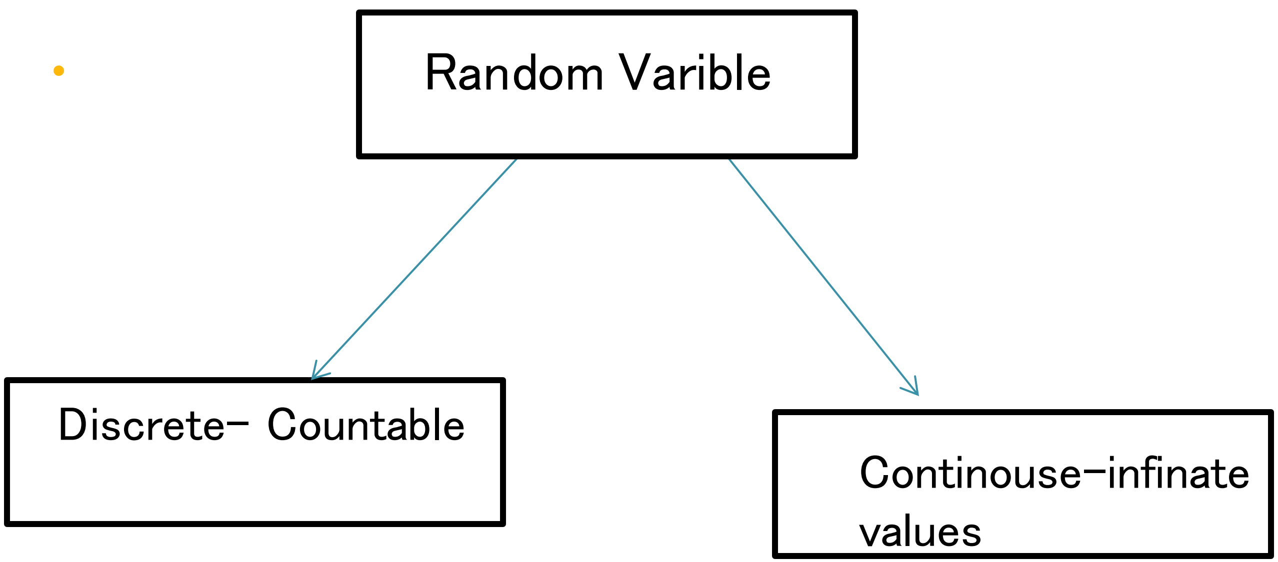
Probability and Distributions

Random variable.

- A **random variable** is a variable that assumes numerical values associated with the random outcome of an experiment.
- Only one Numeric point is Associated with it.



Some important terminology

- The PMF is one way to describe the distribution of a discrete random variable.
- The cumulative distribution function (CDF) of a random variable is another method to describe the distribution of random variables.
- The advantage of the CDF is that it can be defined for any kind of random variable (discrete, continuous, and mixed).

Random Variables

–

- **Discrete random variables**
  - Number of sales
  - Number of calls
  - Shares of stock
  - People in line
  - Mistakes per page

- **Continuous random variables**
  - Length
  - Depth
  - Volume
  - Time
  - Weight

Discrete Distributions

- Relative frequency distributions for "counting" experiments.

- Bernoulli Distribution → Yes-No responses.
- Binomial Distribution → Sums of Bernoulli responses
- Negative Binomial → Number of trials to k^{th} event
- Poisson Distribution → Points in given space
- Geometric Distribution → Number of trials until first *success*
- Multinomial Distribution → Multiple possible outcomes for each trial

Binomial Distribution

- The experiment consists of **n identical trials** (simple experiments).
- Each trial results in one of **two outcomes** (success or failure)
- The random variable y is the number of successes observed during n trials.

Binomial Distribution. . .

- Consider the following scenarios:

  – The number of heads/tails in a sequence of coin flips

  – Vote counts for two different candidates in an election

  – The number of male/female employees in a company

  – The number of accounts that are in compliance or not in compliance with an accounting procedure

  – The number of successful sales calls

  – The number of defective products in a production run

  – The number of days in a month your company's computer network experiences a problem

Canonical Framework. . .

- There is a set of assumptions which, if valid, would lead to a binomial distribution.
- These are: ·
- A set of n experiments or trials are conducted.
- Each trial could result in either a success or a failure.
- The probability p of success is the same for all trials.
- The outcomes of different trials are independent.
- We are interested in the total number of successes in these n trials

Binom

- Under the above assumptions, let X be the total number of successes.
- Then, X is called a binomial random variable, and the probability distribution of X is called the **binomial distribution.**

- X has a binomial distribution with parameters

$$n = 10 \text{ and } p = 0.25.$$

Problem

- What is the probability for the student to get no answer correct?

$$
\begin{aligned}
P(X=0) &= \frac{10!}{0!(10-0)!}(0.25)^0(1-0.25)^{10-0} \\
&= (0.75)^{10} \\
&= 0.0563
\end{aligned}
$$

Binomial Mean and Variance. . .

It can be shown that

$$\mu = E(X) = np$$

and

$$\sigma^2 = V(X) = np(1-p)\,.$$

For the previous example, we have

- $E(X) = 10 \cdot 0.25 = 2.5.$
- $V(X) = 10 \cdot (0.25) \cdot (1 - 0.25) = 1.875.$

Poisson Distribution. . .

- The Poisson distribution is another family of distributions that arises in a great number of business situations.

- It usually is applicable in situations where random "events" occur at a certain rate over a period of *time*.

Consider the following scenarios

— The hourly number of customers arriving at a bank
— The daily number of accidents on a particular stretch of highway
— The hourly number of accesses to a particular web server
— The daily number of emergency calls in Dallas
— The number of typos in a book
— The monthly number of employees who had an absence in a large company
— Monthly demands for a particular product

Canonical Framework. . .

- let λ be the rate at which events occur, t be the length of a time interval, and X be the total number of events in that time interval. Then, X is called a Poisson random variable and the probability distribution of X is called the Poisson distribution.

Poisson Probability–Mass Function. . .

Let X be a Poisson random variable. Then, its probability-mass function is:

$$P(X = x) = e^{-\mu}\frac{\mu^x}{x!} \qquad (2)$$

for $x = 0, 1, 2, \ldots$.

The value of μ is the *parameter* of the distribution. For a given time interval of interest, in an application, μ can be specified as λ times the length of that interval.

Example: Typos

The number of typographical errors in a "big" textbook is Poisson distributed with a mean of 1.5 per 100 pages.

Uniform Distribution. . .

- The uniform distribution is the simplest example of a continuous probability distribution.
- A random variable X is said to be uniformly distributed if its density function is given by:

$$f(x) = \frac{1}{b-a}$$

for $-\infty < a \leq x \leq b < \infty.$

- The values a and b are the parameters of the uniform distribution. It can be shown that.

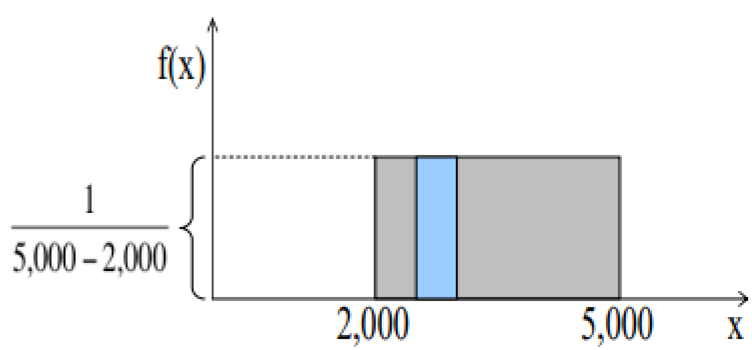
$$E(X) = \frac{a+b}{2} \qquad \text{and} \qquad V(X) = \frac{(b-a)^2}{12}.$$

- Example: Gasoline Sales Suppose the amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons. What is the probability that daily sales will fall between 2,500 gallons and 3,000 gallons?

$$P(2500 < X \leq 3000) = \frac{1}{5000 - 2000}(3000 - 2500)$$

$$= 0.1667 .$$

Visually, we have



Normal Distribution. . .

- The normal distribution is the most important distribution in statistics, since it arises naturally in numerous applications.
- The key reason is that large sums of (small) random variables often turn out to be normally distributed.

Normal Distribution Details

A random variable X is said to have the normal distribution with parameters μ and σ if its density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \qquad (6)$$

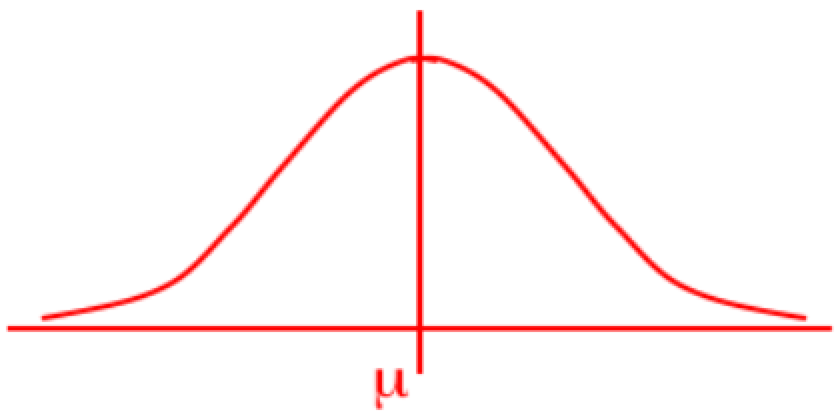
for $-\infty < x < \infty$.

It can be shown that

$$E(X) = \mu \qquad \text{and} \qquad V(X) = \sigma^2 .$$

Thus, the normal distribution is characterized by a mean μ and a standard deviation σ.

- A typical normal density curve looks like this:



- Thank You