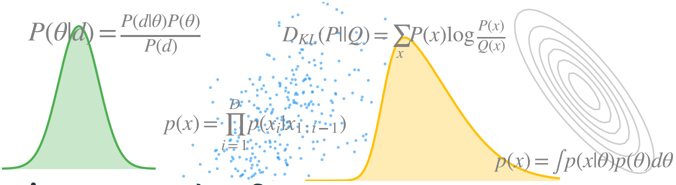


Simulation-Based Inference

An Overview for Applications in Astrophysics & Cosmology

Maximilian von Wietersheim-Kramsta — mwiet.github.io

SBI in Galaxy Evolution 2025, University of Bristol — 27th of May 2025

Institute for Computational Cosmology & Centre for Extragalactic Astronomy, Durham University

Institute for Computational Cosmology

Durham Centre for Extragalactic Astronomy

Durham University

Simulation-Based Inference (SBI): Contents

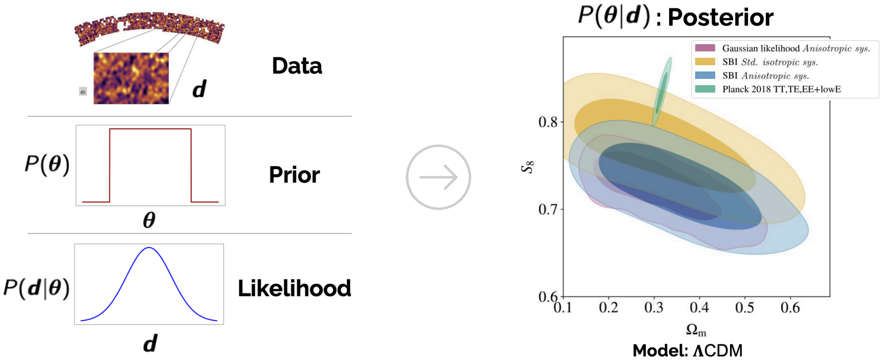
SBI: Motivation & Background

Bayesian Inference

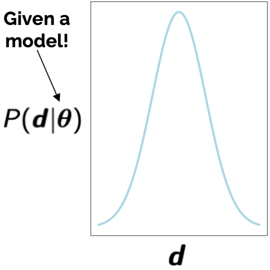
Probability of a statement being true given an update to a prior belief and a specific model.

$\Rightarrow$ Bayes' Theorem:

$$P(\boldsymbol{\theta} \mid \boldsymbol{d}) = \frac{P(\boldsymbol{d} \mid \boldsymbol{\theta})\, P(\boldsymbol{\theta})}{P(\boldsymbol{d})} \tag{1}$$



Modelling Likelihood Distributions



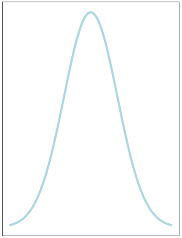
Analytic

e.g. $P(d) \propto \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)$

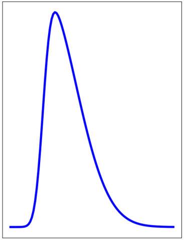
Modelling Likelihood Distributions

Given a model!

$P(\boldsymbol{d}|\boldsymbol{\theta})$



Analytic

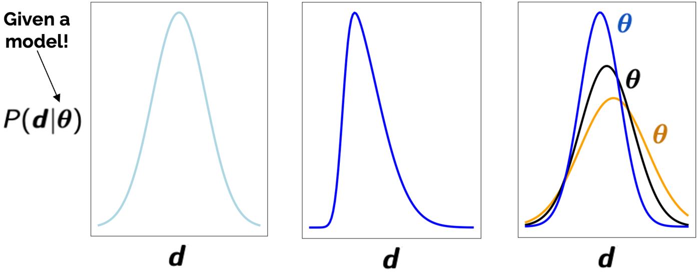


Biased

e.g. $P(d) \propto \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)$

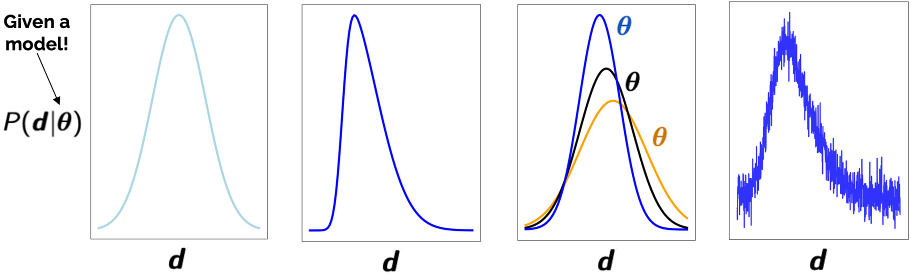
e.g. Systematics

Modelling Likelihood Distributions



Analytic	**Biased**	**Signal-dependent uncertainty**
e.g. $P(d) \propto \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)$	e.g. Systematics	e.g. Non-random noise

Modelling Likelihood Distributions



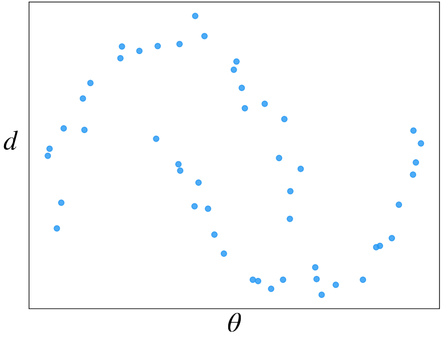
Analytic	**Biased**	**Signal-dependent uncertainty**	**Intractable**
e.g. $P(d) \propto \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)$	e.g. Systematics	e.g. Non-random noise	e.g. Many sources of uncertainty

Zooming Out: the Joint Probability

$$P(\boldsymbol{\theta} \mid \boldsymbol{d}) = \frac{P(\boldsymbol{d} \mid \boldsymbol{\theta})\, P(\boldsymbol{\theta})}{P(\boldsymbol{d})} \propto P(\boldsymbol{\theta}, \boldsymbol{d})\, P(\boldsymbol{\theta}) \tag{2}$$

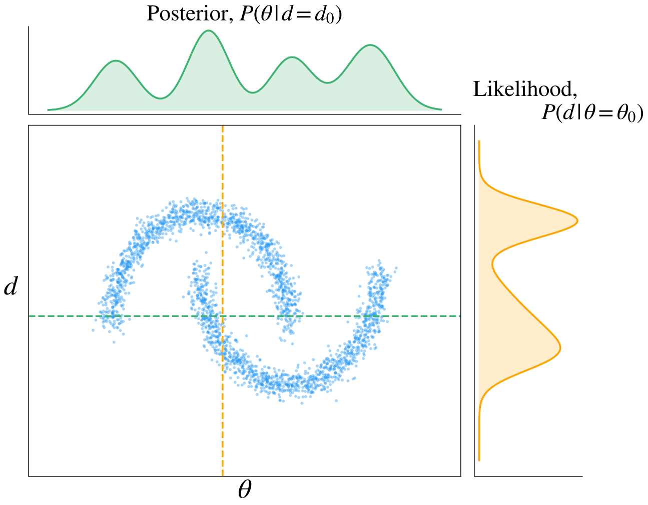
Joint probability: $P(\boldsymbol{\theta}, \boldsymbol{d} \mid \mathrm{Model}\,)$

Simulator: $\boldsymbol{d}_i \sim P(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathrm{Model})$



Zooming Out: the Joint Probability

Joint probability: $P(\boldsymbol{\theta}, \boldsymbol{d} \mid \text{Model})$



Generalised Bayesian Inference (GBI)

Standard Bayesian inference: Derived from Boolean logic when introducing uncertainty
$\rightarrow$ **Joint probability**, $P(\boldsymbol{\theta}, \boldsymbol{d})$ defines everything for a given model.

Generalised Bayesian Inference: Considers the case of imperfect models $\rightarrow$ Based on Decision theory Berger (2013)
$\rightarrow$ Finds optimal belief given an **imperfect model** and a specific goal
$\rightarrow$ Model is characterised through a **Loss function**, $L(\boldsymbol{\theta}, \boldsymbol{d})$, Bissiri et al. (2013):

$$\underbrace{P_G(\boldsymbol{\theta} \mid \boldsymbol{d})}_{\text{Generalised Posterior}} \propto \underbrace{\exp(-\eta L(\boldsymbol{\theta}, \boldsymbol{d}))}_{\text{Generalised Likelihood}} \times \underbrace{P(\boldsymbol{\theta})}_{\text{Prior}}, \tag{3}$$

where η quantifies the degree of match/mismatch between the data and the model ($\eta = 1$ corresponds to standard Bayes' theorem).
$L(\boldsymbol{\theta}, \boldsymbol{d}) = -\ln P(\boldsymbol{d} \mid \boldsymbol{\theta})$ returns Bayes' theorem.

Types of SBI

Types of Simulation-Based Inference

a.k.a. Likelihood-free or implicit likelihood inference

- Approximate Bayesian Computation (ABC)
- Neural Density Estimation (NDE)
    - Neural Posterior Estimation (NPE)
    - Neural Likelihood Estimation (NLE)
    - Neural Ratio Estimation (NRE)
    - Neural Posterior Score Estimation (NPSE)
- Sequential Methods

Simplest Case: Approximate Bayesian Computation

1. Draw simulations from simulator within prior space:

$$\boldsymbol{d}^* \sim P(\boldsymbol{d} \mid \boldsymbol{\theta}^*);\ \ \boldsymbol{\theta}^* \sim P(\boldsymbol{\theta}). \tag{4}$$

2. Define a distance metric between simulated data, $\boldsymbol{d}^*$, and observed data, $\boldsymbol{d}_0$ (e.g. Euclidean):

$$D = D(\boldsymbol{d}_0,\ \boldsymbol{d}^*). \tag{5}$$

3. Accept or reject according to arbitrary threshold, ϵ and repeat Rubin (1984):

$$\mathrm{if}\, D < \epsilon, \mathrm{keep}\, \boldsymbol{\theta}^*. \tag{6}$$

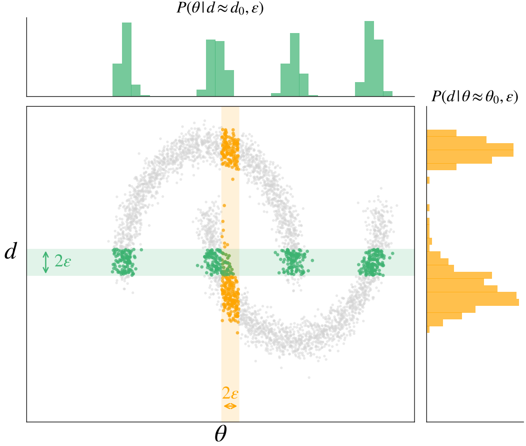
In the context of GBI, the Loss function is:

$$L(\boldsymbol{\theta}, \boldsymbol{d}) = \begin{cases} 0, & \text{if } D < \epsilon, \\ \infty, & \text{otherwise.} \end{cases} \tag{7}$$

Simplest Case: Approximate Bayesian Computation

Converges given:

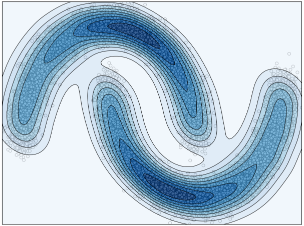
$$\lim_{\epsilon \to 0} P_{\mathrm{ABC}}(\theta \mid d_0) = P(\theta \mid d_0). \tag{8}$$



Simplest Case: Approximate Bayesian Computation

- **Curse of Dimensionality**: Performance degrades significantly with the dimensionality the data vector and model.
- Requires significant **compression**.
- The results sensitive to the choice of distance metric and ϵ.
- **Inefficiency**: Large number of simulations may be rejected, especially for small ϵ.
- **Not amortised**: Reevaluation required for each new d_0.
- **Recent applications**: e.g. galaxy morphology (Cameron and Pettitt, 2012; Tortorelli et al., 2021), diffuse X-ray background (Baxter et al., 2022), galaxy-scale strong lenses (He et al., 2022)

Neural Posterior Estimation (NPE)



$D_{KL}(P \,||\, Q) =$

$\sum_x P(x) \log \frac{P(x)}{Q(x)}$
See Papamakarios and Murray (2016); Lueckmann et al. (2017); Greenberg et al. (2019); Cranmer et al. (2020)

1. Draw simulations:

$$\boldsymbol{d}^* \sim P(\boldsymbol{d} \mid \boldsymbol{\theta}^*); \;\; \boldsymbol{\theta}^* \sim P(\boldsymbol{\theta}). \tag{9}$$

2. Find an estimator of the posterior, $\hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d})$, with its weights, $\boldsymbol{w}$, such that:

$$\boldsymbol{w}^* = \arg\min_w \mathbb{E}_{P(\boldsymbol{d})}[\, D_{\mathrm{KL}}(P(\boldsymbol{\theta} \mid \boldsymbol{d}) \,||\, \hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d}))], \tag{10}$$

$$\boldsymbol{w}^* = \arg\max_w \mathbb{E}_{P(\boldsymbol{\theta},\boldsymbol{d})}[\, \ln(\hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d}))]. \tag{11}$$

3. Train a neural network from this loss function:

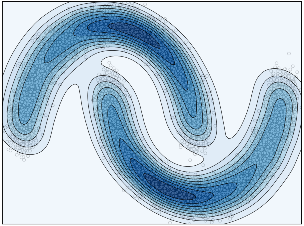
$$L(\boldsymbol{w}) = -\mathbb{E}_{P(\boldsymbol{\theta},\boldsymbol{d})}[\, \ln(\hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d}))] \tag{12}$$

4. Use network to directly sample $\hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d})$.

Neural Posterior Estimation (NPE)

- **Amortised***: Direct mapping from $\boldsymbol{d}$ to $P(\boldsymbol{\theta} \mid \boldsymbol{d})$.
- ***Amortisation gap**: If observed data, d_0, changes, $\hat{P}_w(\boldsymbol{\theta} \mid \boldsymbol{d})$ may not be well characterised at new $d_0 \rightarrow$ Additional training required.
- ***Prior-dependent**: For new parameters priors, retraining is necessary.
- **Recent applications**: e.g. galaxy clustering (Lemos et al., 2023b), exoplanets (Vasist et al., 2023), gravitational waves (Leyde et al., 2024), X-ray spectra (Barret and Dupourqué, 2024), lensed quasars (Erickson et al., 2024)
- **Implemented in**: `sbi` (Tejero-Cantero et al., 2020), `lampe` (Rozet et al., 2021) and `ltu-ili` (Ho et al., 2024).

Neural Likelihood Estimation (NLE)



$D_{KL}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$

See Papamakarios and Murray (2016); Papamakarios et al. (2017); Alsing et al. (2019); Cranmer et al. (2020); Boelts et al. (2022); Glöckler et al. (2022)

1. Draw simulations:

$$\boldsymbol{d}^* \sim P(\boldsymbol{d} \mid \boldsymbol{\theta}^*); \;\; \boldsymbol{\theta}^* \sim P(\boldsymbol{\theta}). \tag{13}$$

2. Find an estimator of the likelihood, $\hat{P}_w(\boldsymbol{d} \mid \boldsymbol{\theta})$, with its weights, $\boldsymbol{w}$, such that:

$$\boldsymbol{w}^* = \arg\min_w \mathbb{E}_{P(\boldsymbol{d})}[\, D_{\mathrm{KL}}(P(\boldsymbol{d} \mid \boldsymbol{\theta}) \,||\, \hat{P}_w(\boldsymbol{d}|\boldsymbol{\theta}))], \tag{14}$$

$$\boldsymbol{w}^* = \arg\max_w \mathbb{E}_{P(\boldsymbol{\theta},\boldsymbol{d})}[\, \ln(\hat{P}_w(\boldsymbol{d}|\boldsymbol{\theta}))]. \tag{15}$$

3. Train a neural network from this loss function:

$$L(\boldsymbol{w}) = -\mathbb{E}_{P(\boldsymbol{\theta},\boldsymbol{d})}[\, \ln(\hat{P}_w(\boldsymbol{d}|\boldsymbol{\theta}))]. \tag{16}$$

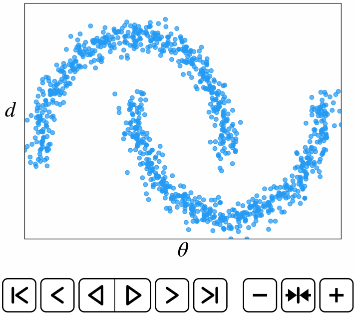
4. Define priors, $P(\boldsymbol{\theta})$, and sample posterior with an MCMC.

Neural Likelihood Estimation (NLE)

- **Prior-independent**: Learnt likelihood can be reused for sampling when varying the priors.
- **Useful for model testing**: Likelihood evaluation allows for model comparison.
- **Sampling required**: Requires sampling with MCMC or other sampler to obtain posteriors.
- **Compression**: Can struggle with high-dimensional data $\rightarrow$ Requiring compression.
- **Recent applications**: e.g. weak lensing (Jeffrey et al., 2024; von Wietersheim-Kramsta et al., 2025), seismology (Saoulis et al., 2025)
- **Implemented in**: `sbi` (Tejero-Cantero et al., 2020), `delfi` (Alsing et al., 2019) and `ltu-ili` (Ho et al., 2024).

NDE: Normalising Flows

- Used for **NPE** and **NLE**, as they naturally encode the normalisation of the probability densities (Papamakarios et al., 2021).
- Learns **invertible and differentiable** transformations between any distribution and a Gaussian.
- Usually train ensembles in parallel for robustness.
- Examples: Masked Autoregressive Flows (MAF), Neural Spline Flows (NSF), etc.



Neural Ratio Estimation (NRE)

1. Typically, learn the likelihood-to-evidence:

$$r(\boldsymbol{d} \mid \boldsymbol{\theta}) = P(\boldsymbol{d} \mid \boldsymbol{\theta})/P(\boldsymbol{d}) = P(\boldsymbol{\theta} \mid \boldsymbol{d})/P(\boldsymbol{\theta}). \tag{17}$$

2. Train a neural classifier, $Q_w(\boldsymbol{d}, \boldsymbol{\theta})$, to distinguish draws from:
   - The "true" joint distribution $P(\boldsymbol{d}, \boldsymbol{\theta}) = P(\boldsymbol{d} \mid \boldsymbol{\theta})\, P(\boldsymbol{\theta})$.
   - The product of the marginal distributions $P(\boldsymbol{d})\, P(\boldsymbol{\theta})$.
3. When optimised, we find:

$$Q^*(\boldsymbol{d}, \boldsymbol{\theta}) = P(\boldsymbol{d}, \theta)/[P(\boldsymbol{d}, \boldsymbol{\theta}) + P(\boldsymbol{d})P(\boldsymbol{\theta})]. \tag{18}$$

$$\hat{r}(\boldsymbol{d} \mid \boldsymbol{\theta}) = \frac{Q_w(\boldsymbol{d}, \boldsymbol{\theta})}{1 - Q_w(\boldsymbol{d}, \boldsymbol{\theta})} \approx \frac{P(\boldsymbol{d} \mid \boldsymbol{\theta})}{P(\boldsymbol{d})}. \tag{19}$$

4. Sample posterior from $P(\boldsymbol{\theta} \mid \boldsymbol{d}_0) \propto \hat{r}(\boldsymbol{d}_0 \mid \boldsymbol{\theta})P(\boldsymbol{\theta})$.

See Hermans et al. (2019); Cranmer et al. (2020); Durkan et al. (2020); Delaunoy et al. (2022); Miller et al. (2022)

Neural Likelihood Estimation (NLE)

- **Robust ratio estimates**: Useful to directly compute likelihood ratios for model testing.
- **Scalability**: Can scale well in high-dimensional parameter spaces, e.g. Truncated Marginal Neural Ratio Estimation (TMNRE).
- **Density-Chasm problem**: joint distribution and marginal distributions may have little overlap in high dimensions.
- **Recent applications**: e.g. CMB (Cole et al., 2022), strong lensing (Anau Montel et al., 2023; Filipp et al., 2024), pulsars (Berteaud et al., 2024), supernova cosmology (Karchev and Trotta, 2024)
- **Implemented in**: `sbi` (Tejero-Cantero et al., 2020), `swyft` (Miller et al., 2022) and `ltu-ili` (Ho et al., 2024).

Neural Posterior Score Estimation (NPSE)

1. Directly estimates the score of the posterior (or likelihood):

$$s(\boldsymbol{\theta} \mid \boldsymbol{d}) = \nabla_{\boldsymbol{\theta}} \ln P(\boldsymbol{\theta} \mid \boldsymbol{d}). \tag{20}$$

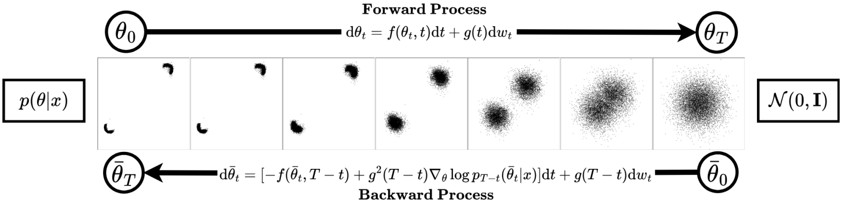
2. "Forward" noising process perturbs samples from the target distribution with noise.
3. A network, $s_w(\boldsymbol{\theta}, \boldsymbol{d}, t)$, is trained to estimate the score of the distributions at different noise levels, t (e.g. with denoising score matching), such that:

$$s_w(\boldsymbol{\theta}, \boldsymbol{d}_0, t) \approx \nabla_{\boldsymbol{\theta}} \ln P_t(\boldsymbol{\theta} \mid \boldsymbol{d}_0). \tag{21}$$

4. Sample posterior by either inverting the network or using Langevin MCMC.

See Hyvärinen and Dayan (2005); Sharrock et al. (2022)

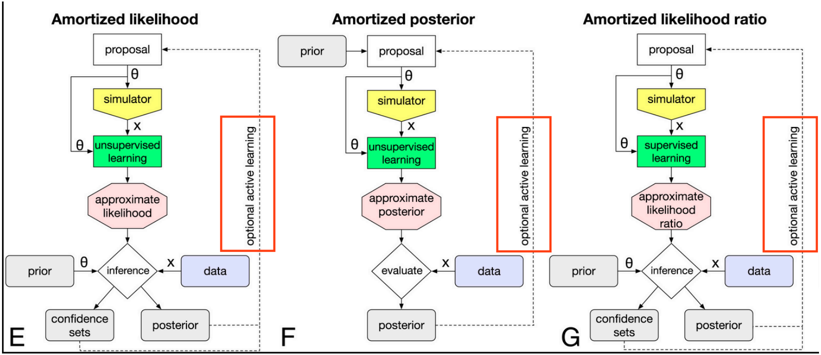
Neural Posterior Score Estimation (NPSE)



Sharrock et al. (2022)

- **Flexible** network architecture, without normalisation requirements (e.g. diffusion models).
- Should scale well to **high-dimensional** data/parameter spaces.
- Naturally incorporates **gradients**.
- Still untested in many contexts.

Sequential SBI and Active Learning



Cranmer et al. (2020)

Sequential SBI and Active Learning

1. Draw initial simulations, $(\boldsymbol{\theta}_i, \boldsymbol{d}_i)$.

2. Train initial density estimate $\hat{P}_{w^{(1)}}(\boldsymbol{\theta} \mid \boldsymbol{x})$ or $\hat{P}_{w^{(1)}}(\boldsymbol{d} \mid \boldsymbol{\theta})$.

3. For subsequent rounds $r > 1$, construct a proposal distribution, $\tilde{P}^{(r)}(\theta)$; e.g., $\tilde{P}^{(r)}(\boldsymbol{\theta}) \propto \hat{P}_{w^{(r-1)}}(\boldsymbol{\theta} \mid \boldsymbol{d}_0)$.

4. Draw new simulations with $\boldsymbol{\theta}_j \sim \tilde{P}^{(r)}(\boldsymbol{\theta})$ and $\boldsymbol{d}_j \sim P(\boldsymbol{d} \mid \boldsymbol{\theta}_j)$.

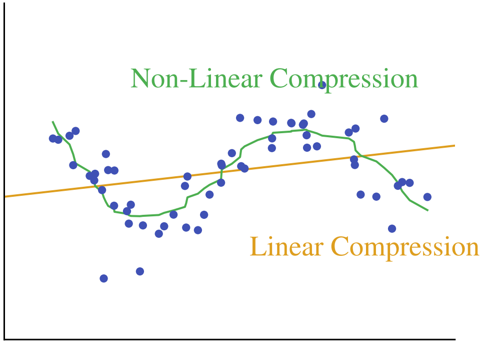
5. Retrain the density estimator using all accumulated simulations.

Data Compression

Data Compression: Linear Compression

- **Principal Component Analysis** (PCA): Projects data onto principal components capturing maximal variance (e.g. Barret and Dupourqué 2024).
- **Score Compression**: Utilizes the score function (gradient of the log-likelihood with respect to parameters, $\nabla_{\boldsymbol{\theta}} \ln P(\boldsymbol{d} \mid \boldsymbol{\theta})$), for compression. Can be lossless in Fisher information (Alsing and Wandelt, 2018).
- **MOPED** and **e-MOPED**: Lossless compression in Fisher information under Gaussian likelihood assumption (Heavens et al., 2000).
- **Canonical Correlation Analysis** (CCA): Finds linear combinations of two sets of model parameters, $\boldsymbol{\theta}$, and data, $\boldsymbol{d}$, that are maximally correlated (e.g. Park et al. 2025).

Data Compression: Linear vs. Non-Linear Compression

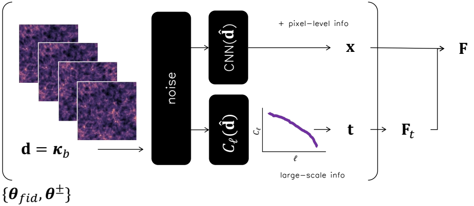


- **Linear methods** may loose non-Gaussian and non-linear information. Sometimes also require an analytic likelihood.

- **Non-linear methods** can address this, but can come at the expense of efficiency and interpretability.

Data Compression: Non-Linear Compression

- **Deep Auto-Enconders**: Unsupervised neural networks consisting of an encoder that maps data to a lower dimensional space.
- **Convolutional Neural Networks** (CNN): Type of encoder that involves filter/kernel optimization (e.g. Lemos et al. 2023b; Jeffrey et al. 2024).
- **Information-Maximising Neural Networks**: Compress with neural networks which find non-linear summary statistics that explicitly maximising the Fisher information (e.g. Charnock et al. 2018; Makinen et al. 2025).



Model Testing & Misspecification with SBI

Bayesian Model Testing

For a given model, $\mathcal{M}$, and observed data, d, the **Bayesian evidence**, $P(d \mid \mathcal{M})$, is defined as :

$$P(d \mid \mathcal{M}) = \int P(d \mid \theta, \mathcal{M})\, P(\theta \mid \mathcal{M})\, d\theta. \tag{22}$$

When comparing two models, $\mathcal{M}_1$ and $\mathcal{M}_2$, one may define the **Bayes factor**:

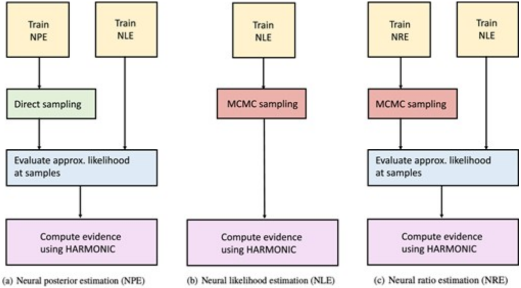
$$B_{12} = \frac{P(d \mid \mathcal{M}_1)}{P(d \mid \mathcal{M}_2)}. \tag{23}$$

$\rightarrow$ Ratio of posterior odds to prior odds of the two models and provides a measure of the evidence in favor of $\mathcal{M}_1$ over $\mathcal{M}_2$.

Naturally incorporates Occam's razor: overly complex models are disfavoured.

Bayesian Model Testing with SBI

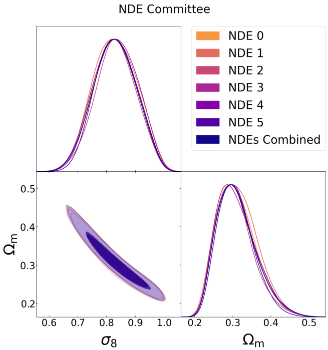
- **NLE-based**: Integrate the learnt likelihood.
- **NLE/NPE or NRE**: If all are available, can compute $\hat{P}(d|\mathcal{M}) \approx \frac{\hat{P}_{w'}(d|\theta_0,\mathcal{M})\,P(\theta_0|\mathcal{M})}{\hat{P}_w(\theta_0|d,\mathcal{M})}$ (Spurio Mancini et al., 2023).
- **floZ**: Training a normalising flow directly on the evidence (Srinivasan et al., 2024).
- **Classifier-based**: Train classifier to distinguish $d \sim \mathcal{M}_1$ from $d \sim \mathcal{M}_2$ (e.g. Jeffrey and Wandelt 2024).



Spurio Mancini et al. (2023)

Model Misspecification in SBI

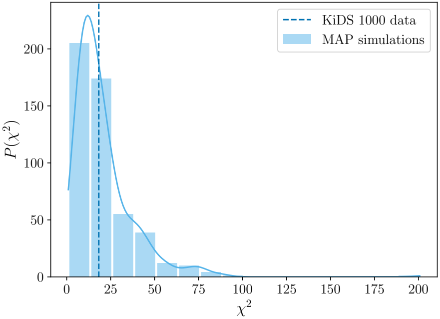
- Model misspecification can cause **biased posterior estimates** despite a self-consistent SBI.
- **Excessive extrapolation:** If the observed data is out-of-distribution (OOD), the trained NNs are no longer describing the correct probability distribution.
- **Mitigation**:
  - GBI framing
  - Error modelling within NDE (Huang et al., 2023; Kelly et al., 2025)
  - Ensembling



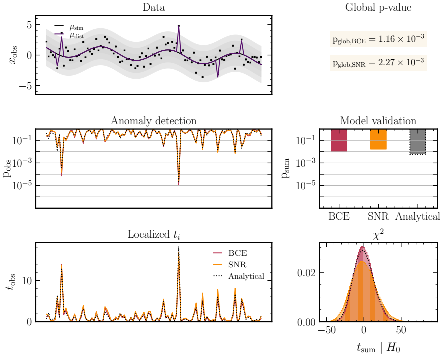
von Wietersheim-Kramsta et al. (2025)

Diagnosing & Testing SBI

Testing for OOD: Goodness-of-Fit Tests

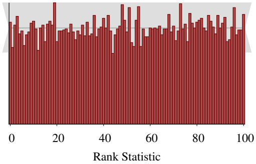


GoF measure under Gaussian likelihood assumption (von Wietersheim-Kramsta et al., 2025)

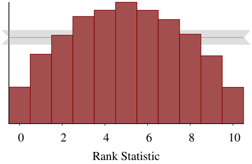


Posterior predictive tests (Anau Montel et al., 2025)

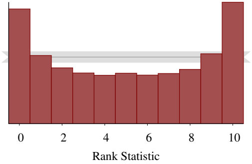
Consistency between SBI & Simulator: Simulation-Based Calibration



Accurate posterior



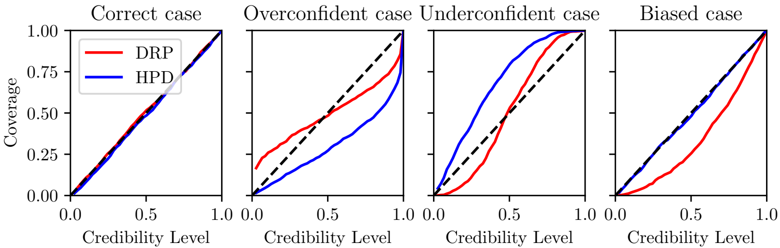
Learnt posterior too wide



Learnt posterior to narrow

Simulation-Based Calibration (Talts et al., 2018)

Consistency between SBI & Simulator: Coverage/TARP



Coverage tests/TARP (Deistler et al., 2022; Lemos et al., 2023a)

Conclusion & Outlooks

Outlooks

- SBI is a powerful method allowing for increasingly complex modelling.
- Some **caveats** $\rightarrow$ Extensive testing and validation required.
- Still limited by computational resources, particularly, with higher dimensional data.

- **Future avenues**:
  - Neural Posterior Score Estimation (NPSE)
  - Transfer Learning to combine constraints
  - Training with multifidelity simulators (Krouglova et al., 2025)
- **Exenstive applications**: large datasets from surveys, simulations and high-resolution imaging.

Questions?

References

Alsing, J., Charnock, T., Feeney, S., and Wandelt, B. (2019). Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458.

Alsing, J. and Wandelt, B. (2018). Generalized massive optimal data compression. , 476(1):L60–L64.

Anau Montel, N., Alvey, J., and Weniger, C. (2025). Tests for model misspecification in simulation-based inference: From local distortions to global model checks. , 111(8):083013.

Anau Montel, N., Coogan, A., Correa, C., Karchev, K., and Weniger, C. (2023). Estimating the warm dark matter mass from strong lensing images with truncated marginal neural ratio estimation. , 518(2):2746–2760.

Barret, D. and Dupourqué, S. (2024). Simulation-based inference with neural posterior estimation applied to X-ray spectral fitting. Demonstration of working principles down to the Poisson regime. , 686:A133.

Baxter, E. J., Christy, J. G., and Kumar, J. (2022). Approximate Bayesian Computation applied to the Diffuse Gamma-Ray Sky. , 516(2):2326–2336.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berteaud, J., Eckner, C., Calore, F., Clavel, M., and Haggard, D. (2024). Simulation-based Inference of Radio Millisecond Pulsars in Globular Clusters. , 974(1):144.

Bissiri, P. G., Holmes, C., and Walker, S. (2013). A General Framework for Updating Belief Distributions. *arXiv e-prints*, page arXiv:1306.6430.

Boelts, J., Lueckmann, J.-M., Gao, R., and Macke, J. H. (2022). Flexible and efficient simulation-based inference for models of decision-making. *eLife*, 11:e77220.

Cameron, E. and Pettitt, A. N. (2012). Approximate Bayesian Computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift. , 425(1):44–65.

Charnock, T., Lavaux, G., and Wandelt, B. D. (2018). Automatic physical inference with information maximizing neural networks. , 97(8):083004.

Cole, A., Miller, B. K., Witte, S. J., Cai, M. X., Grootes, M. W., Nattino, F., and Weniger, C. (2022). Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation. , 2022(9):004.

Cranmer, K., Brehmer, J., and Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062.

Deistler, M., Goncalves, P. J., and Macke, J. H. (2022). Truncated proposals for scalable and hassle-free simulation-based inference. *arXiv e-prints*, page arXiv:2210.04815.

Delaunoy, A., Hermans, J., Rozet, F., Wehenkel, A., and Louppe, G. (2022). Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation. *arXiv e-prints*, page arXiv:2208.13624.

Durkan, C., Murray, I., and Papamakarios, G. (2020). On Contrastive Learning for Likelihood-free Inference. *arXiv e-prints*, page arXiv:2002.03712.

Erickson, S., Wagner-Carena, S., Marshall, P., Millon, M., Birrer, S., Roodman, A., Schmidt, T., Treu, T., Schuldt, S., Shajib, A., Venkatraman, P., and The LSST Dark Energy Science Collaboration (2024). Lens Modeling of STRIDES Strongly Lensed Quasars using Neural Posterior Estimation. *arXiv e-prints*, page arXiv:2410.10123.

Filipp, A., Hezaveh, Y., and Perreault-Levasseur, L. (2024). Robustness of Neural Ratio and Posterior Estimators to Distributional Shifts for Population-Level Dark Matter Analysis in Strong Gravitational Lensing. *arXiv e-prints*, page arXiv:2411.05905.

Glöckler, M., Deistler, M., and Macke, J. H. (2022). Variational methods for simulation-based inference. *arXiv e-prints*, page arXiv:2203.04176.

Greenberg, D. S., Nonnenmacher, M., and Macke, J. H. (2019). Automatic Posterior Transformation for Likelihood-Free Inference. *arXiv e-prints*, page arXiv:1905.07488.

He, Q., Robertson, A., Nightingale, J., Cole, S., Frenk, C. S., Massey, R., Amvrosiadis, A., Li, R., Cao, X., and Etherington, A. (2022). A forward-modelling method to infer the dark matter particle mass from strong gravitational lenses. , 511(2):3046–3062.

Heavens, A. F., Jimenez, R., and Lahav, O. (2000). Massive lossless data compression and multiple parameter estimation from galaxy spectra. , 317(4):965–972.

Hermans, J., Begy, V., and Louppe, G. (2019). Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, page arXiv:1903.04057.

Ho, M., Bartlett, D. J., Chartier, N., Cuesta-Lazaro, C., Ding, S., Lapel, A., Lemos, P., Lovell, C. C., Makinen, T. L., Modi, C., Pandya, V., Pandey, S., Perez, L. A., Wandelt, B., and Bryan, G. L. (2024). LtU-ILI: An All-in-One Framework for Implicit Inference in Astrophysics and Cosmology. *The Open Journal of Astrophysics*, 7:54.

Huang, D., Bharti, A., Souza, A., Acerbi, L., and Kaski, S. (2023). Learning Robust Statistics for Simulation-based Inference under Model Misspecification. *arXiv e-prints*, page arXiv:2305.15871.

Hyvärinen, A. and Dayan, P. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4).

Jeffrey, N. and Wandelt, B. D. (2024). Evidence Networks: simple losses for fast, amortized, neural Bayesian model comparison. *Machine Learning: Science and Technology*, 5(1):015008.

Jeffrey, N., Whiteway, L., Gatti, M., Williamson, J., Alsing, J., Porredon, A., Prat, J., Doux, C., Jain, B., Chang, C., Cheng, T. Y., Kacprzak, T., Lemos, P., Alarcon, A., Amon, A., Bechtol, K., Becker, M. R., Bernstein, G. M., Campos, A., Rosell, A. C., Chen, R., Choi, A., DeRose, J., Drlica-Wagner, A., Eckert, K., Everett, S., Ferté, A., Gruen, D., Gruendl, R. A., Herner, K., Jarvis, M., McCullough, J., Myles, J., Navarro-Alsina, A., Pandey, S., Raveri, M., Rollins, R. P., Rykoff, E. S., Sánchez, C., Secco, L. F., Sevilla-Noarbe, I., Sheldon, E., Shin, T., Troxel, M. A., Tutusaus, I., Varga, T. N., Yanny, B., Yin, B., Zuntz, J., Aguena, M., Allam, S. S., Alves, O., Bacon, D., Bocquet, S., Brooks, D., da Costa, L. N., Davis, T. M., De Vicente, J., Desai, S., Diehl, H. T., Ferrero, I., Frieman, J., García-Bellido, J., Gaztanaga, E., Giannini, G., Gutierrez, G., Hinton,

S. R., Hollowood, D. L., Honscheid, K., Huterer, D., James, D. J., Lahav, O., Lee, S., Marshall, J. L., Mena-Fernández, J., Miquel, R., Pieres, A., Malagón, A. A. P., Roodman, A., Sako, M., Sanchez, E., Sanchez Cid, D., Smith, M., Suchyta, E., Swanson, M. E. C., Tarle, G., Tucker, D. L., Weaverdyck, N., Weller, J., Wiseman, P., and Yamamoto, M. (2024). Dark Energy Survey Year 3 results: likelihood-free, simulation-based wCDM inference with neural compression of weak-lensing map statistics. .

Karchev, K. and Trotta, R. (2024). STAR NRE: Solving supernova selection effects with set-based truncated auto-regressive neural ratio estimation. *arXiv e-prints*, page arXiv:2409.03837.

Kelly, R. P., Warne, D. J., Frazier, D. T., Nott, D. J., Gutmann, M. U., and Drovandi, C. (2025). Simulation-based Bayesian inference under model misspecification. *arXiv e-prints*, page arXiv:2503.12315.

Krouglova, A. N., Johnson, H. R., Confavreux, B., Deistler, M., and Gonçalves, P. J. (2025). Multifidelity Simulation-based Inference for Computationally Expensive Simulators. *arXiv e-prints*, page arXiv:2502.08416.

Lemos, P., Coogan, A., Hezaveh, Y., and Perreault-Levasseur, L. (2023a). Sampling-Based Accuracy Testing of Posterior Estimators for General Inference. *40th International Conference on Machine Learning*, 202:19256–19273.

Lemos, P., Parker, L. H., Hahn, C., Ho, S., Eickenberg, M., Hou, J., Massara, E., Modi, C., Moradinezhad Dizgah, A., Régaldo-Saint Blancard, B., and Spergel, D. (2023b). SimBIG: Field-level Simulation-based Inference of Large-scale Structure. In *Machine Learning for Astrophysics*, page 18.

Leyde, K., Green, S. R., Toubiana, A., and Gair, J. (2024). Gravitational wave populations and cosmology with neural posterior estimation. , 109(6):064056.

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., and Macke, J. H. (2017). Flexible statistical inference for mechanistic models of neural dynamics. *arXiv e-prints*, page arXiv:1711.01861.

Makinen, L. T., Heavens, A., Porqueres, N., Charnock, T., Lapel, A., and Wandelt, B. D. (2025). Hybrid summary statistics: neural weak lensing inference beyond the power spectrum. , 2025(1):095.

Miller, B. K., Cole, A., Weniger, C., Nattino, F., Ku, O., and Grootes, M. W. (2022). swyft: Truncated marginal neural ratio estimation in python. *Journal of Open Source Software*, 7(75):4205.

Miller, B. K., Weniger, C., and Forré, P. (2022). Contrastive Neural Ratio Estimation for Simulation-based Inference. *arXiv e-prints*, page arXiv:2210.06170.

Papamakarios, G. and Murray, I. (2016). Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, page arXiv:1605.06376.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64.

Papamakarios, G., Pavlakou, T., and Murray, I. (2017). Masked autoregressive flow for density estimation. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Park, M., Gatti, M., and Jain, B. (2025). Dimensionality reduction techniques for statistical inference in cosmology. , 111(6):063523.

Rozet, F., Delaunoy, A., and Miller, B. (2021). Lampe: Likelihood-free amortized posterior estimation. *Statistical Software*.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, pages 1151–1172.

Saoulis, A. A., Piras, D., Spurio Mancini, A., Joachimi, B., and Ferreira, A. M. G. (2025). Full-waveform earthquake source inversion using simulation-based inference. *Geophysical Journal International*, 241(3):1740–1761.

Sharrock, L., Simons, J., Liu, S., and Beaumont, M. (2022). Sequential Neural Score Estimation: Likelihood-Free Inference with Conditional Score Based Diffusion Models. *arXiv e-prints*, page arXiv:2210.04872.

Spurio Mancini, A., Docherty, M. M., Price, M. A., and McEwen, J. D. (2023). Bayesian model comparison for simulation-based inference. *RAS Techniques and Instruments*, 2(1):710–722.

Srinivasan, R., Crisostomi, M., Trotta, R., Barausse, E., and Breschi, M. (2024). Bayesian evidence estimation from posterior samples with normalizing flows. , 110(12):123007.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv e-prints*, page arXiv:1804.06788.

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., and Macke, J. H. (2020). sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505.

Tortorelli, L., Siudek, M., Moser, B., Kacprzak, T., Berner, P., Refregier, A., Amara, A., García-Bellido, J., Cabayol, L., Carretero, J., Castander, F. J., De Vicente, J., Eriksen, M., Fernandez, E., Gaztanaga, E., Hildebrandt, H., Joachimi, B., Miquel, R., Sevilla-Noarbe, I., Padilla, C., Renard, P., Sanchez, E., Serrano, S., Tallada-Crespí, P., and Wright, A. H. (2021). The PAU survey: measurement of narrow-band galaxy properties with approximate bayesian computation. , 2021(12):013.

Vasist, M., Rozet, F., Absil, O., Mollière, P., Nasedkin, E., and Louppe, G. (2023). Neural posterior estimation for exoplanetary atmospheric retrieval. , 672:A147.

von Wietersheim-Kramsta, M., Lin, K., Tessore, N., Joachimi, B., Loureiro, A., Reischke, R., and Wright, A. H. (2025). KiDS-SBI: Simulation-based inference analysis of KiDS-1000 cosmic shear. , 694:A223.