

Order-k Markov Transformer LPE Experiments

Automated run via lpe/markov_k.transformer.py

February 14, 2026

Summary

This report summarizes order-k experiments for $k \in \{1, \dots, 7\}$ with independent Beta(1,1) priors per transition state. Training used sequence length equal to rollout length: $100 \cdot 2^k$.

Core Metrics

k	Model NLL	Bayes NLL	Gap (%)	Posterior MAE	LPE Median RelErr (%)
1	0.593086	0.530185	11.86	0.2011	99.71
2	0.652844	0.540651	20.75	0.2308	59688127439485361302011904.00
3	0.631652	0.492707	28.20	0.1713	12082995.17
4	0.614127	0.445917	37.72	0.2331	1618087.60
5	0.690311	0.512228	34.77	0.0618	53197.70
6	0.683737	0.485833	40.74	0.0383	5818550.92
7	0.692454	0.535458	29.32	0.0218	910.38

Table 1: Primary metrics by order k .

Figures

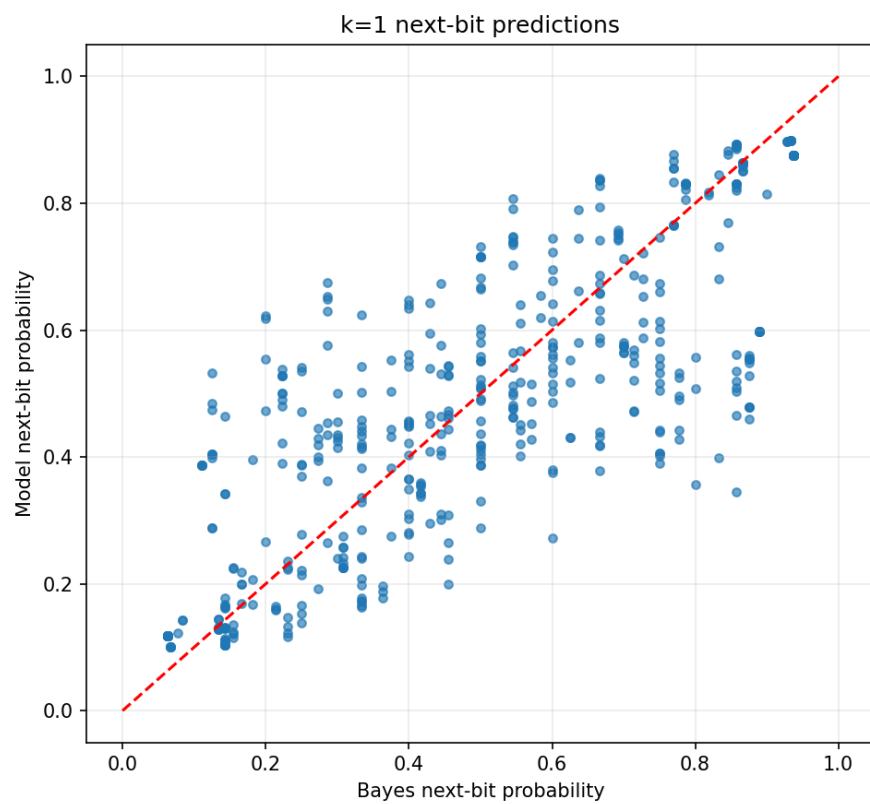


Figure 1: k=1: Step 1: next-bit prediction vs Bayes

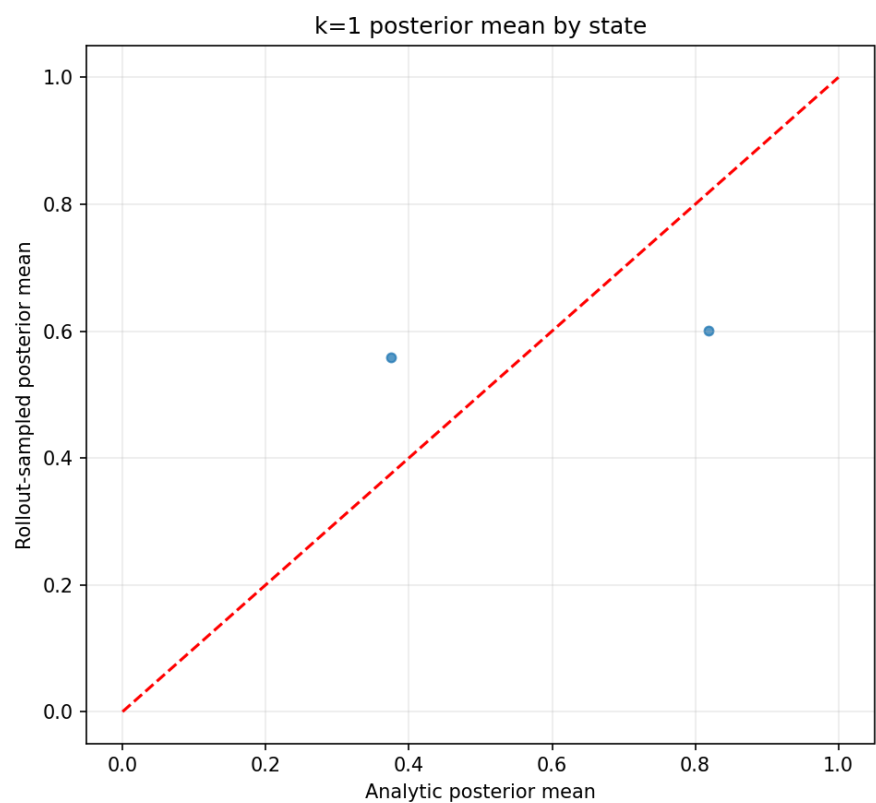


Figure 2: k=1: Step 2: posterior mean comparison

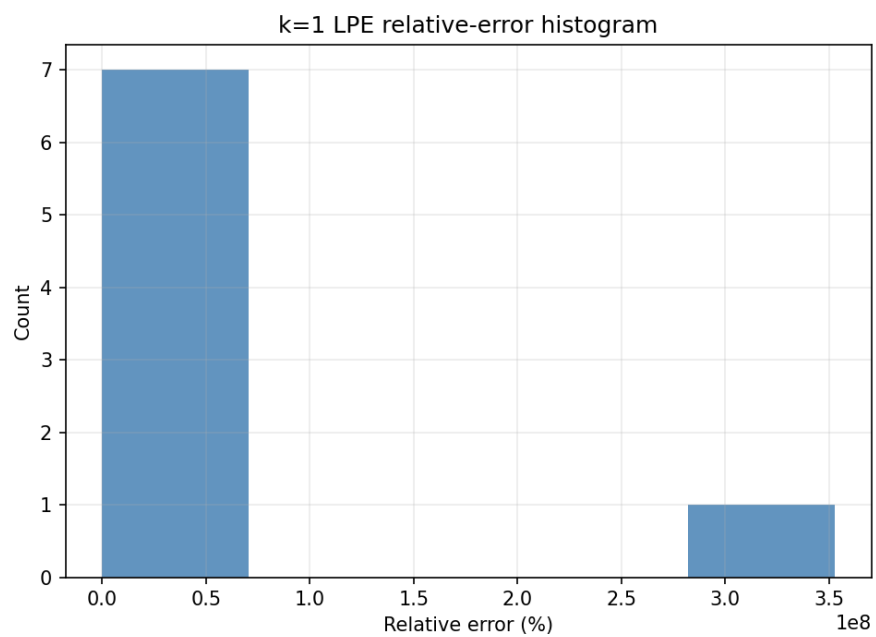


Figure 3: k=1: Step 3: LPE relative-error histogram

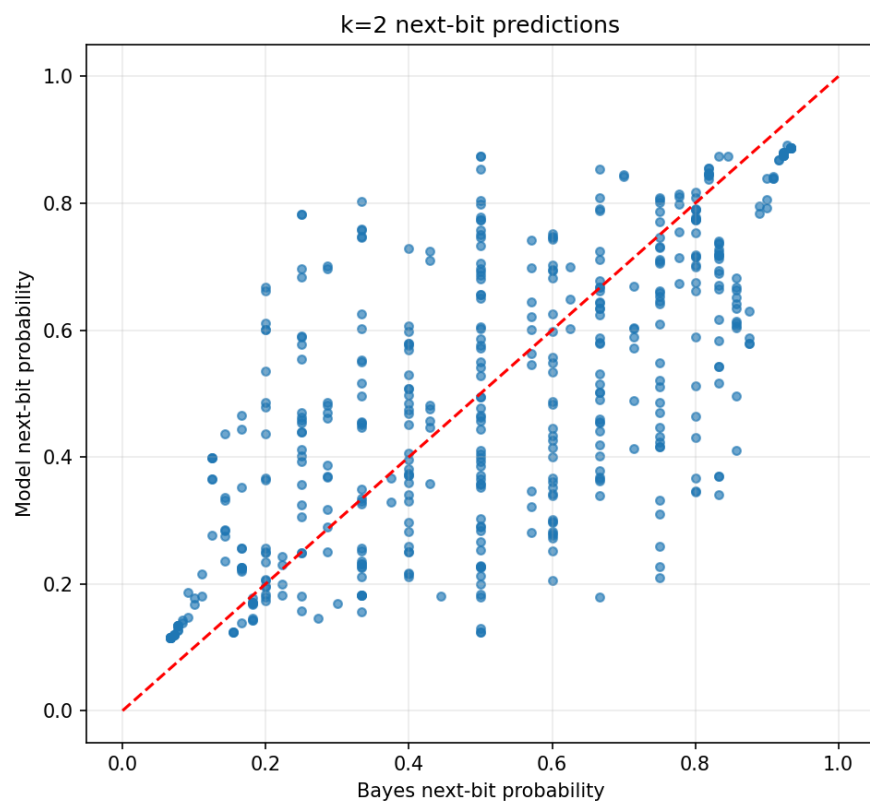


Figure 4: k=2: Step 1: next-bit prediction vs Bayes

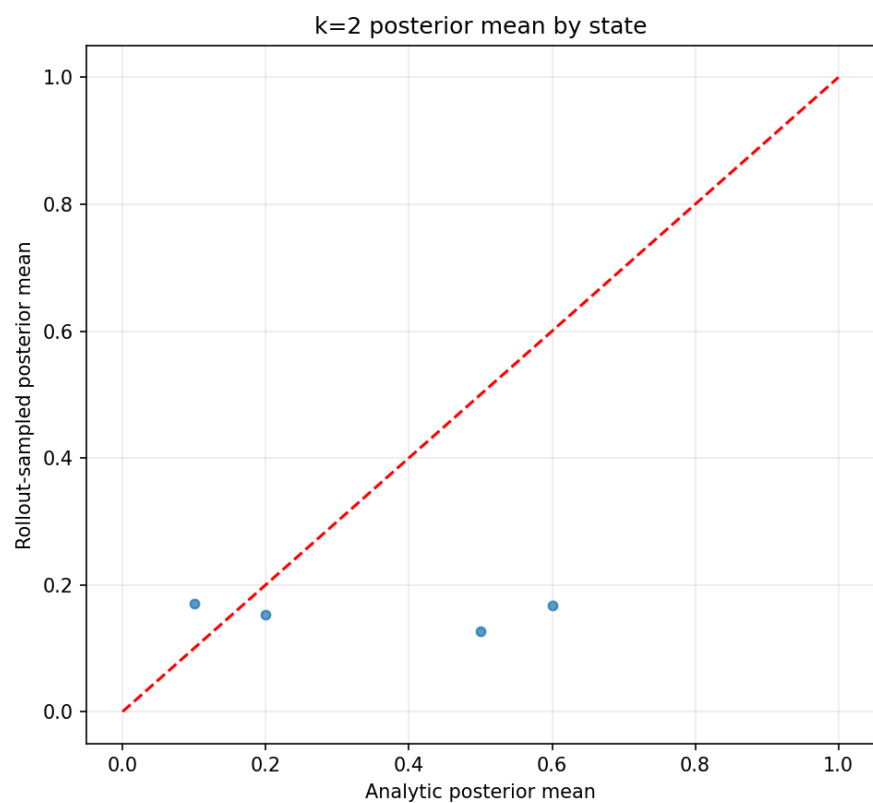


Figure 5: k=2: Step 2: posterior mean comparison

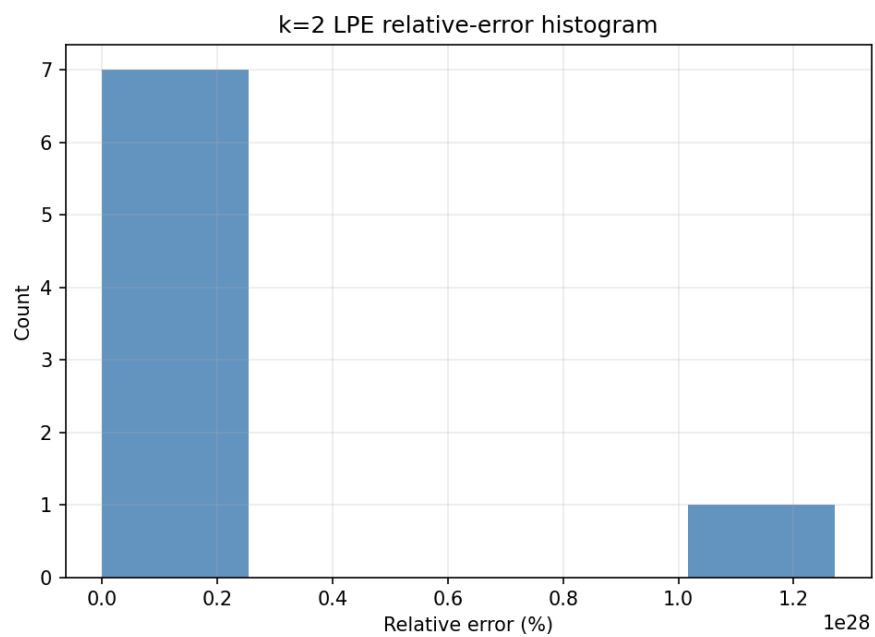


Figure 6: k=2: Step 3: LPE relative-error histogram

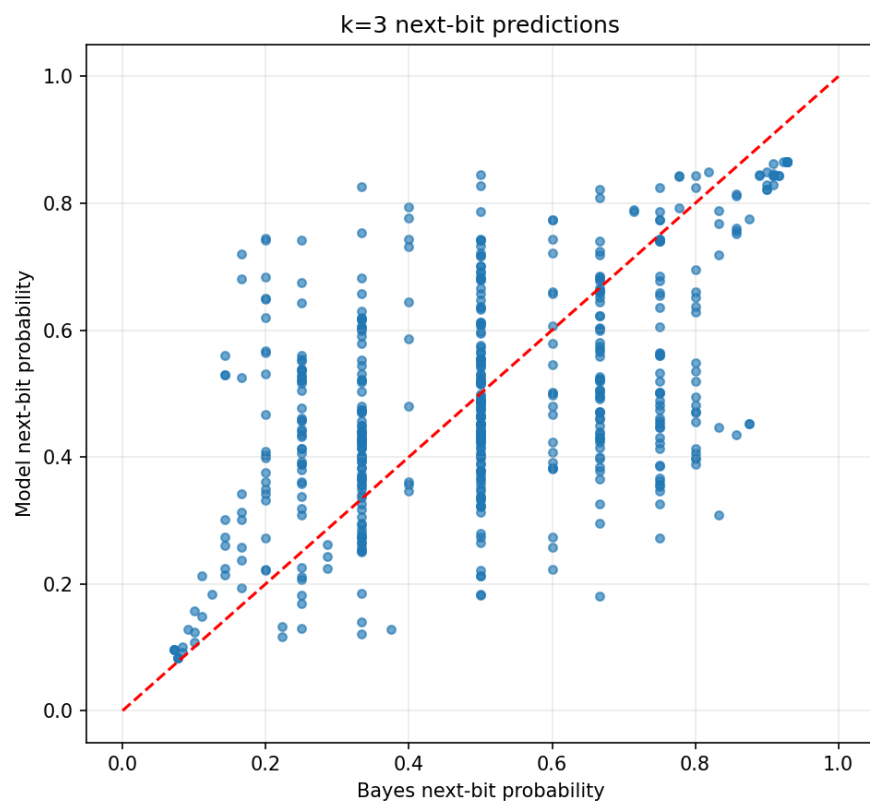


Figure 7: k=3: Step 1: next-bit prediction vs Bayes

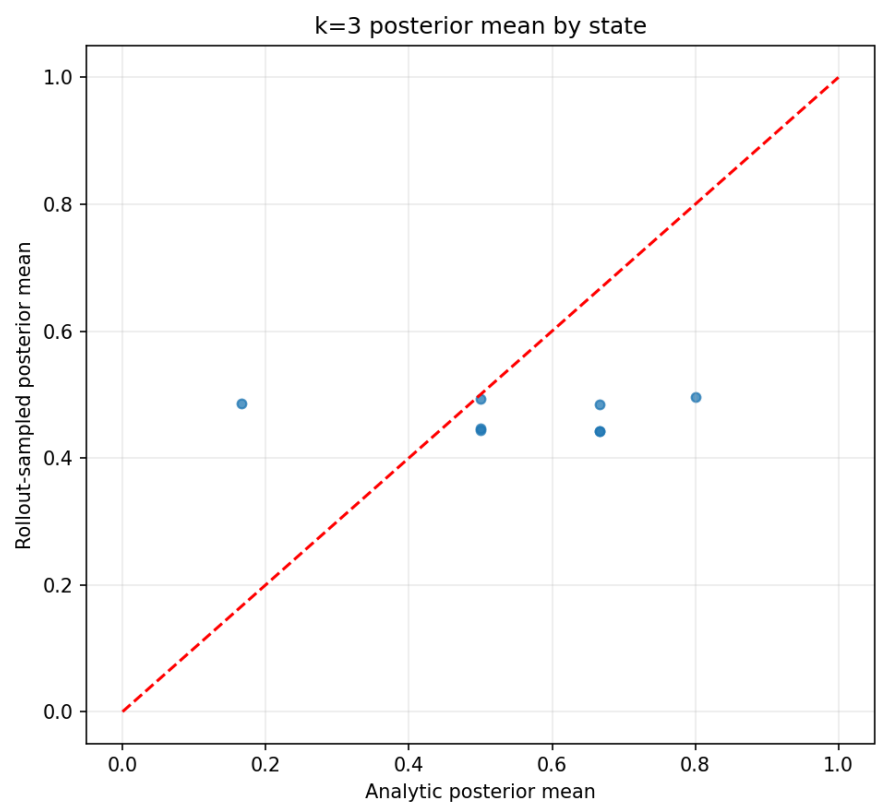


Figure 8: k=3: Step 2: posterior mean comparison

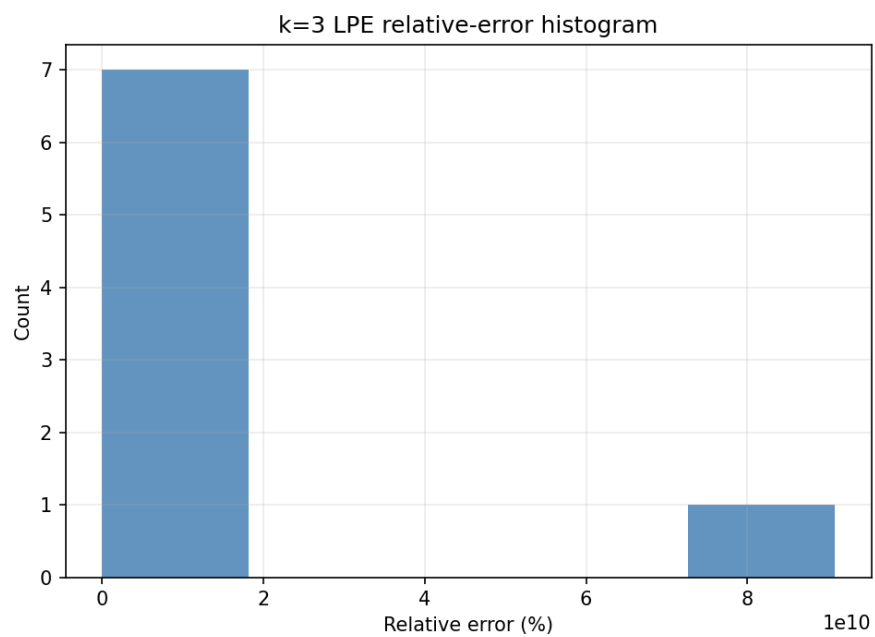


Figure 9: k=3: Step 3: LPE relative-error histogram

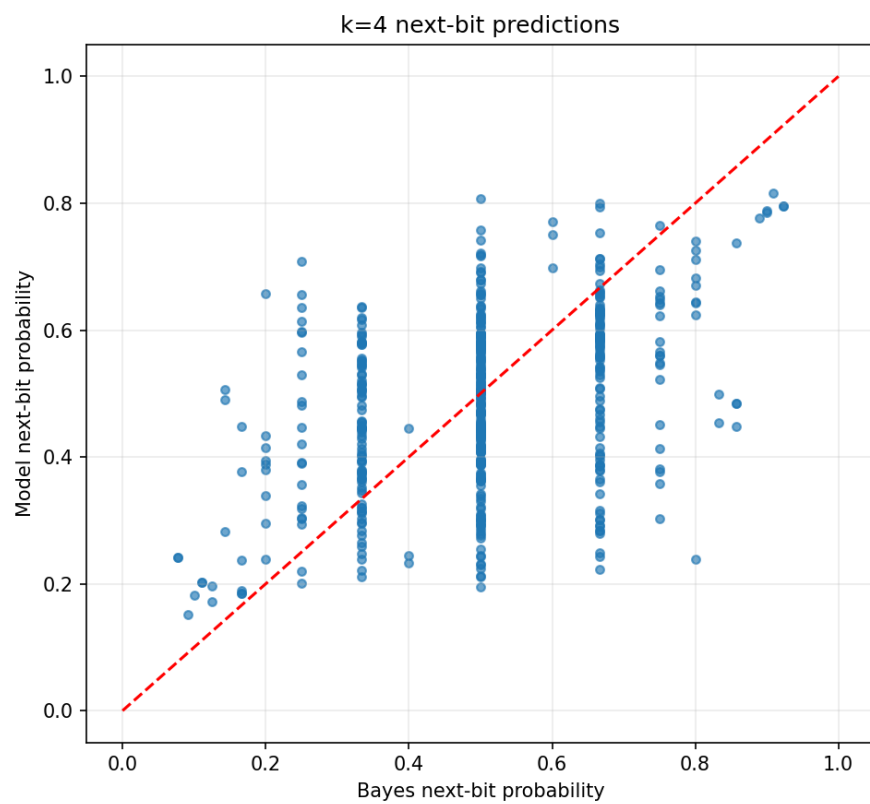


Figure 10: k=4: Step 1: next-bit prediction vs Bayes

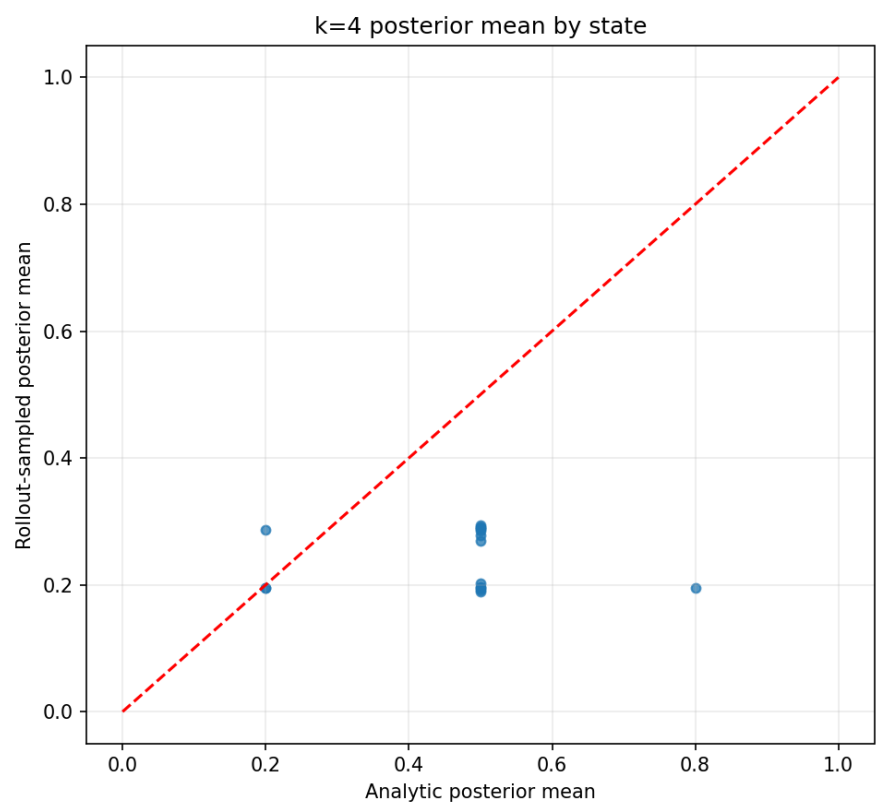


Figure 11: k=4: Step 2: posterior mean comparison

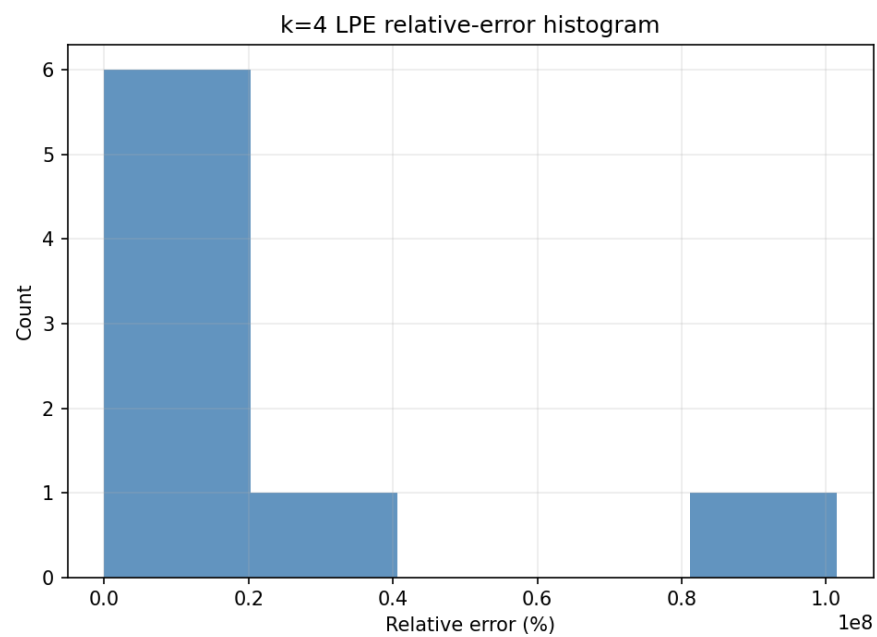


Figure 12: k=4: Step 3: LPE relative-error histogram

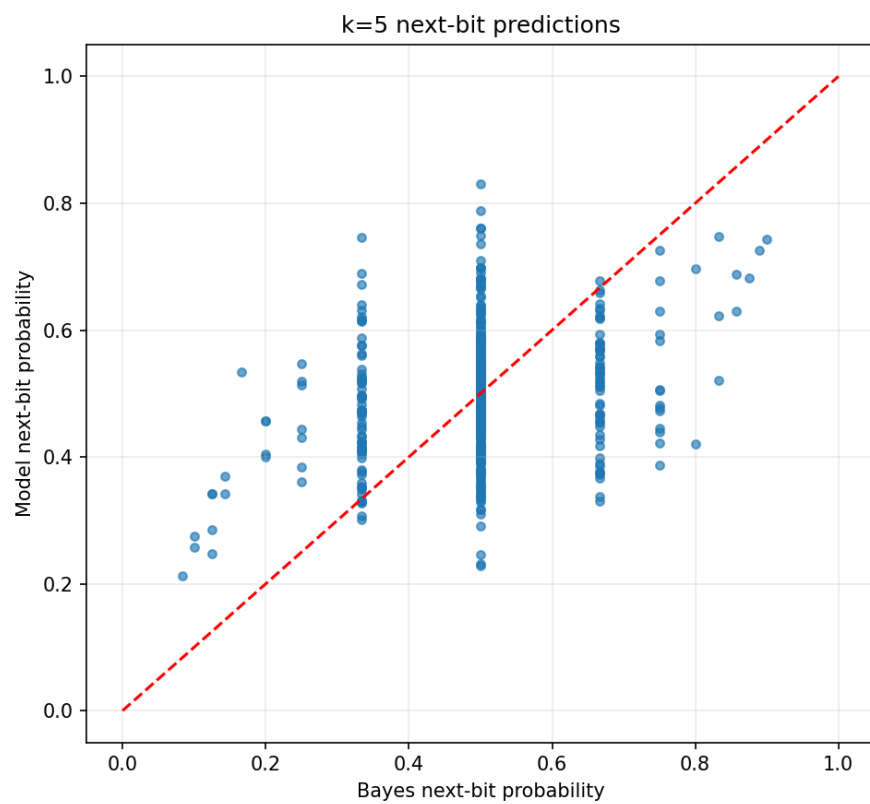


Figure 13: k=5: Step 1: next-bit prediction vs Bayes

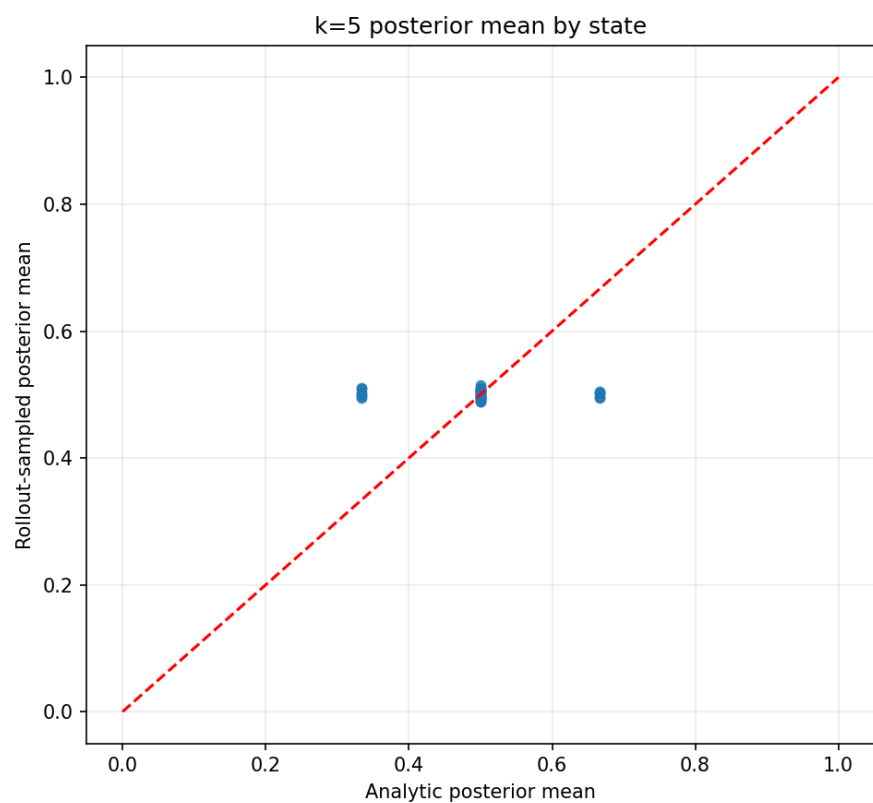


Figure 14: k=5: Step 2: posterior mean comparison

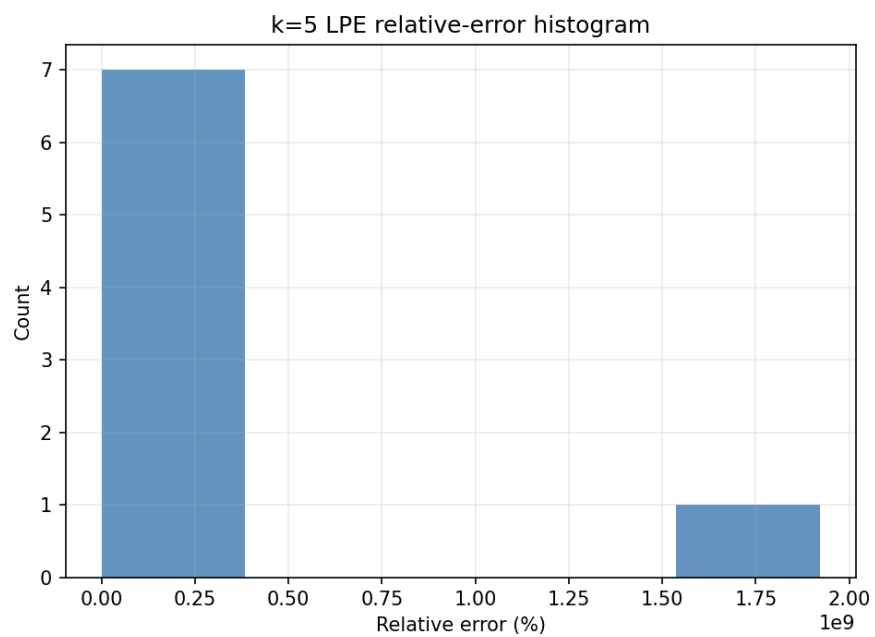


Figure 15: k=5: Step 3: LPE relative-error histogram

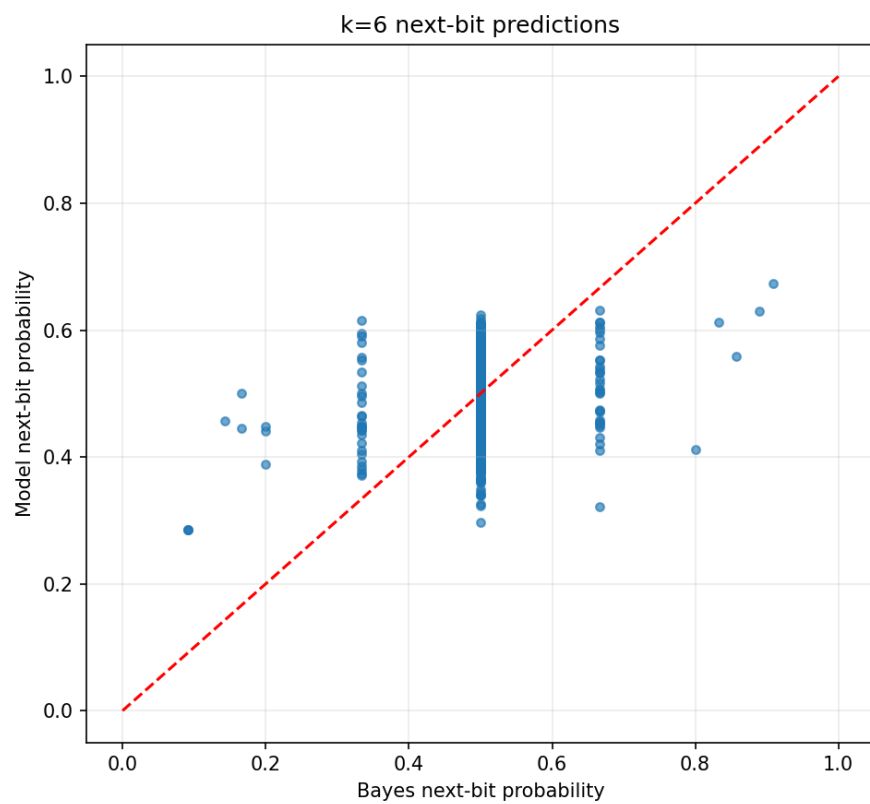


Figure 16: k=6: Step 1: next-bit prediction vs Bayes

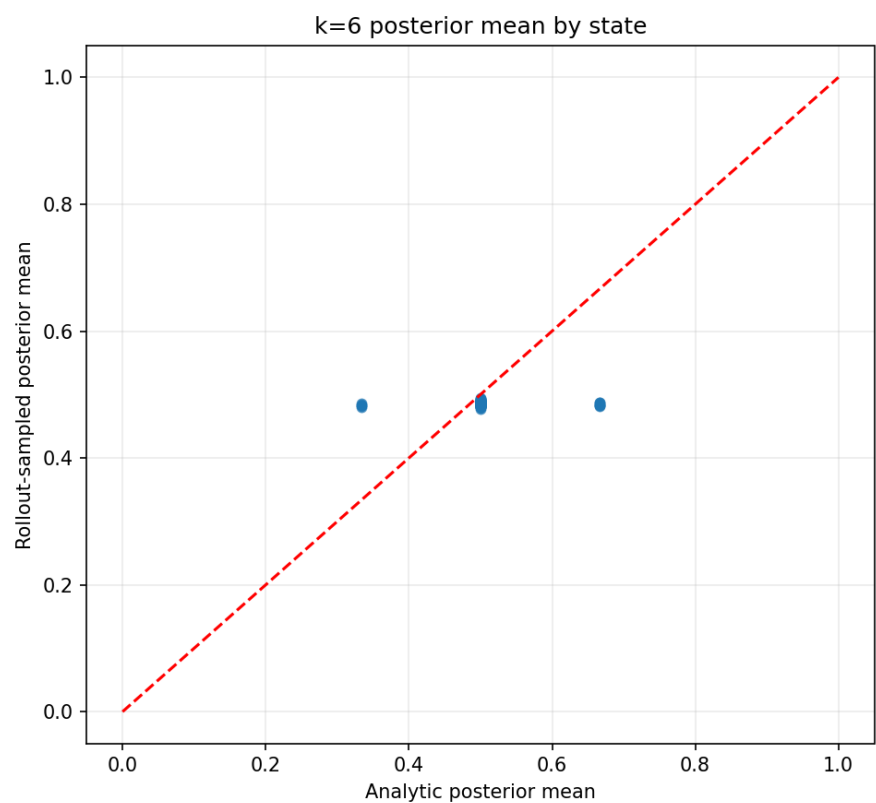


Figure 17: k=6: Step 2: posterior mean comparison

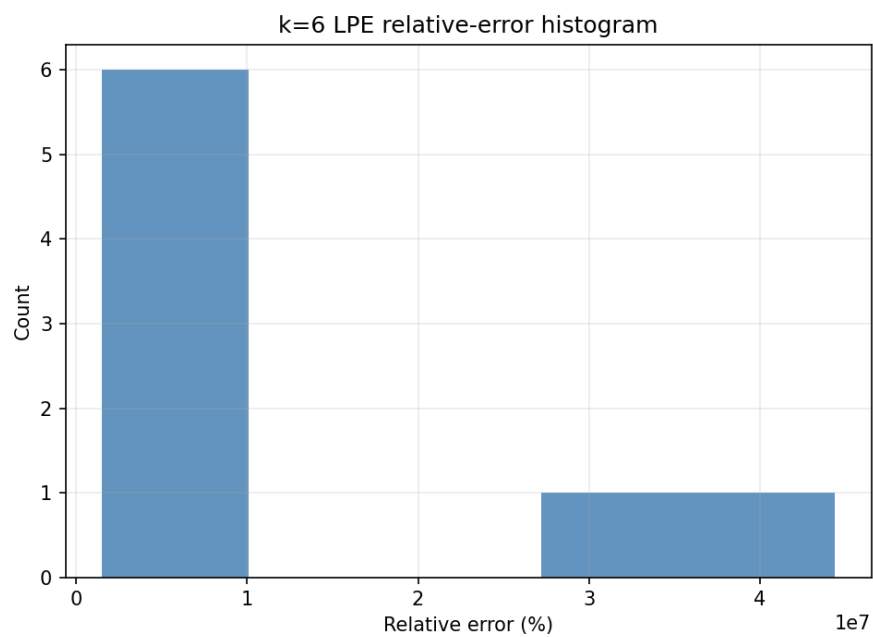


Figure 18: k=6: Step 3: LPE relative-error histogram

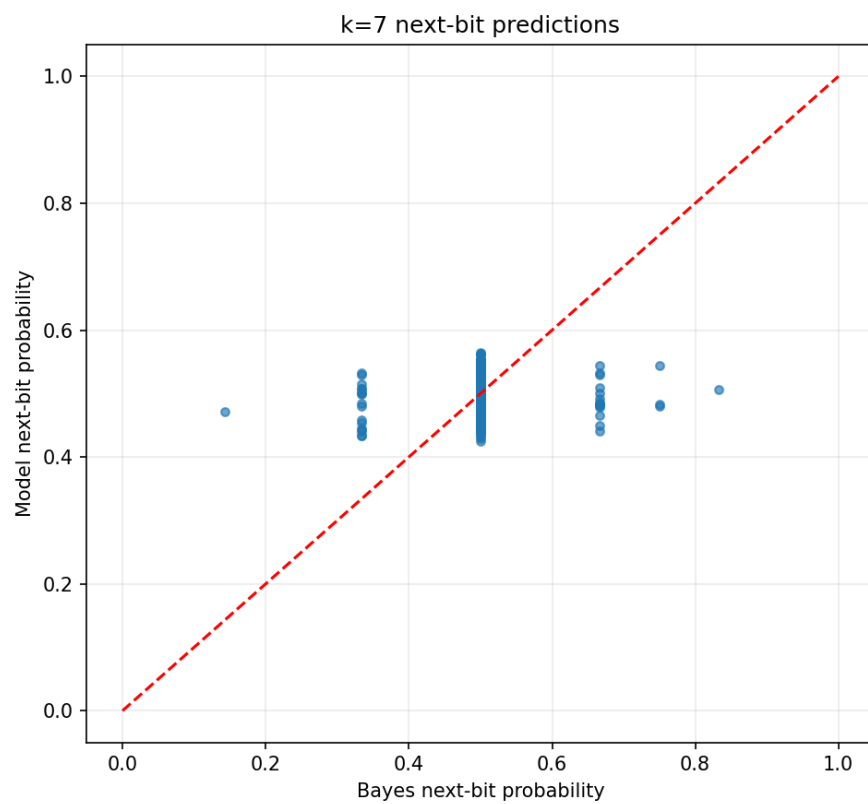


Figure 19: k=7: Step 1: next-bit prediction vs Bayes

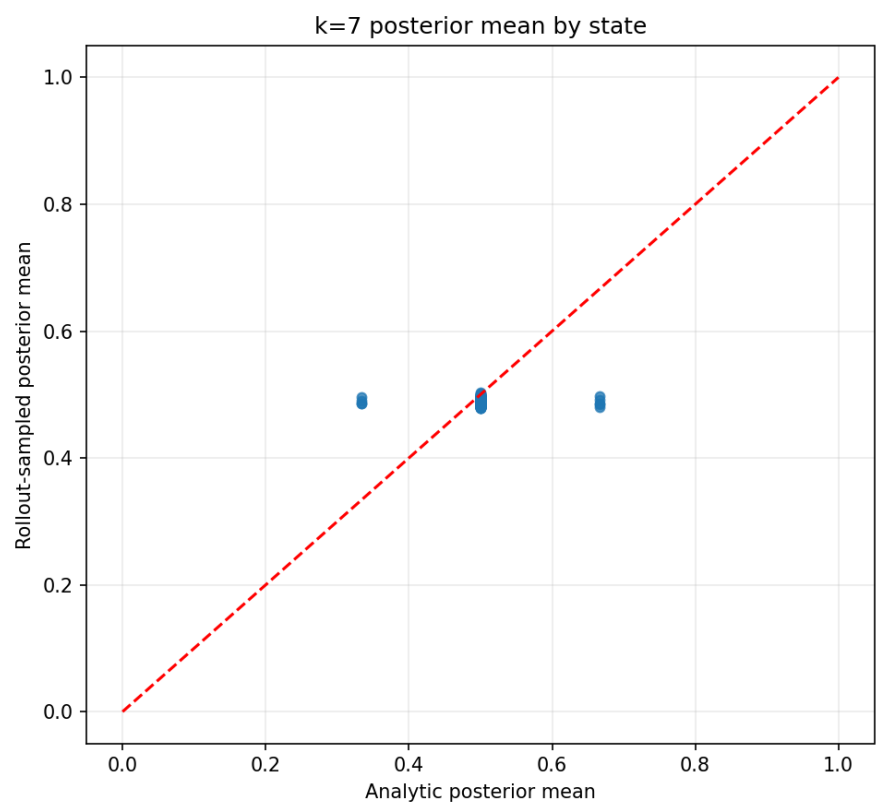


Figure 20: k=7: Step 2: posterior mean comparison

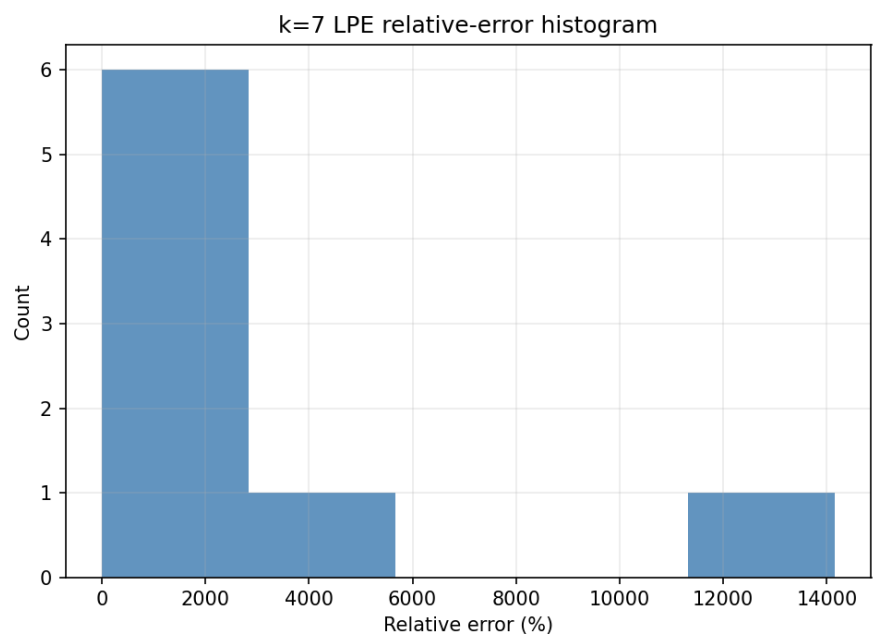


Figure 21: k=7: Step 3: LPE relative-error histogram