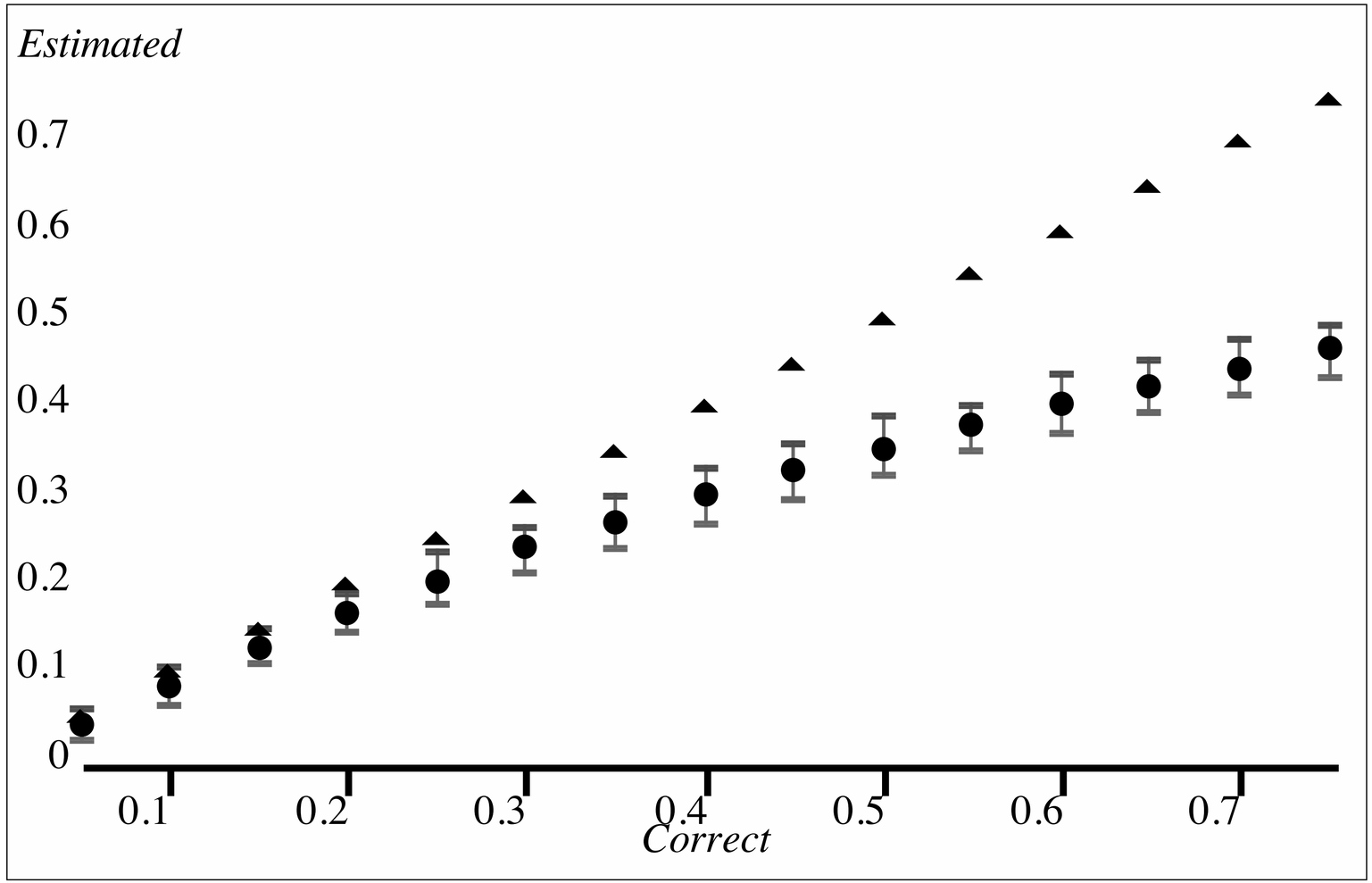
NG86 limitations: underestimation of substitution counts for higher divergence levels

- 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



Codon-substitution models

- In 1994, first tractable mechanistic evolutionary models for codon sequences were proposed by **Muse and Gaut** (MG94), and, independently, by **Goldman and Yang** (GY94) [in the same issue of MBE, back to back]
- Markov models of codon substitution provide a powerful framework for **estimating substitution rates** from coding sequence data, as they
  - *encode our mechanistic understanding of the evolutionary process,*
  - *enable one to compute the phylogenetic likelihood,*
  - *permit hypothesis testing or Bayesian inference,*
  - *systematically account for confounding processes (unequal base frequencies, nucleotide substitution biases, etc.),*
  - *afford many opportunities for extension and refinement (still happening today).*

A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome
S. V. Muse and B. S. Gaut
Mol Biol Evol 11 715-724 (1994)

~1000 citations

A codon-based model of nucleotide substitution for protein-coding DNA sequences.
N. Goldman and Z. Yang
Mol Biol Evol 11 725--736 (1994)

~2250 citations