Chapter 5 - Discrete Probability Distributions:

Vocabulary:

1. **Random Variable: A random variable is a variable (typically represented by x) that has a value for each outcome of a procedure, that is determined by chance.**

2. **Probability Distribution: A probability distribution is a description that gives the probability for each value of the random variable. It is often expressed in the format of a graph, table, or formula.**

   Eg: Probability distribution for rolling a die.

X	P(X)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

3. **Discrete Random Variable: A discrete random variable has either a finite number of values or a countable number of values, where "countable" refers to the fact that there might be infinitely many values, but they can be associated with a counting process so that the number of values is 0 or 1 or 2 or 3, etc.**

   Eg: Number of people in the class, Number of eggs.

4. **Continuous Random Variable: A continuous random variable has infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions.**

   Eg: Height, Weight, Temperature

Requirements for a Probability Distribution:

1. ΣP(X) = 1 where x assumes all possible values. (The sum of all probabilities must be 1, but values such as 0.999 or 1.001 are acceptable because they result from rounding errors.)

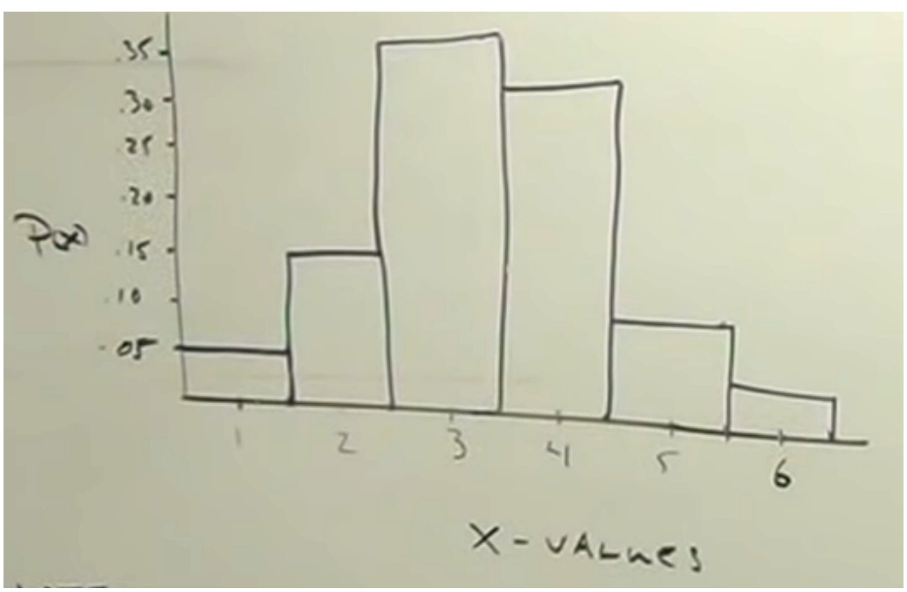
2. $0 \leq P(X) \leq 1$ for every individual value of x. (That is, each probability value must be between 0 and 1 inclusive.)

Histogram From Probability Distribution:

Eg:

1. **Weighted Die:**

X	P(X)	X * P(X)	X^2	X^2*P(X)
1	0.05	0.05	1	0.05
2	0.15	0.30	4	0.60
3	0.35	1.05	9	3.15
4	0.30	1.20	16	4.8
5	0.10	0.50	25	2.5
6	0.05	0.30	36	1.8
		Σ [X.P(X)]= 3.4		Σ [X^2*P(X)] = 12.9



Mean, Variance, SD :

Formula 5-1	$\mu = \Sigma[x \cdot P(x)]$	Mean for a probability distribution
Formula 5-2	$\sigma^2 = \Sigma[(x - \mu)^2 \cdot P(x)]$	Variance for a probability distribution (easier to understand)
Formula 5-3	$\sigma^2 = \Sigma[x^2 \cdot P(x)] - \mu^2$	Variance for a probability distribution (easier computations)
Formula 5-4	$\sigma = \sqrt{\Sigma[x^2 \cdot P(x)] - \mu^2}$	Standard deviation for a probability distribution

Mean:

$\mu = \frac{\Sigma X.f}{N}$ = $\Sigma \frac{x\,.\,f}{N}$ = **Σ [x. $\frac{f}{N}$] = Σ [X.P(X)] is the mean expected value.**

Finding the mean for weighted dice: (Table) μ = 3.4

Variance:

$\sigma^2 = \Sigma\ [X^2.P(X)] - \mu^2$

= 12.9 – 3.4^2

= 12.9 – 11.56

σ^2 = 1.34

Standard Deviation:

σ $= \sqrt{Variance}$

σ = $\sqrt{1.34}$ **= 1.1575**

Usual & Unusual Values:

Values are unusual if they lie outside of :

Maximum Usual Value = $\mu + 2\sigma$

Minimum Usual Value = $\mu - 2\sigma$

$2\sigma - \mu + 2\sigma$

Finding usual and unusual values for weighted dice:

Maximum Usual Value = $\mu + 2\sigma$ **= 3.4 + 2*1.1575 = 5.715**

Minimum Usual Value = $\mu - 2\sigma$ **= 3.4 - 2*1.1575 = -1.085**

If P(A) ≤ 0.5**, "A" is considered to be unusual.**

Eg:

1. **Find If a coin flip 1000 times: find P of Exactly 501 Heads =**

   P of Exactly 501 Heads = 0.0252 since it is less than 0.5 so it is unusual.

   P of 501 or more Heads = 0.487 since it is less than 0.5 so it is unusual.

Binomial Probability Distributions:

A probability that has only two outcomes. Success or Failure.

A binomial probability distribution results from a procedure that meets all the following requirements:

1. The procedure has a fixed number of trials.

2. The trials must be independent. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)

3. Each trial must have all outcomes classified into two categories (commonly referred to as success and failure).

4. The probability of success remains the same in all trials.

Notation for Binomial Probability Distributions

S and F (success and failure) denote the two possible categories of all outcomes.

$P(S) = p$	(p = probability of a success)
$P(F) = 1 - p = q$	(q = probability of a failure)
n	denotes the fixed number of trials.
x	denotes a specific number of successes in n trials, so x can be any whole number between 0 and n, inclusive.
p	denotes the probability of *success* in *one* of the n trials.
q	denotes the probability of *failure* in *one* of the n trials.
$P(x)$	denotes the probability of getting exactly x successes among the n trials.

Caution - When using a binomial probability distribution, always be sure that x and p both refer to the same category being called a success.

$$P(x) = \frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

where
n = number of trials
x = number of successes among n trials
p = probability of success in any one trial
q = probability of failure in any one trial ($q = 1 - p$)

Eg:

1. **Find If a coin flip 1000 times: find P of Exactly 501 Heads =**

   n = 1000
   success = Getting a head
   p = 0.50
   q = 0.50
   x = 501
   p + q = 1
   p = 1 - q or q = 1 - p

   P of Exactly 501 Heads = 0.0252 since it is less than 0.5 so it is unusual.

   $$p(501) \;=\; \frac{1000!}{499! * 501!} \;*\; 0.50^{501} * 0.50^{499}$$

   =

   P of 501 or more Heads = 0.487 since it is less than 0.5 so it is unusual.

2. **The probability of rolling a 4 is 30%, the die is rolled 10 times. Find the prob of rolling exactly eight 4's.**

   n = 10
   success = Rolling a 4
   failure = Rolling 1,2,3,5,6
   p = 0.30
   q = 0.70
   x = 8

   $$p(8) \;=\; \frac{10!}{2! * 8!} \;*\; 0.30^{8} * 0.70^{2}$$

   = 45*0.00006561*0.49

   = 0.0014467005

- Prob of rolling At most eight 4's. (Same data above)

  P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8) (Addition Rule)

  None +

  x = 1

  $$p(1) \;=\; \frac{1!}{9!*1!} \;*\; 0.30^{1} * 0.70^{9}$$

3. You draw 7 cards without replacement and you win the game if you get at least 5 hearts.

   n = 7
   success = Drawing a Heart
   failure = Not Drawing a Heart
   p = 13/52 = 0.20
   q = 0.80
   x = 5

   Here they are asking at least 5 hearts meaning we need to find P (5, 6, 7)

   $$p(5) \;=\; \frac{7!}{2!*5!} \;*\; 0.25^{5} * 0.75^{2}$$

   = 21 * 0.0009765625 * 0.50

   = 0.01025390625

   P (Exactly 4 hearts) = Exact 4 hearts

   P (At most 3 hearts) = 0,1,2,3

   P (More than 2 hearts) = 3,4,5,6,7

   P (Less than 6 hearts) = 0,1,2,3,4,5

Mean, Variance, and Standard Deviation for the Binomial Distribution:

1. **Mean $\mu = np$**

   μ = mean of the population
   n = The number of successes you expect to occur from your procedure.
   p = probability of success in any one trial

2. $\sigma^2 = npq$ **((Variance))**

   n = The number of successes you expect to occur from your procedure.
   p = probability of success in any one trial
   q = probability of failure in any one trial

3. $\sigma = \sqrt{npq}$ **(SD)**

   n = The number of successes you expect to occur from your procedure.
   p = probability of success in any one trial
   q = probability of failure in any one trial

Eg:

1. **A certain population is made up of 80% Mexican Americans, we want to select a jury of 12 people. Find the mean, sd, and variance.**

   Success: Randomly selecting a Mexican american

   Failures: Selecting anyone else

   n = 12 p = 0.80 q = 0.20

   μ = np

   = 12 * 0.8

   = 9.6

2. $\sigma = \sqrt{npq} = \sqrt{12 * 0.80 * 0.20}$ **= 1.38564064606**

3. σ^2 **= 1.92**

Usual & Unusual Values:

Values are unusual if they lie outside of :

Maximum Usual Value = ɯ + 2**σ**

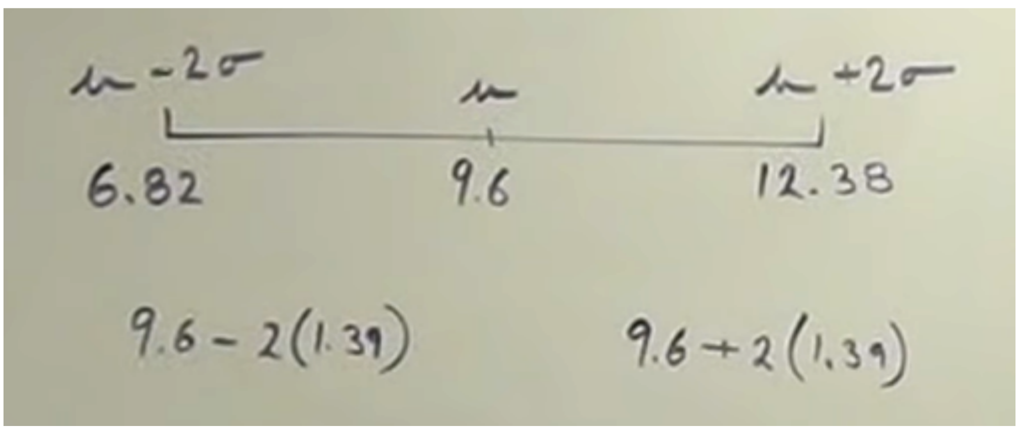
Minimum Usual Value = ɯ - 2**σ**

2**σ** - ɯ + 2**σ**

Finding usual and unusual values for Mexican American population:

Maximum Usual Value = ɯ + 2**σ** = **9.6 + (2 * 1.39) = 12.38**

Minimum Usual Value = ɯ - 2**σ** = **9.6 - (2 * 1.39) = 6.82**



4. **Over a period of time, 870 people were selected for juries from 80% of the Mexican American population. Find mean, SD and Find the usual range.**

   Success: Randomly selecting a Mexican american

   Failures: Selecting anyone else

   n = 870   p = 0.80   q = 0.20

   μ = np

   = 870 * 0.80 = 696

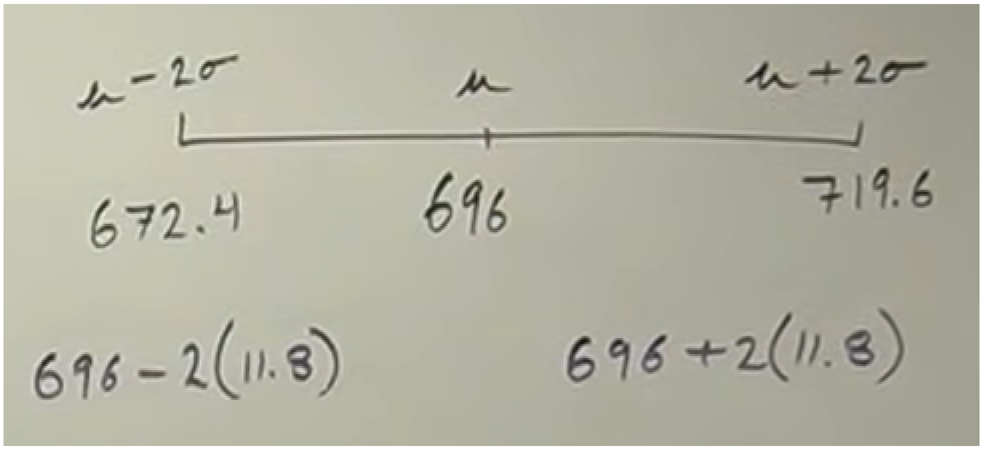
5. $\sigma = \sqrt{npq} = \sqrt{870 * 0.80 * 0.20}$ = **11.798**

6. σ^2 = **139.2**

Finding usual and unusual values for Mexican American population:

Maximum Usual Value = μ + 2σ = 696 + (2 * 11.798) = 719.596

Minimum Usual Value = μ - 2σ = 696 - (2 * 11.798) = 672.404



Only 275 Mexican American were selected for juries = VERY UNUSUAL

Poisson Probability Distributions:

The Poisson distribution is a discrete probability distribution that applies to occurrences of some event over a specified interval. The random variable x is the number of occurrences of the event in an interval. The interval can be time, distance, area, volume, or some similar unit. The probability of the event occurring x times over an interval is given by

$P(x) = \frac{\mu^x * e^{-\mu}}{x!}$ **Where e = 2.71828**

Requirements for the Poisson Distribution:

1. The random variable x is the number of occurrences of an event over some interval.

2. The occurrences must be random.

3. The occurrences must be independent of each other.

4. The occurrences must be uniformly distributed over the interval being used.

Parameters of the Poisson Distribution:

1. **The mean is** μ
2. **The standard deviation is** $\sigma = \sqrt{\mu}$

A Poisson distribution differs from a binomial distribution in these fundamental ways:

1. The binomial distribution is affected by the sample size n and the probability p, whereas the Poisson distribution is affected only by the mean μ

2. In a binomial distribution, the possible values of the random variable x are 0, 1,. . . . . . , n, but a Poisson distribution has possible x values of 0, 1, 2,. . . . . , with no upper limit.

Poisson Distribution as an Approximation to the Binomial Distribution:

The Poisson distribution is sometimes used to approximate the binomial distribution when n is large and p is small. One rule of thumb is to use such an approximation when the following two requirements are both satisfied.

Requirements for Using the Poisson Distribution as an Approximation to the Binomial:

1. **n = >= 100**
2. **np <= 10**

If both requirements are satisfied and we want to use the Poisson distribution as an approximation to the binomial distribution, we need a value for μ **. That value can be calculated by using Formula** μ **= np**

Eg:

In the Illinois Pick 3 game, you pay to select a sequence of three digits, such as 729. If you play this game once every day, find the probability of winning exactly once in 365 days.

Because the time interval is 365 days, n = 365. Because there is one winning set of numbers among 1000 that are possible (from 000 to 999), p = 1/1000. With n = 365 and p = 1/1000, the conditions n >= 100 and np <= 10 are both satisfied, so we can use the Poisson distribution as an approximation to the binomial distribution. We first need the value of μ **which is found as follows:**

$$\mu = np = 365 * 1/1000 = 0.365$$

Having found the value of μ we can now find P (1) by using x = 1, μ = 0.365 and e = 2.71828, as shown here:

$$P(1) = \frac{\mu^x \cdot e^{-\mu}}{x!} = \frac{0.365^1 \cdot 2.71828^{-0.365}}{1!} = \frac{0.253}{1} = 0.253$$

Using the Poisson distribution as an approximation to the binomial distribution, we find that there is a 0.253 probability of winning exactly once in 365 days. If we use the binomial distribution, we get 0.254, so we can see that the Poisson approximation is quite good here.

Ch. 5: Probability Distributions

$\mu = \Sigma x \cdot P(x)$ Mean (prob. dist.)

$\sigma = \sqrt{\Sigma[x^2 \cdot P(x)] - \mu^2}$ Standard deviation (prob. dist.)

$P(x) = \frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x}$ Binomial probability

$\mu = n \cdot p$ Mean (binomial)

$\sigma^2 = n \cdot p \cdot q$ Variance (binomial)

$\sigma = \sqrt{n \cdot p \cdot q}$ Standard deviation (binomial)

$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!}$ Poisson distribution where $e \approx 2.71828$