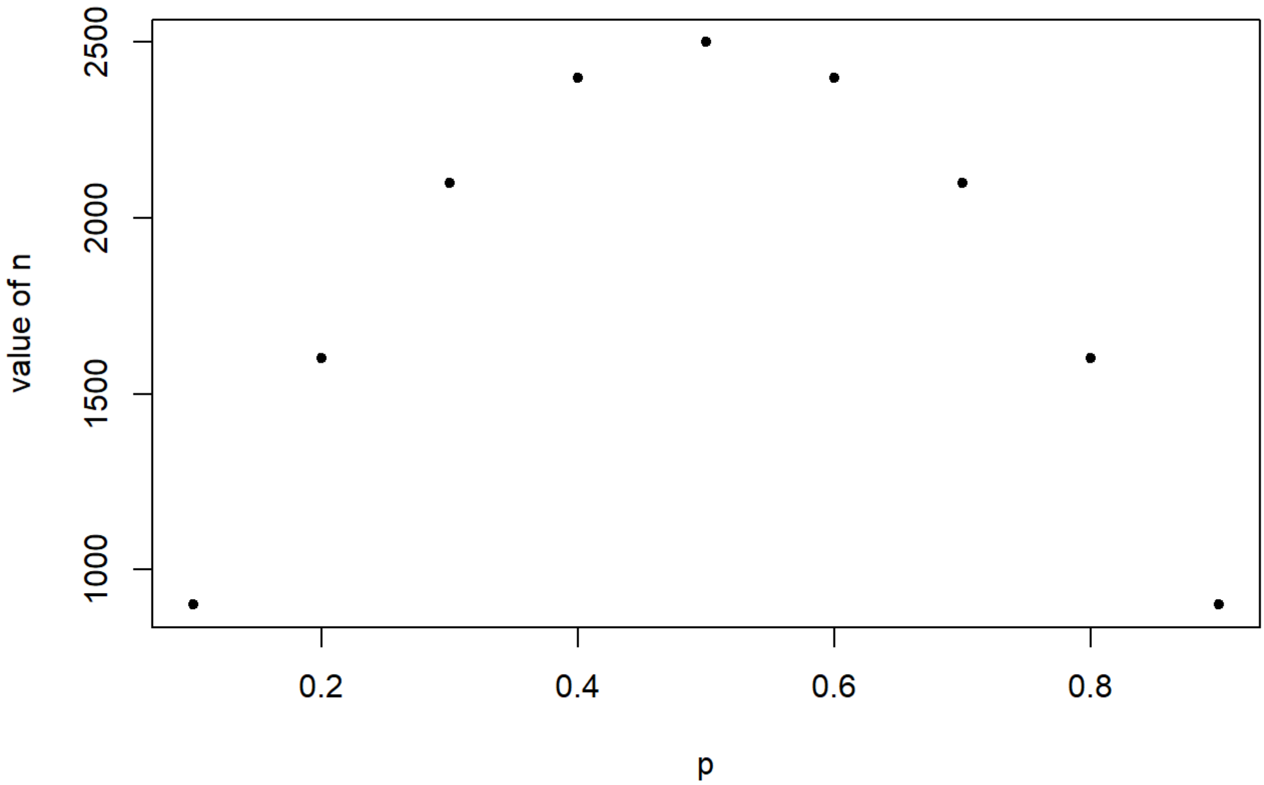
# PSWR

Sucheta

30/11/2021

```r
#Solution 1.a)

var_phat = function(p,n){
  return(p*(1-p)/n)
}

sd_phat = function(p,n){
  return((p*(1-p)/n)^0.5)
}

#solution 1.b)

#SD(phat) = 0.01 = (p(1-p)/n)^0.5
#So, (0.01)^2 = p(1-p)/n
#Therefore, n = p(1-p)/(0.01)^2

sam_n = function(p){
  return(p*(1-p)/(0.01)^2)
}
prob = c(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9)
n_prob = rep(NA,length(prob))
for (i in 1:length(prob)){
  n_prob[i]=sam_n(prob[i])
}
plot(prob,n_prob,ylab='value of n', xlab='p',pch=20)
```




```
#solution 1.c)

#phat depends on the occurrence and non occurrence of an event A. Let 1 denote the
#occurrence(probability = p) and 0 denote the non occurrence(probability = 1-p).
x = c(0,1)

phat = function(sim_size,p,n){
  phat_sam = rep(NA,sim_size)
  for (i in 1:sim_size){
    z = sample(x,n,replace=TRUE,prob=c(1-p,p))
    count = sum(z==1)
    phat_sam[i]=count/n
  }
  return(phat_sam)
}

#Solution c.i)

n=500
sim_size = 1000
p=c(0.01,0.1, 0.25, 0.5, 0.75, 0.9, 0.99)
sd1 = rep(NA,length(p))
for (i in 1:length(p)){
  phat1 = phat(sim_size,p[i],n)
  sd1[i]=sd(phat1)
}
sd1
```

```
## [1] 0.004598524 0.013361047 0.019493896 0.022220503 0.019420617 0.013556824
## [7] 0.004459932
```

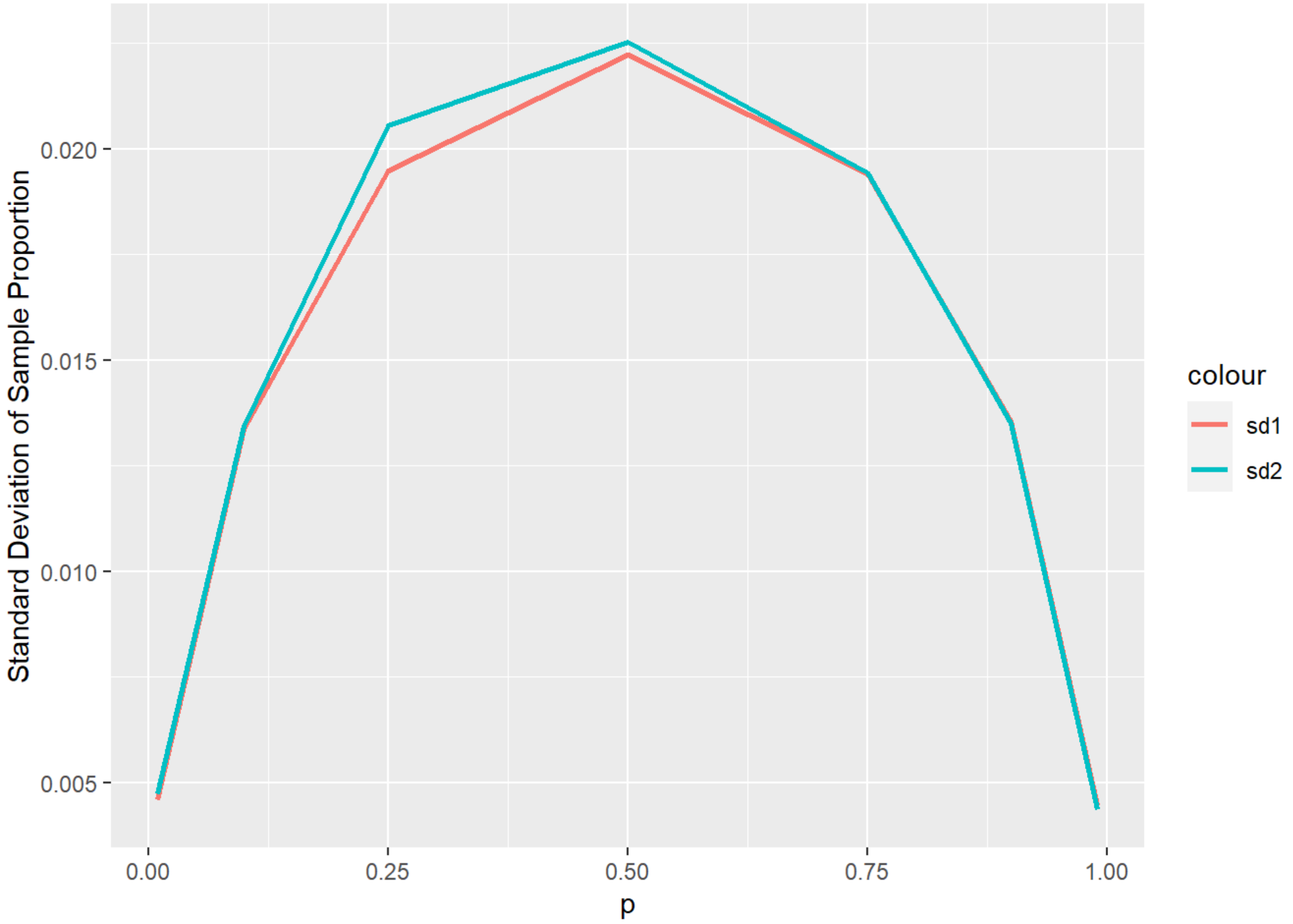
```
#Solution1.c.ii)
n_sd = function(p,sd){
  return(p*(1-p)/(sd)^2)
}

sd2 = rep(NA,length(p))
for (i in 1:length(p)){
  n2 = n_sd(p[i],sd1[i])
  phat2 = phat(sim_size,p[i],n2)
  sd2[i]=sd(phat2)
}
sd2
```

```
## [1] 0.004742179 0.013436129 0.020559681 0.022527167 0.019427393 0.013510451
## [7] 0.004351914
```

```
df = data.frame(p,sd1)
cbind(df,sd2)
```

```
##      p         sd1         sd2
## 1 0.01 0.004598524 0.004742179
## 2 0.10 0.013361047 0.013436129
## 3 0.25 0.019493896 0.020559681
## 4 0.50 0.022220503 0.022527167
## 5 0.75 0.019420617 0.019427393
## 6 0.90 0.013556824 0.013510451
## 7 0.99 0.004459932 0.004351914
```

```
library(ggplot2)
ggplot(df,aes(x=p))+geom_line(aes(y=sd1,colour='sd1'),size=1)+
  geom_line(aes(y=sd2,colour='sd2'),size=1)+
  ylab("Standard Deviation of Sample Proportion")
```




```
#Solution 2.b and 2.c)

n = c(100,500,1000)
mean=c()
var = c()
diff_mean = c()
diff_var=c()
set.seed(11)
df_pois = function(lambda){
  for (j in n){
    pois = rpois(j,lambda)
    mean = append(mean,mean(pois))
    var = append(var,var(pois))
    diff_mean = append(diff_mean,lambda-mean(pois))
    diff_var=append(diff_var,lambda-var(pois))
  }
  return(data.frame(n,mean,var,diff_mean,diff_var))
}

#lambda = 10
df_pois(10)
```

```
##      n   mean       var diff_mean   diff_var
## 1  100  9.540  8.473131     0.460  1.5268687
## 2  500 10.116 10.736016    -0.116 -0.7360160
## 3 1000  9.943  9.577328     0.057  0.4226717
```

```
#As the value of n is increasing the difference between the sample mean and true
#mean and the sample variance and true variance is decreasing.
#we can see as we are double the value of n, our error is getting almost halved.
true_mean = 10
true_var = 10
mean_sam = mean(df_pois(10)$mean)
mean_var = mean(df_pois(10)$var)
print(paste("The expectation of sample mean is : ",mean_sam))
```

```
## [1] "The expectation of sample mean is :  9.90966666666667"
```

```
print(paste("The expectation of sample variance is : ",mean_var))
```

```
## [1] "The expectation of sample variance is :  10.3539952196078"
```

```
#The expectation of sample mean and variance, differ by 0.1 and
#0.35 respectively from the true mean = 10

#lambda = 20
df_pois(20)
```

```
##      n   mean      var diff_mean   diff_var
## 1  100 19.680 20.74505     0.320 -0.7450505
## 2  500 19.834 20.19884     0.166 -0.1988417
## 3 1000 20.134 20.67272    -0.134 -0.6727167
```

```
#As the value of n increases the difference between the sample mean and true
#mean and the sample variance and true variance decreases at first step but then
#again increases. But overall the difference is reduced on increasing n.
true_mean = 20
true_var = 20
mean_sam = mean(df_pois(20)$mean)
mean_var = mean(df_pois(20)$var)
print(paste("The expectation of sample mean is : ",mean_sam))
```

```
## [1] "The expectation of sample mean is :  19.9036666666667"
```

```
print(paste("The expectation of sample variance is : ",mean_var))
```

```
## [1] "The expectation of sample variance is :  19.737532153639"
```

```
#The expectation of sample mean and variance, differ by 0.1 and
#0.36 respectively from the true mean = 20


#lambda = 50
df_pois(50)
```

```
##      n   mean      var diff_mean   diff_var
## 1  100 49.530 45.94859     0.470  4.0514141
## 2  500 50.392 49.87408    -0.392  0.1259158
## 3 1000 50.002 51.04304    -0.002 -1.0430390
```

```
#As the value of n increases the difference between the sample mean and true
#mean and the sample variance and true variance decreases initially. But, then
#it increases very largely, indicating not to increase n for lambda=50.

true_mean = 50
true_var = 50
mean_sam = mean(df_pois(50)$mean)
mean_var = mean(df_pois(50)$var)
print(paste("The expectation of sample mean is : ",mean_sam))
```

```
## [1] "The expectation of sample mean is :  49.7466666666667"
```

```
print(paste("The expectation of sample variance is : ",mean_var))
```

```
## [1] "The expectation of sample variance is :  53.1424211962615"
```

```
#The expectation of sample mean and variance, differ by 0.26 and
#3.14 respectively from the true mean = 50


#We can conclude that for larger values of lambda, increasing the value of n
#increases the difference between the sample mean and true mean and the
#sample variance and true variance. We also observe, The expectation of sample
#mean and variance are close to lambda.
```

2.a) $X \sim \text{Poisson}(\lambda)$

$E(X) = \lambda, \quad Var(X) = \lambda$

let $X_1, X_2, \ldots X_n$ be iid $X$. $\Rightarrow X_i \sim \text{Poisson}(\lambda) \; \forall\, 1 \le i \le n$

Sample mean $= \bar{X} = \dfrac{X_1 + X_2 + \ldots X_n}{n}$

$$E(\bar{X}) = E\left[\frac{X_1 + X_2 + \ldots X_n}{n}\right]$$

$$= \frac{1}{n}\left\{E[X_1] + E[X_2] + \ldots E[X_n]\right\} \quad \text{(Expectation is linear)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{\cancel{n}} \cdot \cancel{n}\, E[X_1] \quad (X_1, X_2, \ldots X_n \text{ are iid})$$

$$= E[X_1] = \lambda \quad (X_1 \text{ is iid } X)$$

$\therefore E[\bar{X}] = \lambda$

$\Rightarrow$ sample mean is an unbiased estimator of $\lambda$.

Sample variances $= S^2 = \dfrac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots (X_n - \bar{X})^2}{n-1}$

$$E[S^2] = E\left[\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots (X_n - \bar{X})^2}{n-1}\right]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] \quad \text{(Expectation is linear)}$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E[X_i^2 + \bar{X}^2 - 2X_i\bar{X}]$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} E[X_i^2] + \sum_{i=1}^{n} E[\bar{X}^2] - 2\sum_{i=1}^{n} E[X_i\bar{X}] \right\}$$

(expectation is linear)

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} E[X_i^2] + nE[\bar{X}^2] - 2E\left[\bar{X} \sum_{i=1}^{n} X_i\right] \right\}$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} E[X_i^2] + nE[\bar{X}^2] - 2E[\bar{X} \cdot n\bar{X}] \right\}$$

$\left(\because \bar{X} = \sum X_i / n\right)$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} E[X_i^2] + nE[\bar{X}^2] - 2nE[\bar{X}^2] \right\}$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} E[X_i^2] - nE[\bar{X}^2] \right\}$$

$Var(X_i) = \lambda \Rightarrow E[(X_i - \lambda)^2] = \lambda \quad \forall 1 \le i \le n$

$\Rightarrow E[(X_i)^2] - (E(X_i))^2 = \lambda$

$\Rightarrow E[X_i^2] = \lambda + \lambda^2$

Also, $Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots X_n}{n}\right)$

$= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)$ $\left(\begin{array}{l} X_i \text{ are iid} \\ \Rightarrow Cov(X_i, X_j) = 0 \quad \forall i \neq j \end{array}\right)$

$= \frac{n}{n^2} Var(X_i) = \frac{Var(X_1)}{n} = \frac{\lambda}{n}$

$$Var(\bar{X}) = E[(\bar{X} - E(\bar{X}))^2]$$

$$= E[\bar{X}^2] - (E[\bar{X}])^2$$

$$\Rightarrow \frac{\lambda}{n} = E[\bar{X}^2] - \lambda^2$$

$$\Rightarrow E[\bar{X}^2] = \frac{\lambda}{n} + \lambda^2.$$

$$\therefore E[S^2] = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\lambda + \lambda^2) - n\left(\frac{\lambda}{n} + \lambda^2\right)\right]$$

$$= \frac{1}{n-1}\left[n(\lambda + \lambda^2) - (\lambda + n\lambda^2)\right]$$

$$= \frac{1}{n-1}\left(n\lambda + \cancel{n\lambda^2} - \lambda - \cancel{n\lambda^2}\right)$$

$$= \frac{\lambda\cancel{(n-1)}}{\cancel{n-1}} = \lambda$$

$$\Rightarrow E[S^2] = \lambda = Var(X)$$

$\Rightarrow S^2$ is an unbiased estimator of $\lambda$.