}$$

$$L(\lambda) = n \log \frac{1}{\mu} - \frac{\Sigma X_i}{\mu} + n \log\left(\frac{1}{2\mu}\right) - \frac{\Sigma y_i}{2\mu}$$

$$\frac{dL}{d\lambda} = -n \frac{1}{\mu} + \frac{\Sigma x_i}{\mu^2} - n \frac{1}{\mu} + \frac{2 \Sigma y_i}{4\mu^2} = \frac{-4\mu n + 4 \Sigma X_i - 4\mu n + 2\Sigma Y_i}{4\mu^2}$$

$$\frac{dL}{d\lambda} = 0 \longrightarrow \hat{\mu}(-4n - 4n) = -\Sigma(4X_i + 2Y_i), \quad \boxed{\hat{\mu} = \frac{\Sigma(2X_i + Y_i)}{4n}}$$

$$E[\hat{\mu}] = E\left[\frac{\Sigma(2X_i + Y_i)}{4n}\right] = \frac{1}{2} \underbrace{E\left[\frac{\Sigma X_i}{n}\right]}_{\mu} + \frac{1}{4} \underbrace{E\left[\frac{\Sigma Y_i}{n}\right]}_{2\mu} = \frac{1}{2}\mu + \frac{1}{4} 2\mu = \mu$$

$$\boxed{E[\hat{\mu}] = \mu}$$

$$Var[\hat{\mu}] = Var\left[\frac{2 \Sigma X_i}{4n} + \frac{\Sigma Y_i}{4n}\right] = \frac{1}{4n^2}\left[\overset{\mu^2}{Var(X_1)} + Var(X_2) \dots\right] + \frac{1}{16n^2}\left[\overset{4\mu^2}{Var(Y_1)} + \dots + Var(Y_n)\right] =$$

$$= \frac{1}{4n} \mu^2 + \frac{1}{16n} 4\mu^2 = \frac{\mu^2 + \mu^2}{4n} = \frac{\mu^2}{2n} \longrightarrow \boxed{Var(\hat{\mu}) = \frac{\mu^2}{2n}}$$

# TEMA 5 RESUMEN

## Método de los momentos:

Consiste en hallar los momentos de grado 1 hasta r de una m.a.s. con r parámetros a estimar e igualarlos a los de la población

$$\underbrace{E[x] = \bar{x}}_{\downarrow\ \mu} \ ; \quad \underbrace{E[x^2] = \sigma^2 + \mu^2}_{E[x^2] - \bar{x}^2 = s^2 = \hat{\sigma}^2} ; \ \ldots$$

## Método de máxima verosimilitud

1) Cálculo de la función de verosimilitud: $\mathcal{L}(\Theta) = f(x_1)\, f(x_2) \ldots f(x_n)$

2) Función soporte $L(\Theta) = \log[\ell(\Theta)]$

3) Derivamos y calculamos el máximo de $L(\Theta)$

$$\frac{dL(\Theta)}{d\Theta} = 0 \longrightarrow \hat{\Theta}$$

| | MOMENTOS | MÁXIMA VEROSIMILITUD |
|---|---|---|
| NORMAL | $\hat{\mu}_m = \bar{x}$ ; $\hat{\sigma}^2_m = s^2$ | $\mu_{mv} = \bar{x}$ , $\hat{\sigma}^2_{mv} = s^2$ |
| POISSON | $\hat{\lambda}_m = \bar{x}$ | $\hat{\lambda}_{mv} = \bar{x}$ |
| BERNOUILLI | $\hat{p}_m = r/n$ | $p_{mv} = r/n$ |

## Distribuciones normales

$X_1, X_2, \ldots X_n$ m.a.s de una normal $N(\mu, \sigma)$

→ MEDIA $\bar{X}$

$$\boxed{E[\bar{X}] = \mu \ ; \quad Var[\bar{X}] = \frac{\sigma^2}{n} \ ; \quad \bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

→ VARIANZA $S^2$

$$\boxed{E[S^2] = \frac{n-1}{n}\sigma^2 \ ; \quad Var[S^2] = \frac{2(n-1)\sigma^4}{n^2}}$$

→ VARIANZA CORREGIDA $\hat{S}^2$

$$\boxed{E[\hat{S}^2] = \sigma^2 \ ; \quad (n-1)\hat{S}^2 = nS^2 \ ; \quad Var(\hat{S}^2) = \frac{2\sigma^4}{n-1}}$$

## DISTRIBUCIÓN $\chi^2$

$$z_1, z_2, \ldots z_n \longrightarrow N(0,1) \Rightarrow z_1^2 + z_2^2 + z_3^2 \ldots + z_n^2 = \chi^2_n$$

$$\boxed{E[\chi^2_n] = n \ ; \quad Var[\chi^2_n] = 2n \ ; \quad \chi^2_n + \chi^2_m = \chi^2_{n+m}}$$

$$\boxed{\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \longrightarrow \chi^2_{n-1}}$$

## ● Aproximaciones a la normal

$\hat{p}$ de una Binomial $(n,p)$

$$\hat{p} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\boxed{\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)}$$

$\hat{\lambda}$ de Poisson

$$\hat{\lambda} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\boxed{\hat{\lambda} \sim N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)}$$

## ● Propiedades de los estimadores

- Centrados o insesgados $\boxed{E[\hat{\theta}] = \theta}$ $\boxed{\text{Sesgo} = E[\hat{\theta}] - \theta}$

- Varianza mínima $\boxed{Var(\hat{\theta}_{MV}) \leq Var(\hat{\theta})}$

- Error cuadrático medio mínimo

$$\boxed{ECM = E[(\hat{\theta} - \theta)^2] = \text{sesgo}^2(\hat{\theta}) + Var(\hat{\theta})}$$

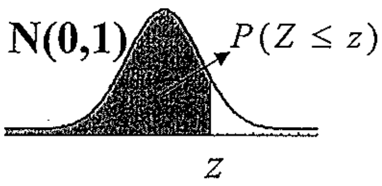
- Consistentes $\boxed{\lim E[\hat{\theta}] = \theta \quad y \quad \lim_{n \to \infty} Var[\hat{\theta}] = 0}$

|||

Asintóticamente centrados y varianza asintóticamente nula.

# *TABLA*

# *Normal Estandar*



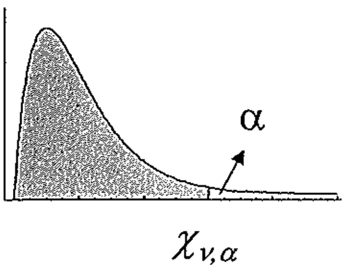
Ejemplo.

$P(Z \leq 1.96) = 0.9750$

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

---

# Tabla $\chi^2$



v: grados de libertad (g.l.)

EJEMPLO

$P(\chi_9 \geq 19{,}02) = 0{,}025$

| | | | | | α | | | | |
|---|---|---|---|---|---|---|---|---|---|
| g.l. | 0,995 | 0,990 | 0,975 | 0,950 | 0,500 | 0,050 | 0,025 | 0,010 | 0,005 |
| 1 | ,00004 | ,00016 | ,00098 | ,00393 | 0,455 | 3,841 | 5,024 | 6,635 | 7,879 |
| 2 | ,01002 | ,0201 | 0,051 | 0,103 | 1,386 | 5,991 | 7,378 | 9,210 | 10,60 |
| 3 | ,0717 | 0,115 | 0,216 | 0,352 | 2,366 | 7,815 | 9,348 | 11,34 | 12,84 |
| 4 | 0,207 | 0,297 | 0,484 | 0,711 | 3,357 | 9,488 | 11,14 | 13,28 | 14,86 |
| 5 | 0,412 | 0,554 | 0,831 | 1,145 | 4,351 | 11,07 | 12,83 | 15,09 | 16,75 |
| 6 | 0,676 | 0,872 | 1,237 | 1,635 | 5,348 | 12,59 | 14,45 | 16,81 | 18,55 |
| 7 | 0,989 | 1,239 | 1,690 | 2,167 | 6,346 | 14,07 | 16,01 | 18,48 | 20,28 |
| 8 | 1,344 | 1,647 | 2,180 | 2,733 | 7,344 | 15,51 | 17,53 | 20,09 | 21,95 |
| 9 | 1,735 | 2,088 | 2,700 | 3,325 | 8,343 | 16,92 | 19,02 | 21,67 | 23,59 |
| 10 | 2,156 | 2,558 | 3,247 | 3,940 | 9,342 | 18,31 | 20,48 | 23,21 | 25,19 |
| 11 | 2,603 | 3,053 | 3,816 | 4,575 | 10,341 | 19,68 | 21,92 | 24,73 | 26,76 |
| 12 | 3,074 | 3,571 | 4,404 | 5,226 | 11,340 | 21,03 | 23,34 | 26,22 | 28,30 |
| 13 | 3,565 | 4,107 | 5,009 | 5,892 | 12,340 | 22,36 | 24,74 | 27,69 | 29,82 |
| 14 | 4,075 | 4,660 | 5,629 | 6,571 | 13,339 | 23,68 | 26,12 | 29,14 | 31,32 |
| 15 | 4,601 | 5,229 | 6,262 | 7,261 | 14,339 | 25,00 | 27,49 | 30,58 | 32,80 |
| 16 | 5,142 | 5,812 | 6,908 | 7,962 | 15,338 | 26,30 | 28,85 | 32,00 | 34,27 |
| 17 | 5,697 | 6,408 | 7,564 | 8,672 | 16,338 | 27,59 | 30,19 | 33,41 | 35,72 |
| 18 | 6,265 | 7,015 | 8,231 | 9,390 | 17,338 | 28,87 | 31,53 | 34,81 | 37,16 |
| 19 | 6,844 | 7,633 | 8,907 | 10,117 | 18,338 | 30,14 | 32,85 | 36,19 | 38,58 |
| 20 | 7,434 | 8,260 | 9,591 | 10,851 | 19,337 | 31,41 | 34,17 | 37,57 | 40,00 |
| 21 | 8,034 | 8,897 | 10,283 | 11,591 | 20,337 | 32,67 | 35,48 | 38,93 | 41,40 |
| 22 | 8,643 | 9,542 | 10,982 | 12,338 | 21,337 | 33,92 | 36,78 | 40,29 | 42,80 |
| 23 | 9,260 | 10,196 | 11,689 | 13,091 | 22,337 | 35,17 | 38,08 | 41,64 | 44,18 |
| 24 | 9,886 | 10,856 | 12,401 | 13,848 | 23,337 | 36,42 | 39,36 | 42,98 | 45,56 |
| 25 | 10,520 | 11,524 | 13,120 | 14,611 | 24,337 | 37,65 | 40,65 | 44,31 | 46,93 |
| 26 | 11,160 | 12,198 | 13,844 | 15,379 | 25,336 | 38,89 | 41,92 | 45,64 | 48,29 |
| 27 | 11,808 | 12,878 | 14,573 | 16,151 | 26,336 | 40,11 | 43,19 | 46,96 | 49,65 |
| 28 | 12,461 | 13,565 | 15,308 | 16,928 | 27,336 | 41,34 | 44,46 | 48,28 | 50,99 |
| 29 | 13,121 | 14,256 | 16,047 | 17,708 | 28,336 | 42,56 | 45,72 | 49,59 | 52,34 |
| 30 | 13,787 | 14,953 | 16,791 | 18,493 | 29,336 | 43,77 | 46,98 | 50,89 | 53,67 |
| 40 | 20,707 | 22,164 | 24,433 | 26,509 | 39,335 | 55,76 | 59,34 | 63,69 | 66,77 |
| 50 | 27,991 | 29,707 | 32,357 | 34,764 | 49,335 | 67,50 | 71,42 | 76,15 | 79,49 |
| 60 | 35,534 | 37,485 | 40,482 | 43,188 | 59,335 | 79,08 | 83,30 | 88,38 | 91,95 |
| 70 | 43,275 | 45,442 | 48,758 | 51,739 | 69,334 | 90,53 | 95,02 | 100,43 | 104,21 |
| 80 | 51,172 | 53,540 | 57,153 | 60,391 | 79,334 | 101,88 | 106,63 | 112,33 | 116,32 |
| 90 | 59,196 | 61,754 | 65,647 | 69,126 | 89,334 | 113,15 | 118,14 | 124,12 | 128,30 |
| 100 | 67,328 | 70,065 | 74,222 | 77,929 | 99,334 | 124,34 | 129,56 | 135,81 | 140,17 |
| 120 | 83,852 | 86,923 | 91,573 | 95,705 | 119,334 | 146,57 | 152,21 | 158,95 | 163,65 |

# TEMA 6 RESUMEN

## Pasos para construir un intervalo de confianza

① Construir estadístico- pivote [función que depende del verdadero valor del parámetro y su estimación]

② Fijar el nivel de confianza ($\alpha=0,05$ o $\alpha=0,01$)

③ Obtener el intervalo con los datos de la m.a.s.

## Intervalo para $\mu$ con $\sigma$ conocido en una normal

$$\boxed{\mu \in \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

## Intervalo para $\mu$ con $\sigma$ desconocido en una normal

$$\boxed{\mu \in \bar{x} \pm t_{n-1,\alpha/2} \frac{\hat{S}}{\sqrt{n}}}$$

## Intervalo para $\sigma^2$ en una normal

$$\boxed{\frac{(n-1)\hat{S}^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{n-1\ \hat{S}^2}{\chi_{\alpha/2}}}$$

## t de Student

Distribución que se construye

$$\boxed{\frac{N(0,1)}{\sqrt{\frac{\chi^2_g}{g}}} \longrightarrow t_g}$$

## Intervalo para $\lambda$ de Poisson

$$\boxed{\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}}$$

# Intervalos de Confianza

## 1 PASOS PARA DEFINIR UN INTERVALO DE CONFIANZA

1) Construir el estadístico pivote $g(\theta,\hat{\theta})$

Es una función que depende del verdadero valor del parámetro y de su estimación

2) Fijar el nivel de confianza

$P(a \le g(\theta,\hat{\theta}) \le b) = 1-\alpha$ ; $1-\alpha$ es la confianza que suele ser: $\begin{cases} 0{,}95 \\ \downarrow \alpha = 0{,}05 \\ 0{,}99 \\ \downarrow \alpha = 0{,}01 \end{cases}$

3) Obtener el intervalo.

$$a \le g(\theta,\hat{\theta}) \le b \longrightarrow \underbrace{g^{-1}(a,\hat{\theta})}_{\theta_1} \le \theta \le \underbrace{g^{-1}(b,\hat{\theta})}_{\theta_2}$$

$$\theta_1 \le \theta \le \theta_2$$

➡ Se ha realizado una encuesta a 400 personas para estimar la proporción p de votantes de un partido. Resultado de la encuesta: $\begin{cases} \text{Si} & 220 \\ \text{No} & 180 \\ & \overline{400} \end{cases}$

$$X \to B(n,p) \longrightarrow N(np, \sqrt{np(1-p)})$$

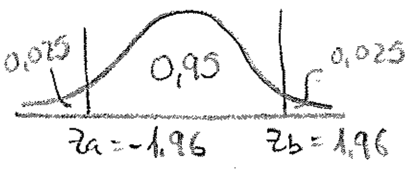
$$\hat{p} = \frac{X}{n} \longrightarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

① Estadístico pivote:

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow N(0,1)$$

②

Fijamos un nivel de confianza del 95 % ⟶



$$-1{,}96 \le \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le 1{,}96 \; ; \quad \hat{p} - 1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + 1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Suponemos la varianza conocida $\equiv -1{,}96 \le \dfrac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le 1{,}96$

En nuestro caso $\hat{p} = \frac{220}{400} = 0{,}55 \longrightarrow$ $\boxed{p \in 0{,}55 \pm 1{,}96\sqrt{\frac{0{,}55 \cdot 0{,}45}{400}}}$

## 2 INTERVALO PARA $\mu$ CON $\sigma$ CONOCIDO EN UNA NORMAL

$X_1, X_2, — X_n \longrightarrow N(\mu, \sigma)$

$\bar{X} \longrightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$ ; Estadístico pivote: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$

Fijando la confianza obtendríamos

$$-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2} \; ;$$

$$\boxed{\mu \in \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Por tanto el intervalo será:

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

## 3 INTERVALO PARA $\mu$ CON $\sigma$ DESCONOCIDO

$X_1, X_2, — X_n \longrightarrow N(\mu, \sigma)$

$\bar{X} \longrightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$ Estadístico pivote $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$

Como $\sigma$ no es conocido tenemos que estimarlo mediante el estimador $\hat{S}^2$ (ya que $E[\hat{S}^2] = \sigma^2$). Por tanto nos queda:

Estadístico pivote $\Big\}$ $\frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \longrightarrow t_{n-1}$ (distribución t de student) que fijando la confianza:

$$-t_{n-1,\alpha/2} \le \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \le t_{n-1,\alpha/2}$$

$$\boxed{\mu \in \bar{X} \pm t_{n-1,\alpha/2}\frac{\hat{S}}{\sqrt{n}}}$$

Por tanto el intervalo será:

$$\bar{X} - t_{n-1,\alpha/2}\frac{\hat{S}}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,\alpha/2}\frac{\hat{S}}{\sqrt{n}}$$

**NOTA!! t de Student**

$$\boxed{\frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \longrightarrow t_{n-1}}$$

Por tanto, como $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow N(0,1)$ y $\frac{(n-1)\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{n-1}$

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\bar{X}-\mu/\frac{\sigma}{\sqrt{n}}}{\sqrt{\frac{(n-1)\hat{S}^2}{(n-1)\sigma^2}}} = \underbrace{\boxed{\frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1}}}_{\text{estadístico pivote}}$$

# 4 INTERVALO PARA $\sigma^2$ DE UNA NORMAL

$X_1, X_2 \dots X_n \longrightarrow N(\mu,\sigma)$

$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} \longrightarrow \frac{(n-1)\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{n-1}$$

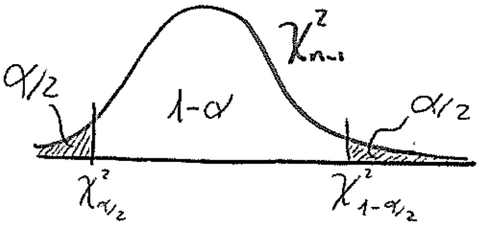
que sería el estadístico pivote

Fijando la confianza y trabajando como si $\chi^2_{n-1}$ fuere simétrica tenemos:

$$P\left(\chi^2_{\alpha/2} \le \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

Por tanto

$$\boxed{\frac{(n-1)\hat{S}^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)\hat{S}^2}{\chi^2_{\alpha/2}}}$$



➡ La resistencia de compresión de 15 probetas de acero elegidos al azar es:

40,15 | 65,10 | 49,50 | 22,40 | 38,20 | 60,40 | 43,40 | 26,35 | 31,20 | 55,60 | 47,25 | 73,20 | 35,90 | 45,25 | 52,40.

a) Calcular un intervalo de confianza para $\mu$ con confianza 99%

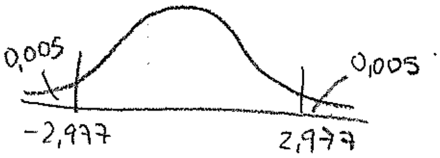
$\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ Por tanto el estadístico pivote será $\longrightarrow \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow N(0,1)$

Como $\sigma$ es desconocido usaremos la var generalizada $\hat{S}$

Por tanto, el nuevo estadístico pivote queda

$$\frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1} \qquad \begin{cases} n=15 \\ \bar{X}=45,75 \\ \hat{S}=14,20 \end{cases} \qquad \frac{45,75-\mu}{\frac{14,2}{\sqrt{15}}} \longrightarrow t_{14}$$

Fijando la confianza en 0,99

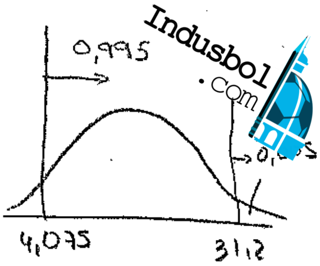


$$-2,977 \le \frac{45,75-\mu}{\frac{14,2}{\sqrt{15}}} \le 2,977 \,; \qquad 45,75-2,977\frac{14,2}{\sqrt{15}} \le \mu \le 45,75+2,977\cdot\frac{14,2}{\sqrt{15}}$$

$$\boxed{34,8 \le \mu \le 56,7}$$

b) Calcular un intervalo con confianza 99% para $\sigma^2$

$$\frac{(n-1)\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{n-1} \left\{ \begin{array}{l} n=15 \\ \hat{S}^2 = 14{,}2^2 = \end{array} \right. \quad \frac{14\cdot 14{,}2^2}{\sigma^2} \longrightarrow \chi^2_{14}$$



Por tanto: $4{,}075 \leq \frac{14\cdot 14{,}2^2}{\sigma^2} \leq 31{,}2$ y despejando:

$$\frac{14\cdot(14{,}2)^2}{31{,}32} \leq \sigma^2 \leq \frac{14\cdot(14{,}2)^2}{4{,}075}, \quad \boxed{90{,}13 \leq \sigma^2 \leq 692{,}75}$$

## 5 INTERVALO PARA $\lambda$ DE UNA POISSON

$X_1, X_2, — X_n \longrightarrow$ Poisson $(\lambda)$

$\lambda = \overline{X} \rightsquigarrow N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$, Por tanto el estadístico pivote será:

$$\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda}{n}}} \longrightarrow N(0,1)$$

Por tanto $P\left(-z_{\alpha/2} \leq \frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \leq z_{\alpha/2}\right) = 1-\alpha$

Para despejar $\lambda$ suponemos la varianza conocida $(\hat{\lambda})$ y tenemos:

$$\boxed{\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}}$$

(6.5) Se desea conocer la aceptación de un nuevo producto. Se realiza una encuesta a 200 personas y 37 manifiestan disposición a comprarlo. Obtener un intervalo de confianza ($\alpha = 0{,}05$) para la prop. $p$ de compradores potenciales del nuevo producto. ¿Cual debería ser el tamaño muestral para reducir el intervalo a la mitad?

$X_1, X_2, \ldots X_{200} \longrightarrow B(n,p)$ ; $\hat{p} = \frac{37}{200} = 0{,}185$

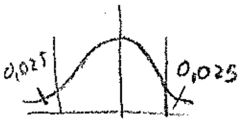
$\hookrightarrow N(np, \sqrt{np(1-p)})$

$\hat{p} = \frac{X}{n}$

$\hat{p} \rightsquigarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ Por tanto el estadístico pivote sera

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow N(0,1)$$

Confianza $(\alpha = 0{,}05)$



$$P\left(-1{,}96 \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1{,}96\right) = 0{,}95; \quad p \in 0{,}185 \pm 1{,}96\sqrt{\frac{0{,}185\cdot(1-0{,}185)}{200}}, \quad \boxed{\begin{array}{l} p \in [0{,}131, 0{,}239] \\ \alpha = 0{,}05 \end{array}}$$

$$\overbrace{\hat{p}-1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \hat{p} \qquad \hat{p}+1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}^{L=2\cdot 1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Suponemos $\hat{p}=$ cte . $L=2L'$

$$\cancel{2}\cdot\cancel{1{,}96}\sqrt{\frac{0{,}185(1-0{,}185)}{200}}=2\cdot\left(\cancel{2}\,\cancel{1{,}96}\sqrt{\frac{0{,}185(1-0{,}185)}{n'}}\right)$$

$$\frac{\cancel{0{,}185(1-0{,}185)}}{200}=4\,\frac{\cancel{0{,}185(1-0{,}185)}}{n'}\ ;\quad \boxed{n'=800}$$

(6.6) Se desea encontrar la prop. de niños correctamente vacunados. Se desea que $|\hat{p}-p|\le 0{,}05$ con probabilidad 0,95. ¿Cuál es el tamaño muestral mínimo requerido?

$$\hat{p}\rightsquigarrow N\left(p,\sqrt{\frac{p(1-p)}{n}}\right) \text{ por tanto } \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\longrightarrow N(0,1)$$

$$P\left(|\hat{p}-p|\le 0{,}05\right)=0{,}95\ ;\ P\left(\frac{|\hat{p}-p|}{\sqrt{\frac{p(1-p)}{n}}}\le\frac{0{,}05}{\sqrt{\frac{p(1-p)}{n}}}\right)=0{,}95$$

$$\Longrightarrow \frac{0{,}05}{\sqrt{\frac{p(1-p)}{n}}}=1{,}96 \longrightarrow \text{Ahora tenemos 2 opciones}$$

1º) Opción conservadora o por Bernoulli. Buscamos p para máx. Varianza

| X | p(X=x) |
|---|---|
| 0 | p |
| 1 | 1-p |

$$Var(X)=p(1-p)\ ;\ \frac{d\,Var(X)}{dp}=1-2p=0\longrightarrow p=0{,}5$$

Por tanto $\longrightarrow \frac{0{,}05}{\sqrt{\frac{0{,}5(1-0{,}5)}{n}}}=1{,}96$ ; $0{,}05=\frac{1{,}96\cdot 0{,}5}{\sqrt{n}}$ ;

$$0{,}05\sqrt{n}=1{,}96\cdot 0{,}5\ ,\ n=\left(\frac{1{,}96\cdot 0{,}5}{0{,}05}\right)^2\ ,\ \boxed{n=384}$$

2º) Opción con muestra piloto de la que se obtiene un $\hat{P}_p$

$$\hat{P}_p\Longrightarrow\frac{0{,}05}{\sqrt{\frac{\hat{P}_p(1-\hat{P}_p)}{n}}}=1{,}96 \text{ y despejando sacamos } n$$

**6.7** Queremos determinar el nivel de radioactividad $\Theta$ de una roca lunar que se mide por nº medio de partículas emitidas por hora. Después de 15h se contabilizan 3547 partículas. Nº partículas emitidas → Poisson ($\Theta$), dar un intervalo para el nivel de radioactividad de la roca con confianza 0,95.

USAR → $Z \to N(0,1)$ entonces $P(Z \leq 1{,}96) = 0{,}975$

$\Theta$ v.a. nº partículas emitidas por h.

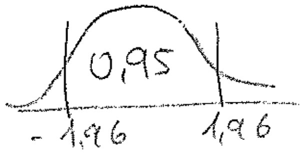
$$\hat{\Theta} = \frac{3547}{15} = 236{,}47 \text{ part}/h$$ (nº medio de part. por hora)

Creo estadístico pivote ya que $\hat{\Theta} \to N\left(\Theta, \sqrt{\frac{\Theta}{n}}\right)$;

$Var(\hat{\Theta}) = \frac{\Theta}{n}$ (por ser poisson)

$E[\hat{\Theta}] = \Theta$

Estadístico pivote: $\frac{\hat{\Theta} - \Theta}{\sqrt{\frac{\Theta}{n}}} \longrightarrow N(0,1)$, y suponiendo conocida la varianza con nivel de confianza 95%



$$P\left(\frac{\hat{\Theta} - \Theta}{\sqrt{\frac{\hat{\Theta}}{n}}} \leq 1{,}96\right) = 0{,}975;$$

$\hat{\Theta} = 236{,}47$, $n = 15$

$$\hat{\Theta} - 1{,}96\sqrt{\frac{\hat{\Theta}}{n}} \leq \Theta \leq \hat{\Theta} + 1{,}96\sqrt{\frac{\hat{\Theta}}{n}} \Longrightarrow \boxed{\Theta \in [228{,}68, 244{,}25]}$$

**6.8** $X_1, X_2, \ldots X_n$ es una m.a.s. de una v.a. exponencial de $f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}$; $x \geq 0$, $\lambda > 0$

El estadístico $U = 2n\bar{X}/\lambda$ tiene distribución $\chi^2_{2n}$; donde $\bar{X} = (X_1 + X_2 + \ldots + X_n)/n$

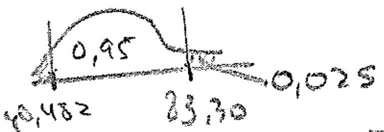
El tiempo de funcionamiento de un equipo es una v.a. con distrib. exponencial. Se toman 30 equipos y sus tiempos de funcionamiento hasta el fallo, obteniéndose $6{,}2 \cdot 10^3$ h de media. Calcular un intervalo con un 95% de confianza para la vida media del equipo.

X v.a. tiempo de vida. $E[X] = \text{vida media} = \lambda$

Tenemos que $\hat{\lambda} = 6{,}2 \cdot 10^3 h = \bar{x}$

Estadístico pivote $\longrightarrow \frac{2n\bar{X}}{\lambda} \longrightarrow \chi^2_{2n}$, $n = 30$

Confianza 95%



$$\Longrightarrow 40{,}482 \leq \frac{2 \cdot 30 \cdot 6{,}2 \cdot 10^3}{\lambda} \leq 83{,}30$$

$$\frac{2 \cdot 30 \cdot 6{,}2 \cdot 10^3}{83{,}3} \leq \lambda \leq \frac{2\,30\,6{,}2\,10^3}{40{,}482} \longrightarrow \boxed{\lambda \in [4{,}465 \cdot 10^3 h; 9{,}189 \cdot 10^3 h] \text{ confianza } 95\%}$$

**6.11** Los núcleos del elemento radioactivo $C^{14}$ se desintegran aleatoriamente. El tiempo que tarda en desintegrarse cada nucleo es una v.a. con distribución exponencial de media $8,27 \cdot 10^3$ años.

a) Inicialmente hay $10^{12}$ nucleos, obtener el número esperado de nucleos sin desintegrar al cabo de 20 000 años.

b) Obtener, para la v.a. nº de núcleos sin desintegrar al cabo de 20000 años un intervalo que contenga el valor de la variable con probabilidad 0,95 e interpretar resultado.

T v.a. tiempo que tarda un nucleo en desintegrarse $\longrightarrow E[T] = 8,27\ 10^3$ años

a) $N_0 = 10^{12}$

Y v.a. nº nucleos sin desintegrar al cabo de 20 000 años de un total de $10^{12}$.   $Y \rightarrow B(n = 10^{12}, p = P(T > 20000))$

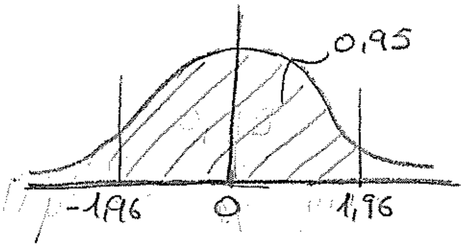
T es exp. $F_T(T) = 1 - e^{-\frac{1}{8,27 \cdot 10^3} T}$

Se pide $E[Y]$

Buscamos $p = P(T > 20000) = 1 - F_T(20\,000) = e^{-\frac{20000}{8,27 \cdot 10^3}} = 0,089$

$$\underset{\substack{\uparrow \\ Y \rightarrow \text{Binomial}(n,p)}}{E[Y] =} n \cdot p = 10^{12} \cdot 0,089 = \boxed{8,9 \cdot 10^{10} \text{ nucleos}}$$

b) $E[Y] = 8,9 \cdot 10^{10}$



$Y \sim\!\!> N(np, \sqrt{np(1-p)})$

$Y \sim\!\!> N(8,9 \cdot 10^{10}, \sqrt{8,9 \cdot 10^{10}(1 - 0,089)}) = (8,9 \cdot 10^{10}, 2,85 \cdot 10^5)$

$$\frac{Y - 8,9 \cdot 10^{10}}{2,85 \cdot 10^5} \longrightarrow N(0,1); \quad -1,96 \leq \frac{Y - 8,9 \cdot 10^{10}}{2,85 \cdot 10^5} \leq +1,96$$

$Y \in [8,9 \cdot 10^{10} - 5,58 \cdot 10^5, 8,9 \cdot 10^{10} + 5,58 \cdot 10^5]$ con $\alpha = 0,05$

$\boxed{Y \in [\ 8,8999 \cdot 10^{10}, 8,9001 \cdot 10^{10}\ ]} \longrightarrow$ Muy preciso ya que el intervalo es pequeño

c) T = 20000 años. Una pieza tiene $10^{10}$ radionucleidos. Estimar por el método de los momentos el nº de nucleos inicial

$E[Y] = n \cdot p$;   $Y_i = 10^{10} = \hat{n} p \xrightarrow[\substack{\uparrow \\ p = 8,9 \cdot 10^{-2}}]{}$   $\boxed{\hat{n} = \frac{10^{10}}{8,9 \cdot 10^{-2}} = 1,123 \cdot 10^{11}}$

$\boxed{E[\hat{n}] = E\left[\frac{Y_i}{P}\right] = \frac{nP}{P} = n}$, $Var[\hat{n}] = Var\left(\frac{Y_i}{P}\right] = \frac{np(1-P)}{P^2} = \boxed{\frac{n(1-P)}{P^2}}$

**6.1** M.A.S, n=12 de una pob normal

30.2, 30.8, 29.3, 29, 30.9, 30.8, 29.7, 28.9, 30.5, 31.2, 31.3, 28.5

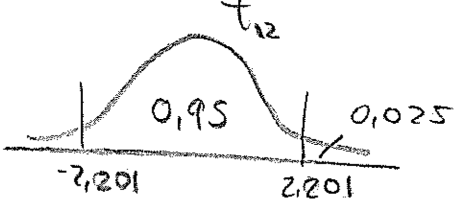
a) Intervalo de confianza de la media. $\alpha=0.05$

b) " " " para la varianza. $\alpha=0,05$

a) $X_1, X_2, \dots X_n \longrightarrow N(\mu, \sigma)$

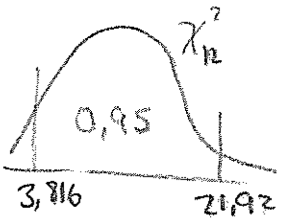
$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \qquad \frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{12}}} \longrightarrow t_{n-1},$$

$$\uparrow \frac{\hat{S}}{\sqrt{n}}$$



$$\begin{cases} \bar{X}=30,09 \\ \hat{S}=0,97 \\ n=12 \end{cases} \qquad \mu \in 30,09 \pm 2,201 \frac{0,97}{\sqrt{12}} \qquad \boxed{29,47 \leq \mu \leq 30,71}$$

b) $$\frac{\hat{S}^2(n-1)}{\sigma^2} \longrightarrow \chi^2_{n-1}$$



$\hat{S}=0,9746$
$n=12$

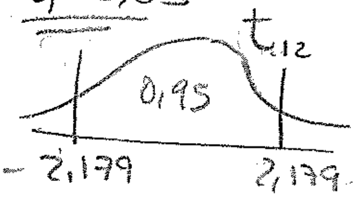
$$2,7 \leq \frac{\hat{S}^2(n-1)}{\sigma^2} \leq 19,02, \quad \frac{\hat{S}^2(n-1)}{21,92} \leq \sigma^2 \leq \frac{\hat{S}^2(n-1)}{3,816} \qquad \boxed{0,476 \leq \sigma^2 \leq 2,738}$$

**6.2** Estimar $\bar{\mu}$, de edad en la que los científicos hacen su principal descubrimiento

MAS: 34, 40, 31, 33, 49, 33, 34, 43, 30, 31, 26, 26, 36 30,37 / 38,[illegible]

$\alpha=0,05$

$$\bar{\mu}=\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \qquad \frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1}$$



$$\mu \in \bar{X} \pm 2,179 \frac{\hat{S}}{\sqrt{n}} \quad \begin{cases} \bar{X}=34,31 \\ \hat{S}=6,51 \\ n=13 \end{cases}$$

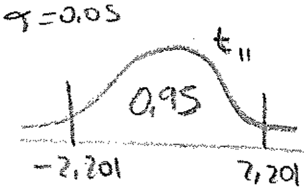
$$\boxed{30,37 \leq \mu \leq 38,24}$$

**6.3** MA.S ≡ n=12 estaciones; $\bar{X}=2340€$, $\hat{S}=815€$. Intervalo de confianza para $\mu$

Nº estaciones que debemos estudiar si se desea que el intervalo tenga una amplitud máxima de 500€

$$\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1}; \quad \mu \in 2340 \pm 2,201 \cdot \frac{815}{\sqrt{12}}$$

$\alpha=0.05$



$$\boxed{1822,17 \leq \mu \leq 2857,83}$$

$$500 \leq 2 \cdot t_{\alpha/2, n-1} \cdot \frac{815}{\sqrt{n}}, \qquad \frac{t_{\alpha/2}}{\sqrt{n}} \geq \frac{500}{2 \cdot 815} = 0,3067$$

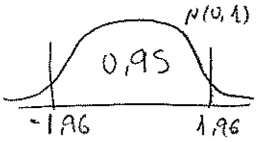
$n=41 \longrightarrow \frac{2,021}{\sqrt{41}} = 0,31$ cumple $\Longrightarrow \boxed{n \approx 41}$

$n=31 \longrightarrow \frac{2,041}{\sqrt{31}} = 0,37$ no cumple

**6.9** 180 llamadas durante las 2 últimas horas.
Obtener intervalo de confianza para el nº llamadas por hora sabiendo que el nº de llamadas durante un periodo T cualquiera sigue una distribución de Poisson.

$\hat{\lambda} = \frac{180}{2} = 90$ llam/h ; $\quad \hat{\lambda} \sim N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right) \Rightarrow \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} \longrightarrow N(0,1)$

← estimamos varianza



$$\lambda \in 90 \pm 1{,}96\sqrt{\frac{90}{2}}$$

$$\boxed{76{,}85 \leq \lambda \leq 103{,}15}$$

# Contrastes de Hipótesis

## 1 CONTRASTE DE HIPÓTESIS. INTRODUCCIÓN, NIVEL DE SIGNIFICACIÓN $\alpha$

Un contraste es una regla de decisión que determina si la hipótesis que formulo es cierta o no.

$H_0$: Llamaremos $H_0$ a la hipótesis inicial o nula

$H_1$: Llamaremos $H_1$ a la hipótesis alternativa, la que hago al fallar $H_0$

→ Se ha realizado una encuesta a 400 personas. p es la prop. de votantes del partido político A. Podemos afirmar que A ganará las elecciones si los resultados de la encuesta fueron
Sí → 220
No → 180

① **Planteamos las hipótesis**

$H_0: p \leq 0{,}5$ (no sabemos si ganará o no)

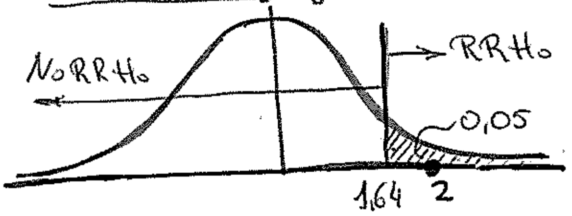
$H_1: p > 0{,}5$ (ganará las elecciones)

¿? El = siempre en $H_0$

② **Medida de discrepancia**

$$\hat{p} \rightsquigarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \Longrightarrow \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0,1)$$

③ **Nivel de significación**: $95\% \rightarrow \alpha = 0{,}05$



④ **Tomo datos de la muestra**

$$\hat{p} = \frac{220}{400} = 0{,}55\,;$$

$p = 0{,}5$ si $H_0$ es cierto

$n = 400$

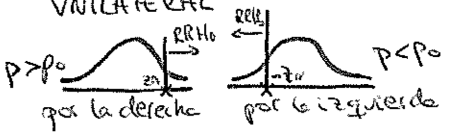
$$Z_0 = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0{,}55-0{,}5}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{400}}} = \underline{\underline{2}} \Longrightarrow$$

Cae en la región de rechazo de $H_0$

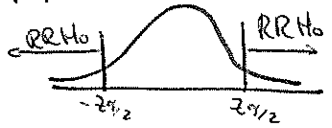
Por tanto, con un nivel de significación $\alpha = 0{,}05$ hay evidencia estadística para decir que el partido A ganará las elecciones

Existen dos tipos de contraste:

$H_0: p = p_0$
$H_1: p > p_0$ o $p < p_0$ $\Longrightarrow$ Contraste UNILATERAL



$H_0: p = p_0$
$H_1: p \neq p_0$ $\Longrightarrow$ Contraste BILATERAL



## 2 TIPOS DE ERRORES

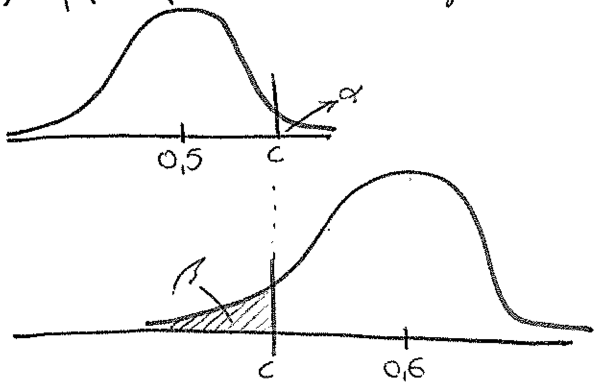
| | | CONTRASTE | |
|---|---|---|---|
| | | Se acepta Ho | Se rechaza Ho |
| REALIDAD | Ho es cierta | OK | ERROR TIPO I<br>$\alpha$ |
| | Ho es falsa | ERROR TIPO II<br>$\beta$ | OK |

$p(\text{error tipo I}) = p(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) = \alpha$

$p(\text{error tipo II}) = p(\text{aceptar } H_0 \mid H_0 \text{ es falsa}) = \beta$

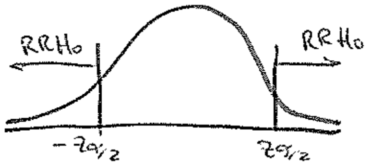
Ho cierta
$p = 0,5$

Ho falsa
$p = 0,6$



Generalmente
$\beta > \alpha$

## 3 NIVEL CRÍTICO O P-VALOR



Sea $H_0: p = p_0$
$H_1: p \neq p_0$

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1) \longrightarrow$$

Rechazare Ho si

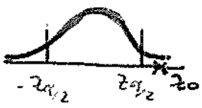
$$\left|\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right| > z_{\alpha/2} \Longrightarrow \hat{p} > p_0 + z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}$$

o

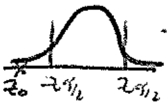
$$\hat{p} < p_0 - z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}$$

Tomo la muestra: $z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ si

→ Aceptamos Ho

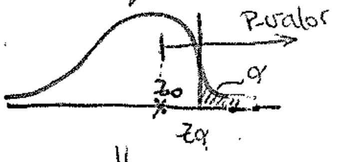


ó



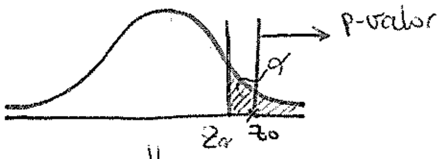
→ Rechazo Ho

El nivel crítico o p-valor es la probabilidad de que la medida de discrepancia sea mayor que la evidencia muestral. Esto es:



⇓

p-valor > α
No rechazo Ho



⇓

p-valor < α
Rechazo Ho

Cuanto más pequeño es el p-valor, más evidencia hay para rechazar la hipótesis nula.

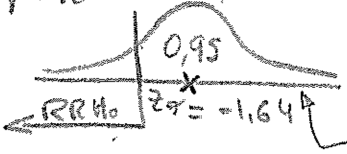
p-valor < 0,01 → Rechazo Ho
p-valor > 0,05 → No Rechazo Ho
0,01 < p-valor < 0,05 → Depende del α escogido

➡ Una ciudad ha tenido en los últimos años un promedio de 15 accidentes graves al año. Desde que se ha aprobado el carnet por puntos se han producido 4 accidentes en 6 meses. Admitiendo que el nº de accidentes es constante en todas las épocas del año, ¿se puede afirmar con un 95% de confianza que se ha reducido el nº de accidentes?

$H_0: \lambda \geq 15$ accidentes/año
$H_1: \lambda < 15$ accidentes/año $\Longrightarrow$ $\begin{cases} H_0: \lambda \geq \frac{15}{2} \text{ accidentes/6 meses} \\ H_1: \lambda < \frac{15}{2} \text{ accidentes/6 meses} \end{cases}$

$\hat{\lambda} \longrightarrow N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$; $\Longrightarrow$ Medida de discrepancia $\frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \longrightarrow N(0,1)$

Tenemos un contraste unilateral por la izquierda:



Rechazo $H_0$ si

$\frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} < -1,64$, $\begin{cases} \lambda = 7,5 \\ \hat{\lambda} = 4 \\ n = 1 \text{ (solo se observa 1 periodo de 6 meses)} \end{cases}$ $\Longrightarrow$ $\frac{4-7,5}{\sqrt{\frac{7,5}{1}}} = -1,278$

Por tanto, no hay evidencia estadística para decir que se ha reducido el nº de accidentes. [ $p$-valor $= p(z < -1,278) > 0,05$ ]

**7.3** Una medicina es efectiva en el 75% de los casos. Se prueba un nuevo medicamento en 100 pacientes, observándose su efectividad en 85 de ellos ¿Es la nueva medicina más efectiva que la estandar? (Contrastar con $\alpha = 0,05$)

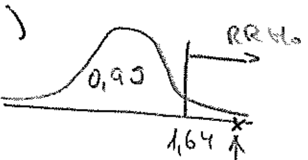
$\begin{cases} H_0: p = 0,75 \\ H_1\ p > 0,75 \end{cases}$

$\downarrow$

$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow N(0,1)$

Si $H_0$ es cierta

$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} < 1,64$



$\hat{p} = 0,85$
$p = 0,75$
$n = 100$ $\longrightarrow$ $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0,85-0,75}{\sqrt{\frac{0,75 \cdot 0,25}{100}}} = 2,3094$

Hay evidencia estadística para rechazar $H_0$
$p$-valor $= p(z \geq 2,3094) = 0,0104$

Otra forma: Rechazaré $H_0$ si $\hat{p} > p + 1,64\sqrt{\frac{p(1-p)}{n}} \overset{p=0,75,\ n=100}{\downarrow} = 0,8210$

$\hat{p} = 0,85 > 0,8210 \longrightarrow$ Rechazo $H_0$

Otra forma: IMPORTANTE

Y v.a. nº pacientes curados de un total de 100

Medicamento antiguo $Y \to B(n=100, p=0{,}75) \rightsquigarrow N(75, \sqrt{100 \cdot 0{,}75 \cdot 0{,}25})$

$p(\text{med antiguo sea más efectivo que el nuevo}) = p(Y \geq 85 \mid Y \rightsquigarrow N(75, 4{,}33)) =$

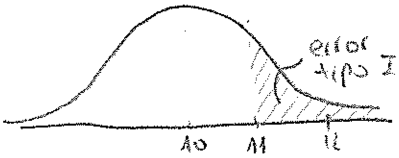
$= P\left(\frac{Y-75}{4{,}33} \geq \frac{85-75}{4{,}33}\right) = p(Z \geq 2{,}3094) = 0{,}0104$

$\Longrightarrow$ Con $\alpha = 0{,}05 > 0{,}0104$ aceptaríamos el nuevo medicamento
Con $\alpha = 0{,}01 < 0{,}0104$ rechazaríamos el nuevo medicamento

$\boxed{7.2}$ Queremos contrastar unilateralmente que $\mu$ de una normal es 10. Tomamos una muestra de $n=16$ y se rechaza la hipótesis en el caso de que la media muestral sea mayor que 11. Sabiendo que $\sigma = 2$, de la población, ¿cuál es la prob. de error tipo I?

$p(\text{error tipo I}) = p(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) =$

$= p\left(\bar{X} > 11 \mid \bar{X} \rightsquigarrow N\left(\underset{\underset{10}{\uparrow}}{\mu}, \underset{\underset{2/\sqrt{16} = \frac{1}{2}}{\uparrow}}{\frac{\sigma}{\sqrt{n}}}\right)\right) = p\left(\frac{\bar{X}-10}{1/2} > \frac{11-10}{\frac{1}{2}}\right) = p(Z > 2) = 0{,}0227$



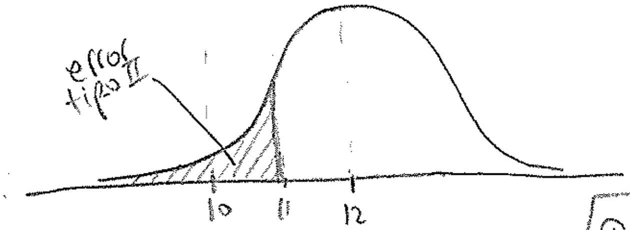
$\boxed{p(\text{error tipo I}) = 0{,}0227}$

¿Cuál sería la prob. de error tipo II del contraste si el valor verdadero de la esperanza fuese 12?

$p(\text{error tipo II}) = p(\text{aceptar } H_0 \mid H_0 \text{ falsa}) =$

$= p\left(\bar{X} \leq 11 \mid \bar{X} \rightsquigarrow N\left(12, \frac{1}{2}\right)\right) =$

$= p\left(\frac{\bar{X}-12}{1/2} \leq \frac{11-12}{1/2}\right) = p(Z \leq -2) = 0{,}0227$



$\boxed{p(\text{error tipo II}) = 0{,}0227}$

# 4 CONTRASTES NORMALES

## ① Contraste para $\mu$ con $\sigma$ conocido

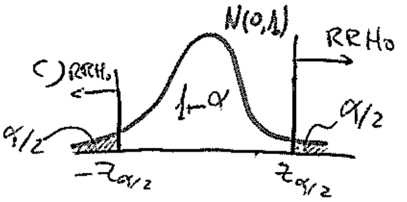
$X_1, X_2, \dots X_n \to N(\mu, \sigma)$

a) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$

b) Medida de discrepancia

$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ ; $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$

c)



Rechazaré Ho si: $\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > z_{\alpha/2}$

d) Toma de datos. $z_0 = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ y comprobamos en qué región estamos.

e) Cálculo del p-valor = $p(z > |z_0|)$

## ② Contraste para $\mu$ con $\sigma$ desconocido

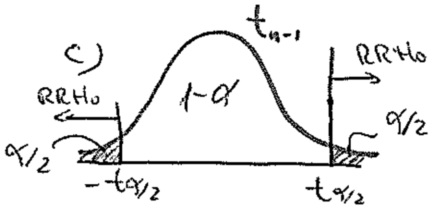
$X_1, X_2, \dots X_n \to N(\mu, \sigma)$

a) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$

b) Medida de discrepancia

$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ ; $\frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \to t_{n-1}$

c)



Rechazaré Ho si:

$$\left|\frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}}\right| > t_{\alpha/2}$$

d) Toma de datos $t_0 = \frac{\bar{X}-\mu_0}{\frac{\hat{S}_0}{\sqrt{n}}}$ y comprobamos en qué región estamos

## ③ Contraste para $\sigma^2$

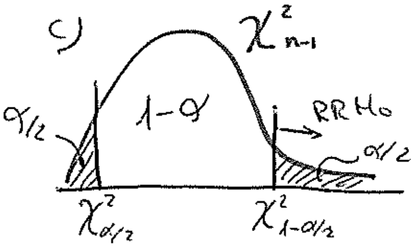
$X_1, X_2 \dots X_n \to N(\mu, \sigma)$

a) $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$

b) Medida de discrepancia

$\frac{(n-1)\hat{S}^2}{\sigma^2} \to \chi^2_{n-1}$

c)



Rechazaré Ho si:

$$\left|\frac{(n-1)\hat{S}^2}{\sigma^2}\right| > \chi^2_{1-\alpha/2}$$

d) Toma de datos

$\chi_0^2 = \frac{(n-1)\hat{S}^2}{\sigma_0^2}$ y comprobamos en qué región cae

➡ La resistencia de compresión de 15 probetas de acero elegidas al azar es:

| | | | | |
|---|---|---|---|---|
| 40,15 | 65,10 | 49,50 | 22,40 | 38,20 |
| 60,40 | 43,40 | 26,35 | 31,20 | 55,60 |
| 47,25 | 73,70 | 35,90 | 45,25 | 52,40 |

¿Es la resistencia media 60? $[\alpha = 0,01]$ desconocido!

a) $\begin{cases} H_0: \mu = 60 \\ H_1: \mu \neq 60 \end{cases}$

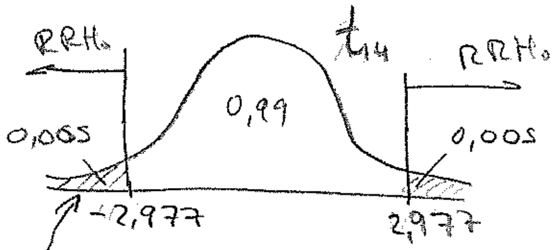
b) $\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ ; $\frac{\bar{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1}$

c) $\bar{X} = 45,75$

$\hat{S} = 14,204$

$n = 15$

$$\frac{45,75-\mu}{\frac{14,204}{\sqrt{15}}} \longrightarrow t_{14}$$



d) Toma de datos

$$\chi_0^2 = \frac{45,75-60}{\frac{14,204}{\sqrt{15}}} = -3,88$$

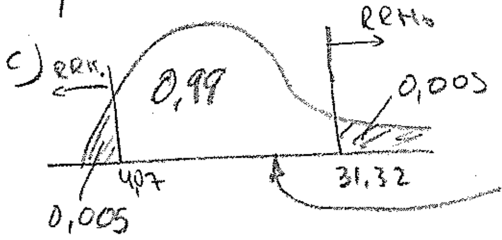
$\Longrightarrow \boxed{\text{Con } \alpha = 0,01 \text{ se rechaza } H_0}$

¿E $\sigma^2 = 200$?

a) $\begin{cases} H_0: \sigma^2 = 200 \\ H_1: \sigma^2 \neq 200 \end{cases}$

b) $\frac{(n-1)\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{n-1}$ ; $\frac{14\,\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{14}$

c)



$\hat{S}^2 = 14,204^2 = 201,75$

d) Toma de datos

$$\frac{14 \cdot 201,75}{200} = 14,12$$

$\Longrightarrow \boxed{\text{No se rechaza } H_0 \text{ con } \alpha = 0,01}$

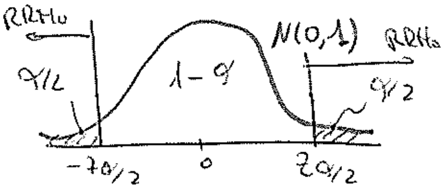
## 5 CONTRASTE PARA λ DE POISSON

$X_1, X_2, \ldots X_n \longrightarrow \text{Poisson}(\lambda)$

① $\begin{cases} H_0: \lambda = \lambda_0 \\ H_1: \lambda \neq \lambda_0 \end{cases}$

② $\hat{\lambda} = \bar{X} \sim N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right) \Longrightarrow \frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda}{n}}} \sim N(0,1)$ : Medida de discrepancia

③ Niv. de significación $\alpha$



Rechazaré $H_0$ si:

$$\left|\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda}{n}}}\right| > z_{\alpha/2}$$

④ Toma de datos

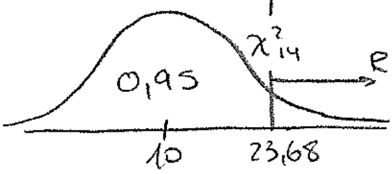
$z_0 = \frac{\hat{\lambda}-\lambda_0}{\sqrt{\frac{\lambda_0}{n}}}$ y vemos en que región está

➔ Calular la probabilidad de error tipo II cuando se quiere contrastar unilateralmente que la varianza de una pob. normal es 10 si el valor verdadero de la misma es 13 y el tamaño muestral 15. Realizar este contraste con:

$\sum_{i=1}^{15}(x_i-\bar{x})^2=150$ y $\alpha=0,05$

Contraste: a) $\left| \begin{array}{l} H_0: \sigma^2=10 \\ H_1: \sigma^2>10 \end{array} \right.$

b) $\frac{(n-1)\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{n-1}$ , $\frac{14\hat{S}^2}{\sigma^2} \longrightarrow \chi^2_{14}$

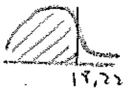


$p(\text{aceptar } H_0) = p\left(\frac{14\hat{S}^2}{\sigma^2} \leq 23,68\right) = \boxed{\sigma^2=10}$

$= p\left(\hat{S}^2 \leq \frac{23,68\cdot 10}{14}\right) = p(\hat{S}^2 \leq 16,91)$

$p(\text{error tipo II}) = p(\text{aceptar } H_0 \mid H_0 \text{ es falsa}) = p(\hat{S}^2 \leq 16,91 \mid \sigma^2=13) =$

$= p\left(\frac{(n-1)\hat{S}^2}{\sigma^2} \leq \frac{16,91(n-1)}{\sigma^2}\right) \underset{\sigma^2=13}{\uparrow}{=} p\left(\chi^2_{14} \leq \frac{16,91\cdot 14}{13}\right) = p(\chi^2_{14} \leq 18,22) \simeq$



≈ No hay precisión suficiente en tablas

ⓐ Contraste $\frac{150}{10} = 15 \Longrightarrow$ Acepto $H_0$

## (7.15)

El tiempo de duración T de un componente es una v.a. con distribución exponencial de media $\mu$. Ensayo de 20 componentes resultando:

10.99 15,79 24,14 34,43 43,72 51,72 56,12 60,27 77,20 88,47
91,07 117,58 130,40 133,12 152,90 159,00 193,62 208,71 308,82 316,07

$2T/\mu \longrightarrow \chi^2_2$. Realiza el contraste $\left\{ \begin{array}{l} H_0: \mu=200\text{ h} \\ H_1: \mu<200\text{ h} \end{array} \right.$ $\alpha=0,05$

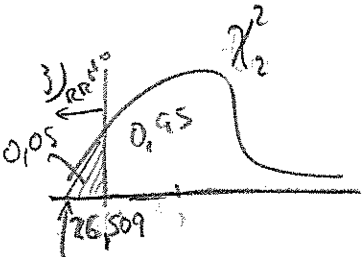
1) $\left| \begin{array}{l} H_0: \mu=200\text{h} \\ H_1: \mu<200\text{h} \end{array} \right.$ (unilateral por la izquierda)

2) $\frac{2T}{\mu} \longrightarrow \chi^2_2$

Medida de discrepancia

$$\frac{2\sum_{i=1}^{20} T_i}{\mu} \longrightarrow \sum_{i=1}^{20} \chi_2^2 = \chi^2_{40}$$

3)



4) Toma de datos

$$\frac{2\sum T_i}{\mu_0} = \frac{2\cdot 2274,14}{200} = 22,7414$$

Hay evidencia estadística suficiente para decir que la media de los componentes electrónicos es menor que 200h

**7.14** Sistema de contamiento de misiles con errores en el 7% contamientos. Se prueba un nuevo sistema con 20 contamientos. El nuevo sistema será mejor si no se producen fallos. Llamamos p a la probabilidad de fallo del nuevo sistema. Obtener la probabilidad de error tipo II

p(error tipo II) = p(aceptar Ho | Ho es falsa)

a) $\begin{cases} H_0: p = 0{,}07 \\ H_1: p < 0{,}07 \end{cases}$ → Es mejor el sist nuevo

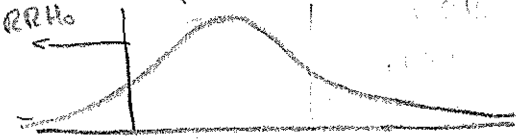
X v.a nº fallos al lanzar 20 → $B(n=20, p)$

$p(\text{aceptar } H_0) = p(X \geq 1 \mid X \to B(20,p)) = 1 - p(X=0 \mid X \to B(n,p)) =$

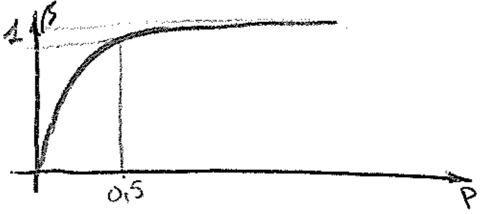
$= 1 - \binom{20}{0} p^0 (1-p)^{20-0} = \boxed{1-(1-p)^{20}}$ Curva característica $= \beta(p)$

$p(\text{error tipo I}) = p(\text{rech } H_0 \mid H_0 \text{ cierta}) = p(X=0 \mid p = 0{,}07) =$

$= (1-p)^{20}$ con $p = 0{,}07$ → $\boxed{p(\text{err. I}) = 0{,}234}$



$\beta(p) = 1-(1-p)^{20}$



| P | β |
|---|---|
| 0,05 | 0,64 |
| 0,01 | 0,18 |
| 0,5 | 0,999 |
| 0,075 | 0,79 |

**NOTA!!** Cuando no nos dan un verdadero valor de p, obtenemos la curva característica
$\beta = p(\text{acep. } H_0)$

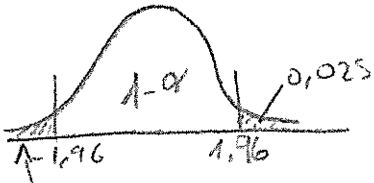
➡ Durante el año 1985 nacieron 25670 bebés: 13333 niños y 12337 niñas. Contrastar la hipótesis de que la probabilidad de nacer niño es igual a la de nacer niña.

$\hat{P}_{♀} = \frac{12337}{25670} = 0{,}48$

1) $\begin{cases} H_0: P_{♀} = 0{,}5 \\ H_1: P_{♀} \neq 0{,}5 \end{cases}$

2) Medida de discrepancia $P_{♀} \sim N\left(P_{♀}, \sqrt{\frac{P_{♀}(1-P_{♀})}{n}}\right)$

$$\frac{\hat{P}_{♀} - P_{♀}}{\sqrt{\frac{P_{♀}(1-P_{♀})}{n}}} \longrightarrow N(0,1)$$



3) Confianza α = 0,05, contraste bilat.

4) Toma de datos de la muestra: $\frac{0{,}48 - 0{,}5}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{25670}}} = -6{,}21$

Hay evidencia estadística para decir que nacen más niños

# 6 COMPARACIÓN DE DOS TRATAMIENTOS

## ● Pasos del método

### ① Definición del modelo de distribución de probabilidad

Hipótesis

Parámetros $\longrightarrow \left\{\begin{matrix}\mu_1\\ \mu_2\end{matrix}\right. \sigma^2$

Las observaciones se descomponen en

$\mu_i \to$ parte predecible

$u_{ij} \to$ parte aleatoria

$$\boxed{\begin{matrix} y_{ij}=\mu_i+u_{ij}\to N(\mu_i,\sigma^2)\\ u_{ij}\to N(0,\sigma^2)\end{matrix}}$$

- Hipótesis de normalidad: Suponemos que las muestras son normales $y_{ij}\longrightarrow N(\mu_i,\sigma^2)$
- Hipótesis de homocedasticidad: $\sigma_1^2=\sigma_2^2=\sigma^2$
- Hipótesis de independencia: Las dos muestras son independientes.

### ② Estimación de los parámetros

- Estimación de medias

$$\left\{\begin{matrix}\mu_1\longrightarrow \bar{y}_{1\circ}=\dfrac{\sum_{j=1}^{n_1} y_{1j}}{n_1}\\ \mu_2\longrightarrow \bar{y}_{2\circ}=\dfrac{\sum_{j=1}^{n_2} y_{2j}}{n_2}\end{matrix}\right.$$

Tendremos que ver si la diferencia entre estas medias es grande o pequeña comparando con la variabilidad.

- Estimación de varianza -residuos-

$y_{ij}=\mu_i+u_{ij}$ ; $u_{ij}\to N(0,\sigma^2)$

Como $u_{ij}=y_{ij}-\mu_i \Rightarrow \boxed{e_{ij}=y_{ij}-\bar{y}_{i\circ} \text{ RESIDUO}}$

Estimaremos la varianza como:

$$\boxed{\hat{S}_R^2=\frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i} e_{ij}^2}{n-2}=\frac{(n_1-1)\hat{S}_1^2+(n_2-1)\hat{S}_2^2}{n_1+n_2-2}}$$

### ③ Diagnosis de las hipótesis

### ④ Aplicación

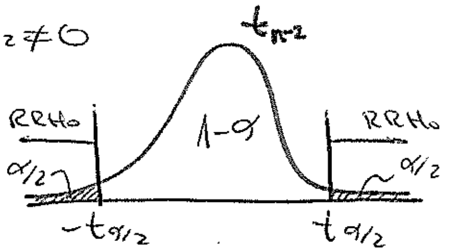
# 7 CONTRASTE DE IGUALDAD DE MEDIAS

Definimos $\bar{y}_{1\circ}-\bar{y}_{2\circ}\longrightarrow N\left(\mu_1-\mu_2,\ \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right)$

$$\frac{(\bar{y}_{1\circ}-\bar{y}_{2\circ})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\longrightarrow N(0,1)\Rightarrow \boxed{\frac{(\bar{y}_{1\circ}-\bar{y}_{2\circ})-(\mu_1-\mu_2)}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\longrightarrow t_{n-2}}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \Longrightarrow \begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$$

$$t_0 = \frac{\bar{y}_{1\circ} - \bar{y}_{2\circ}}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \longrightarrow t_{n-2}$$



Dependiendo de dónde caiga $t_0$, aceptaremos o rechazaremos $H_0$

Se seleccionan 20 individuos y se les asignan dos tipos de dietas. A o B. La tabla muestra la reducción de colesterol a los 2 meses:

| A | B |
|---|---|
| 51,3 | 29,6 |
| 39,4 | 47,0 |
| 26,3 | 25,9 |
| 39,0 | 13,0 |
| 48,1 | 33,1 |
| 34,2 | 22,1 |
| 69,8 | 34,1 |
| 31,3 | 19,5 |
| 45,2 | 43,8 |
| 46,4 | 24,9 |

$$\begin{cases} \bar{\mu}_{1\circ} = 43,1 \\ \bar{\mu}_{2\circ} = 29,3 \end{cases}, \quad \hat{S}_R^2 = \frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2} = 130,96 \Rightarrow \hat{S}_1^2 = 11,44$$

$\hat{S}_1^2 = 12,25^2$

$\hat{S}_2^2 = 10,57^2$

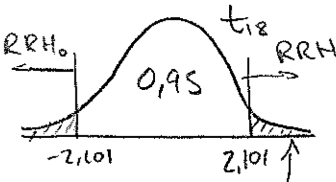
$n_1 = n_2 = 10$

$$\begin{cases} H_0: \mu_1 = \mu_2 \longrightarrow \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 \neq \mu_2 \longrightarrow \mu_1 - \mu_2 \neq 0 \end{cases}$$

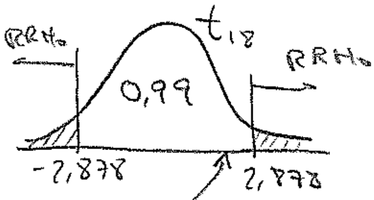
Si $H_0$ es cierta

$$\frac{(\bar{\mu}_{1\circ} - \bar{\mu}_{2\circ}) - \overset{0}{\cancel{(\mu_1 - \mu_2)}}}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \longrightarrow t_{n-2}$$

Con $\alpha = 0,05$



Con $\alpha = 0,01$



$$t_0 = \frac{\bar{\mu}_{1\circ} - \bar{\mu}_{2\circ}}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{43,1-29,3}{11,44\sqrt{\frac{1}{10}+\frac{1}{10}}} = 2,69$$

Con $\alpha = 0,05$: La diferencia entre las medias es significativa y hay suficiente evidencia estadística para decir que los dos tratamientos son diferentes

Con $\alpha = 0,01$: La diferencia de medias no es significativa y no hay evidencia estadística suficiente para decir que los tratamientos son diferentes

**NOTA!!** El no rechazar $H_0$ no significa que $H_0$ sea cierta. No rechazar $H_0$ implica que la diferencia de medias no es lo suficientemente grande como para ser detectada con el tamaño muestral dado.

Calulo del nivel crítico o p-valor ¡¿!

$$p\text{-valor} = p(|t_{18}| > 2{,}69) = 2p(\underbrace{t_{18} > 2{,}69}_{p \in [0{,}005,\, 0{,}01] \to p = 0{,}00735}) = 2 \cdot 0{,}00735 = \underline{0{,}0147}$$

**NOTA TEÓRICA!!** Intervalo de confianza para la diferencia de medias

$$\mu_1 - \mu_2 \in (\bar{y}_{1\circ} - \bar{y}_{2\circ}) \pm t_{\alpha/2}\, \hat{S}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Si $0 \in$ al intervalo → Son iguales No RRHo

Si $0 \notin$ al intervalo → Son diferentes RRHo

En nuestro caso, con $\alpha = 0{,}05$

$$(\mu_1 - \mu_2) \in (43.1 - 29.3) \pm 2{,}102 \cdot 11{,}44 \cdot \sqrt{\frac{1}{10} + \frac{1}{10}} \to [3{,}06,\ 24{,}54]$$

como $0 \notin$ al intervalo → Rechazo Ho

## 8 CONTRASTE DE IGUALDAD DE VARIANZAS

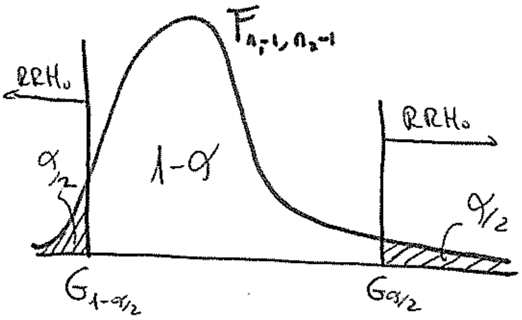
$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \longrightarrow \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \longrightarrow \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{cases}$$

Definimos la distribución F de Fisher ⟶ Es el cociente entre dos $\chi^2$ independientes divididas por sus grados de libertad.

$$\hat{S}_1^2 = \frac{\Sigma(y_{1j} - \bar{y}_{1\circ})^2}{n_1 - 1}; \quad \frac{(n_1 - 1)\hat{S}_1^2}{\sigma_1^2} \to \chi^2_{n_1-1}$$

$$\hat{S}_2^2 = \frac{\Sigma(y_{2j} - \bar{y}_{2\circ})^2}{n_2 - 1}; \quad \frac{(n_2 - 1)\hat{S}_2^2}{\sigma_2^2} \to \chi^2_{n_2-1}$$

$$F = \frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1} = \frac{\dfrac{\hat{S}_1^2}{\sigma_1^2}}{\dfrac{\hat{S}_2^2}{\sigma_2^2}} \longrightarrow F_{n_1-1,\, n_2-1}$$



Si Ho es cierto: $F_0 = \dfrac{\hat{S}_1^2}{\hat{S}_2^2} \longrightarrow F_{n_1-1,\, n_2-1}$

**NOTA TEÓRICA!!** Intervalo de confianza para el cociente de varianzas

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left[ \frac{\hat{S}_1^2}{\hat{S}_2^2 \cdot G_{\alpha/2}},\ \frac{\hat{S}_1^2}{\hat{S}_2^2\, G_{1-\alpha/2}} \right]$$

Si $1 \in$ al intervalo → Son iguales No RRHo

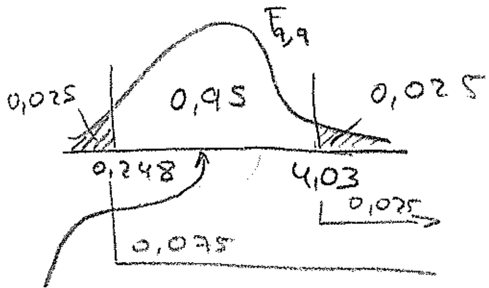
Si $1 \notin$ al intervalo → No son iguales RRHo

→ Ejemplo anterior: Contrastar las varianzas con $\alpha=0{,}05$

$$\begin{cases} H_0: \dfrac{\sigma_1^2}{\sigma_2^2}=1 \\ H_1: \dfrac{\sigma_1^2}{\sigma_2^2}\neq 1 \end{cases}$$

$$\frac{\hat{S}_1^2}{\hat{S}_2^2} \longrightarrow F_{n_1-1,n_2-1}\text{, si } H_0 \text{ es cierta}$$

$$\hat{S}_1^2=12{,}25^2;\ \hat{S}_2^2=10{,}57^2$$



$$F_0=\frac{12{,}25^2}{10{,}57^2}=1{,}37$$

$$\begin{cases} F_{9,9,0{,}025}=4{,}03 \\ F_{9,9,0{,}975}=\dfrac{1}{F_{9,9,0{,}025}}= \\ =\dfrac{1}{4{,}03}=0{,}248 \end{cases}$$

NOTA!!

$$\boxed{F_{m,n,\alpha}=\frac{1}{F_{m,n,(1-\alpha)}}}$$

## 9 CONTRASTE $\chi^2$ DE BONDAD DE AJUSTE

$$\begin{cases} H_0: X_i,\ \forall i=1,2,\dots n \longrightarrow f_x \\ H_1: X_i,\ \forall i=1,2,\dots n \not\longrightarrow f_x \end{cases}$$

| Clases | Fr observada | Fr Esperada |
|---|---|---|
| $C_0\leq X_i\leq C_1$ | $O_1$ | $E_1$ |
| $C_1\leq X_i\leq C_2$ | $O_2$ | $E_2$ |
| ⋮ | ⋮ | ⋮ |
| $C_{k-1}\leq X_i\leq C_k$ | $O_k$ | $E_k$ |
| ⋮ | ⋮ | ⋮ |
| $C_{K-1}\leq X_i\leq C_K$ | $O_K$ | $E_K$ |

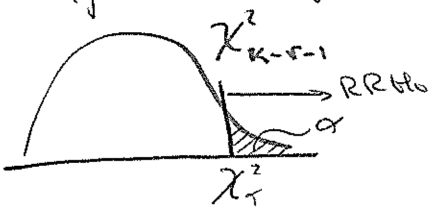
$E_k \longrightarrow$

$$\boxed{\begin{array}{l} E_k=np_k \\ p_k=p(C_{k-1}\leq X\leq C_k) \\ \sum_{k=1}^{K} O_k=\sum_{k=1}^{K} E_k=n \end{array}}$$

$$\boxed{\sum_{k=1}^{K}\frac{(O_k-E_k)^2}{E_k}\longrightarrow \chi^2_{K-r-1}\ \begin{cases} K=\text{nº clases} \\ r=\text{nº param. estimados} \end{cases}}$$

Ej.: $\begin{cases} H_0: X \text{ es normal} \\ H_1: X \text{ no es normal} \end{cases} \longrightarrow \dfrac{\sum_{i=1}^{k}(O_i-E_i)^2}{E_i}\longrightarrow \chi^2_{k-r-1}$

Fijamos la confianza ⟹ SIEMPRE UNILATERAL POR LA DERECHA



Toma datos de la muestra $\chi_0^2=\dfrac{\Sigma(O_i-E_i)^2}{E_i}$ y miro donde cae.

# 10 CONTRASTE K-S DE BONDAD DE AJUSTE

- Sólo es válido para variables continuas.
- Compara la función de distribución teórica con la empírica $F(x) - F_n(x)$
- No requiere agrupar los datos

Pasos:

① Ordenar la muestra

② Calcular la distribución empírica de la muestra $F_n(x)$

③ Calcular la discrepancia máxima

$$\max\{D_n(x)\} = \max\{F_n(x) - F(x)\}$$
$$\max\{F_n(x-1) - F(x)\}$$

④ Comparar con el valor de tablas y si $D_n$ es mayor RRHo.

➡ Se han tomado 12 valores de una variable física X supuesta normal.

30,2 | 30,8 | 29,3 | 29 | 30,9 | 30,8 | 29,7 | 28,9 | 30,5 | 31,2 | 31,3 | 28,5 |

Contrastar mediante el contraste de K-S si X es normal

$$\begin{cases} H_0: X \to \text{Normal} \\ H_1: X \not\to \text{Normal} \end{cases} \qquad \bar{X} = 30{,}09 \qquad \hat{S} = 0{,}97$$

| $X$ | $F_n(x)$ | $F(x)$ | $F_n(X-1) - F(x)$ | $F_n(X) - F(x)$ | $D_n(X)$ |
|---|---|---|---|---|---|
| 28,5 | 1/12 | 0,05 | 0,05 | 0,03 | 0,05 |
| 28,9 | 2/12 | 0,11 | 0,03 | 0,06 | 0,06 |
| 29 | 3/12 | 0,131 | 0,04 | 0,12 | 0,12 |
| 29,3 | 4/12 | 0,21 | 0,04 | 0,12 | 0,12 |
| 29,7 | 5/12 | 0,34 | 0,01 | 0,08 | 0,08 |
| 30,2 | 6/12 | 0,54 | 0,12 | 0,04 | 0,12 |
| 30,5 | 7/12 | 0,66 | 0,16 | 0,08 | 0,16 |
| 30,8 | 9/12 | 0,77 | 0,19 | 0,02 | 0,19 ⟸ ⊛ |
| 30,9 | 10/12 | 0,80 | 0,05 | 0,03 | 0,05 |
| 31,2 | 11/12 | 0,87 | 0,04 | 0,05 | 0,05 |
| 31,3 | 1 | 0,89 | 0,03 | 0,11 | 0,11 |

$X \to N(30{,}09,\ 0{,}97)$

$$p(X \leq 28{,}5) = p\left(z < \frac{28{,}5 - 30{,}09}{0{,}97}\right) = p(z < -1{,}63) = 1 - p(z < 1{,}63) = 0{,}0516$$

$p(X \leq 28{,}9) = p(z \leq -1{,}22) = 0{,}1112$

$p(X \leq 29) = p(z \leq -1{,}12) = 0{,}1314$

$p(X \leq 29{,}3) = p(z \leq -0{,}81) = 0{,}209$

$p(X \leq 29{,}7) = p(z \leq -0{,}4) = 0{,}3436$

$p(X \le 30{,}2) = p(Z \le 0{,}11) = 0{,}5438$

$p(X \le 30{,}5) = p(Z \le 0{,}42) = 0{,}6628$

$p(X \le 30{,}8) = p(Z \le 0{,}73) = 0{,}7673$

$p(X \le 30{,}9) = p(Z \le 0{,}83) = 0{,}7967$

$p(X \le 31{,}2) = p(Z \le 1{,}14) = 0{,}8729$

$p(X \le 31{,}3) = p(Z \le 1{,}24) = 0{,}8925$

$\text{Max } D_n = 0{,}19 < D_{12,\,0{,}05} = 0{,}377$ → No se puede rechazar la hipótesis de normalidad con $\alpha = 0{,}05$

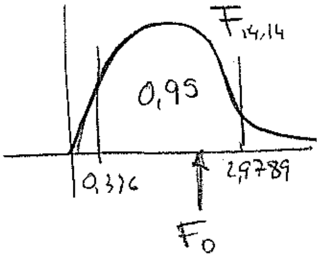
**7.6** $\bar{y}_M = 896{,}6$ ; $\bar{y}_N = 763{,}2$ $n_1 = n_2 = 15$

$\hat{S}_M = 119{,}47$ ; $\hat{S}_N = 111{,}91$

a) Contrastar hip. de mismo variabilidad $\alpha = 0{,}05$

$$\begin{cases} H_0: \sigma_M^2 = \sigma_N^2 \\ H_1: \sigma_M^2 \ne \sigma_N^2 \end{cases} \qquad n_1 = n_2 = 15$$

Si Ho cierta $\dfrac{\hat{S}_M^2}{\hat{S}_N^2} \rightsquigarrow F_{14,14,0{,}05}$ ; $F_0 = \dfrac{119{,}47^2}{111{,}91^2} = 1{,}1397$



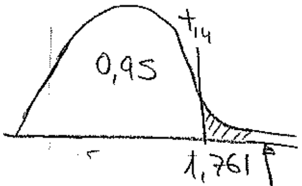
$F_{14,14,0{,}025} = 2{,}9789$

$F_{14,14,0{,}975} = 0{,}336$

No podemos rechazar la hipótesis de homocedasticidad

b) $\mu = 792{,}5$ Contrastar que cada experimento no tiene error sistemático

Ⓜ $$\begin{cases} H_0: \mu_M = 792{,}5 \\ H_1: \mu_M > 792{,}5 \end{cases} \qquad \frac{\hat{\mu}_M - \mu_M}{\frac{\hat{S}_M}{\sqrt{n_1}}} \rightsquigarrow t_{n-1}$$



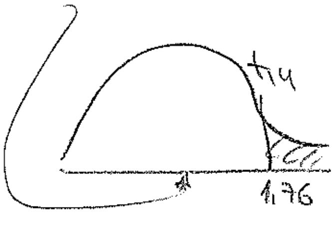
$$t_0 = \frac{896{,}6 - 792{,}5}{\frac{119{,}47}{\sqrt{15}}} = 3{,}37$$

⇒ Tenemos error sistemático

Ⓝ $$\begin{cases} H_0: \mu_N = 792{,}5 \\ H_1: \mu_N > 792{,}5 \end{cases} \qquad t_0 = \frac{763{,}2 - 792{,}5}{\frac{111{,}91}{\sqrt{15}}} = -1{,}01$$

$$\frac{\hat{\mu}_N - \mu_N}{\frac{\hat{S}_N}{\sqrt{n_2}}} \rightsquigarrow t_{n-1}$$

No hay error sistemático

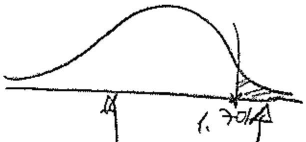


También podríamos haber considerado $\sigma_M^2 = \sigma_N^2 = \sigma^2$; $\hat{\sigma}^2 = \hat{S}_R^2$

$$\hat{S}_R^2 = \frac{(15-1)\hat{S}_M^2 + (15-1)\hat{S}_N^2}{15+15-2} = \frac{14\cdot 119,47^2 + 14\cdot 111,91^2}{28} = 13400$$

$\hat{S}_R = 115,75$

$\rightarrow$ Ⓜ $\rightarrow \dfrac{\hat{\mu}_M - \mu_M}{\frac{115,75}{\sqrt{15}}} \sim t_{28}$



$$t_{0M} = \frac{896,6 - 792,5}{\frac{115,75}{\sqrt{15}}} = 3,49$$

Ⓜ → Error
Ⓝ → No error

Ⓝ $\rightarrow t_{0N} = \dfrac{763,2 - 792,5}{\frac{115,75}{\sqrt{15}}} = -0,98$

(7.5) $X_1, X_2, \ldots X_n \longrightarrow f(x) = \frac{1}{\lambda} e^{-x/\lambda}$; $x \geq 0$; $\lambda > 0$

$U = 2n\bar{X}/\lambda \longrightarrow \chi^2_{2n}$ con $\bar{x} = (x_1 + x_2 + \ldots + x_n)/n$

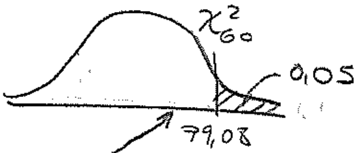
a) Tiempo de funcionamiento es una v.a. con distribución exponencial.

$\bar{X} = 6,2\cdot 10^3 h$. $\alpha = 0,05$ $\quad \left\{ \begin{array}{l} H_0: \lambda = 5\cdot 10^3 \\ H_1: \lambda > 5\cdot 10^3 \end{array} \right.$

$n = 30$

Indicar: p-valor, prob. error tipo II si $\lambda = 7\cdot 10^3$

$U = \dfrac{2n\bar{X}}{\lambda} \sim \chi^2_{2n}$



$U_0 = \dfrac{2\cdot 30\cdot 6,2\cdot 10^3}{5\cdot 10^3} = 74,4$ ——— Con conf. $\alpha = 0,05$ no podemos rechazar la hipótesis nula

$$\boxed{\text{p-valor} = p(\chi^2_{60} > 74,4)}$$

$p(\text{error II}) = p(\text{aceptar } H_0 \mid H_0 \text{ falsa}) = p\left(\text{acep. } H_0 \;\middle|\; \begin{array}{l} \lambda \neq 5\cdot 10^3 \\ \lambda = 7\cdot 10^3 \end{array}\right)$

$$p(\text{aceptar}) = p\left(\frac{2n\bar{X}}{\lambda} \leq 79,08\right) = p\left(\bar{X} \leq \frac{79,08\cdot 5\cdot 10^3}{2\cdot 30}\right) =$$

$= p(\bar{X} \leq 6,59\cdot 10^3)$

$$\Longrightarrow p\left(\bar{X} \leq 6,59\cdot 10^3 \;\middle|\; \begin{array}{l} \lambda \neq 5\cdot 10^3 \\ \lambda = 7\cdot 10^3 \end{array}\right) = p\left(\frac{2n\bar{X}}{\lambda} \leq \frac{6,59\cdot 10^3\cdot 2n}{\lambda}\right) =$$

$$= p\left(\chi^2_{60} \leq \frac{6,59\cdot 10^3\cdot 2\cdot 30}{7\cdot 10^3}\right) = \boxed{p(\chi^2_{60} \leq 52,72)}$$

b) 2° ensayo con 15 equipos. ¿Cual es el valor máximo de la media muestral de estos 15 equipos que permitiría concluir que son peores que de la 1ª empresa?

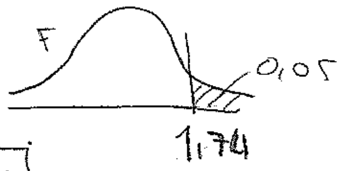
6000h ensayo fallan 6 equipos siendo el promedio 2350h ¿Es necesario seguir con el ensayo?

$$\overline{X}_1 = 6{,}2\cdot10^3 \rightarrow n_1 = 30 \longrightarrow U = \frac{2n_1\overline{X}_1}{\lambda_1} \sim \chi^2_{2n_1}$$

$$\overline{X}_2 \ ; \ n_2 = 15 \longrightarrow U = \frac{2n_2\overline{X}_2}{\lambda_2} \sim \chi^2_{2n_2}$$

$$F = \frac{\chi^2_n/n}{\chi^2_m/m} \Rightarrow \frac{\frac{2n_1\overline{X}_1}{\lambda_1}\Big/\cancel{2n_1}}{\frac{\cancel{2n_2}\overline{X}_2}{\lambda_2}\Big/\cancel{2n_2}} \sim F_{2n_1,2n_2} \ ; \ \frac{\overline{X}_1/\lambda_1}{\overline{X}_2/\lambda_2} \sim F_{2n_1,2n_2}$$

Si $H_0$ es cierta $\lambda_1 = \lambda_2 \longrightarrow \frac{\overline{X}_1}{\overline{X}_2} \longrightarrow F_{60,30}$



$$\alpha\begin{cases} H_0 : \lambda_1 = \lambda_2 \\ H_1 \ \lambda_1 > \lambda_2 \end{cases}$$

$$\frac{\overline{X}_1}{\overline{X}_2} = 1{,}74 \ ; \ \boxed{\overline{X}_2 = \frac{6{,}2\cdot10^3}{1{,}74} = 3{,}56\cdot10^3}$$

$$15\begin{cases} 6 \rightarrow \overline{X}_6 = 2350 \\ 9 \rightarrow \overline{X}_9 > 6000 \end{cases} \qquad \overline{X}_2 \geq \frac{6\cdot2350 + 9\cdot6000}{15} = 4540h > 3{,}56\cdot10^3$$

$\Rightarrow$ Podemos detener el ensayo pues los equipos de la 2ª empresa son iguales que los de la primera

(7.7) Estudiamos 2 neumáticos

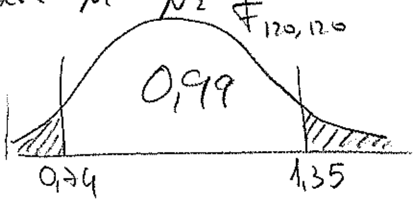
A — $n_A = 121$ — $\overline{X}_A = 27465$ — $\hat{S}_A = 2500$

B — $n_B = 121$ — $\overline{X}_B = 27572$ — $\hat{S}_B = 3000$

Calcular con $\alpha = 0{,}01$

- Intervalo de confianza para $\sigma_1^2/\sigma_2^2$
- Intervalo de confianza para $\mu_1 - \mu_2$

$$\frac{\hat{S}_A^2/\sigma_1^2}{\hat{S}_B^2/\sigma_2^2} \longrightarrow F_{n_A-1,\,n_B-1}$$

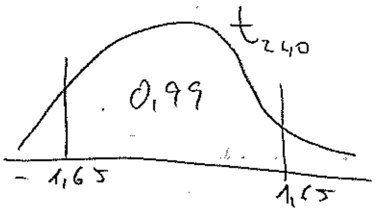


$$0{,}74 \leq \frac{2500^2\cdot\sigma_2^2}{3000^2\ \sigma_1^2} \leq 1{,}35 \ , \quad \boxed{0{,}514 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 0{,}9375}$$

$$\frac{\sigma_1^2}{\sigma_2^2} \leq \frac{2500^2}{0{,}74\cdot3000^2}$$

$$\frac{2500^2}{3000^2\ 1{,}35} \leq \frac{\sigma_1^2}{\sigma_2^2}$$

$$\rightarrow \frac{(\bar{X}_A-\bar{X}_B)-(\mu_1-\mu_2)}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightsquigarrow t_{n-2}$$



$$\hat{S}_R=\sqrt{\frac{120\cdot 2500^2+120\cdot 3000^2}{121+121-2}}=2761,34$$

$$(\mu_1-\mu_2)\in(27465-27572)\pm 1,65\cdot 2761,34\sqrt{\frac{1}{121}+\frac{1}{121}}$$

$$\boxed{-692,76\le\mu_1-\mu_2\le 478,77}$$

**⑦.⑧** Resultados máquina A: 137.5; 140.7; 106.9, 175.1, 177.3, 120.4, 77.9, 104.2

Resultados máquina B: 103.3, 121.7, 98.4, 161.5, 167.8 67.3

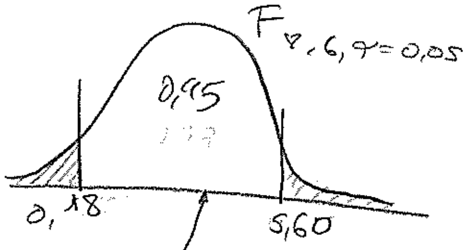
¿Son iguales? suponer MAS → N

$\bar{X}_A=129,97$, $\bar{X}_B=120$ $n_A=8$

$n_B=6$

$\hat{S}_A=64,51$ $\hat{S}_B=38,81$

$$\begin{cases} H_0: \sigma_1^2=\sigma_2^2 \\ H_1: \sigma_1^2\neq\sigma_2^2 \end{cases} \quad \text{Si } H_0 \text{ es cierta: } \frac{\hat{S}_A^2}{\hat{S}_B^2}\rightsquigarrow F_{n_1-1,\,n_2-1}$$

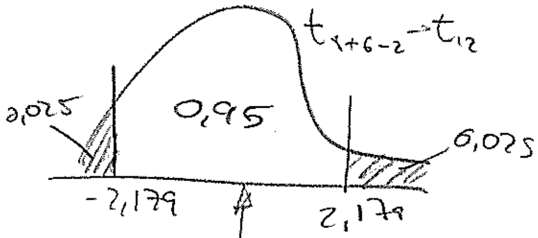


$$F_0=\frac{64,51^2}{38,81^2}=2,76$$

No podemos rechazar hip. de homocedasticidad

$$\begin{cases} H_0: \mu_1-\mu_2=0 \\ H_1\ \mu_1-\mu_2\neq 0 \end{cases}$$

Si $H_0$ cierta $\dfrac{(\bar{X}_A-\bar{X}_B)}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\longrightarrow t_{n-2}$



$$\hat{S}_R=\sqrt{\frac{(8-1)64,51^2+(6-1)38,81^2}{8+6-2}}=55,27$$

$t_0=0,33$

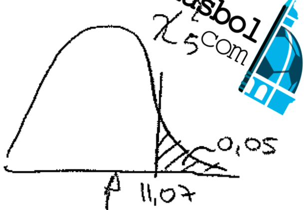
No hay suficiente evidencia estadística para decir que las máquinas sean distintas

ⓐ7.10 Se puede aceptar la hip. de normalidad $(\alpha = 0{,}05)$?

| | $O_i$ | $E_i$ |
|---|---|---|
| 41,5-43,5 | 4 | 4,08 |
| 43,5-45,5 | 7 | 5,58 |
| 45,5-47,5 | 12 | 9,06 |
| 47,5-49,5 | 8 | 11,27 |
| 49,5-51,5 | 6 | 11,27 |
| 51,5-53,5 | 11 | 9,08 |
| 53,5-55,5 | 9 | 5,58 |
| 55,5-57,5 | 3 | 4,08 |
| | 60 | 60 |

$$\underbrace{\sum \frac{(O_i - E_i)^2}{E_i}} \longrightarrow \chi^2_{8-\cancel{2}-1}$$

$$\frac{(4-4{,}08)^2}{4{,}08} + \frac{(7-5{,}58)^2}{5{,}58} + \dots = 7{,}518$$



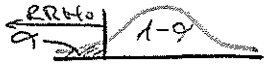
No podemos rechazar la hipótesis de normalidad

1,00
3
1

# TEMA 7 RESUMEN

**Contraste** $\begin{cases} H_0: \text{hip. nula} \\ H_1: \text{hip. alternativa} \end{cases}$ → Contraste unilateral

Por la izquierda



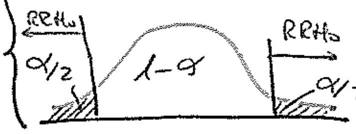
$\begin{cases} H_0: \rho = \rho_0 \\ H_1: \rho < \rho_0 \end{cases}$

Por la derecha

$\begin{cases} H_0: \rho = \rho_0 \\ H_1: \rho > \rho_0 \end{cases}$

Con los datos de la muestra definimos el $\boxed{\text{p-valor}}$

→ Contraste bilateral



$\begin{cases} H_0: \rho = \rho_0 \\ H_1: \rho \neq \rho_0 \end{cases}$

**Errores**

$$\boxed{\begin{array}{l} \text{Tipo I} \rightarrow p(\text{error I}) = p(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = \alpha \\ \text{Tipo II} \rightarrow p(\text{error II}) = p(\text{aceptar } H_0 \mid H_0 \text{ falsa}) = \beta \end{array}}$$

**Diferentes tipos de contrastes:**

|||

Igual que intervalos de confianza.

- $\mu$ de normal con $\sigma$ conocido → Contrasta con normal
- $\mu$ de normal con $\sigma$ desconocido → Contrasta con $t_{n-1}$
- $\sigma^2$ de normal → Contrasta con $\chi^2_{n-1}$
- $\lambda$ de Poisson → Contrasta con normal.

**Comparación de dos tratamientos**

$$\boxed{y_{ij} = \mu_i + u_{ij} \longrightarrow N(\mu_i, \sigma^2)}, \quad \boxed{u_{ij} \rightarrow N(0, \sigma^2)}$$

$$\bar{y}_{1^\circ} = \frac{\sum y_{1j}}{n_1} \; ; \; \bar{y}_{2^\circ} = \frac{\sum y_{2j}}{n_2} \longrightarrow \text{Residuo: } \boxed{e_{ij} = y_{ij} - \bar{y}_{i^\circ}}$$

Varianza residual $\boxed{\hat{S}_R^2 = \frac{\sum\sum e^2}{n-2} = \frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2}}$

- Contraste de igualdad de medias $\boxed{\dfrac{\bar{y}_{1^\circ} - \bar{y}_{2^\circ}}{\hat{S}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow t_{n-2}}$

- Contraste de igualdad de varianzas $\boxed{\dfrac{\hat{S}_1^2}{\hat{S}_2^2} \longrightarrow F_{n_1-1, n_2-1}}$

**Contraste $\chi^2$ de bondad de ajuste**

Sirven para contrastar si una muestra se ajusta a una determinada función de distribución

$O_i$ → Frec. observadas, $E_i$ → Frec. esperadas $\boxed{E_k = nP_k}$ $\sum O_k = \sum E_k = n$

Siempre unilateral por la derecha

$$\boxed{\sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \longrightarrow \chi^2_{K-r-1}}$$

K nº clases
r nº parametros estimados

# Ejercicios Inferencia

T.S

**3.2** X tiene distribución binomial de parámetros n y p, desconocidos. Estimarlos por el metodo de los momentos.

M.A.S: $\{16, 18, 22, 25, 27\}$

Calculamos $\overline{X} = \frac{\Sigma x_i}{n} \underset{\underset{n=5}{\uparrow}}{=} 21{,}6 \leftarrow$ Momento de orden 1 $\longrightarrow 21{,}6 = \hat{n}\hat{p} \quad (1)$

$$E[X] \underset{\underset{X \to B(n,p)}{\uparrow}}{=} n \cdot p$$

Calculamos momento de orden 2 $\rightarrow \frac{\Sigma x_i^2}{n} = 483{,}6 \Longrightarrow$

Como: $Var[X] = E[X^2] - (E(X))^2$; $E[X^2] = Var(X) + (E[X])^2 \underset{\underset{Var(X) = np(1-p)}{\uparrow}}{=} np(1-p) + (np)^2$

$$\Longrightarrow 483{,}6 = \hat{n}\hat{p}(1-\hat{p}) + (\hat{n}\hat{p})^2 \ldots (2)$$

Juntando los dos ecuaciones $\longrightarrow \begin{cases} \hat{n}\hat{p} = 21{,}6 \\ \hat{n}\hat{p}(1-\hat{p}) + (\hat{n}\hat{p})^2 = 483{,}6 \end{cases}$

$\longrightarrow 21{,}6(1-\hat{p}) + 21{,}6^2 = 483{,}6$; $\hat{p} = \frac{21{,}6 + 21{,}6^2 - 483{,}6}{21{,}6} = 0{,}21$

$\hat{n} = \frac{21{,}6}{0{,}21} = 102{,}32 \longrightarrow \boxed{\hat{p} = 0{,}21,\ n = 103}$

* Podríamos haber hecho

$S^2 = 17{,}04$, $Var(X) = np(1-p) \longrightarrow \begin{cases} 17{,}04 = \hat{n}\hat{p}(1-\hat{p}) \longrightarrow 17{,}04 = 21{,}6(1-\hat{p}) \\ 21{,}6 = \hat{n}\hat{p} \end{cases}$

$\hat{p} = 1 - \frac{17{,}04}{21{,}6} = 0{,}21$

$\hat{n} = 102{,}32$

T.5

**3.3** Estimar por el método de los momentos el nº n de grados de libertad de una $\chi^2$ siendo una M.A.S. $\{10, 15, 12, 9, 18\}$

$$\left.\begin{array}{l} E[\chi_n^2] = n \\ \overline{X} = \frac{10+15+12+9+18}{5} = 12{,}8 \end{array}\right\} \longrightarrow \hat{n} = 12{,}8 \longrightarrow \boxed{\hat{n} = 13}$$

T.5

③④ Tiempo T de duración de unos componentes C es una v.a. exponencial de media 4000h. El equipo E lo forman 5 componentes C en serie. Para que E falle basta con que no funcione un componente C.

1) Calcular la prob. de que el tiempo de vida de E sea inferior a 1000h.

T v.a tiempo de vida de un componente C

$$f_{T_C}(T) = \lambda e^{-\lambda t} = \frac{1}{4000} e^{-\frac{1}{4000}t}, \quad t \geq 0$$

$T_E$ v.a. tiempo de vida del equipo E

$$p(T_E < 1000) = 1 - p(\text{todos los comp. C duren más de } 1000h) \overset{5 \text{ comp.}}{=}$$

$$= 1 - [p(T_c > 1000)]^5 = 1 - (1 - F(1000))^5 = 1 - [1 - (1 - e^{-\frac{1000}{4000}})]^5 = 1 - e^{-5/4} = \underline{0{,}713}$$

$$F_{T_C} = \int_0^t f_{T_C}(T)\,dt = -e^{-t}\Big|_0^t = 1 - e^{-\frac{t}{4000}}$$

$$\boxed{p(T_E < 1000h) = 0{,}713}$$

2) Obtener $f_{TE}$ y calcular $E[T_E]$

$$F_{TE} = p(T_E < t) = 1 - (p(T_c > t))^5 = 1 - (1 - F_{TC}(t))^5 =$$

$$= 1 - (1 - 1 + e^{-t/4000})^5 = 1 - e^{-\frac{5t}{4000}}, \quad t > 0$$

$$\boxed{f_{TE} = \frac{dF_{TE}}{dt} = \frac{5}{4000} e^{-\frac{5t}{4000}} = \frac{1}{800} e^{-\frac{1}{800}t}, \quad t > 0}$$

$$\boxed{E[T_E] = 800h}$$

3) 10 equipos E duraron: 73, 150, 162, 368, 592, 985. Cuatro equipos duraron más de 1000h. Estimar el tiempo de vida medio por máx. verosimilitud.

$$f_{TE} = \frac{1}{\lambda} e^{-\frac{1}{\lambda}t}$$

$$F_{TE} = 1 - e^{-\frac{1}{\lambda}t}$$

$$\ell(\lambda) = f_{TE}(73) \cdot f_{TE}(150) \cdots f_{TE}(985) \cdot p(T_E > 1000)^4 =$$

$$= \left(\frac{1}{\lambda}\right)^6 e^{-\frac{1}{\lambda}\sum_{i=1}^{6} t_i} \cdot \left(e^{-\frac{1}{\lambda}1000}\right)^4 = \left(\frac{1}{\lambda}\right)^6 \cdot e^{-\frac{1}{\lambda}2330} e^{-\frac{4000}{\lambda}}$$

$$L(\lambda) = -6\ln(\lambda) - \frac{1}{\lambda}2330 - \frac{4000}{\lambda}$$

$$\frac{dL(\lambda)}{d\lambda} = -6\frac{1}{\hat{\lambda}} + \frac{2330}{\hat{\lambda}^2} + \frac{4000}{\hat{\lambda}^2} = 0; \quad \frac{-6\hat{\lambda} + 6330}{\hat{\lambda}^2} = 0, \quad \hat{\lambda} = \frac{6330}{6} = 1055h$$

$$\boxed{\hat{\lambda} = 1055h}$$

T.5

③⑤ Se toma una muestra de $n=10$ del tiempo en minutos entre dos autobuses $\{9, 10, 6, 4, 15, 6, 5, 4, 10\}$. Su $F(t)=1-e^{-\alpha t}$. Calcular la probabilidad estimada de esperar a un bus más de 10 mins.

Se pide $P(T>10)$ con $T$ v.a. tiempo entre dos buses

$P(T\geq 10)=1-p(T<10)=1-F(10)=1-(1-e^{-\alpha 10})=e^{-\alpha 10}$, por tanto necesitamos estimar $\alpha$

$$f(t)=\alpha e^{-\alpha t}$$

$$\ell(\alpha)=f(9)\,f(10)\cdots f(4)\,f(10)=\alpha^{10}e^{-\alpha\sum_{i=1}^{10}t_i}=\alpha^{10}e^{-\alpha 70}$$

$$L(\alpha)=10\log\alpha-\alpha 70,\quad \frac{dL}{d\alpha}=10\frac{1}{\alpha}-70=0,\quad \frac{10}{\hat{\alpha}}=70,\quad \hat{\alpha}=\frac{1}{7}$$

$$\Longrightarrow P(T>10)=e^{-\alpha 10}=e^{-\frac{10}{7}}=\boxed{0{,}24}$$

T.5

③⑥ Estimar por máx verosimilitud el parametro $a$ si:

$$f(x)=\frac{2a}{1-a}x^{(3a-1)/(1-a)},\quad 0\leq x\leq 1$$

$$\ell(a)=f(x_1)\cdot f(x_2)\cdots f(x_n)=\left(\frac{2a}{1-a}\right)^n\prod_{i=1}^{n}x_i^{(3a-1)/(1-a)}$$

$$L(a)=\log[\ell(a)]=n\log\frac{2a}{1-a}+\frac{(3a-1)}{(1-a)}\sum_{i=1}^{n}\log(X_i)$$

$$\frac{dL(a)}{da}=n\frac{\frac{\cancel{2}(1-a)+\cancel{2}a}{(1-a)^2}}{\frac{\cancel{2}a}{\cancel{1-a}}}+\frac{3(1-a)+(3a-1)}{(1-a)^2}\sum_{i=1}^{n}\log(X_i)=\frac{n\,1}{a(1-a)}+\frac{2}{(1-a)^2}\sum_{i=1}^{n}\log(X_i)$$

$$\frac{dL(a)}{d(a)}=0\longrightarrow\frac{n(1-\hat{a})+2\hat{a}\sum_{i=1}^{n}\log(X_i)}{\hat{a}(1-\hat{a})^2}=0$$

$$n(1-\hat{a})=-2\hat{a}\sum_{i=1}^{n}\log(X_i);\quad \hat{a}\left(-n+2\sum_{i=1}^{n}\log(X_i)\right)=-n$$

$$\boxed{\hat{a}=\frac{n}{n-2\sum_{i=1}^{n}\log X_i}}$$

T.5

③⑦ Obtener un estimador por máx. verosimilitud de $\theta$ siendo

$$f(x) = \begin{cases} \frac{2}{\theta^2} x e^{-\left(\frac{x}{\theta}\right)^2} & ; x > 0 \\ 0 & ; \text{si } x < 0 \end{cases}$$

$$\ell(\theta) = f(x_1) f(x_2) \dots f(x_n) = \frac{2^n}{\theta^{2n}} (x_1 \cdot x_2\, x_3 \dots x_n) e^{-\frac{1}{\theta^2} \sum_{i=1}^{n} x_i^2}$$

$$L(\theta) = n \log(2) - 2n \log(\theta) + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \frac{x_i^2}{\theta^2}$$

$$\frac{dL}{d\theta} = -2n \frac{1}{\hat{\theta}} + \frac{2\hat{\theta} \sum_{i=1}^{n} x_i^2}{\hat{\theta}^4} = 0 \; ; \quad \frac{-\cancel{2} n \hat{\theta}^2 + \cancel{2} \sum_{i=1}^{n} x_i^2}{\hat{\theta}^3} = 0$$

$$\boxed{\hat{\theta} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}}$$

T.5

③⑨ La función de densidad de v.a. $X$ es $f(x) = \begin{cases} x/\theta^2 & ; \text{si } 0 < x < \theta \\ 0 & , \text{en el resto} \end{cases}$

Calcular $\theta$ por máx. verosimilitud

$$\ell(\theta) = f(x_1) f(x_2) \dots f(x_n) = \frac{\prod_{i=1}^{n} x_i}{\theta^{2n}} \; , \quad \theta > x_i$$

$\ell(\theta)$ es decreciente y $\theta > x_i \longrightarrow \boxed{\hat{\theta}_{mv} = \max\{x_1, x_2, \dots x_n\}}$

T.5

③⑪ $X_1, X_2, X_3, X_4, X_5$ : m.a.s de una binomial con $n=10$ y $p$ desconocido. Demostrar que $\bar{X}/10$ es estimador centrado de $p$ y calcular su varianza.

$$X \longrightarrow B(n=10, p) ; \quad E[X] = 10p \Longrightarrow E[X_i] = 10p , \; Var(X) = 10p(1-p) = Var(X_i)$$

$$E\left[\frac{\bar{X}}{10}\right] = \frac{1}{10} E\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right] = \frac{1}{10} \cdot \frac{1}{5} \left(E[X_1] + \dots + E[X_5]\right) =$$

$$= \frac{1}{50} 10p \cdot 5 = p \Longrightarrow \boxed{\hat{p} = \frac{\bar{X}}{10} \text{ es centrado}}$$ (su esperanza coincide con el verdadero valor)

$$\boxed{Var(\hat{p}) = Var\left(\frac{\bar{X}}{10}\right) = \frac{1}{100} Var(\bar{X}) = \frac{1}{10\cancel{0}} \cdot \frac{1}{25} (1\cancel{0} p(1-p)) 5 = \frac{p(1-p)}{50}}$$

T. 7

**3.12** Al medir el radio de un círculo se comete un error aleatorio con distribución $N(0, \sigma^2)$. Se hacen n medidas $\{r_1, r_2, \dots r_n\}$ ¿Es $\hat{a} = \sum_{i=1}^{n} \frac{\pi r_i^2}{n}$ un estimador centrado del area del círculo? En caso negativo calcular su sesgo:

$Area_i = \pi r_i^2$ ; Como el error es una normal $N(0, \sigma^2)$

$a = \pi r^2$ $\quad r_i = r + \varepsilon_i$ , por tanto $r_i \rightsquigarrow N(r, \sigma^2)$

$$E[\hat{a}] = E\left[\sum_{i=1}^{n} \frac{\pi r_i^2}{n}\right] = \frac{\pi}{n} \sum_{i=1}^{n} E[r_i^2] \underset{\uparrow}{=} \frac{\pi}{n}(\sigma^2 + r^2) \cdot \cancel{n} = \pi r^2 + \pi\sigma^2 = a + \pi\sigma^2$$

$E[r_i^2] = Var(r_i) + (E[r_i])^2 = \sigma^2 + r^2$

Como $E[\hat{a}] = a + \pi\sigma^2 \neq a$

$\hat{a}$ no es centrado, sesgo$(\hat{a}) = \pi\sigma^2$

T.5

**3.13** El nº de partículas que emite una fuente radioactiva sigue una distribución de Poisson.

Estimador 1: Contar el nº de partículas emitidas en 2 min y usar este dato en la estimación

Estimador 2: Tomar nota de las partículas emitidas en cada uno de los segundos a lo largo de los 2 minutos disponiendo de una muestra de 120 datos para estimar el parametro

¿Cuál es mejor? Calcular su varianza.

$\lambda$ es nº partículas por segundo

$\hat{\lambda}_1 = \frac{\sum t_i}{120 \text{ seg}}$ ← nº partículas en 120 seg ; $\hat{\lambda}_2 = \frac{\sum t_i}{120 \text{ datos}}$ ← nº partic en cada seg $\Longrightarrow$ Ambos estimadores conducen al mismo resultado

$\boxed{Var(\hat{\lambda}) = \frac{\lambda}{T}}$ → Cuanto mayor tiempo se observe la fuente más precisa es la estimación.

T.5

**3.14** V.a. $X_1 \to N(\mu, \sigma^2)$ ; V.a. $X_2 \to N(2\mu, 3\sigma^2)$

m.a.s. de tamaño $n_1$ de $X_1$. m.a.s. de tamaño $n_2$ de $X_2$.

$\hat{\mu} = a\bar{X}_1 + b\bar{X}_2$. ① Condición de a y b para que $\hat{\mu}$ sea centrado.

② a y b para que $\hat{\mu}$ sea centrado y de var. mínima.

①

$$E[\hat{\mu}] = E[a\bar{X}_1 + b\bar{X}_2] = aE[\bar{X}_1] + bE[\bar{X}_2] = a\mu + b2\mu$$

Condición de est. centrado $E[\hat{\mu}] = \mu \to \mu(a+2b) = \mu \to \boxed{a + 2b = 1}$

② $Var[\hat{\mu}] = Var(a\bar{X}_1 + b\bar{X}_2) = a^2 Var(\bar{X}_1) + b^2 Var(\bar{X}_2) \underset{\uparrow}{=}$ $\quad \bar{X}_1 \to N\left(\mu, \frac{\sigma^2}{n_1}\right)$, $\bar{X}_2 \to N\left(2\mu, \frac{3\sigma^2}{n_2}\right)$

$$= a^2 \frac{\sigma^2}{n_1} + b^2 \frac{3\sigma^2}{n_2} \underset{\substack{\uparrow \\ a = 1-2b}}{=} (1-2b)^2 \frac{\sigma^2}{n_1} + b^2 \frac{3\sigma^2}{n_2}$$

$$\frac{d\,Var(\hat{\mu})}{db} = \frac{2(1-2b)\cdot(-2)}{n_1} + \frac{2b \cdot 3}{n_2} = 0 \quad ; \quad -4n_2 + 8bn_2 + 6bn_1 = 0 \; ; \; b = \frac{4n_2}{6n_1 + 8n_2}$$

$$a = \frac{6n_1 + 8n_2 - 8n_2}{6n_1 + 8n_2} = \frac{6n_1}{6n_1 + 8n_2}$$

$$\boxed{b = \frac{4n_2}{6n_1 + 8n_2} \; ; \; a = \frac{6n_1}{6n_1 + 8n_2}}$$

7.5

(3.15) $X_1, X_2, \ldots X_n$ m.a.s de una normal $\Big\}$ Tienen distinta media pero misma
$Y_1, Y_2 \ldots Y_m$ m.a.s de una normal $\Big\}$ varianza.

Obtener K para que $\dfrac{\sum_{i=1}^{n}(x_i-\bar{X})^2+\sum_{j=1}^{m}(Y_j-\bar{Y})^2}{K}$ tenga ECM mínimo.

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{j=1}^{m}(X_i-\bar{X})^2}{K}; \quad (n-1)\hat{S}_x^2=nS^2 \longrightarrow \hat{S}_x^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$$

$$\Rightarrow \hat{\sigma}^2=\frac{(n-1)\hat{S}_x^2}{K}+\frac{(m-1)\hat{S}_y^2}{K}, \quad \left\{\frac{(n-1)\hat{S}_x^2}{\sigma^2}\longrightarrow \chi^2_{n-1}, \; E[\chi^2_{n-1}]=n-1\right\}$$

$$E[\hat{\sigma}^2]=\frac{\sigma^2}{K}(n-1)+\frac{\sigma^2}{K}(m-1)=\frac{\sigma^2}{K}(n+m-2)$$

$n+m-2=a$

$$\text{Sesgo}[\hat{\sigma}^2]=\frac{\sigma^2}{K}(n+m-2)-\sigma^2=\sigma^2\left(\frac{n+m-2}{K}-1\right)=\sigma^2\left(\frac{a}{K}-1\right)$$

$$\text{Var}[\hat{\sigma}^2]\underset{\uparrow}{=}\text{Var}\left(\frac{(n-1)\hat{S}_x^2}{\sigma^2}\cdot\frac{\sigma^2}{K}\right)+\text{Var}\left(\frac{(m-1)\hat{S}_y^2}{\sigma^2}\frac{\sigma^2}{K}\right)=$$

$\text{Var}[\chi^2_{n-1}]=2(n-1)$

$$=2(n-1)\frac{\sigma^4}{K^2}+2(m-1)\frac{\sigma^4}{K^2}=\frac{2\sigma^4}{K^2}\left[n+m-2\right]=$$

$$=\frac{2\sigma^4}{K^2}a$$

$$ECM=\left(\text{sesgo}^2(\hat{\sigma}^2)\right)^2+\text{Var}(\hat{\sigma}^2)=\left(\sigma^2\left(\frac{a}{K}-1\right)\right)^2+\frac{2\sigma^4}{K^2}a=$$

$$\frac{dECM}{dK}=\cancel{\sigma^4}\left[\cancel{2}\left(\frac{a}{K}-1\right)\left(\frac{-a}{K^2}\right)-\frac{\overset{2}{4a}}{K^3}\right]=0$$

$$\frac{(a-K)(-a)}{K^3}-\frac{2a}{K^3}=0, \quad a^2-aK+2a=0$$

$$K=\frac{a^2+2a}{a}=a+2\underset{\substack{\uparrow\\ a=m+n-2}}{=}m+n \longrightarrow \boxed{K=m+n}$$

T.5

(3.16) M.A.S X, Y: $X \rightsquigarrow N(\mu_1, \sigma_x)$ $Y \rightsquigarrow N(\mu_2, \sigma_y)$ $\sigma_x = 2\sigma_y$. Se pide un estimador de la varianza de Y que sea combinación lineal de las var. muestrales corregidas $\hat{S}_x^2$, $\hat{S}_y^2$, sea centrado y de varianza mínima.

Buscamos $\hat{\sigma}_y^2 = a\hat{S}_x^2 + b\hat{S}_y^2$

Condicion de estim. centrado $\longrightarrow E[\hat{\sigma}_y^2] = \sigma_y^2$

$E[S_x^2] = \sigma_x^2$

$$E[\hat{\sigma}_y^2] = E[a\hat{S}_x^2 + b\hat{S}_y^2] = aE[\hat{S}_x^2] + bE[\hat{S}_y^2] =$$

$$= a\sigma_x^2 + b\sigma_y^2 \underset{\sigma_x^2 = 4\sigma_y^2}{=} \sigma_y^2(4a+b) = \sigma_y^2 \Longrightarrow 4a+b=1$$

$$Var[\hat{\sigma}_x^2] = Var[a\hat{S}_x^2 + b\hat{S}_y^2] = a^2 Var[\hat{S}_x^2] + b^2 Var[\hat{S}_y^2];$$

$$Var[\hat{S}_x^2] = Var\left[\underbrace{\frac{\hat{S}_x^2(n-1)}{\sigma_x^2}}_{\chi^2_{n-1}} \cdot \frac{\sigma_x^2}{n-1}\right] = 2\cancel{(n-1)}\frac{\sigma_x^4}{(n-1)^{\cancel{2}}}$$

$\sigma_x^4 = 16\sigma_y^4$ $\qquad b = 1-4a$

$$\longrightarrow Var[\hat{\sigma}_x^2] = 2a^2\frac{\sigma_x^4}{n-1} + 2b^2\frac{\sigma_y^4}{n-1} = 2\sigma_y^4\left[\frac{16a^2}{n-1} + \frac{b^2}{n-1}\right] =$$

$$= \frac{2\sigma_y^4}{n-1}\left[16a^2 + (1-4a)^2\right]; \quad \frac{d\,Var[\hat{\sigma}_y^2]}{da} = \frac{2\sigma_y^4}{n-1}\left[32a + 2(1-4a)(-4)\right] = 0$$

$$\Longrightarrow 32a - 8 + 32a = 0; \quad a = \frac{8}{64} = \frac{1}{8}, \quad b = 1 - 4\frac{1}{8} = \frac{1}{2}$$

$$\Longrightarrow \boxed{\hat{\sigma}_y^2 = \frac{1}{8}\hat{S}_x^2 + \frac{1}{2}\hat{S}_y^2}$$

T.5

(3.17) n medidas de un patrón de dim. real L. $L_i \longrightarrow N(L, \sigma)$

Obtener un estimador insesgado para la varianza del error y calcular su ECM

$$L_i = L + e_i \Longrightarrow e_i \longrightarrow N(0, \sigma)$$

$$e_i = L_i - L; \quad E[e_i] = 0; \quad Var[e_i] = \sigma$$

Estimamos $\sigma^2$: $\hat{\sigma}^2 = \hat{S}^2 = \frac{\sum(e_i - 0)^2}{n-1} = \frac{\sum e_i^2}{n-1}$ Varianza corregida

$$E[\hat{\sigma}^2] = E\left[\underbrace{\frac{\hat{S}^2(n-1)}{\sigma^2}}_{\chi^2_{n-1}} \cdot \frac{\sigma^2}{n-1}\right] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2 \Longrightarrow$$

Por tanto $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-1}$ es centrado

$$\boxed{ECM = Var(e_i) + \underbrace{sesgo^2(e_i)}_{0}} = Var(\hat{S}^2) = Var\left(\underbrace{\frac{\hat{S}^2(n-1)}{\sigma^2}}_{2(n-1)} \cdot \frac{\sigma^2}{n-1}\right) =$$

$$= \boxed{\frac{2\sigma^4}{n-1}}$$

T.5

③.20 Utilizamos el estimador $\hat{X} = \sum_{i=1}^{n} w_i X_i$ para estimar la media de una pob. , $\sum w_i = 1$. Calcular la media y varianza del estimador

$$\boxed{E(\hat{X}) = E\Big[\underbrace{\sum_{i=1}^{n} w_i}_{1} X_i\Big] = E[X_i] = \mu}$$

$$\boxed{Var(\hat{X}) = \sum w_i^2 Var[X_i] = \sigma^2 \sum w_i^2}$$

T.5

③.21 Para estimar la varianza de una pob. normal se proponen:

$$\hat{S}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \quad y \quad \hat{S}_1^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n+1}$$

Comparar ambos estimadores desde el punto de vista del ECM

$$ECM = (sesgo(\hat{\theta}))^2 + var(\hat{\theta})$$

$\hat{S}^2$

$$E[\hat{S}^2] = \sigma^2 \longrightarrow sesgo[\hat{S}^2] = 0$$

$$Var\left[\frac{\hat{S}^2 \cdot (n-1)}{\sigma^2} \cdot \frac{\sigma^2}{n-1}\right] = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

$$\boxed{ECM[\hat{S}^2] = \frac{2\sigma^4}{n-1}}$$

$$\hat{S}_1^2 = \frac{(n-1)\hat{S}^2}{n+1}$$

$\hat{S}_1^2$

$$E[\hat{S}_1^2] = E\left[\frac{(n-1)\hat{S}^2}{n+1}\right] = \frac{n-1}{n+1} E[\hat{S}^2] = \frac{n-1}{n+1}\sigma^2$$

$$sesgo(\hat{S}_1^2) = \sigma^2\left(\frac{n-1}{n+1} - 1\right) = \sigma^2 \frac{-2}{n+1} = \frac{-2\sigma^2}{n+1}$$

$$Var[\hat{S}_1^2] = Var\left[\frac{(n-1)\hat{S}^2}{n+1}\right] = \left(\frac{n-1}{n+1}\right)^2 Var[\hat{S}^2] = \left(\frac{n-1}{n+1}\right)^2 \frac{2\sigma^4}{n-1}$$

$$ECM[\hat{S}_1^2] = \frac{4\sigma^4}{(n+1)^2} + \frac{(n-1)2\sigma^4}{(n+1)^2} = \frac{2\sigma^4(n-1)}{(n+1)^2} = \boxed{\frac{2\sigma^4}{n+1}}$$

T.6

(322) 15 probetas forman una m.a.s. de una normal.

$$\text{M.A.S.}: \left\{\begin{array}{lllll} 40.15 & 65.10 & 49.5 & 22.4 & 38.2 \\ 60.4 & 43.4 & 26.35 & 31.2 & 55.6 \\ 47.25 & 73.2 & 35.9 & 45.25 & 52.4 \end{array}\right\}$$

① Estimar la resistencia media y su varianza

② Intervalo de confianza del 99% para la resistencia media

③ Intervalo de confianza del 99% para la varianza

④ Queremos estimar la resistencia media con [±6], confianza 95%. Hallar n.

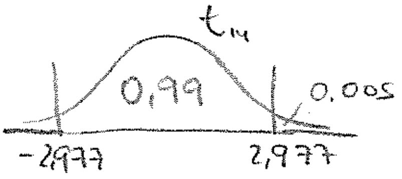
X v.a. resistencia de la prob.

① $X \to N(\mu, \sigma)$

$$\boxed{\hat{\mu} = \bar{X} = 45,75}$$

$$\boxed{\hat{\sigma}^2 = \hat{S}^2 = 201,75}$$

② $\hat{\mu} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$; $\quad \frac{\hat{\mu}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \longrightarrow t_{n-1} \left\{ \begin{array}{l} \hat{\mu} = 45,75 \\ \hat{S} = 14,204 \\ n = 15 \end{array} \right.$

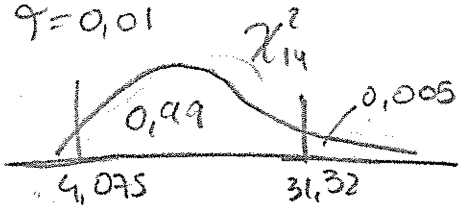


$$-2,977 \leq \frac{45,75-\mu}{\frac{14,204}{\sqrt{15}}} \leq 2,977$$

$$\mu \in 45,75 \pm 2,977 \cdot \frac{14,204}{\sqrt{15}}, \quad \boxed{\mu \in [34,83 \; , \; 56,67]}$$

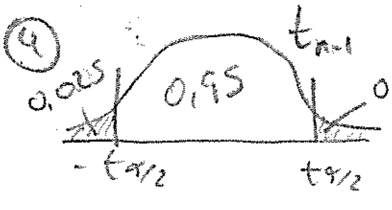
③ $\sigma^2$; $\quad \frac{\hat{S}^2(n-1)}{\sigma^2} \longrightarrow \chi^2_{n-1} \left\{ \begin{array}{l} \hat{S}^2 = 201,75 \\ n = 15 \end{array} \right.$

$\alpha = 0,01$



$$4,075 \leq \frac{201,75 \cdot 14}{\sigma^2} \leq 31,32, \quad \boxed{\sigma^2 \in [90,18 \, , \; 693,13]}$$

④



$$\mu \in 45,75 \pm \underbrace{t_{\alpha/2} \frac{14,204}{\sqrt{n}}}_{\leq 6}$$

$$\sqrt{n} > \frac{14,204}{6} t_{\alpha/2} \left\{ \begin{array}{l} n=15; \; \sqrt{15} = 3,87 \not> 5,044 \\ n=20; \; \sqrt{20} = 4,47 \not> 4,94 \\ n=25; \; \sqrt{25} = 5 > 4,87 \\ n=23; \; \sqrt{23} = 4,79 \not> 4,81 \\ n=24; \; \sqrt{24} = 4,899 > 4,88 \end{array} \right.$$

$$\Longrightarrow \boxed{n = 24}$$

7.6

③.23 Sea $f(x) = \begin{cases} \frac{4}{\sqrt{\pi}} \cdot \frac{1}{\alpha^3} x^2 e^{-(x/\alpha)^2} & ; \; x \geq 0 \\ 0 & ; \; x \leq 0 \end{cases}$ $\alpha \rightsquigarrow N(\mu, \sigma)$

$\alpha > 0$ es el parámetro de la distribución. $E[X] = \frac{2\alpha}{\sqrt{\pi}}$ y $Var[X] = \left(\frac{3}{2} - \frac{4}{\pi}\right)\alpha^2$

① Calcular el estimador máximo verosimil de $\alpha$

② Calcular el estimador por momentos de $\alpha$ y su varianza

③ $n = 100 \longrightarrow \sum_{i=1}^{100} x_i = 342$ y $\sum_{i=1}^{100} x_i^2 = 1339$; 95% confianza

Hallar intervalo de confianza para $\alpha$ con ambos estimadores.

①
$$P(\alpha) = \frac{4^n}{\pi^{n/2}} \frac{1}{\alpha^{3n}} (X_1 \cdot X_2 \cdots X_n)^2 e^{-\frac{1}{\alpha^2} \sum X_i^2}$$

$$L(\alpha) = n \log 4 - \frac{n}{2} \log \pi - 3n \log \alpha - 2 \log (X_1 \cdot X_2 \cdots X_n) - \frac{1}{\alpha^2} \sum X_i^2$$

$$\frac{dL}{d\alpha} = 0 - 0 - 3n \frac{1}{\alpha} - 0 + \frac{\sum X_i^2 \, 2\alpha}{\alpha^4} = \frac{-3n\alpha^2 + 2\sum X_i^2}{\alpha^3}$$

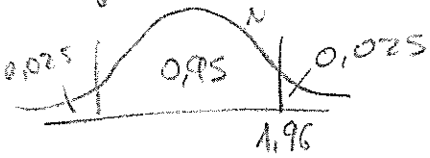
si $\frac{dL}{d\alpha} = 0 \longrightarrow -3n\hat{\alpha}^2 + 2\sum X_i^2 = 0$, $\boxed{\hat{\alpha}_{mv} = \sqrt{\frac{2\sum X_i^2}{3n}}}$

$Var(\bar{X}) = \frac{Var(X)}{n}$

② $\bar{X} = \frac{2\hat{\alpha}}{\sqrt{\pi}}$, $\boxed{\hat{\alpha}_M = \frac{\sqrt{\pi}}{2} \bar{X}}$ To lo don $Var(\hat{\alpha}_M) = Var\left(\frac{\sqrt{\pi}}{2} \bar{X}\right) = \frac{\pi}{4} Var(\bar{X}) =$ $\boxed{= \frac{\pi}{4} \cdot \frac{(3 - 8/\pi)\alpha^2}{2n}}$

③ Con $\hat{\alpha}_{mv} \sim N\left(\alpha, \frac{\hat{\alpha}_{mv}^2}{6n}\right)$, $\frac{\hat{\alpha}_{mv} - \alpha}{\frac{\alpha_{mv}}{\sqrt{6n}}} \longrightarrow N(0,1)$

Confianza 95%



$\hat{\alpha}_{mv} = \sqrt{\frac{2 \cdot 1339}{3 \cdot 100}} = 2{,}98$; $\left|\frac{2{,}98 - \alpha}{\frac{2{,}98}{\sqrt{600}}}\right| \leq 1{,}96$

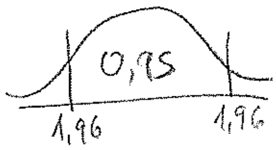
$\alpha \in 2{,}98 \pm 1{,}96 \cdot 0{,}122$

$\boxed{2{,}74 \leq \alpha \leq 3{,}22}$

Con $\alpha_M$ $E[\hat{\alpha}_M] = \frac{\sqrt{\pi}}{2} E[\bar{X}] = \frac{\sqrt{\pi}}{2} \frac{2\alpha}{\sqrt{\pi}} = \alpha$

$\hat{\alpha}_M = \frac{\sqrt{\pi}}{2} \frac{342}{100} = 3{,}03$

$\hat{\alpha}_M \rightsquigarrow N\left(\alpha, \sqrt{\frac{(3\pi - 8)\alpha^2}{8n}}\right)$, $\frac{\hat{\alpha}_M - \alpha}{\sqrt{\frac{(3\pi - 8)\alpha^2}{8n}}} \rightarrow N(0,1)$



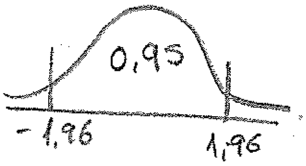
$\alpha \in 3{,}03 \pm 1{,}96 \sqrt{\frac{(3\pi - 8) 3{,}03^2}{8 \cdot 100}}$

$\boxed{2{,}78 \leq \alpha \leq 3{,}28}$

T.6

(3.24) Se observa una muestra de tamaño n de una población normal de media $\mu$ y varianza conocida $\sigma^2=64$. Calcular n para que el intervalo $\bar{x} \pm 0,5$ sea un intervalo de confianza para $\mu$? $\alpha=0,05$

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) , \quad \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$



$$\mu \in \bar{x} \pm \underbrace{1,96\frac{\sigma}{\sqrt{n}}}_{0,5} \Longrightarrow 1,96\frac{\sqrt{64}}{\sqrt{n}} = 0,5 \; ; \quad n = \left(\frac{1.96 \cdot 8}{0,5}\right)^2 = 983,45$$

$$\boxed{n \approx 984}$$

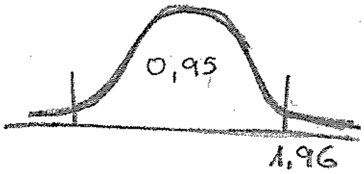
T.6

(3.25) Se desea estimar la proporción de niños entre 0 y 14 años que se encuentran adecuadamente vacunados contra la poliomelitis. Se quiere que la diferencia en valor absoluto entre la estimación final y el verdadero valor sea menor que 0,05, con probabilidad 0,95 ¿cuál es el tamaño muestral mínimo requerido?

X v.a. nº niños correctamente vacunados de un total de n

$$\hat{p} = \frac{X}{n} \; ; \quad \hat{p} \rightsquigarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) , \quad \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0,1)$$

$$P\left(\frac{|\hat{p}-p|}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0,05}{\sqrt{\frac{p(1-p)}{n}}}\right) = P\left(|Z| < \frac{0,05}{\sqrt{\frac{p(1-p)}{n}}}\right) = 0,95$$



$$\Longrightarrow \frac{0,05}{\sqrt{\frac{p(1-p)}{n}}} = 1,96 \longrightarrow n = \frac{1,96^2\, p(1-p)}{0,05^2}$$

No conocemos $p \longrightarrow$ Tomamos valor más desfavorable $p = 1/2$

$$n = \frac{1,96^2 \cdot 0,5^2}{0,05^2} \approx 385$$

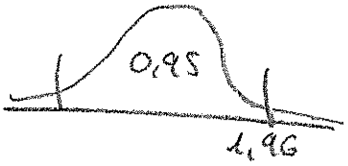
**NOTA!!** Usar $p = \frac{1}{2}$ puede proporcionar tamaños muestrales excesivamente grandes. En la práctica se realiza una preencuesta con la que se estima p.

7.6

**3.26** Una compañía de comida desea lanzar un nuevo producto. De 200 perso 37 manifiestan disposición a comprarlo. Obtener un intervalo de confia za para p con $\alpha = 0,05$ ¿Cuál debería ser el tamaño muestral si se quisiera reducir la longitud del intervalo a la mitad?

$\alpha = 0,05$

$$\hat p \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \; ; \quad \frac{\hat p - p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow N(0,1)$$



$$\left|\frac{\hat p - p}{\sqrt{\frac{\hat p(1-\hat p)}{n}}}\right| \le 1,96 \longrightarrow p \in \hat p \pm 1,96\sqrt{\frac{\hat p(1-\hat p)}{n}} \; , \quad \text{con } \hat p = \frac{37}{200}$$

Estimemos la desviación

$$p \in \frac{37}{200} \pm 1,96\sqrt{\frac{\frac{37}{200}\left(\frac{163}{200}\right)}{200}}$$

$$\Longrightarrow \boxed{0,131 \le p \le 0,239}$$

Suponemos que $\hat p = \frac{37}{200} \longrightarrow \cancel{1,96}\sqrt{\frac{\cancel{\frac{37}{200}\cdot\frac{163}{200}}}{n}} = \frac{\cancel{1,96}}{2}\sqrt{\frac{\cancel{\frac{37}{200}\cdot\frac{163}{200}}}{200}}$

$$\sqrt{\frac{1}{n}} = \frac{1}{2}\sqrt{\frac{1}{200}} \; , \quad \boxed{n = 4\cdot 200 = 800}$$

7.7

**3.27** Para comprobar la resistencia de 2 materiales A y B se toman dos muestras de tamaño 5 de cada uno de ellos

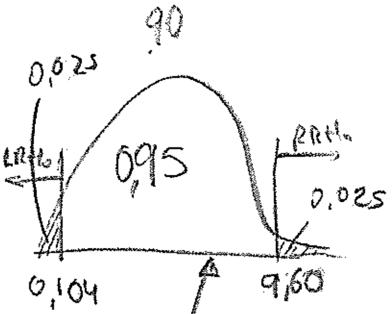
$A \rightarrow 50 \quad 45 \quad 52 \quad 64 \quad 39 \longrightarrow n_1 = 5$

$B \rightarrow 41 \quad 40 \quad 50 \quad 51 \quad 38 \longrightarrow n_2 = 5$

Suponiendo normalidad

① Contrastar la igualdad de las varianzas.

$$\begin{cases} H_0 : \frac{\sigma_A^2}{\sigma_B^2} = 1 \\ H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq 1 \end{cases} \qquad \frac{\hat S_A^2}{\hat S_B^2} \sim F_{4,4} \qquad \begin{array}{l} \hat S_A^2 = 86,5 \\ \hat S_B^2 = 36,5 \end{array}$$



$F_{4,4,0,05} = 6,39$

$F_{4,4,0,975} = \frac{1}{F_{4,4,0,025}} = \frac{1}{9,60} = 0,104$

$\frac{\hat S_A^2}{\hat S_B^2} = 2,37$

No hay suficiente evidencia estadística para rechazar la hipótesis de igualdad de varianzas

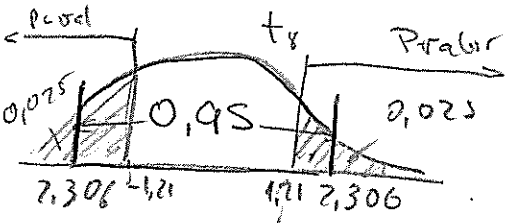
② Contrastar igualdad de las medias e indicar el valor crítico del contraste:

$$\begin{cases} H_0: \mu_A = \mu_B \\ H_1: \mu_A \neq \mu_B \end{cases} \qquad \frac{\bar{y}_A - \bar{y}_B}{\hat{S}_R \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \longrightarrow t_{n-2}$$

$$\hat{S}_R = \frac{(n_A-1)\hat{S}_A^2 + (n_B-1)\hat{S}_B^2}{n_A+n_B-2} = \frac{4 \cdot 86,5 + 4 \cdot 36,5}{5+5-2} = 61,5$$

$\bar{y}_A = 50$ ; $\bar{y}_B = 44$

$$\longrightarrow t_0 = \frac{50-44}{\sqrt{61,5}\sqrt{\frac{1}{5}+\frac{1}{5}}} = 1,21$$

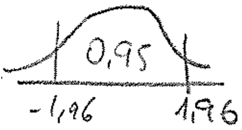


$$\underset{\text{niv. crítico}}{\text{p-valor}} = p(|t_0| > 1,21) = 2 \cdot 0,13 = \boxed{0,26} \Rightarrow \text{Lógicamente mayor que } 0,05$$

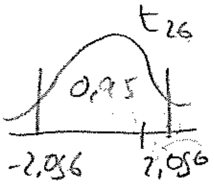
③ n para rechazar Ho

Rechazamos si $|t_0| \geq t_{(n_A+n_B-2,\, 0,025)}$

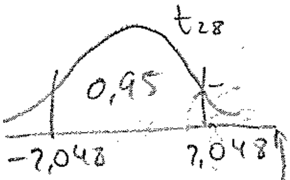
$$t_0 = \frac{6}{\sqrt{61,5}\sqrt{\frac{2}{n_A}}} \rightsquigarrow N(0,1) \quad \text{con conf } 95\%$$



$$\Rightarrow \frac{6}{\sqrt{61,5}\sqrt{\frac{2}{n_A}}} > 1,96 \longrightarrow n_A = 14 \Rightarrow \frac{6}{\sqrt{61,5}\sqrt{\frac{2}{n_A}}} \longrightarrow t_{14+14-2}$$



$$\frac{6}{\sqrt{61,5}\sqrt{\frac{2}{14}}} = 2,02 \rightarrow \text{No RRHo} \ ; \ n_A = 15 \rightarrow \frac{6}{\sqrt{61,5}\sqrt{\frac{2}{15}}} = 2,09$$



㉘ Los circuitos pueden tener 4 tipos de fallos: A, B, C, D

$P(A) = p_A$ ; $P(B) = p_B$ ; $P(C) = p_C$ ; $P(D) = p_D$. MAS: $n=200$ $\begin{cases} 12 \text{ defectos } A \\ 8 \quad " \quad B \\ 10 \quad " \quad C \\ 18 \quad " \quad D \end{cases}$

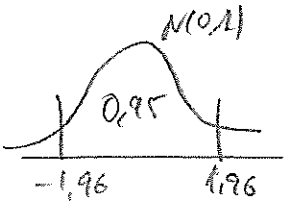
a) Estimar $p_A$, $p_B$, $p_C$ y $p_D$ por máx. verosimilitud.

$$\boxed{\hat{p}_A = \frac{12}{200} \ , \ \hat{p}_B = \frac{8}{200} \ ; \ \hat{p}_C = \frac{10}{200} \ ; \ \hat{p}_D = \frac{18}{200}}$$

b) Intervalo de confianza para cada parámetro:

$$\hat{p}_A \rightsquigarrow N\left(p_A, \sqrt{\frac{p_A(1-p_A)}{n}}\right) \ ; \quad \frac{\hat{p}_A - p_A}{\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n}}} \longrightarrow N(0,1)$$

⟵ Estim. de la var.



$$z_0 = \left| \frac{\frac{12}{200} - p_A}{\sqrt{\frac{\frac{12}{200} \cdot \frac{188}{200}}{200}}} \right| \leq 1,96 \longrightarrow p_A \in \frac{12}{200} \pm 1,96 \sqrt{\frac{\frac{12}{200} \cdot \frac{188}{200}}{200}}$$

$$\boxed{0,027 \leq p_A \leq 0,092}$$

Igualmente

$$p_B \in \frac{8}{200} \pm 1{,}96\sqrt{\frac{\frac{8}{200}\cdot\frac{192}{200}}{200}} \longrightarrow \boxed{0{,}013 \le p_B \le 0{,}067}$$

$$p_C \in \frac{10}{200} \pm 1{,}96\sqrt{\frac{\frac{10}{200}\cdot\frac{190}{200}}{200}} \longrightarrow \boxed{0{,}020 \le p_C \le 0{,}080}$$

$$p_D \in \frac{18}{200} \pm 1{,}96\sqrt{\frac{\frac{18}{200}\cdot\frac{182}{200}}{200}} \longrightarrow \boxed{0{,}050 \le p_D \le 0{,}129}$$

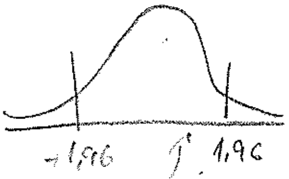
c/ Contrastar que los defectos A y B son equiprobables. 95% confianza

$$\begin{cases} H_0: p_A = p_B = p \\ H_1: p_A \neq p_B \end{cases} \qquad \hat{p} = \frac{n_A\hat{p}_A + n_B\hat{p}_B}{n_A+n_B} = \frac{\hat{p}_A+\hat{p}_B}{2}$$

$$\begin{cases} H_0: p_A - p_B = 0 \\ H_1: p_A - p_B \neq 0 \end{cases} \qquad (\hat{p}_A - \hat{p}_B) \sim N\left(0, \sqrt{\frac{p_A(1-p_A)+p_B(1-p_B)}{n}}\right)$$

$$Var(\hat{p}_A)+Var(\hat{p}_B) = \frac{p_A(1-p_A)}{n} + \frac{p_B(1-p_B)}{n}$$

$$(\hat{p}_A - \hat{p}_B) \sim N\left(0, \sqrt{\frac{2p(1-p)}{n}}\right) \qquad \hat{p} = 0{,}05$$



$$\frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{2\cdot 0{,}05\cdot 0{,}95}{200}}} = 0{,}92$$

$\boxed{\text{No se puede rechazar que } p_A \text{ y } p_B \text{ sean iguales}}$

$$\boxed{\hat{p}_A = \hat{p}_B = 0{,}05}$$

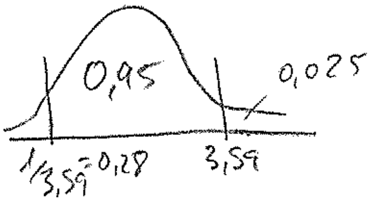
**379** Provedor A y B. $n_A = 10$; $n_B = 12$; $\bar{y}_A = 54000$; $\bar{y}_B = 49000$

$\hat{S}_A^2 = 21000$; $\hat{S}_B = 19000$. Se distribuyen normalmente.

Piezas de A son rentables si tienen una resistencia media al menos 2000 unidades mayor que B y misma variabilidad.

a/ ¿Que pieza es más rentable?

$$\begin{cases} H_0: \sigma_A^2 = \sigma_B^2 \\ H_1: \sigma_A^2 \neq \sigma_B^2 \end{cases} \qquad \frac{\hat{S}_A^2}{\hat{S}_B^2} \sim F_{9,11}; \qquad \frac{\hat{S}_A^2}{\hat{S}_B^2} = \frac{21000^2}{19000^2} = 1{,}22$$



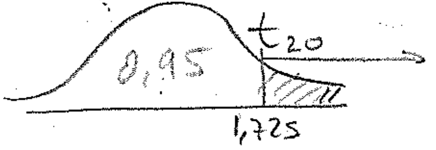
⟹ No se rechaza que la varianza de ambas poblaciones sean iguales

$$\begin{cases} H_0: \mu_A - \mu_B \leq 2000 \\ H_1: \mu_A - \mu_B > 2000 \end{cases}$$

$$\hat{S}_R^2 = \frac{\hat{S}_A^2(n_A-1)+\hat{S}_B^2(n_B-1)}{n_A+n_B-2} = \frac{2100^2\cdot 9+1900^2\cdot 11}{10+12-2}$$

$$= 3970000 \longrightarrow \hat{S}_R = 1992{,}5$$

$$t_0 = \frac{(\hat{\mu}_A-\hat{\mu}_B)-2000}{\hat{S}_R\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} = \frac{54000-49000-2000}{1992{,}5\sqrt{\frac{1}{10}+\frac{1}{12}}} = 3{,}52$$



∴ ⟶ Se rechaza Ho

⟶ $\boxed{\text{Se elige al proveedor A}}$

b) Obtener un intervalo de confianza del 90% para la diferencia de medias.

$$\mu_A-\mu_B \in 5000 \mp 1{,}725\cdot 1992{,}5\sqrt{1/10+1/12}$$

$$\boxed{3528{,}34 \leq \mu_A-\mu_B \leq 6471{,}66 \text{ ; conf } 90\%}$$

(330) 95% de bolsas resisten 6 Kg o más. M.A.S: n=40 → 6 bolsas rotas.
¿podemos rechazar afirmación del fabricante?

X v.a. nº bolsas rotas de un total de n ⟶ $B(n,p)$

Se pide realizar el contraste   p(prob roto) = 1-p(resistir)

$$\begin{cases} H_0: p = 0{,}05 \\ H_1: p > 0{,}05 \end{cases} ; \quad \hat{p} = \frac{6}{40} = 0{,}15$$

$$\hat{p} \longrightarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \longrightarrow N(0,1)$$

$$z_0 = \frac{0{,}15-0{,}05}{\sqrt{\frac{0{,}05\cdot 0{,}95}{40}}} = 2{,}9, \quad \text{p-valor} = P(Z \geq 2{,}9) = 0{,}0018$$

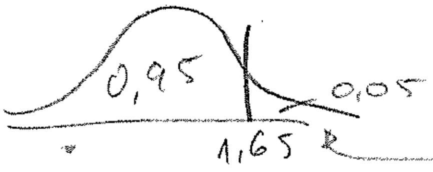
Con confianza $\alpha = 0{,}05$ ó $\alpha = 0{,}01$ ⟶ $\boxed{\text{Rechazamos afirmación del fabricante}}$

(332) $\lambda = 3{,}5$ av./100Km Contrastar $\begin{cases} H_0: \lambda = 3{,}5 \\ H_1: \lambda > 3{,}5 \end{cases}$

100 Km — 9
100 Km — 3
300 Km — 6
400 Km — 11
400 Km — 32
500 Km — 70
500 Km — 9
500 Km — 16
600 Km — 36
800 Km — 32
4200

$$\hat{\lambda} = 4{,}1428 \sim\!\!\to N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$$

$$\frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\lambda}{n}}} \longrightarrow N(0,1); \quad z_0 = \frac{4{,}1428-3{,}5}{\sqrt{\frac{3{,}5}{42}}} = 2{,}22$$



$\boxed{\text{Rechazamos } \lambda = 3{,}5}$