

noise scaling fits

goal: for each combination of metric/method/dataset size, we want to fit the 2 parameters (\bar{u} , I_{\max}) which describe how mutual information scales with UMI count

- the scaling form describes $I(u)$ -- how mutual information scales with the number of UMI -- for a particular method, metric, and with fixed cell number.
- so, for each method and metric, and for each cell number, we should take the I, u pairs (MI value and number of UMI/cell) and fit a curve.
- the form of the curve is

$$I(u) = I_{\max} - \frac{1}{2} \log \frac{1 + u/\bar{u}}{u/\bar{u} + 2^{-2I_{\max}}}$$

- this has two parameters we want to fit: \bar{u} and I_{\max} . We should fit and save these parameters (as well as their confidence intervals) for each method+metric+cell number.
- the two parameters are interpretable: \bar{u} indicates the number of UMIs where information starts to saturate (sort of like a notion of robustness to noise), and I_{\max} tells us how much information we might expect from a perfect measurement. for the paper, we will tabulate these across methods/metrics.