

Lengths

Lecture 10/6 Notes

• Un-gapped multiple alignment (Review)

- Ex. $\begin{matrix} W & W & V & R \\ W & F & Y & . \\ . & . & . & . \\ . & . & . & . \end{matrix}$ } Observed $P(W) = \frac{6}{25} \dots$ use to generate random sequences of proteins $\Rightarrow$ optimal local alignment for many pairs of random sequences $O(\log N)$
 Frequency $P(R) = \frac{5}{25}$ $\Rightarrow$ optimal global alignment for many pairs of random sequences $O(N)$

• Karlin-Altschul Statistics

- ① Scoring matrix needs at least 1 positive score
- ② Expected score of scoring matrix must be negative
- ③ Letters of random sequences must be i.i.d
- ④ Sequences are infinitely long
- ⑤ Alignments do not contain gaps

$\rightarrow a_{ij} = p_i p_j e^{-\lambda S_{ij}}$ is the distribution of A.A pairs in max scoring sequence

• "Lady Testing Tea" - Fisher Null Hypothesis Experiment

- Null Hypothesis: Never proved or established but is possibly disproved by experimentation
- Premise: Provide 8 random order tea cups, 4 prepared w/ milk first, 4 w/ tea first
- $\rightarrow$ Select 4 cups prepared by a given method

- Null Hypothesis: Lady cannot distinguish the tea

- Test Statistic: Number of success in selecting 4 cups

- Possible Combinations: $\binom{8}{4} = \frac{8!}{4!(8-4)!} = 70$

- $\rightarrow$ 0 success: $\frac{1}{70}$ $\binom{4}{0} = 1$ (0 correct) $\binom{4}{4} = 1$ (4 failures) $1 \times 1 = 1$
- $\rightarrow$ 1 success: $\frac{16}{70}$ $\binom{4}{1} = 4$ (1 correct) $\binom{4}{3} = 4$ (3 failures) $4 \times 4 = 16$
- $\rightarrow$ 2 success: $\frac{36}{70}$ $\binom{4}{2} = 6$ (2 correct) $\binom{4}{2} = 6$ (2 failures) $6 \times 6 = 36$
- $\rightarrow$ 3 success: $\frac{16}{70}$ $\binom{4}{3} = 4$ (3 correct) $\binom{4}{1} = 4$ (1 failure) $4 \times 4 = 16$
- $\rightarrow$ 4 success: $\frac{1}{70}$ $\binom{4}{4} = 1$ (4 correct) $\binom{4}{0} = 1$ (0 failures) $1 \times 1 = 1$

- Number of successes is distributed by hypergeometric distribution

- Conventional threshold for rejecting null hypothesis is $< 5\%$ $\Rightarrow$ can only reject null hypothesis with 4 successes ($\frac{1}{70}$)