Exercise4

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

2/29/2020

(1)

$p(N_A, N_C, N_G, N_T) = \binom{N}{N_A}\binom{N-N_A}{N_C}\binom{N-N_A-N_C}{N_G} p_A^{N_A} p_C^{N_C} p_G^{N_G} p_T^{N_T}$

(2)

$\frac{\partial log(p)}{\partial \theta} = -N_A \frac{1}{1-\theta} + N_C \frac{1-2\theta}{\theta-\theta^2} + N_G \frac{2\theta-3\theta^2}{\theta^2-\theta^3} + N_T \frac{3\theta^2}{\theta^3}$
$= -N_A \frac{1}{1-\theta} + N_C \frac{1-2\theta}{\theta-\theta^2} + N_G \frac{2-3\theta}{\theta-\theta^2} + N_T \frac{3}{\theta} = 0$

So $(N_A + 2N_C + 3N_G + 3N_T)\theta = N_C + 2N_G + 3N_T$

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

(3)

$I_n(\theta) = -E(\frac{\partial^2 L(\theta; N_A, N_C, N_G, N_T)}{\partial \theta^2}) = -E(\frac{-b\theta+2a\theta-a}{(1-\theta)^2\theta^2})$

$a = N_C + 2N_G + 3N_T$, $b = N_A + 2N_C + 3N_G + 3N_T$

So $E(a) = N(1+\theta+\theta^2)\theta$, $E(b) = N(1+\theta+\theta^2)$

Then $I(\theta) = I_n(\theta)/N = \frac{1+\theta+\theta^2}{(1-\theta)\theta}$

So $\sqrt{N}(\hat{\theta}-\theta) \to^D N(0, \frac{(1-\theta)\theta}{(1+\theta+\theta^2)})$

(4)

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

Then $ET = \frac{E(N_C+N_G+N_T)}{N} = \frac{N(\theta-\theta^2)+N(\theta^2-\theta^3)+N\theta^3}{N} = \theta$

(5)

So $VarT = Var(N-N_A)/N^2 = VarN_A/N^2 = \frac{(1-\theta)\theta}{N}$

The asymptotic MSE of $\hat{\theta}$ is

$MSE(\hat{\theta}) = Var\hat{\theta} + bias(\hat{\theta}, \theta)^2 \to^D \frac{(1-\theta)\theta}{N(1+\theta+\theta^2)}$

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

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

(6)

Without θ, the MLE of p is

$\hat{p}_A = \frac{N_A}{N}$, $\hat{p}_C = \frac{N_C}{N}$, $\hat{p}_G = \frac{N_G}{N}$, $\hat{p}_T = \frac{N_T}{N}$

With θ, the MLE of p is

$\hat{p}_A = 1-\hat{\theta}$, $\hat{p}_C = \hat{\theta}-\hat{\theta}^2$, $\hat{p}_G = \hat{\theta}^2-\hat{\theta}^3$, $\hat{p}_T = \hat{\theta}^3$

For estimator T, if $a_A = 0$, $a_C = a_G = a_T = \frac{1}{n}$, the estimator of p_A is same as the estimator of p_A without θ.

(7)

We can use the likelihood ratio test (under H_0 we have):

$\Lambda_n = 2\left(\log L_n(\hat{\theta}) - \log L_n\left(\hat{\theta}^c\right)\right)$
$= 2(N_A log\frac{N_A}{N} + N_C log\frac{N_C}{N} + N_G log\frac{N_G}{N} + N_T log\frac{N_T}{N} - N_A log(1-\theta) - N_C log(\theta - \theta^2) - N_G log(\theta^2 - \theta^3) - N_T log(\theta^3)) \to^D \chi^2_4$