A nice explanation of Xavier initialization for weights

<http://andyljones.tumblr.com/post/110998971763/an-explanation-of-xavier-initialization>

Paper for this idea - <http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>

In [mathematical statistics](https://en.wikipedia.org/wiki/Mathematical_statistics), the **Kullback–Leibler divergence** (also called **relative entropy**) is a measure of how one [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) is different from a second, reference probability distribution.

