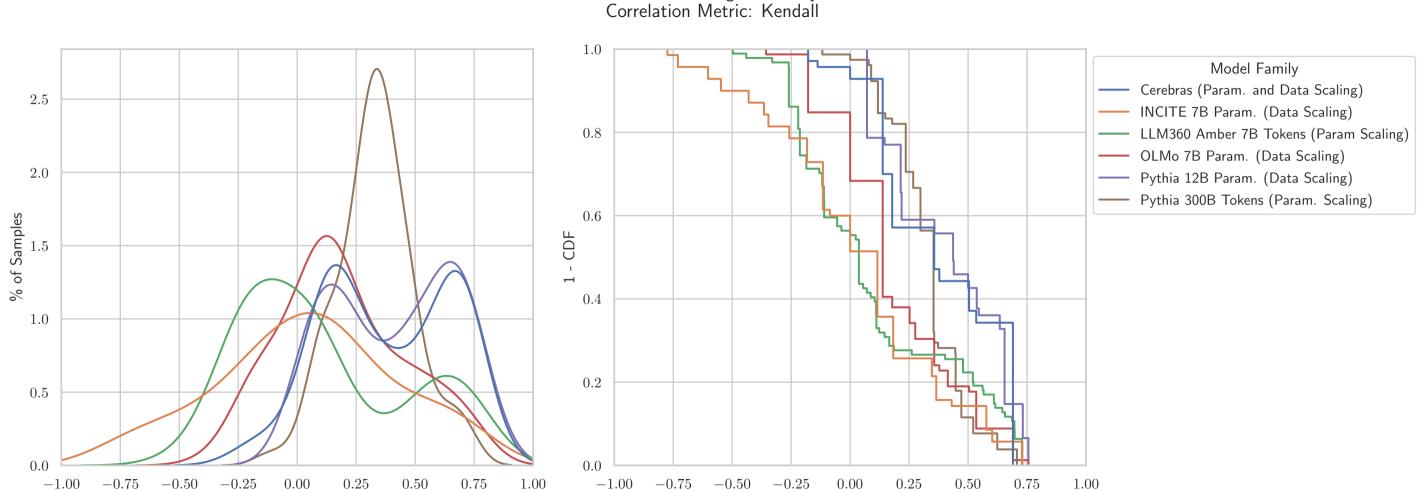
Benchmark: MMLU Jurisprudence Performance Metric: Target Greedily Decoded



Correlation Between FLOPs and Scores (Per Sample)

Correlation Between FLOPs and Scores (Per Sample)