

SBIAX: Density-estimation simulation-based inference in JAX.

Jed Homer^{1,2*} and Oliver Friedrich^{1,2,3*}

¹ University Observatory, Faculty for Physics, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, München, Deutschland. ² Munich Center for Machine Learning. ³ Excellence Cluster ORIGINS, Boltzmannstr. 2, 85748 Garching, Deutschland. * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Open Journals](#)

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

In partnership with



This article and software are linked with research article DOI [10.3847/xxxxx](https://doi.org/10.3847/xxxxx) <- [update this with the DOI from AAS once you know it.](#) published in the Astrophysical Journal <- The name of the AAS journal..

Summary

In a typical Bayesian inference problem, the data likelihood is not known. However, in recent years, machine learning methods for density estimation can allow for inference using an estimator of the data likelihood. This likelihood is created with neural networks that are trained on simulations - one of the many tools for Simulation Based Inference (SBI), ([Cranmer et al., 2020](#)). In such analyses, density-estimation simulation-based inference methods can derive a posterior, which typically involves

- simulating a set of data and model parameters $\{(\xi, \pi)_0, \dots, (\xi, \pi)_N\}$,
- obtaining a measurement $\hat{\xi}$,
- compressing the simulations and the measurements - usually with a neural network or linear compression - to a set of summaries $\{(x, \pi)_0, \dots, (x, \pi)_N\}$ and \hat{x} ,
- fitting an ensemble of normalising flow or similar density estimation algorithms (e.g. a Gaussian mixture model),
- the optional optimisation of the parameters for the architecture and fitting hyper-parameters of the algorithms,
- sampling the ensemble posterior (using an MCMC sampler if the likelihood is fit directly), conditioned on the data-vector, to obtain parameter constraints on the parameters of a physical model, π .

sbi-ax is a software package that implements each of these steps. The code allows for Neural Likelihood Estimation ([Alsing et al., 2019](#); [Papamakarios, 2019](#)), and Neural Posterior Estimation ([Greenberg et al., 2019](#)).

As shown in Homer et al. (2024), SBI can successfully obtain the correct posterior widths and coverages given enough simulations which agree with the analytic solution - this software was used in the research for this publication.

Statement of need

Simulation Based Inference (SBI) covers a broad class of statistical techniques such as Approximate Bayesian Computation (ABC) ([Rubin, 1984](#)), Neural Ratio Estimation (NRE), ([Delaunoy et al., 2022](#)), Neural Likelihood Estimation (NLE) and Neural Posterior Estimation (NPE). These techniques can derive posterior distributions conditioned of noisy data vectors in a rigorous and efficient manner with assumptions on the data likelihood. In particular, density-estimation methods have emerged as a promising method, given their efficiency, using generative models to fit likelihoods or posteriors directly using simulations.

In the field of cosmology, SBI is of particular interest due to complexity and non-linearity of models for the expectations of non-standard summary statistics of the large-scale structure, as

well as the non-Gaussian noise distributions for these statistics. The assumptions required for the complex analytic modelling of these statistics as well as the increasing dimensionality of data returned by spectroscopic and photometric galaxy surveys, limit the amount of information that can be obtained on fundamental physical parameters. Therefore, the study and research into current and future statistical methods for Bayesian inference is of paramount importance for the cosmology, especially in light of current and next-generation survey missions such as DES (Laureijs et al., 2011), DESI (Levi et al., 2019), and Euclid (Laureijs et al., 2011).

The software we present, sbi, is designed to be used by machine learning and physics researchers for running Bayesian inferences using density-estimation SBI techniques. These models can be fit easily with multi-accelerator training and inference within the code. This software - written in jax (Bradbury et al., 2018) - allows for seamless integration of cutting edge generative models to SBI, including continuous normalising flows (Grathwohl et al., 2018), matched flows (Lipman et al., 2023), masked autoregressive flows (Papamakarios et al., 2018; Ward, 2024), and Gaussian mixture models - all of which are implemented in the code. The code features integration with the optuna (Akiba et al., 2019) hyper-parameter optimisation framework which would be used to ensure consistent analyses, blackjax (Cabezas et al., 2024) for fast MCMC sampling, and equinox (Kidger & Garcia, 2021) for neural network methods. The design of sbi allows for new density estimation algorithms to be trained and sampled from, as long as they conform to a simple and typical design pattern demonstrated in sbi.

Whilst excellent software packages already exist for conducting simulation-based inference (e.g. sbi (Tejero-Cantero et al., 2020), sbijax (Dirmeir, 2024)) for some applications it is useful to have a lightweight implementation that focuses on speed, ensembling of density estimators and easily integrated MCMC sampling (e.g. for ensembles of likelihoods) - all of which is based on a lightweight and regularly maintained jax machine learning library such as equinox (Kidger & Garcia, 2021). sbi depends on density estimators and compression modules - as long as log-probability and callable methods exists for these, they can be integrated seamlessly.

Density estimation with normalising flows

The use of density-estimation in SBI has been accelerated by the advent of normalising flows. These models parameterise a change-of-variables $y = f_\phi(x; \pi)$ between a simple base distribution (e.g. a multivariate unit Gaussian $\mathcal{G}[z|\mathbf{0}, \mathbf{I}]$) and an unknown distribution $q(x|\pi)$ (from which we have simulated samples x). Naturally, this is of particular importance in inference problems in which the likelihood is not known. The change-of-variables is fit from data by training neural networks to model the transformation in order to maximise the log-likelihood of the simulated data x conditioned on the parameters π of a simulator model. The mapping is expressed as

$$y = f_\phi(x; \pi),$$

where ϕ are the parameters of the neural network. The log-likelihood of the flow is expressed as

$$\log p_\phi(x|\pi) = \log \mathcal{G}[f_\phi(x; \pi)|\mathbf{0}, \mathbf{I}] + \log |\mathbf{J}_{f_\phi}(x; \pi)|,$$

This density estimate is fit to a set of N simulation-parameter pairs $\{(\xi, \pi)_0, \dots, (\xi, \pi)_N\}$ by minimising a Monte-Carlo estimate of the KL-divergence

$$\begin{aligned}
 \langle D_{KL}(q||p_\phi) \rangle_{\pi \sim p(\pi)} &= \int d\pi p(\pi) \int dx q(x|\pi) \log \frac{q(x|\pi)}{p_\phi(x|\pi)}, \\
 &= \int d\pi \int dx p(\pi, x) [\log q(x|\pi) - \log p_\phi(x|\pi)], \\
 &\geq - \int d\pi \int dx p(\pi, x) \log p_\phi(x|\pi), \\
 &\approx - \frac{1}{N} \sum_{i=1}^N \log p_\phi(x_i|\pi_i),
 \end{aligned} \tag{1}$$

where $q(x|\pi)$ is the unknown likelihood from which the simulations x are drawn. This applies similarly for an estimator of the posterior (instead of the likelihood as shown here) and is the basis of being able to estimate the likelihood or posterior directly when an analytic form is not available. If the likelihood is fit from simulations, a prior is required and the posterior is sampled via an MCMC-sampler given some measurement. This is implemented within the code.

An ensemble of density estimators (with parameters - e.g. the weights and biases of the networks - denoted by $\{\phi_0, \dots, \phi_J\}$) has a likelihood which is written as

$$p_{\text{ensemble}}(\xi|\pi) = \sum_{j=1}^J \alpha_j p_{\phi_j}(\hat{\xi}|\pi)$$

where

$$\alpha_i = \frac{\exp(p_{\phi_i}(\hat{\xi}|\pi))}{\sum_{j=1}^J \exp(p_{\phi_j}(\hat{\xi}|\pi))}$$

are the weights of each density estimator in the ensemble. This ensemble likelihood can be easily sampled with an MCMC sampler. In Figure 1 we show an example posterior from applying SBI, with our software, using two compression methods separately.

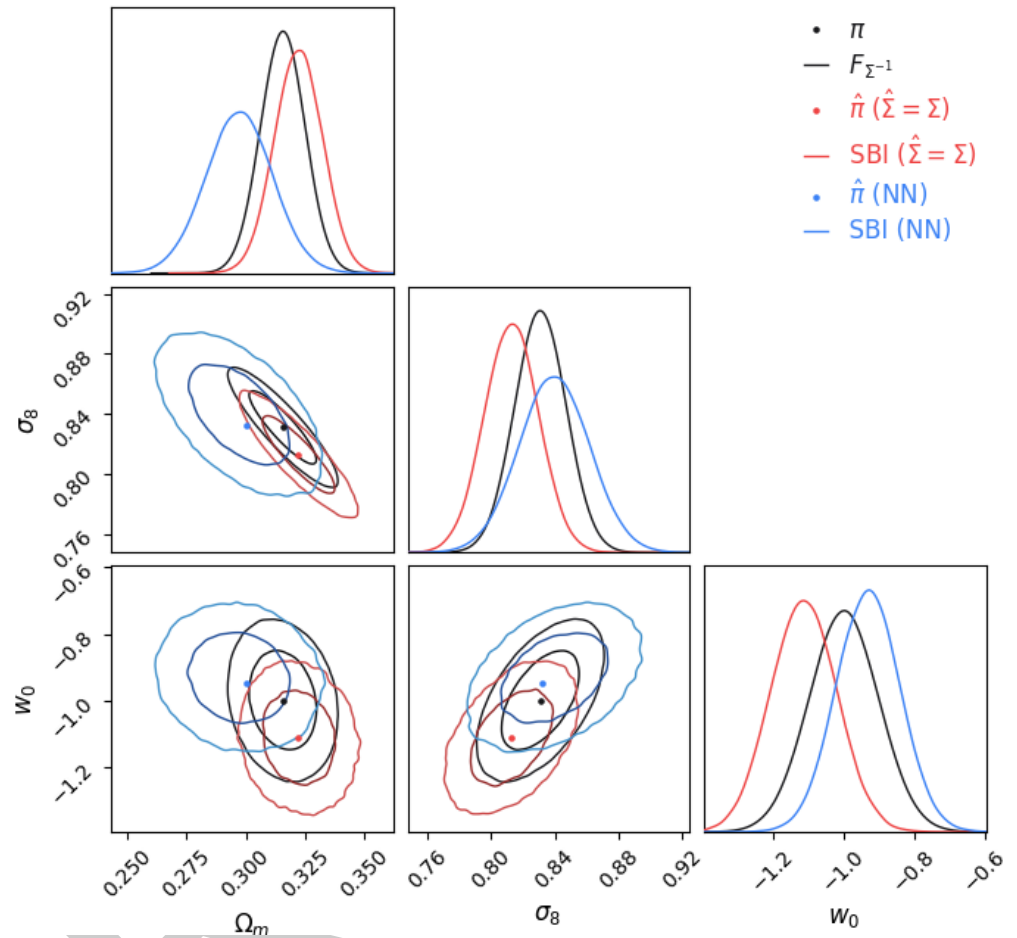


Figure 1: An example of posteriors derived with sbiax. We fit an ensemble of two continuous normalising flows to a set of simulations of cosmic shear two-point functions. The expectation $\xi[\pi]$ is linearised with respect to π and a theoretical data covariance model Σ (in this example) allows for easy sampling of many simulations - an ideal test arena for SBI methods. We derive two posteriors, from separate experiments, where a linear (red) or neural network compression (blue) is used. In black, the true analytic posterior is shown. Note that for a finite set of simulations the blue posterior will not overlap completely with the black and red posteriors - we explore this effect upon the posteriors from SBI methods, due to an unknown data covariance, in Homer et al. (2024).

Acknowledgements

We thank the developers of the packages jax (Bradbury et al., 2018), blackjax (Cabezas et al., 2024), optax (DeepMind et al., 2020), equinox (Kidger & Garcia, 2021), diffrax (Kidger, 2022) and flowjax (Ward, 2024) for their work and for making their code available to the community.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631.

- Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. (2019). Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*. <https://doi.org/10.1093/mnras/stz1960>
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., & Zhang, Q. (2018). *JAX: Composable transformations of Python+NumPy programs* (Version 0.3.13). <http://github.com/jax-ml/jax>
- Cabezas, A., Corenflos, A., Lao, J., & Louf, R. (2024). *BlackJAX: Composable Bayesian inference in JAX*. <https://arxiv.org/abs/2402.10797>
- 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>
- DeepMind, Babuschkin, I., Baumli, K., Bell, A., Bhupatiraju, S., Bruce, J., Buchlovsky, P., Budden, D., Cai, T., Clark, A., Danihelka, I., Dedieu, A., Fantacci, C., Godwin, J., Jones, C., Hemsley, R., Hennigan, T., Hessel, M., Hou, S., ... Viola, F. (2020). *The DeepMind JAX Ecosystem*. <http://github.com/google-deepmind>
- Delaunoy, A., Hermans, J., Rozet, F., Wehenkel, A., & Louppe, G. (2022). *Towards reliable simulation-based inference with balanced neural ratio estimation*. <https://arxiv.org/abs/2208.13624>
- Dirmeir, S. (2024). *SBIJAX: Simulation-based inference in JAX*. (Version 0.3.0). <https://github.com/dirmeir/sbijax>
- Grathwohl, W., Chen, R. T. Q., Bettencourt, J., Sutskever, I., & Duvenaud, D. (2018). *FFJORD: Free-form continuous dynamics for scalable reversible generative models*. <https://arxiv.org/abs/1810.01367>
- Greenberg, D. S., Nonnenmacher, M., & Macke, J. H. (2019). *Automatic posterior transformation for likelihood-free inference*. <https://arxiv.org/abs/1905.07488>
- Homer, J., Friedrich, O., & Gruen, D. (2024). *Simulation-based inference has its own Dodelson-Schneider effect (but it knows that it does)*. <https://arxiv.org/abs/2412.02311>
- Kidger, P. (2022). *On neural differential equations*. <https://arxiv.org/abs/2202.02435>
- Kidger, P., & Garcia, C. (2021). Equinox: Neural networks in JAX via callable PyTrees and filtered transformations. *Differentiable Programming Workshop at Neural Information Processing Systems 2021*.
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. -L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., ... Zucca, E. (2011). *Euclid definition study report*. <https://arxiv.org/abs/1110.3193>
- Levi, M. E., Allen, L. E., Raichoor, A., Baltay, C., BenZvi, S., Beutler, F., Bolton, A., Castander, F. J., Chuang, C.-H., Cooper, A., Cuby, J.-G., Dey, A., Eisenstein, D., Fan, X., Flaugher, B., Frenk, C., Gonzalez-Morales, A. X., Graur, O., Guy, J., ... Zu, Y. (2019). *The dark energy spectroscopic instrument (DESI)*. <https://arxiv.org/abs/1907.10688>
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2023). *Flow matching for generative modeling*. <https://arxiv.org/abs/2210.02747>
- Papamakarios, G. (2019). *Neural density estimation and likelihood-free inference*. <https://arxiv.org/abs/1910.13233>
- Papamakarios, G., Pavlakou, T., & Murray, I. (2018). *Masked autoregressive flow for density estimation*. <https://arxiv.org/abs/1705.07057>

- 148 Rubin, D. B. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the
149 Applied Statistician. *The Annals of Statistics*, 12(4), 1151–1172. [https://doi.org/10.1214/
150 aos/1176346785](https://doi.org/10.1214/aos/1176346785)
- 151 Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J.,
152 Greenberg, D. S., & Macke, J. H. (2020). Sbi: A toolkit for simulation-based inference.
153 *Journal of Open Source Software*, 5(52), 2505. <https://doi.org/10.21105/joss.02505>
- 154 Ward, D. (2024). *FlowJAX: Distributions and normalizing flows in JAX* (Version 16.0.0).
155 <https://doi.org/10.5281/zenodo.10402073>

DRAFT