Lecture Summary

Definition (Expectation)

The **expected value** or **mean** or **first moment** of X is defined to be

$$E(X) = \sum_x x P_x(x)$$

assuming that the sum is well-defined.

- We can think of the expectation as the average of a very large number of independent draws from the distribution (IID draws).
- The fact that $E(X) = \sum_{i=1}^{n} X_i$ is actually a very important theorem we will discuss later.
- Example: $X \sim Bernoulli(p)$. $E(X)$?
- Example: Let X be the number of heads in 3 tosses. $E(X)$?

Expectation of a function of a random variable

Sometimes we are interested in the expectation of a function of a random variable $Y = r(X)$. One way to find the expectation of this random variable:

- Find its pmf $P_Y(y)$
- Compute $\sum_y yP_Y(y)$

e.g. Assume X^2 is a discrete random variable with possible values $\{-3, -1, 0, 1, 3\}$ with probabilities $\{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$. Let $Y = X^2$. What is the expectation of Y?

Law Of The Unconscious Statistician

A simple way to compute the expectation of Y in the example above, or any function of random variables, is the law of unconscious statistician (LOTUS).

Theorem

Let $Y = r(X)$. *Then*

$$E(Y) = E(r(x)) = \sum_{\text{all } x} r(x) P_x(x)$$

if the mean exists.

LOTUS for functions of two or more random variables.

Theorem

Let $Z = r(X, Y)$. *Then*

$$E(Z) = E(r(X, Y)) = \sum_{\text{all } (x,y)} r(x, y) P_{X,Y}(x, y)$$

if the mean exists.

Theorem

Let $Z = r(\mathbf{X} = X_1, \ldots, X_n)$. *Then*

$$E(Z) = E(r(\mathbf{X})) = \sum_{\text{all } (\mathbf{x})} r(\mathbf{x}) P(\mathbf{x})$$

if the mean exists.

Properties of Expectations

Properties of Expectation

- $E(a) = a$
- $E(aX) = aE(X)$
- $E(aX + b) = aE(X) + b$
- But $E(g(X)) \neq g(E(X))$ in most cases!
- Let $X_1, \ldots, X_n$ be a set of *independent* random variables. Then

$$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i]$$

Linearity of Expectation

Linearity of expectation is the property that the expected value of the sum of random variables is equal to the sum of their individual expected values, regardless of whether they are independent.

Theorem

Let $X_1, \ldots, X_n$ *be a set of random variables. Then*

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

Prove it for the case of two discrete variables.

Example

- If $X_1, X_2, \dots X_n$ are i.i.d. $Bernoulli(p)$ random variables then $Y = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$

$$E(Y) = E[X_1] + \cdots + E[X_n] = np$$

Variance of a random variable

Sometimes we are also interested in quantifying how far from the mean

Definition (Variance)

The **variance** of X is defined to be

$$Var(X) = E[(X - E(X))^2]$$

assuming that the sum is well-defined.

Standard deviation

$$\sigma_x = \sqrt{Var(X)}$$

- Let $X \sim Bernoulli(p). Var(X) = ?$
- Let X be the number of heads in three tosses. $Var(X) = ?$

Properties of Variances

- $Var(a) = 0$
- $Var(aX + b) = a^2 Var(X) + b$
- Linearity of variances only holds for independent random variables.

Theorem

Let $X_1, \ldots, X_n$ be a set of independent random variables. Then

$$Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n]$$

Prove it for the case of two discrete variables.

Covariance

Covariance measures how much two r.vs. vary together (i.e., are larger than usual at the same time).

Covariance

The **covariance** of two variables X and Y is defined to be

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

If X, Y are independent, then $Cov(X,Y) = 0$

Correlation

Covariance without dimensions

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Inequalities

Schwartz inequality

$$[E(XY)]^2 \leq E(X^2)E(Y^2)]$$

Cauchy - Schwartz inequality

$$[Cov(X,Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

$$-1 \leq \rho(X,Y) \leq 1$$

Practice Exercises

Section	Exercises
4.1	9, 10
4.2	8
4.3	2,4,6
4.6	5,7