STAT 512 su 2021 hw 6

1. Let $Y_1, \ldots, Y_n$ be a random sample from the distribution with pdf given by

$$f_Y(y) = \frac{\Gamma(2+\beta)}{\Gamma(\beta)} y(1-y)^{\beta-1}, \quad 0 < y < 1.$$

   (a) Find $\mathbb{E}\bar{Y}_n$. *Hint: Try to recognize the pdf.*

   (b) Find $\operatorname{Var} \bar{Y}_n$.

   (c) Show that $\bar{Y}_n$ is a consistent estimator of $2/(2+\beta)$.

   (d) Propose a function of $\bar{Y}_n$ which is a consistent estimator of β.

2. Let $X_1, \ldots, X_n$ be a random sample from the distribution with cdf given by

$$F_X(x) = \begin{cases} 1 - \left(\frac{c}{x}\right)^{\alpha}, & x \geq c \\ 0, & x < c \end{cases}$$

   (a) Find the population pdf.

   (b) Find the pdf of $X_{(1)}$.

   (c) Show that $X_{(1)}$ is a consistent estimator of c by directly showing that

$$\lim_{n \to \infty} P(|X_{(1)} - c| < \varepsilon) = 1$$

   for every $\varepsilon > 0$.

3. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$.

   (a) For $n = 10$, give the value of a such that $P\left(\bar{X}_n - aS_n/\sqrt{n} < \mu < \bar{X}_n + aS_n/\sqrt{n}\right) = 0.99$.

   (b) For $n = 10$, find $P\left(\bar{X}_n > \mu + 2S_n/\sqrt{n}\right)$.

   (c) For $n = 10$, find $P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right)$.

   (d) Give $\lim_{n\to\infty} P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right)$.

4. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$.

   (a) For $n = 10$, give values of a and b such that $P\left(aS_n^2 < \sigma^2 < bS_n^2\right) = 0.90$.

   (b) For $n = 10$, find $P\left((0.473)S_n^2 < \sigma^2 < (3.333)S_n^2\right)$.

   (c) Give $\lim_{n\to\infty} P(-0.01 < S_n^2 - \sigma^2 < 0.01)$.

5. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$. Then according to the central limit theorem

$$\frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}} \to^D Z,$$

   as $n \to \infty$, where $Z \sim \text{Normal}(0, 1)$.

(a) Use the above result to find

$$\lim_{n\to\infty} P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right).$$

(b) Use mgfs to show that $\sqrt{n}\bar{X}_n/\lambda \sim \text{Gamma}(n, 1/\sqrt{n})$.

(c) Use the previous result to compute the probability

$$P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right)$$

for $\lambda = 10$ and $n = 10, 20, 30, 100, 500, 1000, 10000$. The numbers should approach your answer from part (a). *Hint: Use the* `pgamma` *function.*

(d) Suppose we observe $\bar{X}_n = 20.2$ on a sample of size $n = 50$. Give a 95% CI for λ based on

i. the exact pivot quantity result $\sqrt{n}\bar{X}_n/\lambda \sim \text{Gamma}(n, 1/\sqrt{n})$.

ii. the asymptotic pivot quantity result $\sqrt{n}(\bar{X}_n - \lambda)/\lambda \xrightarrow{\text{D}} Z$, $Z \sim \text{Normal}(0, 1)$.

Hint: Use `qgamma(.,shape = ., scale = .)` *to obtain quantiles of the Gamma distribution.*

6. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, and let $\hat{p}_n = n^{-1}(X_1 + \cdots + X_n)$.

(a) Give an exact expression for $P(1/2 < \hat{p}_n)$.

(b) Evaluate your expression from part (a) for $n = 200$ and $p = 4/9$.

(c) Find an approximation to $P(1/2 < \hat{p}_n)$ when $n = 200$ and $p = 4/9$ using the fact that

$$\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \to^D Z,$$

as $n \to \infty$, where $Z \sim \text{Normal}(0, 1)$.

(d) Suppose that for $n = 200$, $\hat{p} = 0.64$. Give an approximate 95% CI for p based on this data.

7. Consider drawing a random sample $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and computing the interval

$$\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}.$$

(a) Give $\lim_{n\to\infty} P\left(\bar{X}_n - S_n/\sqrt{n} < \lambda < \bar{X}_n + S_n/\sqrt{n}\right)$.

(b) For small values of n, the interval $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ will *not* contain λ with the desired probability of $1 - \alpha$. For large n, however, by the central limit theorem and an application of Slutzky's theorem, the probability that $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ contains λ should be close to $1 - \alpha$.

The *coverage* of a confidence interval is the probability that it contains its target. The *nominal coverage* $1 - \alpha$ is the stated and desired coverage, which may differ from the actual coverage.

Conduct some simulations to estimate the coverage of the confidence interval $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ for $\alpha = 0.05$ when $\lambda = 20$ for the sample sizes $n = 5, 10, 15, 25, 50, 100$. For each value of n, generate 1000 realizations of the interval. Here is partial code:

```
covered <- logical(S) # vector to store TRUE/FALSE values

for(s in 1:S)
{
    # generate random sample from Exp(lambda)
    X <- rexp(n,1/lambda)

    X.bar <- mean(X)
    S.n <- sd(X)

    L <- X.bar - qnorm(1-alpha/2) * S.n / sqrt(n)
    U <- X.bar + qnorm(1-alpha/2) * S.n / sqrt(n)

    # check whether interval contained lambda
    covered[s] <- ( L < lambda ) & ( U > lambda )
}

# compute proportion of TRUE values
coverage <- mean(covered)

coverage
```

What coverages do you get for the sample sizes $n = 5, 10, 15, 25, 50, 100$? For what sample sizes do you advise using this confidence interval (for what sample sizes is the coverage close to 0.95)?

Optional (do not turn in) problems for additional study from Wackerly, Mendenhall, Scheaffer, 7th Ed.:

- 8.60, 8.63
- 9.17, 9.20, 9.25 (about consistency)