RICE STAT 533 / GSBS 1283, Spring 2020 Exam 2

- This take home exam must be submitted to `jplong@mdanderson.org` by 5:00pm on March 31. You can write your solutions on paper, take photos of the pages, compile the photos into a single pdf, and email the document. You can also write solutions in latex if you wish and send the pdf.
- Show your work for each question. Partial credit will be given for partially worked out problems.
- You may use the course textbook, your in class notes, homework assignments, and any material on the course website to answer questions. You may **not** use any other internet resources or communicate with other class members.

1. Let $X_1, \dots, X_n$ be i.i.d. from the Weibull distribution $W(a, \theta)$ where $a > 0$ and $\theta > 0$ are unknown.

   (a) (5 points) Show that $R(X, a, \theta) = \prod_{i=1}^n \frac{X_i^a}{\theta}$ is pivotal.

   Answer: The transformation $g(x) = x^a/\theta$ is strictly increasing in x. Therefore from Proposition 1.8 in the textbook $g(X_i) \sim Expo(1)$. Since the X_i are independent R is distributed as the product of n $Expo(1)$ random variables. Thus its distribution does not depend on θ or a and it is pivotal.

   (b) (5 points) Use $R(X, a, \theta)$ to construct a confidence set with confidence coefficient $1 - \alpha$. You do not need to find explicit bounds for the set.

   Answer: Let f_R be the distribution of R. Choose c_1 and c_2 such that $\int_{c_1}^{c_2} f_R(r)dr = 1 - \alpha$. Then

   $$C(x) = \{(a, \theta) : c_1 \le \prod_{i=1}^n \frac{X_i^a}{\theta} \le c_2\}$$

   has confidence coefficient $1 - \alpha$.

2. (10 points) Derive a counter-example to Proposition 7.4 in textbook which states:

   Proposition 7.4. Let $C_j(X)$ $j = 1, 2$ be the confidence sets given in (7.19) with $\widehat{\theta}_n = \widehat{\theta}_{jn}$ and $\widehat{V_n} = \widehat{V}_{jn}$ for $j = 1, 2$ respectively. Suppose that for each j, (7.18) holds for $\widehat{\theta}_{jn}$ and $\widehat{V}_{jn}$ is [strongly] consistent for V_{jn}, the asymptotic covariance matrix of $\widehat{\theta}_{jn}$. If $Det(V_{1n}) < Det(V_{2n})$ for sufficiently large n, where $Det(A)$ is the determinant of A, then

   $$P(vol(C_1(X)) < vol(C_2(X))) \to 1$$

   Hint: There are many possible counter examples. Consider the 1-dimensional case where matrices are scalars so $Det(V) = V$. For the counter-example, it is sufficient to show that

   $$\frac{vol(C_1(X))}{vol(C_2(X))} > 1$$

   with probability 1 for all large n. By the volume formula in the book proof this is equivalent to showing

   $$\frac{\widehat{V}_{1n}}{\widehat{V}_{2n}} > 1 \tag{1}$$

   with probability 1 for all large n. Now find sequences of V_{1n}, V_{2n}, $\widehat{V}_{1n}$ and $\widehat{V}_{2n}$ (deterministic ones exist even for the estimators) such that the conditions of the proposition are satisfied and Equation (1) is satisfied. These quantities do not have to be based on any statistical model for X.

   Answer: Let $\widehat{V}_{2n} = V_{2n} = 1/n$, $V_{1n} = 1/n - \epsilon/n^2$, and $\widehat{V}_{1n} = 1/n + \epsilon/n^2$ where $\epsilon > 0$ and small (eg 1/2). Then $\widehat{V}_{2n}$ is strongly consistent for $\widehat{V}_{n2}$ (they are equal). Further $\widehat{V}_{1n}$ is strongly consistent for V_{1n} because

   $$\frac{\widehat{V}_{1n}}{V_{1n}} = \frac{1/n + \epsilon/n^2}{1/n - \epsilon/n^2} \to 1$$

Further $Det(V_{1n}) < Det(V_{2n})$ because $V_{1n} < V_{2n}$. Finally Equation (1) is satisfied because

$$\frac{\widehat{V}_{1n}}{\widehat{V}_{2n}} = \frac{1/n + \epsilon/n^2}{1/n} > 1$$

where the inequality is with probability 1 (we choose deterministic estimating sequences $\widehat{V}_{1n}$ and $\widehat{V}_{2n}$).

3. Let $X_i \sim N(\mu, \sigma_i^2)$ independent. Note that the X_i are not identically distributed because the variance σ_i^2 is different for each observation. Assume the σ_i^2 are known and μ is unknown.

   (a) (4 points) Construct a $1-\alpha$ confidence interval using the pivot statistic

$$Q(\vec{X}, \mu) = \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2} - \mu.$$

Answer: Q is a linear combination of normals so it is normal. Find the expectation and variance.

$$\mathbb{E}[Q] = \mathbb{E}[\frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2}] - \mu = 0$$
$$Var(Q) = Var(\frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2}) - \mu = \frac{1}{\sum \sigma_i^{-2}}$$

Choose a and b such that

$$P(a < Q < b) = 1 - \alpha$$

Let $Z \sim N(0,1)$. Using the transformation to standard normal we have

$$P(a/\sqrt{1/\sum \sigma_i^{-2}} < Z < b/\sqrt{1/\sum \sigma_i^{-2}}) = 1 - \alpha$$

One approach is to use α splitting. So

$$a = -z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}}$$
$$b = z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}}$$

So the CI is

$$C(x) = \left\{\mu : -z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}} \leq \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2} - \mu \leq z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}}\right\}$$
$$= \left\{\mu : \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2} - z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}} \leq \mu \leq \frac{1}{\sum \frac{1}{\sigma_i^2}} \sum \frac{X_i}{\sigma_i^2} + z_{\alpha/2}\sqrt{1/\sum \sigma_i^{-2}}\right\}$$

(b) (4 points) Show that $Q(\vec{X}, \mu) = \bar{X} - \mu$ is a pivot statistic and construct a $1 - \alpha$ confidence interval using this pivot. $\bar{X} = (1/n) \sum X_i$ is the usual mean.

Answer: $Q = \bar{X} - \mu$ is normal because it is a linear combination of normal random variables. Compute its mean and variance

$$\mathbb{E}[Q] = \mathbb{E}[\bar{X}] - \mu = \mu - \mu = 0$$
$$Var(Q) = \frac{1}{n^2} \sum \text{Var}(X_i) = \frac{1}{n^2} \sum \sigma_i^2$$

So $Q \sim N(0, \frac{1}{n^2} \sum \sigma_i^2)$. Therefore Q is a pivot because its distribution does not depend on μ. Following the same procedure as in part a), a $1 - \alpha$ CI (using α splitting) is

$$C(x) = \left\{ \mu : \bar{X} - z_{\alpha/2} \frac{1}{n} \sqrt{\sum \sigma_i^2} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{1}{n} \sqrt{\sum \sigma_i^2} \right\}$$

(c) (2 points) Which confidence interval is better and why (a) or b))?

Answer: We compare these two intervals by width

$$\text{a) has width } 2z_{\alpha/2} \sqrt{1/\sum \sigma_i^{-2}}$$
$$\text{b) has width } 2z_{\alpha/2} \frac{1}{n} \sqrt{\sum \sigma_i^2}$$

So must determine

$$\sqrt{1/\sum \sigma_i^{-2}} \;?\; \frac{1}{n} \sqrt{\sum \sigma_i^2}$$

Square both sides and multiply by n to obtain

$$\frac{1}{(1/n) \sum \sigma_i^{-2}} \;?\; \frac{1}{n} \sum \sigma_i^2$$

The left side is the harmonic mean of the σ_i^2 while the right side is the arithmetic mean of the σ_i^2. So $\leq$ holds by the harmonic–arithmetic mean inequality. So the confidence interval in a) in narrower and thus preferable.