CM224 HW 2 Solution

Ashish Kumar Singh (UID:105479019)

October 14, 2021

Problem 1. Possible number of haplotypes for 6 SNPs range.

Answer 1. v) Between [81,100]

Solution 1. Since out of the 6 SNPs, 5 are bi-allelic and 1 is tri-allelic, Total number of haplotypes are $2^5 * 3^1 = 32 * 3 = 96$

Problem 2. What is the MLE of a ans b for uniform distribution samples ?

Answer 2. g) $a = 0.1$, $b = 0.84$

Solution 2. For uniform distribution, MLE of a and b is minimum and maximum of all the samples recorded respectively.

Problem 3. Haplotype frequencies with $p_1 = 0.1$ and Langrange multiplier.

Answer 3. j) $p_i = \frac{c_i}{\lambda}$, $\lambda = \frac{n-c_1}{0.9}$

Solution 3. Let $c_1, c_2, ..., c_t$ be the counts of t unique haplotypes and total count be n. Given $p_1, p_2, ..., p_t$ frequencies of haplotypes with $p_1 = 0.1$. Since, p_1 is known, we remove it from all the equations,
Likelihood function for the given data,

$$L(p_2, ..p_t; c_2, ..., c_t) = \binom{n - c_1}{c_2}\binom{n - c_1 - c_2}{c_3}...\binom{n - c_1 - c_2... - c_{t-1}}{c_t} p_2^{c_2} p_3^{c_3}...p_t^{c_t}$$

$$\text{such that } \sum_{i=2}^{n} p_i = 0.9$$

Taking log of likelihood,

$$LL(p_2, ..p_t; c_2, ..., c_t) = \sum_{i=2}^{n} \log \binom{n - c_1 - c_2... - c_{i-1}}{c_i} + \sum_{i=2}^{n} c_i \log p_i$$

From the constraint, we can write $g(\vec{p}) = 0.9 - \sum_{i=2}^{n} p_i = 0$ Using Langrange multiplier λ,

$$f(\vec{p}) = L(p_2, ..p_t; c_2, ..., c_t) + \lambda g(\vec{p})$$

Taking derivative wrt p_i for $i \in (2, t)$ and λ and setting both to zero,

$$\tfrac{c_i}{p_i} - \lambda = 0 \; ; \; 0.9 - \sum_{i=2}^{n} p_i = 0$$

$$\tfrac{c_i}{p_i} = \lambda \; ; \; \sum_{i=2}^{n} p_i = 0.9$$

$$\sum_{i=2}^{n} p_i = \sum_{i=2}^{n} \tfrac{c_i}{\lambda} = 0.9$$

$$\tfrac{c_2 + c_3 + ... + c_t}{\lambda} = 0.9$$

$$\tfrac{n - c_1}{\lambda} = 0.9$$

$$\lambda = \tfrac{n - c_1}{0.9} \text{ and } p_i = \tfrac{c_i}{\lambda}$$