Mid-term Exam: Statistical Inference

1 (10') Suppose that $X_1, \cdots, X_m$ i.i.d. $\sim N(\mu_1, \sigma^2)$, $Y_1, \cdots, Y_n$ i.i.d. $\sim N(\mu_2, \sigma^2)$, and X_i's and Y_j's are independent. Let $\bar{X}, \bar{Y}, S_X^2, S_Y^2$ denote their sample means and sample variances. Determine the distribution of

$$T = \frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{\alpha^2}{m} + \frac{\beta^2}{n}\right)}},$$

where α, β are fixed constants.

2 (20') Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from Normal distribution $N(\theta, 1)$.

(i) (5') Derive the moment estimator of θ^2.

(ii) (5') Derive the MLE of θ^2.

(iii) (5') Derive the UMVUE of θ^2.

(iv) (5') Is the UMVUE an efficient estimator of θ^2? Why?

3 (20') Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from the distribution with p.d.f.

$$f(x; \alpha, \beta) = \frac{1}{\beta} e^{-(x-\alpha)/\beta}, \ x \geq \alpha, \ \alpha \in R, \ \beta > 0.$$

Find sufficient statistic and MLE of

(i) (5') α when β is known.

(ii) (5') β when α is known.

(iii) (10') α and β when both are unknown.

4 (15') In order to study the height of male students in a university, we took a random sample of size 5 and observed their heights (cm): $174, 171, 168, 175, 170$. For simplicity, suppose that the variance of height is $\sigma^2 = 9$ (cm^2).

(i) (5') Estimate the mean height (μ) of male students in this university.

(ii) (5') Construct a 99% confidence interval for μ. ($z_{0.025} = 1.96, z_{0.005} = 2.58, t_{4,0.005} = 4.60$)

(iii) (5') Determine the sample size n such that the length of confidence interval can be reduced by 80%.

5 (20') Let r.v. X be the number of goals scored by teams during the first round matches of the 2002 World Cup and suppose that X follows a Poisson distribution $P(\lambda)$ with p.d.f.

$$f(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \lambda > 0, x = 0, 1, 2, \cdots$$

Let $X_1, \cdots, X_n$ be a random sample.

(i) (5') Determine the distribution of $\sum_{i=1}^n X_i$.

(ii) (10') The observed values of $X_1, \cdots, X_n$ with $n = 95$ are summarized in the following table. Derive the MLE of λ and compute its efficiency.

Goals	0	1	2	3	4	5	6	7	8
Frequency	23	37	20	11	2	1	0	0	1

(iii) (5') Construct an approximately 99% confidence interval for λ. ($z_{0.025} = 1.96, z_{0.005} = 2.58$)

6 (15') The independent random samples $X_i, i = 1, \cdots, 5$ and $Y_i, i = 1, \cdots, 5$ represent resistance measurements taken on two test pieces, and the observed values (in ohms) are as follows:

$$x_1 = 0.3,\ x_2 = 0.2,\ x_3 = 0.1,\ x_4 = 0.2,\ x_5 = 0.1,$$

$$y_1 = 0.2,\ y_2 = 0.1,\ y_3 = 0.3,\ y_4 = 0.2,\ y_5 = 0.2.$$

Assume that $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_i \sim N(\mu_2, \sigma_2^2)$, $i = 1, \cdots, 5$. ($z_{0.025} = 1.96, t_{4,0.025} = 2.78, t_{4,0.01266} = 3.48, t_{4,0.0125} = 3.50, t_{8,0.025} = 2.31, t_{10,0.025} = 2.23, F_{4,4,0.025} = 9.60, F_{4,4,0.975} = 0.10, F_{5,5,0.025} = 7.15, F_{5,5,0.975} = 0.14$)

(i) (10') Construct a 95% confidence interval for $\mu_1 - \mu_2$.

(ii) (5') Construct a 95% confidence region for (μ_1, μ_2).

Solution of Mid-term Exam

Ke Zhu & Bo Yu & Saidi Luo

2018/11/29

1 (10 pts)

We know that$\bar{X} \sim N(\mu_1, \frac{\sigma^2}{m})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma^2}{n})$. Since X_i's and Y_j's are independent, $\bar{X}, \bar{Y}$ are independent. $\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2) \sim N(0, \frac{\alpha^2}{m} + \frac{\beta^2}{n})$, then we can get

$$f(\bar{X}, \bar{Y}) = [\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)]/(\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}\sigma) \sim N(0,1)$$

Since $(m-1)S_X^2/\sigma^2 \sim \chi^2_{m-1}$, $(n-1)S_Y^2/\sigma^2 \sim \chi^2_{n-1}$ and X_i's and Y_j's are independent, S_X^2, S_Y^2 are independent and

$$g(S_X^2, S_Y^2) = (m-1)S_X^2/\sigma^2 + (n-1)S_Y^2/\sigma^2 \sim \chi^2_{m-1+n-1} = \chi^2_{m+n-2}$$

$\bar{X}, S_X^2$ are independent,$\bar{Y}, S_Y^2$ are independent and X_i's and Y_j's are independent, so $(\bar{X}, \bar{Y})$ and (S_X^2, S_Y^2) are independent. $f(\bar{X}, \bar{Y})$ and $g(S_X^2, S_Y^2)$ are independent. Therefore,

$$\begin{aligned}
t_{m+n-2} &= \frac{N(0,1)}{\sqrt{\chi^2_{m+n-2}/(m+n-2)}} \qquad \text{(Need independence)} \\
&= \frac{f(\bar{X}, \bar{Y})}{\sqrt{g(S_X^2, S_Y^2)/(m+n-2)}} \\
&= \frac{[\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)]/(\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}\sigma)}{\sqrt{[(m-1)S_X^2/\sigma^2 + (n-1)S_Y^2/\sigma^2]/(m+n-2)}} \\
&= \frac{[\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)]/\sigma}{\sqrt{\frac{[(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}(\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}})/\sigma} \qquad \text{(Target's distribution)}
\end{aligned}$$

2 (20 pts)

(i) Since $\theta^2 = (EX_1)^2$, the moment estimator of θ^2 is $\bar{X}^2$.

(ii) The log likelihood function of $\mathbf{X} = \mathbf{x}$ is

$$\log L(\theta|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2.$$

Let

$$\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) = n(\bar{x} - \theta) = 0,$$

hence $\theta = \bar{x}$. Furthermore,

$$\frac{\partial^2}{\partial\theta^2}\log L(\theta|\mathbf{x}) = -n < 0.$$

It follows that the MLE of θ is $\bar{X}$. According to the invariant property of MLE, the MLE of θ^2 is $\bar{X}^2$.

(iii) According to the property of exponential family, $\bar{X}$ is complete and sufficient for θ^2. We have $\bar{X} \sim N(\theta, \frac{1}{n})$, thus

$$E(\bar{X}^2 - \frac{1}{n}) = \theta^2.$$

According to L-S theorem, $\bar{X}^2 - \frac{1}{n}$ is the UMVUE of θ^2.

(iv) The fisher information of θ is

$$I(\theta) = E\left[\frac{\partial \log f(X;\theta)}{\partial\theta}\right]^2 = E(X-\theta)^2 = 1.$$

Let $\eta = g(\theta) = \theta^2$. The variance of $\hat{\eta} = \bar{X}^2 - \frac{1}{n}$ is

$$\begin{aligned}
\operatorname{Var}(\hat{\eta}) &= \operatorname{Var}(\bar{X}^2) \\
&= E(\bar{X}^4) - (E(\bar{X}^2))^2 \\
&= (\theta^4 + \frac{6\theta^2}{n} + \frac{3}{n^2}) - (\theta^2 + \frac{1}{n})^2 \\
&= (\theta^4 + \frac{6\theta^2}{n} + \frac{3}{n^2}) - (\theta^4 + \frac{2\theta^2}{n} + \frac{1}{n^2}) \\
&= \frac{4\theta^2}{n} + \frac{2}{n^2}.
\end{aligned}$$

The efficiency of $\hat{\eta} = \bar{X}^2 - \frac{1}{n}$ is

$$\begin{aligned}
e(\hat{\eta}) &= \frac{[g'(\theta)]^2/(nI(\theta))}{\operatorname{Var}(\hat{\eta})} \\
&= \frac{4\theta^2}{n(\frac{4\theta^2}{n} + \frac{2}{n^2})} \\
&= \frac{1}{1 + \frac{1}{2n\theta^2}} \\
&< 1
\end{aligned}$$

Thus, $\hat{\eta}$ is not an efficient estimator of θ^2.

3 (20 pts)

(i)

$$f(x_1, x_2, \ldots, x_n; \alpha, \beta) = \frac{1}{\beta^n} \exp\left(-\frac{\sum_{i=1}^n x_i - n\alpha}{\beta}\right) I_{[\alpha,+\infty)}\left(x_{(1)}\right)$$

so $X_{(1)}$ is a sufficient statistic of α.

(ii)

$$f(x_1, x_2, \ldots, x_n; \alpha, \beta) = \frac{1}{\beta^n} e^{\frac{n\alpha}{\beta}} \exp\left(-\frac{\sum_{i=1}^n x_i}{\beta}\right) I_{[\alpha,+\infty)}\left(x_{(1)}\right)$$

so $\Sigma_{i=1}^n X_i$ is a sufficient statistic of β.

$$\frac{\partial \log f}{\partial \beta} = -\frac{n}{\beta} + \frac{\sum_{i=1}^n x_i - n\alpha}{\beta} \hat{\beta}_{MLE} = \frac{\sum_{i=1}^n x_i - n\alpha}{n}$$

(iii)

$$f(x_1, x_2, \ldots, x_n; \alpha, \beta) = \frac{1}{\beta^n} \exp\left(-\frac{\sum_{i=1}^n x_i - n\alpha}{\beta}\right) I_{[\alpha,+\infty)}\left(x_{(1)}\right)$$

So $(X_{(1),\Sigma_{i=1}^n X_i})$ is sufficient and complete. From (ii),(iii), we conclude

$$\hat{\alpha}_{MLE} = X_{(1)} \hat{\beta}_{MLE} = \frac{\sum_{i=1}^n x_i - n\hat{\alpha}}{n} = \frac{\sum_{i=1}^n x_i - nX_{(1)}}{n}$$

4 (15 pts)

(i) Let X denote the height of a male student in this university. We can estimate μ by either moment estimator or MLE. Moment estimator is $\hat{\mu} = \bar{X}$. MLE is $\hat{\mu} = \bar{X}$ under the assumption of $X \sim N(\mu, 9)$.

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{5}(174 + 171 + 168 + 175 + 170) = 171.6(cm).$$

Note that if we use MLE, a reasonable assumption of distribution is required.

(ii) Since $X_1, X_2, ..., X_n \sim i.i.d. N(\mu, 9), \bar{X} \sim N(\mu, \frac{9}{n})$ $, \frac{\bar{X}-\mu}{\sigma} = \frac{\bar{X}-\mu}{3} \sim N(0, \frac{1}{n})$ so

$$\mathbf{P}_\mu\Big(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \in (-z_{0.005}, z_{0.005})\Big) = 0.99$$

and $n = 5, \bar{X} = 171.6$. Therefore,

$$[\bar{X} - \frac{\sigma}{\sqrt{n}}z_{0.005}, \bar{X} + \frac{\sigma}{\sqrt{n}}z_{0.005} = [168.1, 175.1]$$

is a 99% CI forμ.

(iii) From (ii), the length of the CI is $2\sigma z_{0.005}/\sqrt{n}$. We need a $\tilde{n}$, s.t.

$$\frac{2\sigma z_{0.005}}{\sqrt{\tilde{n}}} \leq 0.2 \cdot \frac{2\sigma z_{0.005}}{\sqrt{n}} \quad \Rightarrow \quad \tilde{n} \geq 25n = 125.$$

5 (20 pts)

(i)

$$\phi_X(t) = \exp(\lambda e^{it}) \quad \Rightarrow \quad \phi_{\Sigma_{i=1}^n X_i}(t) = \exp(n\lambda e^{it}) \quad \Rightarrow \quad \Sigma_{i=1}^n X_i \sim P(n\lambda)$$

(ii)

$$\log f(x_1, ..., x_n|\lambda) = \Sigma_{i=1}^n x_i \log\lambda - \Sigma_{i=1}^n \log(x_i!) - n\lambda$$

$$\frac{\partial \log f}{\partial\lambda} = \frac{\Sigma_{i=1}^n x_i}{\lambda} - n, \quad \frac{\partial^2 \log f}{\partial^2\lambda} = -\frac{\Sigma_{i=1}^n x_i}{\lambda^2} < 0 \quad \Rightarrow \quad \hat{\lambda}_{MLE} = \bar{X} = 1.38$$

$$I(\lambda) = -E(\frac{\partial^2 \log f}{\partial^2\lambda}) = \frac{n}{\lambda}, \quad Var(\bar{X}) = \frac{\lambda}{n} \quad \Rightarrow e_{\bar{X}} = \frac{1/I(\lambda)}{Var(\bar{X})} = 1$$

(iii)

$$\frac{\sqrt{n}(\bar{X}-\lambda)}{\sqrt{\lambda}} \rightarrow_d N(0,1) \quad \Rightarrow \quad P(|\frac{\sqrt{n}(\bar{X}-\lambda)}{\sqrt{\lambda}}| \leq z_{0.005}) \approx 0.99 \quad \Rightarrow \quad P(\lambda \in [1.10, 1.73]) \approx 0.99$$

6 (15 pts)

(i) If $\sigma_1 = \sigma_2 = \sigma$ are unknown,

$$[(\bar{(}X) - \bar{(}Y)) - (\mu_1 - \mu_2)]/[\sigma\sqrt{1/5 + 1/5}] \sim N(0, 1).$$

The pivot statistic

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{4S_X^2 + 4S_Y^2}{8}(\frac{1}{5} + \frac{1}{5})}} \sim t_8.$$

Hence a 95% CI for $\mu_1 - \mu_2$ is

$$\left[(\bar{X}-\bar{Y}) - t_{8,0.025}\sqrt{\frac{4S_X^2+4S_Y^2}{8}(\frac{1}{5}+\frac{1}{5})}, (\bar{X}-\bar{Y}) + t_{8,0.025}\sqrt{\frac{4S_X^2+4S_Y^2}{8}(\frac{1}{5}+\frac{1}{5})}\right] = [-0.133, 0.093].$$

If $\sigma_1 \neq \sigma_2$, we need to calculate

$$\left(\frac{S_X^2}{m}+\frac{S_Y^2}{n}\right)^2 / \left[\frac{S_X^4}{m^2(m-1)}+\frac{S_Y^4}{n^2(n-1)}\right] = 7.784.$$

then its nearest r is 8. For $m = n = 5, \alpha = 0.05$, the approximately 95% CI for μ_1, μ_2 is

$$\left[\overline{X} - \overline{Y} - t_{r,\alpha/2}\sqrt{S_X^2/m + S_Y^2/n}, \overline{X} - \overline{Y} + t_{r,\alpha/2}\sqrt{S_X^2/m + S_Y^2/n}\right] = [-0.133, 0.093].$$

(ii) By the fact that $\frac{\bar{X}-\mu_1}{S_X/\sqrt{m}} \sim t_{m-1}$, $\frac{\bar{Y}-\mu_2}{S_Y/\sqrt{n}} \sim t_{n-1}$(m = n = 5), and by the independence between X and Y,

$$\begin{aligned}
&P(\frac{\bar{X}-\mu_1}{S_X/\sqrt{5}} \in [-t_{4,0.01266}, t_{4,0.01266}], \frac{\bar{Y}-\mu_2}{S_Y/\sqrt{5}} \in [-t_{4,0.01266}, t_{4,0.01266}]) \\
=&P(\frac{\bar{X}-\mu_1}{S_X/\sqrt{5}} \in [-t_{4,0.01266}, t_{4,0.01266}]) \times P(\frac{\bar{Y}-\mu_2}{S_Y/\sqrt{5}} \in [-t_{4,0.01266}, t_{4,0.01266}]) \\
=&(1-2\times 0.01266)^2 \\
=&0.95.
\end{aligned}$$

Then (μ_1, μ_2)'s 95% CI is

$$[\bar{X}-t_{4,0.01266}\frac{S_X}{\sqrt{5}}, \bar{X}+t_{4,0.01266}\frac{S_X}{\sqrt{5}}]\times[\bar{Y}-t_{4,0.01266}\frac{S_Y}{\sqrt{5}}, \bar{Y}+t_{4,0.01266}\frac{S_Y}{\sqrt{5}}] = [0.050, 0.310]\times[0.090, 0.310]$$