D

1. **MLE of Allele Frequency.** Suppose you have the following alleles and their respective counts:
0 - 600
1 - 200

What is the maximum likelihood estimate of the frequency of the 1 allele?

Select from the following options and justify briefly in the box why you choose your answer.

(a) 0.1

(b) 0.75

(c) 0.9

(d) 0.25

$$\frac{\text{frequency}}{\text{total count}} = \frac{200}{200+600} = \frac{2}{8} = \frac{1}{4}$$

by derivation in class

C

2. **MLE question of Haplotype Frequency.**

   Suppose you have the following haplotypes and their respective counts:

   1001 - 300
   1101 - 1200
   0011 - 400
   0101 - 100

   What is the maximum likelihood estimate of the haplotype 1101?

   Select from the following options and justify briefly in the box why you choose your answer.

   (a) 0.1
   (b) 0.25
   (c) 0.6
   (d) 0.2

By derivation

$$\hat{p} = \frac{\text{frequency}}{\text{total } \#} = \frac{1200}{300+1200+400+100} = \frac{12}{20} = \frac{3}{5}$$

3. **Trio Phasing.** Suppose you have the following trio with the genotypes as below

$$\begin{aligned} \text{Parent1} &= 122102 \\ \text{Parent2} &= 011120 \\ \text{Child} &= 112011 \end{aligned}$$

B

Assume there is no recombination. What is the phase of the child in this trio?

Select from the following options and justify briefly in the box why you choose your answer.

(a) 011001, 101010

~~(b) 111001, 001010~~

(c) 011001, 111010

(d) 110010, 001010

*11*01 *11*01 00*010 0*0*10 **100** 0*10**

4. **Clark's algorithm.**
Suppose you have the following genotypes:

AGHAA
GAGHA
HHGHA
AHAAA
HHHHA

AGAAA AAAAA
AGGAA
GAGAA
GAGGA

Perform Clark's algorithm, starting from the first genotype, what is the final haplotype(s) that were added to your known set.

Select from the following options and justify briefly in the box why you choose your answer.

(a) AAAAA
(b) GGGGA
(c) AGAAA
(d) AGGAA

5. **Haplotype phasing with the EM algorithm.**

Suppose you have genotypes $G = \{01210, 10222, 00110\}$. Perform one round of the EM algorithm for haplotype phasing (assume that the haplotypes are equally probable $p_1 = p_2 ... = p_n = 1/n$). After one round of EM, what haplotype(s) have the highest estimated probability?

Select from the following options and justify briefly in the box why you choose your answer.

C

(a) 00110, 10111

(b) 00100, ~~00111, 00010, 00000~~

(c) 10111, 00111, 00110, 00100

(d) ~~10111, 00111, 00100~~

01210	01100	0.25
	00110	0.25
	01110	0.25
	00100	0.25
10222	10111	0.5
	00111	0.5
00110	00100	0.25
	00010	0.25
	00000	0.25
	00110	0.25

$\frac{1}{10}$ - rando

01100	$\frac{0.25}{3}$
00110	$\frac{0.5}{3}$
01110	$\frac{0.25}{3}$
00100	$\frac{0.5}{3}$
10111	$\frac{0.5}{3}$
00111	$\frac{0.5}{3}$
00010	$\frac{0.25}{3}$
00000	$\frac{0.25}{3}$

6. **Likelihood optimization**

Consider the likelihood function of the haplotype frequency estimation problem with no missing data. In this setting, the parameters are the haplotype allele frequencies $p_1, \dots, p_n$, satisfying $\sum_{i=1}^{n} p_i = 1$, and $p_i \geq 0$, and the data is the haplotype counts $c_1, \dots, c_n$, and let $c = c_1 + \dots + c_n$. We showed in class that the likelihood function $L(p_1, \dots, p_n)$ satisfies

$$\log L(p_1, \dots, p_n) = \sum_{i=1}^{n} c_i \log(p_i)$$

A

Assume we are running a gradient ascent with projections algorithm, and we start from the guess $p_1^{(0)} = p_2^{(0)} = \dots = p_n^{(0)} = \frac{1}{n}$. Let $p_1^{(1)}, \dots, p_n^{(1)}$ be the next point that the algorithm reaches. Which of the following is true:

(a) There is $\epsilon > 0$ such that for every i, j we have $p_i^{(1)} - p_j^{(1)} = (c_i - c_j)\epsilon$.

(b) There is $\epsilon > 0$ such that for every i, j we have $p_i^{(1)} - p_j^{(1)} = (c_j - c_i)\epsilon$.

(c) There is $\epsilon > 0$ such that for every i we have $p_i^{(1)} = (c_i - \frac{c}{n})\epsilon$.

(d) There is $\epsilon > 0$ such that for every i, j we have $p_i^{(1)} + p_j^{(1)} = \frac{2}{n} + (c_i + c_j)\epsilon$.

Select from the above options and justify briefly in the box why you choose your answer.

$$\vec{\nabla} L = \left(\frac{c_1}{p_1}, \dots \frac{c_n}{p_n}\right)$$

$$\lambda = \frac{\sum \frac{c_i}{p_i}}{n}$$ For (0), $\lambda = c$

$$u = \left(\frac{c_1}{p_1} - \lambda, \dots \frac{c_n}{p_n} - \lambda\right)$$

$$p_i^{(1)} = (nc_i - c)\epsilon + \frac{1}{n}$$