**Assignment stats 1**

**Problem statement 1:**

Ordered data : 4,4,5,5,6,6,6,6,7,7,7,7,7,8,8,8,8,9,9,10

N = 20

mean = sum(data)/N = 137/20 = 6.85

median = 7

mode =7

std = sqrt((sum(xi-mean)^2)/N) = 1.59

**Problem Statement 2:**

Ordered data: 28, 40, 68, 70, 75, 75, 75, 75, 80, 86, 89, 90, 90, 97, 97, 100, 100, 100, 104, 104, 109, 113, 120, 120, 120, 122, 123, 123, 130, 140, 145, 170, 174, 194, 217

N = 35

mean = 107.514

median = 100

mode = 75

std = 38.773

**Problem Statement 3:**

x = 0, 1, 2, 3, 4, 5

N = 6

f(x) = 0.09, 0.15, 0.40, 0.25, 0.10, 0.01

mean = value \* probability = x \* f(x) = 2.15

variance = 3.04

**Problem statement 4**

f(d) = 20e^(-20(d-12.5)), d>= 12.5

till d=12.6:

P(12.5=<x<=12.6) = = ... (y =-20(x-12.5)) …= e^0 – e^(-2) = 0.865

The proportion that was scrapped is 1-0.865 = 0.135

The proportion that is left is 0.865

CDF is 0, because diameter has to be >= 12.5

Conclusion: proportion are much higher closer to 12.5 due to our function

**Problem statement 5:**

x = 1, 0 (1- not faulty, 0 - faulty)

f(x) = 0.7, 0.3

P(2 are faulty):

P(4/6) = 6!/(4!\*2!)\*0.7^4\*0.3^2 = 15\*0.02 = 0.324

If we are selecting 6 LEDs, average value we can calculate as:

mean = 6\*0.75=4.5

and stdv as:

stdv =sqrt(6\*0.75\*0.25) = 1.06

**Problem Statement 6:**

mean (Gaurav) = 8 \* 0 .75 = 6 (correct questions per day)

mean (Barakha) = 12 \* 0.45 = 5.4

P(5) (Gaurav) = e^(-6)\*6^5/(5!) = 0.16

P (5) (Barakha) = 0.173

P(4) (Barakha) = 0.16

P(6) (Barakha) = 0.155

P(4) (Gaurav) = 0.134

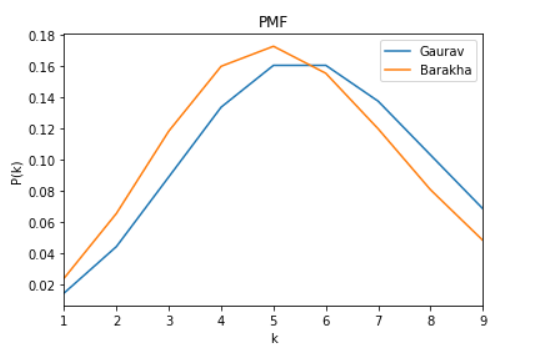
P(6) (Gaurav) = 0.16

Strange factor is that P(5) and P(6) for Gaurav are same. Don’t know how to excplain. I would

except P(6)>P(5) for him to be higher. For him we can conclude that maximum is in 5 and 6 correct questions and the probability is going to decline for other values

For Barakha the highest probability is in 5 correct questions.

Two main governing factors affecting their ability to solve more questions correctly are number of questions solved by day and correction rate.



**Problem Statement 7**

mean = 72 (per hour)

mean = 72/15 = 4.8 (per 4 min)

P(k) = e^(-mean)\*mean^k/k!

a)

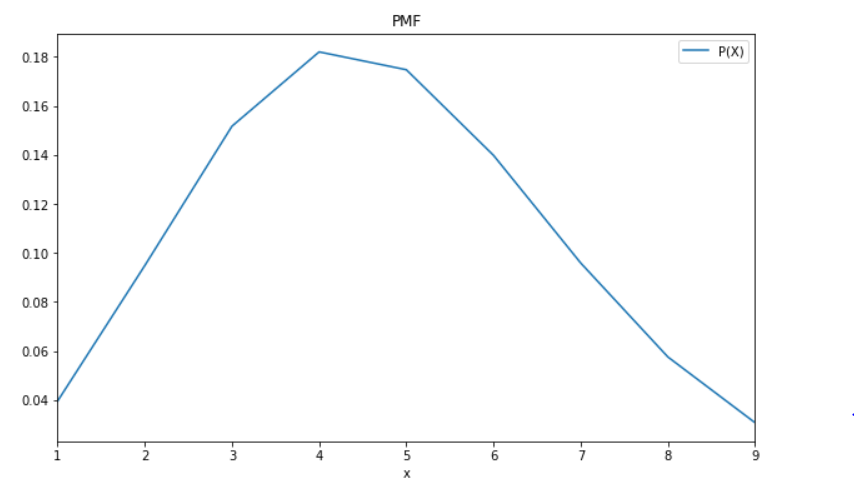
P(5) = e^(-4.8)\*4.8^5/5! = 0.175

b)

P(<=3) = P(1) +P(2) + P(3) = 0.039+0.095+0.152 = 0.286

c)

P(>3) = 1 – P(<=3) = 0.714



**Problem statement 8:**

mean = 0.1 (per 77 words)

mean = 0.1 \* 455/77 = 0.6 (scaled to 455 words)

P(2) = 0.099

When number of words increases, probability for two errors occurring increases:

Number of words = 1000

P(2) = 0.23

When number of words decreases, probability for two errors occurring decreases:

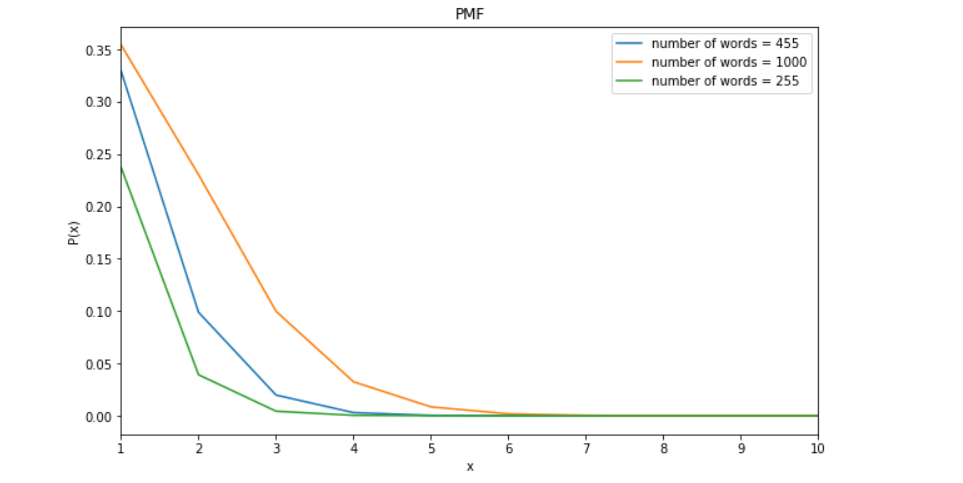
Number of words = 255

P(2) = 0.04

The reason for this is that by increasing/decreasing number of words, **lambda is higher/lower**

When number of words increases/decreases, PMF has higher or equal/lower or equal values

(PMF has smaller/bigger slope)



**Problem statement 9**

= Problem statement 4

**Problem statement 10**

a)

P(Z>1.26) = 1 – P(Z<1.26) = 1 – 0.8962 = 0.1038

P(Z<-0.86) = 0.1949

P(Z>-1.37) = 1 - P(Z<-1.37) = 1 – 0.0853 = 0.9147

P(−1.25 < Z < 0.37) = P(Z<0.37) – P(Z<-1.25) = 0.6443 – 0.1056 = 0.5387

P(Z<=-4.6) :

P(Z = 4.6) = 0 (probability for one point with continuous distributions is zero )

P(Z<4.6) = very close to zero

P(Z <=4.6) = 0

b)

P(Z>z)=0.05 -> P(Z<z) = 0.95

z = 1.65

c)

P(-z<Z<z) = 0.99 -> P(Z<-z)=(1-0.99)/2=0.01/2=0.005

-z = -2.57

z = 2.57

**Problem statement 11**

mean = 10 mA

variance = 4 mA^2

standardization: z = (13-10)/2 = 3/2 = 1.5

P(Z>1.5) = 1 – P(Z<1.5) = 1 – 0.9332 = 0.067

P((9-10)/2<Z<(11-10)/2)=P(-0.5<Z<0.5)= 1 – 2\*P(Z<-0.5) = 1 – 0.617 = 0.383

In continuous distribution you cannot find probability for one value. I will calculate following:

P(Z<z) = 0.98 -> z = 2.05 -> measurement = (2.05+10)\*2 = 24.1

All measurements will be less than 24.1 with probability of 0.98.

**Problem statement 12**

mean = 0.2508

stdv = 0.005

P (0.2485<X<0.2515) = P(-4.6<Z<1.4) = 0.9192

91.92 % of shaft is in proportion with specification

mean(new) = 0.2500

P(-4.6<Z<3) = 0.9987

99.87 % of shaft is in proportion with specification

Conclusion: If mean is closer to the center of the specifications of the shaft, proportion of shafts are more in sync with the specifications