Distributions for "final_norm" in gpt2

