4.4 Probability Distributions

A *random variable* is a real number whose value is based on the random outcome of an experiment, which might be merely an observation or measurement. The statistics and probabilities associated with random variables follow according to a distribution depending on item under observation and the nature of the experiment. Random variables can be either discrete or continuous. Discrete random variables takes on discrete, or countable, values. A continuous random variable can be a real number.

4.4.1 Discrete Distributions

Here are some examples of discrete random variables.

- The number of time that heads might appear in 100 coin flips.
- The number of flips needed to get 1 head.
- The number of customers entering a store in 1 hour.
- The number of products with a defect coming out a machine each day.

The **probability mass** defines the set of probabilities for the random variable taking each possible value.

The expected value of a discrete random variable represents the long-run average value (mean) of that random variable.

$$E[X] = \sum_i x_i P(X = x_i)$$

$0 \cdot 0.5 + 1 \cdot 0.5$

$= 0.5$

The variance of a discrete random variable is

$$Var[X] = E\left[(X - E([X])^2\right] = E\left[X^2\right] - E[X]^2.$$

$0.5 - 0.25 = 0.25$

Bernoulli random variable

A trial that takes one of two outcomes, such as {(success – failure), (head – tail), (true – false), etc. }, has a *Bernoulli distribution*.

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p, \quad (0 \leq p \leq 1) \\ p(1) &= P\{X = 1\} = p \end{aligned}$$

Binomial random variable

In n independent trials of a Bernoulli experiment, the variable X representing the number of successes has a *Binomial distribution* with parameters (n, p), denoted $X \sim Bin(n, p)$.

We need to introduce a new operator to describe the probability distribution. The construct $\binom{n}{k}$ is called n choose k and is the number of different combinations that can be chosen from