

SCVI-TOOLS MODEL REFERENCE

A handbook containing math and code details

*Eduardo da Veiga Beltrame, Adam Gayoso, Romain Lopez,
...*

Created: Mon May 17 22:09:11 UTC 2021

Contents

<i>Introduction</i>	5
---------------------	---

<i>scVI</i>	7
-------------	---

<i>Model</i>	7
--------------	---

<i>Inference</i>	7
------------------	---

<i>Training</i>	8
-----------------	---

<i>Tasks</i>	8
--------------	---

<i>Math to code</i>	8
---------------------	---

<i>totalVI</i>	9
----------------	---

<i>Model</i>	9
--------------	---

<i>Inference</i>	9
------------------	---

<i>Training</i>	10
-----------------	----

<i>Tasks</i>	10
--------------	----

<i>Math to code</i>	10
---------------------	----

<i>A short manual</i>	11
-----------------------	----

<i>Numbered and unnumbered margin notes</i>	11
---	----

<i>Math formulas</i>	11
----------------------	----

<i>Custom environments</i>	12
----------------------------	----

<i>Citations</i>	12
------------------	----

<i>Figures and images</i>	12
---------------------------	----

<i>Tables</i>	13
---------------	----

<i>Cross references</i>	13
-------------------------	----

Bibliography 15

1

Introduction

This document uses the following conventions when depicting graphical models:

- Shaded vertices represent observed random variables.
- Half-shaded vertices represent variables that can be either observed or random. Empty vertices represent latent random variables.
- Edges signify conditional dependency.
- Rectangles (“plates”) represent independent replication.

For an overview graphical models see Graphical models. M. I. Jordan. *Statistical Science (Special Issue on Bayesian Statistics)*, 19, 140-155, 2004. [PDF](#).

For a review of the probability distributions used by the scvi-tools models, see: F. William Townes *Review of Probability Distributions for Modeling Count Data* (2020).¹

¹ Townes, “Review of Probability Distributions for Modeling Count Data,” 2020, <http://arxiv.org/abs/2001.04343>.

2

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 were seen in cell n . scVI is a generative hierarchical Bayesian model for scRNAseq data with conditional distributions parametrized by neural networks for each gene.² 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 number of batches (datasets).

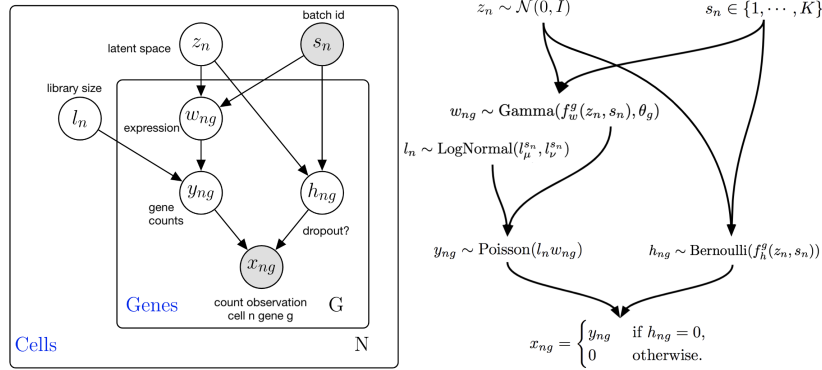
Conditional distribution $p(x_{ng} | z_n, l_n, s_n)$ is a zero-inflated negative binomial distribution (ZINB) to model the kinetics of stochastic gene expression with some entries replaced by zeros. It can also be modelled using Negative binomial or Zero-inflated negative binomial using the `gene_likelihood`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

The neural networks f_w^g and f_h^g use dropout regularization and batch normalization 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 .

² Lopez et al., "Deep Generative Modeling for Single-Cell Transcriptomics," *Nature Methods* 15, no. 12 (2018): 1053–58.

Inference

Detail the inference objective



Training

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

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

3

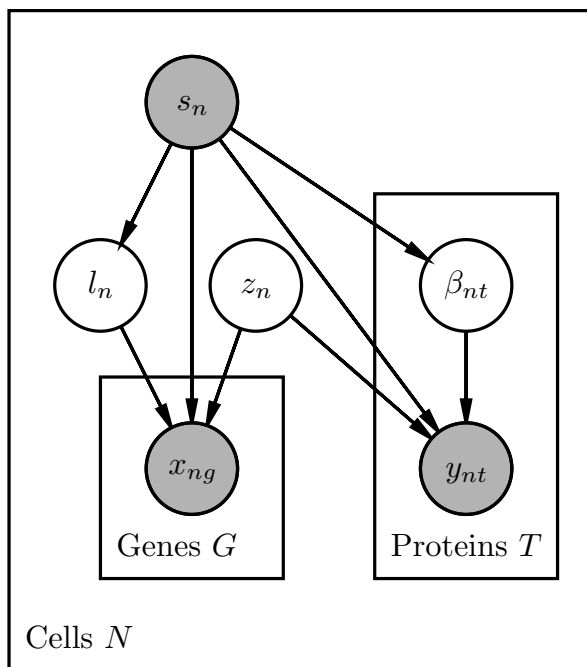
totalVI

Model

Details³

³ Gayoso et al., "Joint Probabilistic Modeling of Single-Cell Multi-Omic Data with totalVI," *Nature Methods*, 2021.

Figure 2: Fancy plot



Inference

Detail the inference objective

Training

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

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

4

A short manual

Below are several features available with the Unbuch setup. Unbuch uses the pandoc markdown syntax. Please see the pandoc documentation for details on that.

Numbered and unnumbered margin notes

To create a side note, use this syntax:⁴

To create a side note, use this syntax:[^sidenote1]

[^sidenote1]:

Aenean id risus at nibh fermentum suscipit vel
risus. Fusce ornare, ante eget placerat congue.

To create an unnumbered margin notes, prefix the text with {-}.

To create an unnumbered margin notes,
prefix the text with ‘{-}’.[^unnumbered-marginnote]

[^unnumbered-marginnote]:

{-} Aenean id risus at nibh fermentum suscipit vel
risus. Fusce ornare, ante eget placerat congue.

⁴ Aenean id risus at nibh fermentum
suscipit vel risus. Fusce ornare, ante
eget placerat congue.

Aenean id risus at nibh fermentum
suscipit vel risus. Fusce ornare, ante
eget placerat congue.

Math formulas

You can use inline and display math environments. These are handled by pdf_latex when compiling to pdf, and they’re handled by katex when displaying html. Use a single \$ for inline math, and a double \$\$ for display.

$$\langle x, y \rangle \leq \|x\| \cdot \|y\|$$

You can define macros using standard latex syntax in templates/shared-macros.tex. This will render in both pdf and html. Here is the use of some such macros:

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}[X = x]$$

Custom environments

Unbuch supports custom numbered and unnumbered environments.

Definition 1. *We call a thing a thing.*

Here is an unnumbered environment.

Example. *Something else.*

For this to work property, the environment Definition must be defined as a latex macro. For html these are automatically converted to div classes that can be styled via css.

Citations

Citations go in the margin where they are cited and also appear at the end of the document.⁵

The syntax for citations is `[@hegel]`. We can also cite multiple papers at once..⁶ Simply separate multiple citations with a semi-colon: `[@hegel; @nietzsche]`. You see that repeated citations are abbreviated to save space.

⁵ Hegel, *Phänomenologie Des Geistes* (Jospeh Anton Goebhardt, 1807).

⁶ Hegel; Nietzsche, *Götzen-dämmerung* (Verlag von C. G. Naumann, 1889).

Figures and images

Ideally, create images in both SVG and PDF format. Include them by their `.svg` extension. The compiler will figure out to substitute the pdf version for the pdf compilation. If vector graphics are not available use JPG or PNG. Those are the only other two formats that are supported.

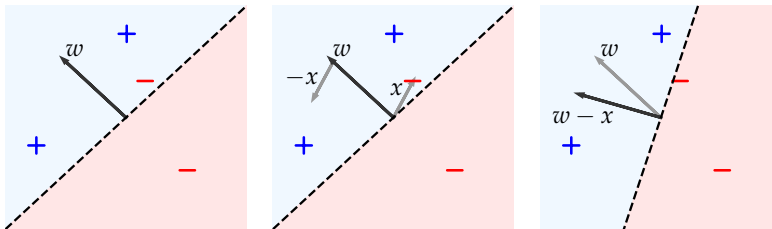


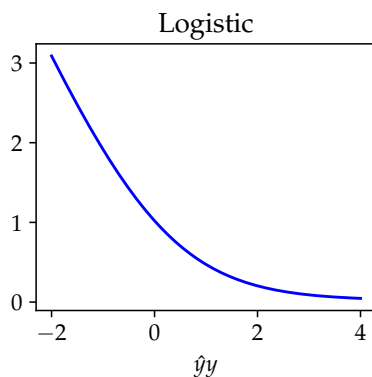
Figure 3: Fancy plot

This image was included as:

`![Fancy plot](assets/perceptron.svg)`

Figures and images can go in the margin.⁷

You can specify the width of a figure by appending something like `{width="50%"}`.



This image was included as:

`![Half width plot](assets/logistic_loss.svg){width="50%"}`

Tables

Tables work just as they would in pandoc markdown. Nothing is different here.

Table 1: A table of something

A	B	C
1	0	1
1	0	0
0	0	1
1	0	1

Cross references

Unbuch offers a minimalistic syntax for cross references. These aren't full-fledged latex cross references and don't work the same way. You can introduce a cross reference at any point by writing `@{counter:name}`. This will assign a number. You can retrieve this number by writing `@{counter:name}` again later on. You can increment the counter by using the same counter multiple times, e.g., by

⁷ Here's an image in the margin.

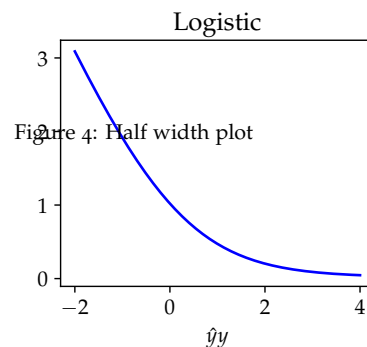


Figure 4: Half width plot

writing `@{theorem:abc}` and `@{theorem:planar}`. Here's an example:

This markdown snippet will turn out as follows:

Theorem 1. *Curabitur at vestibulum velit. Ut ac turpis purus.*

Theorem 2. $1 + 1 = 2$

We saw in Theorem 2 that $1 + 1 = 2$.

This also works with equations.

$$1 + 1 = 2 \quad (1)$$

Bibliography

- Gayoso, Adam, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L Nazer, Aaron Streets, and Nir Yosef. "Joint Probabilistic Modeling of Single-Cell Multi-Omic Data with totalVI." *Nature Methods*, 2021.
- Hegel, Georg Wilhelm Friedrich. *Phänomenologie Des Geistes*. Josph Anton Goebhardt, 1807.
- 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.
- Nietzsche, Friedrich. *Götzen-dämmerung*. Verlag von C. G. Naumann, 1889.
- Townes, F. William. "Review of Probability Distributions for Modeling Count Data," 2020. <http://arxiv.org/abs/2001.04343>.

List of Figures

1	Fancy plot	8
2	Fancy plot	9
3	Fancy plot	12
4	Half width plot	13

List of Tables

1 A table of something 13