

VARIANCE

STAT SIG

DEFN

THE VARIANCE OF X , WITH $\mu = E[X]$ IS

$$\sigma^2 = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & X \text{ DISC} \\ \int x^2 f(x) dx & X \text{ CONT} \end{cases}$$

→ $\sigma_x = \sqrt{\sigma^2} = \sqrt{V[X]}$

THE STANDARD DEVIATION IS

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma_x = \sqrt{V[X]} = SD[X]$$

Thm

THE VARIANCE HAS THREE PROPERTIES:

① $\text{V}[x] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2$

② GIVEN CONSTANTS a, b : $\text{V}[aX + b] = a^2\text{V}[x]$

③ X_1, \dots, X_n INDEPENDENT, a_1, \dots, a_n CONSTANTS THEN

$$\text{V}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \text{V}[X_i]$$

EXAMPLE

$$X \sim \text{Binom}(n, p)$$

$$\mathbb{V}[X] = \sum_{x=0}^n (x - np)^2 \binom{n}{x} p^x (1-p)^{n-x} = \dots$$

AGAIN $X = \sum_{i=1}^n X_i$ X_i INO $\xrightarrow{\text{BERNOULLI}}$ RECALL $E[X_i] = p$

$$E[X_i] = 0 \cdot (1-p) + 1^2 \cdot p = p$$

$$\mathbb{V}[X_i] = E[X_i^2] - (E[X_i])^2 = p - p^2 = p(1-p)$$

$$\mathbb{V}[X] = \mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i] = \sum_{i=1}^n p(1-p) = np(1-p) \checkmark$$

Thm

X_1, X_2, \dots, X_n iid with $E[X_i] = \mu$ and $V[X_i] = \sigma^2$

① $E[\bar{X}_n] = \mu$ WHERE $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

② $V[\bar{X}_n] = \sigma^2/n$ SAMPLE MEAN

③ $E[S_n^2] = \sigma^2$ WHERE $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

SAMPLE
VARIANCE

DEFN

X with mean μ_X and variance σ_X^2

Y with mean μ_Y and variance σ_Y^2

The CORRELATION of X and Y is

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

The CORRELATION of X and Y is

$$\rho = \rho_{X,Y} = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Theorem

①

$$\text{Cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$$

②

$$-1 \leq \rho(x, y) \leq 1$$

③

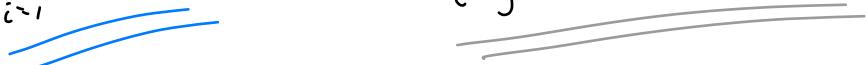
$$\text{IF } X, Y \text{ i.i.d THEN } \text{Cov}[x, y] = \rho = 0.$$

THE REVERSE IS NOT TRUE IN GENERAL.

THEM

$$\mathbb{V}[x+y] = \mathbb{V}[x] + \mathbb{V}[y] + 2\text{Cov}[x, y]$$

OR, IN GENERAL

$$\mathbb{V}\left[\sum_{i=1}^n a_i x_i\right] = \sum_{i=1}^n a_i^2 \mathbb{V}[x_i] + 2 \sum_{i < j} \sum a_i a_j \text{Cov}[x_i, x_j]$$


VARIANCE - COVARIANCE MATRIX

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

$$\mu = \mathbb{E}[X] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix}$$

$$\Sigma = \mathbb{V}[X] =$$

$$\begin{bmatrix} \mathbb{V}(X_1) & \text{cov}(X_1, X_2) & \cdots & \cdots & \text{cov}(X_1, X_k) \\ \text{cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \cdots & \text{cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{cov}(X_k, X_1) & \cdots & \cdots & \cdots & \mathbb{V}(X_k) \end{bmatrix}$$