Covaraince and Sampling Distributions

Learning Outcomes

- Covariance
- Statistics and Inference
- Sampling Distributions
- Central Limit Theorem

Covariance

Covariance

The covariance measures the average dependence between multiple random variables. Let $W = \binom{X}{Y}$ be a random vector. The variance of W is defined as

2x1

covariances of rv's

$$Var(W) = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

2x2

Variances of rv

Covariance

σ_{xy}

Let X_1 and X_2 be 2 random variables with mean μ_1 and μ_2, respectively. The covariance of X_1 and X_2 is defined as

$$\begin{aligned} Cov(X_1, X_2) &= E\{(X_1 - \mu_1)(X_2 - \mu_2)\} \\ &= E(X_1 X_2) - \mu_1 \mu_2 \end{aligned}$$

If X_1 and X_2 are independent random variables, then

$$Cov(X_1, X_2) = 0$$

$$f(x,y) = \begin{cases} xy & 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & \text{o.w.} \end{cases}$$

$$E(X) = \int_0^1 \int_0^1 x \, xy \, dx \, dy = \int_0^1 y \int_0^1 x^2 \, dx \, dy$$

$$\int_0^1 y \left[\frac{x^3}{3} \Big|_0^1 \right] dy$$

$$\int_0^1 \frac{1}{3} y \, dy$$

$$\frac{1}{3} \frac{y^2}{2} \Big|_0^1 = \frac{1}{6}$$

$$E(Y) = \frac{1}{6}$$

$$E(XY) = \int_0^1 \int_0^1 xy \, xy \, dx \, dy$$

$$\int_0^1 y^2 \int_0^1 x^2 \, dx \, dy$$

$$\frac{x^3}{3} \Big|_0^1$$

$$\int_0^1 \frac{1}{3} y^2 \, dy = \frac{1}{3} \frac{y^3}{3} \Big|_0^1 = \frac{1}{9}$$

$$\mathrm{Cov}(X,Y) = \frac{1}{9} - \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{9} - \frac{1}{36} = \frac{4-1}{36} = \frac{3}{36}$$

Correlation

The correlation of X_1 and X_2 is defined as

$$\rho = Cor(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}$$

MGF Property: Independence

Let X and Y be independent random variables. Let $Z = X + Y$, the MGF of Z is

$$M_Z(t) = M_X(t)M_Y(t)$$

$$Z = aX + bY$$

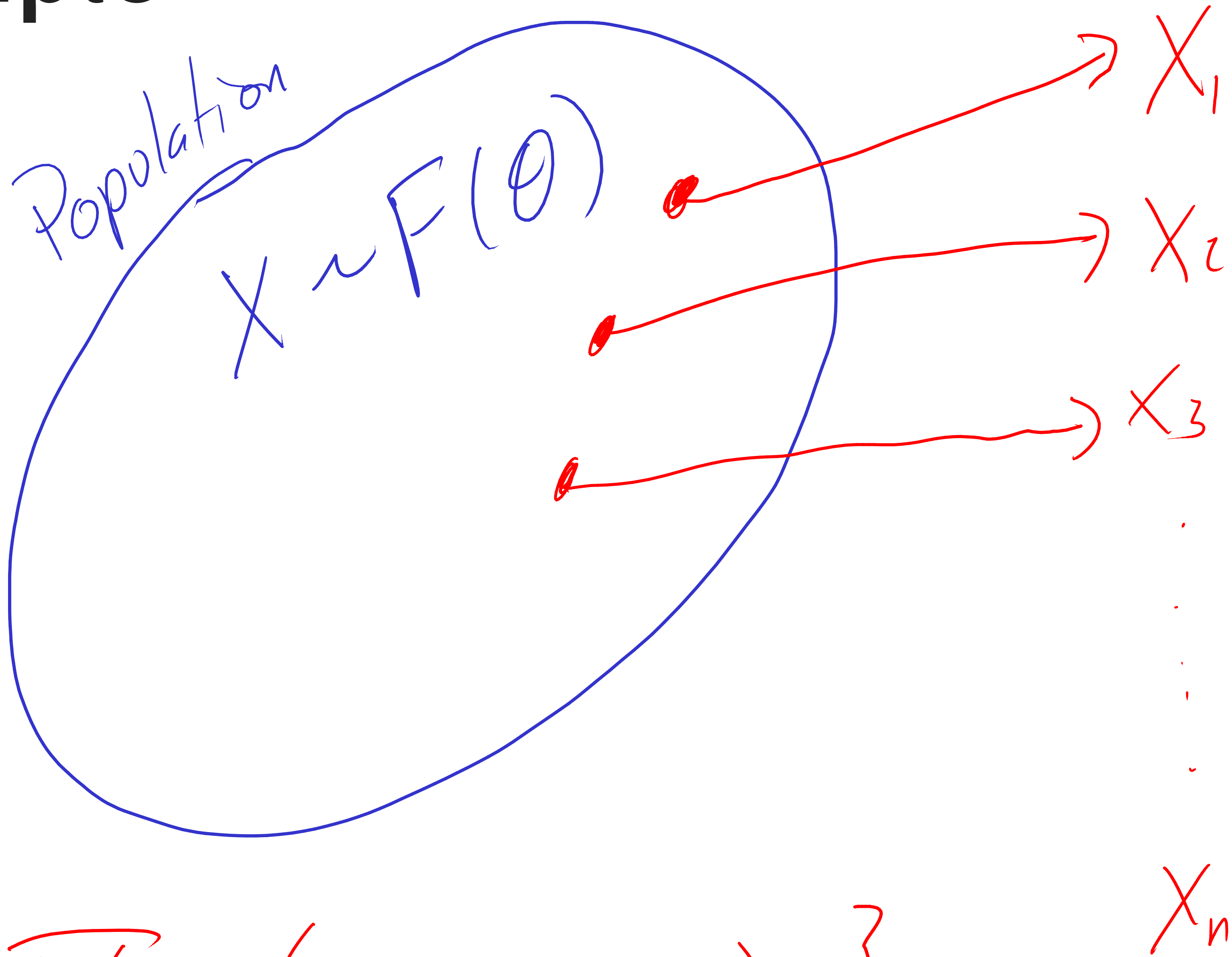
$$E(Z) = aE(X) + bE(Y)$$

$$Var(Z) = a^2 Var(X) + b^2 Var(Y)$$

Statistics and Inference

Sample

$X_i \sim F(\theta)$



$$\mathbb{X} = \{X_1, \ldots, X_n\}$$

Statistics

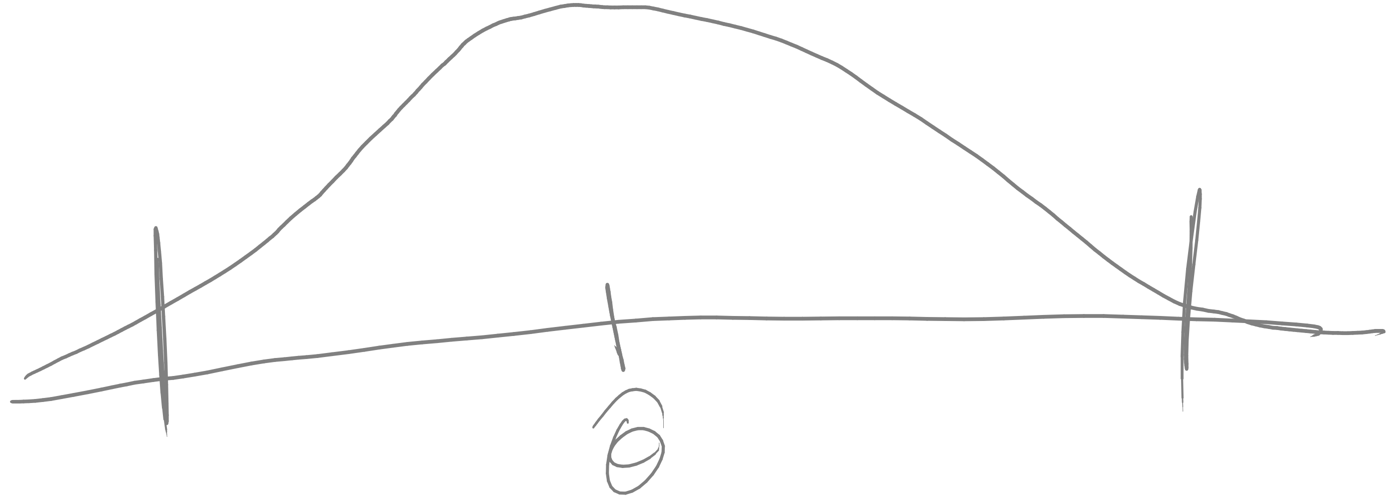
$\mathbb{X} = \{X_1, \ldots, X_n\}$

hat estimate

$\hat{\theta} = g(\mathbb{X})$

$\mathbb{X}$ $\mathbb{X}$ $\mathbb{X}$

θ



Inference

$\theta = 5$

Error

$\alpha = 0.01$

$X \sim F(\theta)$

$\hat{\theta} = 4.8 \pm 0.1$

$\alpha = 0.01$

$\hat{\theta} = 4 \pm 1.2$

Sampling Distributions

iid Random Variables

$$\underline{\mathbb{X}} = \{X_1, \ldots, X_n\}$$

independent

$$X_i \perp X_j$$

identical and independent

$$X_i \sim F(\theta)$$

Sampling Distributions

A sampling distribution is the distribution of a statistic. Many known statistics have a known distribution.

$$\hat{\theta} \sim G(\theta)$$

$\bar{X}$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$E(X_i) = \mu$

$Var(X_i) = \sigma^2$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

$$\sum_{i=1}^{5} E(X_i) = \underset{\mu}{E(X_1)} + \underset{\mu}{E(X_2)} + \underset{\mu}{E(X_3)} + \underset{\mu}{E(X_4)} + \underset{\mu}{E(X_5)}$$

5μ

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} Var\left(\sum X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

s^2

Sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$S^2 \sim \frac{(n-1)}{\sigma^2} \chi^2(n-1)$$

t-distribution

Let $Z \sim N(0,1)$, $W \sim \chi^2_\nu$, $Z \perp W$; therefore:

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu$$

t-tests

F-distribution

Let $W_1 \sim \chi^2_{\nu_1}$ $W_2 \sim \chi^2_{\nu_2}$, and $W_1 \perp W_2$; therefore:

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2}$$

Example

Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, show that $\bar{X} \sim N(\mu, \sigma^2/n)$.

Note: the MGF of X_i is $e^{\mu t + \frac{t^2\sigma^2}{2}}$.

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be identical and independent distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. We define

$$Y_n = \underbrace{\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)} \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then, the distribution of the function Y_n converges to a standard normal distribution function as $n \to \infty$.

$$Y_n \sim N(0,1)$$

Central Limit Theorem

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$n \to \infty$

Example

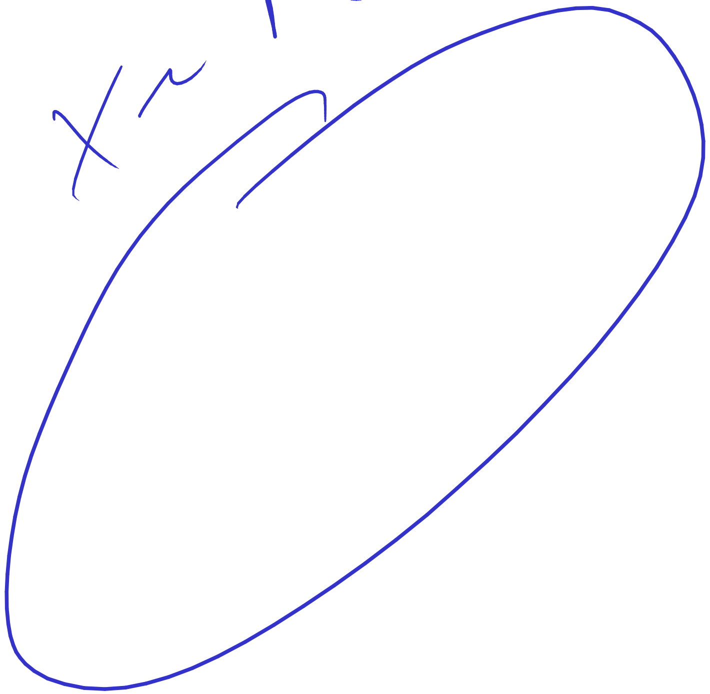
Let $X_1, \ldots, X_n \overset{iid}{\sim} \chi^2_p$, the MGF is $M(t) = (1-2t)^{-p/2}$. Find the distribution of $\bar{X}$.

$E(X) = p$ $\quad Var(X) = 2p = \sigma^2$

$n \to \infty$

$\bar{X} \sim N\left(p, \frac{2p}{n}\right)$

$X \sim F(\theta)$



$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Cauchy Dist

$\cancel{E(X)}$