ECON 340
Economic Research Methods

Div Bhagia

Lecture 11: Independence & Correlation

Random Variables

- *Random variables* take different values under different scenarios.
- Examples: outcome from a coin toss or a die roll, or number of times your wireless network fails before a deadline, etc.
- The likelihood of these scenarios is summarized by the probability distribution.
- Random variables can be *discrete* or *continuous*

Two Random Variables

The *joint probability distribution* of two discrete random variables is the probability that the random variables simultaneously take on certain values.

$$f(x, y) = Pr(X = x, Y = y)$$

	Rain $(X = 1)$	No Rain $(X = 0)$	Total
60-min commute $(Y = 60)$	0.3	0.2	
30-min commute $(Y = 30)$	0.1	0.4	
Total			

Marginal Distribution

The *marginal probability distribution* of a random variable Y is just another name for its probability distribution.

$$f(y) = Pr(Y = y) = \sum_x Pr(X = x, Y = y)$$

Conditional Distribution

The distribution of a random variable Y conditional on another random variable X taking on a specific value is called the conditional distribution of Y given X.

$$f(y|x) = Pr(Y = y|X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)} = \frac{f(x, y)}{f(x)}$$

Commute Times

	Rain $(X = 1)$	No Rain $(X = 0)$	Total
60-min commute ($Y = 60$)	0.3	0.2	
30-min commute ($Y = 30$)	0.1	0.4	
Total			

Conditional Expectation

The *conditional expectation* of Y given X is the mean of the conditional distribution of Y given X.

$$E(Y|X=x) = \sum_y yPr(Y=y|X=x) = \sum_y y \cdot f(y|x)$$

Calculate $E(Y|X=1)$ and $E(Y|X=0)$ in the last example. Comparing these tells us how X affects Y.

Can define conditional variance similarly.

Independence

Two random variables X and Y are independently distributed, or independent, if knowing the value of one of the variables provides no information about the other.

$$Pr(Y = y|X = x) = Pr(Y = y)$$

Example: Two consecutive coin tosses.

Note: We can equivalently say that X and Y are independent if $E(Y|X) = E(Y)$.

Covariance and Correlation

Covariance is a measure of the extent to which two random variables move together.

Let X and Y be a pair of random variables, then the *covariance* of X and Y is given by:

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$

The *correlation* between X and Y is given by:

$$\rho_{XY} = corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad \text{where } -1 \leqslant \rho \leqslant 1$$

Uncorrelated vs Independence

If X and Y are independent, then they are also uncorrelated.

$$E(Y|X) = E(Y) \rightarrow \rho_{XY} = 0$$

However, it is not necessarily true that if X and Y are uncorrelated, then they are also independent.

Sums of Random Variables

X and Y is a pair of random variables, then

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

If X and Y are independent:

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

Portfolio Diversification

You are now contemplating between two stocks with the same average return and spread.

$$\mu_X = \mu_Y \qquad \sigma_X^2 = \sigma_Y^2$$

Should you pick any one stock at random or invest equally in both?

What's next?

- Problem Set 3 is now posted on Canvas (due next week). You can attempt Questions 1 and 2.
- Thursday: Start with Sampling and Estimation.