Population vs. Sample Parameters

Population : entire set of data points that we are interested in studying.

Sample : a subset of the population that is selected for analysis in a statistical study.

It should be representative of the population to ensure the accuracy of the statistical analysis

Population parameters : mean (μ), variance (σ^2), standard deviation (σ), etc.

Sample properties: sample mean ($\bar{X}$), sample variance (s^2), sample standard deviation (s)

Population's statistical properties can be inferred from the underlying probability density function (pdf) $\longrightarrow f_X(x)$ or $f(x)$

① Population mean / expectation / expected value:

$$\mu = \langle X \rangle = E(X) = \underbrace{\int_{-\infty}^{\infty} x^1 \, \underbrace{f(x)\, dx}_{}}_{}$$

probability that X lies between x & $(x+dx)$

first moment of the pdf

(2) Population variance

Quantifies the extent of spread in a random variable following a certain distribution

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

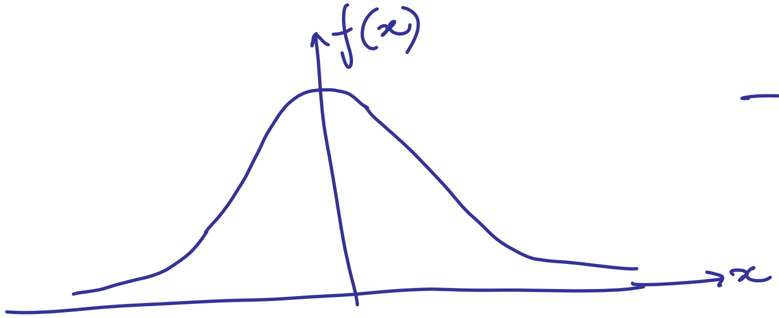
$$\rightarrow E(X^2) = \underbrace{\int_{-\infty}^{\infty} \boxed{x^2} f(x)\, dx}_{\text{second moment of pdf}}$$

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

(3) Standard deviation

Square root of the variance and thus has the same unit as the random variable

$$\sigma = \sqrt{Var(X)} = \sqrt{E(X^2) - (E(X))^2}$$



Normal/ Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

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

④ <u>Covariance</u>

For two random variables X and Y, the covariance provides an understanding of the relationship between them.

$$Cov(X, Y) = E\left[(X - E(X)) \cdot (Y - E(Y))\right]$$

$$= E\left[XY - X\boxed{E(Y)} - \boxed{E(X)} \cdot Y + \boxed{E(X)}\boxed{E(Y)}\right]$$

Expectation operator is a $\boxed{\text{linear operator.}}$

$$Cov(X,Y) = E[XY] - E[X\,E(Y)] - E[Y\,E(X)] + E[E(X)\,E(Y)]$$

$$= E(XY) - E(Y)E(X) - \cancel{E(X)E(Y)} + \cancel{E(X)E(Y)}$$

$$\boxed{Cov(X,Y) = E(XY) - E(X)E(Y)}$$

$Y = X$

$$\underline{Cov(X,X)} = E(X^2) - (E(X))^2 = \underline{Var(X)}$$

Ex: X and Y are two independent random variables with pdf $f_X(x)$ & $f_Y(y)$. Find the covariance of X & Y.

$$f_{XY}(x,y) = f_X(x)\, f_Y(y).$$

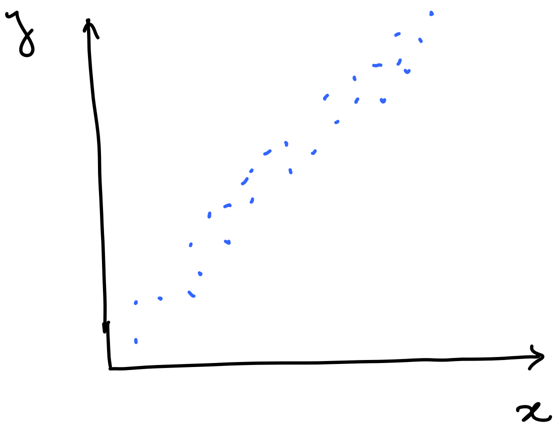
$$\mathrm{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\, dx\, dy - \left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\, dy\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x)\, f_Y(y)\, dx\, dy - \cancel{\left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\, dy\right)}$$

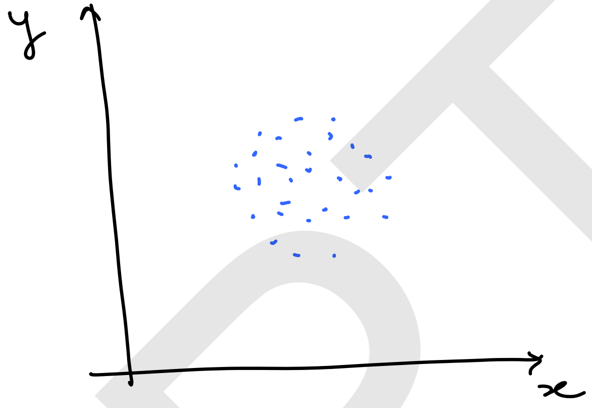
$$\cancel{\left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right) \times \left(\int_{-\infty}^{\infty} y f_Y(y)\,dy\right)}$$

$$= \boxed{0}$$

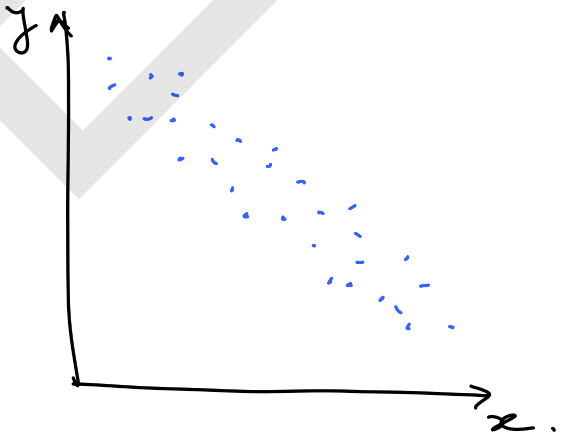


$\mathrm{Cov}(X, Y) > 0$

larger values of X correspond to larger values of Y



$\mathrm{Cov}(X, Y) \simeq 0$



$\mathrm{Cov}(X, Y) < 0$

larger values of X correspond to lower values of Y

5 Correlation coefficient

Normalized measure of the correlation between two random variables.

Pearson's correlation coefficient

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$-1 \leq \rho_{XY} \leq 1$$

Eg: $Y = mX + c$

$$Var(Y) = Var(mX + c) = Var(mX) = m^2 Var(X)$$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

$$= E(X(mX+c)) - E(X)\cdot E(mX+c)$$

$$= E(mX^2 + cX) - E(X)\cdot[mE(X) + c]$$

$$= mE(X^2) + \cancel{cE(X)} - m(E(X))^2 - \cancel{cE(X)}$$

$$= m\underbrace{\left(E(X^2) - (E(X))^2\right)}$$

$$= m\,\mathrm{Var}(X)$$

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{m\,\mathrm{Var}(X)}{\sqrt{\mathrm{Var}(X)\cdot m^2\,\mathrm{Var}(X)}} = \frac{m\,\cancel{\mathrm{Var}(X)}}{m\,\cancel{\mathrm{Var}(X)}} = \boxed{1}$$

if two random variables are perfectly linearly related.