Tutorial 5 Solutions

1. Proof: from the definition of sample mean and sample variance, we have

$$\bar{x}_1 = \frac{1}{m}\sum_{i=1}^{m} x_i, \bar{x}_2 = \frac{1}{n}\sum_{i=m+1}^{m+n} x_i,$$

$$s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_1-\bar{x}_1)^2, s_2^2 = \frac{1}{n-1}\sum_{i=m+1}^{m+n}(x_1-\bar{x}_2)^2.$$

Thus $\frac{m\bar{x}_1+n\bar{x}_2}{m+n}$

$$\begin{aligned}
&=\frac{m\frac{1}{m}\sum_{i=1}^{m} x_i + n\frac{1}{n}\sum_{i=m+1}^{m+n} x_i}{m+n}\\
&=\frac{1}{m+n}\sum_{i=1}^{m+n} x_i\\
&=\bar{x}
\end{aligned}$$

And $\frac{(m-1)s_1^2+(n-1)s_2^2}{m+n-1} + \frac{mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)}$

$$\begin{aligned}
&=\frac{(m-1)\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x}_1)^2 + (n-1)\frac{1}{n-1}\sum_{i=m+1}^{m+n}(x_i-\bar{x}_2)^2}{m+n-1} + \frac{mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)}\\
&=\frac{(m+n)(\sum_{i=1}^{m}(x_i-\bar{x}_1)^2 + \sum_{i=m+1}^{m+n}(x_i-\bar{x}_2)^2) + mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)}\\
&=\frac{(m+n)(\sum_{i=1}^{m}(x_i^2-2x_i\bar{x}_1+\bar{x}_1^2) + \sum_{i=m+1}^{m+n}(x_i^2-2x_i\bar{x}_2+\bar{x}_2^2)) + mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)}\\
&=\frac{(m+n)(\sum_{i=1}^{m} x_i^2 - 2\bar{x}_1\sum_{i=1}^{m} x_i + m\bar{x}_1^2 + \sum_{i=m+1}^{m+n} x_i^2 - 2\bar{x}_2\sum_{i=m+1}^{m+n} x_i + n\bar{x}_2^2) + mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)}\\
&=\frac{(m+n)(\sum_{i=1}^{m+n} x_i^2 - m\bar{x}_1^2 - n\bar{x}_2^2) + mn(\bar{x}_1-\bar{x}_2)^2}{(m+n)(m+n-1)} = \frac{(m+n)\sum_{i=1}^{m+n} x_i^2 - (m\bar{x}_1+n\bar{x}_2)^2}{(m+n)(m+n-1)}\\
&=\frac{(m+n)\sum_{i=1}^{m+n} x_i^2 - (m+n)^2\bar{x}^2}{(m+n)(m+n-1)} = \frac{\sum_{i=1}^{m+n}(x_i^2-\bar{x}^2)}{m+n-1}\\
&=\frac{\sum_{i=1}^{m+n}(x_i^2-2\bar{x}\sum_{i=1}^{m+n} x_i+\bar{x}^2)}{m+n-1} = \frac{\sum_{i=1}^{m+n}(x_i-\bar{x})^2}{m+n-1}\\
&=s^2
\end{aligned}$$

2. Proof: We have

$$
\begin{aligned}
Cov(\bar{X}, S^2) =& Cov(\bar{X} - \mu, S^2) \\
=& Cov(\bar{X} - \mu, \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2) \\
=& Cov(\bar{X} - \mu, \frac{1}{n-1} \sum_{i=1}^{n} ((X_i - \mu) - (\bar{X} - \mu))^2) \\
=& Cov(\bar{X} - \mu, \frac{1}{n-1} \sum_{i=1}^{n} ((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2)) \\
=& Cov(\bar{X} - \mu, \frac{1}{n-1} \sum_{i=1}^{n} ((X_i - \mu)^2 - 2(\bar{X} - \mu)^2 + (\bar{X} - \mu)^2)) \\
=& Cov(\bar{X} - \mu, \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2) \\
=& \frac{1}{n-1} (\sum_{i=1}^{n} Cov(\bar{X} - \mu, (X_i - \mu)^2) - nCov(\bar{X} - \mu, (\bar{X} - \mu)^2)),
\end{aligned}
$$

and $E(\bar{X} - \mu) = E(X_i - \mu) = 0, E(X_i - \mu)^2 = \sigma^2, E(X_i - \mu)^3 = v_3$, $X_i - \mu$ and $X_j - \mu$ are independent if $i \neq j$, thus,

$$
\begin{aligned}
\sum_{i=1}^{n} Cov(\bar{X} - \mu, (X_i - \mu)^2) =& \sum_{i=1}^{n} Cov(\frac{1}{n} \sum_{k=1}^{n} (X_k - \mu), (X_i - \mu)^2) \\
=& \frac{1}{n} Cov(X_i - \mu, (X_i - \mu)^2) \\
=& \frac{1}{n} \sum_{i=1}^{n} (E(X_i - \mu)^3 - E(X_i - \mu) E(X_i - \mu)^2) \\
=& \frac{1}{n} \cdot n v_3 = v_3,
\end{aligned}
$$

$$
\begin{aligned}
Cov(\bar{X}-\mu,(\bar{X}-\mu)^2) =& E(\bar{X}-\mu)^3 - E(\bar{X}-\mu)E(\bar{X}-\mu)^2 \\
=& E(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu))^3 \\
=& \frac{1}{n^3}E(\sum_{i=1}^{n}(X_i-\mu)^3) \\
=& \frac{1}{n^3}\sum_{i=1}^{n}E(X_i-\mu)^3 \\
=& \frac{1}{n^3}\cdot n v_3 = \frac{1}{n^2}v_3.
\end{aligned}
$$

Therefore, we have

$$
Cov(\bar{X},S^2) = \frac{1}{n-1}(v_3 - n\cdot\frac{1}{n^2}v_3) = \frac{v_3}{n}.
$$

3. Proof: Since $f(x)=\frac{mx^{m-1}}{\eta^m}\exp\{-(\frac{x}{\eta})^m\}, x>0, m>0, \eta>0$, thus,

$$
F(x) = \int_0^x \frac{mt^{m-1}}{\eta^m}\exp\{-(\frac{t}{\eta})^m\}dt = -\exp\{-(\frac{x}{\eta})^m\},
$$

we have

$$
f_{x_{(1)}}(x) = n(1-F(x))^{n-1}f(x) = n(\exp\{-(\frac{x}{\eta})^m\})^{n-1}\frac{mx^{m-1}}{\eta^m}\exp\{-(\frac{x}{\eta})^m\} = \frac{mx^{m-1}}{(\eta/n^{\frac{1}{m}})^m}\exp\{(-\frac{x}{\eta/n^{\frac{1}{m}}})^m\},
$$

where is a Weibull distribution and parameters are $(m,\eta/n^{\frac{1}{m}})$.

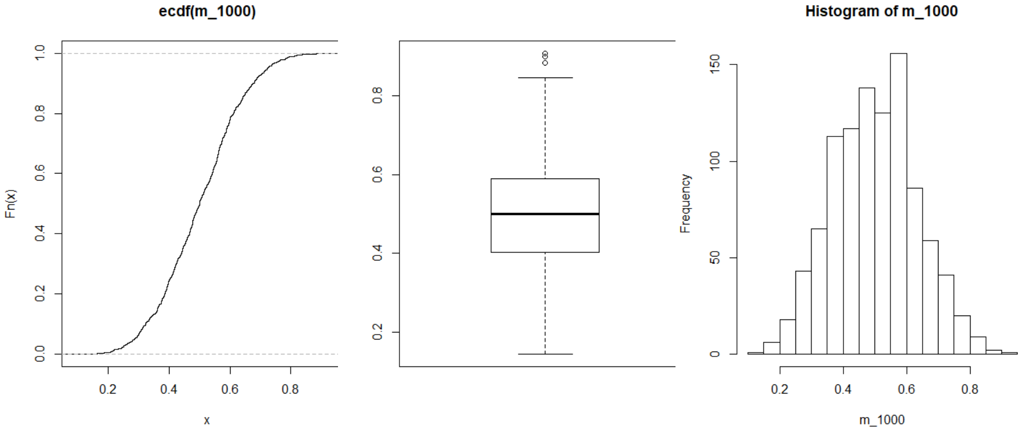
4.

(a) Since $E(\bar{X})=\frac{1}{2}, Var(\bar{X})=\frac{1}{10\times12}$, therefore the asymptotic distribution of $\bar{X}$ from a uniform distribution $U(0,1)$ with sample size 10 is $N(\frac{1}{2},\frac{1}{120})$.
(b) The asymptotic distribution of $m_{0.5}$ is $N(x_{0.5},\frac{1}{4n\cdot p^2(x_{0.5})})$, therefore the asymptotic distribution of $m_{0.5}$ from a uniform distribution $U(0,1)$ with sample size 10 is $N(\frac{1}{2},\frac{1}{40})$.
(c) The R code is

```r
set.seed(1001)
m_1000 <- NULL
for (x in 1:1000){m_1000 <- c(m_1000, median(runif(10, 0, 1)))}
boxplot(m_1000)
hist(m_1000)
plot(ecdf(m_1000))
```

and the plots of ecdf, boxplot and histogram are shown in the figure below.

Figure 1: Plots of ecdf, boxplot and histogram in 4(c).



(a) ecdf. (b) boxplot. (c) histogram.

5. We have

$$y = \sum_{i=1}^{n}(x_i + x_{n+i} - 2\bar{x})^2 = \sum_{i=1}^{n}((x_i - \bar{x})^2 + (x_{n+i} - \bar{x})^2 - 2(x_i - \bar{x})(x_{n+i} - \bar{x})),$$

thus

$$\begin{aligned}
E(y) =& E(\sum_{i=1}^{n}(x_i - \bar{x})^2) + E(\sum_{i=1}^{n}(x_{n+i} - \bar{x})^2) - E(\sum_{i=1}^{n}2(x_i - \bar{x})(x_{n+i} - \bar{x})) \\
=& E(\sum_{i=1}^{2n}(x_i - \bar{x})^2) - 2E(\sum_{i=1}^{n}(x_i x_{n+i} - x_i\bar{x} - x_{n+i}\bar{x} + \bar{x}^2)) \\
=& (2n-1)\sigma^2 - 2(n\mu^2 - n\mu^2 - n\mu^2 + n(\mu^2 + \frac{1}{2n}\sigma^2)) \\
=& (2n-2)\sigma^2.
\end{aligned}$$

6.

(a) The pmf of $Possion(\theta)$ can be expressed as

$$\begin{aligned} P(x;\theta) =& \frac{\theta^x}{x!} e^{-\theta} (x = 0, 1, 2, \cdots) \\ =& e^{\log \frac{\theta^x}{x!} - \theta} \\ =& \frac{1}{x!} e^{\log \theta x - \theta}, \end{aligned}$$

which is a member of exponential family, and the joint pmf of $Possion(\theta)$ is

$$P(x_1, x_2, \cdots, x_n; \theta) = \frac{1}{x_1! x_2! \cdots x_n!} e^{\log \theta \sum_{i=1}^n x_i - n\theta}$$

thus $\sum_{i=1}^n x_i$ is a sufficient statistic for θ.

(b) The pdf of $N(\theta, 1)$ can be expressed as

$$\begin{aligned} f(x;\theta) =& \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \\ =& \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{\mu x - \frac{\mu^2}{2}} \end{aligned}$$

which is a member of exponential family, and the joint pdf of $N(\theta, 1)$ is

$$f(x_1, x_2, \cdots, x_n; \theta) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{\sum_{i=1}^n x_i^2}{2}} e^{\mu \sum_{i=1}^n x_i - \frac{n\mu^2}{2}}$$

thus $\sum_{i=1}^n x_i$ is a sufficient statistic for θ.