STAT 512 hw 5

1. Let $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and consider estimating λ with $nY_{(1)}$.
   (a) Find the pdf of $Y_{(1)}$ and identify its distribution.
   (b) Find $\mathbb{E} nY_{(1)}$ and $\text{Var}(nY_{(1)})$.
   (c) Find $\text{MSE}(nY_{(1)})$ as an estimator of λ.
   (d) Find $\text{MSE}\,\bar{Y}_n$ as an estimator of λ.

2. Let $X_1, \dots, X_5$ be a random sample from a distribution with pdf

$$f_X(x) = \frac{1}{\delta}\mathbf{1}(\delta < x < 2\delta)$$

   for some $\delta > 0$.
   (a) Find the cdf F_X corresponding to the pdf f_X.
   (b) Find the pdf of $X_{(2)}$.
   (c) Find the pdf of the $Y = (X_{(2)} - \delta)/\delta$ of $X_{(2)}$ and give the name of the distribution of Y.
   (d) Give $\mathbb{E}Y$ and $\text{Var}\,Y$.
   (e) Find the MSE of $\hat{\delta} = (3/4)X_{(2)}$ when $\hat{\delta}$ is used as an estimator of δ. *Hint:* $X_{(2)} = \delta Y + \delta$.

3. Let $Y_1, \dots, Y_n$ be a random sample from the distribution with cdf given by

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ (y/a)^b, & 0 \le y \le a \\ 1, & y > a \end{cases}$$

   for some $a, b > 0$, where b is known. Consider the estimator of a given by $\hat{a} = Y_{(n)}$.
   (a) Find the pdf of $Y_{(n)}$.
   (b) Find an expression for $\text{Bias}\,\hat{a}$
   (c) Propose a scaled version of $\hat{a}$ which results in an unbiased estimator, $\tilde{a}$, of a.
   (d) Find the MSE of $\tilde{a}$.
   (e) Find the transformation of a $\text{Uniform}(0, 1)$ rv which will result in a realization of Y.
   (f) Run a simulation using R to confirm the formula you obtained for $\text{MSE}\,\tilde{a}$. Specifically, choose values of a, b, and n and generate $1{,}000$ samples of size n (I recommend choosing $b \le 5$). On each sample, compute the estimator $\tilde{a}$ and record its value. Then compute the average squared distance of your $\tilde{a}$ values from a over the $1{,}000$ simulated data sets. In addition, compute the value of $\text{MSE}\,\tilde{a}$ according to your formula from part (d). The numbers should be quite close to each other. You may make use of the following partial code:

```
a.tilde <- numeric()
for(s in 1:S)
{
  U <- runif(n)
```

```
  Y <- # your formula for generating Y from U
  a.tilde[s] <- # compute a.tilde on the sample
}

mean( (a.tilde - a)^2 )
# compute also MSE of a.tilde according to your formula
```

Here is what to turn in:

i. Your code.

ii. Your simulated value of $\text{MSE}\,\tilde{a}$ as well as its value according to the formula.

iii. A histogram of your `a.tilde` values.

4. Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$ and let $Y = X_1 + \cdots + X_n$. Consider the two estimators of p given by

$$\hat{p} = \frac{Y}{n} \quad \text{and} \quad \tilde{p} = \frac{Y+1}{n+2}.$$

(a) Find $\text{Bias}\,\hat{p}$ and $\text{Bias}\,\tilde{p}$.

(b) Find $\text{Var}\,\hat{p}$ and $\text{Var}\,\tilde{p}$.

(c) Find $\text{MSE}\,\hat{p}$ and $\text{MSE}\,\tilde{p}$.

(d) If the true value of p is 0.50, which estimator has a lower MSE?

(e) If the true value of p is 0.95, which estimator has a lower MSE?

5. Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$. Find a function of $\bar{X}_n$ which is an unbiased estimator of λ^2. *Hint: Begin by finding* $\mathbb{E}\bar{X}_n^2$.

6. Suppose $X_1, \ldots, X_n$ are a random sample from the $\text{Poisson}(\lambda)$ distribution, where λ is unknown.

(a) Find $\mathbb{E}\bar{X}_n$

(b) Find $\mathbb{E}S_n^2$.

(c) Which do you suggest as an estimator for λ? Run a simulation to inform your suggestion: Choose a sample size n and a value of λ and generate $1{,}000$ random samples of size n, computing on each random sample the value of $\bar{X}_n$ and $\bar{S}_n^2$ and storing these. You can do this with a for loop like the following:

```
X.bar <- S.sq <- numeric(S) # S is the number of data sets to simulate
for(s in 1:S)
{
    X <- rpois(n,lambda)
    X.bar[s] <- mean(X)
    S.sq[s] <- var(X)
}
```

Make histograms of the $1{,}000$ values of $\bar{X}_n$ and $\bar{S}_n^2$ from your simulation and use these to argue for using one or the other as an estimator for λ. Turn in your code and the two histograms. *Hint: Use* `rpois` *to generate the Poisson data.*