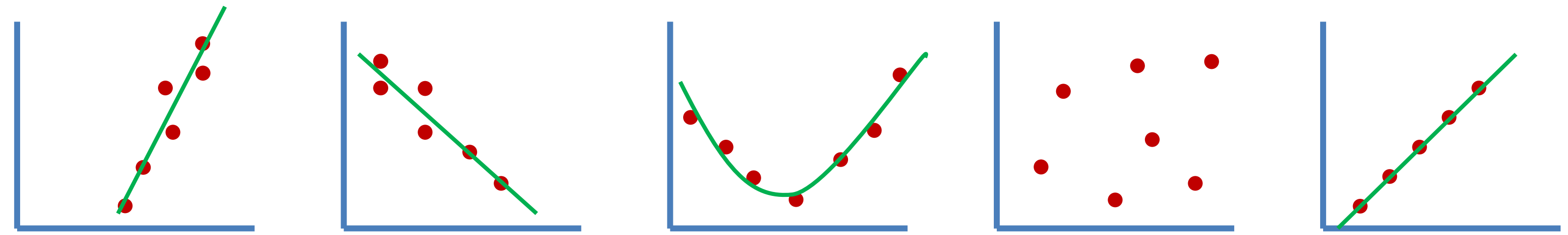
Relación entre dos variables

- Relación o dependencia entre dos variables



Lineal directa | Lineal inversa | No lineal | Sin relación | Funcional

- Con la Esperanza y la Varianza sólo se pueden estudiar las distribuciones marginales.
- Para estudiar la relación entre dos variables:

Covarianza y **Correlación**

Covarianza

- La covarianza es una medida de la dispersión conjunta de dos variables.
- Se define como: $\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mu_X)(Y - \mu_Y))$
- Se calcula como: $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$
  - Puede no existir (es una esperanza).
  - Puede ser positiva, nula o negativa.
  - Si X e Y son independientes $\rightarrow \mathrm{Cov}(X, Y) = 0$.
  - Depende de la unidad de medida.

Propiedades de la covarianza

- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$
  - En particular: $\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y)$
- $\mathrm{Cov}(X,a) = 0$
- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Cov}(aX+b, cY+d) = ac \cdot \mathrm{Cov}(X,Y)$
- $\mathrm{Var}(aX+bY+c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab \cdot \mathrm{Cov}(X,Y)$

Problema 4.12

- Dada la función de cuantía:

Y				
1	0'1	0'5	0	
0	0'2	0'1	0'1	
f	**0**	**1**	**2**	***X***

Hallar la covarianza.

Problema 4.13

- Dada la función de densidad conjunta

$$f(x,y) = \begin{cases} \frac{3}{8}(x^2 + y), & (x,y) \in [0,1] \times [0,2] \\ 0, & resto \end{cases}$$

Hallar la covarianza.

Correlación

- La correlación es una medida de la relación entre dos variables.
- Se define y calcula como: $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$
  - Puede no existir.
  - Toma valores entre -1 y 1.
  - Si X e Y son independientes $\rightarrow \rho = 0$ (incorreladas).
  - Si $|\rho| = 1 \rightarrow$ dependencia funcional.
    - Lineal: $Y = aX + b$, donde $\rho = 1$ si a es positivo y $\rho = -1$ si es neg.
  - Si $|\rho|$ es cercano a 1 $\rightarrow$ se puede encontrar un buen ajuste mediante regresión lineal: $Y \approx aX + b$.

Problema 4.14

- Dada la función de cuantía:

Y

1	0'1	0'5	0
0	0'2	0'1	0'1
f	**0**	**1**	**2**

X

Hallar la correlación.

- Del problema anterior:

X	0	1	2
f_1	0'3	0'6	0'1

Y	0	1
f_2	0'4	0'6

$$\mathrm{E}(XY) = 0'5; \quad \mathrm{E}(X) = 0'8; \quad \mathrm{E}(Y) = 0'6; \quad \mathrm{Cov}(X,Y) = 0'02$$

Problema 4.15

- Dada la función de densidad conjunta

$$f(x,y) = \begin{cases} \frac{3}{8}(x^2+y), & (x,y) \in [0,1] \times [0,2] \\ 0, & resto \end{cases}$$

Hallar la correlación.

- Del problema anterior:

$$f_1(x) = \frac{3}{4}(x^2+1), x \in [0,1]$$

$$f_2(y) = \frac{1}{8}(3y+1), y \in [0,2]$$

$$E(XY) = \frac{11}{16}; \quad E(X) = \frac{9}{16}; \quad E(Y) = \frac{5}{4}; \quad Cov(X,Y) = -0'015625$$

Esperanza condicional

- Dada una distribución bidimensional continua (X, Y), la esperanza de X para un valor dado $Y = y$ es
  - V.A. discreta: $\mathrm{E}(X|Y=y) = \sum_i x_i g_1(x_i|Y=y)$
  - V.A. continua: $\mathrm{E}(X|Y=y) = \int_{-\infty}^{\infty} x g_1(x|Y=y)\, dx$
- De forma análoga para $\mathrm{E}(Y \mid X = x)$
- $\mathrm{E}(\mathrm{E}(X \mid Y)) = \mathrm{E}(X)$

Problema 4.16

- Dada la función de cuantía:

Y				
1	0'1	0'5	0	
0	0'2	0'1	0'1	
f	**0**	**1**	**2**	X

Hallar $E(X|Y = 0)$.

Problema 4.17

- Dada la función de densidad conjunta

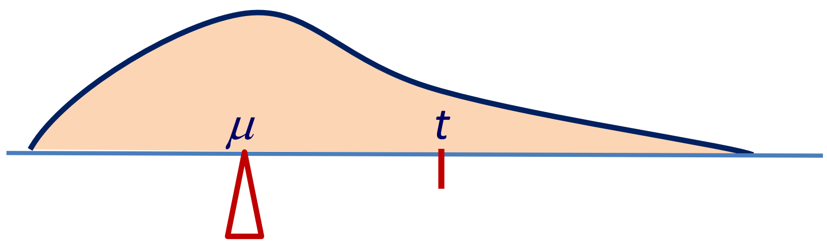
$$f(x,y) = \begin{cases} \frac{3}{8}(x^2 + y), & (x,y) \in [0,1] \times [0,2] \\ 0, & resto \end{cases}$$

Hallar E($X|Y$ = ½).

Desigualdad de Markov

- Las desigualdades de Markov y Chebychev nos permiten establecer una cota superior para la probabilidad sin necesidad de conocer su distribución.
- Se basan en la idea de que la probabilidad de que una v.a. X se aleje de su media es cada vez menor.
- T$^{\text{ma}}$ (**Desigualdad de Markov**): Sea X una v.a. con media μ y $P(X \geq 0) = 1$, entonces

$$P(X \geq t) \leq \frac{\mu}{t}$$



Problema 4.18

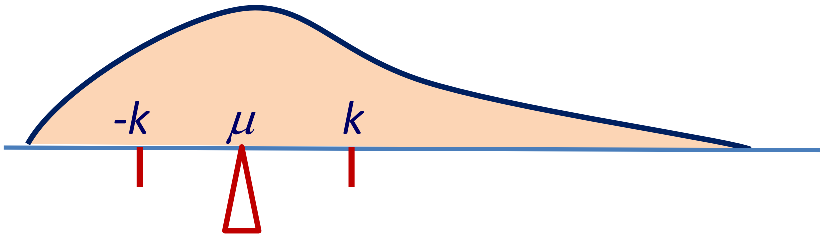
- En un tipo operación, la esperanza de la ganancia es de 10€ (y, sin tener en cuenta el coste de la operación, no se puede perder). Usando la desigualdad de Markov, obtener una cota superior para la probabilidad de ganar más de 15€ en una operación.

Desigualdad de Chebychev

- La desigualdad de *Chebychev* nos permite establecer una cota para la probabilidad usando la varianza como medida de dispersión de la v.a. alrededor de su media.

- T$^{\text{ma}}$ (**Desigualdad de *Chebychev***): Sea X una v.a. con media μ y varianza finita σ^2, entonces para todo $k > 0$ se cumple:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$



Problema 4.19

- En una ciudad, el número de fracturas de cadera en personas mayores de 65 años es de 50 casos al mes en promedio.

  En un estudio se ha obtenido que la varianza de los casos es 25 pero no se ha podido establecer su distribución.

  Usando la desigualdad de *Chebychev* ¿qué cota de probabilidad podemos tener de que este mes el número de casos difiera en más de 10 de la media?

Media muestral

- Una **muestra aleatoria** de tamaño *n* de una v.a. *X* consiste en un conjunto de *n* variables independientes X_i con la misma distribución que *X*.
- La media aritmética de una muestra se llama **media muestral**:

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- Sea μ la media de X_i y σ^2 la varianza, entonces:

$$E(\overline{X}_n) = \mu \quad Var(\overline{X}_n) = \frac{\sigma^2}{n}$$

- Desigualdad de *Chebychev* para la media muestral:

$$P(|\overline{X}_n - \mu| \geq k) \leq \frac{\sigma^2/n}{k^2}$$