Expected value and Variance

Anastasiia Kim

February 10, 2020

Main Concepts Related to Random Variables

Starting with a probabilistic model of an experiment:

- A random variable is a real-valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain 'averages' of interest, such as the mean and the variance
- A random variable can be conditioned on an event or on another random variable
- There is a notion of independence of a random variable from an event or from another random variable

Concepts Related to Discrete Random Variables

Starting with a probabilistic model of an experiment:

- A discrete random variable is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values
- A discrete random variable has an associated probability mass function (PMF) which gives the probability of each numerical value that the random variable can take
- A function of a discrete random variable defines another discrete random variable, whose PMF can be obtained from the PMF of the original random variable

Calculation of the PMF $p(x)$ of a r.v. X for each possible value x of X

$p(x) = P(X = x)$ and $\sum_{\text{all x}} p(x) = 1$

- 1. Collect all the possible outcomes that give rise to the event X = x
- 2. Add their probabilities to obtain $p(x)$

CDF

The cumulative distribution function (CDF) $F(X)$ of a r. v. X is another method to describe the distribution of random variables

- $F(X) = \sum_{i=1}^{x} P(X = i) = \sum_{i=1}^{x} p(i)$
- For all $a \leq b$ $P(a < X \leq b) = F(b) - F(a)$
- $P(X < x) = P(X \leq x) - P(X = x) = F(x) - p(x)$

Mean and Variance of a Discrete Random Variable

- Two numbers are often used to summarize a probability distribution for a random variable X
- The mean is a measure of the center or middle of the probability distribution
- The variance is a measure of the dispersion, or variability in the distribution
- These two measures do not uniquely identify a probability distribution! Two different distributions can have the same mean and variance

Expectation of a random variable

If X is a discrete random variable having a probability mass function p(x), then the expectation (mean), or the expected value, of X, denoted by $E(X)$, is defined by

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

Properties of expectation:

- $E(X)$ is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it

The expected value from the roll of a die is $7/2$. Note that the expected value is not necessarily a value that can be 'expected' to turn up.

Example

If the weather is good (which happens with probability 0.6). Alice walks the 2 miles to class at a speed of $V = 5$ miles per hour, and otherwise rides her motorcycle at a speed of $V = 30$ miles per hour. What is the mean of the time T to get to class?

- Derive the PMF of T
- Calculate the expected value

Properties of the first moment E(X)

- For any constants a and b, $E(aX + b) = aE(X) + b$
- The nth moment is given by

$$E(X^n) = \sum_x x^n p(x)$$

- Let $g(X)$ be a function of X. The expected value of the random variable $g(X)$ is given by

$$E(g(X)) = \sum_x g(x)p(x)$$

Optimizing the choice between several candidate decisions

A contestant on a quiz show is presented with two questions, which he is to attempt to answer in some order he chooses. If he decides to try question 1 first, then he will be allowed to go on to question 2 only if his answer to question 1 is correct. If his initial answer is incorrect, he is not allowed to answer the other question. Question 1 will be answered correctly with probability 0.8, and the contestant will receive $100, while question 2 will be answered correctly with probability 0.5, and he will receive $200 as a prize. Which question should he attempt to answer first so as to maximize his expected winnings? Assume that the events that he knows the answer to question 1 and 2 are independent events.

- The answer is not obvious because there is a tradeoff: attempting first the more valuable but also more difficult question 2 carries the risk of never getting a chance to attempt the easier question 1

Variance

The variance of a random variable X is a measure of dispersion or scatter in the possible values for X

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

- For any constants a and b, $Var(aX + b) = a^2 Var(X)$
- The standard deviation is $\sqrt{Var(X)}$