Biophysical Reviews (2019) 11:95–102
https://doi.org/10.1007/s12551-019-00499-1

REVIEW

Bayesian statistical learning for big data biology

Christopher Yau[1,2] · **Kieran Campbell**[3,4]

Received: 7 January 2019 / Accepted: 8 January 2019 / Published online: 7 February 2019

Abstract
Bayesian statistical learning provides a coherent probabilistic framework for modelling uncertainty in systems. This review describes the theoretical foundations underlying Bayesian statistics and outlines the computational frameworks for implementing Bayesian inference in practice. We then describe the use of Bayesian learning in single-cell biology for the analysis of high-dimensional, large data sets.

Keywords Bayesian · Computational biology · Statistical modelling

Introduction

Statistics provides a theoretical foundation for rigorous and coherent data analysis by providing a mathematical framework in which to unify models of how data are produced by systems or experiment with techniques to handle uncertainty associated with these processes (Friedman et al. 2001). Whilst there is no single universal statistical approach, one philosophy that has gathered strength in the last 30 years is Bayesian statistical inference (Lindley 1972; Robert 2007; Bernardo and Smith 2009; Gelman et al. 2013). Bayesian statistics offers certain capabilities that enable it to be amenable to a variety of complex statistical applications and constraints, notably in machine learning, where other statistical frameworks would find difficulty. As a consequence, Bayesian approaches are now widely used in a variety of scientific and technological applications including biological research.

In this review, we will examine the fundamental concepts that underpin Bayesian Statistics and consider a concise but otherwise precise overview of the mechanics of applying Bayesian methodology. We will then consider applications of Bayesian techniques in the field of single-cell biology in which technological advances have enabled the high-throughput collection of massive quantities of data that have given us an unprecedented insight into cell function.

This article is part of a Special Issue on 'Big Data' edited by Joshua WK Ho and Eleni Giannoulatou.

✉ Christopher Yau
c.yau@bham.ac.uk

[1] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

[2] The Alan Turing Institute, London, UK

[3] Department of Statistics, University of British Columbia, Vancouver, Canada

[4] Department of Molecular Oncology, BC Cancer Agency, Vancouver, Canada

Fundamentals of Bayesian modelling

Bayesian modelling requires three components (Fig. 1a). The first is *data* (D) corresponding to measurements that are taken from the system of interest. Data can range from simple scalar values or, in big data applications, potentially complex structured tuples of multidimensional tensors (Rukat et al. 2017, 2018). The second component is a *generative model* (M) which describes a stochastic process by which the observed data arises. The generative model can be mechanistically inspired and based upon real-world physical laws and measurement processes, or may be given by generic statistical models that attempt to describe the dependencies between observed data sources and possibly unseen (latent) factors. Finally, an object of inference (θ) that we wish to learn about is required. This could be a set of unknown parameters that govern the properties of the generative model which need to be estimated or predictions of future data under alternate conditions.

We can define the *posterior probability* of the object of inference given the observed data in terms of the *likelihood* and *prior probabilities* and the *evidence* via Bayes' theorem:

$$\underbrace{p(\theta|D,M)}_{\text{Posterior}} = \frac{\overbrace{p(D|\theta,M)}^{\text{Likelihood}} \times \overbrace{p(\theta)}^{\text{Prior}}}{\underbrace{p(D|M)}_{\text{Marginal Likelihood}}} \quad (1)$$

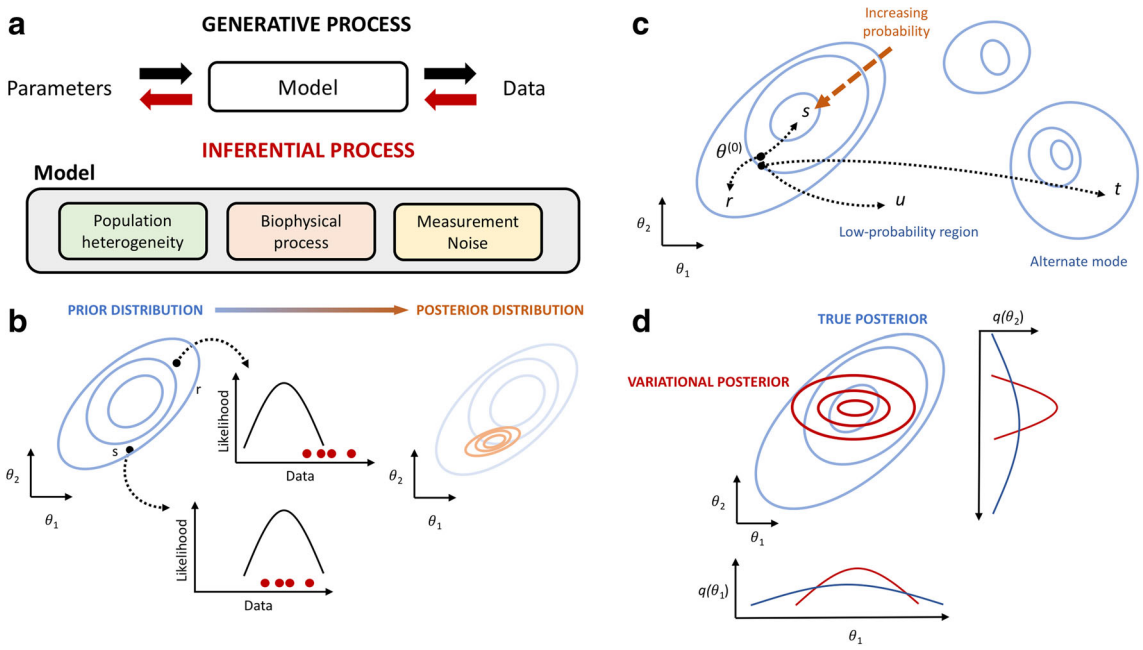

Fig. 1 **a** Overview of Bayesian modelling. Data is assumed to be generated by a stochastic model which describes various underlying processes and is specified by some unknown parameters. Bayesian inference seeks to recover those parameters from the observed data. **b** Prior beliefs are expressed as a probability distribution over parameters $\theta = (\theta_1, \theta_2)$ which are updated when data is collected via the likelihood function to give a posterior distribution over θ. **c** Real-world posterior distributions often contain a number of separated high probability regions. An ideal Metropolis-Hastings algorithm would possess a proposal mechanism that allows regular movement between different high-probability regions without the need to tranverse through low-probability intermediate regions. **d** Variational methods build approximations of the true posterior distribution. In this example, a mean-field approximation breaks the dependencies between the parameters (θ_1, θ_2) so the variational posterior models each dimension separately

What this says is that, given a generative model M and data D, the posterior probability distribution over the object of inference θ is given by our prior belief that θ takes on a certain value, scaled by the likelihood that the generative model under those beliefs would give rise to the data observed (Fig. 1b). The denominator corresponds to a normalising term to ensure that the probability distributions are valid but also describes the *marginal likelihood* of the data under the assumed model. The latter quantity is useful if alternate generative models are available and one can use the marginal likelihood as a means of determining which generative model is likely to be the most consistent with nature. Ratios of marginal likelihoods for different models, say M_1 and M_2, $P(D|M_2)/P(D|M_1)$, are known as *Bayes Factors*.

Bayesian statistics can be seen as a coherent system for probability-based belief updating. We begin with some prior knowledge about θ, we collect data and then we combine the data with our prior beliefs to give our posterior beliefs—what we believe about θ *after* seeing data. Importantly, since θ is an unobserved quantity, Bayesian inference describes our lack of certainty in its value via a probability distribution. If we take an interval of possible values for θ (a *posterior credible interval*), we can compute the amount of probability mass contained within that interval from the posterior distribution and obtain the probability that the true parameters lie in that region. This interpretation is often considered more natural than the coverage (confidence) intervals used in frequentist-based statistics.

Bayesian computation

The implementation of Bayesian computation centres on the calculation of the marginal likelihood, $p(D|M)$. This quantity is required to evaluate the posterior probability $p(\theta|D, M)$ and requires a multidimensional integral over all parameters associated with the statistical model. Direct computation is typically intractable, due to the *curse of dimensionality* for any problem of even moderate dimensionality, which results in a combinatorial explosion in the number of configurations that must be summed/integrated over. The challenges are analogous to the computation of the *partition function* in statistic mechanics and Bayesian statisticians have utilised techniques inspired by statistical mechanics to overcome this obstacle in Bayesian computation.

Monte Carlo methods

Markov Chain Monte Carlo (MCMC) simulations (Gilks et al. 1995; Brooks et al. 2011) generate sequences of random numbers such that their long-term statistical properties converge towards the target posterior distribution of interest. The predominant MCMC implementation derives from the Metropolis algorithm formulation in the 1953 paper by Metropolis et al. (1953, whose work was motivated by statistical mechanics applications involving sampling low-energy configurations of complex molecular systems). The technique was later extended in generality by Hastings (1970) to give the *Metropolis-Hastings* (M-H) algorithm. The key insight by Metropolis et al. (1953) was to derive a sampling algorithm which did not require the evaluation of the partition function (marginal likelihood) but only point-wise evaluation of the Boltzmann factors. Given a current configuration of the system θ, the Metropolis algorithm proceeds by proposing a new state θ' via any *proposal distribution* and then evaluate the Boltzmann factor $\exp(-E(\theta')/kT)$ at the proposed new state. If the new state results in a lower energy configuration then move to that new state, if it results in a higher energy configuration then choose to move to the new state with a probability which is given by the ratio of the Boltzmann factors: $\alpha = \exp(-(E(\theta') - E(\theta))/kT)$. By treating the negative logarithm of the unnormalised posterior probability distribution as an energy function, $E(\theta) = -\log p(\theta|D)$, the Metropolis algorithm (and its derivatives) has been co-opted by Bayesian statisticians as a means of efficiently performing from complex posterior distributions.

MCMC algorithms provide theoretical guarantees that the stationary distribution of the random number sequences will asymptotically converge to the posterior distribution of interest. The determination of when convergence occurs and designing efficient proposal schemes to enable that convergence to be achieved in the shortest time is highly challenging and remains an area of ongoing research. The critical design choice in the M-H algorithm is the proposal mechanism. If the proposed states are randomly chosen, they are less likely to yield high-probability configurations and will be rejected. If the new states are too similar to the current state, then their probabilities will be similar but the configurations will not be fundamentally different leading to poor exploration of the overall probability space (Fig. 1c). The proposal mechanism must therefore balance the need to search the configuration space globally whilst maintaining a sufficient locality to provide a useful acceptance rate.

A variety of modern MCMC variants now exist (Girolami and Calderhead 2011; Chen et al. 2014; Hoffman and Gelman 2014; Shahbaba et al. 2014). For instance, originally conceived by Duane et al. (Duane et al. 1987) for lattice field theory simulations of quantum chromodynamics, Bayesians have generalised *Hamiltonian Monte Carlo* (HMC) methods (Neal et al. 2011) which exploit geometric information to greatly increase the sampling efficiency of MCMC algorithms. Whilst standard M-H algorithms can be described as a *propose-and-check* approach, HMC biases proposals along trajectories that are likely to lead to high-probability configurations. Probabilistic programming languages such as Stan (Carpenter et al. 2016) and PyMC3 (Salvatier et al. 2016) contain prebuilt implementations of HMC and variants freeing modellers from many of the detailed requirements of building HMC algorithms.

Variational methods

The computational requirements of MCMC methods can be prohibitive in applications that involve large, high-dimensional data sets or complex models. As the dimensionality of θ increases, the convergence complexity of MCMC algorithms also increases when sampling from high-dimensional posteriors (Mengersen et al. 1999; Rajaratnam and Sparks 2015). An alternative is to abandon the theoretical guarantees of MCMC methods and to construct analytically tractable approximations $q_\nu(\theta|D)$ to the true posterior distribution $p(\theta|D)$—this is the motivation underlying *Variational Bayesian* methods (Blei et al. 2017).

In the construction of variational approximations, it is typical to assume that the approximating distribution has a simplified structure (Fig. 1d). The frequently used *mean-field* approximation assumes a fully factorisable form of the approximate posterior, $q_\nu(\theta|D) = \prod_{t=1}^{T} q_\nu^{(t)}(\theta_t|D)$ where the dependencies between the different elements of θ are uncoupled and each factor $q_\nu^{(t)}$ is typically given by a simple distribution (e.g. Gaussian, Gamma). If the approximating distribution q_ν is parameterised by ν, the variational approach seeks to optimise these *variational parameters* to minimise the difference—measured using the Kullback-Leibler (KL) divergence—between the true and approximate posterior distributions. Therefore, unlike Monte Carlo methods which use stochastic sampling, variational methods transform the inference problem into an optimisation task. The latter means that assessing the convergence of variational methods is relatively straightforward and typically requires significantly less time for complex models than MCMC approaches.

Classic variational algorithms used analytically derived optimisation steps (coordinate ascent VI) but, more recently, stochastic variational inference (SVI) methods employ stochastic gradient descent algorithms instead (Hoffman et al. 2013; Titsias and Lázaro-Gredilla 2014). SVI uses cheap to compute, "noisy" estimates of natural gradients based on a subset of data points instead of the true gradients

which require a pass through all data points. This exploits the fact that the expected value of these noisy gradients is equal to the true gradient and so convergence of the SVI algorithm can be guaranteed under certain conditions. As a consequence, SVI allows the application of variational methods to a wider class of models and by operating on *mini-batches* of data in each optimisation step provides substantial speed-ups in large data settings.

Amortised variational inference uses *inference networks* within variational inference algorithms for latent variable models—where each data item is associated with its own set of parameters (Zhang et al. 2018). In such situations, a typical variational approximation would result in each data item also being associated with its own compliment of variational parameters; thus, with larger data sets, there would be an increase in the number of variational parameters to optimise. Inference networks replace these *local* variational parameters with a single set of global parameters associated with a neural network. The goal is to use variational inference to optimise the parameters associated with this neural network and to use the optimised network to *predict* the local variational parameters. Inference networks therefore offer another layer of approximation that breaks the dependency between the computational requirements of the variational inference algorithm and the size of the data set.

Whilst the accuracy of variational approximations is often impossible to quantify (Yao et al. 2018), they provide the current mainstay inferential approach for high-dimensional Bayesian modelling. Current machine learning development libraries, such as TensorFlow (Abadi et al. 2016) and PyTorch (Paszke et al. 2017), provide a substantial body of tools for the construction of neural networks, and optimisation algorithms for the implementation of variational inference algorithms.

Bayesian applications in single-cell biology

The recent availability of a plethora of relatively low-cost experimental methods and protocols for high-throughput screening of individual cells has lead to an explosion in single-cell biological data (Theillet 1998). A single-cell experiment can generate data that is both high-dimensional and large in sample size with recent studies involving single-cell RNA sequencing routinely able to produce cell numbers on the order of 10^5 cells measuring $10^3 - 10^4$ genes. International endeavours, such as The Human Cell Atlas (HCA) project (Regev et al. 2017), will seek to catalogue and sequence all known human cell types in the coming years.

The benefit of single-cell measurements is to remove the averaging effect when measurements are taken on

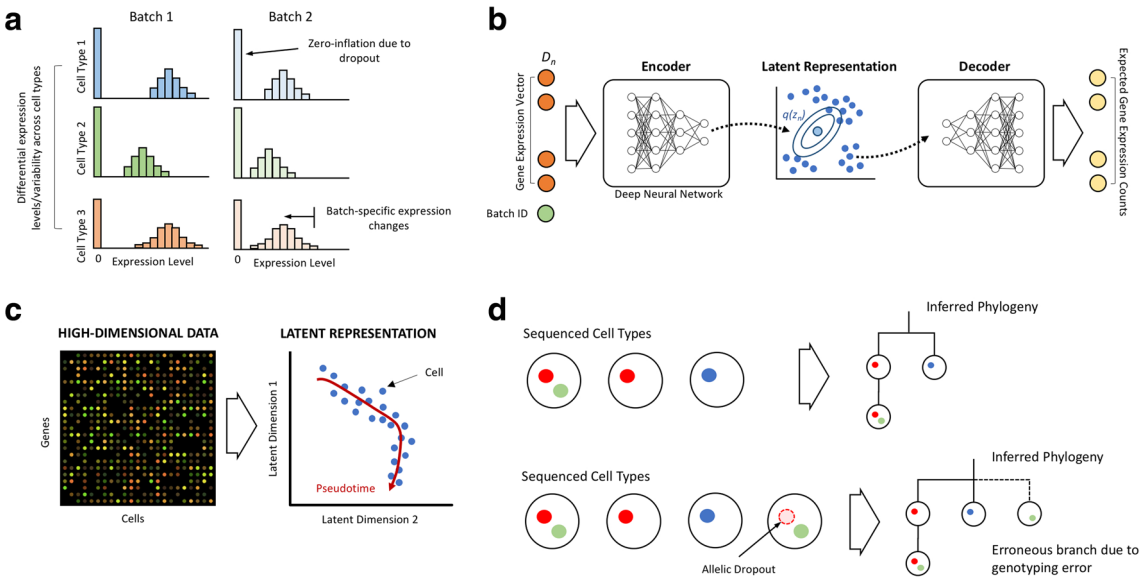

Fig. 2 **a** Single-cell differential expression analysis aims to identify differences in expression level and variability between cell types. Confounding effects such as dropout and batch effects must be accounted for in order to avoid false conclusions. **b** Variational autoencoders use deep neural networks to *encode* input expression data vectors into low-dimensional latent representations whilst simultaneously learning *decoders* that can generate realistic expression data from these latent representations. **c** Pseudotemporal model aims to identify latent uni-dimensional representations that correspond to physical time variation from high-dimensional cross-sectional single-cell data. **d** Probabilistic approaches to tumour phylogeny inference are essential in the presence of sequencing noise since genotyping errors can lead to uncertainties in phylogenetic reconstruction. Here, the presence of allelic dropout leading to genotyping error in a single-cell type could lead to alternate phylogenetic histories and different interpretations of the importance of acquired mutations

populations of cells which can obscure important stochastic dynamics operating in individual cells. However, the challenge when working with single cells is the inherent sensitivity of cells to physical manipulation and the difficulties of robustly measuring minuscule quantities of potentially unstable molecules, e.g. RNA. Consequently, single-cell data from any technical platform is inherently noisy, contains various levels of missingness and may harbour many sources of bias—all of which could have both a biological or technical origin (Stegle et al. 2015; Poirion et al. 2016). Probabilistic modelling of single-cell data, based on a Bayesian framework, provides a coherent strategy for encapsulating these complexities.

Differential expression

Differential expression (DE) aims to identify genes that are up- or downregulated between cell types (Fig. 2a). Whilst standard frequentist-based hypothesis testing procedures can be employed, Bayesian DE alternatives offer certain benefits. Kharcenko et al. (2014) introduced a generative model that includes drop-outs for differential expression analysis. Dropouts are frequent occurrences in single-cell expression data due to the low quantities of mRNA involved which means the presence of some transcripts cannot always be reliably detected by sequencing—the result is a zero expression measurement for cells that might actually be expressing a gene at a low level. BASiCS (Bayesian Analysis of Single-Cell Sequencing data (Vallejos et al. 2015; Vallejos et al. 2016)) jointly models highly variable genes and differential expression between cell populations which allows it to detect *differential variability*—an effect often masked by both standard differential expression methods. A related analysis is identifying differential splicing in which exon usage varies between cells or cell populations. Differential splicing can be difficult to detect in single-cell RNA-seq data due to amplification biases, shallow read depths, and other technical artefacts. To solve this, the Bayesian method BRIE (Bayesian regression for isoform estimation (Huang and Sanguinetti 2017)) leverages sequence-derived features as an informative prior distribution in a hierarchical model to greatly increase the accuracy of inference.

Deep learning representations

One particular analytical problem is the identification of latent structure within these high-dimensional data sets in order to understand the underlying fundamental biological processes. Ideas inspired from *deep learning* (LeCun et al. 2015) have recently emerged in single-cell biology as a means of extracting low-dimensional representations from high-dimensional data (Ding et al. 2018; Lopez et al. 2018). For instance, scVI (Lopez et al. 2018) uses a hierarchical Bayesian model—a variational autoencoder (Kingma and Welling 2013), incorporating deep neural networks and stochastic variational inference stochastic optimisation to aggregate information across similar cells and genes whilst simultaneously adjusting for batch effects and lack of measurement sensitivity (Fig. 2b). The benefit of the deep neural networks is that the functional relationship between the measured gene expression and the latent representations does not need to be prespecified by the modeller and scVI is able to exploit the vast array of data available to *learn* these relationships from the data itself. Implementations using modern machine learning development frameworks allow a vast array of high-performance computational machinery (such as graphics processing units) to be exploited permitting methods such as scVI to make short work of data sets involving millions of cells.

Temporal modelling

High-throughput single-cell molecular technologies provide an instantaneous measurement of the molecular state of individual cells. Genuine time series measurements of individual cells undergoing dynamic processes, such as differentiation or cell cycle, are difficult due to the inherently destructive nature of the measurement process and asynchronicity of cellular progression. To circumvent this, analytical methods have been developed that use a cross-sectional "snapshot" of cells' gene expression to assign a *pseudotime* to each cell--a surrogate measure of progression through the process of interest (Fig. 2c). Downstream analyses such as differential expression (Campbell and Yau 2017b; Sander et al. 2017) can then be performed using the pseudo times in lieu of physical time information.

A majority of Bayesian pseudotime inference methods build upon the Gaussian Process Latent Variable (GPLVM) framework. The first model for single-cell RNA-seq was DeLorean (Reid and Wernisch 2016) that uses a $\text{Matern}_{3/2}$ kernel with a Gaussian likelihood on suitably log-transformed data. DeLorean uses the probabilistic programming language Stan (Carpenter et al. 2016) for inference that performs an adaptive version of Hamiltonian Monte Carlo. This was recently reimplemented in the method GrandPrix (Ahmed et al. 2019) with fast inducing point variational inference implemented in the GPFlow framework (Matthews et al. 2017) to achieve order-of-magnitude faster inference. A related model is the PseudoGP framework (Campbell and Yau 2016) that uses the posterior distributions from probabilistic pseudotime to quantify the uncertainty in downstream analyses such as differential expression. Branching differentiation processes can also be modelled using Gaussian processes (Boukouvalas et al. 2018; Penfold et al. 2018).

Further, Bayesian pseudotime methods have been developed based on dimensionality reduction techniques other than GPLVM. A popular class of these are factor analysis models that seek a probabilistic mapping from the latent space (pseudotimes) through a linear or parametric nonlinear function. Such an approach was successfully applied in the Ouija framework (Campbell and Yau 2018) that uses a sigmoidal mapping to learn pseudotimes from small marker gene panels along with interpretable parameters corresponding to activation times of genes. A related model is MFA (Campbell and Yau 2017a) that implements a mixture of linear factor analysers to infer bifurcations from single-cell gene expression data, using MCMC sampling for inference. Finally, a Bayesian variant of unidimensional scaling (BUDS, Nguyen and Holmes 2017) has been proposed for ordering single cell with an emphasis on visualising uncertainty.

Tumour evolution

Bayesian approaches have also been developed for single-cell–based modelling of cancer evolution (Zafar et al. 2018; Goh et al. 2019). Here, the data corresponds to genome sequences of tumour samples and the unobserved object of inference is the evolutionary tree relating the different cancer cell populations within the tumour (Yuan et al. 2015; Roth et al. 2016) or a mutation tree representing the partial (temporal) order of the mutation events (Jahn et al. 2016; Ross and Markowetz 2016). Since an arbitrary number of evolutionary mechanisms may be possible, the information included in the priors help to *regularise* the inferential problem to make it tractable by limiting the space of possible evolutionary trajectories. Uncertainty propagation is also of specific help in this problem. Allelic dropout in single-cell sequencing can cause mutations to become undetected and lead to errors in the genotyping of individual cells. Errors in cellular mutation profiles could fundamentally alter the inferred evolutionary trees hence joint modelling of sequencing errors and evolutionary trajectories is critical (Fig. 2d).

Discussion

Bayesian methodology is a conceptually natural approach to apply to biological research applications. Modern probabilistic programming language environments for Bayesian computation have further facilitated its application by providing interfaces for specifying potentially highly complex models even for non-experts. This review has described the underlying theoretical framework as well as the computational techniques required to implement Bayesian modelling with a focus on applications in single-cell biology. Nonetheless, further research into improved and faster Bayesian computation techniques for big data biology is required. Despite its strong theoretical foundations, Bayesian approaches are still relatively underused in biological sciences. Bayesian modelling requires considerable thought to be given to the constitution of the generative models and the specification of prior beliefs. Probabilistic programming languages have simplified model development by allowing users to focus on model specification rather than the computational implementations but there remains a considerable "art" to designing good models and expertise is gained through experience. Research to develop more *automatic* tools for Bayesian model specification would be beneficial. Posterior uncertainty characterisation intrinsically means that there is no "right answer" in Bayesian modelling—only a distribution over possibilities. Probabilistic outcomes can be difficult to interpret even for seasoned experts and non-experts may find such summaries challenging to palate. Finally, in high-dimensional, large data settings, recent computational advances have made Bayesian inference feasible for increasingly larger problems but often remains more computationally taxing than alternative approaches that might forego uncertainty characterisation for point estimation. However, as described in many of the single-cell applications, without formal uncertainty modelling, erroneous inferences can be made in the presence of confounding factors or noisy/missing data.

Acknowledgements CY is supported by a UK Medical Research Council Research Grant (Ref: MR/P02646X/1) and by The Alan Turing Institute under the EPSRC grant EP/N510129/1.

Compliance with Ethical Standards

Conflict of interests Christopher Yau declares that he has no conflict of interest. Kieran Campbell declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M et al (2016) Tensorflow: a system for large-scale machine learning. OSDI 16:265–283

Ahmed S, Rattray M, Boukouvalas A (2019) Grandprix: scaling up the bayesian gplvm for single-cell data. Bioinformatics (Oxford England) 35:47–54. https://doi.org/10.1093/bioinformatics/bty533
Bernardo JM, Smith AF (2009) Bayesian theory, vol 405. Wiley, New York
Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: A review for statisticians. J Am Stat Assoc 112(518):859–877
Boukouvalas A, Hensman J, Rattray M (2018) Bgp: identifying gene-specific branching dynamics from single-cell data with a branching gaussian process. Genome Biol 19(1):65
Brooks S, Gelman A, Jones G, Meng XL (2011) Handbook of Markov chain Monte Carlo. CRC Press, Boca Raton
Campbell KR, Yau C (2016) Order under uncertainty: robust differential expression analysis using probabilistic models for pseudotime inference. PLoS Comput Biol 12(11):e1005,212
Campbell KR, Yau C (2017) Probabilistic modeling of bifurcations in single-cell gene expression data using a bayesian mixture of factor analyzers. Wellcome Open Res 2:19. https://doi.org/10.12688/wellcomeopenres.11087.1
Campbell KR, Yau C (2017) Switchde: inference of switch-like differential expression along single-cell trajectories. Bioinformatics 33(8):1241–1242
Campbell KR, Yau C (2018) A descriptive marker gene approach to single-cell pseudotime inference. Bioinformatics 35(1):28–35
Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A (2016) Stan: a probabilistic programming language. J Stat Softw 20:1–37
Chen T, Fox E, Guestrin C (2014) Stochastic gradient Hamiltonian Monte Carlo. In: International conference on machine learning, pp 1683–1691
Ding J, Condon A, Shah SP (2018) Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun 9(1):2002. https://doi.org/10.1038/s41467-018-04368-5, http://www.ncbi.nlm.nih.gov/pubmed/29784946
Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid monte carlo. Phys Lett B 195(2):216–222
Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer Series in Statistics, New York
Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis. Chapman and Hall/CRC, Boca Raton
Gilks WR, Richardson S, Spiegelhalter D (1995) Markov chain Monte Carlo in practice. Chapman and Hall/CRC, Boca Raton
Girolami M, Calderhead B (2011) Riemann manifold langevin and Hamiltonian Monte Carlo methods. J R Stat Soc Ser B Stat Methodol 73(2):123–214
Goh G, McGranahan N, Wilson GA (2019) Computational methods for analysis of tumor clonality and evolutionary history. In: Cancer bioinformatics. Springer, pp 217–226
Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications
Hoffman MD, Blei DM, Wang C, Paisley J (2013) Stochastic variational inference. J Mach Learn Res 14(1):1303–1347
Hoffman MD, Gelman A (2014) The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15(1):1593–1623
Huang Y, Sanguinetti G (2017) Brie: transcriptome-wide splicing quantification in single cells. Genome Biol 18(1):123
Jahn K, Kuipers J, Beerenwinkel N (2016) Tree inference for single-cell data. Genome Biol 17(1):86
Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. Nat Methods 11(7):740–742. https://doi.org/10.1038/nmeth.2967
Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv:1312.6114
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436
Lindley DV (1972) Bayesian statistics, a review, vol 2. SIAM
Lopez R, Regier J, Cole MB, Jordan MI, Yosef N (2018) Deep generative modeling for single-cell transcriptomics. Nat Methods 15(12):1053–1058. https://doi.org/10.1038/s41592-018-0229-2, http://www.ncbi.nlm.nih.gov/pubmed/30504886
Matthews DG, Alexander G, Van Der Wilk M, Nickson T, Fujii K, Boukouvalas A, León-Villagrá P, Ghahramani Z, Hensman J (2017) Gpflow: a Gaussian process library using tensorflow. J Mach Learn Res 18(1):1299–1304
Mengersen KL, Robert CP, Guihenneuc-Jouyaux C (1999) Mcmc convergence diagnostics: a reviewww. Bayesian Stat 6:415–440
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21(6):1087–1092
Neal RM et al (2011) Mcmc using Hamiltonian dynamics. Handbook of Markov Chain Monte Carlo 2(11):2
Nguyen LH, Holmes S (2017) Bayesian unidimensional scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations. BMC Bioinf 18(10):394
Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in pytorch
Penfold CA, Sybirna A, Reid JE, Huang Y, Wernisch L, Ghahramani Z, Grant M, Surani MA (2018) Branch-recombinant gaussian processes for analysis of perturbations in biological time series. Bioinformatics 34(17):i1005–i1013. https://doi.org/10.1093/bioinformatics/bty603
Poirion OB, Zhu X, Ching T, Garmire L (2016) Single-cell transcriptomics bioinformatics and computational challenges. Front Genet 7:163
Rajaratnam B, Sparks D (2015) Mcmc-based inference in the era of big data: a fundamental analysis of the convergence complexity of high-dimensional chains. arXiv:1508.00947
Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M et al (2017) Science forum: the human cell atlas. Elife 6:e27,041
Reid JE, Wernisch L (2016) Pseudotime estimation: deconfounding single cell time series. Bioinformatics 32(19):2973–2980
Robert C (2007) The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media
Ross EM, Markowetz F (2016) Onconem: inferring tumor evolution from single-cell sequencing data. Genome Biol 17(1):69
Roth A, McPherson A, Laks E, Biele J, Yap D, Wan A, Smith MA, Nielsen CB, McAlpine JN, Aparicio S et al (2016) Clonal genotype and population structure inference from single-cell tumor sequencing. Nat Methods 13(7):573–576
Rukat T, Holmes C, Yau C (2018) Probabilistic boolean tensor decomposition. In: International conference on machine learning, pp 4410–4419
Rukat T, Holmes CC, Titsias MK, Yau C (2017) Bayesian Boolean matrix factorisation. In: International conference on machine learning, pp 2969–2978
Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in python using pymc3. PeerJ Comput Sci 2:e55
Sander J, Schultze JL, Yosef N (2017) Impulsede: detection of differentially expressed genes in time series data using impulse models. Bioinformatics 33(5):757–759
Shahbaba B, Lan S, Johnson WO, Neal RM (2014) Split Hamiltonian Monte Carlo. Stat Comput 24(3):339–349

Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16(3):133

Theillet C (1998) Full speed ahead for tumor screening. Nat Med 4(7):767–768

Titsias M, Lázaro-Gredilla M (2014) Doubly stochastic variational bayes for non-conjugate inference. In: International conference on machine learning, pp 1971–1979

Vallejos CA, Marioni JC, Richardson S (2015) Basics: Bayesian analysis of single-cell sequencing data. PLoS Comput Biol 11:e1004,333. https://doi.org/10.1371/journal.pcbi.1004333

Vallejos CA, Richardson S, Marioni JC (2016) Beyond comparisons of means: understanding changes in gene expression at the single-cell level. Genome Biol 17:70. https://doi.org/10.1186/s13059-016-0930-3

Yao Y, Vehtari A, Simpson D, Gelman A (2018) Yes, but did it work?: evaluating variational inference. arXiv:1802.02538

Yuan K, Sakoparnig T, Markowetz F, Beerenwinkel N (2015) Bitphylogeny: a probabilistic framework for reconstructing intra-tumor phylogenies. Genome Biol 16(1):36

Zafar H, Navin N, Nakhleh L, Chen K (2018) Computational approaches for inferring tumor evolution from single-cell genomic data. Curr Opin Syst Biol 7:16–25

Zhang C, Butepage J, Kjellstrom H, Mand S (2018) Advances in variational inference. IEEE Transactions on Pattern Analysis and Machine Intelligence