Random Variables

Agenda

- Random Variable
- Discrete and Continuous Random Variables
- Expected Value and Variance

What is a Random Variable?

- A random variable assigns a numerical value to each possible outcome of a random experiment.
  - Describes prob for an uncertain outcome of a random event
- Random because for any run of the experiment, one of several outcomes can be realized
  - With each outcome there is a chance (probability) associated
- Are these random variables?
  - Today's weather
  - The color of a car chosen at random
  - 1 if the next car we see is blue, 2 if it's green, 4 if black…
  - The result of a coin flip (heads or tails)
  - For a coin flip, assign 1 if heads, 0 if tails
  - The price of Microsoft stock

Varieties of Random Variables

- Discrete
  - Finite number of outcomes
  - The number of chocolates in a box

- Continuous
  - All possible values in some range
  - The amount of dessert I will eat today
  - Large number of outcomes -> continuous distribution

Discrete or Random?

- Are they discrete or random?
- Today's weather
  - Continuous
- The color of a car chosen at random
  - Discrete
- Height of a student in class
  - Continuous
- Number of insurance policies issued by a company on a day
  - Discrete
- Sometimes, it is good to approximate discrete as continuous

Discrete Random Variable

Discrete Random Variables: Roll of a Die

Outcome x	Probability p	P(X=x)=p
1	1/6	P(X=1) = 1/6
2	1/6	P(X=2) = 1/6
3	1/6	P(X=3) = 1/6
4	1/6	P(X=4) = 1/6
5	1/6	P(X=5) = 1/6
6	1/6	P(X=6) = 1/6

- X: random variable, the value on the die when you roll it.
- x: an outcome, either 1, 2, 3, 4, 5 or 6.
- For each value of RV, there is prob of occurrence of value
- This table is called probability distribution of a RV

Properties of probability distribution

- P(X=x) the probability that X will have outcome x.
- P(x) >=0
- i.e. The probabilities must sum to 1.
- This is the probability mass function for this random variable
  - Every PMF of a RV must satisfy two properties mentioned above

Discrete Random Variables: Roll of a Die

Outcome x	Probability p	P(X=x)=p
10	1/6	P(X=10) = 1/6
20	3/12	P(X=20) = 3/12
30	1/12	P(X=30) = 1/12
40	1/6	P(X=40) = 1/6
50	1/6	P(X=50) = 1/6
60	1/6	P(X=60) = 1/6

- X: random variable, the value on the die when you roll it.
- x: an outcome, either 10, 20, 30, 40, 50 or 60.
- P(X=x) the probability that X will have outcome x. The probabilities must sum to 1.

Discrete Random Variables: Roll of a Die

Outcome x	Probability p	P(X=x)=p
x_1	p_1	$P(X=x_1) = p_1$
x_2	p_2	$P(X=x_2) = p_2$
x_3	p_3	$P(X=x_3) = p_3$
x_4	p_4	:
:	:	:
x_n	p_n	$P(X=x_n) = p_n$

- X: random variable, the value on the die when you roll it.
- x: an outcome, either x_1, x_2, x_3, x_4,…,x_n.
- P(X=x) the probability that X will have outcome x.

Cumulative Distribution Function

Outcome x	Probability p	P(X=x)=p	Cumulative Prob
1	1/6	P(X=1) = 1/6	1/6
2	1/6	P(X=2) = 1/6	2/6
3	1/6	P(X=3) = 1/6	3/6
4	1/6	P(X=4) = 1/6	4/6
5	1/6	P(X=5) = 1/6	5/6
6	1/6	P(X=6) = 1/6	1

- Often interested in cumulative probabilities
- Prob (X<=x) is cumulative probability
- Obtained by summing up probabilities up to the outcome
- F(x) = P(X<=x) =

Summarizing a random variable

- Suppose you went on a trip to Hawaii, and noted following temperature on each day:
  - 30, 32, 35, 38, 40
  - How would you summarize this information to tell someone about temperature in Hawaii?
  - Next week temp is 25, 30, 35, 40, 45. Are the two series same?
  - Is week 2 hotter than week 1?
  - Is week 2 more extreme than week 1?

Summarizing a random variable

- RV needs to be summarized
- Most times, summarization is done using two numbers:
  - Central tendency or mean or expected value
  - Dispersion or spread
  - There are others as well, but for now we will focus on these two

Expected Value

- Central tendency
  - Measure to find out expected(average) value
  - Mean is prob weighted average value that is 'expected' to occur
  - For discrete RV: $\mu = \mathrm{E(X)} = \sum x * p(x)$

Computing the Mean or Expected Value

Outcome x	Probability p
10	1/6
20	1/6
30	1/6
40	1/6
50	1/6
60	1/6

- How do you compute the mean?

$$\mu = 10 \times \frac{1}{6} + 20 \times \frac{1}{6} + 30 \times \frac{1}{6} + \ldots + 60 \times \frac{1}{6} = 35$$

$$\mu = x_1 \times p_1 + x_2 \times p_2 + x_3 \times p_3 + \ldots + x_n \times p_n = \sum_i x_i p_i$$

Computing the Mean or Expected Value

Outcome x	Probability p
10	1/6
20	3/12
30	1/12
40	1/6
50	1/6
60	1/6

- How do you compute the mean?

$$\mu = 10\times\frac{1}{6}+20\times\frac{3}{12}+30\times\frac{1}{12}+\ldots+60\times\frac{1}{6}=34.1667$$

$$\mu = x_1\times p_1+x_2\times p_2+x_3\times p_3+\ldots+x_n\times p_n = \sum_i x_i p_i$$

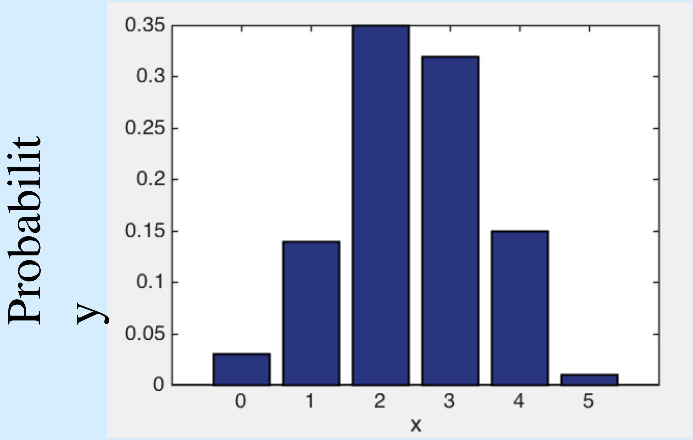
Computing the Spread

Outcome x	Probability p
25	1/5
30	1/5
35	1/5
40	1/5
45	1/5

- Mean?
- How might you compute the spread?
- $\sum(x-\mu) * p(x)$????

Computing the Spread

Outcome x	Probabilit y p
0	.03
1	.14
2	.35
3	.32
4	.15
5	.01



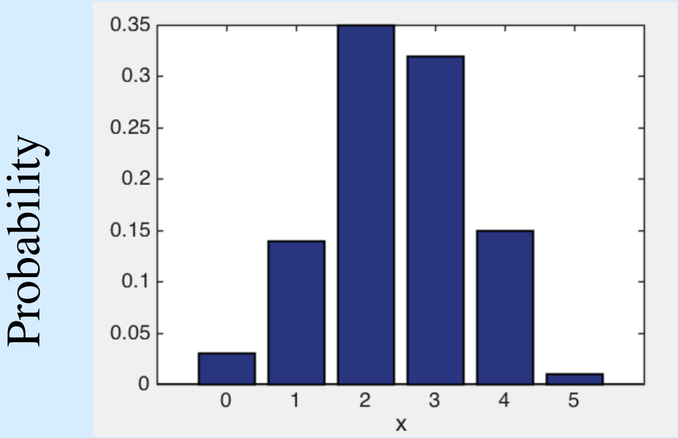
- How might you compute the spread?

$$\sum_i p_i \left| x_i - \mu \right| ?$$

- Sum of absolute differences from the mean?

Computing the Spread

Outcome x	Probability p
0	.03
1	.14
2	.35
3	.32
4	.15
5	.01



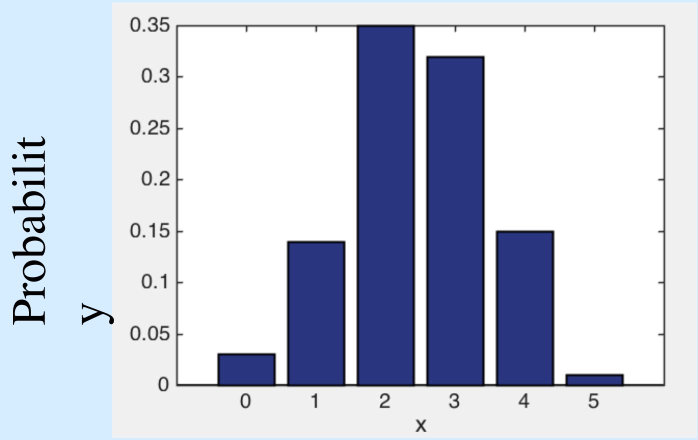
- How might you compute the spread?

$$\sum_i p_i (x_i - \mu)^2 = \mathrm{Var}(X)$$

- Sum of squared differences from the mean

Computing the Spread

Outcome x	Probability p
0	.03
1	.14
2	.35
3	.32
4	.15
5	.01



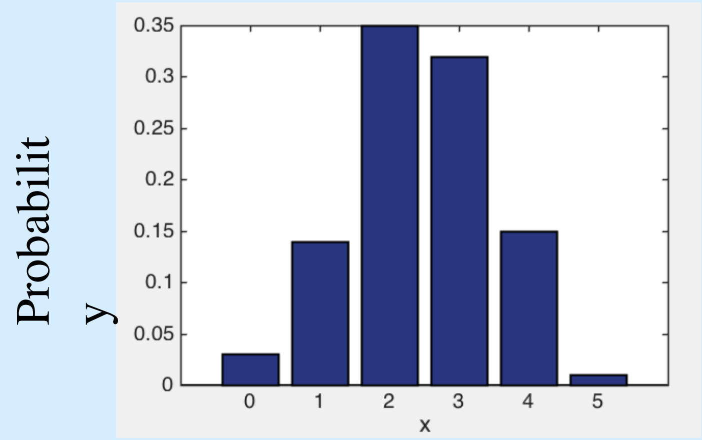
$$\mathrm{Var(X)} = \sum_i p_i \left(x_i - \mu \right)^2$$

$$= .03 \times (0 - 2.45)^2 + .14 \times (1 - 2.45)^2 + \ldots + .01 \times (5 - 2.45)^2$$

$$= 1.0675$$

Computing the Spread

Outcome x	Probability p
0	.03
1	.14
2	.35
3	.32
4	.15
5	.01



$$\sigma = \mathrm{STD}(X) = \sqrt{\mathrm{Var}(X)}$$

$$= \sqrt{1.0675} = 1.0332$$

Variance…Putting it formally

- Measure of spread in a random variable
- Expected square deviation of RV from its mean
- It is also an expectation (but of higher order)
- Prob of outcomes are weights for computation
- Var(X) = $E(X - E(X))^2 = \sum \big(X - E(X)\big)^2 p(X)$
- Can also be written as
  - Var(X) = E(X^2) - $E(X)^2$
- Check if both are equal
- St. Dev = $\sqrt{Var(X)}$

Temperature Example

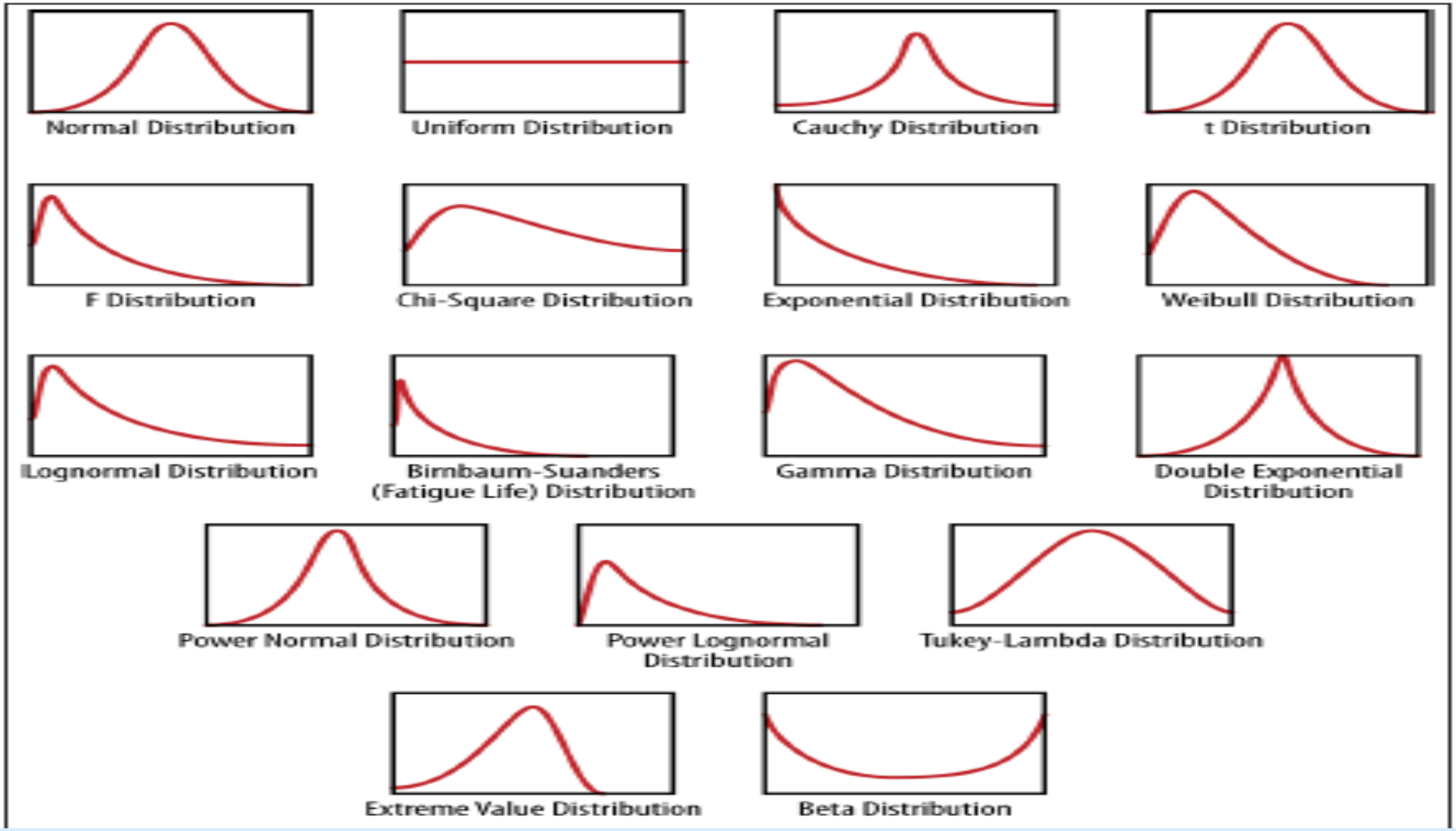
X	P(X)	X-mean	$(X-mean)^2$	$(X-mean)^2$*p(x)	$(X)^2$	$(X)^2$*p(x)
20	0.20	-10	100	20	400	80
25	0.20	-5	25	5	625	125
30	0.20	0	0	0	900	180
35	0.20	5	25	5	1225	245
40	0.20	10	100	20	1600	320

- Mean = 30
- Var(X) = $E(X-E(X))^2$
  - $= .2^*(-10)^2 + .2^*(-5)^2 + \ldots$
  - = 50
- Also, Var(X) = E(X^2) - $E(X)^2$
  - = (80+125+...) – 900
  - = 50
- St Dev = $\sqrt{50}$ = 7.07

Some Properties

- Expected value of a linear combination of a RV:
  - $E(aX + b) = aE(X) + b$; a and b are constants
- Variance of a linear combination of RV:
  - $V(aX + b) = a^2 * V(X)$

Common Distributions



- There are lots of distribution functions
  - Both discrete and continuous in nature
- We will study commonly used distributions:
  - Discrete : Bernoulli, Binomial, Poisson
  - Continuous : Uniform, Normal

Bernoulli

- Let us flip a coin.

Outcome x	Probability p
1	p
0	1-p

$$X \sim \text{Bernoulli}(p)$$

- X is a RV with Bernoulli distribution
- Example: In Hyderabad, every person has 60% chance of having a vehicle.
  - 1 if they have vehicle, 0 otherwise
  - X~Bernoulli(.60)

Bernoulli Distribution (Expected Value and Variance)

- For X, $E(X) = p$ and $V(X) = p*(1-p)$

Binomial Distribution

- Extension of Bernoulli trials to n times, each with the same probability of success.

Binomial

- Select 10 people in Hyderabad at random.
- Let X = number of them having a vehicle
- Of 10 randomly chosen people, what is the probability that first 8 will have vehicle and the next 2 not?
  - P(YYYYYYYYNN) = $(.60)^8*(.40)^2$

- Probability of first 3 having a vehicle and next 7 not?
  - P(YYYNNNNNNN) = $(.60)^3*(.40)^7$

- $P(X=3)$?
- P(YYYNNNNNNN) = $(.60)^3 * (.40)^7$
- P(YYNYNNNNNN) = $(.60)^3 * (.40)^7$
- P(NNNYNNNYNY) = $(.60)^3 * (.40)^7$
- And so on….
- In order to arrive at final prob, we would have to add up all possible cases. How many such arrangements are there?

Detour: Permutations & Combinations

- Imagine 10 AMPBA students sitting in front row. If I have to pick three of them, in how many ways can I do it?
- How many ways to arrange 10?

  10! = 10*9*8*7*6*5*4*3*2*1
- Let us say I pick 3 people. In how many ways I can arrange them?
  - 3!
- For the seven I did not pick, how many ways?
  - 7!

- Therefore, number of ways to pick are

$$\frac{10!}{3!7!} = 120$$

- Formally speaking, to choose x out of n (n>x):

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{x!n-x!}$$

Coming back to Binomial

- Number of such cases are $\frac{10!}{3!7!} = 120$
- Thus, P(X=3) = 120* $(.60)^3$*$(.40)^7$ = 0.04
- More generally, we can write

$$P(X = x) = \frac{n!}{x!n-x!}(p)^x(1-p)^{n-x}$$

Formalizing Binomial Distribution

- There are n independent trials.
- Each trial has two outcomes only: 1 or 0 with prob p and 1-p respectively
- Parameters of a binomially distributed random variable are Number of trials (n) and Prob of success (p).
- It is denoted as X ~ Bin(n,p)
- PMF of X is given by

$$P(X=x)=\frac{n!}{x!n-x!}(p)^x(1-p)^{n-x}$$

- For X, E(X) = np and V(X) = n*p*(1-p)

Example

- Let us do a coin toss 20 times.
- P(H) = P(1-H) = 0.5
- X = number of tosses with heads
- What is P(X=10)?

$$P(X=x) = \frac{n!}{x!n-x!}(p)^x(1-p)^{n-x}$$

$$P(X=10) = \frac{20!}{10!20-10!}(.5)^{10}(1-.5)^{20-10} = .1762$$

Example

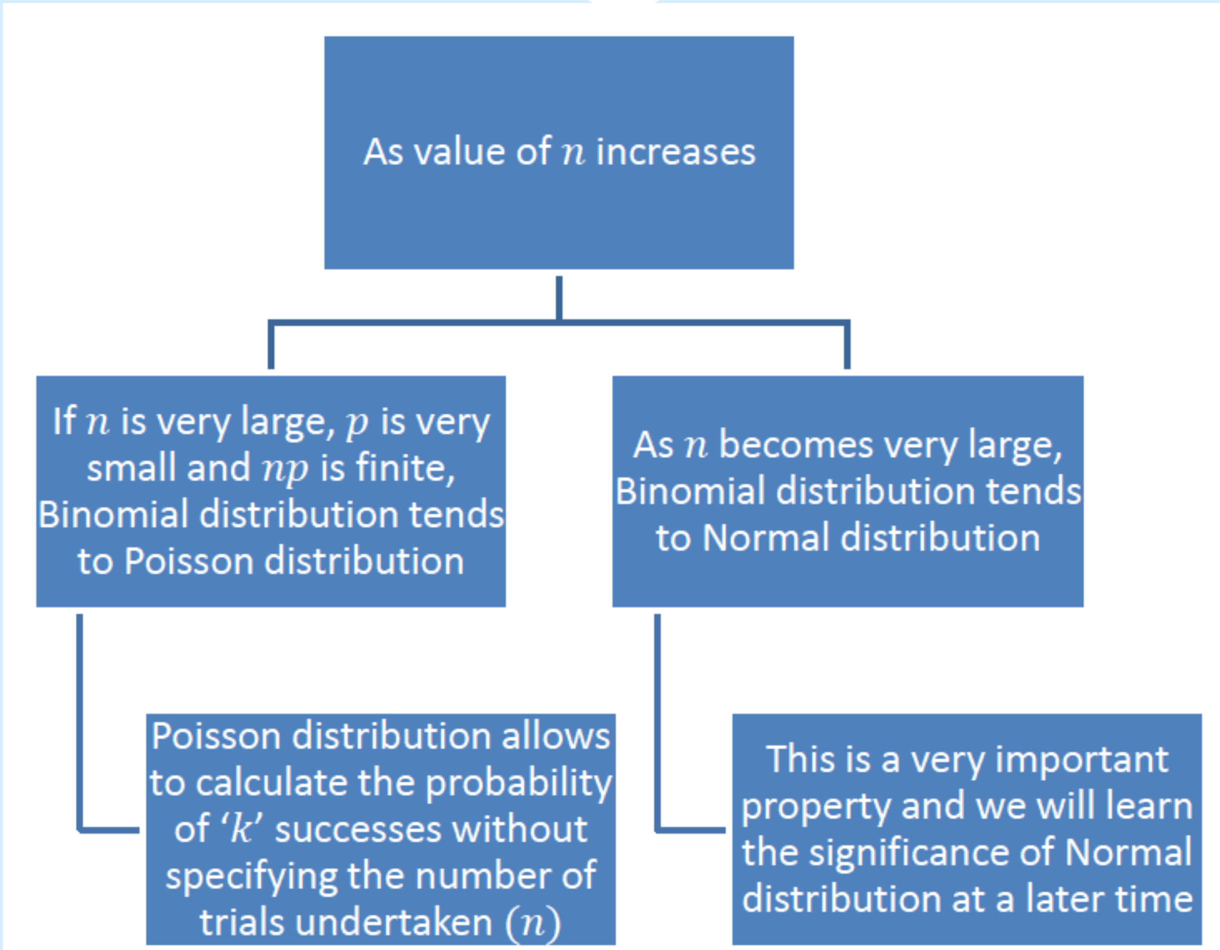
- P(X≤10)?

$$P(X=0)+P(X=2)+\cdots+P(X=10)=.5881$$

Applications

- There are many situations where binomial distribution can evaluate probability of events:
  - If an insurance company knows the probability of a claim being fraudulent (or not), binomial distribution can help determine the probability that there would be more than 'k' fraudulent claims in the next 1000 claims
  - If an advertising company knows the probability of a customer buying a certain product, if he/she receives an advertisement on their Facebook account, binomial distribution can help them determine whether 'k' purchases would be made from a total of 10000 Facebook ads sent. Effectively, they can determine how many Facebook advertisements to send, if they want to achieve a particular sales target
  - Suppose that a hardware store manager has 3 different suppliers. If he knows the probability of a newly purchased bolt being defective, based on the supplier it is bought from, binomial distribution can help him/her check whether there will be less than 'k' defective bolts in a box of 2000 bolts. This would help his decide which supplier to buy the bolts from.

Binomial Distribution Properties



Poisson Distribution

Consider a made up example

- There are more than 50 Million active phone connections in India used daily. On average, 1000 phones break because of falling while traveling.
- Compute the prob that 5 phones will break while traveling today.
- It can be approached using Binomial distribution
  - p = P(phone breaks while traveling) = 1000/50 Million
  - n = 50 Million
- Thus, Prob(X=5) = $P(X=x)=\frac{n!}{x!n-x!}(p)^{x}(1-p)^{n-x}$
- Consider solving this using a high end computer!!!!

Poisson Distribution

- For such situations, we make use of the fact that if n is large, p is small, and np is a finite number, then we can use the following PMF function:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2,\ldots$$

- Where λ = np where n is number of trials and p is number of successes
- E(X) = λ and V(X) = λ
- It is essentially an extension of Binomial distribution

Example

- A LIC salesman sells on the average 3 insurance policies per week. Calculate the probability that in a given week he will sell
- A) Some policies
- B) 2 or more policies but less than 5 policies
- C) Assuming that there are 5 working days per week, what is the probability that in a given day he/she will sell one policy?

Example

Here, $\mu = 3$

(a) "Some policies" means "1 or more policies". We can work this out by finding 1 minus the "zero policies" probability:

$$P(X > 0) = 1 - P(x_0)$$

Now $P(X) = \dfrac{e^{-\mu}\mu^x}{x!}$ so $P(x_0) = \dfrac{e^{-3}3^0}{0!} = 4.9787 \times 10^{-2}$

Therefore the probability of 1 or more policies is given by:

$$\begin{aligned} \text{Probability} &= P(X \geq 0) \\ &= 1 - P(x_0) \\ &= 1 - 4.9787 \times 10^{-2} \\ &= 0.95021 \end{aligned}$$

Example

(b) The probability of selling 2 or more, but less than 5 policies is:

$$P(2 \leq X < 5)$$

$$= P(x_2) + P(x_3) + P(x_4)$$

$$= \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} + \frac{e^{-3}3^4}{4!}$$

$$= 0.61611$$

(c) Average number of policies sold per day: $\frac{3}{5} = 0.6$

So on a given day, $P(X) = \frac{e^{-0.6}(0.6)^1}{1!} = 0.32929$

Example

- Cars pass through a busy road at an average rate of 300 per hour.
- Find the probability that none passes in a given minute.
- What is the expected number passing in two minutes?
- Find the probability that this expected number actually pass through in a given two-minute period.

Example

The average number of cars per minute is: $\mu = \frac{300}{60} = 5$

(a) $P(x_0) = \frac{e^{-5}5^0}{0!} = 6.7379 \times 10^{-3}$

(b) Expected number each 2 minutes = $E(X) = 5 \times 2 = 10$

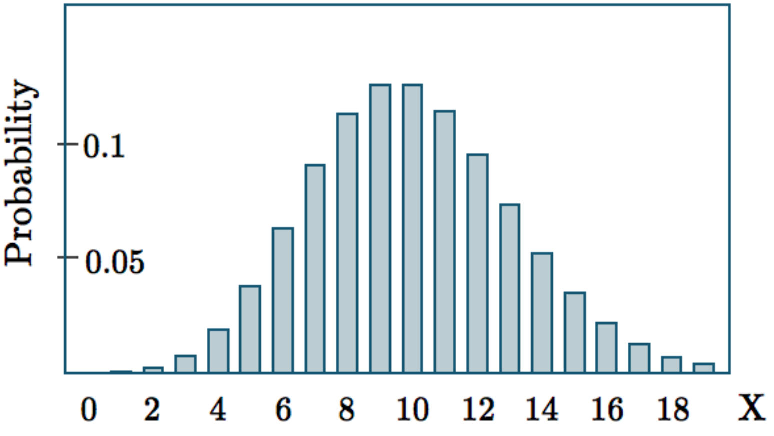
(c) Now, with $\mu = 10$, we have: $P(x_{10}) = \frac{e^{-10}10^{10}}{10!} = 0.12511$

Example

Based on the function

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

we can plot a histogram of the probabilities for the number of cars for each 2 minute period:



Histogram of the Poisson distribution

Applications

- Prob of a house being destroyed by earthquake in SF area
- Prob of a meteor hitting earth in next year
- Prob of finding a defective part in a machine that produces million components and each component have a prob of 0.1