MTH211A: Theory of Statistics

Mid-semester Examination

Time: 120 minutes | Solution set | Total marks: 50

Name: ______________________ Roll number: ____________

1. *Answer all questions.*
2. *All notations used are as discussed in class.*

Q.1 Let $X_1, \cdots, X_n$ be a random sample from $\texttt{normal}(\mu_x, \sigma_x^2)$, $Y_1, \cdots, Y_m$ be a random sample from $\texttt{normal}(\mu_y, \sigma_y^2)$, $\{X_1, \cdots, X_n, Y_1, \cdots, Y_m\}$ are mutually independent, and $\sigma_x/\sigma_y = \alpha$, $\alpha > 0$ is known. Write the pdf of $T = S_x^2/S_y^2$. [6]

As $X_i \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$ we have $\dfrac{n\cancel{(n-1)}\, S_x^2}{\sigma_x^2} \sim \chi^2_{(n-1)}$

Similarly, $\dfrac{m S_y^2}{\sigma_y^2} \sim \chi^2_{(m-1)}$. Further, as $\underset{\sim}{X} \perp\!\!\!\perp \underset{\sim}{Y}$, we have $S_x^2 \perp\!\!\!\perp S_y^2$. Now, by defn.

$$W = \frac{\dfrac{n S_x^2}{\sigma_x^2} \Big/ (n-1)}{\dfrac{m S_y^2}{\sigma_y^2} \Big/ (m-1)} = \frac{n(m-1)}{m(n-1)} \cdot \frac{1}{\alpha^2} \frac{S_x^2}{S_y^2} \sim F_{n-1, m-1} \text{ distribution.}$$

So, $T = \dfrac{\cancel{\alpha} S_y^2}{S_x} = \dfrac{\alpha^2 m(n-1)}{n(m-1)} W$ has the following PDF :

$$f_T(t) = f_W\left(\frac{n(m-1)}{m(n-1)} \cdot \frac{1}{\alpha^2} t\right) \cdot |J| \quad \text{where} \quad J = \frac{\partial w}{\partial t} = \frac{\alpha^2 m(n-1)}{n(m-1)}$$

$$= \frac{\Gamma\left(\frac{m+n-2}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{n-1}{m-1}\right)^{(n-1)/2} \left\{\frac{n(m-1)}{m(n-1)\alpha^2} t\right\}^{\left(\frac{m-1}{2}\right)-1} \left(1 + \frac{m-1}{n-1} \cdot \frac{n(m-1)}{m(n-1)\alpha^2} t\right)^{-\left(\frac{m+n-2}{2}\right)} \cdot |J|$$

$$= \frac{\cancel{1}(\alpha^2)^{-\left(\frac{m-1}{2}\right)+1}}{\text{Beta}\left(\frac{m-1}{2}, \frac{n-1}{2}\right)} \cdot \frac{(n-1)^{\frac{m-n}{2}+2}}{(m-1)^{\frac{m-n}{2}+2}} \cdot \left(\frac{n}{m}\right)^{\left(\frac{m-1}{2}\right)-2} (t)^{\frac{m-1}{2}-1} \left\{1 + \frac{(m-1)^2 n}{(n-1)^2 m \alpha^2} t\right)^{-\left(\frac{m+n}{2}\right)+1},$$

$t > 0$.

x ———— x.

Q.2 Consider a trivariate continuous random variable $\mathbf{W} = (X, Y, Z)'$ with pdf

$$f_{\mathbf{W}}(\mathbf{w}) = f_{\mathbf{W}}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1} y^{\beta-1} z^{\gamma-1}, \text{ with } x, y, z \in (0,1),\ x+y+z=1,\ \alpha, \beta, \gamma > 0.$$

(a) Show that the marginal distributions of X, Y and Z are Beta. Find the parameters.

(b) Based on a random sample $\mathbf{w}_1, \ldots, \mathbf{w}_n$, find the expectation of the trace of the sample variance covariance matrix $T_n = n^{-1} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\top}$.

[5+4]

(a) Marginal distn.

$$f_X(x) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma\alpha\,\Gamma\beta\,\Gamma\gamma} x^{\alpha-1} \int_0^{1-x} y^{\beta-1} (1-x-y)^{\gamma-1}\, dy \quad \text{as } z = (1-x-y)$$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma\alpha\,\Gamma\beta\,\Gamma\gamma} x^{\alpha-1} (1-x)^{\gamma-1+\beta-1} \int_0^{1-x} \left(\frac{y}{1-x}\right)^{\beta-1} \left(1 - \frac{y}{1-x}\right)^{\gamma-1} dy.$$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma\alpha\,\Gamma\beta\,\Gamma\gamma} x^{\alpha-1} (1-x)^{\gamma+\beta-1} \int_0^{1} u^{\beta-1} (1-u)^{\gamma-1}\, du$$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma\alpha\,\cancel{\Gamma\beta}\,\cancel{\Gamma\gamma}} \cdot \frac{\cancel{\Gamma\beta}\,\cancel{\Gamma\gamma}}{\Gamma(\beta+\gamma)} x^{\alpha-1} (1-x)^{\gamma+\beta-1} \quad \Bigg|\ \text{where } u = \frac{y}{1-x} \Rightarrow du = \frac{dy}{(1-x)}$$

$0 < x < 1$

$X \sim \text{Beta}(\alpha, \beta+\gamma)$. Similarly, $Y \sim \text{Beta}(\beta, \alpha+\gamma)$ and $Z \sim \text{Beta}(\gamma, \alpha+\beta)$.

(b)
$$T_n = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} X_i^2 & \cancel{X_iY_i}\ X_iY_i & X_iZ_i \\ X_iY_i & Y_i^2 & Y_iZ_i \\ X_iZ_i & Y_iZ_i & Z_i^2 \end{pmatrix}$$

$$\therefore \operatorname{tr}\{\mathbb{E}(T_n)\} = \mathbb{E}[\operatorname{tr}(T_n)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right] + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right]$$

$$= \mathbb{E}(X_i^2) + \mathbb{E}(Y_i^2) + \mathbb{E}(Z_i^2) \quad \text{— (1)}$$

If $X \sim \text{Beta}(\alpha, \beta+\gamma)$

$$E(X^2) = \frac{1}{\text{Beta}(\alpha, \beta+\gamma)} \int_0^1 x^{\alpha+1} (1-x)^{\beta+\gamma-1}\, dx = \frac{\Gamma(\alpha+2)\,\Gamma(\beta+\gamma)\ \Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta+\gamma+2)\ \Gamma\alpha\ \Gamma(\beta+\gamma)}$$

$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta+\gamma)(\alpha+\beta+\gamma+1)}.$$

Similarly, (1) $= \dfrac{\alpha(\alpha+1) + \beta(\beta+1) + \gamma(\gamma+1)}{(\alpha+\beta+\gamma)(\alpha+\beta+\gamma+1)}$

x ———— Y

Q.3 Let $X_1, \cdots, X_n$ be a random sample from a family of distributions with pdf

$$f_X(x) = \frac{2x}{\theta} \exp\left\{-\frac{x^2}{\theta}\right\} \quad \text{with } x > 0, \text{ and } \theta > 0.$$

Find the conditional distribution of $X_{(r)}$ given $X_{(s)}$, where $r < s$. [5]

Joint distribution of $W = X_{(r)}$ and $T = X_{(s)}$, $r < s$ is

$$f_{W,T}(w,t) = \left(F_X(w)\right)^{r-1} \left\{F_X(t) - F_X(w)\right\}^{s-r-1} \quad ; w < t$$

$$\times f_X(w) f_X(t) \left\{1 - F_X(t)\right\}^{n-s} \times \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}$$

Marginal distributn. of $T = X_{(s)}$ is.

$$f_T(t) = \left(F_X(t)\right)^{s-1} f_X(t) \left\{1 - F_X(t)\right\}^{n-s} \frac{n!}{(s-1)!\,(n-s)!} \quad t > 0.$$

$\therefore$ The conditional distribution of W given T is:

$$f_{W|T=t}(w) = \frac{f_{W,T}(w,t)}{f_T(t)} \quad w < t.$$

$$= \frac{(s-1)!}{(s-r-1)!\,(r-1)!} \left\{F_X(t) - F_X(w)\right\}^{s-r-1} f_X(w) \left(1 - F_X(t)\right)^{s-1} \quad -① \quad w < t.$$

Now, $$F_X(t) = \frac{2}{\theta} \int_0^t u\, e^{-u^2/\theta}\, du \qquad \frac{u^2}{\theta} = z$$

$$= \int_0^{t^2/\theta} e^{-z}\, dz \qquad \frac{2u\,du}{\theta} = dz$$

$$= 1 - e^{-t^2/\theta}$$

From ① we have:

$$f_{W|T=t}(w) = \frac{(s-1)!\, e^{-\frac{t^2(s-1)}{\theta}}}{(r-1)!\,(s-r-1)!} \left[e^{-w^2/\theta} - e^{-t^2/\theta}\right]^{s-r-1} \frac{2w}{\theta} e^{-w^2/\theta}; \quad 0 < w < t.$$

Q.4 Based on a random samples $X_1, \cdots, X_n$, find a minimal sufficient statistic for the parameter $\theta > 0$ of the uniform$(-\theta, \theta)$ distribution. [4]

To ~~ob~~ obtain ~~the~~ a minimal sufficient statistic, we consider the ratio of joint pdf for two realizations $\underset{\sim}{x}$ and $\underset{\sim}{y}$:

$$\lambda_\theta(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\mathbb{I}(-\theta < x_{(1)} \leq x_{(n)} < \theta)}{\mathbb{I}(-\theta < y_{(1)} \leq y_{(n)} < \theta)} \quad -(*).$$

The condition $-\theta < x_{(1)} \leq x_{(n)} < \theta$

is equivalent to $-\theta < x_i < \theta$ for all $i = 1, \cdots, n$

$\iff |x_i| < \theta$ for all $i = 1, \cdots, n$

$\iff Z_{(n)} < \theta$ where $Z_i = |x_i|$

So, $(*)$ can be re-written as:

$$\lambda_\theta(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\mathbb{I}(Z_{(n)} < \theta)}{\mathbb{I}(W_{(n)} < \theta)} \quad \text{where } Z_i = |x_i| \text{ and } W_i = |y_i|.$$

The above fn. is const. w.r.t. θ iff $Z_{(n)} = W_{(n)}$,

i.e. $\max_{i \in \{1, \cdots, n\}} |x_i| = \max_{i \in \{1, \cdots, n\}} |y_i|$.

$\therefore T(\underset{\sim}{x}) = \max \{|x_1|, \cdots, |x_n|\}$ is minimal sufficient for θ.

X ——— X

Q.5 Let $X_1, \cdots, X_n$ be iid samples from some distribution with the j-th order raw moment as μ'_j, for $j = 1, 2, \ldots$. Derive the covariance between μ'_1 and μ'_2 in terms of the raw moments. [4]

$\mu_1' = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\mu_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

$$\begin{aligned}
\mathrm{Cov}(\mu_1', \mu_2') &= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} x_i , \frac{1}{n}\sum_{i=1}^{n} x_i^2\right) \\
&= \frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^{n} x_i , \sum_{i=1}^{n} x_i^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(x_i, x_j^2) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left\{\mathrm{Cov}(x_i, x_i^2) + \sum_{j\neq i}\mathrm{Cov}(x_i, x_j^2)\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Cov}(x_i, x_i^2) = \frac{1}{n}\mathrm{Cov}(x_1, x_1^2) \\
&= \frac{1}{n}\left\{E(x_1^3) - E(x_1)E(x_1^2)\right\} \\
&= \frac{1}{n}\left\{\mu_3' - \mu_1'\mu_2'\right\}.
\end{aligned}$$

$\times \longrightarrow \times$

Q.6 Let $X_1, \cdots, X_n$ be a random sample from a distribution with pdf $f_X(x) = \alpha\beta^\alpha x^{-(\alpha+1)}$ for $x > \beta$ and $\alpha, \beta > 0$. Prove or disprove the following: [$3 \times 3 = 9$]

(a) $\mathbf{T}_1 = (\sum_{i=1}^n \log X_i, \max_{i\in\{1,\ldots,n\}} X_i)'$ is sufficient for (α, β).

(b) Let $\alpha = 1$, then $T_2 = X_{(1)}/X_{(n)}$ is ancillary for β.

(c) Let $\beta = 1$, then $T_3 = \sum_{i=1}^n Y_i$ is minimal sufficient for α, where $Y_i = |\log(X_i)|$, $i = 1, \ldots, n$.

(a) To find ~~the~~ a minimal sufficient statistic, we take the ratio of joint PDFs for 2 diff realizations $\underset{\sim}{x}$ and $\underset{\sim}{y}$:

$$\frac{f_{\underset{\sim}{X}}(\underset{\sim}{x})}{f_{\underset{\sim}{X}}(\underset{\sim}{y})} = \frac{\left(\prod_{i=1}^n x_i\right)^{-(\alpha+1)} \mathbb{I}(x_{(1)} > \beta)}{\left(\prod_{i=1}^n y_i\right)^{-(\alpha+1)} \mathbb{I}(y_{(1)} > \beta)} = \lambda_{\underset{\sim}{\theta}}(\underset{\sim}{x}, \underset{\sim}{y}) \text{ where } \underset{\sim}{\theta} = \binom{\alpha}{\beta}.$$

Clearly, $\lambda_{\underset{\sim}{\theta}}(\underset{\sim}{x}, \underset{\sim}{y})$ is a const. fn. of $\underset{\sim}{\theta}$ if $\prod_{i=1}^n x_i = \prod_{i=1}^n y_i$ and $x_{(1)} = y_{(1)}$.

Thus, $\underset{\sim}{T}(\underset{\sim}{x}) = \left[\prod_{i=1}^n x_i, x_{(1)}\right]$ is jointly minimal sufficient.

$\underset{\sim}{T_1} \neq h(\underset{\sim}{T})$, ~~where~~ ~~$h((\underset{\sim}{x}))$ = [log x~~

as for 2 different realizations, for e.g. $\underset{\sim}{x} = \begin{pmatrix}1\\2\\4\\8\end{pmatrix}$ and $\underset{\sim}{y} = \begin{pmatrix}1.5\\2\\8/3\\8\end{pmatrix}$ we have $T_1(\underset{\sim}{x}) = T_1(\underset{\sim}{y})$ but $T(\underset{\sim}{x}) \neq T(\underset{\sim}{y})$.

So, T_1 is not sufficient.

(b) Let $\alpha = 1$, then $f_X(x) = \beta x^{-2}$ with $x > \beta$.

Let $U = \frac{X}{\beta}$ then $f_U(u) = f_X(\beta u) \cdot |J|$ with $J = \beta$

$= u^{-2}$ with $u > 1$: Does not depend on β.

So, $X \sim$ Scale family.

As $\frac{X_{(1)}}{X_{(n)}} = S(\underset{\sim}{x}) = S(c\underset{\sim}{x})$, $S(\underset{\sim}{x})$ is ancillary for β.

(c) Let $\beta = 1$, then $f_X(x) = \alpha x^{-(\alpha+1)}$ with $x > 1$.

In this case $\lambda_\alpha(\underset{\sim}{x}, \underset{\sim}{y}) = \left(\frac{\prod y_i}{\prod x_i}\right)^{\alpha+1}$, and will be free of α if $\prod x_i = \prod y_i$. So, $T(\underset{\sim}{x}) = \prod_{i=1}^n x_i$ is minimal sufficient.

Now, $T_3(\underset{\sim}{x}) = \sum_{i=1}^n |\log(x_i)| = \sum_{i=1}^n \log(x_i) = \log(\prod x_i) = \log(T(\underset{\sim}{x}))$ is also minimal sufficient as log is a one-one function.

Q.7 Suppose $X_1, \cdots, X_n$ are n iid observations from the normal (θ, θ^2) distribution, $\theta \in \mathbb{R}$. Define $Y_i = \log|X_i|$, for $i = 1, \ldots, n$. Prove or disprove: $Y_{(n)} - Y_{(1)}$ is ancillary for θ. [4]

Let $X \sim N(\theta, \theta^2)$ then $\frac{X}{\theta} = U \sim N(1,1)$ distribution which is free of θ. Thus, $X \sim$ Scale family of distn.s.

Claim: $T(\underline{x}) = Y_{(n)} - Y_{(1)}$ satisfies. $T(\underline{x}) = T(c\underline{x})$. for any $c>0$.

Let $Z_i = c X_i$ then $|Z_i| = c|X_i|$ as $c>0$.

Then $Y_i^c = \log c + \log|X_i|$ and $Y_{(n)}^c = \log c + \max \log|X_i|$.

$$\therefore\ T(c\underset{\sim}{X}) = Y_{(n)}^c - Y_{(1)}^c = \max \log|X_i| - \overset{\min}{\cancel{\max}} \log|X_i|.$$

$$= Y_{(n)} - Y_{(1)}. \quad \text{implying } T(c\underline{x}) = T(\underline{x}).$$

$\therefore$ T is ancillary for θ.

$X \longrightarrow Y$

Q.8 Let $X_1, \cdots, X_n$ be a random sample from normal$(0, \sigma^2)$ distribution. Prove or disprove the following: [3 × 3 = 9]

(a) $T_1(\mathbf{X}) = \sum_{i=1}^n X_i$ is complete.

(b) $T_2(\mathbf{X}) = \sum_{i=1}^n X_i^2$ is complete.

(c) $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))'$ is complete.

(a) $\mathbb{E}\left(T_1(\underset{\sim}{x})\right) = 0 \quad \forall \ \sigma^2 > 0$ but $P\left(T_1(\underset{\sim}{x}) = 0\right) = 0$

$\therefore \ T_1(\underset{\sim}{x})$ is not complete.

(b) $\sum_{i=1}^n X_i^2 = T_2(\underset{\sim}{x}) \sim \sigma^2 \chi^2_{(n)}$ distribution, as $\frac{X_i}{\sigma} \overset{IID}{\sim} N(0,1)$ $i=1, \cdots, n$.

$f_{T_2}(t) = f_U\left(t/\sigma^2\right) \cdot |J|$ where

$J = \frac{\partial u}{\partial T_2} = \frac{1}{\sigma^2}$; $t > 0$; $U = \sum_{i=1}^n \frac{X_i^2}{\sigma^2} = \frac{T_2}{\sigma^2}$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \left(\frac{t}{\sigma^2}\right)^{\frac{n}{2}-1} e^{-t/2\sigma^2} \cdot \frac{1}{\sigma^2} \qquad t > 0.$$

$$= \frac{(\sigma^2)^{-n/2}}{2^{n/2}\Gamma\frac{n}{2}} \exp\left\{-\frac{t}{2\sigma^2} + \left(\frac{n}{2}-1\right)\log t\right\}, \quad t > 0$$

which belongs to the exponential family, with $w_1(\sigma^2) = -\frac{1}{2\sigma^2}$ and ~~$T_1(x)$~~ $t_1 = t$.

Thus, T_2 is complete sufficient.

(c) $T(\underset{\sim}{x})$ is not complete, as $\mathbb{E}_{\sigma^2}\left[g\left(T(\underset{\sim}{x})\right)\right] = \mathbb{E}_{\sigma^2}\left[T_1(\underset{\sim}{x})\right] = 0$

but $P_{\sigma^2}\left(T_1(\underset{\sim}{x}) = 0\right) = 0$. Here $g\left(\binom{x}{y}\right) = x$.

$\times$ ——————→ $\times$