Distributions for "pre_attention_norm" in gpt2

