COLUMBIA UNIVERSITY

DEPARTMENT OF BIOSTATISTICS

P8109 – STATISTICAL INFERENCE

***exercise sheet* 1 (model answers)**

Question 1 (1 MARK)

We have $\sum_{i=1}^{6} Z_i^2 \sim \chi_6^2$ so that, using the CDF command in SPSS,

$$\Pr\left\{\sum_{i=1}^{6} Z_i^2 \le 6\right\} = .5768 .$$

Question 2 (1 MARK)

By definition of the *F*-distribution, $X = W_1 / W_2$, where $W_1 \sim \chi^2_{\nu_1}$, $W_2 \sim \chi^2_{\nu_2}$, and W_1 and W_2 are independent r.v.s. Now, $1/X = W_2 / W_1$. By definition, therefore $1/X \sim F_{\nu_2,\nu_1}$.

Question 3 (1+2+2=5 MARKS)

(a) $T \sim t_\nu$

(b) (i) We have, since *Z* and *W* are independent,

$$\begin{aligned} \mathcal{E}T &= \mathcal{E}\left(Z.\frac{1}{\sqrt{W/\nu}} \right) \\ &= \left(\mathcal{E}Z\right)\left(\mathcal{E}\frac{1}{\sqrt{W/\nu}} \right) \end{aligned}$$

Now, $\mathcal{E}Z = 0$ and $\mathcal{E}W^{-1/2} = \dfrac{\Gamma\left(\frac{\nu}{2}-\frac{1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} 2^{-1/2}$ so that $\mathcal{E}T = 0$.

(ii) Also,

$$\begin{aligned}\operatorname{var} T &= \mathcal{E}T^2 - \mathcal{E}^2 T = \mathcal{E}T^2 \\ &= \mathcal{E}\frac{Z^2}{W/\nu} \\ &= \mathcal{E}Z^2 \mathcal{E}\left(\frac{1}{W/\nu}\right),\end{aligned}$$

where $\mathcal{E}Z^2 = \operatorname{var} Z = 1$ and $\mathcal{E}W^{-1} = \dfrac{\Gamma\left(\frac{\nu}{2}-1\right)}{\Gamma\left(\frac{\nu}{2}\right)} 2^{-1} = \dfrac{1}{\left(\frac{\nu}{2}-1\right)} 2^{-1} = \dfrac{1}{\nu-2}$. Hence

$$\operatorname{var} T = 1\left(\frac{\nu}{\nu-2}\right) = \frac{\nu}{\nu-2}.$$

Question 4 (1+1+1+1=4 MARKS)

(a) We have

$$\begin{aligned}\mathcal{E}e^{tX^*} &= \mathcal{E}e^{t(X-\lambda)/\sqrt{\lambda}} \\ &= \mathcal{E}e^{t(X-\lambda)/\sqrt{\lambda}} \\ &= \mathcal{E}\left(e^{-t\sqrt{\lambda}} e^{tX/\sqrt{\lambda}}\right) \\ &= e^{-t\sqrt{\lambda}} \mathcal{E}e^{tX/\sqrt{\lambda}}.\end{aligned}$$

(b) We have

$$\begin{aligned}\mathcal{E}e^{tX^*} &= e^{-t\sqrt{\lambda}} \mathcal{E}e^{tX/\sqrt{\lambda}} \\ &= e^{-t\sqrt{\lambda}} \sum_{x=0}^{\infty} e^{tx/\sqrt{\lambda}} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-t\sqrt{\lambda}-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{t/\sqrt{\lambda}}\right)^x}{x!}.\end{aligned}$$

(c) Since $\sum_{x=0}^{\infty} a^x / x! \equiv e^a$, we have

$$
\begin{aligned}
\mathcal{E}e^{tX^*} &= e^{-t\sqrt{\lambda}-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{t/\sqrt{\lambda}}\right)^x}{x!} \\
&= e^{-t\sqrt{\lambda}-\lambda} \times e^{\lambda e^{t/\sqrt{\lambda}}} \\
&= \exp\left\{-t\sqrt{\lambda} - \lambda + \lambda e^{t/\sqrt{\lambda}}\right\}
\end{aligned}
$$

(d) Using $\sum_{x=0}^{\infty} a^x / x! \equiv e^a$ again,

$$
\begin{aligned}
\mathcal{E}e^{tX^*} &= \exp\left\{-t\sqrt{\lambda} - \lambda + \lambda e^{t/\sqrt{\lambda}}\right\} \\
&= \exp\left[-t\sqrt{\lambda} - \lambda + \lambda\left\{1 + \frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + o\left(\frac{1}{\lambda}\right)\right\}\right] \\
&= \exp\left(-t\sqrt{\lambda} - \lambda + \lambda + t\sqrt{\lambda} + \frac{t^2}{2} + o(1)\right) \\
&\to e^{t^2/2} \quad \text{as} \quad \lambda \to \infty
\end{aligned}
$$

Question 5 (2 MARKS)

Using the hint, we have

$$
\begin{aligned}
\mathcal{E}e^{t\frac{X-np}{\sqrt{npq}}} &= \mathcal{E}e^{t\frac{(Z_1+\ldots+Z_n)-np}{\sqrt{npq}}} \\
&= \mathcal{E}e^{t\frac{(Z_1-p)+\ldots+(Z_n-p)}{\sqrt{npq}}} \\
&= \mathcal{E}e^{t\frac{(Z_1-p)}{\sqrt{npq}}} \times \ldots \times \mathcal{E}e^{t\frac{(Z_n-p)}{\sqrt{npq}}} \quad \text{(by independence of the } Z_i \text{'s)} \\
&= \left\{\mathcal{E}e^{t\frac{(Z_i-p)}{\sqrt{npq}}}\right\}^n \\
&= \left[\mathcal{E}\left\{1 + \frac{t(Z_i - p)}{\sqrt{npq}} + \frac{t^2(Z_i - p)^2}{2npq} + o\left(\frac{1}{n}\right)\right\}\right]^n \\
&= \left\{\mathcal{E}1 + \frac{t}{\sqrt{npq}}\mathcal{E}(Z_i - p) + \frac{t^2}{2npq}\operatorname{var} B_i + o\left(\frac{1}{n}\right)\right\}^n \\
&= \left\{1 + 0 + \frac{t^2}{2npq}(pq) + o\left(\frac{1}{n}\right)\right\}^n \\
&= \left\{1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right\}^n
\end{aligned}
$$

As $n \to \infty$, $\mathcal{E} e^{t \frac{X-np}{\sqrt{npq}}} \to e^{t^2/2}$, which is the m.g.f. of the N(0, 1) r.v.

Question 6 (1+1+1+1=4 MARKS)

(a) $\bar{Y}$ is normal with mean 0 and variance $1/5$;

(b) U has a chi-squared distribution with 5 d.f.;

(c) $V = \sum_{i=1}^{5}(Y_i - \bar{Y})^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi_4^2$;

(d) $\sum_{i=1}^{5}(Y_i - \bar{Y})^2 \sim \chi_4^2$ and Y_6^2 is an independent χ_1^2 variable. Therefore, $\sum_{i=1}^{5}(Y_i - \bar{Y})^2 + Y_6^2 \sim \chi_5^2$

Question 7 (1+1+1=3 MARKS)

(a) We have

$$T = \frac{\bar{Y} - 0}{S/\sqrt{n}} = \frac{\bar{Y}}{S/\sqrt{10}} \sim t_9$$

Therefore,

$$\frac{10\bar{Y}^2}{S^2} = T^2 \sim F_{1,9}$$

(b) We have

$$\frac{S^2}{10\bar{Y}^2} \sim F_{9,1}$$

(c) From F-tables, if $f \sim F_{9,1}$ then

$$\Pr\{f \le 240.54\} = .95 .$$

Therefore,

$$\Pr\left\{\frac{S^2}{10\bar{Y}^2} \le 240.54\right\} = .95$$

$$\Pr\left\{\frac{S^2}{\bar{Y}^2} \le 2405.4\right\} = .95$$

$$\Pr\left\{-\sqrt{2405.5} \le \frac{S}{\bar{Y}} \le \sqrt{2405.4}\right\} = .95$$

Hence, $c = \sqrt{2405.4} = 49.04$

Question 8 (1+1+1=3 MARKS)

(a) We have

$$W = \sum_{i=1}^{n-1} Z_i^2 \, .$$

(b) (i) We have

$$\frac{(n-1)S^2}{\sigma^2} \overset{D}{=} W$$

(ii) Since $W = \sum_{i=1}^{n-1} Z_i^2$, by the Central Limit theorem,

$$W \mathrel{\dot{\underset{\cdot}{\sim}}} N\left[(n-1)\mathcal{E}Z_i^2, (n-1)\operatorname{var} Z_i^2\right].$$

Now, since $Z_i^2 \sim \chi_1^2$,

$$\mathcal{E}Z_i^2 = 1 \quad \text{and} \quad \operatorname{var} Z_i^2 = 2$$

Therefore,

$$W \mathrel{\dot{\underset{\cdot}{\sim}}} N\left[(n-1), 2(n-1)\right] \;\; \Rightarrow \;\; S^2 = \frac{\sigma^2 W}{n-1} \mathrel{\dot{\underset{\cdot}{\sim}}} N\left[\sigma^2, \frac{\sigma^4}{(n-1)^2} .2(n-1)\right].$$

Hence,

$$\frac{S^2 - \sigma^2}{\sigma^2\sqrt{2/(n-1)}} \xrightarrow{D} Z_i \sim N(0, 1)$$

Question 9 (1+2=3 MARKS)

(a) We have

$$
\begin{aligned}
p(x;p) &= \binom{n}{x} p^x (1-p)^{n-x} I\left(x \in \{0, 1, ..., n\}\right) \\
&= I\left(x \in \{0, 1, ..., n\}\right) \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n \\
&= I\left(x \in \{0, 1, ..., n\}\right) \binom{n}{x} (1-p)^n \exp\left\{\log\left(\frac{p}{1-p}\right)^x\right\} \\
&= \underbrace{I\left(x \in \{0, 1, ..., n\}\right) \binom{n}{x}}_{b(x)} \underbrace{(1-p)^n}_{a(\theta)} \exp\left\{\underbrace{x}_{d(x)} \underbrace{\log\left(\frac{p}{1-p}\right)}_{c(\theta)}\right\}
\end{aligned}
$$

Hence, the binomial (*n*, *p*) belongs to the one-parameter exponential family.

(a) We have

$$
\begin{aligned}
f(x;\alpha,\beta) &= \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} I(x > 0) \\
&= \frac{I(x>0)}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \\
&= \frac{I(x>0)}{\beta^\alpha \Gamma(\alpha)} \exp\left(\log x^{\alpha-1}\right) e^{-x/\beta} \\
&= \underbrace{I(x>0)}_{b(x)} \underbrace{\frac{1}{\beta^\alpha \Gamma(\alpha)}}_{a(\alpha,\beta)} \exp\left\{\underbrace{(\alpha-1)}_{c_1(\alpha,\beta)} \underbrace{\log x}_{d_1(x)} \underbrace{-\frac{1}{\beta}}_{c_2(\alpha,b)} \underbrace{x}_{d_2(x)}\right\}
\end{aligned}
$$

Hence, the gamma(α,β) distribution belongs to the two-parameter exponential family.

Question 10 (2+2 = 4 MARKS)

(a) We have

$$h(X) = h(\mu) + (X-\mu)h^{\prime}(\mu) + \frac{(X-\mu)^2}{2}h^{\prime\prime}(\mu) + \ldots$$

$$\therefore \quad \mathcal{E}h(X) \approx \mathcal{E}h(\mu) + h^{\prime}(\mu)\mathcal{E}h(X-\mu) + \frac{h^{\prime\prime}(\mu)}{2}\mathcal{E}(X-\mu)^2 .$$

$$= h(\mu) + \frac{h^{\prime\prime}(\mu)}{2}\operatorname{var} X .$$

(b) From (a), we write $X = P_s$, $\mu = P$, and $h(P) = P/(1-P) = -1 + (1-P)^{-1}$, so that $h^{\prime}(P) = (1-P)^{-2}$. Now,

$$\hat{O} = h(P_s) \approx h(P) + (P_s - P)h^{\prime}(P)$$

$$\therefore \quad \mathcal{E}h(P_s) \approx h(P) = \frac{P}{1-P}$$

Hence, the estimated mean of $\widehat{O}$ is $\frac{P_s}{1-P_s} = \frac{.368}{1-.368} = .58$.

Moreover,

$$\operatorname{var} h(P_s) \approx \{h^{\prime}(P)\}^2 \operatorname{var} P_s$$

$$= \frac{1}{(1-P)^4} \cdot \frac{P(1-P)}{n}$$

$$= \frac{P}{n(1-P)^3}.$$

Hence, the estimated variance of $\widehat{O}$ is $\frac{P_s}{n(1-P_s)^3} = \frac{.368}{378(1-.368)^3} = 3.86\times 10^{-3}$.