Correcting mutation rates for observer bias

Matty Allan

February 29, 2024

Computing the mutation rates that would be observed given real mutation rates

Suppose that, for a sequence of length L, you know

- the real (unobserved) mutation rate for each position j: μ_j ($0 \leq \mu_j < 1$; $1 \leq j \leq L$)

- the real (unobserved) fraction of reads that start at position a and end at position b: η_{ab} ($0 \leq r_{ab} \leq 1$; $1 \leq a \leq b \leq L$)

To reframe the problem in terms of probabilities, three events can happen:

S The read spans position j, i.e. $a \leq j \leq b$.

M Base j is mutated in the read.

G The read contains no pair of mutations that are closer than the minimum gap g.

You want to calculate the mutation rate that you would observe at each position j due to the observer bias: m_j.