Solutions for Stat 512 — Take home exam III

1. If $Y \sim F_{v_1,v_2}$, then prove: $\dfrac{\left(\frac{v_1}{v_2}\right)Y}{1+\left(\frac{v_1}{v_2}\right)Y} \sim Beta(\frac{v_1}{2},\frac{v_2}{2})$. (10 pts)

Solution: (Hint: use transformation technique)

Let $U = \dfrac{\left(\frac{v_1}{v_2}\right)Y}{1+\left(\frac{v_1}{v_2}\right)Y}$, then $g^{-1}(u) = \dfrac{\left(\frac{v_2}{v_1}\right)u}{1-u} = \dfrac{v_2}{v_1}u(1-u)^{-1}$ and $\dfrac{dg^{-1}(u)}{du} = \dfrac{v_2}{v_1}\dfrac{1}{(1-u)^2}$.

Now,

$$
\begin{aligned}
f_U(u) &= f_Y(g^{-1}(u))\left|\frac{dg^{-1}(u)}{du}\right| \\
&= \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}\frac{\left(\frac{\left(\frac{v_2}{v_1}\right)u}{1-u}\right)^{\frac{v_1}{2}-1}}{\left(1+\frac{u}{1-u}\right)^{\frac{v_1+v_2}{2}}}\cdot\frac{v_2}{v_1}\frac{1}{(1-u)^2} \\
&= \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}\left(\frac{v_1}{v_2}\right)^{1-\frac{v_1}{2}}u^{\frac{v_1}{2}-1}\left(\frac{1}{1-u}\right)^{-1-\frac{v_2}{2}}\cdot\left(\frac{v_1}{v_2}\right)^{-1}\cdot\left(\frac{1}{1-u}\right)^2 \\
&= \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}u^{\frac{v_1}{2}-1}(1-u)^{\frac{v_2}{2}-1}, \qquad u \in (0,1)
\end{aligned}
$$

Hence, $\dfrac{\left(\frac{v_1}{v_2}\right)Y}{1+\left(\frac{v_1}{v_2}\right)Y} \sim Beta(\frac{v_1}{2},\frac{v_2}{2})$.

****

2. Assume now that the amount of fill dispensed by the bottling machine is exponentially distributed with $\beta = 2$, that is $Y_1,\ldots,Y_n \sim Exp(2)$.

a. What is the asymptotic distribution of $\bar{Y}$ when $n = 3$? Using the standard normal table to find out $P|\bar{Y} - 2| \leq 3$. (10 pts)
(Hint: Using CLT; asymptotic distribution means the limiting distribution, that is, the distribution when $n \to \infty$).

b. Obtain the exact distribution of $\bar{Y}$ using MGF technique. Is this a common distribution that you've seen before? (10 pts)

Solution:

a.

Suppose CLT can be applied here, then $\dfrac{\bar{Y} - 2}{\sqrt{\dfrac{4}{3}}} \xrightarrow{d} N(0, 1)$ which implies the asymptotic distribution for $\bar{Y}$ is $N(2, \dfrac{4}{3})$. Now, $P|\bar{Y} - 2| \leq 3 = P(-1 \leq \bar{Y} \leq 5) = P\left(\dfrac{-1-2}{\sqrt{\dfrac{4}{3}}} \leq \dfrac{\bar{Y}-2}{\sqrt{\dfrac{4}{3}}} \leq \dfrac{5-2}{\sqrt{\dfrac{4}{3}}}\right) \approx P(-2.6 \leq Z \leq 2.60) = 1 - 2P(Z \leq -2.6) = 1 - 2 * 0.0047 = 0.9906$.

b.

The MGF of $\bar{Y}$ is:

$$
\begin{aligned}
m_{\bar{Y}}(t) &= E\left(e^{t\bar{Y}}\right) \\
&= E\left(e^{t\frac{\sum Y_i}{n}}\right) \\
&= m_{\sum Y_i}\left(\frac{t}{n}\right) \\
&= \left[m_Y\left(\frac{t}{n}\right)\right]^n \\
&= \left(1 - \frac{\beta t}{n}\right)^{-n}
\end{aligned}
$$

Hence, $\bar{Y} \sim gamma(n, \dfrac{\beta}{n})$ where $\beta = 2$ in this question.

****

3. The flow of water through soil depends on, among other things, the porosity (volume proportion of voids) of the soil. To compare two types of sandy soil, $n_1 = 50$ measurements are to be taken on the porosity of soil A and $n_2 = 100$ measurements are to be taken on the porosity of soil B. Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Population means for the two types of soil are μ_1, μ_2. Denote the samples for the soil A be $X_1, \ldots, X_{50}$, the samples for the soil B be $Y_1, \ldots, Y_{100}$.

a. What is the asymptotic distribution for $\bar{X} - \bar{Y}$? (10 pts)

b. What is the probability that sample variance of soil A is at least twice as large as the sample variance of soil B? (You do not need to find out the exact value, here is the format of your final answer: $P(F_{15,20} \geq (\leq) 10)$) (10 pts)

solution:

a.

Asymptotic distribution for $\bar{X}$ is $N(\mu_1, \frac{0.01}{50})$.
Asymptotic distribution for $\bar{Y}$ is $N(\mu_2, \frac{0.02}{100})$ based on CLT.
Since $X_i's$ are independent with $Y_i's$, $\bar{X}$ is independent with $\bar{Y}$. Hence, it is clear that $\bar{X} - \bar{Y} \xrightarrow{d} N(\mu_1 - \mu_2, \frac{0.04}{100})$.

b.

$$
\begin{aligned}
P(S_1^2 \geq 2S_2^2) &= P\left(\frac{S_1^2}{S_2^2} \geq 2\right) \\
&= P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \geq \frac{2/\sigma_1^2}{1/\sigma_2^2}\right) \\
&= P(F_{49,99} \geq 4) \\
&\approx 2.4 \times 10^{-9}
\end{aligned}
$$

****

4. Suppose Y follows a binomial distribution with parameter n and p, then $\frac{Y}{n}$ is an unbiased estimator of p. Now, we want to estimate the variance of Y which is $np(1-p)$, we proposed a new estimator: $n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$.

a. Show that the proposed estimator for the variance of Y is biased. (10 pts)

b. Adjust the estimator so that the new estimator is an unbiased estimator of $np(1-p)$. (10 pts)

Solution:

a.

$$\begin{aligned} E(\widehat{p}) = E\left[n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\right] &= E\left[Y - \frac{Y^2}{n}\right] \\ &= EY - \frac{EY^2}{n} \\ &= np - \frac{n^2p^2 + np(1-p)}{n} \\ &= np - np^2 - p + p^2 \\ &= p(1-p)(n-1) \neq np(1-p) \end{aligned}$$

That means, the proposed estimator is biased.

b.

From part (a), $E(\hat{p}) = p(1-p)(n-1)$. That implies $E\left(\frac{n}{n-1}\hat{p}\right) = np(1-p)$. So $\frac{n}{n-1}\hat{p} = \frac{n^2}{n-1}\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$ is unbiased for the variance of Y.

****

5. Suppose that $Y_1, \ldots, Y_n$ denote a random sample of size n from a population with an exponential distribution $exp(\beta)$. $Y_{(1)} = min(Y_1, \ldots, Y_n)$ denotes the smallest-order ststistic.

a. Show that $\hat{\beta}_1 = nY_{(1)}$ is an unbiased estimator for β. (10 pts)

b. Show that $\hat{\beta}_2 = \bar{Y}$ is also an unbiased estimator for β. (10 pts)

c. Compare the variance of the two estimators. (10 pts)

Solution:

a. For $\hat{\beta}_1 = nY_{(1)}$, let's derive find the distribution of $Y_{(1)}$:

$$f_{Y_{(1)}}(y) = \frac{n!}{(n-1)!} \cdot \frac{1}{\beta} e^{-\frac{y}{\beta}} \left(e^{-\frac{y}{\beta}} \right)^{n-1}$$
$$= \frac{n}{\beta} e^{-\frac{ny}{\beta}}, \qquad y \geq 0$$

Hence, $Y_{(1)} \sim exp(\frac{\beta}{n}) \Longrightarrow E(Y_{(1)}) = \frac{\beta}{n} \Longrightarrow E(nY_{(1)}) = \beta$. That means, $\hat{\beta}_1$ is unbiased for β.

b.

$$E(\hat{\beta}_2) = E(\bar{Y}) = E\left(\frac{\sum Y_i}{n}\right) = \frac{1}{n}\sum E(Y_i) = \frac{1}{n} \cdot n\beta = \beta$$

Hence, $\hat{\beta}_2$ is unbiased for β.

c.

For $\hat{\beta}_1$, since $Y_{(1)}$ follows exponential distribution, $Var(Y_{(1)}) = \frac{\beta^2}{n^2}$. Hence $Var(nY_{(1)}) = \beta^2$.

For $\hat{\beta}_2$, $Var(\bar{Y}) = \frac{\beta^2}{n}$. That is, variance of $\hat{\beta}_2$ is smaller than variance of $\hat{\beta}_1$ for $n > 1$.

****

Extra credit question: $Y_1, \ldots, Y_n$ are i.i.d. from $U(\theta, \theta + 1)$. Define $\hat{\theta}_1 = Y_{(1)}$. Adjust $\hat{\theta}_1$ so that it becomes an unbiased estimator for θ and compute the associated variance for the adjusted estimator. (10 pts)

Solution:

Firstly let's find the distribution of $\hat{\theta}_1$:

$$f_{Y_{(1)}}(y) = n \cdot (1 - y + \theta)^{n-1}, \qquad y \in (\theta, \theta + 1)$$

Hence,

$$
\begin{aligned}
E(Y_{(1)}) &= \int_{\theta}^{\theta+1} y \cdot n \cdot (1-y+\theta)^{n-1} dy \\
&= n \int_{\theta}^{\theta+1} y(1-y+\theta)^{n-1} dy \qquad \text{(integration by parts)} \\
&= -y(1+\theta-y)^n \Big|_{\theta}^{\theta+1} + \int_{\theta}^{\theta+1} (1+\theta-y)^n dy \\
&= -(\theta+1)(0) + \theta + \left(-\frac{1}{n+1}\right)(1+\theta-y)^{n+1}\Big|_{\theta}^{\theta+1} \\
&= \theta + \frac{1}{n+1}
\end{aligned}
$$

That means, $E\left(Y_{(1)} - \frac{1}{n+1}\right) = \theta$. Now let's derive the variance of $Y_{(1)} - \frac{1}{n+1}$:

$$
\begin{aligned}
Var\left(Y_{(1)} - \frac{1}{n+1}\right) &= Var\left(Y_{(1)}\right) \\
&= E\left(Y_{(1)}^2\right) - \left[E(Y_{(1)})\right]^2
\end{aligned}
$$

We derived $E(Y_{(1)}) = \theta + \frac{1}{n+1}$. Now let's derive $E\left(Y_{(1)}^2\right)$:

$$
\begin{aligned}
E\left(Y_{(1)}^2\right) &= \int_{\theta}^{\theta+1} y^2 \cdot n \cdot (1-y+\theta)^{n-1} dy \\
&= -y^2(1-y+\theta)^n \Big|_{\theta}^{\theta+1} + \int_{\theta}^{\theta+1} 2y(1-y+\theta)^n dy \qquad \text{(Again, integration by parts)} \\
&= \theta^2 + \frac{2\theta}{n+1} + \frac{2}{(n+1)(n+2)}
\end{aligned}
$$

Now,

$$
\begin{aligned}
Var\left(Y_{(1)} - \frac{1}{n+1}\right) &= \theta^2 + \frac{2\theta}{n+1} + \frac{2}{(n+1)(n+2)} - \left(\theta + \frac{1}{n+1}\right)^2 \\
&= \frac{n}{(n+1)^2(n+2)}
\end{aligned}
$$