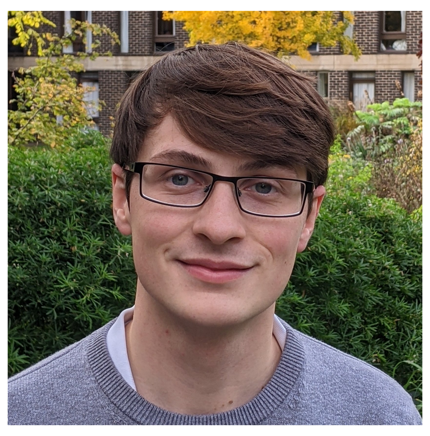
Fully Bayesian Forecasts with Neural Bayes Ratio Estimation

Thomas Gessey-Jones (tg400@cam.ac.uk) and Will Handley (wh260@cam.ac.uk)

0. Abstract



Sensitivity forecasts inform the design of experiments and the direction of theoretical efforts. To arrive at representative results, Bayesian forecasts should marginalize their conclusions over uncertain parameters and noise realizations rather than picking fiducial values. However, this is typically computationally infeasible with current methods for forecasts of an experiment's ability to distinguish between competing models. We thus propose a novel simulation-based methodology utilizing neural Bayes ratio estimators capable of providing expedient and rigorous Bayesian model comparison forecasts without relying on restrictive assumptions.

1. Motivation

In many scientific contexts, the question of interest is most clearly formulated as one of Bayesian model comparison [1, 2]. For example, if a signal has been detected (e.g. gravitational waves [3]) or if there is significant evidence for new physics (e.g. non-zero neutrino mass in CMB analysis [4]).

To design experiments to answer such questions, we thus need fast and versatile methods for forecasting the expected results from Bayesian model comparison analyses, which marginalize their conclusions over the uncertainties in potential data to ensure robust results. However, current methods are either too slow to be used to explore the potential data space (e.g., nested sampling on simulated data [5]) or have limited applicability due to restrictive assumptions (e.g., Savage–Dickey forecasts [6, 7]). We thus propose a new method using simulation-based inference and neural Bayes ratio estimation, which simultaneously addresses the above problems and no longer requires an explicit likelihood.

2. Neural Bayes Ratio Estimation

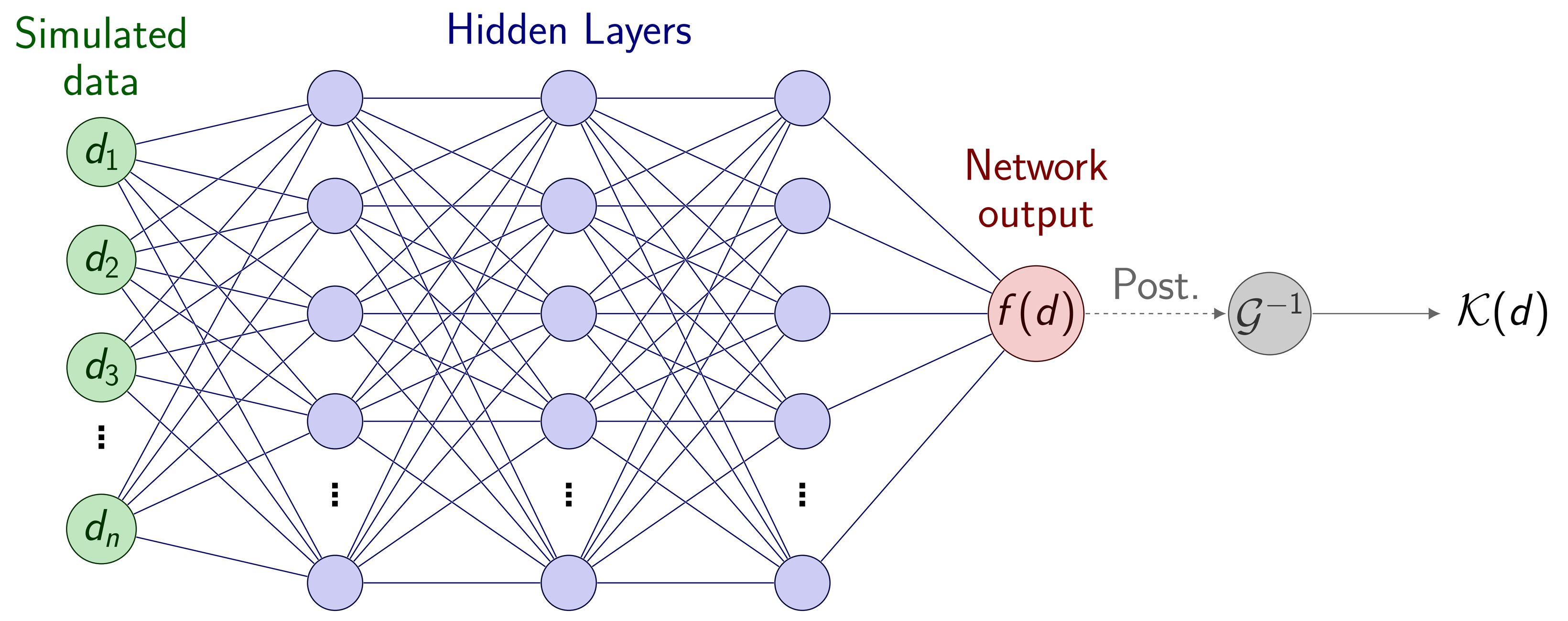


Fig. 1: Illustration of a neural Bayes ratio estimator (a.k.a., evidence network). The network is trained to classify between simulated data generated from two competing models. For this training, the loss function used is intentionally chosen such that the output of the network converges to a known invertible function $\mathcal{G}$ of the Bayes ratio between the two models $\mathcal{K}(d)$, allowing the network output to be transformed directly into $\mathcal{K}$. Introduced in Jeffrey and Wandelt [8], these networks have been shown to be able to calculate accurate $\mathcal{K}$ for actual scientific problems after training times less than the cost of a single nested sampling run.

3. Fully Bayesian Forecast Methodology

1. Create data simulators for the two competing models (including experimental considerations such as noise or selection effects).
2. Using the two simulators, generate training, validation, and testing sets of labelled simulated data and train a neural Bayes ratio estimator.
3. Validate the network using a blind coverage test [8] to verify the network's accuracy (see Fig 2.) and, if possible, also compare its output to a sample of $\mathcal{K}$ values calculated from traditional Bayesian techniques (see Fig 3.).
4. Exploiting the amortized nature of neural Bayes ratio estimators evaluate $\mathcal{K}$ over the parameter and noise space (or equivalently over the uncertainties in potential data).
5. Using these $\mathcal{K}$ samples marginalize the conclusions of the Bayesian model comparison forecast to arrive at a representative result.

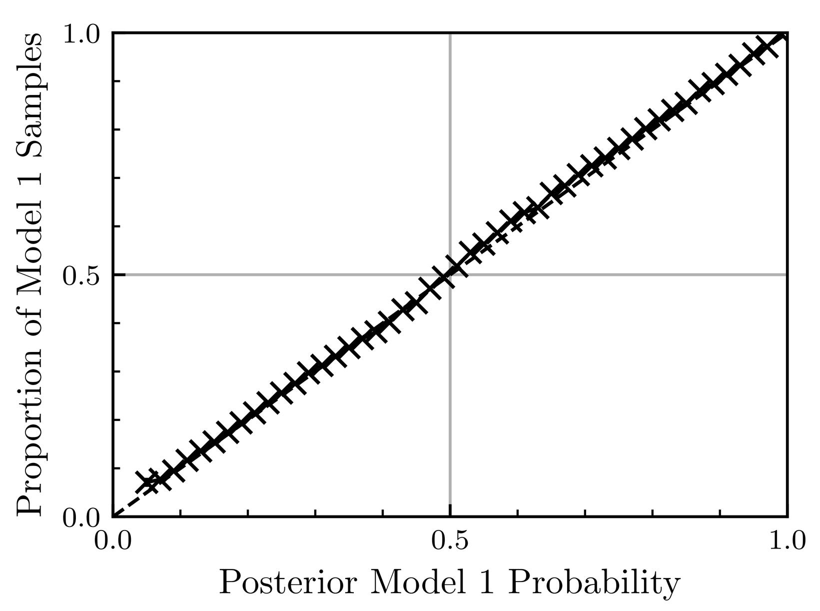


Fig. 2: Blind coverage test verifying the accuracy of a converged neural Bayes ratio estimator. See Jeffrey and Wandelt [8] for details.

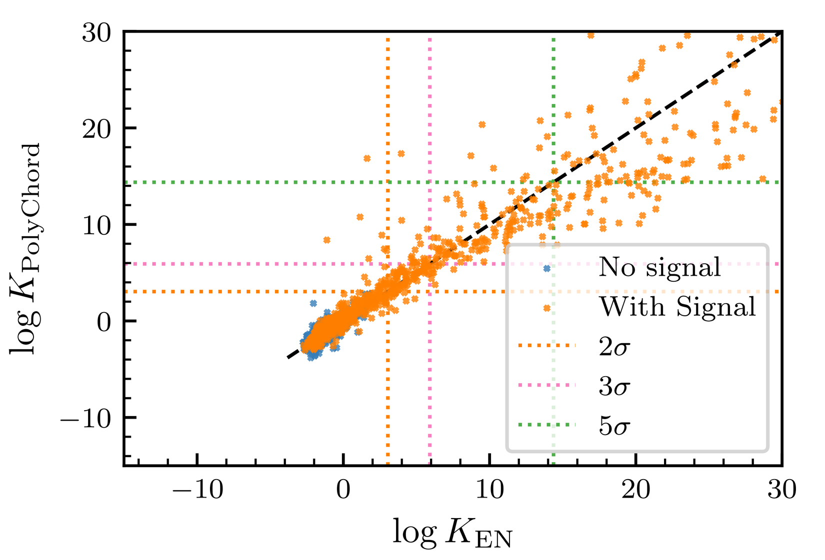


Fig. 3: Comparison of $\mathcal{K}$ values calculated using a neural Bayes ratio estimator and the nested-sampling code `Polychord`, showing good agreement out to 5σ significance.

4. Example - Detecting the 21-cm Global Signal

To illustrate the methodology, we forecast the expected probability of a REACH-like [9] experiment detecting the global 21-cm signal at $\geq 3\sigma$ statistical significance given our current knowledge of high-redshift astrophysics and the experimental foregrounds. We find a 46% chance of making this detection (assuming a noise level between the pessimistic and expected case outlined in the REACH mission paper). Furthermore, our methodology facilitates an investigation of how this detection chance varies with uncertain astrophysics (see Fig 4.), and hence the features of the 21-cm signal the experiment is sensitive to. This analysis took only 5.54 GPU hours to perform, a $10^{5.2}$ cost-weighted speed-up compared to the 45,000,000 CPU hours an equivalent nested-sampling-based approach would have required.

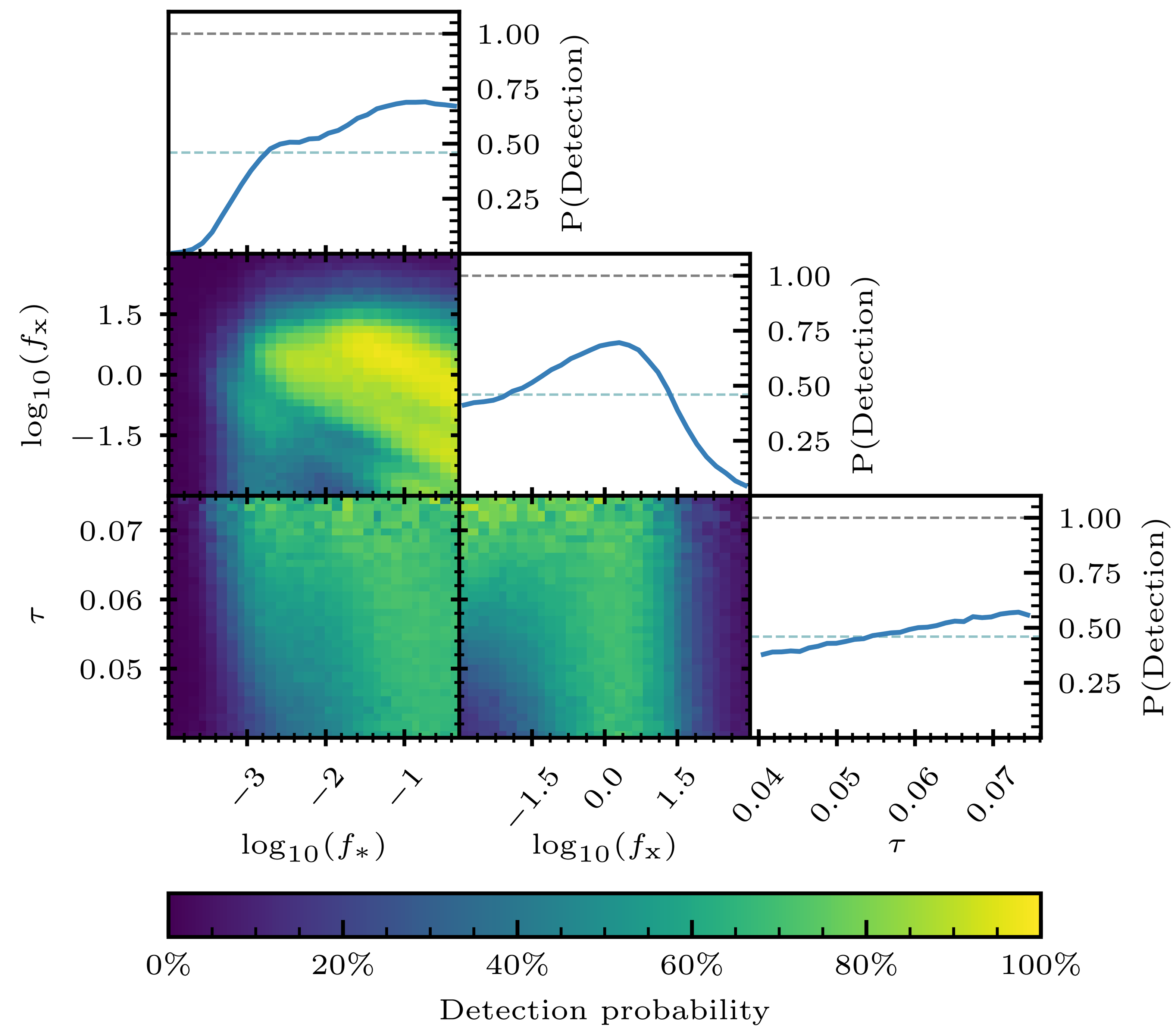


Fig. 4: Triangle plot depicting the variation of the probability of a global 21-cm signal detection at $\geq 3\sigma$ statistical significance by a REACH-like experiment for select unknown astrophysical parameters. The observation band is assumed to cover the redshift range 7.5 to 28.0, with frequency resolution $\Delta\nu = 1\,$MHz. We adopt the Hills et al. [10] physical foreground model, assume white noise of $\sigma_{\mathrm{noise}} = 0.015\,$K, and use globalemu [11] as our global 21-cm signal model.

5. Conclusions and Further Directions

Neural Bayes ratio estimators make fully Bayesian forecasts of model comparison questions computationally feasible without the need for restrictive assumption or even an explicit likelihood. In a follow-up study, we shall illustrate how the outlined methodology can be used to optimize an experiment's design to maximize the probability of getting a definitive answer to the scientific question of interest.

6. References, Our Paper, and Code

[1] Pia Mukherjee, David Parkinson, Pier Stefano Corasaniti, et al. In: *MNRAS* 369.4 (July 2006), pp. 1725–1734.

[2] Roberto Trotta. In: *MNRAS* 378.1 (June 2007), pp. 72–82.

[3] Matthew Evans, Alessandra Corsi, Chaitanya Afle, et al. In: arXiv:2306.13745 (June 2023).

[4] Peter Ade, James Aguirre, Zeeshan Ahmed, et al. In: *JCAP* 2019.2, 056 (Feb. 2019), p. 056.

[5] Greg Ashton, Noam Bernstein, Johannes Buchner, et al. In: *Nature Reviews Methods Primers* 2, 39 (May 2022), p. 39.

[6] James M. Dickey. In: *The Annals of Mathematical Statistics* 42.1 (1971), pp. 204–223.

[7] Roberto Trotta. In: *MNRAS* 378.3 (July 2007), pp. 819–824.

[8] Niall Jeffrey and Benjamin D. Wandelt. In: *Machine Learning: Science and Technology* 5.1, 015008 (Mar. 2024), p. 015008.

[9] E. de Lera Acedo, D. I. L. de Villiers, N. Razavi-Ghods, et al. In: *Nature Astronomy* 6 (July 2022), pp. 984–998.

[10] Richard Hills, Girish Kulkarni, P. Daniel Meerburg, et al. In: *Nature* 564.7736 (Dec. 2018), E32–E34.

[11] H. T. J. Bevins, W. J. Handley, A. Fialkov, et al. In: *MNRAS* 508.2 (Dec. 2021), pp. 2923–2936.

Paper:

arXiv:2309.06942

Code:

github.com/ThomasGesseyJones