**Motivation: Likelihoods with an unknown ancestral tree**

Given a set of aligned sequences, e.g.

Sequence 1: ... T ... G ... A ... A ...
Sequence 2: ... T ... G ... A ... A ...
Sequence 3: ... A ... A ... T ... A ...
Sequence 4: ... T ... G ... A ... G ...
Sequence 5: ... A ... A ... A ... A ...

the probability of having evolved from some initial sequence (here ... A ... A ... A ... A ...) may be expressed:

$$P \left( \mathcal{A} \right) = \sum_{x \in \text{Histories}} P(x) P(x)$$

which in turn may be expressed by conditioning on the most recent event:

$$P \left( \mathcal{A} \right) = \frac{1}{\theta + 4} P \left( \mathcal{A}^{\theta} \right) + \theta \frac{1}{\theta + 4} P \left( \mathcal{A} \right)$$

If we apply this conditioning-trick recursively, computing $P \left( \mathcal{A} \right)$ reduces to computing a weighted sum over all paths from "12345" to "5" in the ancestral graph below:

![Ancestral Graph Diagram]

- number of nodes = number of distinct terms in $P$-recursion,
- number of paths "12345 ... → $x$" = number of execution paths when evaluating $P \left( \mathcal{A} \right)$ via tail-recursion (without memoization/tabling).

**Importance Sampling of Ancestral Paths**

We can approximate probabilities of aligned sequences—e.g. $P \left( \mathcal{A} \right)$—by sampling ancestral histories $X_1, \ldots, X_N \sim Q \ll P$ and relying on the following approximation:

$$P \left( \mathcal{A} \right) = \sum_{x \in \mathcal{X} \cap \mathcal{P} \cap \mathcal{H}} P(x) \frac{Q(x)}{Q(X)} = E_{X \sim Q} \left[ \frac{P(X)}{Q(X)} \right] \approx \frac{1}{N} \sum_{i=1}^{N} \frac{P(X_i)}{Q(X_i)}$$

For this approach to work effectively, $Q$ should satisfy:

1. $Q$ must approximate $P$ well on the space of histories;
2. sampling $X_i \sim Q$ should be fast;
3. computing $Q(X_i)$ should be fast.

**Sequential Sampling Schemes**

Existing proposal distributions are all sequential: they construct paths step-by-step from the bottom up. They differ by how the next step in a path is sampled.

- Stephens and Donnelly (S&D): simple combinatorial sampling
- Griffiths and Taveré (G&T): $Q \approx \text{Histories}$ & $P \approx \text{Histories}$

**Path Density Bias and Path Counting**

Any step-by-step scheme which does not penalize choices which “lead to fewer choices down the line”, will be biased in favour of low-density regions of path-space, e.g.

$$b(T) = \sum_{S \subseteq \mathcal{P}, i \in S} h(\{T_i \mid i \in S\}) h(\{T_i \mid i \not\in S\}) \left( \sum_{i \in S} k_i - 2 \right) \left( \sum_{i \notin S} k_i - 1 \right)$$

whereby we here encode rooted unordered trees as nested systems of sets, e.g.

$$\mathcal{A}^{\mathcal{A}} = \{ \{\{1, 2\}\} \}, \{3, 4, 5\}.$$