IMPERIAL

Inference methods for studying single cells

a tutorial lecture

Philipp Thomas
June 2024

Resources

Tutorial 1 Machine learning-based inference

Tutorial 2 Approximate Bayesian Computation

Tutorial 3 Model selection

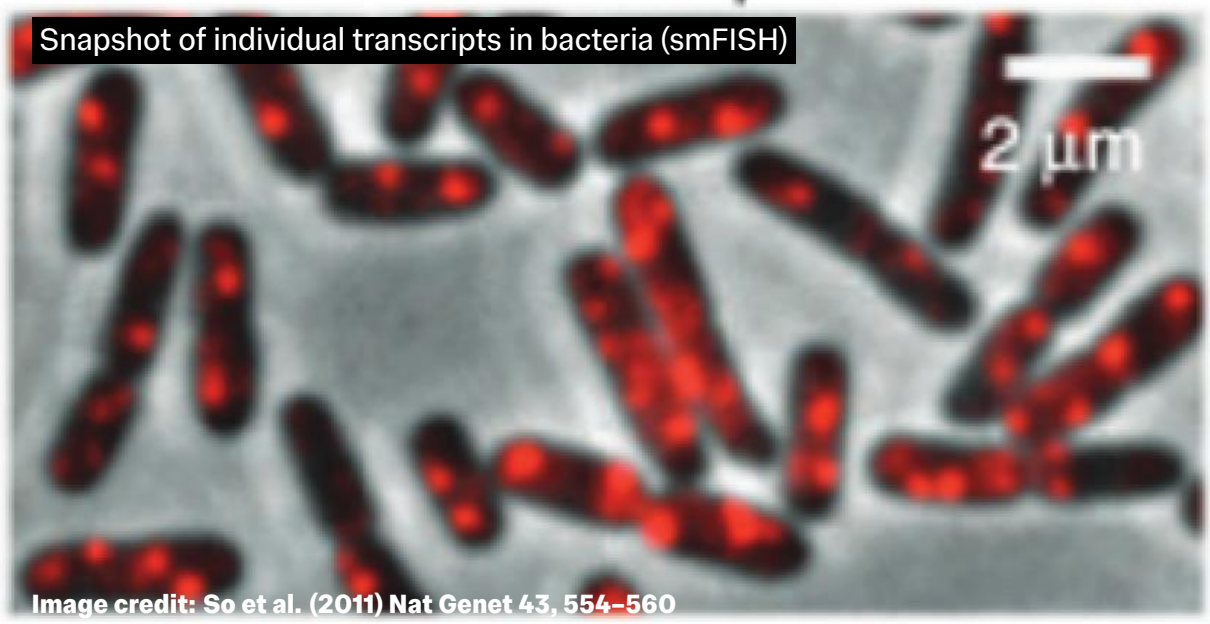
Can be found at:

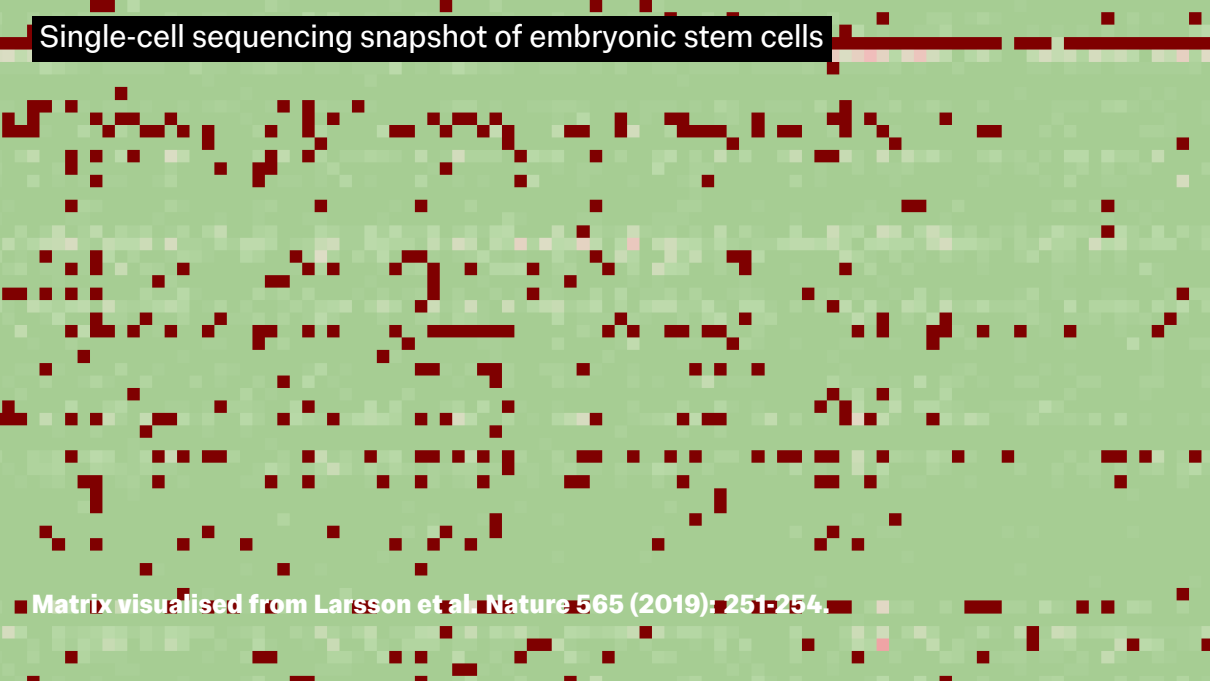
https://shorturl.at/hDps9

Python requirements:
numpy, scipy, matplotlib, pandas, scikit-learn

Single-cell data


Timelapse data of bacteria expressing green fluorescent protein
Image credit: Hasty lab, UC San Diego.


Snapshot of individual transcripts in bacteria (smFISH)
2 µm
Image credit: So et al. (2011) Nat Genet 43, 554–560


Single-cell sequencing snapshot of embryonic stem cells
Matrix visualised from Larsson et al. Nature 565 (2019): 251-254.

Single-cell data

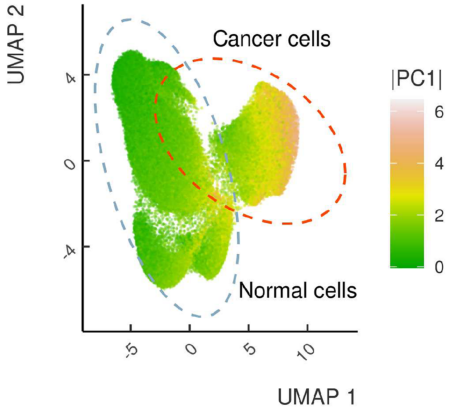
Focus: analysis of snapshot data

high-throughput
transcript numbers for thousands of genes in thousands of cells

beyond average
harness cell-to-cell variation (distributions) to understand processes inside individual cells

mathematical models
capture distributions to make sense of data and biological mechanisms

inference
parameterise models, which allows testing hypotheses with data



Agenda

01 Data, modelling and likelihood
02 Maximum likelihood estimation
03 Moment-based inference
04 Machine learning-assisted inference
05 Bayesian inference
06 Approximate Bayesian Computation
07 Model selection
08 Applications to real data

Galileo Galilei: “Measure what can be measured, and make measurable what cannot be measured.”

Modelling gene expression

Modelling gene expression

Chemical Master Equation

Biochemical kinetics of a general reaction:

$$\nu_{1,\mathsf{r}}^{-}\mathsf{S}_1 + \nu_{2,\mathsf{r}}^{-}\mathsf{S}_2 + \ldots \xrightarrow{\theta_\mathsf{r}} \nu_{1,\mathsf{r}}^{-}\mathsf{S}_1 + \nu_{2,\mathsf{r}}^{-}\mathsf{S}_2 + \ldots$$

Master equation:

$$\frac{\mathsf{dp}}{\mathsf{dt}}(\mathsf{m}|\theta) = \sum_{\mathsf{r}} \left[\mathsf{w}_\mathsf{r}(\mathsf{m} - \nu_\mathsf{r}, \theta_\mathsf{r})\mathsf{p}(\mathsf{m} - \nu_\mathsf{r}|\theta) - \mathsf{w}_\mathsf{r}(\mathsf{m}, \theta_\mathsf{r})\mathsf{p}(\mathsf{m}|\theta)\right],$$

where $\nu_\mathsf{r} = \nu_\mathsf{r}^{+} - \nu_\mathsf{r}^{-}$ is the net-change of an reaction.

Infinite system of coupled ordinary differential equations:
cannot generally be solved.

Modelling gene expression

Poisson model

Biochemical kinetics of transcription and degradation:

$$\varnothing \xrightarrow{k_1} M, \quad M \xrightarrow{k_2} \varnothing$$

Master equation:

$$\dot{p}(m) = k_1 p(m-1) - k_1 p(m) + k_2 (m+1) p(m+1) - k_2 m p(m),$$

In steady state:

$$\dot{p}(m) = 0 \implies 0 = k_1 p(m-1) - k_2 m p(m)$$

Poisson distribution:

$$P(m) = p_0 \times \mu \times \mu/2 \times \mu/3 \times \ldots = p_0 \frac{\mu^m}{m!},$$

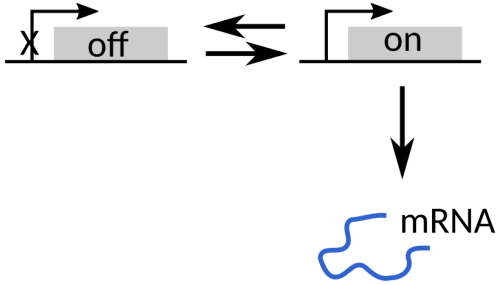
where $\mu = \frac{k_1}{k_2}$ and $p_0 = e^{-\mu}$.

Modelling gene expression

Telegraph model

Biochemical kinetics of transcription and degradation:

$$G_{off} \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} G_{on}, \quad G_{on} \xrightarrow{k_{tx}} G_{on} + M, \quad M \xrightarrow{k_{dm}} \varnothing$$



Solved by Beta-Poisson distribution ($k_2 = 1$):

$$m|p \sim \text{Poisson}(pk_{tx}), \quad p \sim \text{Beta}(k_{on}, k_{off})$$

Determinants of gene expression noise

- **burst frequency** $a = k_{on}/k_{dm}$
- **burst size** $b = k_{tx}/k_{off}$
- **switching frequency** $s = k_{on}/k_{off}$

Modelling gene expression

What's the likelihood?

Definition: Assuming independent data points, the probability that the data X was generated under the model θ is:

$$\mathcal{L}(\theta|\mathsf{X}) = \mathsf{p}(\mathsf{x}_1|\theta) \times \mathsf{p}(\mathsf{x}_2|\theta) \times \ldots \times \mathsf{p}(\mathsf{x}_\mathsf{N}|\theta) = \prod_{\mathsf{i}=1}^{\mathsf{N}} \mathsf{p}(\mathsf{x}_\mathsf{i}|\theta)$$

For example, we have two parameters θ_1 and θ_2 then if

$$\mathcal{L}(\theta_1|\mathsf{X}) > \mathcal{L}(\theta_2|\mathsf{X})$$

then θ_1 fits the data "better".

- Likelihood includes effects of
  - sampling noise (N), i.e., measurement could also arise by chance
  - biological noise (p), i.e., variability across
- Likelihood measures **goodness of fit** rather than the probability that θ is the truth.

Maximum likelihood principle

Maximum likelihood principle

Negative log-likelihood acts as an error function:

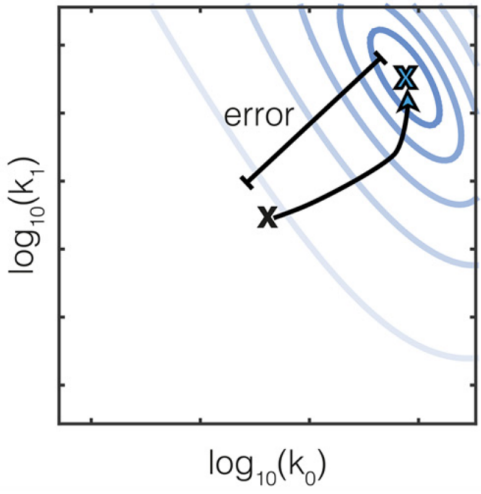
$$\widehat{\mathrm{err}}\,(\theta\,;\mathsf{x}) = -\ln \mathcal{L}(\theta|\mathsf{X}) = -\sum_{\mathsf{i}=1}^{\mathsf{N}} \ln \mathsf{p}(\mathsf{x_i} \mid \theta)$$

For example, assume normal distribution $\mathsf{p}(\mathsf{x_i}|\theta) \propto \mathsf{e}^{-(\mathsf{x_i}-\theta)^2}$ with mean θ, then

$$\widehat{\mathrm{err}}\,(\theta\,;\mathsf{x}) = \sum_{\mathsf{i}=1}^{\mathsf{N}} (\mathsf{x_i} - \theta)^2 + \mathrm{const}$$

Idea: Maximising the likelihood is minimising the error between the model and data.

$$\theta^* = \underset{\theta}{\mathrm{argmax}}\,\mathcal{L}(\theta|\mathsf{X})$$



log-likelihood surface: blue cross is maximum of likelihood

Maximum likelihood principle

Example: Birth-death process

Poisson distribution: $P(m|\theta) = e^{-\mu}\frac{\mu^m}{m!}$

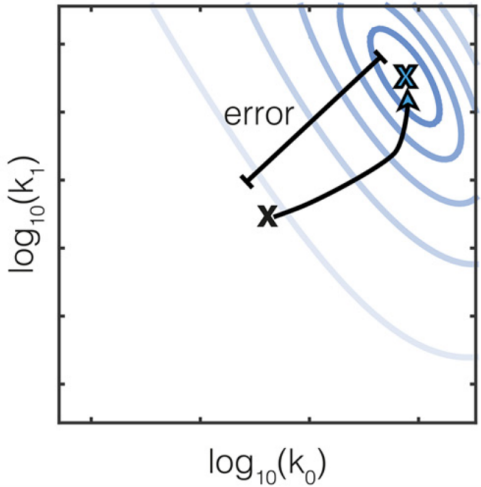
The error is

$$\widehat{\text{err}}\,(\theta\,;x) = \sum_{i=1}^{N}(x_i \ln \mu - \mu) + \text{const}$$

Minimising this gives

$$\nabla_\mu \widehat{\text{err}}\,(\theta\,;x) = \sum_{i=1}^{N}(x_i/\mu - 1) = 0$$

and hence $\mu^* = \frac{1}{N}\sum_{i=1}^{N} x_i$



log-likelihood surface: blue cross is maximum of likelihood

Maximum likelihood principle

Example: Telegraph process

Solve the Chemical Master Equation:

Beta-Poisson distribution[a]

$$P(m|\theta) = \frac{k_1^m \Gamma(m+k_{on})\Gamma(k_{off}+k_{on})}{\Gamma(m+1)\Gamma(k_{on})\Gamma(m+k_{off}+k_{on})} {}_1F_1(m+k_{on}; m+k_{off}+k_{on}; -k_{tx})$$

Likelihood can be obtained in terms of special functions.

Maximise likelihood:

$$\nabla \widehat{\mathrm{err}}(\theta; x) = -\sum_i \ln p(x_i|\theta)$$

has no closed-form solution.

[a] Raj et al. PLoS Biology 4 (2006): e309.

Maximum likelihood principle

Error guarantees

Asymptotic error guarantees (large sample size):

- Unbiased
- Smallest variances

Limitations:

- Implicitly assumes large sample size (how much is sufficient?)
- Computationally intractable optimisation; non-convex problem can suffer local likelihood maxima
- Computationally intractable likelihoods

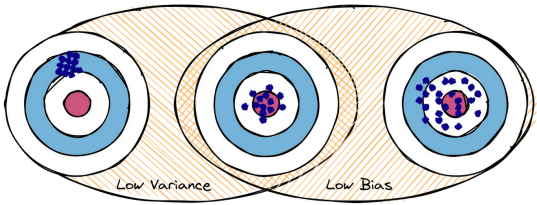


Illustration of bias and variance
(Credit: Ivan Reznikov on medium.com)

Moment-based inference

Moment-based inference

Approximation methods

Problem: Likelihood is not available

Idea: Distribution of $\hat{\mu} = \frac{1}{\mathsf{N}} \sum_{\mathsf{i}} \mathsf{x}_{\mathsf{i}}$ and $\hat{\Sigma} = \frac{1}{\mathsf{N}} \sum_{\mathsf{i}} (\mathsf{x}_{\mathsf{i}} - \mu)^2$ is asymptotically normal

Likelihood of moment data:

$$(\hat{\mu}, \hat{\Sigma}) \sim \mathsf{Normal}((\mu(\theta), \Sigma(\theta)), \mathsf{s}(\mathsf{N}))$$

where s is the matrix standard errors of mean and variance:

$$\mathsf{s}(\theta, \mathsf{N}) = \frac{1}{\mathsf{N}} \begin{bmatrix} \Sigma & \mathsf{E}[(\mathsf{x} - \mu)^3] \\ \mathsf{E}[(\mathsf{x} - \mu)^3] & \mathsf{E}[(\mathsf{x} - \mu)^4] - \frac{\mathsf{n}-3}{\mathsf{n}-1}\Sigma^2 \end{bmatrix}$$

This means

$$\begin{aligned} -2 \ln \mathcal{L}(\theta) = & \ln(|\mathsf{s}(\theta, \mathsf{N})|) \\ & + (\hat{\mu} - \mu(\theta))\mathsf{s}_{11}^{-1}(\theta)(\hat{\mu} - \mu(\theta))^{\mathsf{T}} \\ & + (\hat{\Sigma} - \Sigma(\theta))\mathsf{s}_{22}^{-1}(\theta)(\hat{\Sigma} - \Sigma(\theta))^{\mathsf{T}} + \dots \end{aligned}$$

Observations:

- Readily generalised to higher order moments
- Roughly minimised by matching mean and variance of model and data
- Penalty for high variance of summary statistics (N small)

Moment-based inference

Moments of telegraph process

- Obtain equations for the moments:

$$m_i(\theta) = E_p[x^i|\theta]$$

- Solve them in steady state:

$$m_1 = \frac{k_1 k_{on}}{k_{off} + k_{on}}$$

$$m_2 = m_1 \frac{k_1 (k_{on} + 1)}{(k_{off} + k_{on} + 1)}$$

$$m_3 = m_2 \frac{k_1 (k_{on} + 2)}{(k_{off} + k_{on} + 2)}$$

- Express parameters in terms of moments

Result:

burst frequency:

$$a = \frac{\left(2m_1^2 - m_2\right) m_3 - m_1 m_2^2}{m_2 \left(m_1^2 - 2m_2\right) + m_1 m_3},$$

burst size:

$$b = m_1 \left(\frac{m_1}{m_1^2 - m_2} + \frac{2m_2}{m_3 - m_1 m_2} \right)$$

These expressions match mean, variance and skewness of model and data ($N \gg 1$).

Moment-based inference

Approximation methods

Need for further approximation methods when moments cannot be obtained in closed-form.

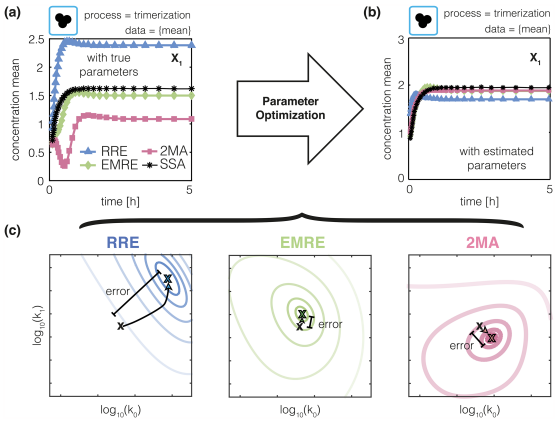
Linear noise approximation: Mean given by deterministic rate equations:

$$\frac{d}{dt}\mu = \sum_r \nu_r f_r(\mu)$$

Covariance matrix:

$$\frac{d}{dt}\Sigma = J\Sigma + \Sigma J^T + D$$

where J is the Jacobian matrix, $D = \nu \mathrm{diag}(f)\nu^T$ is noise matrix.



Approximations can fit data but have different likelihood profiles

Moment-based inference

Example: enzyme degradation process

Model:

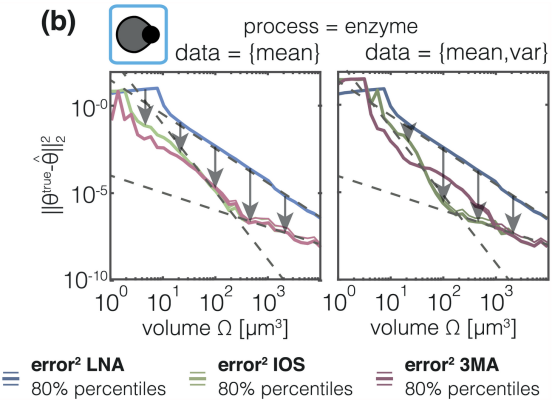
$$G \to G + M$$
$$M \to M + P$$
$$M \to \varnothing$$
$$P + E \rightleftharpoons C \to E$$

Solve using linear noise approximation or higher-order approximations.

Accuracy increases with the system size Ω that determines the total number of molecules.

Fröhlich, Thomas et al. PLoS Computational Biology (2016) 12, e1005030.

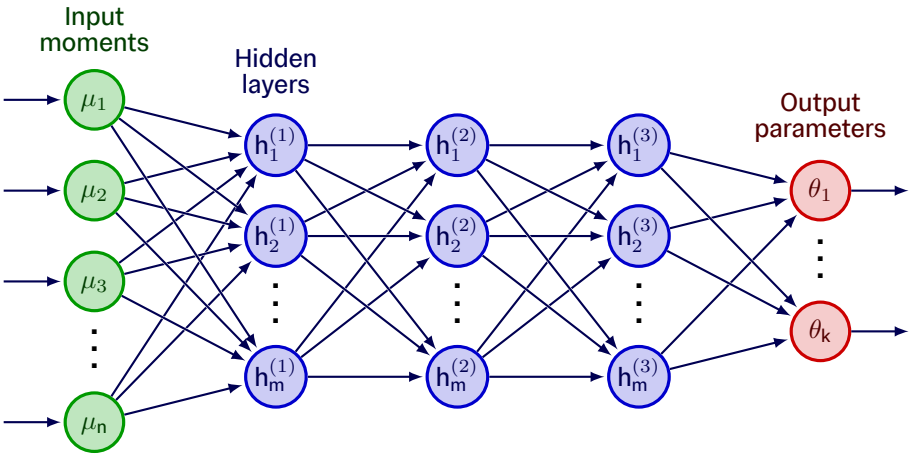
Unsurprisingly, more accurate approximations yield more accurate parameter estimates.



Machine learning-assisted inference

Machine learning-assisted inference

Neural network method



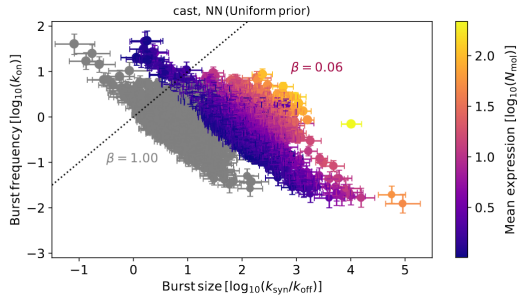
Machine learning-assisted inference

Neural network method

Regression-based algorithm

1. Simulate from the proposal (prior) $\theta' \sim \pi(\theta)$
2. Use θ' and simulate your model $x_i' \sim p(\cdot|\theta)$ for $i = 1, ..., N$.
3. Compute a number of summary statistics from your data $S_1(x'), \ldots, S_m(x')$
4. Pair the prior samples with the summary statistics to train a neural network.

Justification to follow... now onto the coding example



Bayesian inference

Bayesian inference

Bayes' theorem

Views both the data D and the parameters as random variables with joint distribution:

$$\mathrm{P}(\mathrm{D}, \theta) = \mathrm{P}(\mathrm{D}|\theta)\mathrm{P}(\theta)$$

using conditional probability:

$$\mathrm{P}(\theta|\mathrm{D}=\mathrm{X}) = \underbrace{\mathrm{P}(\mathrm{D}=\mathrm{X}|\theta)}_{\substack{=\mathcal{L}(\theta|\mathrm{X}) \\ \text{likelihood}}} \underbrace{\mathrm{P}(\theta)}_{\substack{\pi(\theta) \\ \text{prior}}} / \underbrace{\mathrm{P}(\mathrm{D}=\mathrm{X})}_{\text{evidence}}$$

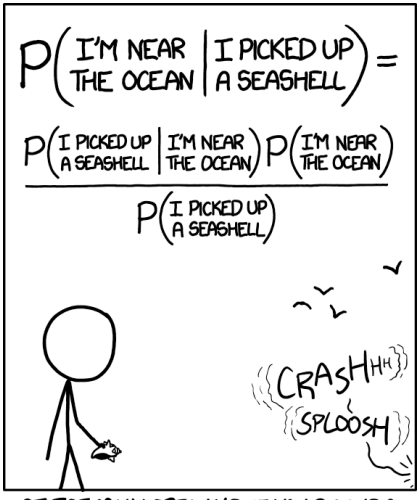


Image credit: XKCD needs a mechanistic model

Bayesian inference

Example: Poisson distribution

$$\mathcal{L}(X|\mu) = \prod e^{-\mu} \frac{\mu^{x_i}}{x_i!} \propto e^{-\mu N} \mu^{\sum_i x_i} = e^{-\mu N} \mu^{N\bar{\mu}}$$

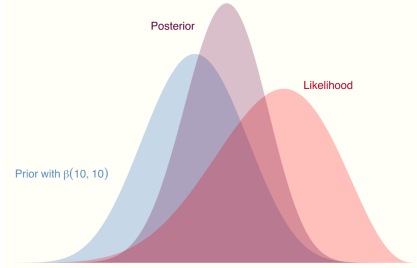
Maximum likelihood occurs when $\mu = \bar{\mu}$

Prior: $\mu \sim \Gamma(\alpha, \beta) : \pi(\mu) \propto \mu^{\alpha-1} e^{-\beta\mu}$

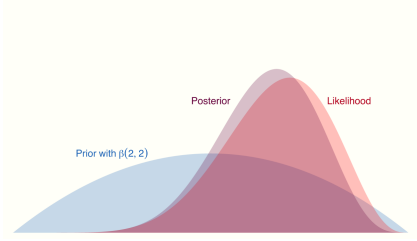
Posterior:

$$\begin{aligned} p(\mu|X) &= \mathcal{L}(\mu|X)\pi(\mu) \\ &\propto e^{-\mu N} \mu^{N\bar{\mu}} \mu^{\alpha-1} e^{-\beta\mu} \\ &= e^{-\mu N - \beta\mu} \mu^{N\bar{\mu}+\alpha-1} \end{aligned}$$

$$\mu|X \sim \Gamma(\alpha + N\bar{\mu}, N + \beta)$$



tight prior



wide prior

Bayesian inference

MCMC

Markov Chain Monte Carlo (MCMC)

- Idea: construct a stochastic process that has $p(\theta|X)$ as a stationary distribution
- **Advantages:** exact, many excellent libraries exist
- **Drawback:** requires analytically form of the likelihood
- MCMC will not be covered here

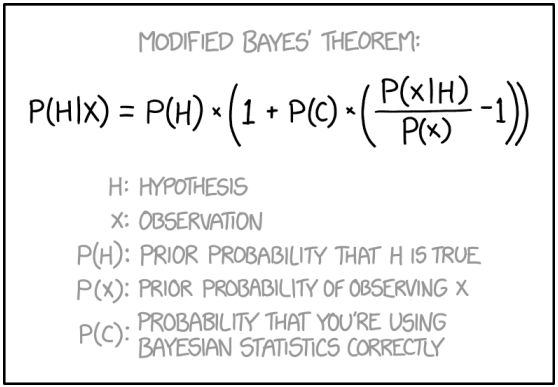


Image credit: XKCD on model uncertainty

Approximate Bayesian Computation

Simulation-based Bayesian Inference

Exact rejection sampling

Exact algorithm (Rubin 1984[a])

1. Simulate from the prior $\theta' \sim \pi(\theta)$
2. Use θ' and simulate your model $x'_i \sim p(\cdot|\theta)$ for $i = 1, ..., N$.
3. If $x' = X$ then accept θ'

Each accepted θ' is such that $\theta' \sim \pi(\theta|X)$ exactly.

[a] Rubin, DB (1984). "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician". The Annals of Statistics. 12 (4): 1151–1172. doi:10.1214/aos/1176346785.

Justification

- Joint distribution of accepted samples and parameters is $f(\theta', x') = \underbrace{\pi(\theta)}_{1.} \underbrace{\prod_{i=1}^{N} p(x'_i|\theta)}_{2.} \underbrace{\mathbb{I}_X(x')}_{3.}$
- Then the marginal distribution is $\int dx' f(\theta', x') = \pi(\theta')\mathcal{L}(\theta'|X) \propto \pi(\cdot|X)$

Problems

- Since x' is discrete, this is manageable if N is small.
- If N is large, the chance of generating the exact same sequence is essentially zero.

Approximate Bayesian Computation

ABC rejection sampling

Exact algorithm (Rubin 1984)

1. Simulate from the prior $\theta' \sim \pi(\theta)$
2. Use θ' and simulate your model $x'_i \sim p(\cdot|\theta)$ for $i = 1, ..., N$.
3. If $x' = X$ then accept θ'

Each accepted θ' is such that $\theta' \sim \pi(\theta|X)$ when S is a sufficient set of summary statistics and $\epsilon = 0$.

ABC algorithm

1. Simulate from the prior $\theta' \sim \pi(\theta)$
2. Use θ' and simulate your model $x'_i \sim p(\cdot|\theta)$ for $i = 1, ..., N$.
3. Compute a number of summary statistics from your data $S_1(x'), \ldots, S_m(x')$
4. If $d(S(x'), S(X)) < \epsilon$ then accept θ' for a given distance function d.

Justification

- The algorithm is exact when the summary statistics are sufficient and $\epsilon = 0$.
- When $\epsilon > 0$ and the summary statistics are arbitrary, this is a computationally tractable approximation.

Model selection

Frequentist vs. Bayesian inference

Akaike Information Criterion (AIC)

$$\mathrm{AIC} = 2\mathrm{k} - \max \ln \mathcal{L}(\theta|\mathrm{X})$$

Occam's razor: compromise between model fit (max-L) and model complexity (number of parameters k).

ABC model selection algorithm

1. Simulate model m and parameter θ from the prior m, $\theta' \sim \pi$
2. Use θ' and simulate model m according to $\mathrm{x}'_\mathrm{i} \sim \mathrm{p}_\mathrm{m}(\cdot|\theta)$ for $\mathrm{i} = 1, ..., \mathrm{N}$.
3. Compute a number of summary statistics from your data $\mathrm{S}_1(\mathrm{x}'), \ldots, \mathrm{S}_\mathrm{m}(\mathrm{x}')$
4. If $\mathrm{d}(\mathrm{S}(\mathrm{x}'), \mathrm{S}(\mathrm{X})) < \epsilon$ then accept θ' and m for a given distance function d.

The frequency of accepted models determines the model probability under the prior π.

Occam's razor in-built: simpler models are easier to fit.

Comparison of inference methods

Likelihood-based vs. Likelihood-free approaches

Maximum likelihood

Provides point estimate.

Bayesian inference

Provides distribution over parameters based on prior.

Moment-based inference

Uses moments instead of likelihoods.

Approximates the maximum likelihood estimate.

ABC

Simulates moments from prior.

Provides approximate posterior distributions.

Maximum posterior agrees with maximum likelihood if prior is flat.

Neural network simulation-based inference

Simulates moments from prior.

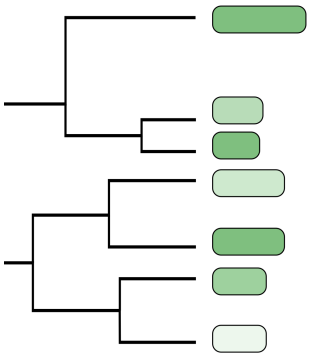
Approximates the maximum of the posterior distribution.

Application to single cell data

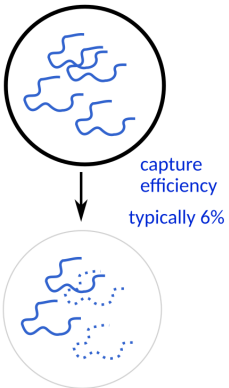
Extrinsic noise models

Stochastic concentration homeostasis: Thomas, Philipp, and Vahid Shahrezaei (2021). Royal Society Interface 18:20210274.

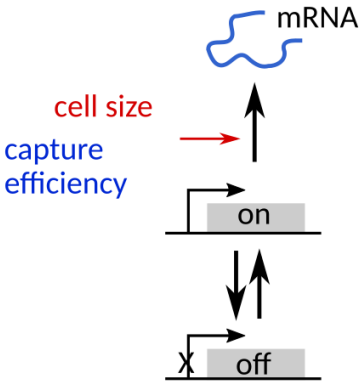
Agent-based modelling



Sequencing (technical) noise



Extrinsic noise models

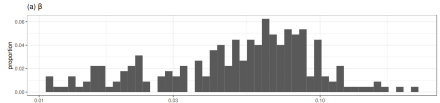


Application to single cell data

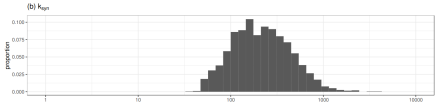
Modelling capture efficiency improves estimates of transcription rate

Mouse fibroblast data from Larsson et al. Nature 565 (2019): 251-254.

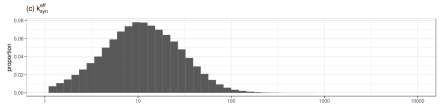
cell-specific capture efficiency



gene-specific transcription rate



effective transcription rate $\beta_{ij} = k_i c_j s_j$



Gene expression matrix of telegraph model:

$$m_{ij} | (p_i, \beta_{i,j}) \sim \text{Poisson}(p_i \beta_{i,j}),$$
$$p_i \sim \text{Beta}(k_{on,i}, k_{off,i})$$

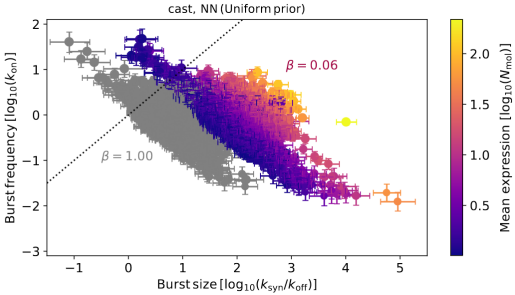
i is gene- and j is cell index.

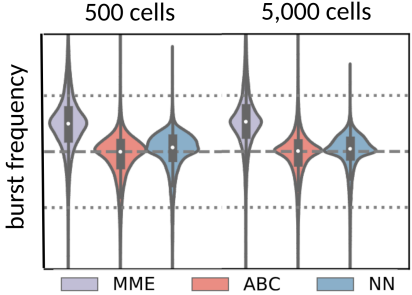
Decomposing transcription rate into cell- and gene-specific components.

- k's gene-specific rate constants
- $\beta_{i,j}$ effective transcription rate comprising
  - transcription rate
  - cell size
  - cell-specific technical noise

Application to single cell data

Inference of gene expression parameters





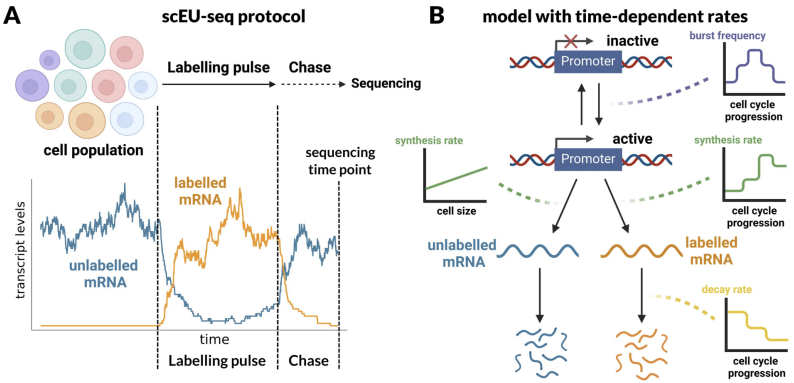
Extrinsic noise models more accurately predict burst size and frequency.

ABC and neural network approaches have similar accuracy.

Tang, Wenhao, et al. "Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics." Bioinformatics 39.7 (2023): btad395.

Application to single cell data

Agent-based modelling of time-resolved scRNA

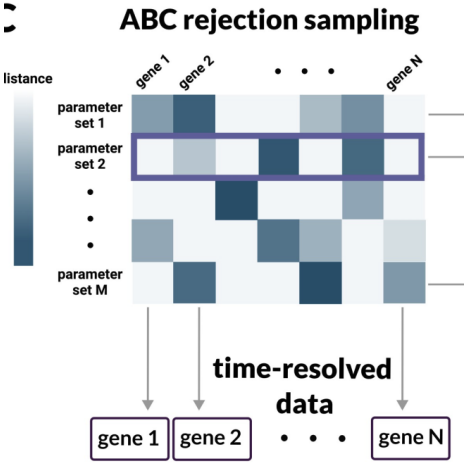


We fitted **every possible model to more than 2,000 genes** under a range of conditions.
Sounds alright?

Volteras, Shahrezaei, Thomas. Cell Systems (2024): in press. bioRxiv 10.1101/2023.10.24.563709

Application to single cell data

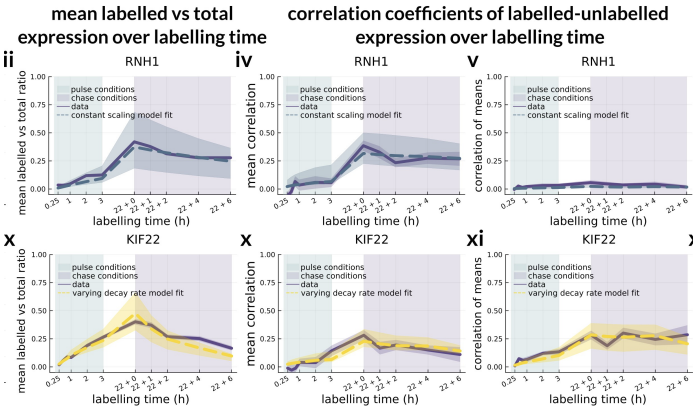
Amortised inference framework of time-resolved scRNA



ABC rejection sampling scales easily to genome-scale datasets.

Application to single cell data

ABC fits time-resolved scRNA data



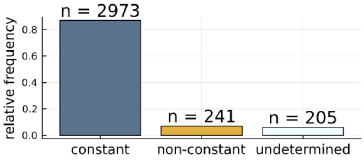
Different models are required to fit individual genes.

Human cell line data (RPE1) from Battich et al. Science 367 (2020): 1151-1156.

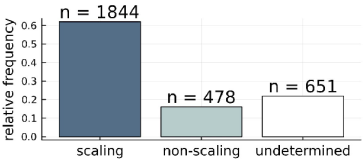
Application to single cell data

Genome-wide regulation underlies distinct regulatory mechanisms

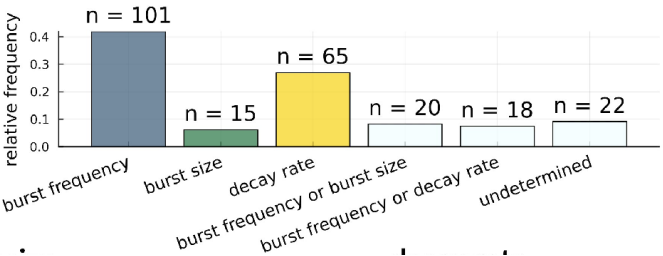
i. overall gene classification



ii. constant gene classification

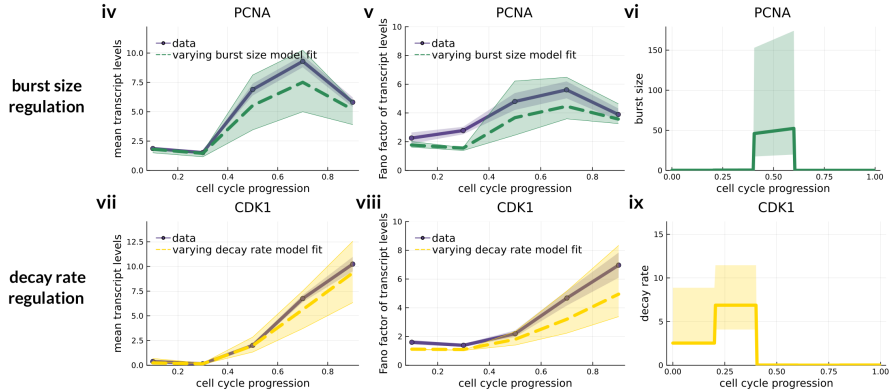


iii. non-constant gene classification



Application to single cell data

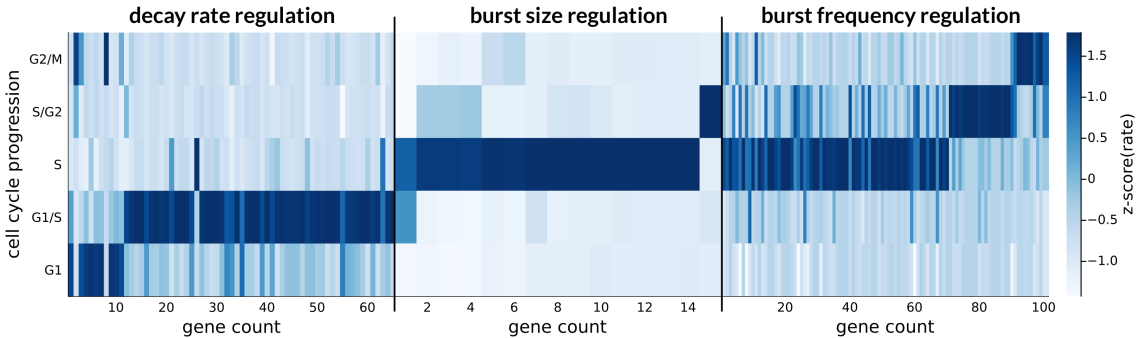
Cell cycle dependence of transcription regulation



Transcription and degradation rates of cell cycle-dependent peak at distinct cell cycle stages.

Application to single cell data

Waves of transcriptional regulation



Global regulation of cell-cycle-dependent transcripts occurs in waves.

Summary

likelihood-based inference

can be computationally challenging

or infeasible depending on model and data complexity

simulation-based methods

ABC and ML are flexible tools for mechanistic modelling of data

grounded in Bayesian inference and provide uncertainty quantification

Outlook

ABC

ML emulators
$\rightarrow$ for more efficient sampling in ABC

Tankhilevich et al. Bioinformatics 36.10 (2020): 3286-3287.

ML-assisted inference

Bayesian neural networks
$\rightarrow$ posterior distributions

Tang et al. Bioinformatics 39.7 (2023): btad395.

Moment-based inference

Error guarantees and bounds on parameters

Li et al. arxiv (2024): 0.48550/arXiv.2406.17434.

IMPERIAL

References:

Volteras, Shahrezaei, Thomas
”Global transcription regulation revealed from dynamical correlations in time-resolved single-cell RNA-sequencing.”
Cell Systems (2004, in press)

Tang, et al. ”Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics.” **Bioinformatics** (2023) 39:btad395.

Fröhlich, Thomas et al. ”Inference for stochastic chemical kinetics using moment equations and system size expansion.” **PLoS Computational Biology** (2016) 12, e1005030.

Collaborators:

inference of scRNA-seq data: Dimitris Volteras, Wenhao Tang, Andreas Joergensen, Vahid Shahrezaei

inference of tree-structured data: Fern Hughes, Alasdair Daniels, Alexis Barr

optimisation-based inference: Zekai Li, Mauricio Barahona

Thank you