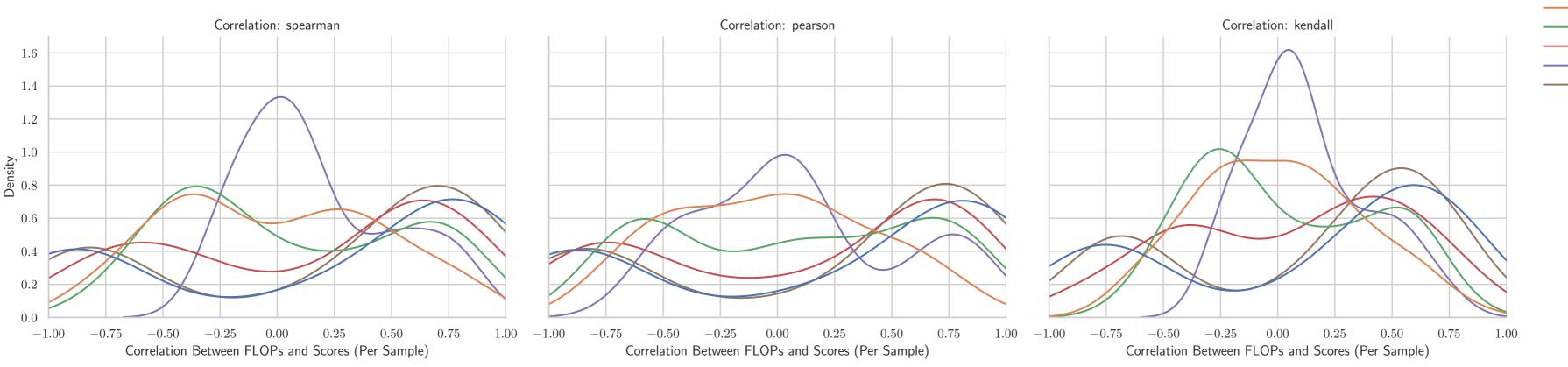
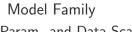
$\begin{array}{c} {\rm MMLU~Machine~Learning} \\ p_{\theta}^{\rm Choices}({\rm Correct~Choice}) \end{array}$





Cerebras (Param. and Data Scaling)INCITE 7B Param. (Data Scaling)

LLM360 Amber 7B Tokens (Param Scaling)

OLMo 7B Param. (Data Scaling)

— Pythia 12B Param. (Data Scaling)

— Pythia 300B Tokens (Param. Scaling)