

Lecture 21:

Covariance, Total Expectation

Lecture 20 Summary

- $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ measures the expected deviation squared from the mean.
- $\sigma(X) = \sqrt{\text{Var}(X)}$
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- Two r.v.s X, Y are independent if for all a, b $\Pr[X=a, Y=b] = \Pr[X=a] \cdot \Pr[Y=b]$.
- If X, Y are independent then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
- If X, Y are independent, then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
- If X_1, \dots, X_n are pairwise indep. then $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$
- If $X \sim \text{Bin}(n, p)$ then $\text{Var}(X) = n \cdot p \cdot (1-p)$.

Covariance

If X and Y are r.v.s their co-variance is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

measures whether greater values of X mainly correspond with greater values of Y .

Facts: $\text{Cov}(X+a, Y+b) = \text{Cov}(X, Y)$

$$\text{Cov}(c \cdot X, d \cdot Y) = c \cdot d \cdot \text{Cov}(X, Y).$$

Covariance

If X and Y are r.v.s their co-variance is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

Fact: $\text{Cov}(X, Y) = E[XY] - (EX) \cdot (EY)$

Proof: let $\mu_x = EX$ $\mu_y = EY$

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y \\ &= E[XY] - \mu_x \mu_y = E[XY] - (EX)(EY).\end{aligned}$$

Covariance

If X and Y are r.v.s their co-variance is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

Fact: $\text{Cov}(X, Y) = E[XY] - (EX) \cdot (EY)$

Corollary: If X and Y are independent then

$$\text{Cov}(X, Y) = 0.$$

Recap

Theorem: If X and Y are independent r.v.s, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof:

$$\begin{cases} E[(X+Y)^2] = E[X^2] + 2E[XY] + E[Y^2] \\ (E[X+Y])^2 = (E[X] + E[Y])^2 \\ = (E[X])^2 + 2(E[X])(E[Y]) + (E[Y])^2 \end{cases}$$

$$\begin{aligned} \text{Var}(X+Y) &= E[X^2] - (E[X])^2 + 2E[XY] - 2(E[X])(E[Y]) \\ &\quad + E[Y^2] - (E[Y])^2 \\ &= \text{Var}(X) + 0 + \text{Var}(Y) \end{aligned}$$

↑
The only point where we used that X & Y are indep

Theorem: If X and Y are ~~independent~~ r.v.s, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$$

Proof:

$$\begin{cases} \mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] \\ (\mathbb{E}[X+Y])^2 = (\mathbb{E}X + \mathbb{E}Y)^2 \\ = (\mathbb{E}X)^2 + 2(\mathbb{E}X)(\mathbb{E}Y) + (\mathbb{E}Y)^2 \end{cases}$$

$$\begin{aligned} \text{Var}(X+Y) &= \mathbb{E}[X^2] - (\mathbb{E}X)^2 + 2\mathbb{E}[XY] - 2(\mathbb{E}X)(\mathbb{E}Y) \\ &\quad + \mathbb{E}[Y^2] - (\mathbb{E}Y)^2 \\ &= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) \end{aligned}$$

Variance of a Poisson RV

$$X \sim \text{Pois}(\lambda) \quad \Pr[X=i] = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

We saw: $E[X] = \lambda$

$$\text{Var}(X) = ?$$

Recall:

$$\text{Bin}\left(n, \frac{\lambda}{n}\right) \xrightarrow{n \rightarrow \infty} \text{Pois}(\lambda)$$

$$Y \sim \text{Bin}\left(n, \frac{\lambda}{n}\right)$$

$$E[Y] = n \cdot \frac{\lambda}{n} = \lambda = E[X].$$

$$\text{Var}(Y) = n \cdot \frac{\lambda}{n} \cdot \left(1 - \frac{\lambda}{n}\right) \xrightarrow{n \rightarrow \infty} \lambda = \text{Var}[X].$$

$$E[X^2] = ?$$

Normalizing a Random Variable

If X is a r.v. with mean μ and stdev σ

Then $\tilde{X} = \frac{X - \mu}{\sigma}$ is the normalized version of X .

$$E[\tilde{X}] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{E[X - \mu]}{\sigma} = 0$$

$$\begin{aligned} \text{Var}(\tilde{X}) &= E[\tilde{X}^2] = E\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = \frac{E[(X - \mu)^2]}{\sigma^2} \\ &= \frac{\text{Var}(X)}{\text{Var}(X)} = 1. \end{aligned}$$

Correlation

Def'n: The correlation of X, Y is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Claim: $\text{Corr}(X, Y) = E[\tilde{X} \cdot \tilde{Y}]$

Proof:

$$\tilde{X} = \frac{X - \mu(X)}{\sigma(X)}$$

$$\tilde{Y} = \frac{Y - \mu(Y)}{\sigma(Y)}$$

$$\begin{aligned} E[\tilde{X} \cdot \tilde{Y}] &= E\left[\left(\frac{X - \mu(X)}{\sigma(X)}\right) \cdot \left(\frac{Y - \mu(Y)}{\sigma(Y)}\right)\right] = \frac{E[(X - \mu(X))(Y - \mu(Y))]}{\sigma(X)\sigma(Y)} \\ &= \text{Corr}(X, Y). \end{aligned}$$

Def'n: The Correlation of X, Y is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Claim: $\text{Corr}(X, Y) = E[\tilde{x} \cdot \tilde{y}]$

Theorem: $-1 \leq \text{Corr}(X, Y) \leq 1$ and equality holds
iff $Y = \alpha X + \beta$.

Proof: $\text{Corr}(X, Y) = E[\tilde{x} \cdot \tilde{y}]$

$$(i) \quad 0 \leq E[(\tilde{x} - \tilde{y})^2] = E[\tilde{x}^2] + E[\tilde{y}^2] - 2E[\tilde{x} \cdot \tilde{y}]$$

$$\therefore 2E[\tilde{x} \cdot \tilde{y}] \leq E[\tilde{x}^2] + E[\tilde{y}^2] = 2. \quad \text{Hence, } E[\tilde{x} \cdot \tilde{y}] \leq 1.$$

$$(ii) \quad 0 \leq E[(\tilde{x} + \tilde{y})^2] = E[\tilde{x}^2] + E[\tilde{y}^2] + 2E[\tilde{x} \cdot \tilde{y}]$$

$$\therefore 2E[\tilde{x} \cdot \tilde{y}] \geq -2. \quad \text{Hence } E[\tilde{x} \cdot \tilde{y}] \geq -1.$$

$$\text{Corr}(X, Y) = 1 \iff E[(\tilde{x} - \tilde{y})^2] = 0 \iff \tilde{x} = \tilde{y} \iff \frac{x - \mu_x}{\sigma_x} = \frac{y - \mu_y}{\sigma_y}$$

$$\text{Corr}(X, Y) = -1 \iff E[(\tilde{x} + \tilde{y})^2] = 0 \iff \tilde{x} = -\tilde{y} \iff \frac{x - \mu_x}{\sigma_x} = -\frac{y - \mu_y}{\sigma_y}$$

Example

X, Y are two r.v.s whose joint dist is

		X		
		-1	0	+1
Y	+1		$\frac{1}{5}$	
	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
	-1		$\frac{1}{5}$	

Q: Are X and Y independent?

Q: What's the $\text{Cov}(X, Y)$?

$$E[XY] =$$

$$E[X] =$$

$$E[Y] =$$

Independent \Rightarrow Uncorrelated

but

Independent $\not\Leftarrow$ Uncorrelated

Example

X, Y are two r.v.s whose joint dist is

		X		
		-1	0	$+1$
Y	$+1$		$1/7$	$1/7$
	0	$1/7$	$1/7$	$1/7$
	-1	$1/7$	$1/7$	

Q: Are X and Y independent?

Q: What's the $\text{Cov}(X, Y)$?

$$E[XY] =$$

$$E[X] =$$

$$E[Y] =$$

Conditioning on Events

Let X be a r.v. over Ω

and let A be an event.

$$\Pr[X=x | A] = \frac{\Pr[(X=x) \cap A]}{\Pr[A]}$$

If A_1, \dots, A_n are disjoint events that partition Ω

Then,

$$\begin{aligned} \Pr[X=x] &= \Pr[(X=x) \cap A_1] + \dots + \Pr[(X=x) \cap A_n] \\ &= \Pr[A_1] \cdot \Pr[X=x | A_1] + \dots + \Pr[A_n] \cdot \Pr[X=x | A_n]. \end{aligned}$$

Law of Total Expectation

Theorem: Let X be a r.v. over Ω

and A_1, \dots, A_n be disjoint events that partition Ω

$$\text{then } E[X] = \sum_{i=1}^n \Pr[A_i] \cdot E[X | A_i]$$

Proof:

$$\begin{aligned} E[X] &= \sum_x x \cdot \Pr[X=x] \\ &= \sum_x x \cdot \sum_{i=1}^n \Pr[A_i] \cdot \Pr[X=x | A_i] \\ &= \sum_{i=1}^n \Pr[A_i] \cdot \sum_x x \cdot \Pr[X=x | A_i] \\ &= \sum_{i=1}^n \Pr[A_i] \cdot E[X | A_i] \end{aligned}$$

Example: Computing the Expectation of a Geometric R.V.

$$X \sim \text{Geo}(p)$$

We saw using
algebraic tricks
that $\mathbb{E}X = \frac{1}{p}$

Random Experiment: toss a coin
with heads prob. p until
you get H.

Let's see another proof.

$$X \sim \text{Geo}(p)$$

$$\Omega = \{H, TH, TTH, \dots\}$$

$$\Pr[X=i] = (1-p)^{i-1} \cdot p \quad \text{for } i=1, 2, \dots$$

Idea: Partition the sample space to two parts:

A = first toss is heads, \bar{A} = first toss is tails.

$$\Pr[A] = p \quad \Pr[\bar{A}] = 1-p.$$

$$X|A = 1 \quad X|\bar{A} = \left\{ \begin{array}{l} 2, \text{ w.p. } p \\ 3, \text{ w.p. } p \cdot (1-p) \\ \vdots \end{array} \right\} = 1+Y$$

where
 $Y \sim \text{Geo}(p)$

$$\mathbb{E}[X] = \Pr[A] \cdot \mathbb{E}[X|A] + \Pr[\bar{A}] \cdot \mathbb{E}[X|\bar{A}]$$

$$= p \cdot 1 + (1-p) \cdot \mathbb{E}[1+Y] = p \cdot 1 + (1-p) \cdot (1 + \mathbb{E}[Y]) = 1 + (1-p) \cdot \mathbb{E}[X]$$

Hence $p \cdot \mathbb{E}[X] = 1$ and we get $\mathbb{E}[X] = \frac{1}{p}$.

Computing the Variance of a Geometric RV

$$X \sim \text{Geo}(p) \quad \Omega = \{H, TH, TTH, \dots\}$$

$$\Pr[X=i] = (1-p)^{i-1} \cdot p \quad \text{for } i=1, 2, \dots$$

Idea: Partition the sample space to two parts:

A = first toss is heads, \bar{A} = first toss is tails.

$$\Pr[A] = p \quad \Pr[\bar{A}] = 1-p.$$

$$X|A = 1 \quad X|\bar{A} = 1+Y \quad \text{where } Y \sim \text{Geo}(p)$$

$$\mathbb{E}[X^2] = \Pr[A] \cdot \mathbb{E}[X^2|A] + \Pr[\bar{A}] \cdot \mathbb{E}[X^2|\bar{A}]$$

$$= p \cdot 1 + (1-p) \cdot \mathbb{E}[(1+Y)^2]$$

$$= p + (1-p) \cdot (1 + 2\mathbb{E}[Y] + \mathbb{E}[Y^2])$$

$$= 1 + \frac{(1-p) \cdot 2}{p} + (1-p) \mathbb{E}[X^2]$$

$$\therefore \mathbb{E}[X^2] \cdot p = \frac{2-p}{p} \quad \therefore \mathbb{E}[X^2] = \frac{2-p}{p^2} \quad \therefore \text{Var}[X] = \frac{1-p}{p^2}$$

Law of Iterated Expectation

Total Expectation: Let X be a r.v. over Ω

and A_1, \dots, A_n be disjoint events that partition Ω

$$\text{then } E[X] = \sum_{i=1}^n \Pr[A_i] \cdot E[X | A_i]$$

Special Case: let X, Y be r.v.s over Ω , then

$$E[X] = \sum_{y \in \text{range}(Y)} \Pr[Y=y] \cdot E[X | Y=y]$$

Law of Iterated Expectation - Example

You pick $N \sim \text{Pois}(\lambda)$.

Then you flip a coin w. heads prob p , N times.

X - the number of heads you got.

Calculate: $E[X]$

Solution:
$$E[X] = \sum_{n=1}^{\infty} P[N=n] \cdot E[X | N=n]$$
$$= \sum_{n=1}^{\infty} P[N=n] \cdot p \cdot n = E[p \cdot N] = p \cdot E[N] = p \cdot \lambda.$$