



Homework 4

Brian Cervantes Alvarez

May 19, 2024

ST 563 Theory of Statistics III

Problem 1

Solution

Given X_1, \dots, X_n as iid $\mathcal{N}(\theta, 1)$, we know the 95% confidence interval for θ is given by $\bar{X} \pm \frac{1.96}{\sqrt{n}}$, where \bar{X} is the sample mean of X_1, \dots, X_n . We define p as the probability that an additional independent observation X_{n+1} will fall within this interval,

$$p = \mathbb{P} \left(X_{n+1} \in \left[\bar{X} - \frac{1.96}{\sqrt{n}}, \bar{X} + \frac{1.96}{\sqrt{n}} \right] \right).$$

We need to determine whether p is greater than, less than, or equal to 0.95.

Recall,

- $X_{n+1} \sim \mathcal{N}(\theta, 1)$
- $\bar{X} \sim \mathcal{N}(\theta, \frac{1}{n})$

Next, $X_{n+1} - \bar{X}$ is the difference of two independent normally distributed variables, which follows,

$$X_{n+1} - \bar{X} \sim \mathcal{N}(0, 1 + \frac{1}{n})$$

Transforming $X_{n+1} - \bar{X}$ into a standard normal variable, we have,

$$Z = \frac{X_{n+1} - \bar{X}}{\sqrt{1 + \frac{1}{n}}} \sim \mathcal{N}(0, 1)$$

The event that X_{n+1} falls within the confidence interval can be rewritten as:

$$\mathbb{P} \left(-\frac{1.96}{\sqrt{n}} \leq X_{n+1} - \bar{X} \leq \frac{1.96}{\sqrt{n}} \right) = \mathbb{P} \left(|Z| \leq \frac{1.96\sqrt{n}}{\sqrt{n+1}} \right)$$

Here, $\frac{1.96\sqrt{n}}{\sqrt{n+1}} < 1.96$ because $\sqrt{n+1} > \sqrt{n}$, implying that the interval width defined by $\frac{1.96\sqrt{n}}{\sqrt{n+1}}$ on the standard normal is narrower than the typical ± 1.96 used to define a 95% confidence interval in

a $\mathcal{N}(0, 1)$ distribution. Since the interval for $|Z|$ is narrower than the standard ± 1.96 , the coverage of this interval under the standard normal curve is less than 95%.

Therefore, $p < 0.95$.



Problem 2

Solution

Part A

Consider $X_1, \dots, X_n \sim \mathcal{N}(0, \sigma_X^2)$ and $Y_1, \dots, Y_m \sim \mathcal{N}(0, \sigma_Y^2)$. Assume that the X and Y are independent.

The joint likelihood function, considering the independence between X and Y , is given by:

$$L(\sigma_X^2, \sigma_Y^2) \propto (\sigma_X^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n X_i^2}{2\sigma_X^2}\right) (\sigma_Y^2)^{-m/2} \exp\left(-\frac{\sum_{i=1}^m Y_i^2}{2\sigma_Y^2}\right).$$

Setting $\lambda = \sigma_Y^2/\sigma_X^2$, we can express σ_Y^2 as $\lambda\sigma_X^2$. Thus, substituting and simplifying, the likelihood becomes:

$$L(\sigma_X^2, \lambda) = (\sigma_X^2)^{-(n+m)/2} \exp\left(-\frac{1}{2\sigma_X^2} \left(\sum_{i=1}^n X_i^2 + \lambda \sum_{i=1}^m Y_i^2\right)\right).$$

Under the unrestricted model, the MLEs are obtained directly:

$$\hat{\sigma}_X^2 = \frac{\sum X_i^2}{n}, \quad \hat{\sigma}_Y^2 = \frac{\sum Y_i^2}{m}, \quad \hat{\lambda} = \frac{\hat{\sigma}_Y^2}{\hat{\sigma}_X^2}$$

Under the restricted model, the MLE of σ_X^2 is,

$$\hat{\sigma}_{X,0}^2 = \frac{\sum X_i^2 + \lambda_0 \sum Y_i^2}{n + \lambda_0 m}$$

The LRT statistic is then given by,

$$\lambda = \left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_{X,0}^2}\right)^{n/2} \left(\frac{\hat{\sigma}_Y^2}{\lambda_0 \hat{\sigma}_{X,0}^2}\right)^{m/2} \lambda_0^{-m/2}$$

Therefore, we reject H_0 if $\Lambda(x, y) < k$, where k is a critical value chosen to control the Type I error rate at a level α .

Part B

Under the null hypothesis H_0 , the ratio $\frac{Y_i^2}{\lambda_0 \sigma_X^2}$ follows a scaled chi-squared distribution, χ_m^2 , and similarly, $\frac{X_i^2}{\sigma_X^2}$ follows χ_n^2 . Thus, we can express the test statistic $\lambda(X, Y)$ using these distributions,

$$\begin{aligned}\sum Y_i^2 / (\lambda_0 \sigma_X^2) &\sim \chi_m^2 \\ \sum X_i^2 / \sigma_X^2 &\sim \chi_n^2\end{aligned}$$

The LRT statistic becomes,

$$\lambda(X, Y) = \left(\frac{(\sum Y_i^2 / \lambda_0 \sigma_X^2) / m}{(\sum X_i^2 / \sigma_X^2) / n} \right)^{m/2} \cdot \left(\frac{m}{n} \right)^{m/2} \cdot \left(\frac{n}{m+n} \right)^{n/2} \cdot \left(\frac{m}{m+n} \right)^{m/2}$$

This simplifies to,

$$\lambda(X, Y) = \left(\frac{n/m + F}{m/n + 1/F} \right)^{m/2}$$

where $F = \frac{\sum Y_i^2 / n}{\sum X_i^2 / m}$ under H_0 follows an $F_{m,n}$ distribution.

The rejection region for the LRT in terms of the F-distribution is derived from the transformation of the LRT statistic:

$$\lambda(X, Y) \text{ is rejected if } \lambda(X, Y) < c_\alpha$$

Rewriting this condition by using the F-statistic:

$$\begin{aligned}\left[\frac{n + m \cdot F}{n \cdot F + m} \right]^{m/2} &< c_\alpha \\ (n + m \cdot F)^{m/2} &< c_\alpha (n \cdot F + m)^{m/2}\end{aligned}$$

The exact form of c_α depends on the desired level of the test α and would typically be determined by referring to critical values from the F-distribution, $P(F > F_\alpha) = \alpha$



Part C

Let us define $a = \frac{m}{n+m}$ and $b = \frac{\sum Y_i^2}{\sum X_i^2}$

Then, we can consider the function,

$$h(\lambda) = \left(\frac{1}{a + b/\lambda} \right)^{n/2} \left(\frac{1}{(1-a) + a(1-a)/b} \right)^{m/2}$$

To make it easier, we can log transform and simplify,

$$\log h(\lambda) = -\frac{n}{2} \log(a + b/\lambda) - \frac{m}{2} \log((1-a) + a(1-a)/b)$$

To continue, the first derivative of $\log h(\lambda)$ with respect to λ gives insight into the behavior of the function:

$$\frac{d}{d\lambda} \log h(\lambda) = -\frac{n}{2} \frac{b}{\lambda^2(a\lambda + b)}$$

This derivative shows the function is an upside-down bowl, confirming that $h(\lambda)$ has a single maximum over $\lambda > 0$.

The critical points are derived by setting the derivative to zero and solving for λ ,

$$-\frac{n}{2} \frac{b}{\lambda^2(a\lambda + b)} = 0$$

However, this derivative does not zero out, indicating no critical points from this derivative directly. Thus, we need to look at the endpoints $\lambda \rightarrow 0$ and $\lambda \rightarrow \infty$ to understand the behavior.

So, the confidence set for λ can be formulated from the quantiles of the F-distribution:

$$\lambda \in \left(\frac{\sum Y_i^2/m}{\sum X_i^2/n} \cdot F_{m,n}^{-1}(1 - \alpha/2), \frac{\sum Y_i^2/m}{\sum X_i^2/n} \cdot F_{m,n}^{-1}(\alpha/2) \right)$$

This interval reflects the range of λ values for which we are $1 - \alpha$ confident that the true ratio of variances lies, given the data.



Problem 3

Solution

Part A

Given $X \sim \text{Beta}(\theta, 1)$, the density function of X is defined as,

$$f_X(x; \theta) = \theta x^{\theta-1}, \quad 0 < x < 1.$$

Define a new random variable $Y = -(\log X)^{-1}$. To find the distribution of Y , consider the transformation:

$$Y = -(\log X)^{-1} \Rightarrow X = e^{-1/Y}.$$

The Jacobian of the transformation from X to Y is,

$$\frac{dX}{dY} = \frac{d}{dY} e^{-1/Y} = e^{-1/Y} \cdot \frac{1}{Y^2}.$$

Thus, the density function of Y transforms as follows,

$$f_Y(y; \theta) = f_X(e^{-1/y}; \theta) \left| \frac{dX}{dY} \right| = \theta (e^{-1/y})^{\theta-1} e^{-1/y} \cdot \frac{1}{y^2} = \theta e^{-\theta/y} \frac{1}{y^2}.$$

The probability that Y lies between $\frac{Y}{2}$ and Y is,

$$P\left(\frac{Y}{2} \leq \theta \leq Y\right) = \int_0^{2\theta} \theta e^{-\theta/y} \frac{1}{y^2} dy = e^{-1/2} - e^{-1} = 0.239.$$



Part B

Since X^θ is uniformly distributed on $[0, 1]$, the transformed variable $T = X^\theta$ serves as a pivotal quantity. From this, we construct a pivotal interval as follows,

$$P(a < X^\theta < b) = b - a,$$

which leads to the following pivotal interval for θ ,

$$\left\{ \theta : \frac{\log b}{\log x} \leq \theta \leq \frac{\log a}{\log x} \right\}.$$

Given X^θ , the interval will have confidence 0.239 as long as $b - a = 0.239$.



Part C

The pivotal interval from Part B is a special case of intervals formed using pivotal quantities. To minimize the interval length $\log(b) - \log(a)$ subject to $b - a = 1 - \alpha$, or equivalently $\log(1 - \alpha + a) - \log a$, we choose a to maximize its lower bound. The optimal choice is $b = 1$ and $a = \alpha$, resulting in:

$$\left\{ \theta : 0 \leq \theta \leq \frac{\log(1 - 0.239)}{\log x} \right\}.$$

This provides a shorter and more efficient interval with a confidence coefficient of 0.239.



Problem 4

Solution

Part A

The interval for μ is based on \bar{x} and s , using the t-distribution:

$$\mathcal{C}_A(x) = \left\{ (\mu, \sigma^2) : \bar{x} - t_{n-1, 1-\alpha/4} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1, 1-\alpha/4} \frac{s}{\sqrt{n}} \right\}$$

where $t_{n-1, 1-\alpha/4}$ is the critical value from the t-distribution with $n-1$ degrees of freedom, reflecting the upper $1-\alpha/4$ quantile.

The interval for σ^2 is formed using the sample variance and the chi-squared distribution:

$$\mathcal{C}_A(x) = \left\{ (\mu, \sigma^2) : \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/4}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{n-1, \alpha/4}^2} \right\}$$

where $\chi_{n-1, 1-\alpha/4}^2$ and $\chi_{n-1, \alpha/4}^2$ are the critical values from the chi-squared distribution at the $1-\alpha/4$ and $\alpha/4$ quantiles, respectively.

With the Bonferroni correction, these intervals are designed to collectively maintain an overall confidence level of $1-\alpha$ when considering both parameters simultaneously:

$$\left\{ (\mu, \sigma^2) : \mu \in \left[\bar{x} - t_{n-1, 1-\alpha/4} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/4} \frac{s}{\sqrt{n}} \right], \sigma^2 \in \left[\frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/4}^2}, \frac{(n-1)s^2}{\chi_{n-1, \alpha/4}^2} \right] \right\}$$



Part B

If we replace the interval for μ with an interval based on a known variance σ^2 , the confidence intervals adjust accordingly:

$$\mathcal{C}_B(x) = \left\{ (\mu, \sigma^2) : \bar{x} - z_{1-\alpha/4} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/4} \frac{\sigma}{\sqrt{n}} \right\}$$

Here, $z_{1-\alpha/4}$ is the critical value from a $\mathcal{N}(0, 1)$, appropriate for when σ is known. Plus, the coverage probability for each parameter individually remains high, but when combined under the Bonferroni approach, it makes sure that,

$$P(\mathcal{C}_B(x) \text{ covers } (\mu, \sigma^2)) \geq 1 - \alpha$$

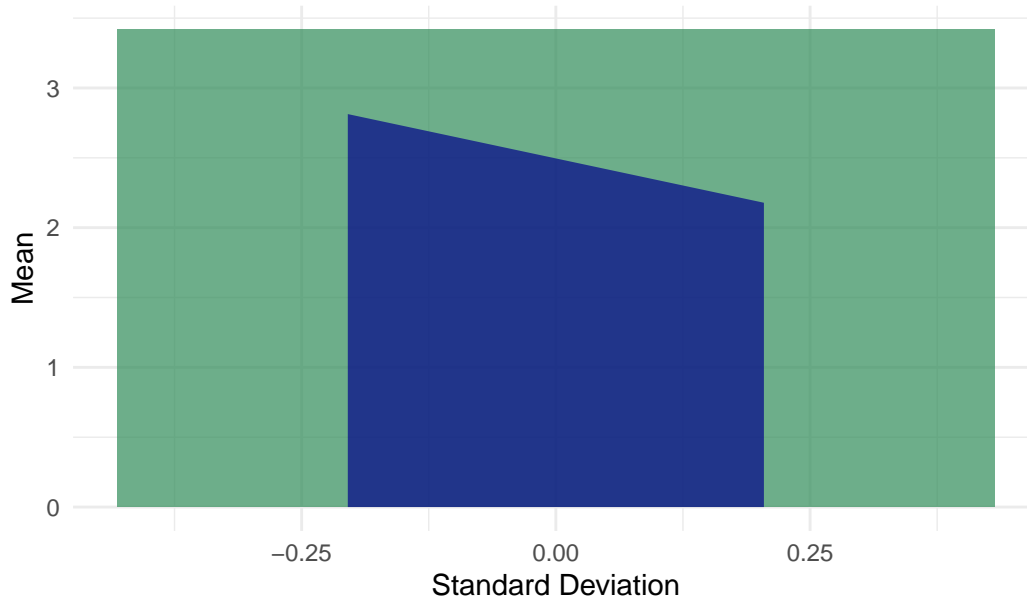
Part C

We can understand the graphical representation and comparison of the confidence regions C_a and C_b by visualizing the areas of these sets.

$$\text{Area of } C_a = 2t_{n-1,1-\alpha/4} \frac{s}{\sqrt{n}} \times \left(\sqrt{(n-1)} \left(\frac{1}{\sqrt{\chi_{n-1,\alpha/4}^2}} - \frac{1}{\sqrt{\chi_{n-1,1-\alpha/4}^2}} \right) \right)$$

$$\text{Area of } C_b = z_{\alpha/4} \frac{s}{\sqrt{n}} \times \left(\sqrt{(n-1)} \left(\frac{1}{\sqrt{\chi_{n-1,\alpha/4}^2}} - \frac{1}{\sqrt{\chi_{n-1,1-\alpha/4}^2}} \right) \right) \times \left(\frac{n-1}{\chi_{n-1,1-\alpha/4}^2} + \frac{n-1}{\chi_{n-1,\alpha/4}^2} \right)$$

Visualization of C_a and C_b



This graph helps in understanding the coverage and size differences between these two approaches, with C_a typically being more conservative due to its construction method involving the t-distribution, which accounts for the uncertainty in estimating the standard deviation s .



Problem 5

Solution

Given X_1, \dots, X_n are i.i.d. $Uniform(0, \theta)$, the maximum order statistic $Y = \max(X_1, \dots, X_n)$ has the distribution function,

$$F_Y(y) = P(Y \leq y) = P(\max(X_1, \dots, X_n) \leq y) = \prod_{i=1}^n P(X_i \leq y) = \left(\frac{y}{\theta}\right)^n$$

for $0 \leq y \leq \theta$. The transformation Y/θ is then,

$$P\left(\frac{Y}{\theta} \leq u\right) = P(Y \leq \theta u) = (\theta u / \theta)^n = u^n$$

for $0 \leq u \leq 1$, indicating that Y/θ is indeed a pivotal quantity, as its distribution does not depend on θ .

Next, the coverage probability of the interval is given by,

$$P\left(Y \leq \theta \leq \frac{Y}{a^{1/n}}\right) = P\left(\frac{Y}{\theta} \leq 1 \leq \frac{Y}{\theta} a^{1/n}\right) = P\left(\frac{Y}{\theta} \leq a^{1/n}\right)$$

Since Y/θ follows a Beta distribution, the probability that $\frac{Y}{\theta} \leq a^{1/n}$ is,

$$P\left(\frac{Y}{\theta} \leq a^{1/n}\right) = (a^{1/n})^n = a$$

Setting $a = 1 - \alpha$, the probability is $1 - \alpha$, confirming the coverage of the interval.

To show it is the shortest interval, note that any interval of the form $\{\theta : cY \leq \theta \leq dY\}$ where $c < d$ would need to satisfy,

$$P(cY \leq \theta \leq dY) = 1 - \alpha$$

Now, the factor d/c needs to be minimized to minimize the length $d - c$. Calculating for the interval given,

$$\frac{d}{c} = \frac{1}{a^{1/n}} \rightarrow \min \frac{d}{c} \text{ as } n \rightarrow \infty$$

Thus, the provided interval $\{\theta : Y \leq \theta \leq \frac{Y}{a^{1/n}}\}$ is the shortest such interval.



Problem 6

Solution

Part A

X follows a mixed normal distribution,

$$X \sim \begin{cases} N(\theta, 100) & \text{with probability } p \\ N(\theta, 1) & \text{with probability } 1 - p \end{cases}$$

The rejection rule for the null hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ is:

$$\text{Reject } H_0 \text{ if } X > \theta_0 + z_\alpha \sigma$$

where σ depends on which population is sampled (either $\sigma = 10$ or $\sigma = 1$).

To derive the $1 - \alpha$ confidence set by inverting the test, we consider that H_0 is not rejected if,

$$X \leq \theta_0 + z_\alpha \sigma$$

Rearranging this yields,

$$\theta_0 \geq X - z_\alpha \sigma$$

Hence, the confidence set for θ where we would fail to reject H_0 at level α is,

$$\theta \in [X - z_\alpha \sigma, \infty)$$

Here, σ is either 1 or 10 depending on which component of the mixture we consider sampled from. To account for the mixed nature of the distribution, the correct interpretation of σ is dependent on prior knowledge or estimation of which distribution X came from.



Part B

The proposed test is formulated as follows,

$$\text{Reject } H_0 \text{ if } X > c$$

For X under H_0 :

$$X \sim pN(0, 100) + (1 - p)N(0, 1)$$

The probability of rejecting H_0 is calculated as:

$$P_{\theta_0}(X > c) = pP\left(\frac{X}{10} > \frac{c}{10}\right) + (1 - p)P(X > c)$$

Substituting $Z \sim N(0, 1)$, this is expressed as:

$$P_{\theta_0}(X > c) = pP\left(Z > \frac{c}{10}\right) + (1 - p)P(Z > c)$$

For $c = z_{(\alpha-p)/(1-p)}$, the computation goes as:

$$P_{\theta_0}(X > c) = p + (1 - p)P\left(Z > z_{(\alpha-p)/(1-p)}\right) = p + (1 - p)\frac{\alpha - p}{1 - p} = \alpha$$

Again, this ensures the test maintains the desired level α .

The error lies in the misinterpretation of the mixture model's influence on the test's effectiveness across different parameters. While it's correct that the test maintains level α , the assumption about its power or optimality across different values of θ and mixing probabilities p without thorough analysis could be misleading. For the test to be UMP, it must be the most powerful among all tests of the same size for every possible alternative. However, given the mixed distribution and the dependency of the critical value on p , claiming it to be UMP needs careful validation through the calculation of the power function across different θ values and not just the control of the type I error rate.



Problem 7

Consider a random sample X_1, X_2, \dots, X_n from a Pareto distribution with scale parameter x_m (minimum value) and shape parameter α . The probability density function is given by,

$$f(x; x_m, \alpha) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, \quad x \geq x_m$$

Assuming x_m is known, construct a 95% pivotal confidence interval for α^2 .

Solution

Using the pdf of the Pareto distribution, one can derive the likelihood function for the shape parameter α ,

$$L(\alpha|x_1, \dots, x_n) = \prod_{i=1}^n \frac{\alpha x_m^\alpha}{x_i^{\alpha+1}} = \alpha^n x_m^{n\alpha} \prod_{i=1}^n x_i^{-\alpha-1}$$

Taking the log to simplify, we get,

$$\log L(\alpha|x_1, \dots, x_n) = n \log \alpha + n\alpha \log x_m - (\alpha + 1) \sum_{i=1}^n \log x_i$$

Differentiating with respect to α and setting to zero for maximization, we find the (MLE) for α ,

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log \left(\frac{x_i}{x_m} \right)}$$

Given the MLE of α , we know that $\hat{\alpha}$ asymptotically follows a normal distribution,

$$\hat{\alpha} \sim N \left(\alpha, \frac{\alpha^2}{nI(\alpha)} \right)$$

where $I(\alpha)$ is the Fisher information, which for the Pareto distribution is,

$$I(\alpha) = -E \left[\frac{\partial^2}{\partial \alpha^2} \log f(X; \alpha) \right] = \frac{n}{\alpha^2}$$

Now, we can form a pivot using the asymptotic normality of $\hat{\alpha}$,

$$Z = \frac{\hat{\alpha} - \alpha}{\sqrt{\frac{\alpha^2}{nI(\alpha)}}} \sim N(0, 1)$$

To find the confidence interval for α^2 , we solve for α^2 in terms of Z ,

$$\alpha^2 = \left(\hat{\alpha} - Z \sqrt{\frac{\hat{\alpha}^2}{nI(\hat{\alpha})}} \right)^2$$

Plugging in the percentiles of the standard normal distribution for Z , we find the 95% confidence interval for α^2 ,

$$\left[\left(\hat{\alpha} - 1.96 \sqrt{\frac{\hat{\alpha}^2}{nI(\hat{\alpha})}} \right)^2, \left(\hat{\alpha} + 1.96 \sqrt{\frac{\hat{\alpha}^2}{nI(\hat{\alpha})}} \right)^2 \right]$$

Therefore, we have shown how to construct a 95% confidence interval for α^2 using the MLE of the shape parameter α of the Pareto distribution.