Nei-Gojobori dN/dS estimate (NG86)

- For each codon `C` we define **`ES(C)`** and **`EN(C)`** - the numbers of synonymous and non-synonymous *sites* of a codon
  - e.g., **`ES(GAA) = 1/3, EN(GAA) = 8/3.`**
- May also define them as fractions of substitutions that do not lead to stop codons,
  - e.g., **`ES(GAA) = 1/3, EN(GAA) = 7/3.`**
- The sum of **`ES`** and **`EN`** over all codons in a sequence gives an estimate of expected synonymous and non-synonymous **sites** in a sequence.
- For two sequences (the target of the original method), we average **`ES(C)`** and **`EN(C)`** at each site.
- **`EN/ES`** is thus the ***expected ratio of non-synonymous to synonymous substitutions counts under neutral evolution***

Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions
M. Nei and T. Gojobori
Mol. Biol. Evol. 3 418--426 (1986)

>5,300 citations

NG86 limitations: underestimation of substitution counts for higher divergence levels

- Simulated 100 replicates of 1000 nucleotide long sequences for various divergence levels (substitutions/site)
- Even for divergence of `0.25` (1/4 sites have mutation on average), p-distance already significantly underestimates the true level: `0.2125 (0.19-0.241 95% range)`
- Underestimation becomes progressively worse for larger divergence levels

Simulated divergence vs that estimated by p-distance