PL 05 - Discrete Probability Distributions

Video 1: Random Variables

- A random variable is a variable that takes on numerical values as a result of a random experiment or measurement.
- Random variables must have numerical values.

- Discrete Random Variable

  - has a finite no. of values or an infinite sequence of values (0, 1, 2...) and the differences between the outcomes are meaningful.

- Continuous Random Variable

  - has nearly infinite no. of outcomes that cannot be easily counted and the differences between the outcomes are not meaningful.

Video 3: Discrete Random Variable Probabilities

(1) $0 \leq P(x) \leq 1$: No probabilities less than 0 or greater than 1

(2) $\Sigma P(x) = 1$: Sum of all RV probabilities $P(x)$ must be equal to 1.

Video 4 : Expected Value

Weighted Averages

Assignment	Grade%	Weight	Grade * Weight	Subscore
Homework	89%	.30	89% * .3	26.7%
Quizzes	79%	.20	79% * .2	15.8%
Midterm	84%	.25	84% * .25	21.0%
Final term	92%	.25	92% * .25	23.0%
Final Grade				86.5%

- EXPECTED VALUE

- The expected value is simply the mean of a random variable; the average expected outcomes.

- It does not have to be a value the discrete random variable can assume.

- Formula : $\boxed{E(X) = \mu = \Sigma x P(x)}$

$E(X)$ - is the expected value or mean of the outcomes x.

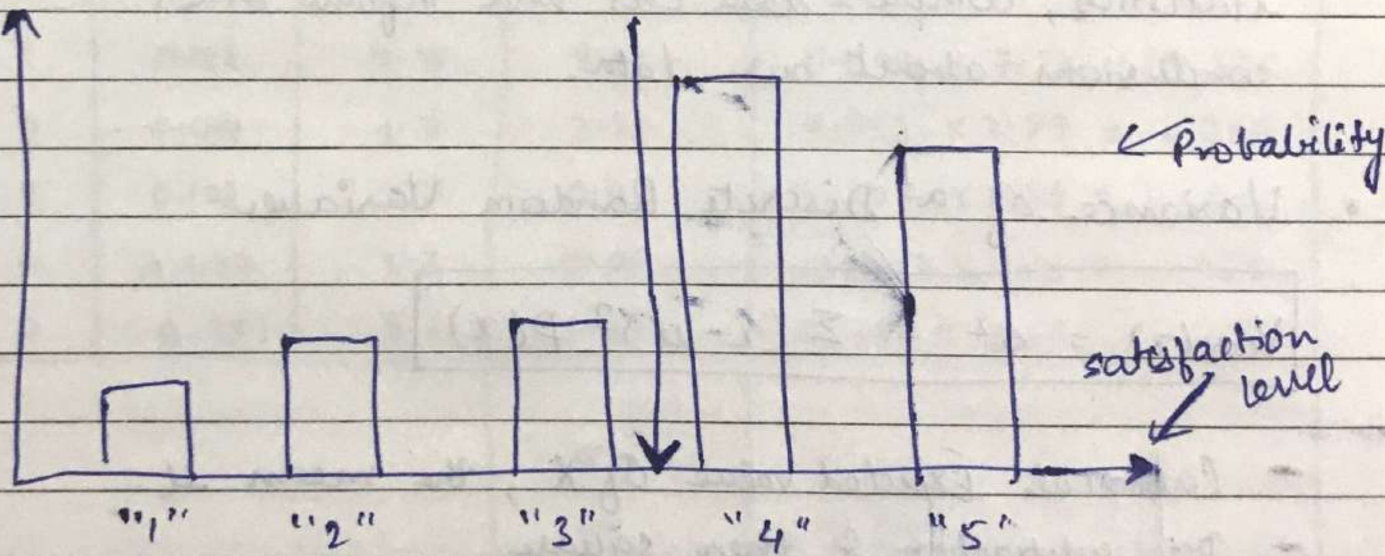
μ - is the mean

$\Sigma x P(x)$ - is the sum of each random variable value x multiplied by its own probability $P(x)$.

e.g : What is the overall satisfaction level of this class ?

x (satisfaction level)	Count (no. of students)	P(x)	x P(x)
1 (very dissatisfied)	5	5/108 = .046	1 × 0.046 = .046
2	10	10/108 = .093	2 × .093 = .186
3	11	11/108 = .102	3 × .102 = .306
4	44	44/108 = .407	4 × .407 = 1.628
5 (very satisfied)	38	38/108 = .351	5 × .351 = 1.755
Σ	108	1	$\boxed{3.70}$

Probability Distribution



$$E(X) = \mu = 3.70$$

Video 5: Discrete Random Variable Variance

- Though the Expected value tells us the mean of a random variable, oftentimes we need to know the variability, or how spread out the random variable is from its mean.

- We can use the variance and standard deviation of a random variable to learn how dispersed its is relative to its mean.

- This information can be used to calculate other statistics, compare data sets and inform other conclusions about our data.

- Variance of a Discrete Random Variable

$$\boxed{Var(x) = \sigma^2 = \Sigma (x-\mu)^2 P(x)}$$

- Calculate Expected value of X ; the mean μ
- Do subtraction & then square
- Multiply the squared difference by the probability of the random variable $P(x)$
  (In this step we are weighting the deviations using the probabilities)

- Standard Deviation of a Discrete Random Variable

$Var(x) = \sigma^2 = \Sigma(x-\mu)^2 P(x)$

$$\boxed{\sqrt{var(x)} = \sigma = \sqrt{\Sigma(x-\mu)^2 P(x)}}$$

e.g Class Satisfaction Survey

- Expected value, μ is calculated in video 4.

x	$P(x)$	μ	$(x-\mu)^2$	$(x-\mu)^2 P(x)$
1	0.046	3.7	7.29	0.046 × 7.29 = 0.335
2	0.093	3.7	2.89	0.093 × 2.89 = .269
3	0.102	3.7	0.49	0.102 × 0.49 = .05
4	0.407	3.7	0.09	0.407 × 0.09 = .037
5	0.351	3.7	1.69	0.351 × 1.69 = .488
			σ^2	1.18
			σ	1.09

Video 6: The Binomial Distribution

- A Binomial Experiment has the following characteritics.

1. The process consists of a sequence of n trials.
2. Only two exclusive outcomes are possible in each trial. One outcome is called a "success" and other a "failure".
3. The probability of a success denoted by p, does not change from trial to trial. The probability of failure is $1-p$.
4. The trials are independent. The outcome of previous trials to not influence future trials.

e.g

Trial	1	2	3	4	5
Outcome	S	S	F	S	S

→ 1 Failure & 4 Success in an experiment of 5 trials.

→ But the Failure can appear in any trial.

Trial	1	2	3	4	5
Outcome	F	S	S	S	S
Outcome	S	F	S	S	S
Outcome	S	S	F	S	S
Outcome	S	S	S	F	S
Outcome	S	S	S	S	F

∴ No. of combinations for 1 Failure
$= C(5, 1)$
$= 5$ ←

Problem : A manufacturer is making a product with a 20% defect rate. If we select 5 randomly chosen items at the end of the assembly line, what is the probability of having 1 defective item in our sample ?

Solution Space :

The answer is NOT 20% or .2 or 1/5.
B'coz the Probability of a defective product on picking up ONE product is 20%.
In other words, it is 20% chance that ~~the~~ if a product is picked up would turn out to be defective

But, the catch here is that we need to figure out the probability ~~of~~ or chance of Getting 1 defective product, when we pick up 5 products.

So here, picking up 5 products, is No. of Trials.
So with every trial, the probability of defective product is 20%.

And there can be 5 ways/combinations to pick a defective product. $C(5,1) = 5$

- Here: A success is a Defective Product.

$p = 0.2$
$p-1 = 0.8$

$C(5,1) = 5$

$n = 5$

Trial	1	2	3	4	5	Probability
1. Outcome	.2	.8	.8	.8	.8	0.08192
2. Outcome	.8	.2	.8	.8	.8	0.08192
3. Outcome	.8	.8	.2	.8	.8	0.08192
4. Outcome	.8	.8	.8	.2	.8	0.08192
5. Outcome	.8	.8	.8	.8	.2	0.08192

$$C(n,x)\, p^{x} (1-p)^{n-x}$$

$$= C(5,1) \times .2^{1} \times .8^{4}$$

$$= 5 \times 0.08192$$

$$= 0.4096 \leftarrow$$

Probability of 1 success or 1 defective product is 0.4096 or 41 %

- FORMULA :

$$\boxed{C(n,x)\, p^{x} (1-p)^{n-x}}$$

n = no. of trials

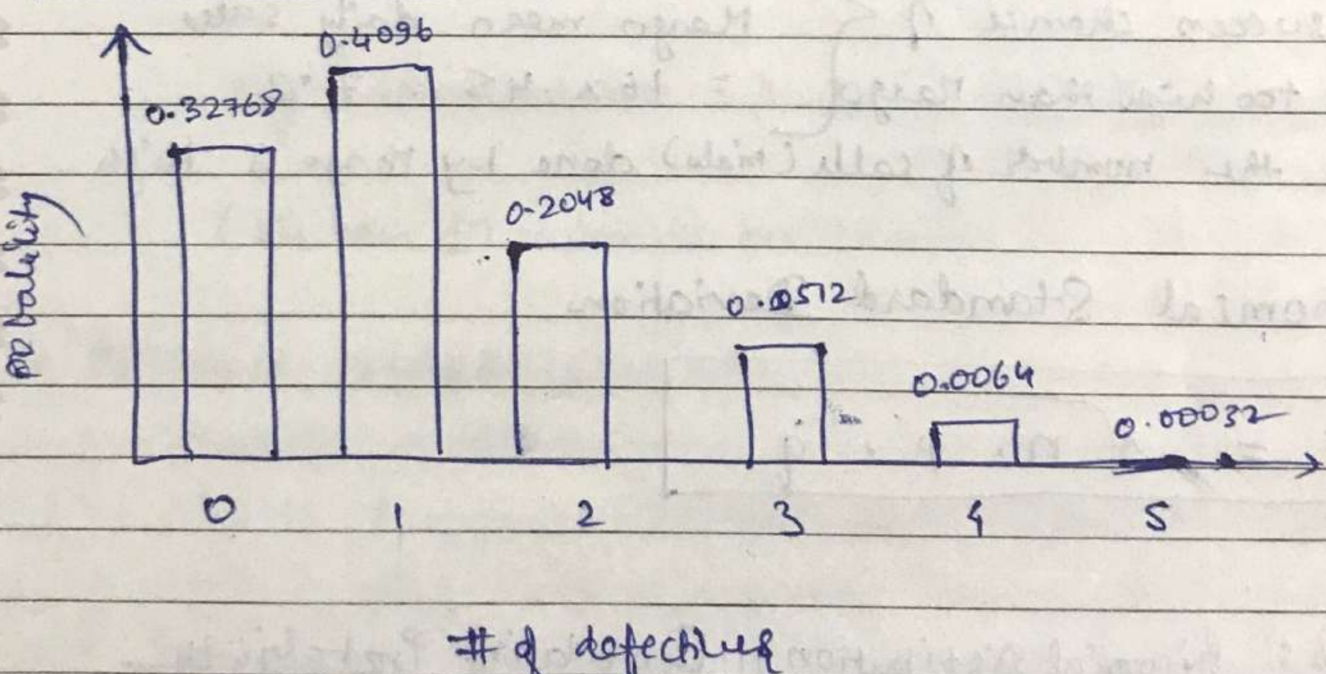
x = no. of successes

p = probability of success in any trials (constant)

- Now, moving a level up.

n	# of successes	# failures	calculation	probability	# of defectives
5	0	5	$C(5,0) \times .2^0 \times .8^5$	.32768	0
5	1	4	$C(5,1) \times .2^1 \times .8^4$	.4096	1
5	2	3	$C(5,2) \times .2^2 \times .8^3$	.2048	2
5	3	2	$C(5,3) \times .2^3 \times .8^2$	.0512	3
5	4	1	$C(5,4) \times .2^4 \times .8^1$	.0064	4
5	5	0	$C(5,5) \times .2^5 \times .8^0$	.00032	5

- Binomial Distribution for this



- Sum of all these Probabilities is 1.

Lesson The probability of any given outcome is a combination of both the no of trials and the success rate.

Video 7: Binomial Distribution: Mean & Standard Deviation

- In a Binomial Random variable, there are only two outcomes.

- Binomial Mean (Expected) Value

$$\boxed{\mu = n \cdot p}$$

e.g

	n	p	q
Joan	10	.75	.25
Margo	16	.45	.55

The Reason why mean daily sales is so close even after success chance of Joan is too high than Margo because the number of calls (trials) done by Margo is high.

Joan's mean daily sales $= 10 \times .75 = 7.5$

Margo mean daily sales $= 16 \times .45 = 7.2$

- Binomial Standard Deviation

$$\boxed{\sigma = \sqrt{n \cdot p \cdot q}}$$

Video 8: Binomial Distribution: Cumulative Probability

- The probability that more than 5 users are using OS X.

$$P(>5) = 1 - \text{binomcdf}(\underbrace{25}_{\substack{\text{no. of objects} \\ \text{no. of trials}}}, \underbrace{0.073}_{\substack{p \\ \text{success probability}}}, \underbrace{5}_{\substack{\text{no. of success} \\ \text{no. of samples}}})$$

Video 11 : Poisson Distributions : Introduction

- Queing theory Problem / Waiting line

- POISSON Distribution - focuses on the no. of discrete events or occurrences over a specified interval or continuum (time, length, distance)

$$\lambda = \frac{\text{\# of occurrences}}{\text{specified interval}}$$

Here, $E(x)$ = expected value $= \mu = \lambda$

e.g. Checkout line $\lambda = \frac{10 \text{ customers}}{\text{per 15 minutes}} = 10$

(We have 10 customers per 15 mins)

- POISSON CHARACTERISTICS
  - Discrete outcomes ($x = 0, 1, 2, 3 \ldots$)
  - The no. of occurrences in each interval can range from zero to infinity (theoritically) : $0 \leq x \leq \infty$
  - Describes the distribution of infrequent (rare) events.
  - Each event is independent of the other events
  - Describes discrete events over an interval (time, distance, etc)
  - Expected no. of occurrences $E(x)$ are assummed to be constant throughout the experiment.

- POISSON FORMULA

$$\boxed{P(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}} \text{ or } \boxed{P(x) = \frac{\mu^{x} e^{-\mu}}{x!}}$$

$x = 0, 1, 2, 3, \ldots \infty$ # no. occurrences of interest

λ or μ = long run average = $\frac{\text{\# occurrences}}{\text{interval}}$

$e = 2.7182828$ (base of natural logs)

- e.g. What is the probability that exactly 7 customers enter the line between 4:30 & 4:45 ?

$x = 7$

$\lambda = 10$

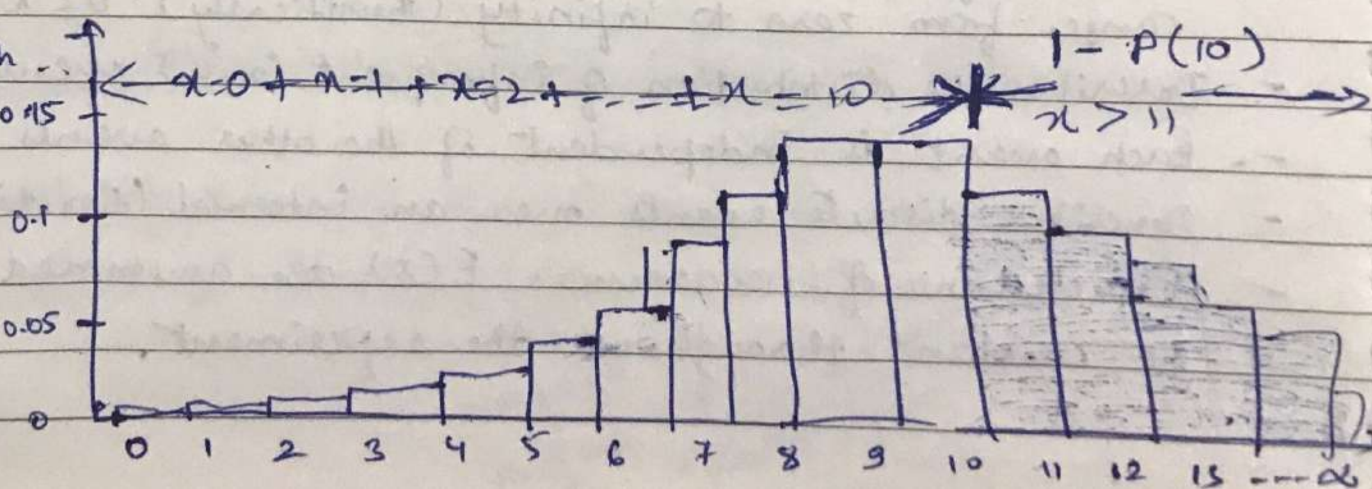
$e = 2.7182828$

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

$$P(7) = \frac{10^{7} e^{-10}}{7!}$$

$$\Rightarrow P(7) = 0.09$$

- Graph



- To summarize.

The Probability that exactly 7 customers enter your line between 4:30 & 4:45? 0.09

The Probability that more than 10 people arrive?
1- poissoncdf (10,10) = 0.417

- Scenario changed:

How many customers do you expect to arrive between 4:30 & 4:40?

Expected

$$\frac{10 \text{ customers}}{15 \text{ minutes}} = \frac{?}{10 \text{ minutes}} \Rightarrow \text{Expected customers in 10 mins} = 6.67$$

[PS: Simple Unitary calculations]

Video 12: Poisson Distribution: Problems

- On Average, 1.6 customers walk up to ATM during any 10 min interval b/w 9 pm & midnight.

→ What is λ for this problem? 1.6

→ Probability of exactly 3 customers using ATM in any 10 min?

$$P(3) = \frac{1.6^3 \times e^{-1.6}}{3!} = 0.139 \text{ or } 14\%$$

→ Probability of 3 or fewer people? possioncdf (1.6, 3) = 0.921