PROBABILITY AND STATISTICS I
HOMEWORK XIII

Jianyu Dong 2019511017

May, 29 2021

1

Since $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^m$ are random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and the two samples are independent, we have that the sample means $\overline{X}_n$ follows $N(\mu_1, \frac{\sigma_1^2}{n})$ and $\overline{Y}_m$ follows $N(\mu_2, \frac{\sigma_2^2}{m})$. So that we have $\overline{X}_n - \overline{Y}_m$ follows $N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$.

Since $\sigma_1 = \sigma_2$, let $\sigma_1 = \sigma_2 = \sigma$. Thus, we have that $\overline{X}_n - \overline{Y}_n$ follows $N(\mu_1 - \mu_2, \frac{\sigma^2}{n+m})$.

Let the sample variance be

$$S^2 = \frac{n-1}{n+m-2}S_X^2 + \frac{m-1}{n+m-2}S_Y^2$$

Since all parameters are unknown, by the theorem, we have that $\frac{(\overline{X}_n - \overline{Y}_n) - (\mu_1 - \mu_2)}{S\sqrt{1/m + 1/n}}$ follows t-distribution with parameter n+m-2. So we could formulate that

$$\mathbf{P}\left(-t_{\alpha/2,n+m-2} < \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} < t_{\alpha/2,n+m-2}\right) = 1 - \alpha,$$

where $t_{\alpha/2,n+m-2}$ is the 1-α/2 quantile of the t distribution with degrees of freedom n+m-2. Rewriting the inequality, we obtain that

$$\mathbf{P}\left((\overline{X}_n - \overline{Y}_m) - t_{\alpha/2,n+m-2}S\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_1 - \mu_2 < (\overline{X}_n - \overline{Y}_m) + t_{\alpha/2,n+m-2}S\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = 1 - \alpha.$$

Thus, the two-sided coefficient $1 - \alpha$ confidence interval for $\mu_1 - \mu_2$ is

$$\left((\overline{X}_n - \overline{Y}_m) - t_{\alpha/2,n+m-2}S\sqrt{\frac{1}{n} + \frac{1}{m}},\ (\overline{X}_n - \overline{Y}_m) + t_{\alpha/2,n+m-2}S\sqrt{\frac{1}{n} + \frac{1}{m}}\right).$$

2

Since $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^m$ are random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and the two samples are independent, we have that $\frac{S_X^2/S_Y^2}{\sigma_1^2/\sigma_2^2}$ follows F distribution with n-1 and m-1 degrees of freedom. So we could formulate that

$$\mathbf{P}\left(F_0^{-1}(\alpha/2) < \frac{S_X^2/S_Y^2}{\sigma_1^2/\sigma_2^2} < F_0^{-1}(1-\alpha/2)\right) = 1 - \alpha,$$

where F_0 is the c.d.f. of the F distribution with n-1 and m-1 degrees of freedom and S_X^2, S_Y^2 are the sample variances of $(X_i)_{i=1}^n$, $(Y_j)_{j=1}^m$,

$$S_X^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2,\ S_Y^2 = \sum_{j=1}^m (Y_j - \overline{Y}_m)^2.$$

Rewriting the inequality, we have that

$$\mathbf{P}\left(\frac{S_X^2/S_Y^2}{F_0^{-1}(1-\alpha/2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_X^2/S_Y^2}{F_0^{-1}(\alpha/2)}\right).$$

Thus, the two-sided coefficient $1 - \alpha$ confidence interval for σ_1^2/σ_2^2 is

$$\left(\frac{S_X^2/S_Y^2}{F_0^{-1}(1-\alpha/2)},\ \frac{S_X^2/S_Y^2}{F_0^{-1}(\alpha/2)}\right).$$

3

Since $(X_i)_{i=1}^n$ are random samples from $N(\mu, \sigma 2)$ and $\sigma^2 = 10$, we have $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ follows N(0,1). According to the question, we have that

$$\mathbf{P}\left(\overline{X} - \frac{1}{2} < \mu < \overline{X} + \frac{1}{2}\right) = 0.954.$$

Rewriting the inequality, we obtain that

$$\mathbf{P}\left(-\frac{1}{2\sigma/\sqrt{n}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{1}{2\sigma/\sqrt{n}}\right) = 0.954.$$

So that

$$\Phi(\frac{1}{2\sigma/\sqrt{n}}) = 0.977.$$

According to the table, we have that

$$\frac{1}{2\sigma/\sqrt{n}} = 2.00.$$

Since $\sigma = \sqrt{10}$, we have that n is 160.

4

Since $(X_i)_{i=1}^n$ are random samples from $\Gamma(3, \beta)$, by the theorem, we have that $Y_i = 2\beta X_i$ follows $\chi^2(6)$. By the theorem, we have that $\sum_{i=1}^n Y_i$ follows $\chi^2(6n)$. By the theorem, we have the mean of $\sum_{i=1}^n Y_i$ is 6n. To obtain a coefficient 1-α confidence interval, we have that

$$\mathbf{P}(-k_{6n,1-\alpha/2} < n\overline{Y}_n < k_{6n,\alpha/2}) = 1 - \alpha,$$

where $k_{6n,\alpha/2}$ is the value that $\mathbf{P}(n\overline{Y}_n > k_{6n,\alpha/2}) = \alpha/2$ and $k_{6n,1-\alpha/2}$ is the value that $\mathbf{P}(n\overline{Y}_n > k_{6n,1-\alpha/2}) = 1 - \alpha/2$.

Rewriting the inequality, we obtain that

$$\mathbf{P}\left(\frac{k_{6n,1-\alpha/2}}{2n\overline{X}_n} < \beta < \frac{k_{6n,\alpha/2}}{2n\overline{X}_n}\right) = 1 - \alpha,$$

where $\overline{X}_n$ is the sample mean.

Thus, the two-sided coefficient $1 - \alpha$ confidence interval for β is

$$\left(\frac{k_{6n,1-\alpha/2}}{2n\overline{X}_n}, \frac{k_{6n,\alpha/2}}{2n\overline{X}_n}\right)$$

5

Since $(X_i)_{i=1}^n$ are random samples from $N(\mu, 16)$, we have that $\frac{\overline{X}-\mu}{4/\sqrt{n}}$ follows N(0,1). To obtain a coefficient 1-α confidence interval with length $\leqslant L$, we have that

$$\mathbf{P}\left(\overline{X}_n - \frac{L}{2} < \mu < \overline{X}_n + \frac{L}{2}\right) = 1 - \alpha.$$

Rewriting the inequality, we obtain that

$$\mathbf{P}\left(-\frac{\sqrt{n}L}{8} < \frac{\overline{X}-\mu}{4/\sqrt{n}} < \frac{\sqrt{n}L}{8}\right) = 1 - \alpha.$$

Let $z_{\alpha/2}$ be the value that $\mathbf{P}(X > z_{\alpha/2}) = \alpha/2$, so that we get

$$\frac{\sqrt{n}L}{8} = z_{\alpha/2}.$$

Thus, the smallest n is

$$n = \left\lceil \frac{64 z_{\alpha/2}^2}{L^2} \right\rceil.$$

6

Since X has a Possion distribution with mean θ, we have the p.d.f. of X given θ is

$$p(X = x; \theta) = \frac{e^{-\theta}\theta^x}{x!}, \ \ for\ x = 0, 1, 2, \dots$$

zero elsewhere.

Since $(X_i)_{i=1}^{10}$ are independent and each X_i follows Possion distribution with mean θ, then we have $\sum_{i=1}^{10} X_i$ follows the Possion distribution with mean 10θ.

We have that the sample mean $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$ is an estimator of θ. We could determine the c.d.f. of $\overline{X}$ is

$$F_{\overline{X}}(\overline{X}; \theta) = \mathbf{P}(10\overline{X} \leqslant 5) = \sum_{n=0}^{5} \frac{e^{-10\theta}(10\theta)^n}{n!}.$$

Then we could determine that

$$\frac{d}{d\theta}\mathbf{P}(10\overline{X} \leqslant 5) = \sum_{n=0}^{5} 10e^{-10\theta}\frac{(10\theta)^{n-1}}{(n-1)!}(1 - \frac{10\theta}{n}) < 0.$$

By the definition, there is

$$F_{\overline{X}}(\overline{X}-; \underline{\theta}) = \sum_{n=0}^{4} \frac{e^{-10\underline{\theta}}(10\underline{\theta})^n}{n!} = 0.95$$

$$F_{\overline{X}}(\overline{X}; \overline{\theta}) = \sum_{n=0}^{5} \frac{e^{-10\overline{\theta}}(10\overline{\theta})^n}{n!} = 0.05$$

which implies that

$$e^{-10\underline{\theta}}\left(1 + 10\underline{\theta} + \frac{10^2}{2}\underline{\theta}^2 + \frac{10^3}{6}\underline{\theta}^3 + \frac{10^4}{24}\underline{\theta}^4\right) = 0.95$$

$$e^{-10\overline{\theta}}\left(1 + 10\overline{\theta} + \frac{10^2}{2}\overline{\theta}^2 + \frac{10^3}{6}\overline{\theta}^3 + \frac{10^4}{24}\overline{\theta}^4 + \frac{10^5}{120}\overline{\theta}^5\right) = 0.05$$

Then we could determine that

$$\underline{\theta} = 0.1970, \ \overline{\theta} = 1.0513$$

Thus, the confidence interval with coefficient of at least 0.9 is (0.1970,1.0513).

7

Since we have the 95% confidence interval for the parameter μ of Possion(μ) distribution is (2,3), there is

$$\mathbf{P}(2 < \mu < 3) = 95\%.$$

Then we let $z = e^{-\mu}$, so we get that $\mu = -\log z$ which is strictly decreasing. Thus, there is

$$\mathbf{P}(2 < -\log z < 3) = 95\%.$$

Rewriting the inequality, there is that

$$\mathbf{P}(e^{-3} < z < e^{-2}) = 95\%.$$

So that the 95% confidence interval for $\mathbf{P}(X = 0) = e^{-\mu}$ is (e^{-3}, e^{-2}).

8

8.1

Let $\overline{X}_{16} = \frac{1}{16}\sum_{i=1}^{16} X_i$ be the sample mean and $S_X^2 = \sum_{i=1}^{16}(X_i - \overline{X}_{16})^2$ be the sample variance.

By the theorem, we have that $Y = \frac{\overline{X}_{16}-\mu}{S_X/\sqrt{16}}$ follows the t distribution with 15 degrees of freedom. So that we have

$$\mathbf{P}\left(-t_{0.025,15} < \frac{\overline{X}_{16} - \mu}{S_X/\sqrt{16}} < t_{0.025,15}\right) = 0.95$$

where $t_{0.025,15}$ is the value that $\mathbf{P}(Y > t_{0.025,15}) = 0.025$

Rewriting the inequality, we get that

$$\mathbf{P}\left(\overline{X}_{16} - t_{0.025,15}\frac{S_X}{4} < \mu < \overline{X}_{16} + t_{0.025,15}\frac{S_X}{4}\right) = 0.95.$$

According to the question, we have that

$$\mathbf{P}(1.6 < \mu < 7.8) = 0.95.$$

Thus, we could determine the mean of the dataset is

$$\overline{X}_{16} = \frac{1.6 + 7.8}{2} = 4.7.$$

8.2

Searching for table we have that $t_{0.025,15} = 2.131$. Then we have that $4.7 - 2.131 \times \frac{S_X}{4} = 1.6$. Thus, we get that $S_X = 5.819$

If we perfer to have 99% confidence interval for μ, by the theorem, we have that

$$\mathbf{P}\left(-t_{0.005,15} < \frac{\overline{X}_{16} - \mu}{S_X/\sqrt{16}} < t_{0.005,15}\right) = 0.99$$

Rewriting the inequality, we have that

$$\mathbf{P}\left(\overline{X}_{16} - t_{0.005,15}\frac{S_X}{4} < \mu < \overline{X}_{16} + t_{0.005,15}\frac{S_X}{4}\right) = 0.99$$

Searching for the table we have that $t_{0.005,15} = 2.947$. Then we could determine that

$$\mathbf{P}\left(0.413 < \mu < 8.987\right) = 0.99$$

Thus, the 99% confidence interval for μ is (0.413,8.987).