• Expansion de la varianza

$$\boxed{\sigma^2 = \frac{1}{N}\sum (x_i-\bar{x})^2 n_i} = \frac{1}{N}\sum \left(x_i^2+\bar{x}^2-2x_i\bar{x}\right)n_i =$$

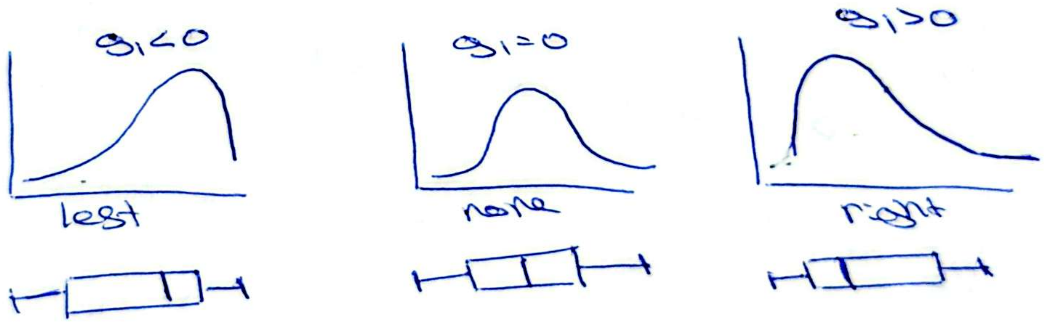
$$= \frac{1}{N}\left[\sum x_i^2 n_i + \bar{x}^2 \sum n_i - 2\bar{x}\sum x_i n_i\right] =$$

$$= \frac{1}{N}\left[\sum x_i^2 n_i + \bar{x}^2 N - 2\bar{x}\,\cancel{\sum}\,\bar{x} N\right] =$$

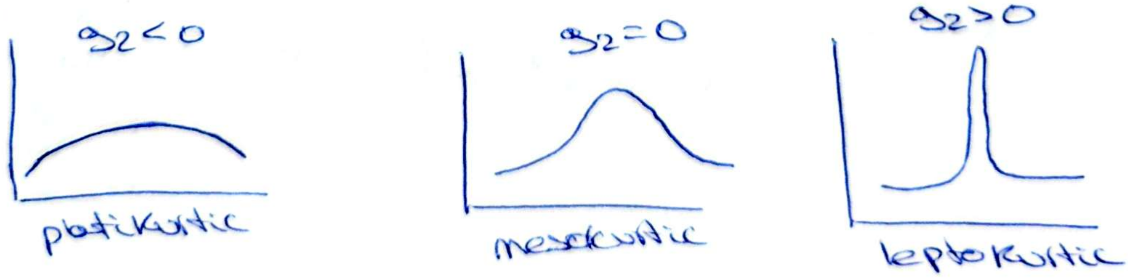
$$= \frac{1}{N}\sum x_i^2 n_i + \bar{x}^2 - 2\bar{x}^2 = \boxed{\frac{1}{N}\sum x_i^2 n_i - \bar{x}^2}$$

• Coeficientes de forma

→ Fisher (asimetria) skewness : $g_1 = \frac{1}{N\sigma^3}\sum (x_i-\bar{x})^3$



→ Kurtosis : $g_2 = \frac{1}{N\sigma^4}\sum (x_i-\bar{x})^4 - 3$



→ we considere gaussian kind is

$-2 < g_1 < 2$

$-2 < g_2 < 2$

- Transformación de variables

→ Transformación afín: cambiar variables

$$z = a x + b$$

b ← traslación

a ↑ transformación lineal

- ejemplo: medida en metros con zapatos → convertirlos sin zapatos

→ Tipificación / standarización: comparación relativa

$$z = \frac{x - \bar{x}}{\sigma}$$

- ejemplo: que es relativamente más alto un docente o una hormiga dentro de su grupo

→ cambio de variable sobre las medidas a,b,d

$\boxed{\text{media}}$ $\bar{x} = \frac{1}{N}\sum x_i \rightarrow \boxed{\bar{z}} = \frac{1}{N}\sum(ax_i + b) =$

$= a \frac{1}{N}\sum x_i + \frac{1}{\cancel{N}} b \cancel{N}$ ($\frac{1}{N}\sum x_i = \bar{x}$)

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

$\boxed{\text{varianza}}$

$\sigma_x^2 = \frac{1}{N}\sum(x_i - \bar{x})^2 \rightarrow \boxed{\sigma_z^2} = \frac{1}{N}\sum\left[ax_i + \cancel{b} - (a\bar{x} + \cancel{b})\right]^2$

$= \frac{1}{N} a^2 \sum(x_i - \bar{x})^2$ (σ_x^2)

$\boxed{= a^2 \sigma_x^2}$

$\boxed{g_1 \text{ y } g_2}$

↳ son invariantes bajo transf afines
son invariantes sobre transformaciones afines

$\boxed{CV}$ → No es invariante

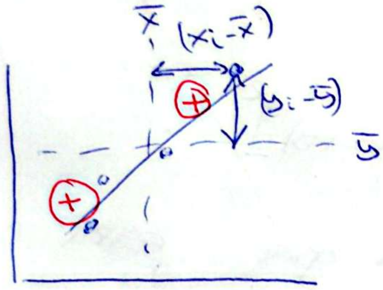
$CV_x = \frac{\sigma_x}{\bar{x}} \rightarrow CV_z = \frac{\sigma_z}{\bar{z}} = \frac{a\sigma_x}{a\bar{x} + b}$

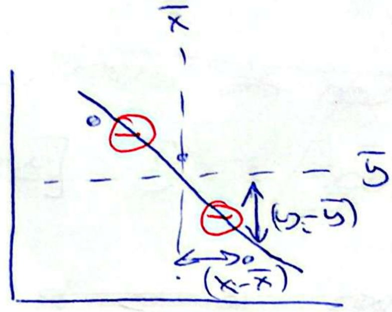
pero sí sobre transformaciones lineales puras

Regression Lineal I

$$\boxed{Y - \overline{Y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \overline{x})}$$

$$\sigma_{xy} = \frac{1}{N}\sum n_i (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{N}\sum n_i x_i y_i - \overline{x}\,\overline{y}$$





Pearson correlation coefficient

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

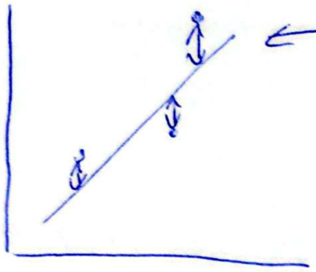
coefficient of determination r^2:

% of ~~error~~ errors in the prediction explained by variances precission

Predictions are valid only in the range of study

Tema X

Regression lineal II: minimización de la función error



$g(x) = ax + b$

$$E(a,b) = \sum\left[y_i - g(x_i)\right]^2$$

a y b tq $\dfrac{\partial E}{\partial a} = 0$ & $\dfrac{\partial E}{\partial b} = 0$

$$\frac{\partial E}{\partial a} = 0$$

$$\cancel{2}\sum\left[y_i - (ax_i + b)\right]\cdot(-x_i) = 0$$

$$-\sum y_i x_i + \sum a x_i x_i + \sum b x_i = 0$$

$$-\sum y_i x_i + a\sum x_i^2 + b\sum x_i = 0$$

$$-\sum y_i x_i + a\sum x_i^2 + b\bar{x}$$

$$-\sum y_i x_i + a\sum x_i^2 + (\bar{y} - a\bar{x})\bar{x} = 0$$

$$\frac{1}{N}\left[-\sum y_i x_i + \bar{y}\bar{x} + a\left[\sum x_i^2 - \bar{x}^2\right] = 0\right]$$

$$-\sigma_{xy} + a\sigma_x^2 = 0$$

$$\boxed{a = \frac{\sigma_{xy}}{\sigma_x^2} \qquad b = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x}}$$

$$\frac{\partial E}{\partial b} = 0$$

$$\cancel{2}\sum\left[y_i - (ax_i + b)\right] = 0$$

$$\frac{1}{N}\left[\sum y_i - a\sum x_i - \sum b = 0\right]$$

$$\bar{y} - a\bar{x} - b = 0$$

$$\boxed{b = \bar{y} - a\bar{x}}$$

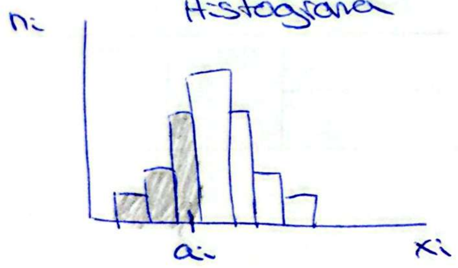
$$g(x) = y = \frac{\sigma_{xy}}{\sigma_x^2}x + \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x}$$

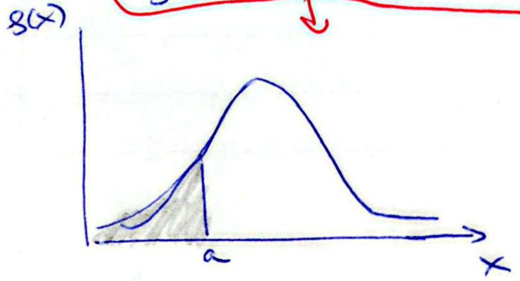
$$\boxed{y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})}$$

cvd

Probabilidad I: variable aleatoria continua

Estadística	Probabilidad
Datos finitos → Histograma	función integrable $f(x)=Ae^{-z^2/2}$ $\quad z=\frac{x-\mu}{\sigma}$





Estadística	Probabilidad
$N \to$ grande	$N \to \infty$
$\sum_i$	$\int dx$
$N = \sum_i n_i$	$\text{Norma} = \int_{-\infty}^{\infty} f(x)\,dx$
$\bar{x} = \frac{1}{N}\sum_i^N n_i x_i$	$\mu = \frac{1}{\text{Norma}}\int_{-\infty}^{\infty} f(x)\, x\, dx$
$s^2 = \frac{1}{N}\sum_i^N n_i (x_i - \bar{x})$	$\sigma^2 = \frac{1}{\text{Norma}}\int_{-\infty}^{\infty} f(x)\,(x-\mu)\,dx$
$F(x_i < a_i) = \frac{1}{N}\sum_{i=1}^{a_i} n_i x_i$ Suma con ordenador o a mano	$P(x<a) = \frac{1}{\text{Norma}}\int_{-\infty}^{a} f(x)\,dx$ Suma con ordenador o tablas
condición de F ~~frec~~ acum	condición de f ~~acum~~ prob
$F(x_1 < x < x_N) = \frac{1}{N}\sum_{i=1}^{N} n_i = 1$	$P(-\infty < x < \infty) = \frac{1}{\text{Norma}}\int_{-\infty}^{\infty} f(x)\,dx = 1$
⇓ distribución de frecuencias	⇓ distribución de probabilidad PDF

Continuous random variable

↳ can take any value in a real interval
↳ it make no sense to calculate the prob of an isolated value and we calculate prob. for intervals

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad \Big\} \quad P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

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

example

$$f(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \geq 0 \end{cases}$$

Normalization: $$\int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\int_{-\infty}^{0} 0\,dx}_{0} + \int_0^{\infty} e^{-x}\,dx = \left[-e^{-x}\right]_0^{\infty} = -e^{-\infty} + e^{0} = 1$$

$$P(0 \leq x \leq 2) = \int_0^2 e^{-x}\,dx = \left[-e^{-x}\right]_0^2 = 0,8646$$

$$\mu = \int_{-\infty}^{\infty} x f(x) = \cancel{\int_{-\infty}^{0} 0\,dx} + \int_0^{\infty} x e^{-x}\,dx = \left[-e^{-x}(1+x)\right]_0^{\infty} = 1$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \cancel{\int_{-\infty}^{0} 0\,dx} + \int_0^{\infty} x^2 e^{-x} dx - \mu^2 = \left[-e^{-x}(x^2+2x+2)\right]_0^{\infty} - 1^2 = 1$$

Uniform

$f(x) = \frac{1}{b-a}$ $\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(b-a)^2}{12}$

Normal $N(\mu, \sigma)$

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

μ: mean

σ: standar desviation

to avoid integrating the normal density to compute prob it is comun to use tables

↳ z-values

+ standarization $z = \frac{x-\mu}{\sigma}$

$$N(\mu=6, \sigma=1{,}5) \rightarrow P(x<5) = \Bigg(\underbrace{\frac{x-6}{1{,}5}}_{z} < \frac{5-6}{1{,}5}\Bigg) =$$

$$= P(z < -0{,}67) = 0{,}2514$$

table z value
$(-0{,}6\,;\ 0{,}07) = 0{,}2514$

Descrete Random Variables

$f(x_i)$: probability function
$= P(X = x_i)$

$F(x_i)$: distribution function, cumulative of $f(x_i)$
$= P(X \leq x_i) = f(x_0) + f(x_1) + f(x_2) + \dots + f(x_i)$

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

$$\sigma^2 = \sum_{i=1}^{n} x^2 f(x) - \mu^2 \qquad \sigma = \sqrt{\sigma^2}$$

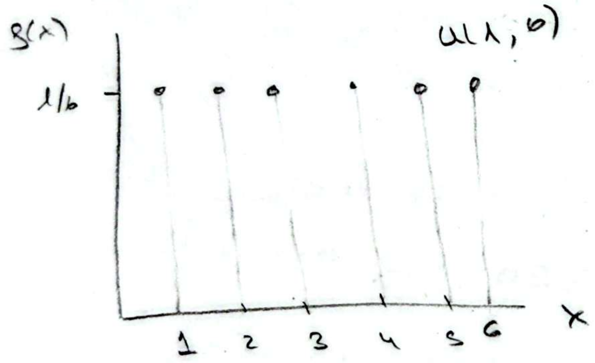
Uniform $\sim U(a,b)$ ← min, ← max

$Ran(X) = \{a, a+1, \dots, b\}$

$$\boxed{f(x) = \frac{1}{b-a+1}}$$

$$\mu = \sum_{i=0}^{b-a} \frac{a+i}{b-a+1} = \frac{a+b}{2}$$

$$\sigma^2 = \sum_{i=0}^{b-a} \frac{(a+i-\mu)^2}{b-a+1} = \frac{(b-a+1)^2-1}{12}$$



when all the values of a random variabl X have equal prob
↓
(rolling dice)

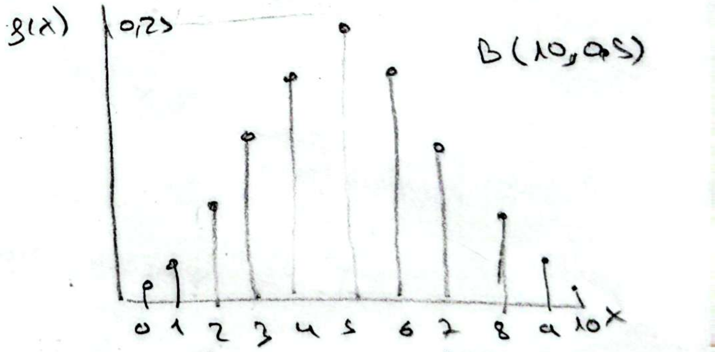
Binomial

$Ran(X) = \{0, 1, \dots, n\}$

n: num of repetitions
p: prob of succes

$$\boxed{f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

$\mu = n \cdot p$
$\sigma^2 = n \cdot p \cdot (1-p)$



$f(4) = \binom{10}{4} 0{,}5^4 (1-0{,}5)^{10-4} = \frac{10!}{4!\,6!} 0{,}5^4 \cdot 0{,}5^6$
$= 0{,}2051$

$\mu = 10 \cdot 0{,}5 = 5$

sequence of (n) repetitions of same trial with same prob (p)
↳ each trial is identical and produce two possible outcoms
↳ each trial is independent
↳ each trial has same prob
(tossing n coins)

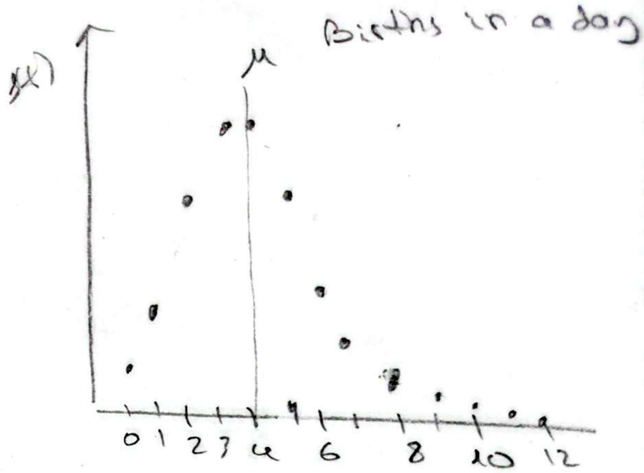
Poisson

$$f(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$Ran(X) = \{0, 1, \ldots, \infty)$

λ: average rate of event for interval unit (change if the interval changes)

$\mu = \lambda \qquad \sigma^2 = \lambda$



$f(5) = e^{-4} \frac{4^5}{5!} = 0{,}1563$

- \# of events occurring in a fixed interval of time/space
  - ↳ Births in a months
  - ↳ errors in one house
  - ↳ red blood cells in a volume of blood
- events occur independently
- the same average rate of event λ for every interval unit

$\boxed{\text{law of rare events}}$

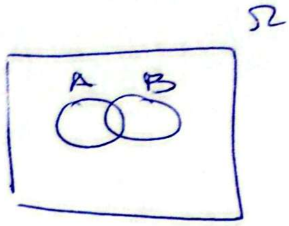
Poisson = Binomial when $\begin{cases} \text{number of trials } n \to \infty \\ \text{prob of success } p \to 0 \end{cases}$

$$\lim_{\substack{n \to \infty \\ p \to 0}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

$\uparrow$ $\lambda = np$

! it works for $n \geq 30$ and $p \leq 0{,}1$!

Escaneado con CamScanner



Ω: a set of all possibilities

A: a sub-set of possibilities

B: a different sub-set

$\bar{A}$: complementary

$A \cup B$: union

$A \cap B$: intersection

$P(\Omega) = 1$

$P(A) = \frac{N_A}{N_\Omega}$

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

~~P(A+B) = CP(A∪B~~

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

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

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

independent: $P(A|B) = P(A)$ or $P(A \cap B) = P(A) \cdot P(B)$

incompatibles: $P(A \cap B) = 0$ or $P(A) + P(B) = P(A \cup B)$

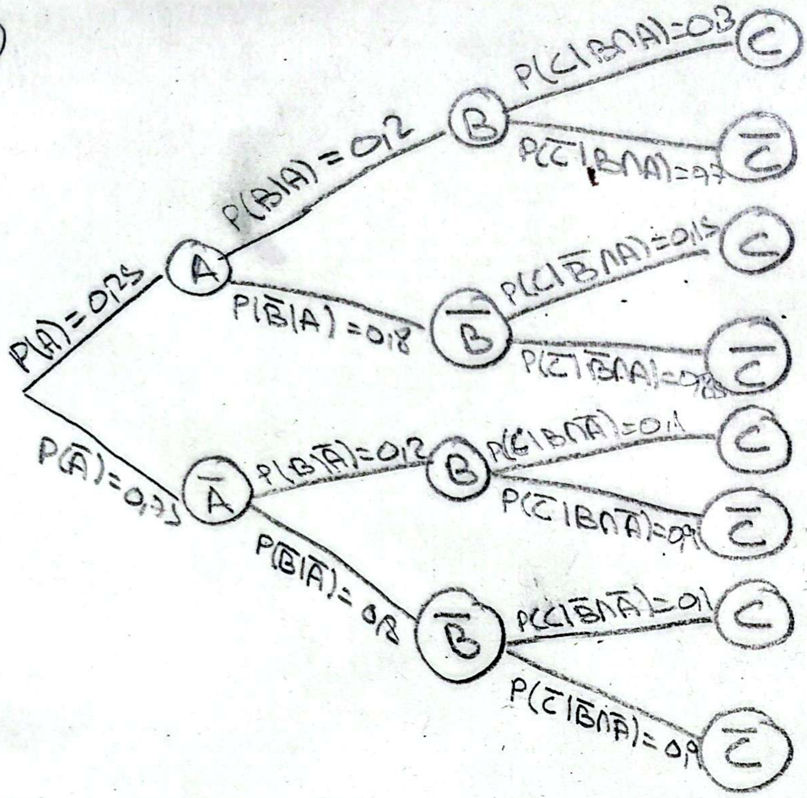
Bayes

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

Total

$$P(B) = \sum_A \underbrace{P(A)\,P(B|A)}_{\sum_A P(B \cap A)}$$

41



$P(A \cap B \cap C) = 0{,}25 \cdot 0{,}2 \cdot 0{,}3 = 0{,}015$

$P(A \cap B \cap \overline{C}) = 0{,}25 \cdot 0{,}2 \cdot 0{,}7 = 0{,}035$

$P(A \cap \overline{B} \cap C) = 0{,}25 \cdot 0{,}8 \cdot 0{,}15 = 0{,}03$

$P(A \cap \overline{B} \cap \overline{C}) = 0{,}25 \cdot 0{,}8 \cdot 0{,}85 = 0{,}17$

$P(\overline{A} \cap B \cap C) = 0{,}75 \cdot 0{,}2 \cdot 0{,}1 = 0{,}015$

$P(\overline{A} \cap B \cap \overline{C}) = 0{,}75 \cdot 0{,}2 \cdot 0{,}9 = 0{,}135$

$P(\overline{A} \cap \overline{B} \cap C) = 0{,}75 \cdot 0{,}8 \cdot 0{,}1 = 0{,}06$

$P(\overline{A} \cap \overline{B} \cap \overline{C}) = 0{,}75 \cdot 0{,}8 \cdot 0{,}9 = 0{,}54$

Teorema probabilidad total

$$P(C) = P(A \cap B \cap C) + P(A \cap \overline{B} \cap C) + P(\overline{A} \cap B \cap C) + P(\overline{A} \cap \overline{B} \cap C) =$$

$$= 0{,}015 + 0{,}03 + 0{,}015 + 0{,}06 = 0{,}12$$

Teorema de Bayes

all the lines

$$P(B|C) = \frac{\sum P(B \cap C)}{P(C)} = *$$

all the lines

$$P(B \cap C) = \sum P(B_i) \cdot P(C|B_i) =$$

$$= P(B \cap A) \cdot P(C|B \cap A) + P(B \cap \overline{A}) P(C|B \cap \overline{A}) =$$

$$= P(A) P(B|A) \cdot P(C|B \cap A) + P(\overline{A}) P(B|\overline{A}) P(C|B \cap \overline{A}) =$$

$$= 0{,}25 \cdot 0{,}2 \cdot 0{,}3 + 0{,}75 \cdot 0{,}2 \cdot 0{,}1 = 0{,}03$$

sum of all the lines that contain $B \cap C$

$$* = \frac{0{,}03}{0{,}12} = 0{,}25$$

$$P(\bar{A} \mid \bar{B} \cap \bar{C}) = \frac{P(\bar{A} \cap \bar{B} \cap \bar{C})}{P(\bar{B} \cap \bar{C})} = \frac{0{,}54}{0{,}71}$$

Teorema prob total

$$P(\bar{B} \cap \bar{C}) = \sum_{\text{all even}} P(\bar{B}) \cdot P(\bar{B} \mid \bar{C}) = \ldots$$

Sum all the lines that contain $\bar{B} \cap \bar{C}$

$$= P(\bar{A} \cap \bar{B} \cap \bar{C}) + P(A \cap \bar{B} \cap \bar{C}) =$$

$$= 0{,}54 + 0{,}17 = 0{,}71$$

(e) Son B y C independientes ?

$$P(B \mid C) = P(B)$$

$\uparrow$ 0,25

Teorema de la Probabilidad total

$$P(B) = \sum_{\text{even}} P(B_i) = P(A) \cdot P(B \mid A) + P(\bar{A}) \cdot P(B \mid \bar{A}) =$$

Sum all the lines that contain B

$$= 0{,}25 \cdot 0{,}2 + 0{,}75 \cdot 0{,}2 = 0{,}2$$

$$0{,}25 \stackrel{?}{=} 0{,}2 \quad \text{NO}$$

$\Downarrow$

no son independientes