MATH 180A (Lecture A00)

mathweb.ucsd.edu/~ynemish/teaching/180a

Today: Expectation and variance of sums. Covariance and correlation

Next: ASV 9.1

Week 10:

- Homework 7 due Sunday, March 19

Linearity of expectation

<u>Thm</u>. Let $X_1, \dots, X_n$ be random variables defined on the same probability space. Let $g_1, g_2, \dots, g_n$ be functions of one variable. Then

$$E(g_1(X_1) + g_2(X_2) + \cdots + g_n(X_n)) = E(g_1(X_1)) + E(g_2(X_2)) + \cdots + E(g_n(X_n))$$

In particular, $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

Important: independence does not matter

Expectation of a sum = sum of expectations

ALWAYS!

Linearity of expectation

Example (Binomial distribution).

Let $X_1, \dots, X_n$ be independent random variables, $X_i \sim Ber(p)$

$$S_n = X_1 + \dots + X_n, \quad S_n \sim Bin(n,p).$$

$$E(S_n) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + \dots + E(X_n) = \overbrace{p + \dots + p}^{n \text{ times}} = np$$

Example Adam must pass both written test and road test for his driver's license. He passes written test with probability $\frac{4}{10}$, independently of other tests. For the road test, the probability of success is $\frac{7}{10}$. What is the total expected number of tests Adam must take before earning his license?

Denote $X = \#$ written tests before he passes

$Y = \#$ road tests before he passes

$E(X+Y) - ?$

$X \sim Geom\left(\frac{4}{10}\right), \ Y \sim Geom\left(\frac{7}{10}\right)$

$E(X) = \frac{10}{4}, \ E(Y) = \frac{10}{7}$

$E(X+Y) = \frac{10}{4} + \frac{10}{7}$

Expectation of a product of independent random variables

Thm. Let $X_1, \dots, X_n$ be independent random variables.

Let $g_1, g_2, \dots, g_n$ be functions of one variable.

Then

$$E(g_1(X_1)\, g_2(X_2) \cdots g_n(X_n))$$
$$= E(g_1(X_1))\, E(g_2(X_2)) \cdots E(g_n(X_n))$$

Corollary If $X_1, \dots, X_n$ are independent, then

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)$$

Variance of a sum of independent random variables

Example

Binomial: $X_1, \dots, X_n$ independent identically distributed (iid)

$X_i \sim Ber(p)$, $Var(X_i) = p(1-p)$, $S_n = X_1 + \dots + X_n$, $S_n \sim Bin(n,p)$

$$Var(S_n) = Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + \dots + Var(X_n) = np(1-p)$$

Sample mean: $X_1, \dots, X_n$ independent identically distributed (iid)

$E(X_i) = \mu$, $Var(X_i) = \sigma^2$

$$E\left(\frac{X_1 + \dots + X_n}{n}\right) = \mu, \quad Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + \dots + X_n)$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Covariance

Suppose that we have a random variable X.

- E(X) – mean value, average of a large number of independent realizations
- Var(X) – variance, fluctuations of X, how far the realizations are spread around the mean

Covariance "describes" strength and type of dependence between two random variables.

Def. Let X and Y be random variables defined on the same probability space. The covariance of X and Y is given by
$$Cov(X,Y) = E\Big((X-E(X))(Y-E(Y))\Big)$$

Computations: $Cov(X,Y) = E(XY) - E(X)E(Y)$

Covariance

Example Let X, Y be discrete random variables with the joint PMF $P(X=k, Y=\ell)$ given by the table

$k \backslash \ell$	0	1	2	
-1	0.1	0	0.1	0.2
0	0.3	0.2	0	0.5
1	0	0.2	0.1	0.3
	0.4	0.4	0.2	

$$E(X) = (-1)\cdot 0.2 + 0\cdot 0.5 + 1\cdot 0.3 = 0.1$$

$$E(Y) = 0\cdot 0.4 + 1\cdot 0.4 + 2\cdot 0.2 = 0.8$$

$$E(XY) = (-2)\cdot 0.1 + 1\cdot 0.2 + 2\cdot 0.1 = 0.2$$

$$\mathrm{Cov}(X,Y) = 0.2 - 0.1\cdot 0.8 = 0.12$$

Some heuristics

By definition, $Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$

- $(X-E(X))(Y-E(Y))$ is positive if

  $(X-E(X))$ and $(Y-E(Y))$ have the same sign
- $(X-E(X))(Y-E(Y))$ is negative if

  $(X-E(X))$ and $(Y-E(Y))$ have opposite signs

Thus,

- $Cov(X,Y)>0$ means that on average $X-E(X)$ and $Y-E(Y)$ have the same sign, positively correlated
- $Cov(X,Y)<0$ means that on average $X-E(X)$ and $Y-E(Y)$ have opposite signs, negatively correlated
- If $Cov(X,Y)=0$, we say that X and Y are uncorrelated

Example

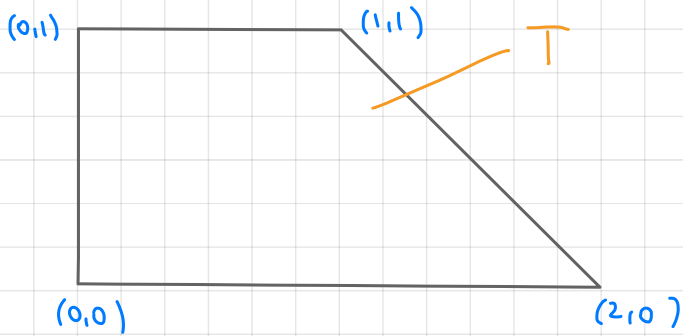
Let (X,Y) be a uniformly distributed random point on the trapezoid with vertices $(0,0),(2,0),(1,1),(0,1)$

Is $\mathrm{Cov}(X,Y)$

(a) positive

(b) negative

?



Joint density: $f(x,y) = \frac{2}{3}$ for $(x,y) \in T$

$$E(X) = \iint_T x \ \frac{2}{3} \ dx\,dy = \frac{7}{9}, \quad E(Y) = \iint_T y \cdot \frac{2}{3}\, dx\,dy = \frac{4}{9}$$

$$E(XY) = \iint_T xy \cdot \frac{2}{3}\, dx\,dy = \frac{11}{36}$$

$$\mathrm{Cov}(X,Y) = \frac{11}{36} - \frac{7}{9} \cdot \frac{4}{9} = -\frac{13}{324}$$

Uncorrelated vs Independent

- X and Y are independent $\Rightarrow$ $\mathrm{Cov}(X,Y)=0$

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

- $\mathrm{Cov}(X,Y)=0 \not\Rightarrow$ X and Y are independent

Example of random variables X, Y that are <u>not</u> independent, but $\mathrm{Cov}(X,Y)=0$

Let $X \sim N(0,1)$, $Y = X^2$. Then

$$E(X)=0, \quad E(X^2)=1, \quad E(X^3)=0$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\ &= E(X^3) - E(X)E(X^2) = 0\end{aligned}$$

Variance of a sum. Properties of covariance

Thm Let $X_1, \dots, X_n$ be random variables with finite variances

Then $\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \operatorname{Cov}(X_i, X_j)$

For example, $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$

Properties of covariance:

- $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$
- $\operatorname{Cov}(aX+b, Y) = a\operatorname{Cov}(X,Y)$
- $\operatorname{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j \operatorname{Cov}(X_i, Y_j)$

bilinear