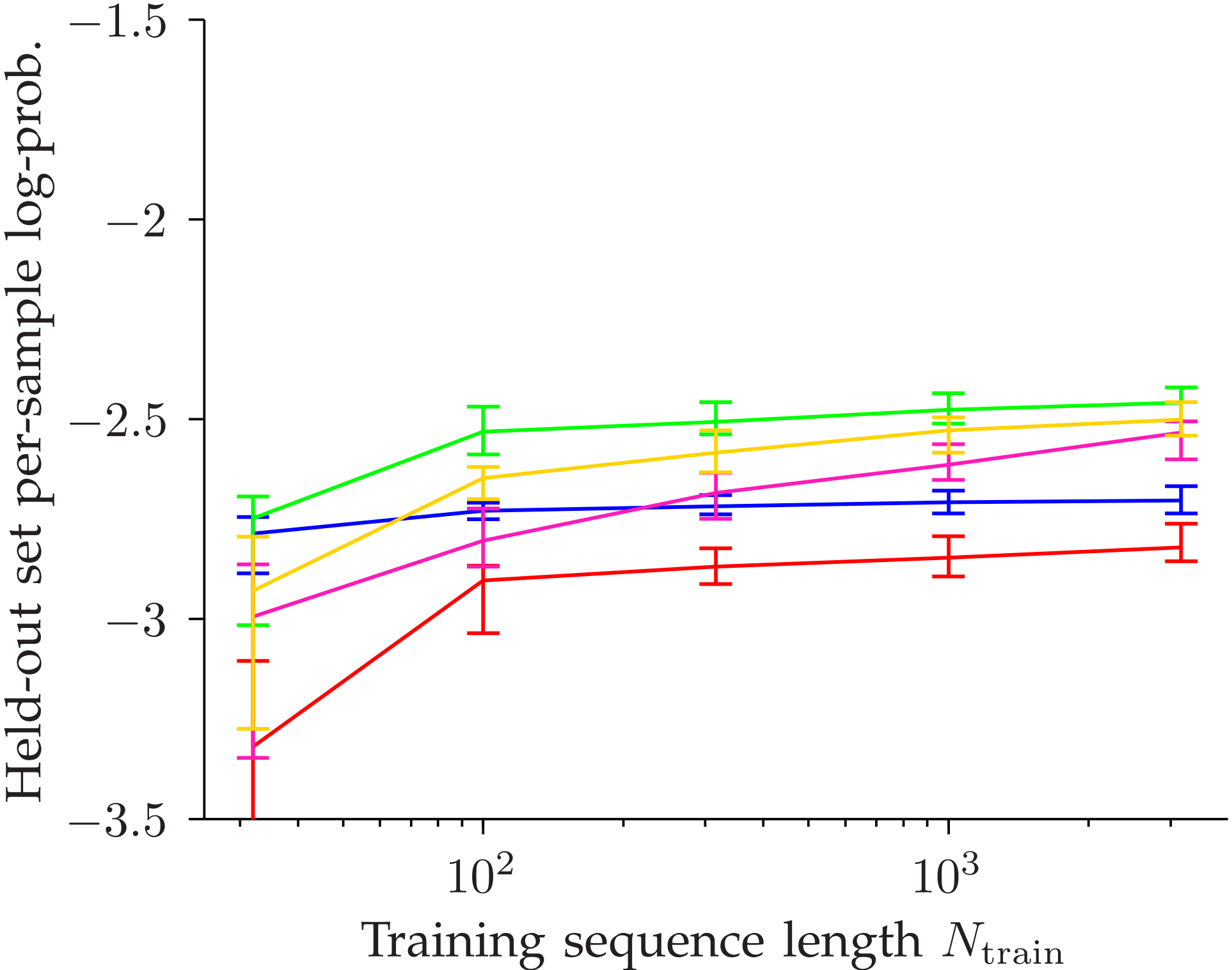
−1.5
−2
−2.5
−3
−3.5
Held-out set per-sample log-prob.
10^2
10^3
Training sequence length N_{train}