computing transition probabilities and visits of the random walk

Pete Rigas, Lambert Lab

Wednesday, April 8

coding up the probability measure μ

from the expression in last week's meeting,

$$\mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda_{\mathrm{mis}}) = \frac{\prod_{i\sim j} \exp\Big(- \frac{1}{(|\{\mathrm{mat}\}|+2)^3} \mathcal{J}_{ij}\sigma_i\sigma_j \Big) \cdots}{1 + \lambda_c e^{-\beta\epsilon_{\mathrm{PAM}}} + \cdots}$$

$$\frac{\cdots \prod_{i\not\sim j} \exp\Big(- \frac{1}{(|\{\mathrm{mis}\}|+2)^3} \mathcal{J}_{ij}(1 - \sigma_i\sigma_j) \Big)}{\cdots N_{\mathrm{mis}} \sum_{\mathrm{mis}} \lambda_{\mathrm{mis}} e^{-\lambda_{\mathrm{mis}} X_{\mathrm{mis}}}} ,$$

we can simplify terms in our implementation by setting $|\{\mathrm{mat}\}| = N - |\{\mathrm{mis}\}|$, as to avoid having an extra parameter in our Matlab function

framing computations of the transition probabilities in a more approachable manner

algorithmically, with $\mu(N, N_{\text{mis}}, |X_{\text{mis}}|, |\text{mis}|, \lambda_{\text{mis}})$, we can achieve robust control over the visits of the random walk to the position of binding by:

- defining the N_{mis} parameters with a sequence $\{N_{\text{mis}}(i)\}_{1 \leq i \leq 20}$, that is monotonically decreasing,
- a fixed choice of the λ_c parameter, as long as it is $<< 1$,
- and finally, a monotonically decreasing set of parameters $\{\lambda_{\text{mis}}(i)\}_{1 \leq i \leq 20}$ corresponding to the exponential mismatch term in the partition function

simplifying μ for 1, 2, or 3, base pair mismatches, for a position of binding $N \equiv 15$

naturally, we can express μ for each possibility of base pair mismatches with the following, from $\mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda_{\mathrm{mis}})$, as, observing that $\exp(-\frac{(25-1)}{2} \approx 6.14 \cdots \times 10^{-6}$:

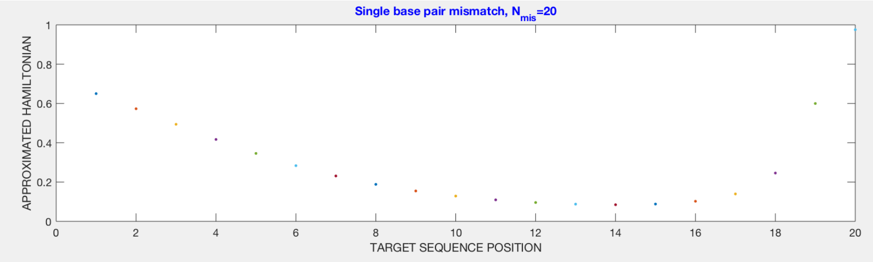
- <u>1 mismatch:</u> $\mu_1\bigg(15, (25-1)^3, |X_{\mathrm{mis}}|_1 = (1), 1, 6.14 \times 10^{-6})\bigg)$

$$= \frac{\prod_{i \sim j} \exp\bigg(- \frac{1}{(14+2)^3} \mathcal{J}_{ij}(1)\bigg) \cdots}{1 + \lambda_c e^{-\beta \epsilon_{\mathrm{PAM}}} + \cdots}$$

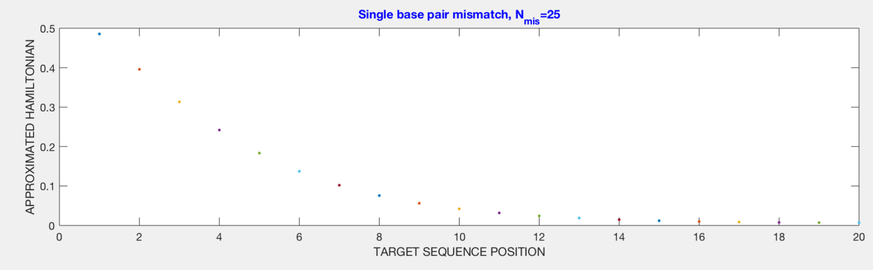
$$\frac{\cdots \prod_{i \not\sim j} \exp\bigg(- \frac{1}{(1+2)^3} \mathcal{J}_{ij}(1 - (-1))\bigg)}{\cdots (25-1)^3 (6.14 \times 10^{-6}) e^{-(6.14 \times 10^{-6})1}} \ ,$$

short remark: motivation for plots in the next slide

goal: to numerically justify the choice of N_{mis}, in addition to motivating the choice of other parameters to construct transition probability sequences, refer to the plots below, which demonstrate that choosing $N_{\mathrm{mis}} \equiv 20$ artificially increases the value of μ at the last base pair (20) which we dismiss as a poor parameter choice plotting μ across all 20 base pair sequences for binding position $N \equiv 15$, $N_{\mathrm{mis}} \equiv 20$



plotting μ across all 20 base pair sequences for binding position $N \equiv 15$, $N_{\text{mis}} \equiv 25$



towards an explicit formula

- candidate formula 1: for a mismatch at position j, given the transition probability at i, we can iteratively compute the transition probability at the base pair mismatch at j with

$$p_{t_j} = \Bigg(1 - \mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda_{\mathrm{mis}})\Bigg) p_{t_i}$$

- candidate formula 2: alternatively, one can calculate the same transition probability at j with the recursive formula,

$$p_{t_j} = abs\Bigg(\Bigg(1 - \big(\mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda_{\mathrm{mis}}) + \cdots \\ \cdots \mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda'_{\mathrm{mis}})\Bigg)\Bigg) p_{t_i} \ ,$$

  namely subtracting μ for distinct λ_{mis}, λ'_{mis}

future directions

Note from previous slide: That's why it is important to make a "good" choice of parameters so that μ is a probability measure, otherwise the transition probabilities could be negative which we dismiss as degenerate sequences!

Next week, I will continue working to

- **generalize** the procedure to 2 and 3 base pair mismatches, by calculating μ for the 1 base pair mismatch (as I have already shown), and then appending another 1, or 2, base pair mismatches in the summation to the partition function $\mathcal{Z}$
- develop more robust formulae for computing the transition probabilities, to feed to my code to compute random walk visits