$$F(a) = \frac{e^{-\lambda_a}}{\sum_b e^{-\lambda_b}}$$ ← λ_a = parameter determining frequency of a.a. a. "scaled selection coeff."

Read: Ramsey et al., Genetics 2011

$$\pi_{a\to b} = \frac{1-[F(a)/F(b)]^{1/N}}{1-F(a)/F(b)}$$

$$K = \mu N \sum_a \left[F(a) \sum_b \pi_{a\to b} \right]$$

$$\pi_{a\to b} = \frac{1-\exp\left[\frac{1}{N}\ln\frac{F(a)}{F(b)}\right]}{1-F(a)/F(b)} \approx \frac{1}{N}\frac{\ln F(a)-\ln F(b)}{1-F(a)/F(b)}$$

If we insert this into K, then N drops out.

We can do the same for codons:

$$K_s = \mu N \sum_i F(i) \sum_{j,\, i\to j \text{ is syn.}} \pi_{i\to j}$$

here, i, j are codons, $F(i)$ is equil. codon frequency for i.

Better write like this: $K_s = \mu N \sum_i \sum_{\substack{j,\\ i\to j \text{ syn}}} F(i)\, \underset{\substack{\parallel \\ 1/N}}{\pi_{i\to j}}$ in neutral case

Now, the syn sites N_s: $N_s = \cancel{\mu N} \sum_i \sum_{\substack{j\\ i\to j \text{ syn}}} F(i)$

Same for K_N, N_N

More notation: $S(i)$ = set of sites j with $i\to j$ syn.

Simulation: Simplest case: all codon for each a.a. have same fitness. a.a. fitnesses are drawn at random from $F(a) \sim e^{-\lambda d(a)}$

integer ordering drawn at random for each site. λ is linear fit.

$$d_S = \frac{K_S}{N_S} = \mu$$

$$d_N = \frac{K_N}{N_N} = \frac{\mu N \sum_i \sum_{j \in \mathcal{N}} F(i) \pi_{i \to j}}{\sum_i \sum_{j \in \mathcal{N}} F(i)}$$