STAT 4102 Midterm 1 Solution

2020 Fall

1 (25 points)

Let X be a random variable with a t-distribution with 10 degrees of freedom. Determine the distribution of $1/X^2$, and prove your answer.

Answer: Let Z, Y be independent random variables, and be $N(0,1)$ and χ^2_{10} distributed, respectively. Then we know

$$X =_d \frac{Z}{\sqrt{Y/10}},$$

where '$=_d$' means that the distributions of the left hand side and the right hand side are the same. For $1/X^2$, we know its distribution is the same as the distribution of $\frac{Y/10}{Z^2}$. Since Z^2 is χ^2_1 distributed, $1/X^2$ is $F_{10,1}$ distributed by the definition of F-distribution.

2 (25 points)

Suppose that $X_1, ..., X_n \overset{i.i.d}{\sim} U(\theta, 10\theta)$, $\theta > 0$.

1. Show that $\frac{2}{11}\bar{X}_n$ is an unbiased estimator of θ. (10 points)
2. Compute its MSE. (10 points)
3. Find the constant k such that $k\bar{X}_n$ has the smallest MSE (estimating θ). Is it unbiased? (5 points)

Answer:

1. Because $EX_1 = \frac{11}{2}\theta$ and $X_1, ..., X_n$ are i.i.d., we have $E\frac{2}{11}\bar{X}_n = \frac{2}{11}EX_1 = \theta$. Thus it's unbiased.

2.

$$\begin{aligned}
\text{MSE}(\frac{2}{11}\bar{X}_n) &= \text{Bias}^2(\frac{2}{11}\bar{X}_n) + \text{Var}(\frac{2}{11}\bar{X}_n) \\
&= \text{Var}(\frac{2}{11}\bar{X}_n) = \frac{4}{121n}\text{Var}(X_1) \\
&= \frac{4}{121n}\frac{(9\theta)^2}{12} = \frac{108}{484}\theta^2.
\end{aligned}$$

3.

$$\begin{aligned}\mathrm{MSE}(k\bar{X}_n) &= \mathrm{Bias}^2(k\bar{X}_n) + \mathrm{Var}(k\bar{X}_n) \\ &= (\frac{11}{2}k - 1)^2\theta^2 + k^2\frac{27}{4n}\theta^2\end{aligned}$$

Since $\theta^2 > 0$, we know the MSE achieves its smallest value when $k = \frac{11}{2}/(\frac{31}{4n} + (\frac{11}{2})^2) = \frac{22n}{27+121n}$. With this k, the estimator is biased because the expectation of it is $k\frac{11}{2}\theta \neq \theta$.

3 (25 points)

A recent poll asked n = 1000 people if they were baseball fans. Out of 1000 people surveyed, 560 of them responded "YES".

1. Find a two-sided 95% confidence interval for p (could be a large sample one), the proportion of the entire population who are baseball fans. Make sure you check the assumptions/conditions whenever you apply a formula/theorem. The final answer should be two numbers (i.e. you will need to find the values of some quantiles). (18 points)
2. Suppose the confidence interval you calculated is (a, b). Here, a, b are real numbers, not random variables. Your friend look at it and says "so the probability that p is in (a, b) is 95%". Do you think your friend's right? If not, please write several sentences telling him/her why he/she is wrong, and correct the statement. (7 points)

Answer:

1. Let Y be the number of people responded "Yes" out of the 1000 people surveyed. Then we can assume that $Y \sim$ Bin(n,p), where $n = 1000$. Since Y have the same distribution as the sum of n i.i.d. Bernoulli(p) distributed random variables, by the CLT, $Y \sim_{approx.} N(np, np(1-p))$. Therefore,

$$1 - \alpha \approx P(z_{1-\alpha/2} \leq \sqrt{n}\frac{Y/n - p}{\sqrt{p(1-p)}} \leq z_{\alpha/2}).$$

Since we have a large sample, $\hat{p} := Y/n$ is close to p. By replacing the p in the denominator by $\hat{p}$, we still get the same approximate probability (this argument can be made rigorous by using theorem 9.3, but mentioning that we did an approximation here and it's only possible when the sample size is large is enough, since chapter 9 is not required in this exam). That is

$$\begin{aligned}1 - \alpha &\approx P(z_{1-\alpha/2} \leq \sqrt{n}\frac{Y/n - p}{\sqrt{\hat{p}(1-\hat{p})}} \leq z_{\alpha/2}) \\ &= P(z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} \leq Y/n - p \leq z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}) \\ &= P(-z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} + Y/n \geq p \geq -z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} + Y/n).\end{aligned}$$

Because $z_{\alpha/2} = 1.96$ and $z_{1-\alpha/2} = -1.96$, we get the 95% confidence interval $(0.53, 0.59)$.

2. My friend's wrong. The probability of $(0.53, 0.59)$ containing p is either 1 or 0, can't be 95%, since p is just a fixed number, even though unknown. The correct statement is that we have 95% confidence that $(0.53, 0.59)$ contains p.

4 (25 points)

Suppose $X_1, ..., X_{n_X}, Y_1, ..., Y_{n_Y}, Z_1, ..., Z_{n_Z} \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, where $n_X, n_Y, n_Z > 1$ are integers.

1. Find the constant k such that

$$T = \frac{\bar{X}/3 + \bar{Y}/3 + \bar{Z}/3 - \mu}{kS_X}$$

is t-distributed and find the degrees of freedom. Here $S_X = \sqrt{S_X^2}$, the square root of sample variance of X's. (15 points)

2. Construct a $1 - \alpha$ confidence interval for μ based on the pivot T. (10 points)

Answer:

1.Let $k = \frac{1}{3}\sqrt{\frac{1}{n_X} + \frac{1}{n_Y} + \frac{1}{n_Z}}$, then we have $\frac{(\bar{X}+\bar{Y}+\bar{Z})/3-\mu}{k} \sim N(0, \sigma^2)$. By our theorem, $(n_X - 1)S_X^2/\sigma^2 \sim X_{n_X-1}^2$, thus $S_X \sim \sigma\sqrt{\chi_{n_X-1}^2/(n_X - 1)}$. Since for the normal case, the sample variance and the sample mean are independent, and also the X's Y's and Z's are joint independent, we have that S_X is independent of $\bar{X}+\bar{Y}+\bar{Z}$. Combining these together, we have that $T \sim \frac{N(0,1)}{\sqrt{\chi_{n_X-1}^2/(n_X-1)}}$, where the normal and chi-squared are independent. Thus T is t-distributed by definition and the degrees of freedom is $n_X - 1$.

2.

$$\begin{aligned}
1 - \alpha &= P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) \\
&= P(-t_{\alpha/2}kS_X \leq (\bar{X} + \bar{Y} + \bar{Z})/3 - \mu \leq t_{\alpha/2}kS_X) \\
&= P((\bar{X} + \bar{Y} + \bar{Z})/3 + t_{\alpha/2}kS_X \geq \mu \geq (\bar{X} + \bar{Y} + \bar{Z})/3 - t_{\alpha/2}kS_X)
\end{aligned}$$

Done.