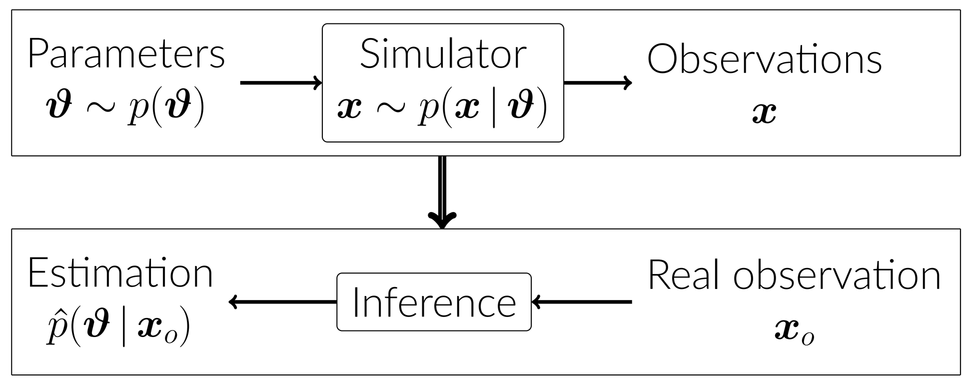
Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation

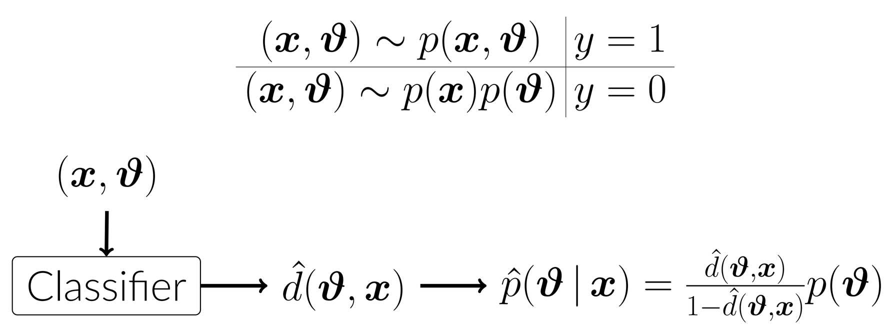
Arnaud Delaunoy* Joeri Hermans* François Rozet Antoine Wehenkel Gilles Louppe

Simulation-based inference

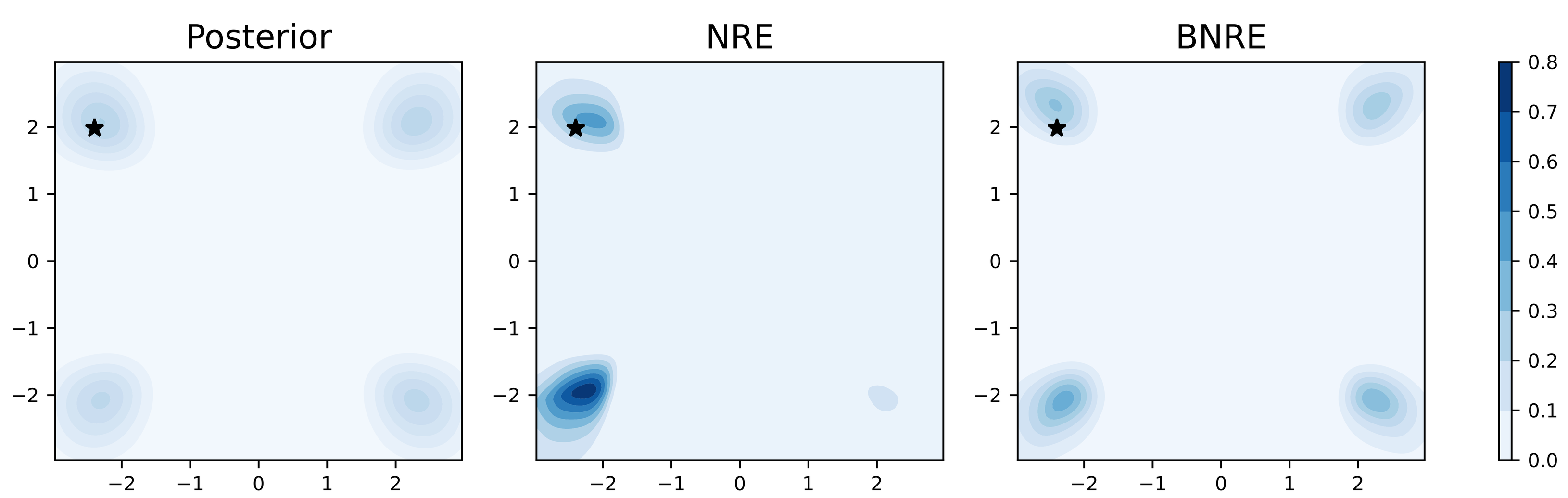


Neural Ratio Estimation

Neural ratio estimation consists in training a classifier to discriminate between samples from the joint density, $p(\boldsymbol{x}, \boldsymbol{\vartheta})$, and the marginal densities, $p(\boldsymbol{x})p(\boldsymbol{\vartheta})$.



Motivation



We observe that the posterior density obtained with NRE is sharper than the true density. Consequently **NRE may exclude parameter values that are actually plausible**. BNRE aims to mitigate this issue by producing more conservative posterior approximations.

Balanced Neural Ratio Estimation

Idea: restrict the hypothesis space to balanced classifiers.

Definition: A classifier $\hat{d}$ is balanced if $\mathbb{E}_{p(\boldsymbol{\vartheta}, \boldsymbol{x})}\left[\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})\right] + \mathbb{E}_{p(\boldsymbol{\vartheta})p(\boldsymbol{x})}\left[\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})\right] = 1.$

Algorithm 1 Balanced Neural Ratio Estimation (BNRE)

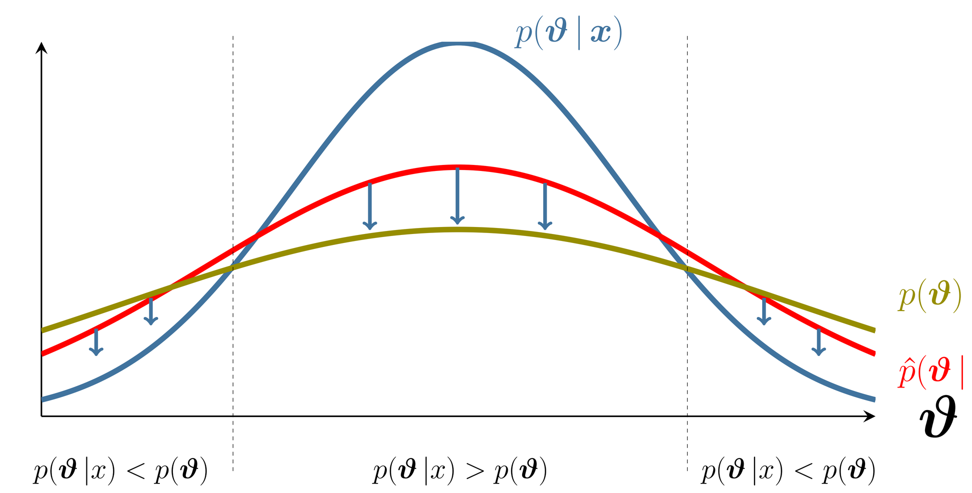
repeat
- Sample from the joint $\{\boldsymbol{\vartheta}_i,\ \boldsymbol{x}_i \sim p(\boldsymbol{\vartheta}, \boldsymbol{x}),\ y_i = 1\}_{i=1}^{n/2}$
- Sample from the marginals $\{\boldsymbol{\vartheta}_i,\ \boldsymbol{x}_i \sim p(\boldsymbol{\vartheta})p(\boldsymbol{x}),\ y_i = 0\}_{i=n/2+1}^{n}$
- $\mathcal{L}[\hat{d}_\psi] = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{d}_\psi(\boldsymbol{\vartheta}_i, \boldsymbol{x}_i) + (1-y_i)\log(1-\hat{d}_\psi(\boldsymbol{\vartheta}_i, \boldsymbol{x}_i))$
- $\mathcal{B}[\hat{d}_\psi] = \frac{2}{n}\sum_{i=1}^{n/2} \hat{d}_\psi(\boldsymbol{\vartheta}_i, \boldsymbol{x}_i) + \frac{2}{n}\sum_{i=n/2+1}^{n} \hat{d}_\psi(\boldsymbol{\vartheta}_i, \boldsymbol{x}_i)$
- $\psi =$ `minimizer_step(params=`ψ`, loss=`$\mathcal{L}[\hat{d}_\psi] + \lambda(\mathcal{B}[\hat{d}_\psi] - 1)^2$`)`

until convergence
return $\hat{d}_\psi(\boldsymbol{\vartheta}, \boldsymbol{x})$.

Theorems

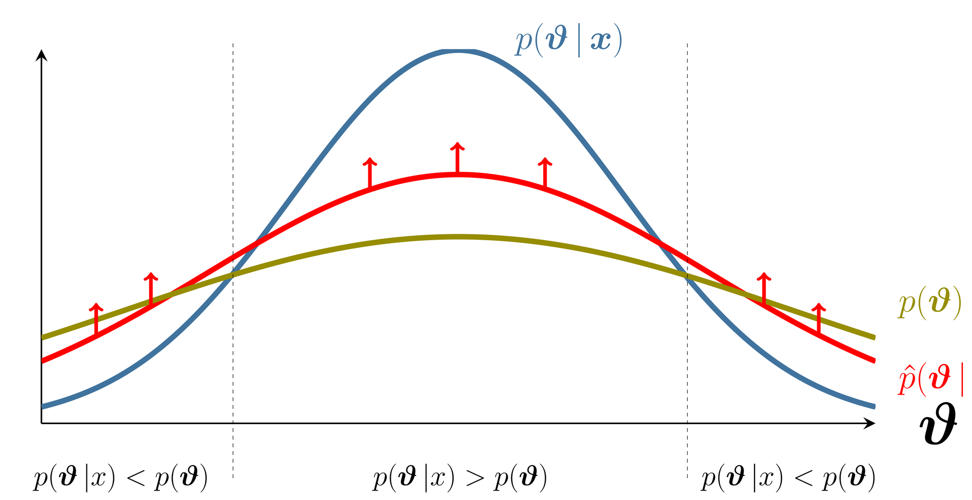
Theorem 1: Any balanced classifier $\hat{d}$ satisfies $\mathbb{E}_{p(\boldsymbol{\vartheta}, \boldsymbol{x})}\left[\frac{d^*(\boldsymbol{\vartheta}, \boldsymbol{x})}{\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})}\right] \geq 1.$

$$\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x}) \leq d^*(\boldsymbol{\vartheta}, \boldsymbol{x}) \Leftrightarrow \frac{\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})}{1-\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})} \leq \frac{d^*(\boldsymbol{\vartheta}, \boldsymbol{x})}{1-d^*(\boldsymbol{\vartheta}, \boldsymbol{x})} \Leftrightarrow \hat{p}(\boldsymbol{\vartheta} \mid \boldsymbol{x}) \leq p(\boldsymbol{\vartheta} \mid \boldsymbol{x})$$

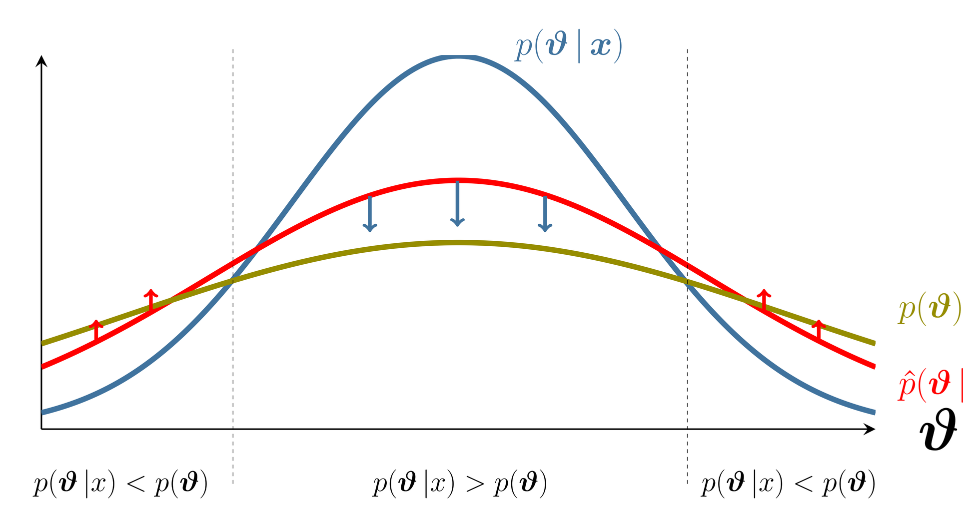


Theorem 2: Any balanced classifier $\hat{d}$ satisfies $\mathbb{E}_{p(\boldsymbol{\vartheta})p(\boldsymbol{x})}\left[\frac{1-d^*(\boldsymbol{\vartheta}, \boldsymbol{x})}{1-\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})}\right] \geq 1.$

$$1-\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x}) \leq 1-d^*(\boldsymbol{\vartheta}, \boldsymbol{x}) \Leftrightarrow \frac{\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})}{1-\hat{d}(\boldsymbol{\vartheta}, \boldsymbol{x})} \geq \frac{d^*(\boldsymbol{\vartheta}, \boldsymbol{x})}{1-d^*(\boldsymbol{\vartheta}, \boldsymbol{x})} \Leftrightarrow \hat{p}(\boldsymbol{\vartheta} \mid \boldsymbol{x}) \geq p(\boldsymbol{\vartheta} \mid \boldsymbol{x})$$



Theorems 1 + 2



Theorem 3: The Bayes optimal classifier $d^*(\boldsymbol{\vartheta}, \boldsymbol{x})$ is balanced.

Theorems 1 and 2 state that balanced classifiers tend to be more conservative than non-balanced ones. Theorem 3 states that restricting the classifier hypothesis space to balanced classifiers does not modify the optimum and hence that BNRE keeps the asymptotic properties of NRE. In conclusion, **BNRE provides more conservative inferences in low-budget regimes while still asymptotically converging to the Bayes optimal classifier.**

Results

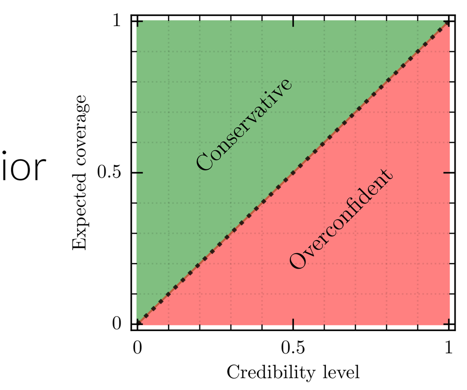
The expected coverage is expressed as

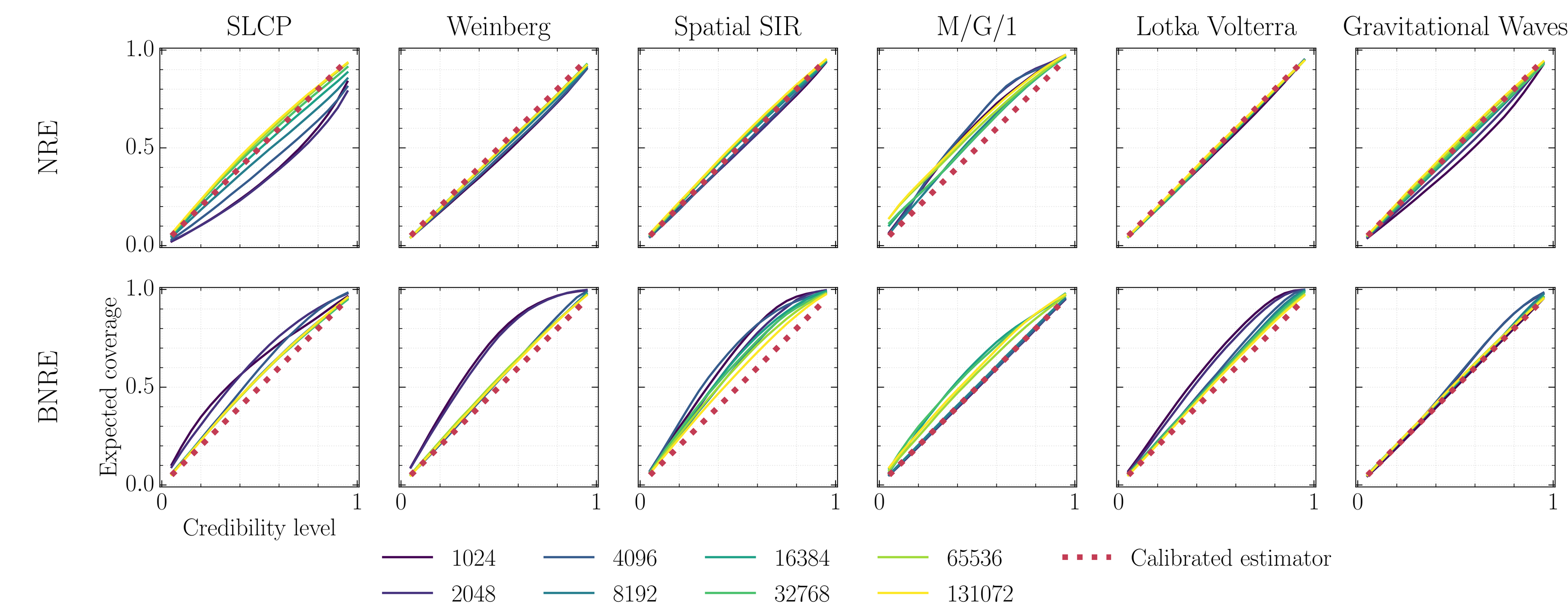
$$\text{expected coverage}(\hat{p}, \alpha) = \mathbb{E}_{p(\boldsymbol{\vartheta}, \boldsymbol{x})}\left[\mathbb{1}\left[\boldsymbol{\vartheta} \in \Theta_{\hat{p}(\boldsymbol{\vartheta} \mid \boldsymbol{x})}(1-\alpha)\right]\right],$$

where the function $\Theta_{\hat{p}(\boldsymbol{\vartheta} \mid \boldsymbol{x})}(1-\alpha)$ yields the $1-\alpha$ highest posterior density region of $\hat{p}(\boldsymbol{\vartheta} \mid \boldsymbol{x})$.

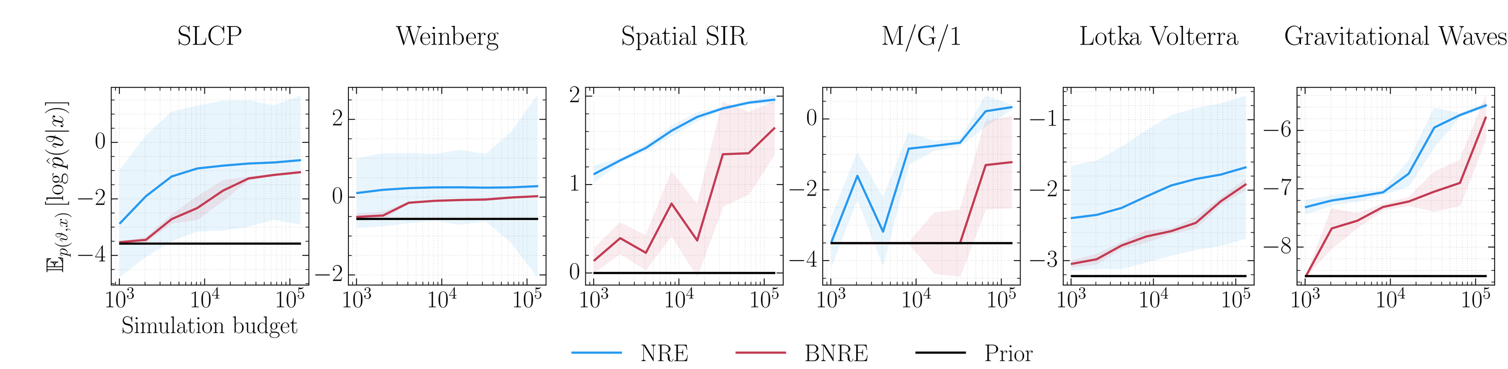
A **conservative model** is a model such that

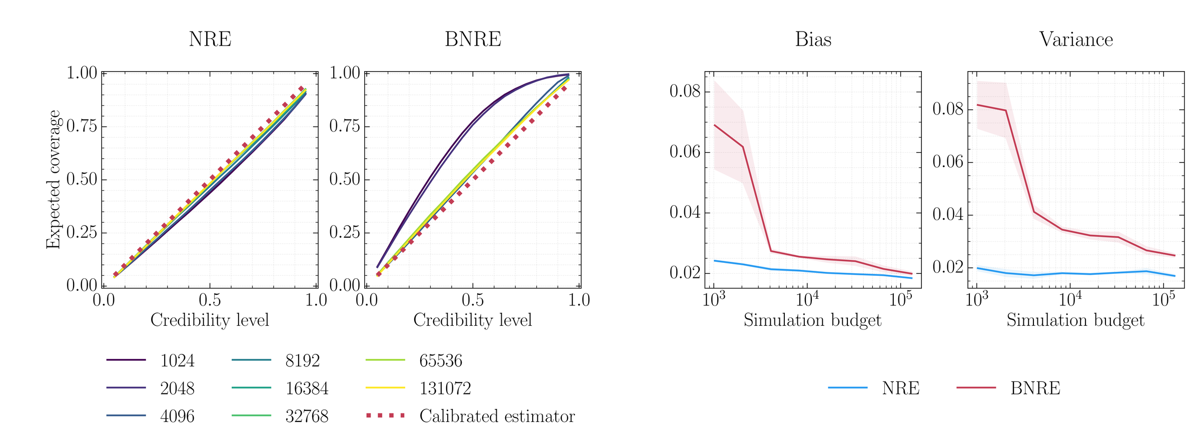
$$\text{expected coverage}(\hat{p}, \alpha) \geq 1-\alpha, \quad \forall \alpha$$





BNRE produces more conservative posterior approximations than NRE.





Enforcing the balancing condition comes at the price of a small loss in information gain. However, BNRE eventually converges towards NRE as the simulation budget increases.

Take-home messages

- BNRE aims to produce more conservative posterior approximations than NRE by enforcing the classifier to be balanced.
- Empirically, BNRE produces conservative posterior approximations on all the benchmarks. However, theoretical guarantees hold only in expectation and are on the discriminator. Approximate posterior conservativeness is hence not guaranteed.
- BNRE should not be viewed as a way to obtain conservative posterior estimators with 100% reliability, but rather as a way to increase the reliability of the posterior estimators with minimal effort and no computational overhead.