STAT 5703 Homework 2 Exercise 4

Shijie He(sh3975), Yunjun Xia(yx2569), Shuyu Huang(sh3967)

3/1/2020

Part 1

Joint distribution of (N_A, N_C, N_G, N_T) is:

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

Part 2

$$\begin{aligned} l(N_A, N_C, N_G, N_T; \theta) &= log f(N_A, N_C, N_G, N_T; \theta) \\ &= log(\frac{n!}{n_A!n_C!n_G!n_T!}) + n_A log(1-\theta) + n_C log(\theta-\theta^2) + n_G log(\theta^2-\theta^3) + n_T log(\theta^3) \end{aligned}$$

$$\begin{aligned} \frac{\partial l(N_A, N_C, N_G, N_T; \theta)}{\partial \theta} &= -\frac{n_A}{1-\theta} + (1-2\theta)\frac{n_C}{\theta-\theta^2} + (2\theta-3\theta^2)\frac{n_G}{\theta^2-\theta^3} + 3\theta^2\frac{n_T}{\theta^3} = 0 \\ \theta(-n_A - 2n_C - 3n_G - 3n_T) &= -n_C - 2n_G - 3n_T \\ \hat{\theta}_{MLE} &= \frac{n_C + 2n_G + 3n_T}{n_A + 2n_C + 3n_G + 3n_T} \end{aligned}$$

Part 3

by property of MLE estimator: (Note: $I(\theta_0)$ is fishier information)

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow[n\to\infty]{D} N(0, I(\theta_0)^{-1})$$

$$
\begin{aligned}
I_n(\theta) &= -\mathbb{E}[\frac{\partial^2}{\partial\theta\partial\theta} logL(X_1, ...X_n;\theta)] \\
&= -\mathbb{E}[-\frac{n_A}{(1-\theta)^2} + \frac{n_C(-2\theta^2+2\theta-1)}{\theta^2(1-\theta^2)} + \frac{n_G(-3\theta^2+4\theta-2)}{\theta^2(1-\theta)^2} - \frac{3n_T}{\theta^2}] \\
&= \frac{1}{(1-\theta)^2}\mathbb{E}[N_A] - \frac{(-2\theta^2+2\theta-1)}{\theta^2(1-\theta^2)}\mathbb{E}[N_C] - \frac{(-3\theta^2+4\theta-2)}{\theta^2(1-\theta)^2}\mathbb{E}[N_G] + \frac{3}{\theta^2}\mathbb{E}[N_T] \\
&= \frac{1}{(1-\theta)^2}Np_A - \frac{(-2\theta^2+2\theta-1)}{\theta^2(1-\theta^2)}Np_C - \frac{(-3\theta^2+4\theta-2)}{\theta^2(1-\theta)^2}Np_G + \frac{3}{\theta^2}Np_T \\
&= N(\frac{1+\theta+\theta^2}{\theta(1-\theta)}) \\
&= nI(\theta)
\end{aligned}
$$

$$
\begin{aligned}
I(\theta) &= \frac{I_n\theta}{N} = \frac{1+\theta+\theta^2}{\theta(1-\theta)} \\
I(\theta)^-1 &= \frac{\theta(1-\theta)}{1+\theta+\theta^2}
\end{aligned}
$$

Therefore the asymptotic distribution is

$$
\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow[n\to\infty]{D} N(0, \frac{\theta(1-\theta)}{1+\theta+\theta^2})
$$

Part 4

Since T is unbiased estimator for θ

$$
\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[a_A N_A + a_C N_C + a_G N_G + a_T N_T] \\
&= \sum_{x:A,C,G,T} a_x \mathbb{E}[N_x] \\
&= \sum_{x:A,C,G,T} a_x n p_x \\
&= n \sum_{x:A,C,G,T} a_x p_x \\
&= n(a_A(1-\theta) + a_C(\theta-\theta^2) + a_G(\theta^2-\theta^3) + a_T(\theta^3)) \\
&= \theta
\end{aligned}
$$

Therefore, we can find

$$
a_A = 0, a_C = a_G = a_T = \frac{1}{n}
$$

Part 5

$$
\begin{aligned}
Var(T) &= Var(\frac{1}{n}(n_C + n_G + n_T)) \\
&= \frac{1}{n^2}Var(n_C + n_G + n_T) \\
&= \frac{1}{n^2}Var(n - n_A) \\
&= \frac{1}{n^2}np_A(1-p_A) \\
&= \frac{\theta(1-\theta)}{n}
\end{aligned}
$$

Relative Efficency:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2 \xrightarrow[n\to\infty]{D} \frac{\theta(1-\theta)}{n(1+\theta+\theta^2)}$$

$$MSE(T) = Var(T) = \frac{\theta(1-\theta)}{n}$$

$$eff(\hat{\theta}, T) = \frac{MSE(\hat{\theta})}{MSE(T)} = 1 + \theta + \theta^2$$

Part 6

The MLE without dependence of θ:

$$\begin{cases} \hat{p_A} = \frac{N_A}{n} \\ \hat{p_C} = \frac{N_C}{n} \\ \hat{p_G} = \frac{N_G}{n} \\ \hat{p_T} = \frac{N_T}{n} \end{cases}$$

The MLE with dependence of θ:

$$\begin{cases} \hat{p'_A} = 1-\theta \\ \hat{p'_C} = \theta - \theta^2 \\ \hat{p'_G} = \theta^2 - \theta^3 \\ \hat{p'_T} = \theta^3 \end{cases}$$

For estimator T, assuming $a_A = 0, a_C = a_G = a_T = \frac{1}{N}$,

$$T = \frac{N_C + N_G + N_T}{n}$$

T is an estimator for $1 - p_A$, $1 - \hat{p_A}$, is identical to estimator without assumption on θ.

Part 7

We would like to use likelihood ratio test to test the hypothesis. Test Statistics:

$$\begin{aligned} \Lambda_n &= 2\{\ell(\hat{\mathbf{P}})_{\mathbf{p}(\theta)} - \ell(\hat{\mathbf{P}})_{\mathbf{p}(\theta')}\} \\ &= 2(N_A log(1-\theta) + N_C log(\theta - \theta^2) + N_G log(\theta^2 - \theta^3) + N_T log(\theta^3) \\ &- (N_A log(\frac{N_A}{N}) + N_C log(\frac{N_C}{N}) + N_G log(\frac{N_G}{N}) + N_T log(\frac{N_T}{N}))s \end{aligned}$$

where $\Lambda_n \xrightarrow[n\to\infty]{D} \chi^2_4$