https://doi.org/10.1038/s42254-024-00698-0

Using Bayesian parameter estimation to learn more from data without black boxes

Rachel C. Kurchin

In an age of expensive experiments and hype around new data-driven methods, researchers understandably want to ensure they are gleaning as much insight from their data as possible. Rachel C. Kurchin argues that there is still plenty to be learned from older approaches without turning to black boxes.

At the core of the scientific enterprise is learning from data, often through describing it by parameterizing a mathematical model. And yet, across a wide swath of science, despite the existence of a wide variety of types of data and models, the go-to technique for this parameterization is almost always the same: minimizing the squared error (classical regression). A key assumption of this approach is that there exists a single optimal point in the parameter space. However, there is an alternative that doesn't force this assumption on us: Bayesian parameter estimation (BPE). It instead works with a probability distribution in the space of the model parameters. Doing so offers a variety of other advantages, including facile incorporation of information from other sources and prospective evaluation of the impact of additional data collection.

The value of BPE may not be immediately obvious in a data-abundant and data-enthusiastic age, in which it is tempting to turn to contemporary machine learning methods such as neural networks to distill insights from data when classical regression is insufficient. Indeed, these can be powerful tools in certain contexts, but they should not necessarily be treated as 'frontline' tools, as there are not currently reliable methods to interpret, or bound errors of, these methods. BPE, coupled with traditional numerical simulation, is an appealing path to thinking outside the black box.

The Bayesian approach

The most common approach to parameter estimation is to frame it as an optimization problem over the parameters, with the goal of minimizing some distance function (typically the sum of squared errors) between the model predictions and the measured data by identifying the 'best fit' set of parameters. The minimum achievable value of the distance function is often used as a goodness-of-fit metric to assess the appropriateness of the model for the data and potentially to suggest the presence of effects unaccounted for by the model. Its landscape in the neighbourhood of the best-fit point can also give some insights regarding confidence intervals and local correlation structure.

By contrast, BPE instead produces, as a fundamental output of its analysis, a probability distribution in the space of the model parameters. This distribution, in the Bayesian interpretation, reflects, for each combination of parameter values, the strength of belief that is justified in it being the 'correct' one to describe the given data. BPE is built on Bayes' theorem, which states

$$\underbrace{P(H|E)}_{\text{posterior}} = \frac{\overbrace{P(E|H)}^{\text{likelihood}}\,\overbrace{P(H)}^{\text{prior}}}{\underbrace{P(E)}_{\text{evidence}}},$$

where E represents some evidence (in this case, the observed data) and H a hypothesis (in this case, a point in the parameter space).

In essence, Bayes' theorem shows how to 'convert' a likelihood (the probability of observing some evidence in a world where our hypothesis is true) to a posterior (the probability of our hypothesis being true, given we have observed the evidence). An important note is that the likelihood function operates fundamentally in the observation space, in which the data (E) lives. The posterior, conversely, exists in the parameter space, in which the hypotheses (H) live.

To frame subsequent discussion in a concrete context, consider the case (Fig. 1) of an object launched from an initial height of 0, with an unknown initial velocity v_0 and on some planet with an unknown value of g. We thus wish to estimate the two parameters v_0 and g. The model function is the kinematic equation

$$y^{\text{model}}(t; v_0, g) = v_0 t + \frac{1}{2} g t^2.$$

The prior distribution is a choice that must be made by the practitioner. In principle, it can be a way to include one's prior beliefs in the analysis; in practice, given enough data, often all that matters about the prior is that it is nonzero in all plausible regions of the parameter space. For our purposes, we presume a uniform prior over a rectangular region of parameter space.

In the general case of Bayesian inference, the likelihood function can take many forms. For the specific case of fitting a model to experimental observations, if we presume that the observations (t_i, y_i^{meas}) come equipped with (potentially pointwise) uncertainty or error estimates Δy_i, a Gaussian likelihood is a reasonable choice:

$$P(\{y^{\text{meas}}\}|v_0, g) \propto \prod_i \exp\left(-\frac{(y^{\text{model}}(t_i; v_0, g) - y_i^{\text{meas}})^2}{2(\Delta y_i)^2}\right).$$

This choice is also the most analogous to least-squares regression, which is equivalent to maximum likelihood estimation under the assumption of normally distributed errors. Likelihoods associated with two different height measurements can be found in Fig. 1a and d.

Finally, we need to compute the evidence, which serves as the normalization constant of our posterior distribution. There are a variety of approaches to do so. The conceptually simplest is to divide the

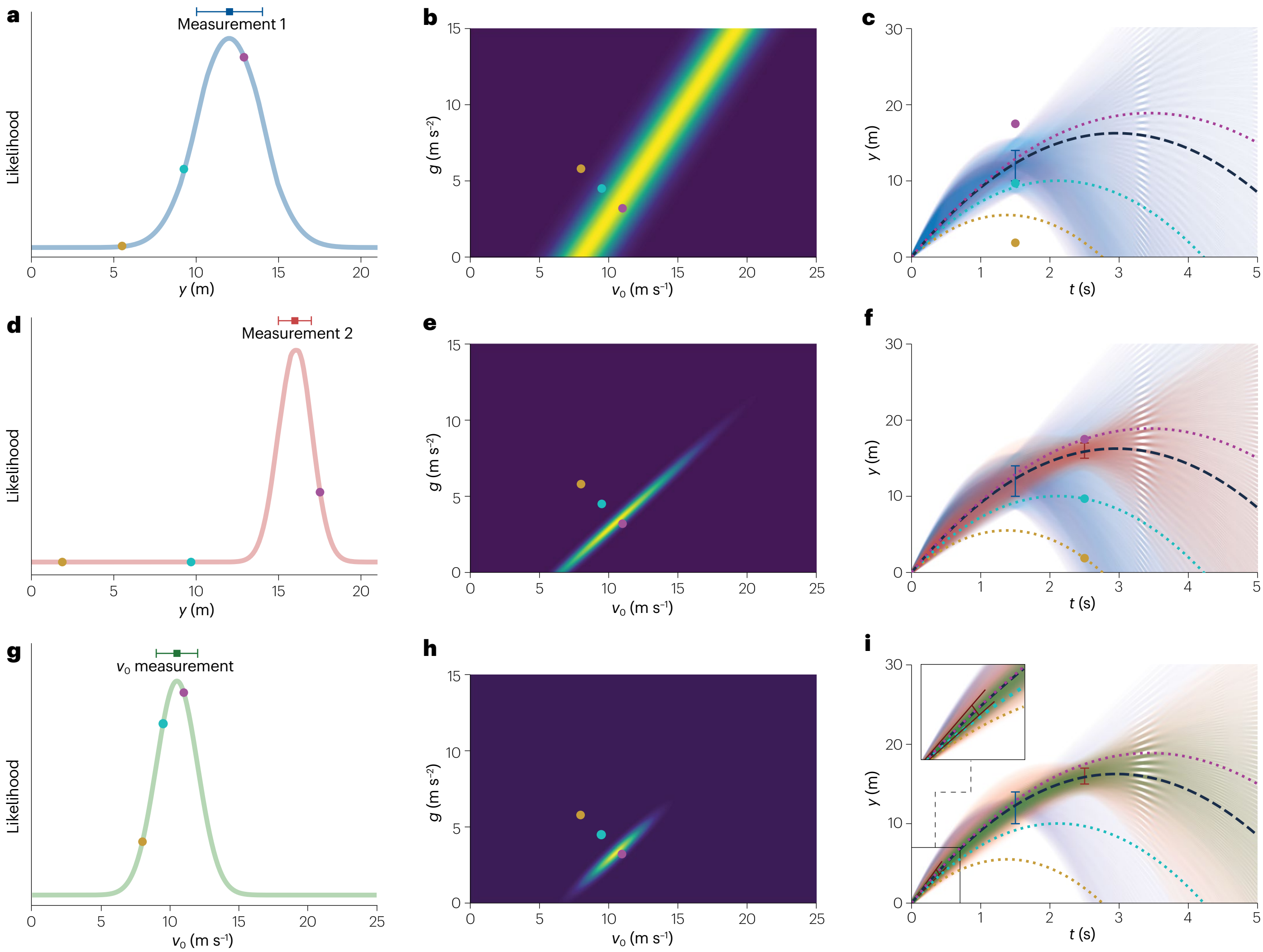

Fig. 1 | Bayesian parameter estimation for experiments on a projectile launched from a height of 0 with unknown initial velocity and subject to unknown gravitational acceleration. a–c, Likelihood, with associated uncertainty, for one measurement (part **a**), the posterior after one measurement visualized across the parameter space (part **b**) and possible trajectories with opacity proportional to posterior probability (part **c**). **d–f**, Likelihood for a second measurement (part **d**), the posterior after two measurements (part **e**), and trajectories (part **f**) with the trajectories from part **c** underlaid at lower opacity for ease of comparison. **g–i**, Likelihood for a direct measurement of the initial velocity (part **g**), the posterior conditioned on this measurement (part **h**), and trajectories (part **i**) with trajectories from parts **c** and **f** underlaid. The trajectory corresponding to the 'true' parameters is shown in the dotted black lines. Throughout, three representative points in parameter space (coloured dots) are shown to illustrate the correspondence of these specific points across the visualizations. Measurements are indicated with vertical (parts **c**, **f**) and angular (part **i**) error bars. y, height; v_0, initial velocity; g, acceleration due to gravity; t, time.

parameter space into a regular grid, in which case the normalization comes 'for free' by summing the product of the prior and the likelihood over every grid point. However, this approach is only tractable if hard bounds on all parameters are known, and the dimensionality of the parameter space and computational cost of the model are sufficiently low.

Other approaches fall broadly into two categories. The first is using Markov Chain Monte Carlo to build up a numerical approximation to the evidence (or directly to the posterior distribution) via iterative sampling. The second is variational inference, in which the distribution is approximated via a set of basis functions[1]. There are active efforts in methods development in both categories, aimed at improving convergence behaviour and reducing computational cost.

A distribution as the fundamental object

There are a variety of advantages to working directly with a probability distribution in the parameter space. With this distribution in hand, it is easy to interrogate the broader parameter landscape, beyond just the local curvature implied by the confidence intervals and correlation accessible in standard regression. Insights can also be gleaned into identifiability of a system given the specific observed data. However, this landscape could be qualitatively accessible as a loss landscape without the need for a probabilistic approach, and so the real strength comes from the ability to do operations most naturally done on distributions. For example, it is straightforward to marginalize a posterior to account for nuisance parameters, or condition it based on information from other sources. This is demonstrated in the bottom row of Fig. 1, where a direct measurement of initial velocity is incorporated by conditioning the posterior. Similarly, if the goal were only to learn about the gravitational acceleration, one could marginalize over v_0 by integrating it out of the distribution to create a single-variable posterior over g. Within a regression approach, it would be much less straightforward to account for these heterogeneous measurement types.

The ability to marginalize and condition distributions also allows straightforward interrogation of the impact of a given data point or set of points on our degree of confidence in the result, either post hoc or prospectively. The latter case can be useful in active learning schemes that, for example, seek to maximize information gained from a finite experimental measurement budget (this approach forms the conceptual link between Bayesian parameter estimation and the related technique of Bayesian optimization). In addition, for many model selection tasks, a distribution is easier to work with, and can be assessed conditionally. Furthermore, in cases where the parameters being estimated are best described as a distribution rather than as single scalar values, a probabilistic approach like BPE is the more natural and flexible choice.

Outlook

There are examples of researchers using BPE across a wide variety of fields such as physiology[2], fluid mechanics[3], energy devices[4–6], and particle physics[7]. The fields of astronomy[8], astrophysics[9], and cosmology[10] have an especially strong culture of using Bayesian methods in their analysis of observational data and model formulation and fitting. This could be related to a culture of broad sharing of large datasets across the community, and interest in distributions of properties across populations of astronomical objects as well as inference of the parameters of underlying models. It likely is also because they deal with fundamental theories and questions at extreme length scales with many unknowns, extremely constrained observation, and virtually no ability to conduct controlled experiments in the conventional sense of directly probing a system to influence its behaviour. However, the advantages of probabilistic methods such as BPE coupled with traditional modelling offer an approach worth serious consideration by a broader swath of the science and engineering research communities.

Rachel C. Kurchin ✉
Carnegie Mellon University, Materials Science and Engineering Department, Pittsburgh, PA, USA.
✉ e-mail: rkurchin@cmu.edu

References

1. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian Data Analysis* (Chapman and Hall/CRC, 1995).
2. Bartsoen, L. et al. Bayesian parameter estimation of ligament properties based on tibio-femoral kinematics during squatting. *Mech. Syst. Signal Process.* **182**, 109525 (2023).
3. Ray, J., Lefantzi, S., Arunajatesan, S. & Dechant, L. Bayesian parameter estimation of ak-ε model for accurate jet-in-crossflow simulations. *AIAA J.* **54**, 2432–2448 (2016).
4. Brandt, R. E. et al. Rapid semiconductor device characterization through Bayesian parameter estimation. *Joule* **1**, 843–856 (2017).
5. Kurchin, R. C. et al. How much physics is in a current–voltage curve? Inferring defect properties from photovoltaic device measurements. *IEEE J. Photovolt.* **10**, 1532–1537 (2020).
6. Aitio, A., Marquis, S. G., Ascencio, P. & Howey, D. Bayesian parameter estimation applied to the Li-ion battery single particle model with electrolyte dynamics. *IFAC-PapersOnLine* **53**, 12497–12504 (2020).
7. Wesolowski, S., Klco, N., Furnstahl, R., Phillips, D. & Thapaliya, A. Bayesian parameter estimation for effective field theories. *J. Phys. G Nucl. Part. Phys.* **43**, 074001 (2016).
8. Thrane, E. & Talbot, C. An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publ. Astron. Soc. Aust.* **36**, e010 (2019).
9. Loredo, T. J. In *Statistical Challenges in Modern Astronomy* (eds Feigelson, E. D. & Babu, G. J.) 275–297 (Springer, 1992).
10. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).

Acknowledgements

The author gratefully acknowledges J. Wang, J. Freudenburg, and P. Komiske for helpful conversations, suggestions and topical references.

Competing interests

The author declares no competing interests.