MET CS544 A1 Foundations of Analytics
Homework#4

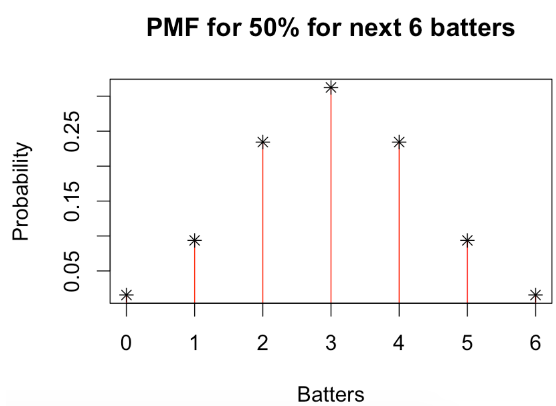
Problem#1

Suppose a pitcher in Baseball has 50% chance of getting a strike-out.
Using the binomial distribution,
a) Compute and plot the probability distribution for striking out the next 6 batters.
b) Plot the CDF for the above
c) Repeat a) and b) if the pitcher has 70% chance of getting a strike-out.
d) Repeat a) and b) if the pitcher has 30% chance of getting a strike-out.
e) Infer from the shape of the distributions

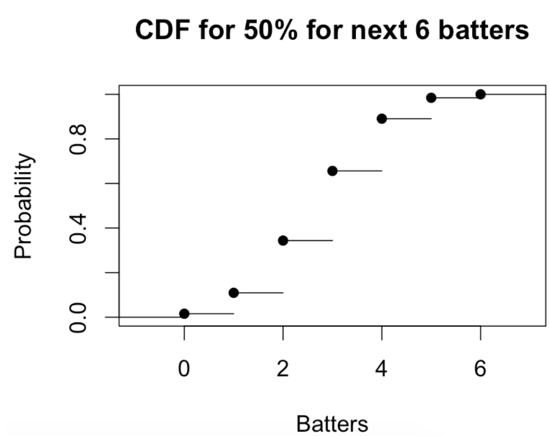
Solution:

a) The Probabilities for the next 6 batters having 50% strike out are given below using binomial distribution,

```
> strike.outs.50
[1] 0.015625 0.093750 0.234375 0.312500 0.234375 0.093750 0.015625
> sum(strike.outs.50)
[1] 1
```

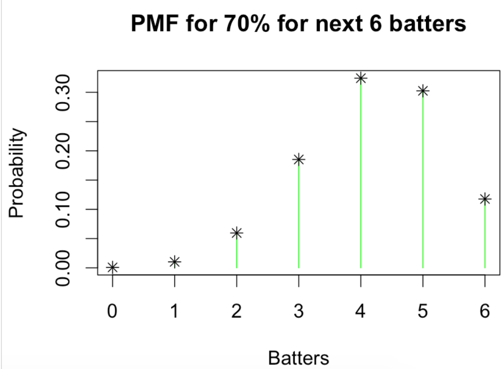


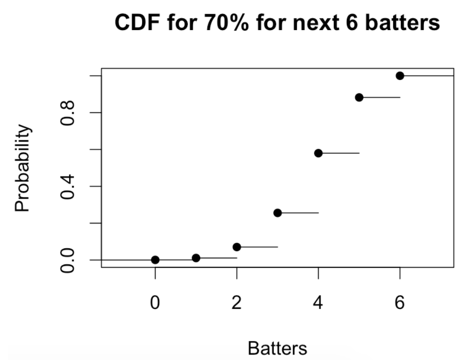
b) The cumulative distributive function is shown here,



c) The Probabilities for the next 6 batters having 70% strike out are given below using binomial distribution,

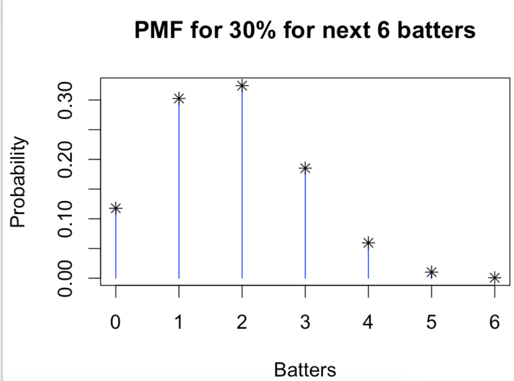
```
> strike.outs.70
[1] 0.000729 0.010206 0.059535 0.185220 0.324135 0.302526 0.117649
> sum(strike.outs.70)
[1] 1
```

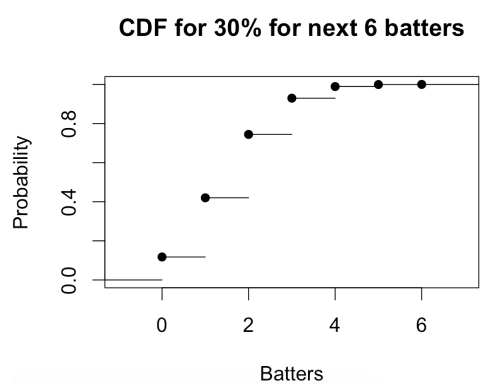




d) The Probabilities for the next 6 batters having 30% strike out are given below using binomial distribution,

```
> strike.outs.30
[1] 0.117649 0.302526 0.324135 0.185220 0.059535 0.010206 0.000729
> sum(strike.outs.30)
[1] 1
```





e) The inferences are

1) The PMF(a) is symmetrical since the chance of getting a strike out and and surviving a strike out are same.
2) The PMF(c) is left-skewed since the desired environment (getting strike out) is more than surviving it.
3) The PMF(d) is right-skewed since the desired environment (getting strike out) is less than surviving it.
4) Hence, the more the probability of desired event, the more the left-skewed the data is and vise versa.
5) The more the probability of desired event, the slower the CDF approaches 1 since all the winning probability are closer to 1 and # of trials and vise versa.

Problem#2

Suppose that 80% of the flights arrive on time. Using the binomial distribution,

a) What is the probability that four flights will arrive on time in the next 10 flights?
b) What is the probability that four or fewer flights will arrive on time in the next 10 flights?
c) Compute the probability distribution for the next 10 flights.
d) Show the PMF and the CDF for the next 10 flights.

Solution:

a) The probability that four flights arrive on time is 0.0055

```
> flights.on.time=4
> x1=dbinom(flights.on.time,size=total.flights,prob=probability)
> x1
[1] 0.005505024
>
```

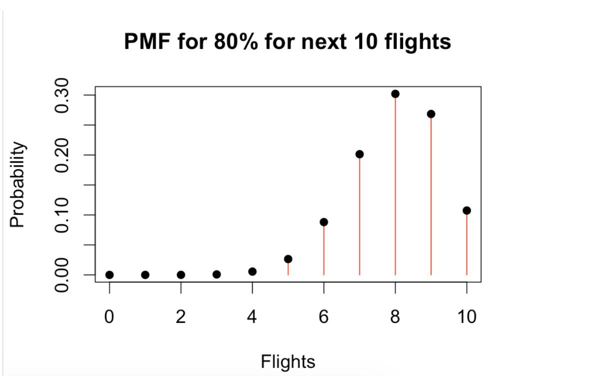
b) The probability for four or fewer flights arrive on time is 0.00637

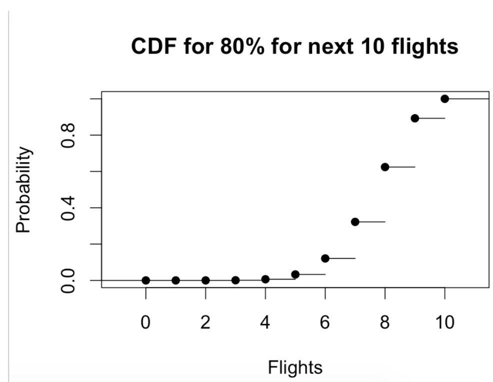
```
> x2=dbinom(flights.on.time,size=total.flights,prob=probability)
> x2
[1] 0.0000001024 0.0000040960 0.0000737280 0.0007864320 0.0055050240
> sum(x2)
[1] 0.006369382
> x2=pbinom(4,size=total.flights,prob=probability)
> x2
[1] 0.006369382
```

c) The PMF for next ten flights is shown

```
> flights.on.time=c(0:total.flights)
> x3=dbinom(flights.on.time,size=total.flights,prob=probability)
> x3
 [1] 0.0000001024 0.0000040960 0.0000737280 0.0007864320 0.0055050240 0.0264241152 0.0880803840 0.2013265920
 [9] 0.3019898880 0.2684354560 0.1073741824
```

d)The PMF and CDF plots are shown below





Problem#3

Suppose that on average 10 cars drive up to the teller window at your bank between 3 PM and 4 PM and the random variable has a Poisson distribution. During this time period,

a) What is the probability of serving exactly 3 cars?
b) What is the probability of serving at least 3 cars?
c) What is the probability of serving between 2 and 5 cars (inclusive)?
d) Calculate and plot the PMF for the first 20 cars.

Solution:

a) The probability of serving 3 cars exactly is

```
> cars.serving1=dpois(x,lambda=mu)
> cars.serving1
[1] 0.007566655
```

b) The probability of serving atleast 3 cars is

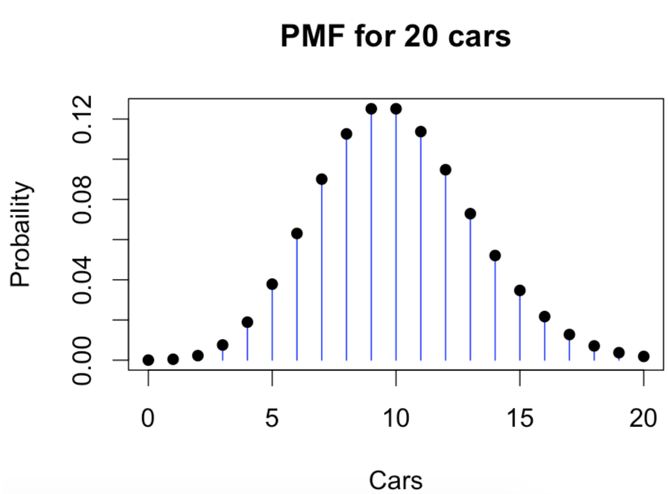
```
> cars.serving2=1-(dpois(0,lambda=mu)+dpois(1,lambda=mu)+dpois(2,lambda=mu))
> cars.serving2
[1] 0.9972306
```

c) The probability of serving between 2 and 5 cars (inclusive) is

```
> cars.serving3=ppois(x2,lambda=mu)-ppois(x1,lambda=mu)
> cars.serving3
[1] 0.06431657
```

d) The PMF for the first 20 cars is

```
> cars.serving
 [1] 4.539993e-05 4.539993e-04 2.269996e-03 7.566655e-03 1.891664e-02 3.783327e-02 6.305546e-02 9.007923e-02
 [9] 1.125990e-01 1.251100e-01 1.251100e-01 1.137364e-01 9.478033e-02 7.290795e-02 5.207710e-02 3.471807e-02
[17] 2.169879e-02 1.276400e-02 7.091109e-03 3.732163e-03 1.866081e-03
```



Problem#4

Suppose that your exams are graded using a uniform distribution between 60 and 100.

a) What is the probability of scoring i) 60? ii) 80? iii) 100?
b) What is the mean and standard deviation of this distribution?
c) What is the probability of getting a score of at most 70?
d) What is the probability of getting a score greater than 80 (use the lower.tail option)?
e) What is the probability of getting a score between 90 and 100 (inclusive)?

Solution:

a) The probability of scoring 60,80,100 is

```
> value
   60    80   100
0.025 0.025 0.025
```

b) The mean and standard deviation of the distribution are

```
> uniform.mean
[1] 80
> uniform.sd
[1] 11.54701
```

c) The probability of getting a score of at most 70 is

```
> x=70
> punif(x,min=minimum,max=maximum)
[1] 0.25
```

d) The probability of getting a score greater than 80 (use the lower.tail option) is

```
> x=80
> punif(x,min=minimum,max=maximum,lower.tail=FALSE)
[1] 0.5
```

e) The probability of getting a score between 90 and 100 (inclusive) is

```
> x1=90
> x2=100
> punif(x2,min=minimum,max=maximum)-punif(x1,min=minimum,max=maximum)
[1] 0.25
```

Problem#5

Suppose that visitors at a theme park spend an average of $100 on souvenirs. Assume that the money spent is normally distributed with a standard deviation of $10.

a) Show the PDF of this distribution covering the three standard deviations on either side of the mean.

b) What is the probability that a randomly selected visitor will spend more than $120?

c) What is the probability that a randomly selected visitor will spend between $80 and $90 (inclusive)?

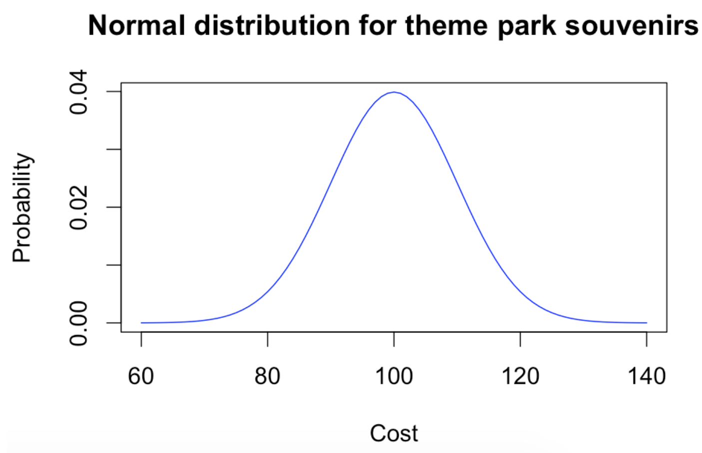
d) What are the probabilities of spending within one standard deviation, two standard deviations, and three standard deviations, respectively?

e) Between what two values will the middle 90% of the money spent will fall?

f) Show a plot for 10,000 visitors using the above distribution.

Solution:

a) The PDF for this distribution problem is



b) The probability that a randomly selected visitor will spend more than $120 is

```
> pdf=pnorm(120,mean=mu,sd=sigma,lower.tail=FALSE)
> pdf
[1] 0.02275013
```

c) The probability that a randomly selected visitor will spend between $80 and $90 (inclusive) is

```
> x1=80
> x2=90
> pdf=pnorm(x1,mean=mu,sd=sigma,lower.tail=FALSE)-pnorm(x2,mean=mu,sd=sigma,lower.tail=FALSE)
> pdf
[1] 0.1359051
```

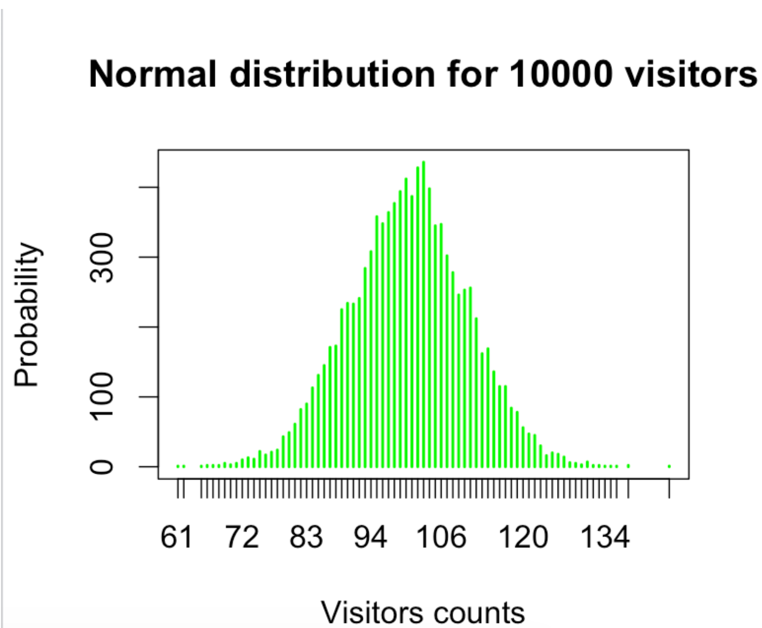
d) The probabilities of spending within one standard deviation, two standard deviations, and three standard deviations, respectively are

```
> sd1=c(mu-sigma,mu+sigma)
> sd2=c(mu-2*sigma,mu+2*sigma)
> sd3=c(mu-3*sigma,mu+3*sigma)
> pdf.sd1=pnorm(sd1[1],mean=mu,sd=sigma,lower.tail=FALSE)-pnorm(sd1[2],mean=mu,sd=sigma,lower.tail=FALSE)
> pdf.sd2=pnorm(sd2[1],mean=mu,sd=sigma,lower.tail=FALSE)-pnorm(sd2[2],mean=mu,sd=sigma,lower.tail=FALSE)
> pdf.sd3=pnorm(sd3[1],mean=mu,sd=sigma,lower.tail=FALSE)-pnorm(sd3[2],mean=mu,sd=sigma,lower.tail=FALSE)
> pdf.sd1
[1] 0.6826895
> pdf.sd2
[1] 0.9544997
> pdf.sd3
[1] 0.9973002
```

e) The middle 90% of the money spent will fall between the values 83.55 and 116.45

```
> middle.money.spent
[1]  83.55146 116.44854
> middle.90<-pnorm(middle.money.spent[2],mean=mu,sd=sigma,lower.tail=TRUE)-pnorm(middle.money.spent[1],mean=mu,sd=sig
ma,lower.tail=TRUE)
> middle.90
[1] 0.9
```

f) The plot 10,000 visitors using the distribution is



Problem#6

Suppose your cell phone provider's customer support receives calls at the rate of 18 per hour.

a) What is the probability that the next call will arrive within 2 minutes?
b) What is the probability that the next call will arrive within 5 minutes?
c) What is the probability that the next call will arrive between 2 minutes and 5 minutes (inclusive)?
d) Show the CDF of this distribution.

Solution:

a) The probability that the next call will arrive within 2 minutes is

```
> x=2
> pdf1=pexp(x/60,rate=lambda)
> pdf1
[1] 0.4511884
```

b) The probability that the next call will arrive within 5 minutes

```
> x=5
> pdf2=pexp(x/60,rate=lambda)
> pdf2
[1] 0.7768698
```

c) The probability that the next call will arrive between 2 minutes and 5 minutes (inclusive) is

```
> pdf3=pdf2-pdf1
> pdf3
[1] 0.3256815
```

d) The CDF of this distribution is