

Per-bin accuracy with Wilson 95% CIs
MATH-500, Qwen2.5-1.5B

