Statistical Analysis ST1002

Myra O'Regan

October 14, 2015

Theoretical Distributions- why study them?

- We want to check if our data is like any of them
- We want to use them for simulation purposes
- Some are used in Statistical theory
- Underlying process
- pdf function - probability density function
- Enables us to calculate probabilities.
- Parameters needed to describe a distribution
- Mean called expected value

Two types of distributions

Discrete Dist Variables only take on integer values

- No. of heads on 10 tosses of a coin
- No. of visitors to website

Continuous dist Variables can take on all possible values

- Heights of people
- Queueing time for jobs

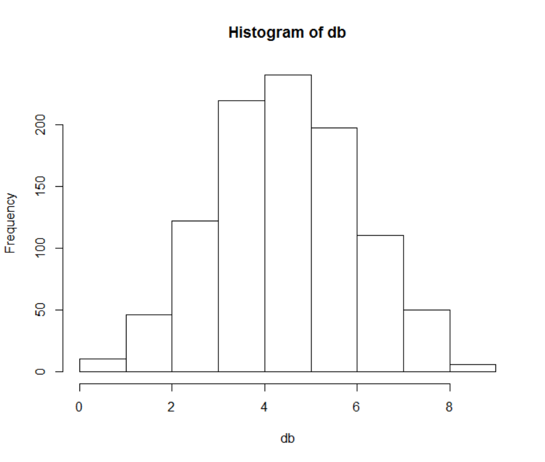
Discrete Distributions

- Bernouilli Distribution
- Random variable with two possible outcome 0 or 1
- Transmitted or lost signals
- Benign or malicious attachments
- Defined by 1 parameter p= probability of a 1
- Mean = p
- Variance = 1-p

Binomial Distribution

- A sequence of independent bernouilli trials
- Count the number of successes
- Two parameters
- n: number of trials
- p: probability of success constant over trials
- $P(X = x) = \binom{n}{k} p^x (1-p)^{n-x}, x = 0, 1, ..., n$
- (1-p) sometimes written as q
- Mean = np; Variance = np(1-p)

Figure : Binomial with p=.5, n=10



Example of Binomial

Toss a coin 10 times - count heads

- What are the possible values for variable?
- $P(X = x) = \binom{n}{k} p^x (1-p)^{n-x}, x = 0, 1, ..., n$
- $P(X = 4) = \binom{10}{4} (0.5)^4 (1-0.5)^{10-4} = 0.21$
- Interpretation

0	1	2	3	4	5	6	7	8	9	10
1	10	50	117	209	240	199	124	40	9	1

Poisson Distribution

- The number of rare events occuring within a fixed period of time has a Poisson distribution
- Number of customers opening accounts per day for an internet service
- Number of mistakes on a page
- $P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ where x=0,1,2,3,,
- 1 parameter $\lambda =$ average number of events in an interval
- Mean $= \lambda$
- Variance $= \lambda$

Example of Poisson

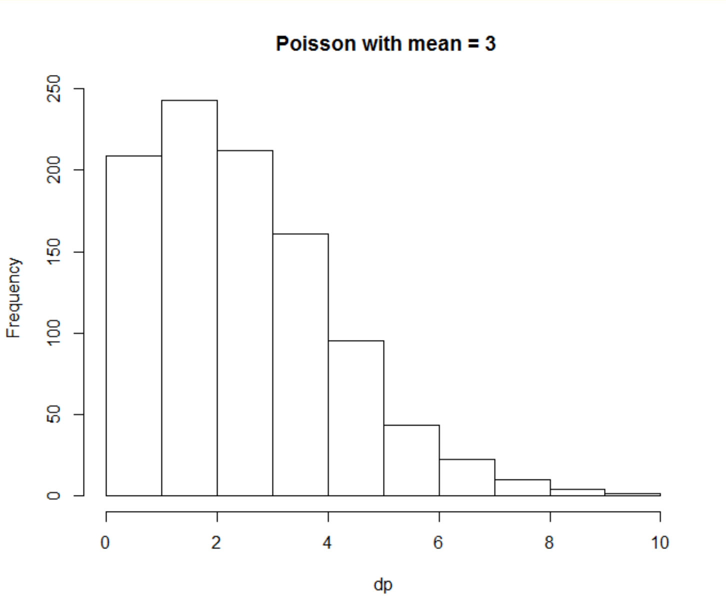
On average 3 customers sign up to a particular internet site per day

- P(0 customers)$= e^{-3}\frac{3^0}{0!} = 0.05$
- P(5 customers) $= e^{-3}\frac{3^5}{5!} = 0.10$

Simulation of Poisson

- Simulate 1000 observations with mean =3

0	1	2	3	4	5	6	7	8	9
48	158	230	202	168	106	50	26	10	2


Poisson with mean = 3
Frequency
0
50
100
150
200
250
0
2
4
6
8
10
dp

Poisson approximation to Binomial

- $Binomial(n, p) \approx Poisson(\lambda)$
- where $n \geqslant 30, p \leqslant 0.05, np = \lambda$

Uniform distribution

- Every value is equally likely in an interval (a,b)
- Can be discrete or continuous

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

$$Mean = \frac{a+b}{2}$$

$$variance = \frac{(b-a)^2}{12}$$

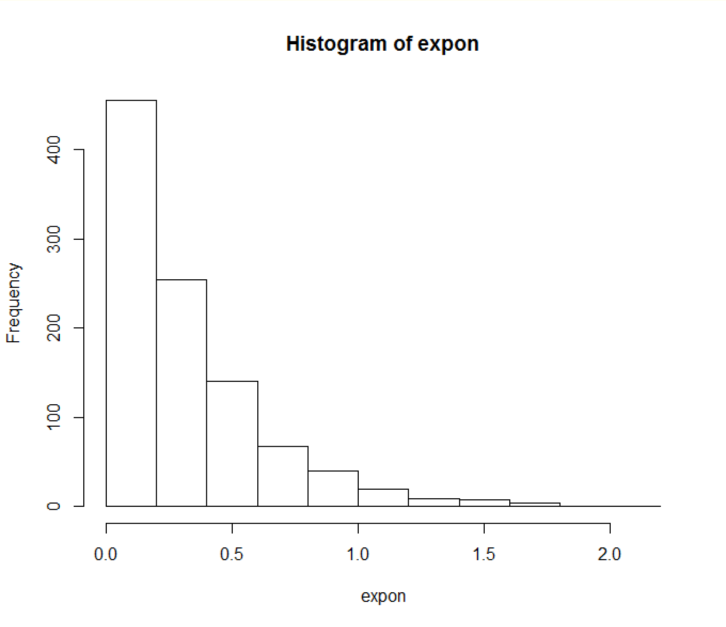
Exponential Distribution

- Often used to model wait time
- Time between Poisson events

$$f(x) = \lambda e^{-\lambda x} \text{ x} > 0$$

where λ is the average in a particular interval.

- λ is the variance as well
- Exponential variables lose memory
- Having waited for t minutes gets forgotten and it does not affect future waiting time


Histogram of expon
Frequency
0
100
200
300
400
0.0
0.5
1.0
1.5
2.0
expon

Summary

- We have looked at a couple distributions.
- We have seen how to calculate probabilities
- We will see in labs how to generate random numbers from certain distributions
- Many many more distributions
- We will look at one more special one called the Normal distribution