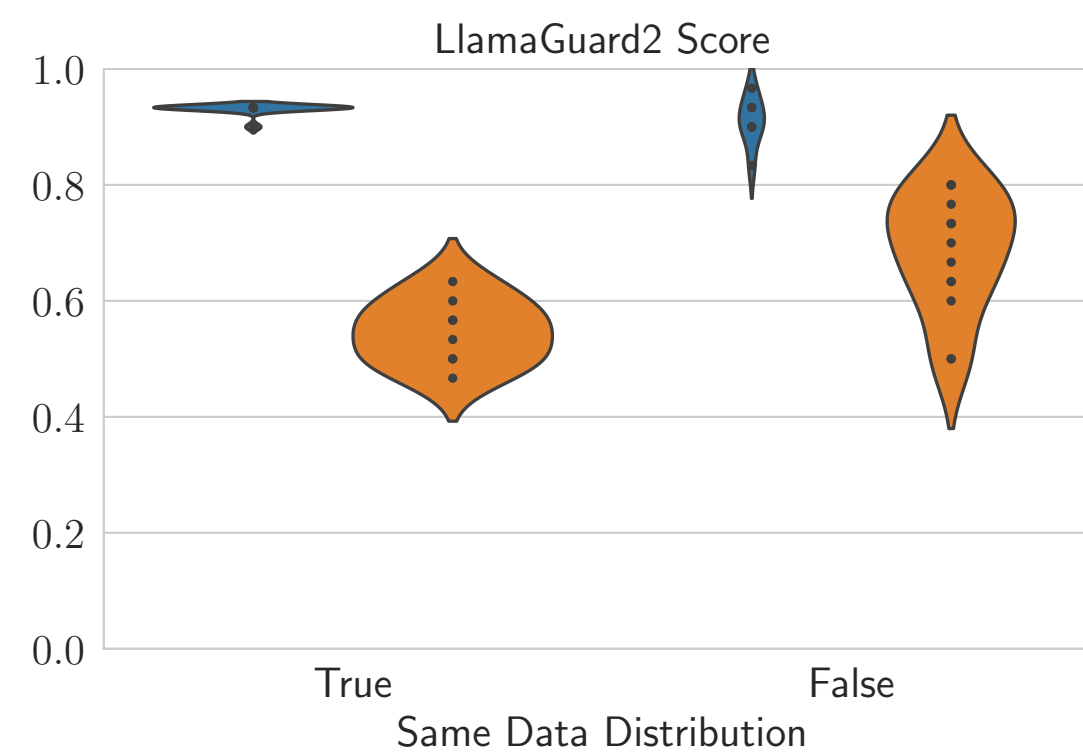
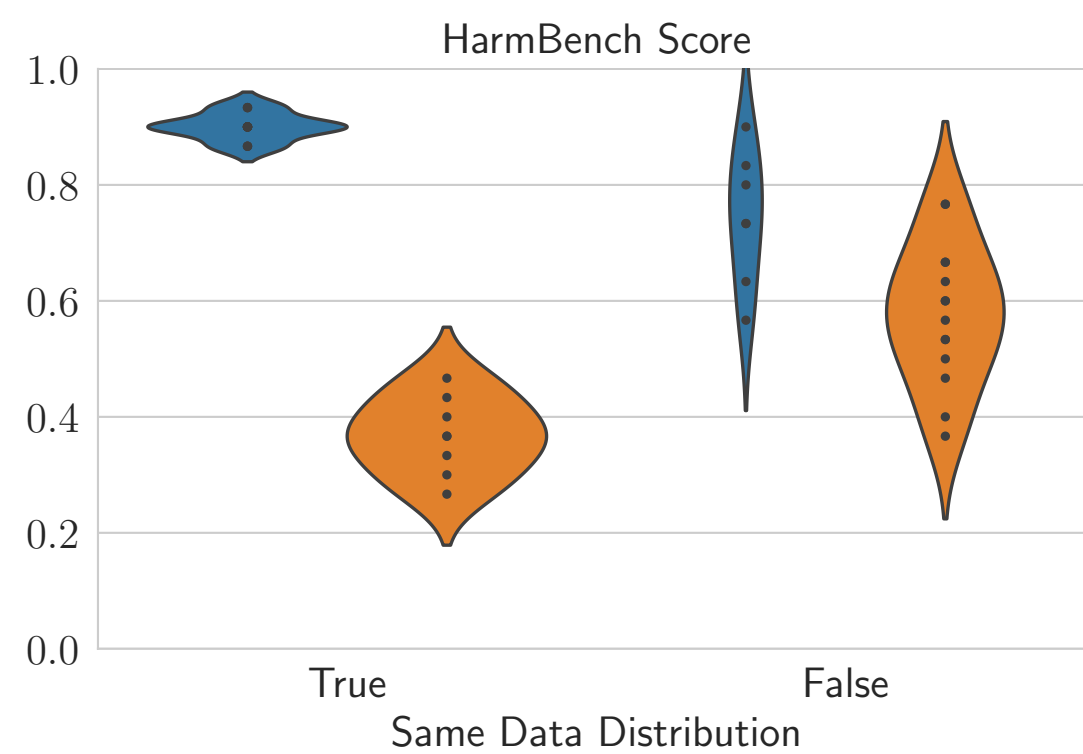
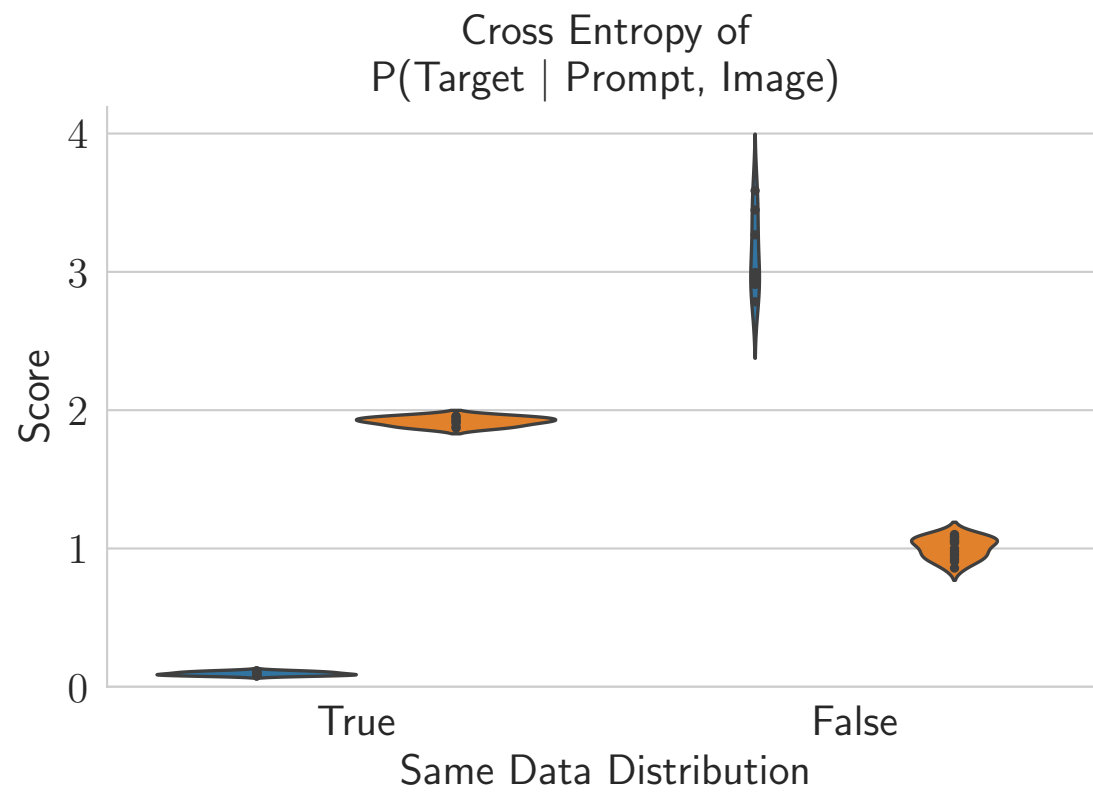


Universality of Image Jailbreaks



Attack Dataset (Train Split)

- advbench
- rylan_anthropic_hhh