STAT3005 Assignment 1
Lam Hoi Chun (1155192755)

Exercise 1.1

1(a).

$\because \text{Var}(X_i) = E(X_i^2) - E(X_i)^2$ and $E(X_i)^2 = \mu_i^2 = 0$

$\therefore \text{Var}(X_i) = \sigma^2 = E(X_i^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$\therefore \hat{\sigma}_1^2$ is a sensible estimator.

i.e. It is sensible to use the sample variance of a random variable X with mean $= 0$ to estimate the population variance.

Besides, it is a non-parametric estimator. Although it assumes n number of parameters, the distribution of X cannot be defined by a finite number of parameters. However, it also can be treated as a parametric estimator since it assumes a parametric form of μ_i, which is equal to 0. So, the conclusion depends on how we really define a non-parametric estimator.

1(b).

Using Theorem 2.5, we know that $\frac{X_{(i)}-q}{\sigma_p/\sqrt{n}} \overset{d}{\sim} \text{N(0,1)}$.

By changing the subject, $X_{(i)} - q \approx \text{N}\big(0, \frac{\sigma_p^2}{n}\big) \approx 0$ when $n \to \infty$

i.e. $X_{(i)} - q \overset{\text{pr}}{\to} 0 \Rightarrow X_{(i)} \overset{\text{pr}}{\to} q \Rightarrow X_{(i)} \overset{\text{d}}{\to} q$, where $q \sim \text{N}(0, \sigma^2)$

Now consider $p = P(X \leq q) = P(Z \leq \frac{q-\mu}{\sigma}) = \Phi(\frac{q-\mu}{\sigma})$, where $\mu = \mu_1 = \ldots = \mu_n$.

$\Rightarrow \Phi^{-1}(p) = \frac{q-\mu}{\sigma}$

Then, we can rewrite the variable as: $q = \mu + \sigma\Phi^{-1}(p)$

Note that $X_{(i)} = X_{(\lfloor np \rfloor)} \overset{\text{pr}}{\to} q = \mu + \sigma\Phi^{-1}(p)$

$\therefore X_{(\lfloor np \rfloor)} \overset{\text{pr}}{\to} \mu + \sigma\Phi^{-1}(p)$

$$\begin{aligned}
\therefore \hat{\sigma}_2^2 &= \big[\tfrac{X_{(\lfloor 0.75n \rfloor)} - X_{(\lfloor 0.25n \rfloor)}}{2\Phi^{-1}(0.75)}\big]^2 \\
&= \big[\tfrac{(\mu+\sigma\Phi^{-1}(0.75))-(\mu+\sigma\Phi^{-1}(0.25))}{2\Phi^{-1}(0.75)}\big]^2 \\
&= \big[\tfrac{\sigma(\Phi^{-1}(0.75)-\Phi^{-1}(0.25))}{2\Phi^{-1}(0.75)}\big]^2 \\
&= \big[\tfrac{\sigma(\Phi^{-1}(0.75)-(-\Phi^{-1}(0.75)))}{2\Phi^{-1}(0.75)}\big]^2 \\
&= \big[\tfrac{2\sigma\cdot\Phi^{-1}(0.75)}{2\Phi^{-1}(0.75)}\big]^2 \\
&= \sigma^2
\end{aligned}$$

1(b).(Cont'd)

$\therefore \hat{\sigma}_2^2$ is a sensible estimator.

i.e. It is sensible to use the sample IQR to estimate the population variance.

Besides, it is a parametric estimator as it assumes a model that can be described by 2 parameters: a normal distribution when mean μ and variance σ^2.

1(c).

Note that $\mu_i - \mu_{i-1} = g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)$

By the definition of derivative: $g'\left(\frac{i}{n}\right) = \frac{g(\frac{i}{n})-g(\frac{i}{n}-\frac{1}{n})}{1/n}$ when $n \to \infty$

$\therefore \mu_i - \mu_{i-1} \approx 0$ when $n \to \infty$

$\therefore X_i - X_{i-1} = (\mu_i - \mu_{i-1}) + (\epsilon_i - \epsilon_{i-1}) \approx (\epsilon_i - \epsilon_{i-1})$ when $n \to \infty$

$$
\begin{aligned}
&\quad E(X_i - X_{i-1})^2 \\
&= \mathrm{Var}(X_i - X_{i-1}) + E(X_i - X_{i-1})^2 \\
&= \mathrm{Var}(X_i - X_{i-1}) + [E(X_i) - E(X_{i-1})]^2 \\
&= \mathrm{Var}(X_i - X_{i-1}) + (\mu_i - \mu_{i-1})^2 \\
&= \mathrm{Var}(X_i - X_{i-1}) + 0^2 \\
&= \mathrm{Var}(X_i) + \mathrm{Var}(X_{i-1}) - \mathrm{Cov}(X_i, X_{i-1}) \\
&= \sigma^2 + \sigma^2 - 0 \\
&= 2\sigma^2
\end{aligned}
$$

$\therefore \sigma^2$ can be estimated by $\frac{E(X_i-X_{i-1})^2}{2} = \frac{1}{2}\left[\sum_{i=2}^{n} \frac{(X_i-X_{i-1})^2-0}{n-1}\right] = \frac{1}{2(n-1)}\sum_{i=2}^{n}(X_i - X_{i-1})^2$

$\therefore \hat{\sigma}_3^2$ is a sensible estimator.

Besides, it is a non-parametric estimator as it assumes a model that cannot be defined by a finite number of parameters, which is a differentiable function g.

Exercise 1.1

2(a)(i). $\sigma^2 = 5^2 = 25$

2(a)(ii). $\sigma^2 = \frac{[1-(-1)]^2}{12} = \frac{1}{3}$

2(a)(iii). Note that the variance of t-distribution $= \frac{v}{v-2}$ for $v > 2$.

$\therefore$ Variance of $t_4 = \frac{4}{4-2} = 2$

2(b).

	RMSE($\log \widehat{\sigma}_1^2$)	RMSE ($\log \widehat{\sigma}_2^2$)	RMSE ($\log \widehat{\sigma}_3^2$)
DGP (i)	1.5979723	0.4256686	0.3308837
DGP (ii)	0.1594576	0.5433290	0.2571831
DGP (iii)	4.1580413	3.6005726	0.6455057

2(c). $\hat{\sigma}_1^2$ only performs well for DGP (ii) but poorly in other two DGPs. The reason may be that the assumption of $\mu_i = 0$ is only reasonable for DGP (ii).

$\hat{\sigma}_2^2$ is relatively stable but less accurate, particularly for DGP (iii). The reason may be that the noise of DGP (iii) $\not\sim N$.

$\hat{\sigma}_3^2$ is generally a better estimator as it consistently shows lower RMSE across all DGPs. The reason may be that it has the weakest assumption. So, it will be a fit for more distribution, resulting that it is the most reliable estimator across different DGPs.

Exercise 1.2

1. $E(T)$

$= E\big[\sum_{i=2}^{n} \frac{i}{n+1}(R_i - R_{i-1})\big]$

$= E\big(\sum_{i=2}^{n} \frac{i}{n+1}\big) \cdot E(R_i - R_{i-1})$

$\because E(R_i) = \frac{n+1}{2}$, for all $i = 2, \ldots, n$

$\therefore E(R_i - R_{i-1}) = 0$, for all $i = 2, \ldots, n$

$\therefore E(T) = E\big(\sum_{i=2}^{n} \frac{i}{n+1} \cdot 0\big) = 0$

$$
\begin{aligned}
T &= \textstyle\sum_{i=2}^{n} \frac{i}{n+1}(R_i - R_{i-1}) \\
&= \tfrac{1}{n+1}\big[\textstyle\sum_{i=2}^{n} iR_i - \sum_{i=1}^{n-1}(i+1)R_i\big] \\
&= \tfrac{1}{n+1}\big[\textstyle\sum_{i=2}^{n} iR_i - \sum_{i=1}^{n-1} iR_i - \sum_{i=1}^{n-1} R_i\big] \\
&= \tfrac{1}{n+1}\big[\big(\textstyle\sum_{i=1}^{n} iR_i - \sum_{i=1}^{n-1} iR_i - (1)R_1\big) - \sum_{i=1}^{n-1} R_i\big] \\
&= \tfrac{1}{n+1}\big[nR_n - R_1 - \textstyle\sum_{i=1}^{n-1} R_i\big] \\
&= \tfrac{1}{n+1}\big[nR_n - R_1 - R_1 - \textstyle\sum_{i=2}^{n-1} R_i\big] \\
&= \tfrac{1}{n+1}\big[nR_n - 2R_1 - \textstyle\sum_{i=2}^{n-1} R_i\big]
\end{aligned}
$$

$$
\begin{aligned}
&\tfrac{1}{n+1}\big[nR_n - 2R_1 - \textstyle\sum_{i=2}^{n-1} R_i\big] \\
&= \tfrac{1}{n+1}\big[nR_n - 2R_1 - \big(\textstyle\sum_{i=1}^{n} R_i\big) + R_n + R_1\big] \\
&= \tfrac{1}{n+1}\big[nR_n - R_1 - \tfrac{n(n+1)}{2} + R_n\big] \\
&= \tfrac{1}{n+1}\big[(n+1)R_n - R_1 - \tfrac{n(n+1)}{2}\big] \\
&= R_n - \tfrac{1}{n+1}R_1 - \tfrac{n}{2}
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(T) \\
&= \mathrm{Var}(R_n - \tfrac{1}{n+1}R_1 - \tfrac{n}{2}) \\
&= \mathrm{Var}(R_n) + \tfrac{1}{(n+1)^2} \cdot \mathrm{Var}(R_1) - 2(1)\big(\tfrac{1}{n+1}\big) \cdot \mathrm{Cov}(R_1, R_n) \\
&= \big[1 + \tfrac{1}{(n+1)^2}\big] \cdot \mathrm{Var}(R_n) - \tfrac{2}{n+1} \cdot \mathrm{Cov}(R_1, R_n) \\
&= \big[\tfrac{n^2+2+2}{(n+1)^2}\big] \cdot \big[\tfrac{n^2-1}{12} - \tfrac{E(A)}{12n}\big] - \tfrac{2}{n+1} \cdot \big[-\tfrac{n+1}{12} + \tfrac{E(A)}{12n(n-1)}\big] \\
&= \big[\tfrac{(n^2+2n+2)(n^2-1)}{12(n+1)^2} + \tfrac{2(n+1)}{12(n+1)}\big] - \big[\tfrac{n^2+2n+2}{12n(n+1)^2} + \tfrac{2}{12n(n+1)(n-1)}\big] \cdot E(A) \\
&= \tfrac{(n^2+2n+2)n}{12(n+1)} - \tfrac{n^2+2n+2}{12(n+1)^2(n-1)} \cdot E(A)
\end{aligned}
$$

Exercise 1.2

2. Using the corresponding R code with random seed = 123 , the results are:

	Absolute error of m	**Absolute error of** v	**Relative error of** v
$n=2$	0.006998698	0.00004886125	0.0001219348
$n=4$	0.003271484	0.04938548	0.0352682729
$n=6$	0.016165597	0.1302010	0.0417196422
$n=8$	0.044365777	0.2942849	0.0533562305

$\therefore E(T)=0$ and $\mathrm{Var}(T)=\frac{(n^2+2n+2)n}{12(n+1)}-\frac{n^2+2n+2}{12(n+1)^2(n-1)}\cdot E(A)$ are verified as the differences between the absolute error of mean and variance are both very insignificant. Also, the relative error of variance are mostly less than 5%.

Exercise 1.3

1.

$\text{Cov}(R_1 + R_2, R_1 - R_2) = \text{Cov}(R_1, R_1) - \text{Cov}(R_1, R_2) + \text{Cov}(R_2, R_1) - \text{Cov}(R_2, R_2)$

Note that $\text{Cov}(R_1, R_1) = \text{Var}(R_1)$, $\text{Cov}(R_2, R_2) = \text{Var}(R_2)$ and $\text{Cov}(R_1, R_2) = \text{Cov}(R_2, R_1)$

$\therefore$ $\text{Cov}(R_1 + R_2, R_1 - R_2)$= $\text{Var}(R_1)$ - $\text{Var}(R_2)$

$= [\frac{n^2-1}{12} - \frac{E(A)}{12n}] - [\frac{n^2-1}{12} - \frac{E(A)}{12n}]$

$= 0$

2.

For simplicity, we consider $n = 2$.

$\therefore$ The possibles pair of ranks for R_1 and R_2 are (1,2) or (2,1), each with probability $= 0.5$.

$\therefore P(R_1 + R_2 = 3) = 1$ and $P(R_1 - R_2 = -1) = P(R_1 - R_2 = 1) = 0.5$

Then, to check whether $R_1 + R_2$ and $R_1 - R_2$ are independent:

We have to verify that whether $P(R_1 + R_2 = x, R_1 - R_2 = y) = P(R_1 + R_2 = x) \cdot P(R_1 - R_2 = y)$

Consider $(x, y) = (3, 1)$:

$P(R_1 + R_2 = 3, R_1 - R_2 = 1) = 0.5$ and $P(R_1 + R_2 = 3) \cdot P(R_1 - R_2 = 1) = 1 \cdot 0.5 = 0.5$

Consider $(x, y) = (3, -1)$:

$P(R_1 + R_2 = 3, R_1 - R_2 = -1) = 0.5$ and $P(R_1 + R_2 = 3) \cdot P(R_1 - R_2 = -1) = 1 \cdot 0.5 = 0.5$

$\therefore$ $R_1 + R_2$ and $R_1 - R_2$ are independent for $n = 2$.

However, by similar approach, it can be showed that $R_1 + R_2$ and $R_1 - R_2$ are not independent for $n = 3$, as $P(R_1 + R_2 = 3, R_1 - R_2 = -1) = \frac{1}{6}$, while $P(R_1 + R_2 = 3) = \frac{1}{3}$ and $P(R_1 + R_2 = -1) = \frac{1}{3}$.

Therefore, we have to check the case with finite n by a more general approach:

It can be observed that R_1, R_2 have n possible values respectively, each with probability $= \frac{1}{n}$

$\Rightarrow (R_1, R_1)$ has $n(n-1)$ possible pairs, each with probability $= \frac{1}{n(n-1)}$

$\Rightarrow (R_1 + R_2, R_1 - R_2)$ has $n(n-1)$ possible pairs, each with probability $= \frac{1}{n(n-1)}$

In general, by the marginal probability of R_1 and R_2: $P(R_1 = a) \cdot P(R_2 = b) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}$, while $P(R_1 + R_2 = a + b, R_1 - R_2 = a - b) = \frac{1}{n(n-1)} \neq \frac{1}{n^2}$ for $n \neq 2$.

$\therefore$ It is shown that $R_1 + R_2$ and $R_1 - R_2$ are generally not independent, except the case of $n = 2$.