Homework 06
Joseph Blubaugh
jblubau1@tamu.edu
STAT 641-720

I. 1) a) $PDF : \frac{e^{-\frac{x}{5}}}{5}$ b) mean = $\frac{1}{5}$, sd = $\frac{1}{5}$

2) Using two different test for testing normality, both tests have low p-values so the normal distribution is not a good fit.

```
    One-sample Kolmogorov-Smirnov test

data:  dt
D = 0.4142, p-value = 0.0002132
alternative hypothesis: two-sided


    Shapiro-Wilk normality test

data:  dt
W = 0.8448, p-value = 0.001408
```

3)

a) 0.9705618
b) 0.96
c) 0.852

4) The exact distribution is the most accurate followedb by the central limit theorem and the simulated values are the least accurate.

II.

1) The bootstrap standard error is much lower than the sample error because of the sample size difference.

```
Sample Data Std Error: 0.2055469

Bootstrap Std Err: 0.006455202
```

2)

```
Sample Median:  2.709055

Sample Std Dev:  1.404836

MAD:  0.4608814
```

3) Overall the bootstrap sample results for mean and median are pretty close to the historically modeled version. In the boostrap simulation the mean and standard deviation are noticeably lower than the theoretical model.

III.

1) This is a binomial(n = 100, p = .2) distribution.

a) Mean = 20, Sd = 4
b) Mean = .2, Sd = .04
c) exact probability 0.4405384, normal approximation 0.2676

2)

a)

```
Sample Quantiles:  4.155514 5.794235 8.288465 14.51986 16.96354 21.66505
Chi Squared Quantiles:  4.168159 5.898826 8.342833 14.68366 16.91898 21.66599
```

b)

```
Gamma Sample Quantiles:  0.0504419 0.3434575 1.874509 23.92845 40.22023 123.8628
Chi Squared Quantiles: 4.168159 5.898826 8.342833 14.68366 16.91898 21.66599
```

c)

```
T Distibution Sample Quantiles:  2.049167 3.216188 5.586994 17.03645 25.23898
58.85611 Chi Squared Quantiles:  4.168159 5.898826 8.342833 14.68366 16.91898
21.66599
```

d) The skewness and heavy tails impacts the variance by stretching out the distribution.

3)

a)

$$
\begin{aligned}
P(Yes) &= P(Yes \cap Q_1 \cup Yes \cap Q_2) \\
&= P(Yes \cap Q_2) + P(Yes \cap Q_2) \\
&= P(Yes|Q_1)P(Q_1) + P(Yes|Q_2)P(Q_2) \\
&= p\theta + (1-p)(1-\theta)
\end{aligned}
$$

b)

$$\begin{aligned}\pi &= p\theta + (1-\theta)(1-p)\\ &= p\theta - p(1-\theta) + (1-\theta)\\ &= p(2\theta-1) + (1-\theta)\\ p &= \frac{\pi-(1-\theta)}{(2\theta-1)}\\ \hat{p} &= \frac{\hat{\pi}-(1-\theta)}{2\theta-1}\end{aligned}$$

c)

$$\begin{aligned}\hat{p} &= \frac{\hat{\pi}-(1-\theta)}{2\theta-1}\\ Var(\hat{p}) &= Var(\frac{\hat{\pi}}{2\theta-1} - \frac{(1-\theta)}{(2\theta-1)})\\ &= Var(\frac{\hat{\pi}}{2\theta-1}\\ &= \frac{1}{(2\theta-1)^2}Var(\hat{\pi})\\ &= \frac{1}{(2\theta-1)^2}\frac{\pi(1-\pi)}{n}\end{aligned}$$

d) As θ goes to 0 or 1 from 1/2 $Var(\hat{p})$ approaches 0.

IV.

1) C, the best estimator is always the one with the smallest error

2) B, bootstrapping is really good at approximating distributions when the sample is small

3) A, both of these conditions are better than one or the other individually

4) C, for a symmetrical distribution mean is superior to median

5) E, none of these choices are correct

6) D, the estimator with the smallest mse will be the closest value to the true mean

7) B, since the sample is not small, using the sample distribution is best

8) D, bias is the difference between the actual parameter value and the sample parameter value