

Histograms of attention logits  $\frac{1}{\sqrt{d'}} \langle Qx_i, Kx_j \rangle$   
(uniform\_albert-xlarge-v2\_params-random\_dense-off\_layers-48)

