Lecture 21: Covariance and Correlation

Definition: $\mathrm{Cov}(X,Y) = E((X-EX)(Y-EX)) = E(XY) - E(X)E(Y)$

$X-EX$: X relative to its mean

$Y-EX$: Y relative to its mean

X and Y are any two random variables ~~of~~ on the same space

Since $E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$ (by Linearity)

(Imagine drawing a random sample, suppose we had a lot of i.i.d. pairs X, Y. The pairs are i.i.d. but with each pair X_i, Y_i, they have some joint distribution. They may not be independent. If they are independent: $\mathrm{Cov}(X,Y) = E(X-EX)E(Y-EX)$.)

Properties:

(1) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

(2) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ (it's symmetric)

(3) $\mathrm{Cov}(X,c) = 0$, if c is a constant.

bi-linearity

(4) $\mathrm{Cov}(cX,Y) = c\,\mathrm{Cov}(X,Y)$

(5) $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(XY) + \mathrm{Cov}(XZ)$

$E(XY) - E(X)E(Y)$

$E(X(Y+Z)) - E(X)E(Y+Z)$

$E(XY) + E(XZ) - E(X)E(Y) + E(X)E(Z)$

$E(XY) - E(X)E(Y) + E(XZ) - E(X)E(Z)$

$\mathrm{Cov}(XY) + \mathrm{Cov}(XZ)$

bi-linearity – Imagine treating one coordinate as fixed and working with the other coordinate, it looks like linearity.

(6) $\mathrm{Cov}(X+Y, Z+W) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(X,W) + \mathrm{Cov}(Y,Z) + \mathrm{Cov}(Y,W)$

linear combination of random variables

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i,j} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$$

(7) Variance of sum

$$\underbrace{Var(X_1+X_2)}_{\text{Covariance of } (X_1+X_2) \text{ with itself}} = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

⇒ If the Covariance is 0, then the variance of the sum is the sum of the variances.

⇒ One case where that's true is if they are independent.

$Var(X_1+X_2) = Var(X_1) + Var(X_2)$,

since $Cov(X_1, X_2) = 0$ if X_1, X_2 independent.

Note: $Cov(X_1, X_2)$ can be 0 even when X_1, X_2 not independent.

$$Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$$

Theorem: If X, Y are independent, then they are uncorrelated, i.e., $Cov(X, Y) = 0$

Converse is false. That is, if $Cov(X, Y) = 0$, X and Y may or may not be independent.

Example $Z \sim N(0,1)$, $X = Z$, $Y = Z^2$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$
$$= E(Z^3) - E(Z)E(Z^2) = 0 \quad \left[\text{Odd moments of standard normal are 0}\right]$$

but very dependent: Y is a function X, and Y also determines magnitude of X.

Correlation

$$Corr(X, Y) = \frac{Cov(X, Y)}{SD(X)\,SD(Y)} = Cov\left(\frac{X - EX}{SD(X)}, \frac{Y - E(Y)}{SD(Y)}\right)$$

dimensionless quantity

Standardize both X and Y

Correlation means standardize them first, then take the covariance.

<u>Theorem</u>: $-1 \leq Corr(X,Y) \leq 1$ [form of Cauchy-Schwarz]

<u>proof</u>: WLOG (without Loss of Generality) assume X, Y are standardized. Let $Corr(X,Y) = \rho$.

$\to$ mean 0, variance 1

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$
$$= 1+1+2\rho$$
$$= 2+2\rho \text{ (since, } X, Y \text{ are standardized)}$$

$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$
$$= 2-2\rho$$

$$Var(X+Y) \geq 0 \Rightarrow 2+2\rho \geq 0 \Rightarrow 1+\rho \geq 0 \Rightarrow \rho \geq -1$$
$$Var(X-Y) \geq 0 \Rightarrow 2-2\rho \geq 0 \Rightarrow 1-\rho \geq 0 \Rightarrow \rho \leq 1$$

or, $\boxed{-1 \leq \rho \leq 1}$

In general, it is easier to work with covariances than correlations, but correlations are more intuitive and standardized with everything between -1 and 1.

<u>Example</u> Covariance in a multinomial.

$(X_1, X_2, \ldots, X_k) \sim Mult(n, \vec{p})$. Find $Cov(X_i, X_j)$ for all i, j.

$\to$ k diff. categories
X_j is # people or objects in jth category
n objects/people

If $i=j$, $cov(X_i, X_j) = Var(X_i) = np_i(1-p_i)$ [Success to be being in category $i \to$ binomial]

Now, let $i \neq j$.

Find $cov(X_1, X_2)$: $\rightsquigarrow$ if we knew that there were more people in the first category, then there's fewer left over who could be in the second category.

Let $c = cov(X_1, X_2)$.

$Var(X_1+X_2) = np_1(1-p_1) + np_2(1-p_2) + 2c$

$n(p_1+p_2)(1-(p_1+p_2))$ [follows from the lumping property] [says merge the first two categories together into one bigger category. Then, it's still binomial. Now, we're defining success to being a member of category 1 or category 2.]

or, $n(p_1+p_2)(1-p_1-p_2) = np_1(1-p_1) + np_2(1-p_2) + 2c$

$\Rightarrow \cancel{np_1} - \cancel{np_1^2} - np_1p_2 + \cancel{np_2} - np_1p_2 - \cancel{np_2^2} = \cancel{np_1} - \cancel{np_1^2} + \cancel{np_2} - \cancel{np_2^2} + 2c$

$\Rightarrow -\cancel{2}np_1p_2 = \cancel{2}c$

$\Rightarrow cov(X_1, X_2) = -np_1p_2$.

In general, $Cov(X_i, X_j) = -np_ip_j$, for $i \neq j$

Example $X \sim Bin(n,p)$, Write as $X = X_1 + \dots + X_n$, X_j are i.i.d. $Bern(p)$.

$Var(X_j) = E(X_j^2) - (EX_j)^2$
$= p - p^2 = p\underbrace{(1-p)}_{q} = pq$

$Var(X) = npq$ since $Cov(X_i, X_j) = 0$ for $i \neq j$

adding up n of independent bernoulli trials

Variance of a binomial $\to$ n times the variance of one of these Bernoullis.

We can think of X_j's – they are Bernoulli's, but they are also indicator random variables. It's the indicator of success on the jth trial.

Let I_A be indicator r.v. of event A.

$I_A^2 = I_A$, $I_A^3 = I_A$ (Since 0, 1)

$I_A I_B = I_{A \cap B}$ (1 if and only if both are 1)

Example $X \sim HGeom(w, b, n)$ → Hypergeometric jar of w white balls, b black balls, we take a sample of size n & we want the distribution of the number of white balls in the sample

$X = X_1 + \dots + X_n$,

we can ~~X_j~~ interpret this as drawing balls from the jar one at a time without replacement. We would get a binomial if we did it with replacement, but hypergeometric would be without replacement.

$$X_j = \begin{cases} 1, & \text{if } j^{th} \text{ ball is white} \\ 0, & \text{otherwise} \end{cases}$$

The ~~problem~~ reason that it's more difficult than previous problem is that $X_1, \dots, X_n$ are dependent indicator random variables, because it's without replacement.

$$Var(X) = \underbrace{n\,Var(X_1)}_{\text{Symmetry}} + 2\underbrace{\binom{n}{2}}_{\text{Covariance terms}} Cov(X_1, X_2)$$

(again by symmetry)

Using Bernoulli

$$Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$$

$E(X_1X_2)$ → expected value of indicator $\boxed{\text{fundamental bridge}}$ that's the prob. that the first two balls are both white $\left(\frac{w}{w+b}\right)\left(\frac{w-1}{w+b-1}\right)$ given that the first ball is white

$E(X_1)E(X_2)$ → prob. of the first ball is white times prob. of the second ball is white $\left(\frac{w}{w+b}\right)\left(\frac{w}{w+b}\right)$

fundamental bridge

$E(X_1, X_2)$
$I_A I_B = I_{A \cap B}$