

Expectation of Linear Function of Multiple RV's

- ▶ Linearity is well preserved in expectation.

$$E(a \cdot X + b \cdot Y + c) = a \cdot E(X) + b \cdot E(Y) + c$$

Expectation of Product of Multiple RV's

- ▶ Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

Expectation of Product of Multiple RV's

- ▶ Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

- ▶ But if X and Y are independent, then

$$\begin{aligned} E(XY) &= \int \int xyf(x, y)dxdy = \int \int xyf_X(x)f_Y(y)dxdy \\ &= \int xf_X(x)dx \int yf_Y(y)dy = E(X)E(Y) \end{aligned}$$

Expectation of Product of Multiple RV's

- ▶ Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

- ▶ But if X and Y are independent, then

$$\begin{aligned} E(XY) &= \int \int xyf(x, y)dx dy = \int \int xyf_X(x)f_Y(y)dx dy \\ &= \int xf_X(x)dx \int yf_Y(y)dy = E(X)E(Y) \end{aligned}$$

- ▶ And for independent RV's, in general

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Covariance

- When two random variables X and Y are not independent, it is often of interest to assess how strongly they are related to one another.
- A popular measurement to characterize the dependence of two rv's is called **correlation**. To calculate correlation of two rv's, we'll have calculate the **covariance** of the two rv's.
- The **covariance** between two rv's X and Y is

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) \cdot p(x, y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy & X, Y \text{ continuous} \end{cases}\end{aligned}$$

Short cut

- Proposition:

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

- What happens if we set $Y=X$?

Covariance and Variance

- ▶ As we can see, variance is a special case of covariance, where $X = Y$.

Covariance and Variance

- ▶ As we can see, variance is a special case of covariance, where $X = Y$.
- ▶ Variance of linear function of multiple RV's is given by

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \cdot \text{Cov}(X, Y)$$

Covariance and Variance

- ▶ As we can see, variance is a special case of covariance, where $X = Y$.
- ▶ Variance of linear function of multiple RV's is given by

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \cdot \text{Cov}(X, Y)$$

- ▶ This is a special case of

$$\begin{aligned} \text{Cov}(aX + bY, cZ + dW) = & ac \cdot \text{Cov}(X, Z) + ad \cdot \text{Cov}(X, W) \\ & + bc \cdot \text{Cov}(Y, Z) + db \cdot \text{Cov}(Y, W) \end{aligned}$$

Example

Ex. Suppose the joint distribution of X and Y are

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the covariance of X and Y?

$$f_X(x) = \int_y f(x, y) dy = \int_0^{1-x} 24xy dy = 12x(1-x)^2$$

$$f_Y(y) = 12y(1-y)^2$$

$$E(X) = \int_0^1 x \cdot 12x(1-x)^2 dx = \frac{2}{5} = E(Y)$$

$$E(XY) = \int \int_{x,y} xy f(x, y) dx dy = \int_0^1 \int_0^{1-y} 24x^2 y^2 dx dy = \frac{2}{15}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75}$$

Correlation

- The **correlation coefficient** of X and Y , denoted by $\text{Corr}(X, Y)$ or $\rho_{X,Y}$ is defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

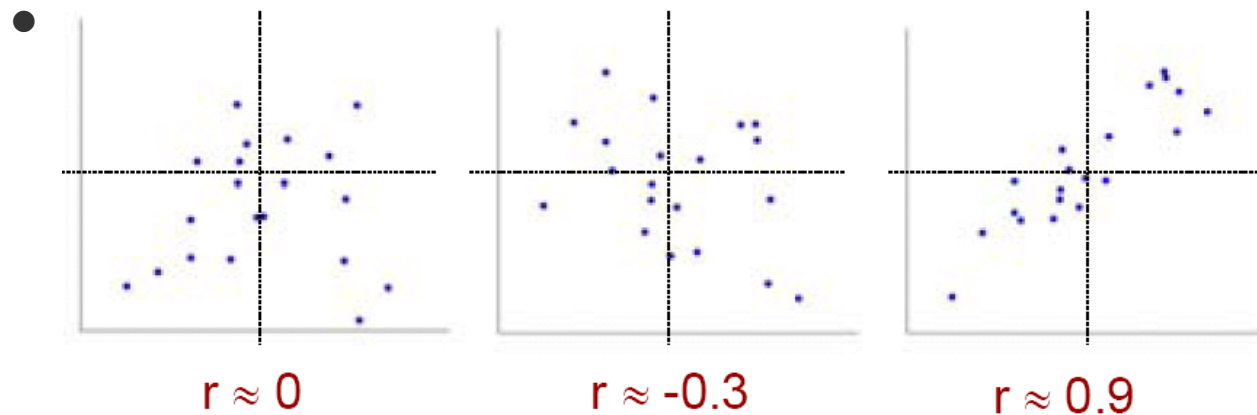
- Because of Cauchy-Schwarz inequality, we have

$$\text{Cov}^2(X, Y) \leq \text{Var}(X)\text{Var}(Y) \implies |\rho_{X,Y}| \leq 1$$

- The correlation coefficient $\rho_{X,Y}$ is **NOT** a completely general measure of the strength of a relationship. $\rho_{X,Y}$ is actually a measure of the degree of **linear** relationship between X and Y .

Remarks

- If X and Y are independent, then $\rho_{X,Y} = 0$ (why?). But $\rho_{X,Y} = 0$ does **NOT** imply independence.
- $\rho_{X,Y} = 1$ or -1 **iff** $Y = aX + b$ for some numbers a and b with $a \neq 0$.



Relationship Between Correlation and Independence

- ▶ Independence leads to uncorrelatedness.

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

Relationship Between Correlation and Independence

- ▶ Independence leads to uncorrelatedness.

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

- ▶ But not vice versa!

Relationship Between Correlation and Independence

- ▶ Independence leads to uncorrelatedness.

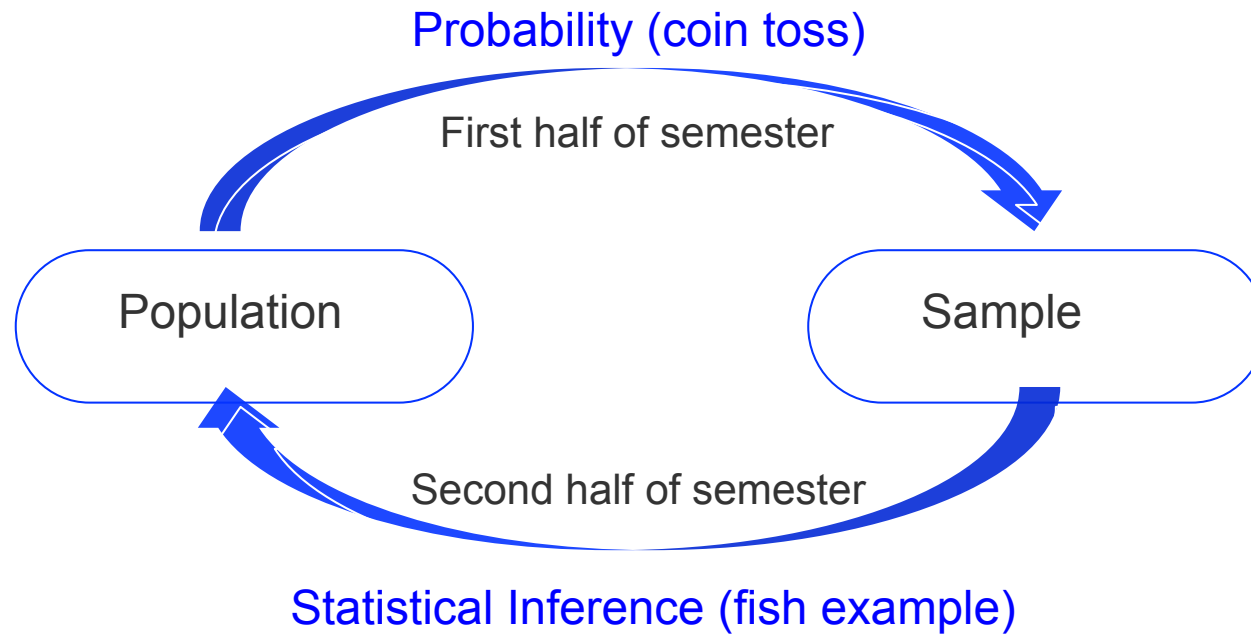
$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

- ▶ But not vice versa!
- ▶ We will talk about this more in regression.

Population and Sample

- ▶ We will start changing our discussion from probability to statistics, which means we need to think about samples and how they relate to the underlying population.
- ▶ Recall the relationship between population and sample (probability and inference) that we visualized in the first lecture.

Probability and Inference



RV or a Particular Number

- ▶ In the first chapter, we use lowercase letters to represent the sample, x_1, x_2, x_3, \dots . That means we have already observed the data and each of the letters can be replaced by a particular number.
- ▶ Before the data becoming available, there is uncertainty as to what value we will observe, so we view each observation as a RV, thus denoted by uppercase letter X_1, X_2, X_3, \dots .

Sample and Statistics

- ▶ A statistic is any quantity whose value can be calculated from sample data, such as Sample Mean and Sample Variance.
- ▶ Before obtaining data, a statistic is also a RV. The bulk of statistical inference is to find the distribution of the statistics, or the so-called *Sampling Distributions*.
- ▶ To make things easier, we often need to assume the observed data are *Simple Random Samples*, which means they are IID (Independently Identically Distributed).

Introduction to IID

- A sequence of random variables, X_1, X_2, \dots, X_n , is **independent and identically distributed (i.i.d.)** if each random variable has the same probability distribution as the others and all are **mutually independent**.
- In statistical analysis, we often assume the sampled data X_1, X_2, \dots, X_n , are i.i.d. from a common distribution $f(x)$. And usually, we end up analyzing a **linear combination** of the X_i 's, that is

$$Y = a_1X_1 + \dots + a_nX_n = \sum_{i=1}^n a_iX_i$$

Sample Mean***

- Let X_1, X_2, \dots, X_n , be an i.i.d. sequence of rv's from a distribution with mean value μ and standard deviation σ .
- Notice that the sample mean or the sample total ($T = X_1 + X_2 + \dots + X_n$) can also be viewed as a special case of linear combination of X_1, X_2, \dots, X_n . In the i.i.d. case,

$$E(T) = E(X_1) + E(X_2) + \dots + E(X_n) = n\mu$$

$$\text{Var}(T) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) = n\sigma^2$$

- It is also easy to verify that for sample mean,

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

$$\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma^2/n \implies \sigma_{\bar{X}} = \sigma/\sqrt{n}$$

Invariance of Normal RV under Linear Transformation

- ▶ When $X_1, X_2, X_3, X_4, \dots$ are normal random variables, then the linear combination of them

$$a_1X_1 + a_2X_2 + \dots + a_nX_n = \sum_{i=1}^n a_iX_i$$

is still a normal random variable.

- ▶ In particular, sample mean \bar{X} is still a random variables.
- ▶ Remark: 1. No IID assumption is necessary; 2. This property is for Normal only.

Sample Mean of IID Normal

- ▶ If X_1, X_2, \dots, X_n IID $\sim N(\mu, \sigma^2)$, then what is the distribution of \bar{X} ?

Sample Mean of IID Normal

- ▶ If X_1, X_2, \dots, X_n IID $\sim N(\mu, \sigma^2)$, then what is the distribution of \bar{X} ?

▶

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

Sample Mean of IID Normal

- ▶ If X_1, X_2, \dots, X_n IID $\sim N(\mu, \sigma^2)$, then what is the distribution of \bar{X} ?

▶

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

- ▶ But how can we derive the distribution of Sample Mean when the sample are not normal? We need to use Central Limit Theorem.

CLT

- Theorem:

The Central Limit Theorem (CLT)

Let X_1, X_2, \dots, X_n , be an i.i.d. sequence from a distribution with mean μ and variance σ^2 . Then if n is sufficiently large, the sample mean \bar{X} has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$; And the sample total has approximately a normal distribution with $\mu_T = n\mu$, $\sigma_T^2 = n\sigma^2$. The larger the value of n , the better the approximation.

- Rule of Thumb: if $n > 30$, the CLT can be used.