1

scVI

Model

The output of a scRNAseq experiment is a matrix of counts with N rows (the number of cells) and G columns (the number of genes), where each entry x_{ng} is an integer representing how many transcripts of gene g where seen in cell n. scVI is a generative hierarchical Bayesian model for scRNAseq data with conditional distributions parametrized by neural networks for each gene.[1] There are technical variables to account for different batches (s_n) and for library size (l_n, which can be interpreted as cell size or sequencing depth). Thus the number of networks being trained is $2 \cdot G \cdot K$, where K is the total the number of batches (datasets).

[1] Lopez et al., "Deep Generative Modeling for Single-Cell Transcriptomics," *Nature Methods* 15, no. 12 (2018): 1053–58.

Conditional distribution $p\left(x_{ng} \mid z_n, l_n, s_n\right)$ is a zero-inflated negative binomial distribution (ZINB) to model the kinetics of stochastic gene expression with some entries replaced by zeros. It can also be modelled using Negative binomial or Zero-inflated negative binomial using the `gene_likelihood` argument.

The neural networks f_w^g and f_h^g use dropout regularization and batch nomalization to model gene expression while accounting for library sizes and batch effects respectively. Each network typically has 3 fully connected-layers, with 128-256 nodes each. The activation functions are ReLU, exponential, or linear. f_w has a final softmax layer to represent normalized expected frequencies of gene expression as in. Weights for some layers are shared between f_w and f_h.

Inference

Detail the inference objective

Training

Any details that aren't clear in manuscripts but are important for training.

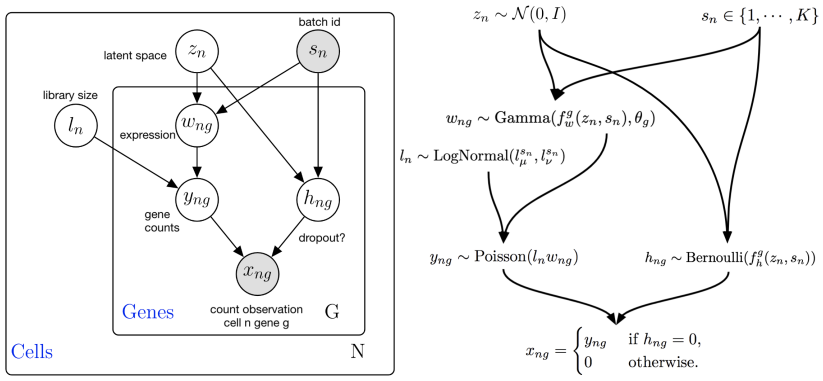

Figure 1: Fancy plot

Tasks

Here we put the mathematical description of tasks.

Math to code

Table for each variable, what it's variable name is in the code

References

Lopez, Romain, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. "Deep Generative Modeling for Single-Cell Transcriptomics." *Nature Methods* 15, no. 12 (2018): 1053–58.