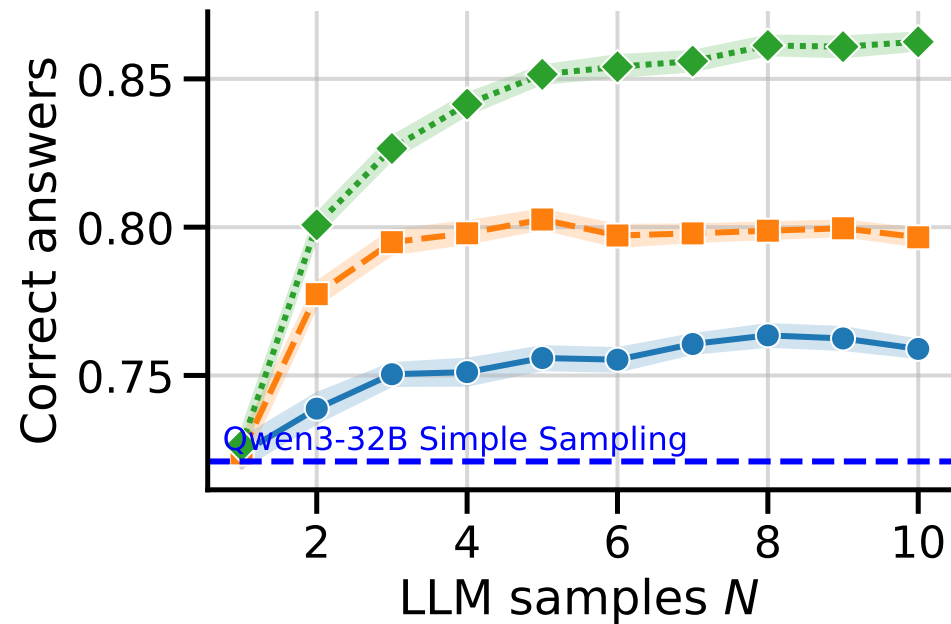


Judge

Qwen3-32B DeepSeek-R1-8B gemini-2.5-flash

Dataset = PRM800K



Dataset = AIME

