

# Discrete Probability Distributions

Manideep Bangaru

24/11/2020

A **Probability Discrete Distribution** for a discrete variable is a mutually exclusive list of all the possible numerical outcomes along with the probability of occurrence of each outcome

## Expected value of a discrete variable

Expected value ( $\mu$ ) of a discrete variable,

$$\mu = E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

## Variance and Standard Deviation of a discrete variable

Variance of a discrete variable,

$$\sigma^2 = \sum_{i=1}^n [x_i - E(X)]^2 P(X = x_i)$$

Standard Deviation of a discrete variable,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^n [x_i - E(X)]^2 P(X = x_i)}$$

## Applying the concepts

```
# Distribution A
A_x = c(0,1,2,3,4)
P_x = c(0.50,0.20,0.15,0.10,0.05)

# Distribution B
B_y = c(0,1,2,3,4)
P_y = c(0.05,0.10,0.15,0.20,0.55)

# a. Compute the expected value for each distribution
# b. Compute the standard deviation for each distribution
# c. Compare the results of distributions A and B

ExStats <- function(x,Px){
  E_x <- sum(x*Px)
  Var_x <- sum((x - E_x)^2)*Px)
  Std_x <- Var_x ^ 0.5

  print(paste0('Expectation = ',toString(E_x)))
  print(paste0('Variance = ',Var_x))
  print(paste0('Standard Deviation = ',Std_x))
}

ExStats(A_x,P_x)
```

```
## [1] "Expectation = 1"
## [1] "Variance = 1.5"
## [1] "Standard Deviation = 1.22474487139159"
```

```
ExStats(B_y,P_y)
```

```
## [1] "Expectation = 3.2"
## [1] "Variance = 1.572"
## [1] "Standard Deviation = 1.2537942414926"
```

Standard deviation of Distribution B is greater than A

## Binomial Distribution

A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes (the prefix **“bi”** means two, or twice). For example, a coin toss has only two possible outcomes: heads or tails and taking a test could have two possible outcomes: pass or fail.

### Binomial distribution must meet the following criteria

- **The number of observations or trials is fixed.** In other words, you can only figure out the probability of something happening if you do it a certain number of times. This is common sense—if you toss a coin once, your probability of getting a tails is 50%. If you toss a coin a 20 times, your probability of getting a tails is very, very close to 100%

- **Each observation or trial is independent.** In other words, none of your trials have an effect on the probability of the next trial
- **The probability of success** (tails, heads, fail or pass) is exactly the same from one trial to another
- The probability of an observation being classified as the event of interest,  $\pi$ , is constant from observation to observation. Thus, the probability of an observation being classified as not being the event of interest,  $1 - \pi$ , is constant over all observations

Binomial Distribution,

$$P(X = x | n, \pi) = \frac{n!}{x!(n-x)!} \pi^x \pi^{n-x}$$

Mean of Binomial Distribution,

$$\mu = E(X) = n\pi$$

Standard deviation of Binomial Distribution,

$$\sigma = \sqrt{\sigma^2} = \sqrt{n\pi(1 - \pi)}$$

## Poisson Distribution

A poisson distribution helps to predict the probability of certain events from happening on when you know how often the event has occurred.

It gives us the probability of a given number of events happening in a fixed interval of time.

### Practical uses of Poisson Distribution

A textbook store rents an average of 200 books every Saturday night. Using this data, you can predict the probability that more books will sell (perhaps 300 or 400) on the following Saturday nights. Another example is the number of diners in a certain restaurant every day. If the average number of diners for seven days is 500, you can predict the probability of a certain day having more customers.

Poisson Distribution,

$$P(X = x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

## Poisson Distribution Vs Binomial

- If your question has an average probability of an event happening per unit (i.e. per unit of time, cycle, event) and you want to find probability of a certain number of events happening in a period of time (or number of events), then use the **Poisson Distribution**
- If you are given an exact probability and you want to find the probability of the event happening a certain number out times out of x (i.e. 10 times out of 100, or 99 times out of 1000), use the **Binomial Distribution**