Title
Autocorrelated Sampling in Cognition: Implementing MCMC Algorithms as Cognitive Models and Fitting Them Without Likelihoods

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Castillo, Lucas
Qian, Cheng Stella
Sanborn, Adam

Publication Date
2026

Autocorrelated Sampling in Cognition:
Implementing MCMC Algorithms as Cognitive Models and Fitting Them Without Likelihoods

Lucas Castillo (lucas.castillo@warwick.ac.uk)[1], C. Stella Qian[1] & Adam N. Sanborn[1]
[1]Department of Psychology, University of Warwick

Keywords: Bayesian Models of Cognition; Rational Process Models; MCMC; Sampling; Simulation-based Inference

Overview

One of the fundamental questions in cognitive science is how people achieve such high levels of performance given the very limited resources they have access to. While human behaviour is similar to the Bayesian ideal both in low-level domains such as vision (Yuille & Kersten, 2006) and high-level domains such as categorization (Xu & Tenenbaum, 2007); optimality is in fact impossible given existing constraints.

Two factors make implementing Bayes' rule intractable: first, it requires pointwise multiplication of distributions (the likelihood and the prior), which is computationally expensive if done for all possible hypotheses. Second, it requires the area under the curve of the posterior to add up to 1, which in turn requires adding up the probability of the observed evidence given all possible hypotheses to obtain a normalizing constant.

To circumvent these problems while performing Bayesian inference, computer scientists have developed a family of algorithms, Markov Chain Monte Carlo (MCMC; Brooks et al., 2011), which operate by drawing samples directly from the posterior distribution (thus avoiding distribution multiplication). Samples are produced in an autocorrelated fashion (a new sample is created by perturbing the previous one with random noise), and added to a chain by evaluating its posterior density relative to the sample prior to it. This way, the two intractable tasks that Bayes' rule entails are substituted by two easy ones: modifying existing samples and comparing their relative 'goodness'. As a trade-off, samples are less informative (due to their autocorrelation), and the algorithms are sensitive to starting point biases.

An exciting development in cognitive science has been to show that these inference algorithms describe human behaviour remarkably well. MCMC algorithms have been used to explain human behaviour in perception (Gershman et al., 2012), causal reasoning (Bramley et al., 2017; Davis & Rehder, 2020), forecasting (Spicer et al., 2024), numerical estimates (Lieder et al., 2018) and probability judgments (Dasgupta et al., 2017), among others.

For example, Davis and Rehder's (2020) *mutation sampler* model explains human causal judgments as arising from people considering a hypothesis, evaluating it given the evidence, then changing the hypothesis slightly, for a limited number of iterations. This approach explains a key phenomenon of human causal judgements, *Markov violations*, by which people give weight to irrelevant evidence when judging the probability of an event (for example, if A causes B which causes C, the state of A should be ignored when estimating the probability of C given B). Similarly, Lieder et al. (2018) show how MCMC algorithms can explain the fact that people's numerical estimates (e.g., guessing the year Genghis Khan died) are influenced by preceding questions including a number (because a small number of autocorrelated samples is influenced by the starting point).

Despite MCMC algorithms having been successful in several domains, their uptake has been sluggish. We believe a key reason for this is the existing difficulty in their implementation. Researchers wanting to incorporate MCMC to their cognitive models will find that the algorithms themselves are somewhat tricky to code, and, in addition, that they should be implemented in a low-level programming language so that they can run quickly. What is more, models using autocorrelated sampling algorithms will lack an explicit likelihood function, and therefore model fitting will need to be carried out via simulation. While resources exist on how to do this (Cranmer et al., 2020; Turner & Van Zandt, 2012), this again constitutes an additional hurdle.

In this Tutorial, we aim to address these problems. We introduce attendees to the `samplr` package (Castillo et al., 2025), implemented in the R programming language (R Core Team, 2024) and `C++`, which includes different statistical sampling algorithms, as well as a cognitive process model which incorporates these into a larger architecture (the Autocorrelated Bayesian Sampler; Zhu et al., 2024), and a set of diagnostic functions which can be used to distinguish among different sampling approaches. We also provide an introduction as to how models without a tractable likelihood function can be modelled, using Approximate Bayesian Computation (Turner and Van Zandt, 2012; although other approaches are available).

Target Audience and Goals

Our Tutorial will be of interest to a broad audience, regardless of previous experience with MCMC algorithms. Attendees who have not worked with MCMC algorithms before will learn about how these algorithms operate, and will become equipped with a set of tools – the `samplr` package – for their easy implementation. More experienced attendees will find in this Tutorial a place to discuss different ap-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

proaches to fitting and comparing these models to human data. This is an open-source package (https://github.com/lucas-castillo/samplr), and experienced users will be able to guide its future development by suggesting features.

The only prerequisite will be for attendees to have a working knowledge of the R programming language.

Structure of the Workshop

This half-day workshop is divided into two sections, tackling the two hurdles previously identified. Materials and a more detailed schedule will be provided prior to the day. Attendees will need to install R and the `samplr` package prior to the workshop (available on CRAN); organizers will troubleshoot issues regarding software installation if necessary. To support larger groups, the hands-on components will be organized around short worked examples followed by structured exercises.

Section 1. Sampling

The main aim in this section will be to introduce and use MCMC algorithms.

- First, we will discuss sampling approaches in cognitive modelling, highlighting their success in several domains (high- and low-level cognition). This subsection will include invited talks by other researchers using these methods (e.g., Dr Neil Bramley, University of Edinburgh; Prof. Bob Rehder, New York University).
- In a second, more hands-on, segment, we will introduce the `samplr` package. This interactive component will involve attendees producing their own simulated data from the algorithms available in the package, and exploring the different patterns of behaviour they display by using the diagnostic tools that come with the package.

Section 2. Simulation-based Inference and Future Steps

The aim of this section will be to go from having observed data from experimental participants (provided by us) and simulated data from models (generated by attendees in section one), to having a way in which to do simulation-based model comparison.

- A first interactive part will have attendees carrying out their own model selection workflow. This will start with an introduction by us where we introduce several approaches to simulation-based inference before honing in onto a single one, Approximate Bayesian Computation (Turner & Van Zandt, 2012). Attendees will then practice using this approach to perform model selection on a set of real participant data.
- Finally, we will have a discussion touching on where future avenues for research using these methods may lead, as well as fostering collaborations between attendees, and between attendees and organizers.

Organizers

Lucas Castillo is a Research Fellow at the University of Warwick, and is the lead developer of the *samplr* package. **C. Stella Qian** is a Research Fellow at the University of Warwick. **Adam N. Sanborn** is a Professor at the University of Warwick.

Acknowledgements

The authors acknowledge funding from the United Kingdom's Economic and Social Research Council (ESRC; Grant No. UKRI1676), awarded to ANS.

References

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338. https://doi.org/10.1037/rev0000061

Brooks, S., Andrew Gelman, Jones, G., & Meng, X.-L. (Eds.). (2011). *Handbook for Markov chain Monte Carlo*. Taylor & Francis.

Castillo, L., Li, Y.-X., & Sanborn, A. N. (2025). *The samplr package: A tool for modeling human cognition with sampling algorithms*. https://doi.org/10.31234/osf.io/ax8hm_v1

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25. https://doi.org/10.1016/j.cogpsych.2017.05.001

Davis, Z. J., & Rehder, B. (2020). A process model of causal reasoning. *Cognitive Science*, *44*(5), e12839.

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and Perceptual Inference. *Neural Computation*, *24*(1), 1–24. https://doi.org/10.1162/NECO_a_00226

Lieder, F., Griffiths, T. L., M. Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, *25*(1), 322–349. https://doi.org/10.3758/s13423-017-1286-8

R Core Team. (2024). *R: A language and environment for statistical computing*. manual. Vienna, Austria, R Foundation for Statistical Computing. https://www.R-project.org/

Spicer, J., Zhu, J.-Q., Chater, N., & Sanborn, A. N. (2024). How do people predict a random walk? Lessons for models of human cognition. *Psychological Review*, *131*(5), 1069–1113. https://doi.org/10.1037/rev0000493

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*(2), 69–85. https://doi.org/10.1016/j.jmp.2012.02.005

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zhu, J.-Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2024). The autocorrelated Bayesian sampler. *Psychological Review*, *131*(2), 456–493. https://doi.org/10.1037/rev0000427