

Judge

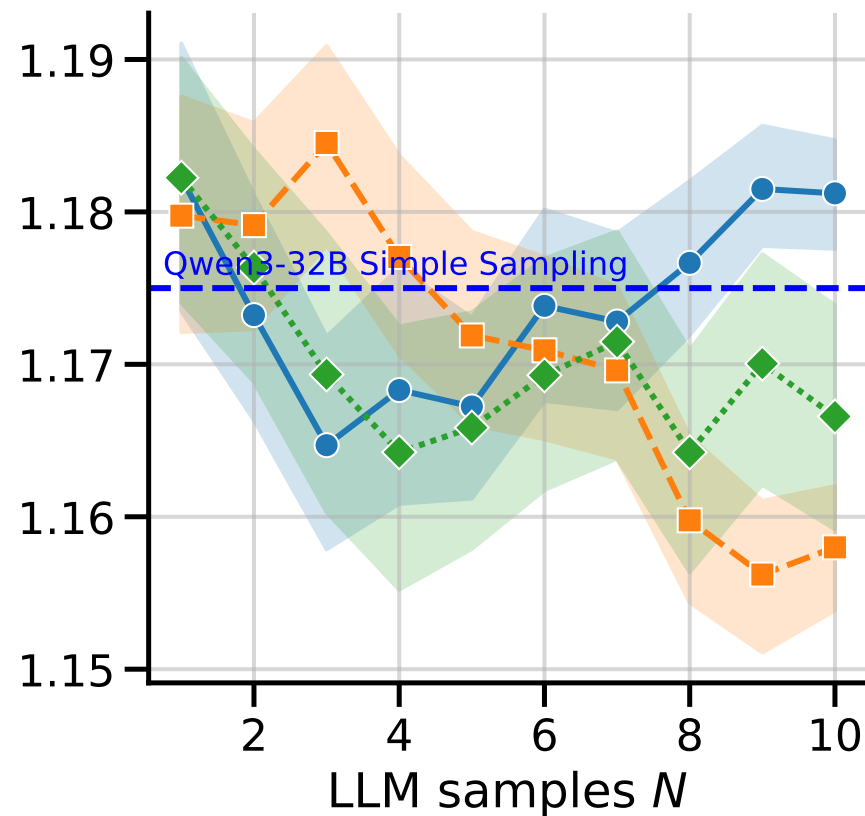
● Qwen3-32B

■ DeepSeek-R1-0528-Qwen3-8B

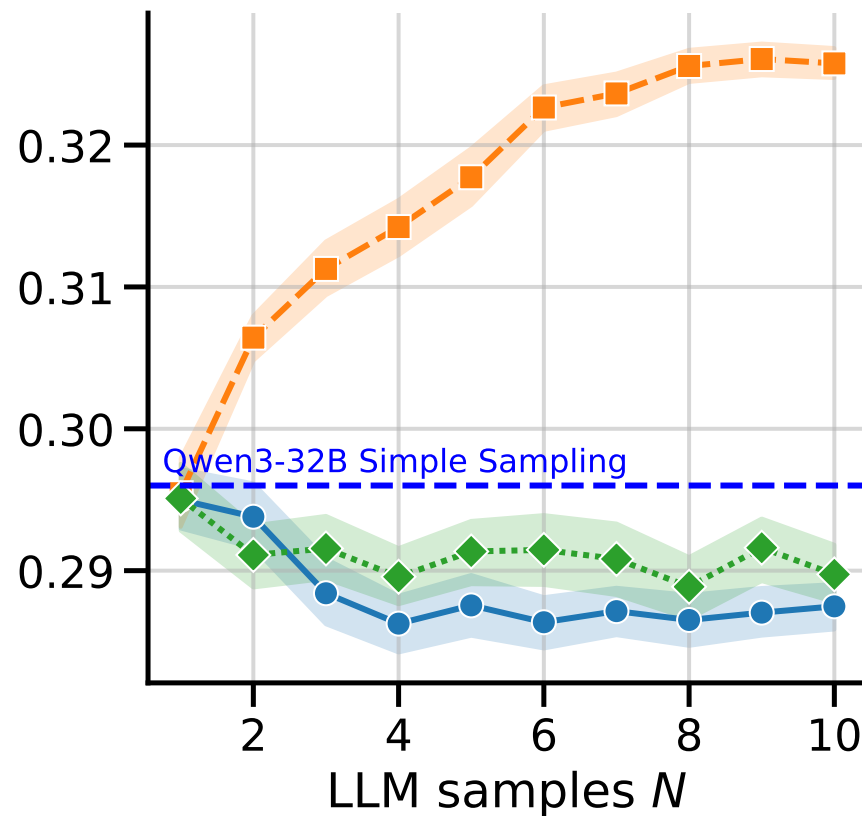
◆ gemini-2.5-flash

Wasserstein Distance

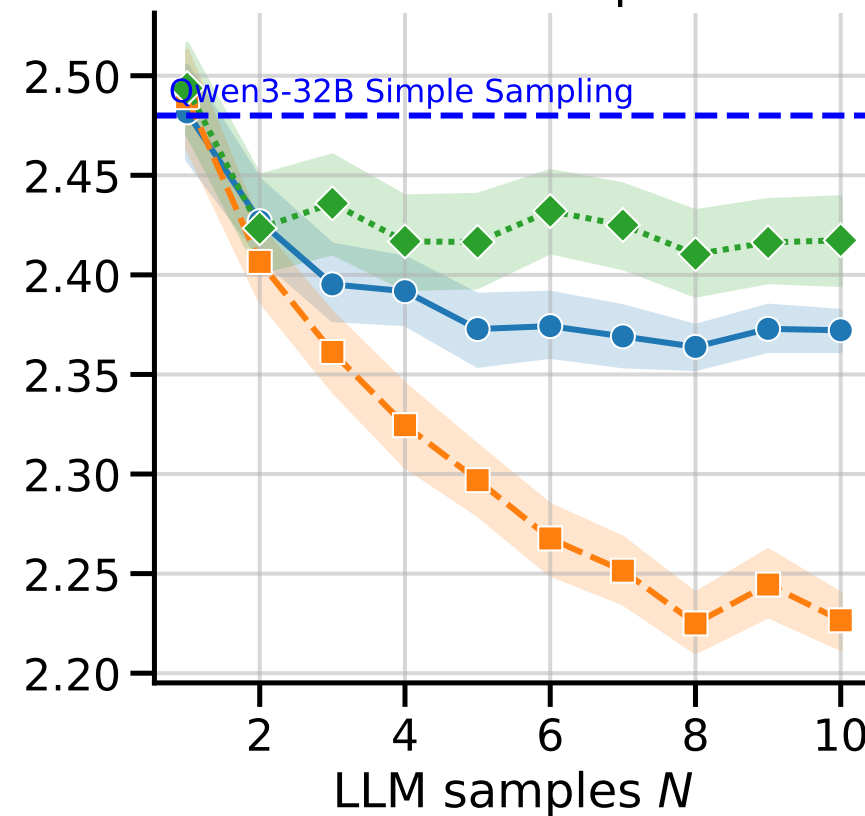
Dataset = CSC



Dataset = MP



Dataset = Paraphrase



Dataset = VariErrNLI

