Group: Zining Fan, Mutian Wang, Siyuan Wang
UNI: zf2234, mw3386, sw3418

Graded Homework 2- Exercise 4

1.

$$
\begin{aligned}
&P\{N_A = n_A, N_C = n_C, N_G = n_G, N_T = n_T\} \\
=&\binom{n}{n_A}\binom{n-n_A}{n_C}\binom{n-n_A-n_C}{n_G}(1-\theta)^{n_A}(\theta-\theta^2)^{n_C}(\theta^2-\theta^3)^{n_G}(\theta^3)^{n_T}
\end{aligned}
$$

2.

$$
\begin{aligned}
\log L(\theta) =& n_A \log(1-\theta) + n_C \log(\theta-\theta^2) + n_G \log(\theta^2-\theta^3) + n_T \log\theta^3 + C \\
=& n_A \log(1-\theta) + n_C \log(1-\theta) + n_G \log(1-\theta) \\
&+ n_C \log\theta + 2n_G \log\theta + 3n_T \log\theta + C \\
&\text{where } C \text{ is a constant}
\end{aligned}
$$

MLE equation:

$$
\begin{aligned}
0 =& \frac{d}{d\theta}\log L(\theta) \\
=& -\frac{n_A+n_C+n_G}{1-\theta} + \frac{n_C+2n_G+3n_T}{\theta}
\end{aligned}
$$

Solve for θ, and we have

$$
\hat{\theta} = \frac{N_C+2N_G+3N_T}{N_A+2N_C+3N_G+3N_T}
$$

3. $\hat{\theta}$ is consistent and asymptotically normal distributed, so we only need to calculate $I(\theta)$.

$$
\begin{aligned}
I(\theta) =& -E\left[\frac{d^2}{(d\theta)^2}\log L(\theta)\right] \\
=& -E\left[\frac{d}{d\theta} - \frac{n_A+n_C+n_G}{1-\theta} + \frac{n_C+2n_G+3n_T}{\theta}\right] \\
=& E\left[\frac{n_A+n_C+n_G}{(1-\theta)^2} + \frac{n_C+2n_G+3n_T}{\theta^2}\right] \\
=& \frac{E[n_A]+E[n_C]+E[n_G]}{(1-\theta)^2} + \frac{E[n_C]+2E[n_G]+3E[n_T]}{\theta^2} \\
=& n\cdot\frac{1+\theta+\theta^2}{\theta-\theta^2}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\hat{\theta} \xrightarrow[n\to\infty]{D}& N\left(\theta, \frac{1}{n\cdot I(\theta)}\right) \\
=& N\left(\theta, \frac{\theta(1-\theta)}{n^2(1+\theta+\theta^2)}\right)
\end{aligned}
$$

4. We claim $a_A = 0, a_C = a_G = a_T = \frac{1}{n}$ can satisfy that $E[T] = \theta$. In this case, $T = \frac{N_C+N_G+N_T}{n}$.

$$
\begin{aligned}
E[T] &= \frac{E[N_C + N_G + N_T]}{n} \\
&= \frac{E[n - N_A]}{n} \\
&= 1 - \frac{E[N_A]}{n} \\
&= 1 - \frac{n(1-\theta)}{n} \\
&= \theta
\end{aligned}
$$

Therefore, T is an unbiased estimator.

5.

$$
\begin{aligned}
var(T) =& var\left(\frac{N_C + N_G + N_T}{n}\right) \\
=& var\left(\frac{n - N_A}{n}\right) \\
=& \frac{1}{n^2} \cdot var\left(N_A\right) \\
=& \frac{1}{n^2} \cdot n(1-\theta)\theta \\
=& \frac{\theta(1-\theta)}{n}
\end{aligned}
$$

$$
var(\hat{\theta}) \xrightarrow[n\to\infty]{} \frac{1}{n \cdot I(\theta)} = \frac{\theta(1-\theta)}{n^2(1+\theta+\theta^2)}
$$

$$
\text{Asymptotic\_Eff}(T, \hat{\theta}) = \frac{\text{MSE}(T)}{\text{MSE}(\hat{\theta})} = \frac{var(T)}{var(\hat{\theta})} = n(1+\theta+\theta^2) \xrightarrow[n\to\infty]{} \infty
$$

Thus T is much worse than $\hat{\theta}$.

6. MLE without common parameter θ is still consistent and asymptotically normal distributed. This MLE includes four estimators, $\hat{p_i} = \frac{n_i}{n}$, where $i \in \{A, C, G, T\}$ and $\sum \hat{p_i} = 1$.

   Since $\hat{\theta}$ knows more information than $\hat{p_i}$, we think $\hat{\theta}$ has a lower asymptotic variance.

   Previous point shows that $\hat{\theta}$ is "infinitely" better than T, so we think $\hat{p_i}$ should also be better than T.

   In conclusion, the new MLE estimator is consistent. As for asymptotic efficiency, it is worse than $\hat{\theta}$ but better than T.

7. We will use likelihood ratio test.

   $H_0 : \mathbf{p} = \mathbf{p}(\theta)$ v. $H_1 : \mathbf{p}$ does not depend on a common parameter θ.

   H_0 means $\hat{\theta}$ is correct, and H_1 means $\hat{p_i}$ in previous point is correct.

   Test statistic $\Lambda = 2\log L(H_1) - 2\log L(H_0)$. $\log L(H_0)$ refers to the log likelihood under H_0, which has already been shown earlier. $\log L(H_1)$ is very similar.

   $\Lambda \sim \chi^2_2$ and $p - value = P\{\chi^2_2 \geq \Lambda\}$.