18.650 Problem Set 2 Spring 2017
Statistics for Applications
Due Date: Fri 2/24/2017, prior to 4:00pm
Where: Electronically to Stellar website (preferred)
or Problem Set Box (outside 4-174)

Problems from John A. Rice, Third Edition. [*Chapter.Section.Problem*]
Total points: 63 = 16 + 16 + 19 + 12

1. (See Problem 8.10.3 of Rice). As stated by Rice (p. 313)

   [16 points: 4 points per part]

   > One of the earliest applications of the Poisson distribution was made by Student (1907) in studying errors madein counting yeast cells or blood corpuscles with a haemacytometer. In this study yeast cells were killed and mixed iwth water and gelatin; the mixture was then spread on a glass and allowed to cool. Four different concentrations were used. Couns were made on 400 squares, and the data are summarized [in table on p. 313].

   The following R code creates variables corresponding to the columns of the table:

```
> x.NumberofCells<-c(0:12)
> Counts.conc1<-c(213,128,37,18,3,1,0,0,0,0,0,0,0)
> Counts.conc2<-c(103,143,98,42,8,4,2,0,0,0,0,0,0)
> Counts.conc3<- c(75,103,121,54,30,13,2,1,0,1,0,0,0)
> Counts.conc4<- c(0,20,43,53,86,70,54,37,18,10,5,2,2)
```

   (a). For a given Concentration level, let $X_1, X_2, \ldots, X_{400}$ denote the counts of the 400 squares. If the random variables X_i were i.i.d. Poisson(λ) random variables then

   $$E[X_i] = Var[X_i] = \lambda$$

   Using method-of-moments, we can compute two estimates of λ for each concentration:

   $$\hat{\lambda}_A = \hat{\mu}_1 = \overline{X} = \sum_{i=1}^n X_i/n$$

   $$\hat{\lambda}_B = \hat{\mu}_2 - (\hat{\mu}_1)^2 = (\sum_{i=1}^n X_i^2/n) - \overline{X}^2$$

   The following R commands compute the first and second sample moments of the number of cells per square for the 400 squares of Concentration 1.

```
> mu1hat.conc1=sum(Counts.conc1*x.NumberofCells)/sum(Counts.conc1)
> mu2hat.conc1=sum(Counts.conc1*(x.NumberofCells)^2)/sum(Counts.conc1)
```

Compute these estimates for each of the four different concentrations, completing the following table:

Estimate	Conc. 1	Conc. 2	Conc. 3	Conc. 4
$\hat{\lambda}_A$				
$\hat{\lambda}_B$				

```
> mu1hat.conc1=sum(Counts.conc1*x.NumberofCells)/sum(Counts.conc1)
> mu2hat.conc1=sum(Counts.conc1*(x.NumberofCells)^2)/sum(Counts.conc1)
> lambdahatA.conc1=mu1hat.conc1
> lambdahatB.conc1=mu2hat.conc1-mu1hat.conc1^2
> mu1hat.conc2=sum(Counts.conc2*x.NumberofCells)/sum(Counts.conc2)
> mu2hat.conc2=sum(Counts.conc2*(x.NumberofCells)^2)/sum(Counts.conc2)
> lambdahatA.conc2=mu1hat.conc2
> lambdahatB.conc2=mu2hat.conc2-mu1hat.conc2^2
> mu1hat.conc3=sum(Counts.conc3*x.NumberofCells)/sum(Counts.conc3)
> mu2hat.conc3=sum(Counts.conc3*(x.NumberofCells)^2)/sum(Counts.conc3)
> lambdahatA.conc3=mu1hat.conc3
> lambdahatB.conc3=mu2hat.conc3-mu1hat.conc3^2
> mu1hat.conc4=sum(Counts.conc4*x.NumberofCells)/sum(Counts.conc4)
> mu2hat.conc4=sum(Counts.conc4*(x.NumberofCells)^2)/sum(Counts.conc4)
> lambdahatA.conc4=mu1hat.conc4
> lambdahatB.conc4=mu2hat.conc4-mu1hat.conc4^2
> table0<-cbind(
+     c(lambdahatA.conc1, lambdahatB.conc1),
+     c(lambdahatA.conc2, lambdahatB.conc2),
+     c(lambdahatA.conc3, lambdahatB.conc3),
+     c(lambdahatA.conc4, lambdahatB.conc4))
> dimnames(table0)<-list(
+     c("lambdahatA","lambdahatB"),
+     c("Conc1","Conc2","Conc3","Conc4"))
> print(table0)
```

```
               Conc1    Conc2 Conc3  Conc4
lambdahatA 0.6825000 1.322500  1.80 4.6800
lambdahatB 0.8116938 1.283494  1.96 4.4576
```

(b). If the true underlying distribution is a Poisson distribution with fixed λ, compute the bias of each estimate as a function of λ and n (the sample size), i.e.,

$$Bias(\hat{\lambda}_A) = E[\hat{\lambda}_A] - \lambda$$
$$Bias(\hat{\lambda}_B) = E[\hat{\lambda}_B] - \lambda$$

Do the estimates differ in terms of their bias? If so, is this bias apparent in the table of part (a)?

$$E[\hat{\lambda}_A] = E[\overline{X}] = E[X_1] = \lambda$$

so $Bias(\hat{\lambda}_A) = \lambda - \lambda = 0$, $\hat{\lambda}_A$ is unbiased.

$$
\begin{array}{rcl}
E[\hat{\lambda}_B] & = & E[(\hat{\mu}_2 - \hat{\mu}_1)^2] \\
 & = & E[\frac{1}{n}\sum_1^n X_i^2 - (\frac{1}{n}\overline{X})^2] \\
 & = & \frac{1}{n}E[\sum_1^n (X_i^2 - \overline{X})^2] \\
 & = & \frac{1}{n}E[(n-1)S^2] \\
 & = & \frac{n-1}{n}E[S^2] \\
 & = & \frac{n-1}{n}Var[X_1] \\
 & = & \frac{n-1}{n}\lambda]]
\end{array}
$$

So $Bias(\hat{\lambda}_B) = \frac{n-1}{n}\lambda - \lambda = -\frac{\lambda}{n}$, so $\hat{\lambda}_B$ is biased.

The estimates differ in terms of their bias, $\hat{\lambda}_B$ is always biased to underestimate λ while $\hat{\lambda}_A$ is unbiased.

The magnitude of the bias is $0.25\%\lambda = \frac{1}{400}\lambda$. In the table from part (a), this bias is not readily apparent. The estimates $\hat{\lambda}_B$ are larger than those for $\hat{\lambda}_A$ for two concentrations (the first and the third); if the bias were readily apparent we might expect to see $\hat{\lambda}_B$ systematically lower than $\hat{\lambda}_A$ at all concentrations.

(c). Conduct a simulation of the sampling distributions of $\hat{\lambda}_A$ and $\hat{\lambda}_B$ to evaluate the bias and variability of the estimates.

Assume the underlying data follow a Poisson distribution with $\lambda = lambdahatA.conc1$.

The following R code should work:

```
> # Simulate n=400 samples with lambdahatA.conc1, with n.simulations=1000
> sample.n=400
> n.simulations=1000
> set.seed(1)
> xmat=matrix(rpois(sample.n*n.simulations, lambda=lambdahatA.conc1),
+             nrow=sample.n, ncol=n.simulations)
> simulated.hatmu1=colMeans(xmat)
> simulated.hatmu2=colMeans(xmat^2)
> simulated.lambdahatA=simulated.hatmu1
> simulated.lambdahatB=simulated.hatmu2 - (simulated.hatmu1)^2
> # Graph histograms of the sampling distributions
> par(mcol=c(2,1))
> hist(simulated.lambdahatA, breaks=100)
> hist(simulated.lambdahatB, breaks=100)
> # Compute means and standard deviations of the sampling distributions
> #   mean(simulated.lambdahatA)
> #   mean(simulated.lambdahatB)
> #   sqrt(var(simulated.lambdahatA))
> #   sqrt(var(simulated.lambdahatB))
```

Based on the results of this simulation compare $\hat{\lambda}_A$ and $\hat{\lambda}_B$ in terms of their bias and variability. Which estimate has a sampling distribution which is closer to the true Poisson parameter?

The bias of each estimate are estimated by

```
> bias.A=mean(simulated.lambdahatA)- lambdahatA.conc1 # Bias of lambdahatA
> bias.B=mean(simulated.lambdahatB)- lambdahatA.conc1 # Bias of lambdahatB
> bias.A
```

```
[1] 0.00081
```

```
> bias.B
```

```
[1] -0.001968837
```

Note that the average bias of both are small. We would expect the average bias of $\hat{\lambda}_B$ to be less than zero and that for $\hat{\lambda}_A$ to be consistent with having a theoretical value of zero. Results will vary depending on the random numbers generated. Because the bias of $\hat{\lambda}_B$ is very small in magnitude, its simulated average could be smaller than that of $\hat{\lambda}_A$.

The variability of the estimates can be compared using estimates of their standard deviations:

```
>     sqrt(var(simulated.lambdahatA))
```

```
[1] 0.04016086
```

```
>     sqrt(var(simulated.lambdahatB))
```

```
[1] 0.05976472
```

These are the standard errors of estimates (i.e., estimates of the standard deviation of the sampling distribution of the estimates). $\hat{\lambda}_A$ has a lower standard error.

To assess which estimate has a sampling distribution closer to the true parameter, we can compute their mean-squared errors which is the sum of the variance plus the squared bias.

```
> MSE.A= var(simulated.lambdahatA) + (bias.A)^2
> MSE.B= var(simulated.lambdahatB) + (bias.B)^2
> MSE.A
```

```
[1] 0.00161355
```

```
> MSE.B
```

```
[1] 0.003575698
```

$\hat{\lambda}_A$ has the lower mean-square-error.

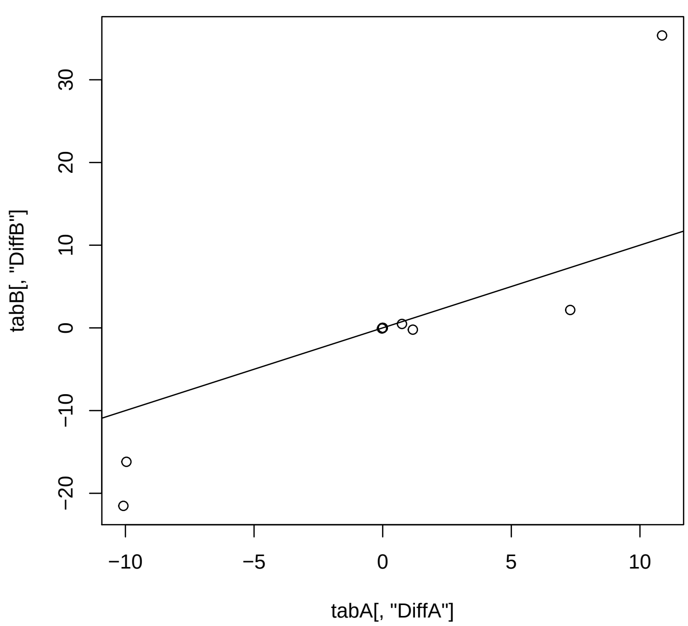
When comparing mean-squared-errors, it is convenient to compute the root-mean-squared errors (RMSE) which is on the same scale as the estimate

```
> sqrt(MSE.A)

[1] 0.04016902

> sqrt(MSE.B)

[1] 0.05979714
```

(d). Using $\hat{\lambda}_A$, compute expected counts for concentration 1, and compute the difference between the observed and expected counts. Repeat with $\hat{\lambda}_B$. Plot the differences for one estimate versus the other.

You can use the following R code:

```
> ExpectedCountsA.conc1= 400*dpois(x.NumberofCells,lambda=lambdahatA.conc1)
> tabA<-cbind(Counts.conc1, ExpectedCountsA.conc1, DiffA=Counts.conc1-ExpectedCountsA.conc1)
> ExpectedCountsB.conc1= 400*dpois(x.NumberofCells,lambda=lambdahatB.conc1)
> tabB<-cbind(Counts.conc1, ExpectedCountsB.conc1, DiffB=Counts.conc1-ExpectedCountsB.conc1)
> par(mfcol=c(1,1))
> plot(tabA[,"DiffA"], tabB[,"DiffB"])
> abline(a=0,b=1)
```



Does this analysis suggest that $\hat{\lambda}_B$ is better than $\hat{\lambda}_A$ in terms of fitting the observed data?

Points close to the line with slope equal to 1 correspond to observed counts that have comparable estimates of expected counts for the two fits. When the difference is positive, points above the line corresond to cases where the fit with $\hat{\lambda}_B$ is worse than $\hat{\lambda}_A$. When the difference is negative, points below the line correspond to cases where $\hat{\lambda}_B$ is worse than $\hat{\lambda}_A$. This plot has these features so it suggests that $\hat{\lambda}_B$ fits worse than lambdahat. Later in the course we will study goodness-of-fit tests which use these differences to measure quality of fit.

2. Problem 8.10.27; instead of 5 components, suppose there are 6 and that the first one fails at 45 days.

   [16 points total: 4 points per part] Suppose that certain electronic components have lifetimes that are exponentially distributed

   $$f(t \mid \tau) = (1/\tau)exp(-t/\tau),\ t \geq 0.$$

   Five new components are put on test. The first one fails at 45 days, and no further observations are recorded.

   By the Hint (Example A of Section 3.7), let $T_1, T_2, \ldots, T_n$ be the time until failure of n components (we shall set $n = 6$ below).

   These random variables are i.i.d. with cumulative distribution function

   $$F_T(t) = 1 - exp(-t/\tau)$$

   Let $T_{MIN} = min(T_1, T_2, \ldots, T_n)$ be the shortest time to failure. (If a system operates with components 1-n connected in a series, then T_{MIN} is the time until failure of the sytem.)

   The cumulative distribution function of T_{Min}, $F_{T_{MIN}}(t)$ must satisfy:

   $$\begin{aligned} [1 - F_{T_{MIN}}(t)] &= P(T_{MIN} > t) \\ &= P(T_1 > t, T_2 > t, \ldots, T_n > t) = \prod_{i=1}^{n} P(T_i > t) \\ &= \prod_{j=1}^{n}[1 - F_T(t)] = [1 - F_T(t)]^n \end{aligned}$$

   It follows that the probability density function of T_{MIN} is

   $$\begin{aligned} f_{T_{MIN}}(t) &= -\frac{d}{dt}[1 - F_{T_{MIN}}(t)] \\ &= n[1 - F_T(t)]^{n-1}\frac{d}{dt}[F_T(t)] \\ &= n[exp(-t/\tau)]^{n-1}(1/\tau)exp(-t/\tau) \\ &= (n/\tau)exp[-t(n/\tau)] \end{aligned}$$

   That is, $T_{MIN} \sim Exponential(rate = n/\tau)$.

   (a). What is the likelihood function of τ? The data consists of $T_{MIN} = 45$ with $n = 6$, so

   $$\begin{aligned} lik(\tau) &= f_{T_{MIN}}(t = 45) = (n/\tau)exp[-t(n/\tau)]|_{t=45,n=6} \\ &= (6/\tau)exp[-45(6/\tau)] = (6/\tau)exp[-270/\tau] \end{aligned}$$

   (b). What is the mle of τ?

   The mle maximizes $lik(\tau)$ and $\ell(\tau) = \log lik(\tau)$, which is the solution to

$$
\begin{aligned}
0 = \ell'(\tau) &= \frac{d}{d\tau}\ell(\tau) \\
&= \frac{d}{d\tau}[\ln(n/\tau)] + \frac{d}{d\tau}[-t(n/\tau)] \\
&= -\frac{1}{\tau} - tn(\frac{1}{\tau})^2(-1) \\
\Longrightarrow \hat{\tau} &= tn = T_{MIN} \times n = (45 \times 6) = 270.
\end{aligned}
$$

(c). What is the sampling distribution of the mle?

The mle is $\hat{\tau} = T_{MIN} \times n$ which is n times the minimum of n i.i.d. $Exponential(rate = 1/\tau)$ random variables. The distribution of T_{MIN} is $Exponential(rate = 6/\tau)$ The mle is 6 times the minimum of n i.i.d. exponential random variables.

$$
\begin{aligned}
1 - F_{\hat{\tau}}(u) &= P(\hat{\tau} \geq u) \\
&= P(T_{MIN} \times n \geq u) = P(T_{MIN} \geq u/n) \\
&= [1 - F_T(u/n)]^n \\
&= [exp(-(u/n)/\tau]^n \\
&= [exp(-u/\tau)]
\end{aligned}
$$

So $F_{\hat{\tau}}(u) = 1 - exp(-u/\tau)$ which is the cdf of an $Exponential(rate = 1/\tau)$ random variable.

(d). What is the standard error of the mle?

The standard error of the mle is the formula for the square root of the variance of the mle, plugging in the mle estimate for the value of the true parameter

$$
\begin{aligned}
StError(\hat{\tau}) &= \sqrt{Var(\hat{\tau})}|_{\tau=\hat{\tau}} \\
&= \sqrt{\tau^2}|_{\tau=\hat{\tau}} = \hat{\tau} = 270.
\end{aligned}
$$

The variance formula follows from p. A2 Appendix A for

$Gamma(\alpha = 1, \lambda = 1/\tau)$, which is $Exponential(rate = 1/\tau)$.

3. Problem 8.10.53

   [4 points each for parts (a), (b), (d); 7 points for part (c)]

   Let $X_1, \ldots, X_n$ be i.i.d. uniform on $[0, \theta]$

   (a). Find the method of moments estimate of θ and its mean and variance.

   The first moment of each X_i is

$$
\begin{aligned}
\mu_1 &= E[X_i \mid \theta] = \int_0^\theta xf(x \mid \theta)dx \\
&= \int_0^\theta x\frac{1}{\theta}dx = \frac{1}{\theta}[x^2/2]|_{x=0}^\theta \\
&= \frac{1}{\theta}[\theta^2/2] = \theta/2
\end{aligned}
$$

   The method of moments estimate solves

$$\overline{x} = \frac{1}{n}\sum_1^n x_i = \mu_1 = \theta/2$$

   So

$$\hat{\theta}_{MOM} = 2\overline{x}.$$

   The mean and variance of $\hat{\theta}_{MOM} = 2\overline{x}$ can be computed from the mean and variance of the i.i.d. X_i:

$$
\begin{aligned}
\mu_1 &= E[X_i] = \theta/2 \text{ (shown above)} \\
\sigma^2 &= \mu_2 - (\mu_1)^2 = E[X_i^2] - E[X_i]^2 \\
&= \int_0^\theta x^2 f(x \mid \theta)dx - [\theta/2]^2 \\
&= \tfrac{1}{\theta}[x^3/3]|_{x=0}^{x=\theta} - [\theta/2]^2 \\
&= [\theta^2/3] - [\theta^2/4] = \theta^2/12
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E[\hat{\theta}_{MOM}] &= E[2\overline{X}] = 2 \times \tfrac{1}{n}\sum_1^n E[X_i] = 2 \times \tfrac{n}{n}\theta = \theta \\
Var[\hat{\theta}_{MOM}] &= Var[2\overline{X}] = 4 \times Var[\overline{X}] \\
&= 4 \times Var[X_i]/n \\
&= 4 \times (\theta^2/12)/n = \theta^2/(3n).
\end{aligned}
$$

(b). Find the mle of θ.

The mle $\hat{\theta}_{MLE}$ maximizes the likelihood function

$$
\begin{aligned}
lik(\theta) &= f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^n [\tfrac{1}{\theta} \times \mathbf{1}_{[0,\theta]}(x_i)] \\
&= \tfrac{1}{\theta^n}\mathbf{1}_{[0,\theta]}(max(x_1, \ldots, x_n)) \\
&= \tfrac{1}{\theta^n}\mathbf{1}_{[max(x_1,\ldots,x_n),\infty)}(\theta)
\end{aligned}
$$

Where $\mathbf{1}_{[a,b]}(x) = \begin{cases} 1, & if \quad x \in [a,b] \\ 0, & if \quad x \notin [a,b] \end{cases}$

The likelihood function is maximized by the minimizing θ. Since $\theta \geq max(x_1, \ldots, x_n)$ the likelihood is maximized with

$$\hat{\theta} = max(x_1, \ldots, x_n)$$

(c). Find the probability density of the mle, and calculate its mean and variance. Compare the variance, the bias, and the mean squared error to those of the method of moments estimate.

The probability density of the mle is the probability density of the maximum member of the sample $X_1, \ldots, X_n$

We compute the cdf (cumulative distribution function) of the mle first.

$$
\begin{aligned}
F_{\hat{\theta}_{MLE}}(t) &= P(\hat{\theta}_{MLE} \leq t) \\
&= P(max(X_1, \ldots, X_n) \leq t) = P(X_1 \leq t, X_2 \leq t, \ldots X_n \leq t) \\
&= [P(X_i \leq t)]^n = [\tfrac{t}{\theta}]^n
\end{aligned}
$$

The density of $\hat{\theta}_{MLE}$ is the derivative of the cdf:

$$
\begin{aligned}
f_{\hat{\theta}_{MLE}}(t) &= \tfrac{d}{dt}F_{\hat{\theta}_{MLE}}(t) = n[\tfrac{t}{\theta}]^{n-1}[\tfrac{1}{\theta}] \\
&= \frac{nt^{n-1}}{\theta^n}, \quad 0 < t < \theta
\end{aligned}
$$

The mean and variance of $\hat{\theta}_{MLE}$ can be computed directly

$$
\begin{aligned}
E[\hat{\theta}_{MLE}] &= \textstyle\int_0^\theta [t f_{\hat{\theta}_{MLE}}(t)]dt \\
&= \int_0^\theta [t\frac{nt^{n-1}}{\theta^n}]dt \\
&= \frac{n}{\theta^n}[\frac{t^{n+1}}{n+1}]|_{t=0}^{t=\theta} \\
&= \frac{n}{\theta^n}[\frac{\theta^{n+1}}{n+1}] \\
&= \frac{n}{n+1}\theta \\
Var[\hat{\theta}_{MLE}] &= E[\hat{\theta}^2_{MLE}] - (E[\hat{\theta}_{MLE}])^2 \\
&= \textstyle\int_0^\theta [t^2 f_{\hat{\theta}_{MLE}}(t)]dt - (E[\hat{\theta}_{MLE}])^2 \\
&= \int_0^\theta [t^2\frac{nt^{n-1}}{\theta^n}]dt \\
&= \frac{n}{\theta^n}[\frac{t^{n+2}}{n+2}]|_{t=0}^{t=\theta} - (E[\hat{\theta}_{MLE}])^2 \\
&= \frac{n}{\theta^n}[\frac{\theta^{n+2}}{n+2}] - (E[\hat{\theta}_{MLE}])^2 \\
&= \frac{n}{n+2}\theta^2 - (E[\hat{\theta}_{MLE}])^2 \\
&= \frac{n}{n+2}\theta^2 - [\frac{n}{n+1}\theta]^2 \\
&= \theta^2 \times [\tfrac{n}{n+2} - \tfrac{n^2}{(n+1)^2}] \\
&= \theta^2 \times \tfrac{n}{(n+2)(n+1)^2}
\end{aligned}
$$

Now, to compare the variance, the bias, and the mean squared error to those of the method of moments estimate, we apply the formulas:

$$
\begin{aligned}
Bias(\hat{\theta}) &= E[\hat{\theta}] - \theta \\
MSE(\hat{\theta}) &= Var(\hat{\theta}) + [Bias(\hat{\theta})]^2
\end{aligned}
$$

So, we can compare:

$$
\begin{aligned}
Var(\hat{\theta}_{MLE}) &= \theta^2 \times \tfrac{n}{(n+2)(n+1)^2} \\
Bias(\hat{\theta}_{MLE}) &= E[\hat{\theta}_{MLE}] - \theta = -\tfrac{1}{n+1}\theta \\
MSE(\hat{\theta}_{MLE}) &= Var(\hat{\theta}_{MLE}) + [Bias(\hat{\theta}_{MLE})]^2 \\
&= \theta^2 \times [\tfrac{n}{(n+2)(n+1)^2} + (-\tfrac{1}{n+1})^2] \\
&= \theta^2 \times [\tfrac{2n+2}{(n+2)(n+1)^2}] = \theta^2 \times \tfrac{2}{(n+2)(n+1)}
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat{\theta}_{MOM}) &= \theta^2/3n \\
Bias(\hat{\theta}_{MOM}) &= E[\hat{\theta}_{MOM}] - \theta = 0 \\
MSE(\hat{\theta}_{MOM}) &= Var(\hat{\theta}_{MOM}) + [Bias(\hat{\theta}_{MOM})]^2 \\
&= \theta^2/3n
\end{aligned}
$$

Note that as n grows large, the MLE has variance and MSE which decline to order $O(n^{-2})$ while the MOM estimate declines slower, to order $O(n^{-1})$.

(d). Find a modification of the mle that renders it unbiased.

To adjust $\hat{\theta}_{MLE}$ to make it unbiased, simply multiply it by the factor $(\frac{n+1}{n})$

$\hat{\theta}^*_{MLE} = (\frac{n+1}{n})\hat{\theta}_{MLE}$

4. Problem 8.10.57

[3 points part (a); 4 points part (b); 5 points part (c)]

Note: if a statistic T with expectation $\mu = E[T]$ is used to estimate a constant parameter θ, the mean-squared error is defined as

$$\begin{aligned} MSE &= E[(T-\theta)^2] \\ &= E[(T-\mu+\mu-\theta)^2] \\ &= E[(T-\mu)^2] + E[(+\mu-\theta)^2] + 2E[(T-\mu)(\mu-\theta)] \\ &= Var(T) + [\mu-\theta]^2 + (0) \\ &= Variance \ \ + Squared\ Bias \end{aligned}$$

(See Section 4.4.2 of Rice)

Solution:

(a). $E[s^2] = E[\frac{\sigma^2}{n-1}\chi^2_{n-1}] = \frac{\sigma^2}{n-1} \times E[\chi^2_{n-1}] = \frac{\sigma^2}{n-1} \times (n-1) = \sigma^2$
and $E[\hat{\sigma}^2] = E[\frac{n-1}{n}s^2] = \frac{n-1}{n}E[S^2] = \frac{n-1}{n}\sigma^2$.
So, s^2 is unbiased.

(b). The MSE of an estimate is the sum of its variance and its squared bias.

First, compute the variances of each estimate:

$$\begin{aligned} Var[s^2] &= Var[\frac{\sigma^2}{n-1}\chi^2_{n-1}] = (\frac{\sigma^2}{n-1})^2 \times Var[\chi^2_{n-1}] \\ &= (\frac{\sigma^2}{n-1})^2 \times 2(n-1) = 2\sigma^4/(n-1) \\ Var[\hat{\sigma}^2] &= Var[(\frac{n-1}{n})s^2] = (\frac{n-1}{n})^2 Var[s^2] \\ &= (\frac{n-1}{n})^2 \times 2\sigma^4/(n-1) \\ &= 2(\frac{n-1}{n^2}) \times \sigma^4 \end{aligned}$$

Then, compute the MSEs of each estimate:

$$\begin{aligned} MSE(s^2) &= Var[s^2] + [Bias(s^2)]^2 = Var[s^2] = 2\sigma^4/(n-1) \\ MSE(\hat{\sigma}^2) &= Var[\hat{\sigma}^2] + [Bias(\hat{\sigma}^2)]^2 \\ &= 2\sigma^4(\frac{n-1}{n^2}) + (-\frac{1}{n})^2\sigma^4 \\ &= 2\sigma^4(\frac{n-1/2}{n^2}) = \sigma^2 \times \left(\frac{2n-1}{n^2}\right) \end{aligned}$$

Simple algegra proves that $MSE(\hat{\sigma}^2) < MSE(s^2)$.

(c). For what values of ρ does $W = \rho\sum_{i=1}^n (x_i - \overline{x})^2$ have minimal MSE?

$$\begin{aligned} MSE(W) &= Var[W] + [Bias(W)]^2 \\ &= \rho^2 Var[\sum_1^n (x_i - \overline{x})^2] + [E(W) - \sigma^2]^2 \\ &= \rho^2[2(n-1)\sigma^4] + [\rho(n-1)\sigma^2 - \sigma^2]^2 \\ &= \sigma^4[\rho^2(2(n-1) + (n-1)^2) - 2\rho(n-1) + 1] \\ &= \sigma^4[\rho^2(n-1)(n+1) - 2\rho(n-1) + 1] \end{aligned}$$

Minimizing with respect to ρ we solve:

$$\begin{aligned} \frac{d}{d\rho}MSE(W) &= 0 \\ \Longrightarrow \rho &= \frac{1}{n+1} \end{aligned}$$

So $\rho = 1/(n+1)$ is the value that minimizes the MSE.