MATH3423 Statistical Inference

Classwork 2

1. (**6 marks**) Let X_1, ...,X_n be a random sample from N(μ,σ^2). Define

$$\bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$$

$$\bar{X}_{n-k} = \frac{1}{n-k}\sum_{i=1+k}^{n} X_i$$

and

$$S_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \bar{X}_k)^2,$$

$$S_{n-k}^2 = \frac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \bar{X}_{n-k})^2,$$

Answer the following:

(a) What is the distribution of $((k-1)S_k^2 + (n-k-1)S_{n-k}^2)/\sigma^2$?

(b) What is the distribution of S_k^2/S_{n-k}^2 ?

(c) What is the distribution of $(\bar{X}_k + \bar{X}_{n-k})/2$?

(d) What is the distribution of $(\bar{X}_n - \mu)/(S_n/\sqrt{n})$?

If $\mu = 0$ $\sigma = 1$,

(e) What is the distribution of X_1/X_2 ?

(f) What is the distribution of $(X_1 + X_2)^2/(X_1 - X_2)^2$?

Solution:

(a) Since $X_i \overset{\text{i.i.d.}}{\sim}$ N(μ,σ^2), according to theorem 4.2 and theorem 7.3 in lecture notes, we have

$$\begin{aligned}
\bar{X}_k &\sim N(\mu, \frac{\sigma^2}{k}) && (1)\\
\bar{X}_{n-k} &\sim N(\mu, \frac{\sigma^2}{n-k}) && (2)\\
\frac{(k-1)S_k^2}{\sigma^2} &\sim \chi^2(k-1) && (3)\\
\frac{(n-k-1)S_{n-k-1}^2}{\sigma^2} &\sim \chi^2(n-k-1) && (4)
\end{aligned}$$

Since X_i are i.i.d. and S_k^2 and $\bar{X}_k$ involves X_1,X_2,...,X_k, and S_{n-k}^2 and $\bar{X}_{n-k}$ only involves, X_{k+1},X_{k+2},...,X_n, we have S_k^2 and S_{n-k}^2 are independent and $\bar{X}_k$ and $\bar{X}_{n-k}$ are independent. Therefore,

$$\frac{(k-1)S_k^2 + (n-k-1)S_{n-k}^2}{\sigma^2} \sim \chi^2(k-1+n-k-1) = \chi^2(n-2)$$

(b) By definition of F distribution, and (3) and (4) above,

$$\frac{S_k^2}{S_{n-k}^2} = \frac{\frac{(k-1)S_k^2/\sigma^2}{k-1}}{\frac{(n-k-1)S_{n-k-1}^2/\sigma^2}{n-k-1}} \sim F(k-1, n-k-1)$$

(c) Since $\bar{X}_k$ and $\bar{X}_{n-k}$ are independent, and (3) and (4) above, we have

$$\begin{aligned}
E((\bar{X}_k + \bar{X}_{n-k})/2) &= (E(\bar{X}_k) + E(\bar{X}_{n-k}))/2 = \mu\\
Var((\bar{X}_k + \bar{X}_{n-k})/2)) &= Var(\bar{X}_k)/4 + Var(\bar{X}_{n-k})/4 + Cov(\bar{X}_k, \bar{X}_{n-k})/2 = (\frac{1}{4k} + \frac{1}{4(n-k)})\sigma^2\\
(\bar{X}_k + \bar{X}_{n-k})/2 &\sim N(\mu, (\frac{1}{4k} + \frac{1}{4(n-k)})\sigma^2)
\end{aligned}$$

(d) Since $\dfrac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$ and they are independent

$$\Rightarrow \quad \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} / \sqrt{\frac{(n-1)S_n^2}{(n-1)\sigma^2}} = \frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

If $\mu = 0$ $\sigma = 1$,

(e) Notice $\dfrac{X_1}{\sqrt{X_2^2/1}} \sim t$, and $X_2 \sim N(0,1)$ is symmetrically distributed $\Rightarrow \quad \dfrac{X_1}{\sqrt{X_2^2/1}} = \dfrac{X_1}{X_2} \sim t_1$

(f) It is easy to know $(X_1,X_2)^T$ follows bivariate normal, note that $\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}X_1\\ X_2\end{pmatrix} = \begin{pmatrix}X_1+X_2\\ X_1-X_2\end{pmatrix} \Rightarrow X_1+X_2$ and X_1-X_2 also follows bivariate normal, since $Cov(X_1+X_2, X_1-X_2) = 0$, they are independent. Notice $X_1+X_2 \sim N(0,2)$ and $X_1-X_2 \sim N(0,2) \Rightarrow \quad \dfrac{1}{2}(X_1+X_2)^2 \sim \chi_1^2, \quad \dfrac{1}{2}(X_1-X_2)^2 \sim \chi_1^2$

$$\Rightarrow \frac{\frac{1}{2}(X_1+X_2)^2}{\frac{1}{2}(X_1-X_2)^2} = \frac{(X_1+X_2)^2}{(X_1-X_2)^2} \sim F_{1,1}$$

2. (**6 marks**) Let $X_1,\ldots,X_n$ be i.i.d. r.v.'s from the $U(0,\theta)$, $\theta \in \Omega = (0,\infty)$, distribution. Answer the following questions.

(a) Find the p.d.f. of $X_{(n)}$ and $E(X_{(n)})$, where $X_{(n)} = \max(X_1,\ldots,X_n)$;

(b) Find the p.d.f. of $X_{(1)}$ and $E(X_{(1)})$, where $X_{(1)} = \min(X_1,\ldots,X_n)$;

(c) Find two unbiased estimators for θ.

Solution:(a)

$$P(X_{(n)} \le t) = P(X_1 \le t, X_2 \le t, ..., X_n \le t) = \Pi_{i=1}^n P(X_i \le t) = \begin{cases} 0, & if \quad t \le 0 \\ (\frac{t}{\theta})^n, & if \quad 0 < t \le \theta \\ 1, & if \quad t > \theta \end{cases}$$

Therefore,

$$f_{X_{(n)}}(t) = \frac{d}{dt}P(X_{(n)} \le t) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & if \quad 0 < t \le \theta \\ 0, & otherwise \end{cases}$$

$$E(X_{(n)}) = \int_0^\theta t f_{X_{(n)}}(t)dt = \int_0^\theta \frac{nt^n}{\theta^n}dt = \frac{n}{n+1}\theta$$

(b)

$$P(X_{(1)} \le t) = 1-P(X_{(1)} \ge t) = 1-P(X_1 \ge t, X_2 \ge t, ..., X_n \ge t) = 1-\Pi_{i=1}^n P(X_i \ge t) = \begin{cases} 0, & if \quad t \le 0 \\ 1-(1-\frac{t}{\theta})^n, & if \quad 0 < t \le \theta \\ 1, & if \quad t > \theta \end{cases}$$

Therefore,

$$f_{X_{(1)}}(t) = \frac{d}{dt}P(X_{(1)} \le t) = \begin{cases} -n(1-\frac{t}{\theta})^{n-1}(-\frac{1}{\theta}) = \frac{n}{\theta}(1-\frac{t}{\theta})^{n-1}, & if \quad 0 < t \le \theta \\ 0, & otherwise \end{cases}$$

$$E(X_{(1)}) = \int_0^\theta t f_{X_{(1)}}(t)dt = \int_0^\theta \frac{n}{\theta}(1-\frac{t}{\theta})^{n-1}dt = \frac{1}{n+1}\theta$$

(c) From (a) and (b), we know $E(X_{(1)}+X_{(n)}) = E(X_{(1)}) + E(X_{(n)}) = \theta$, so one unbiased estimator for θ is $X_{(1)}+X_{(n)}$. Note that $E(X) = \frac{\theta}{2}$, and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is an unbiased estimator for $E(X)$, so $2\bar{X}$ is another unbiased estimator for θ