PNAS
www.pnas.org

Supplementary Information for

Learning complex models with invertible neural networks: a likelihood-free Bayesian approach

Stefan T. Radev, Ulf K. Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe

Stefan Radev.
E-mail: stefan.radev@psychologie.uni-heidelberg.de

This PDF file includes:

Supplementary text
Fig. S1
Captions for Movies S1 to S3
Captions for Databases S1 to S2
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Movies S1 to S3
Databases S1 to S2

Supporting Information Text

Results

Performance metrics. In the following, the computation of the performance metrics used throughout the main text is detailed.

Normalized Root Mean Squared Error. The normalized root mean squared error (NRMSE) between a sample of true parameters $\{\theta^{(i)}\}_{i=1}^n$ and a sample of estimated parameters $\{\hat{\theta}^{(i)}\}_{i=1}^n$ is given by:

$$NRMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\theta^{(i)} - \hat{\theta}^{(i)}\right)^2}{\theta_{max} - \theta_{min}}} \quad [1]$$

Due to the normalization factor $\theta_{max} - \theta_{min}$, the NRMSE is scale-independent, and thus suitable for comparing the recovery across parameters having different numerical ranges. The NRMSE is zero when the estimates are exactly equal to the true values.

Coefficient of Determination . The coefficient of determination R^2 gives the proportion of variance in a sample of true parameters $\{\theta^{(i)}\}_{i=1}^n$ that is "explained" by a sample of estimated parameters $\{\hat{\theta}^{(i)}\}_{i=1}^n$. It is computed as:

$$R^2 = 1 - \sum_{i=1}^{n} \frac{\left(\theta^{(i)} - \hat{\theta}^{(i)}\right)^2}{\left(\theta^{(i)} - \bar{\theta}^{(i)}\right)^2} \quad [2]$$

where $\bar{\theta}$ denotes the mean of the true parameter samples. When R^2 equals 1, it means that the estimates are perfect reconstructions of the true parameters.

Kullback-Leibler Divergence. The Kullback-Leibler divergence (D_{KL}) quantifies the increase in entropy incurred by approximating a target probability distribution P with a distribution Q. Its general form for absolutely continuous distributions is given by

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad [3]$$

where p and q denote the pdfs of P and Q. In the case where P and Q are both multivariate Gaussian distributions, the KL divergence can be computed in closed form (1):

$$D_{KL}(P||Q) = \frac{1}{2}\left[\log \frac{\det \mathbf{\Sigma}_q}{\det \mathbf{\Sigma}_p} + \mathrm{Tr}(\mathbf{\Sigma}_q^{-1}\mathbf{\Sigma}_p) - d + (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbf{\Sigma}_q^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)\right] \quad [4]$$

where $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_q$ denote the covariance matrices of p and q, $\boldsymbol{\mu}_p$ and $\boldsymbol{\mu}_q$ the respective mean vectors, and d the number of dimensions of the Gaussian. In the case of diagonal Gaussian distributions, Eq.4 reduces to:

$$D_{KL}(P||Q) = \sum_{i=1}^{d}\left(\log \frac{\sigma_{q,i}}{\sigma_{p,i}} + \frac{\sigma_{p,i}^2 + (\mu_{q,i} - \mu_{p,i})^2}{2\sigma_{q,i}^2} - \frac{1}{2}\right) \quad [5]$$

Even though the KL divergence is not a proper distance metric, as it is not symmetric in its arguments, it can be used to quantify the error of approximation and serve as a metric for comparing different methods.

Simulation-Based Calibration. Simulation-based calibration is a recently proposed method for validating the accuracy of posterior samples generated by a Bayesian sampling method (2). It is based on the so called *self-consistency* of the Bayesian joint distribution. Given a sample from the prior distribution $\tilde{\theta} \sim p(\theta)$ and a sample from the data-generating process $\tilde{x} \sim p(x|\tilde{\theta})$, one can integrate $\tilde{\theta}$ and $\tilde{x}$ out of the Bayesian joint distribution to recover back the prior of θ:

$$p(\theta) = \int p(\theta, \tilde{\theta}, \tilde{x}) d\tilde{x} d\tilde{\theta} \quad [6]$$

$$= \int p(\theta, \tilde{x}|\tilde{\theta}) p(\tilde{\theta}) d\tilde{x} d\tilde{\theta} \quad [7]$$

$$= \int p(\theta|\tilde{x}) p(\tilde{x}|\tilde{\theta}) p(\tilde{\theta}) d\tilde{x} d\tilde{\theta} \quad [8]$$

If the Bayesian sampling method produces samples from the exact posterior, the equality implied by Eq.8 should hold regardless of the particular form of the posterior. Thus, any violation of this equality indicates some error incurred by the sampling method.

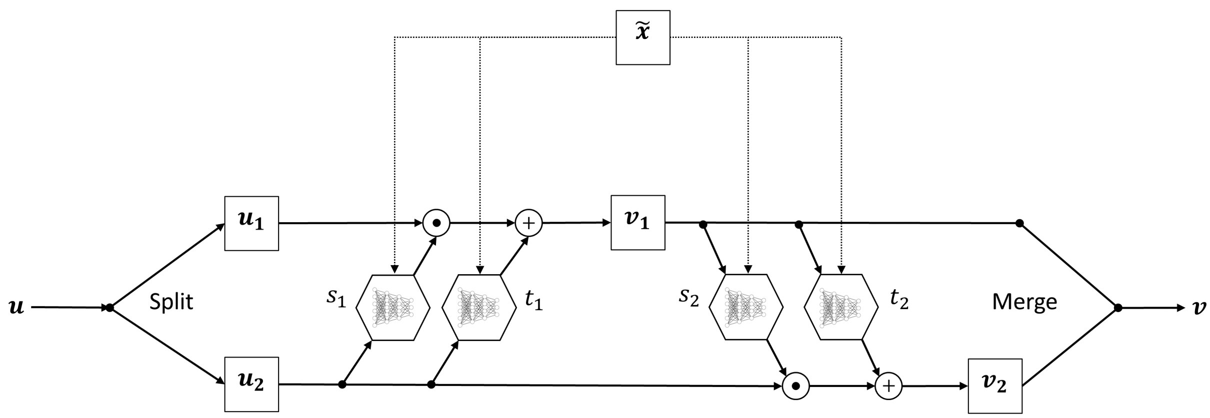

Fig. S1. Second figure

Movie S1. Type caption for the movie here.

Movie S2. Type caption for the other movie here. Adding longer text to show what happens, to decide on alignment and/or indentations.

Movie S3. A third movie, just for kicks.

Additional data table S1 (dataset_one.txt)
Type or paste caption here.

Additional data table S2 (dataset_two.txt)
Type or paste caption here. Adding longer text to show what happens, to decide on alignment and/or indentations for multi-line or paragraph captions.

References

1. Hershey JR, Olsen PA (2007) Approximating the kullback leibler divergence between gaussian mixture models in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. (IEEE), Vol. 4, pp. IV–317.
2. Talts S, Betancourt M, Simpson D, Vehtari A, Gelman A (2018) Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*.