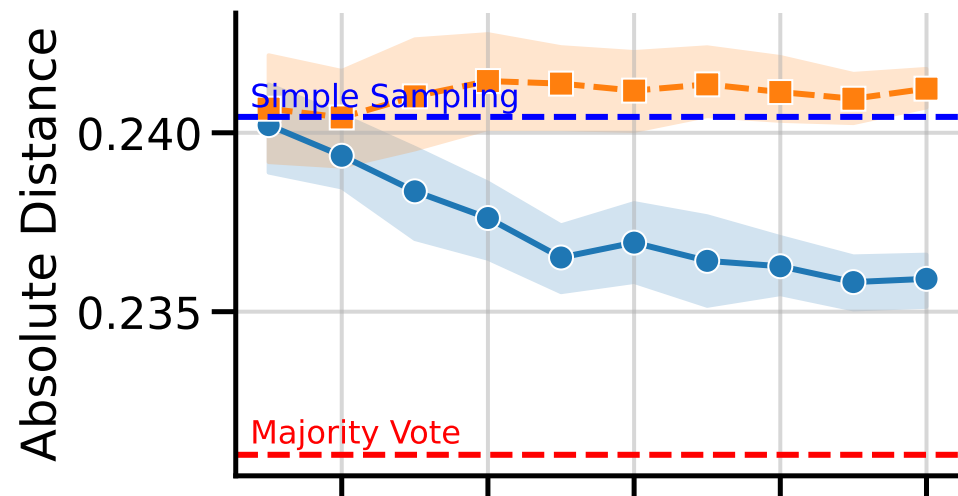


Judge

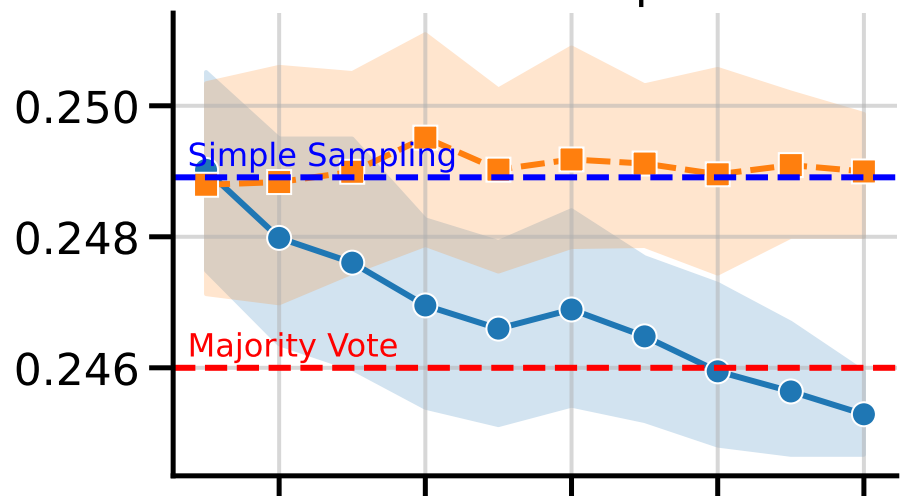
—●— Qwen3-32B

—■— DeepSeek-R1-8B

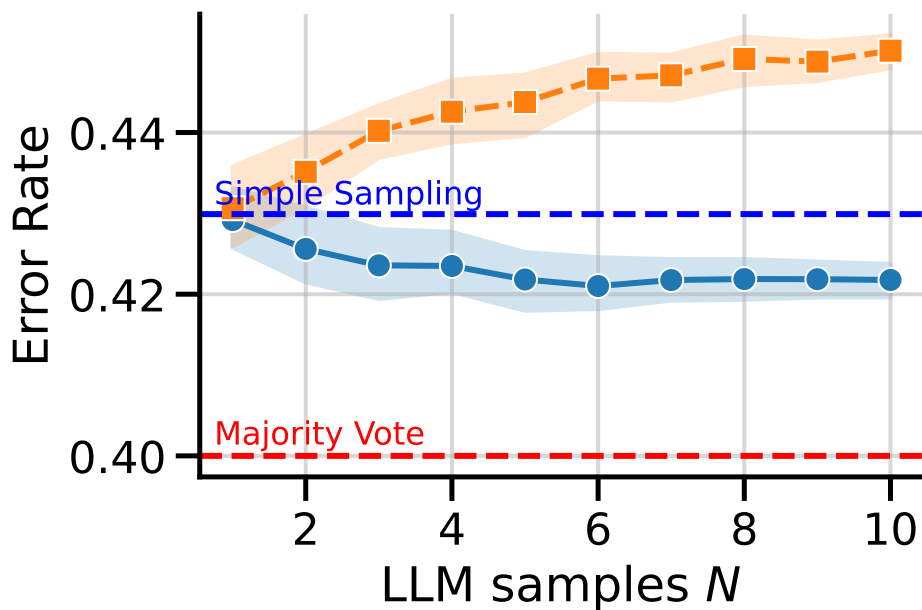
Dataset = CSC



Dataset = Paraphrase



Dataset = MP



Dataset = VariErrNLI

