

Comparing neural simulations by neural density estimation

Jan Boelts

Computational Neuroengineering,
Technical University of Munich

Jan-Matthis Lueckmann

Computational Neuroengineering,
Technical University of Munich

Pedro J. Goncalves

Research Center caesar,
an associate of the Max Planck Society, Bonn

Henning Sprekeler

Modelling of Cognitive Processes,
Technical University of Berlin

Jakob H. Macke

Computational Neuroengineering,
Technical University of Munich

Abstract

A common problem in computational neuroscience is the comparison of competing models in the light of observed data. Bayesian model comparison provides a statistically sound framework for this comparison based on the evidence each model provides for the data. In practice, however, models are often defined through complex simulators so that methods relying on likelihood functions are not applicable. Previous approaches in the field of Approximate Bayesian Computation (ABC) rely on rejection sampling to circumvent the likelihood, but are typically computationally inefficient. We propose an efficient method to perform Bayesian model comparison for simulation-based models. Using recent advances in posterior density estimation, we train a mixture-density network to map features of the observed data to the parameters of the posterior over models. We show that the method performs accurately on two tractable example problems, and present an application to a use case scenario from computational neuroscience – the comparison of ion channel models.

Keywords: model comparison; Approximate Bayesian Computation; density estimation

Introduction

Complex simulations play an important role in cognitive science, computational neuroscience, and other fields, and are used to develop and evaluate hypotheses for mechanisms underlying the processes under investigation (Gerstner, Sprekeler, & Deco, 2012). For hypothesis-building, we often want to decide which of several candidate models provides the best explanation of empirical data, i.e. we want to perform model comparison. However, in practice, complex mechanistic models are often defined implicitly through stochastic simulators, e.g., a set of differential equations to be integrated.

Bayesian model comparison offers a statistically sound framework for comparing competing hypotheses (Dienes, 2011; Bishop, 2006). Given candidate models m_i with parameters θ and observed data x_o the posterior probability of a model is

$$p(m_i|x_o) \propto p(m_i) p(x_o|m_i) = p(m_i) \int p(x_o|\theta, m_i) p(\theta|m_i) d\theta.$$

where $p(x_o|\theta, m_i)$ denotes the likelihood of the data given the model m_i and parameters θ , and the integral is called marginal

likelihood of the model (model evidence). In the context of complex simulator models, however, it is often impossible to compute the likelihood function $p(x_o|\theta, m)$, implying that conventional approaches to Bayesian model comparison cannot be applied.

Classical approaches to solving this problem originate in the field of Approximate Bayesian Computation (ABC, Sunnåker et al., 2013) and rely on rejection sampling (Rubin et al., 1984). Basic rejection sampling was extended to a sequential Monte Carlo approach (ABC SMC) by Toni, Welch, Strelkova, Ipsen, and Stumpf (2009), improving the sampling efficiency using importance weights. Overall, however, these approaches do not scale to high-dimensional applications and often rely on the choice of summary statistics, distance functions and acceptance thresholds.

Recently, several studies used conditional density estimation to infer the parameters of a simulation-based model (Papamakarios & Murray, 2016; Lueckmann et al., 2017; Greenberg, Nonnenmacher, & Macke, 2019). In contrast to rejection sampling approaches, these methods approximate the posterior over the parameters of a model in parametrized form, avoiding the choice of distance functions and providing improved simulation efficiency. Furthermore, these improve sampling efficiency by drawing new samples from adaptively chosen proposal distributions, giving rise to the name of sequential neural posterior estimation (SNPE).

However, previous studies concentrated on neural density estimation for posterior inference, and did not provide methodology for *model comparison*. Here, we investigate how neural density estimation can be used for *model comparison*, in order to provide methodology for comparing complex simulation-based models in neuroscience and beyond.

Model comparison via neural density estimation

The goal of neural posterior estimation is to learn a parametric approximation to the exact posterior, using simulations from the model. SNPE approaches originate in regression correction approaches to ABC (Beaumont, Zhang, & Balding, 2002), which proposed to learn properties of the posterior, such as its mean and variance, directly using a regression from simulated parameters to generated data. More recently, this idea was developed further using mixture-density networks (MDNs) to estimate the entire conditional density of the parameters given the observed data $p(\theta|x_o)$ (Papamakarios & Murray, 2016;



Lueckmann et al., 2017). Greenberg et al. (2019) use SNPE with flow-based conditional density estimators, and their methods allows for full flexibility in adaptively sampling simulations.

Here, we extend the idea of learning posterior over model parameters of a single model, to learning the posterior over multiple competing models to perform model comparison: Our goal is to approximate the posterior probability of each model i , given the observed data, i.e. $p(m_i|x_o)$. Instead of doing rejection sampling, approximating the posterior probability (or the evidence) with a selection of sampled parameters, our approach follows the rationale of posterior density estimation as outlined in Figure 1:

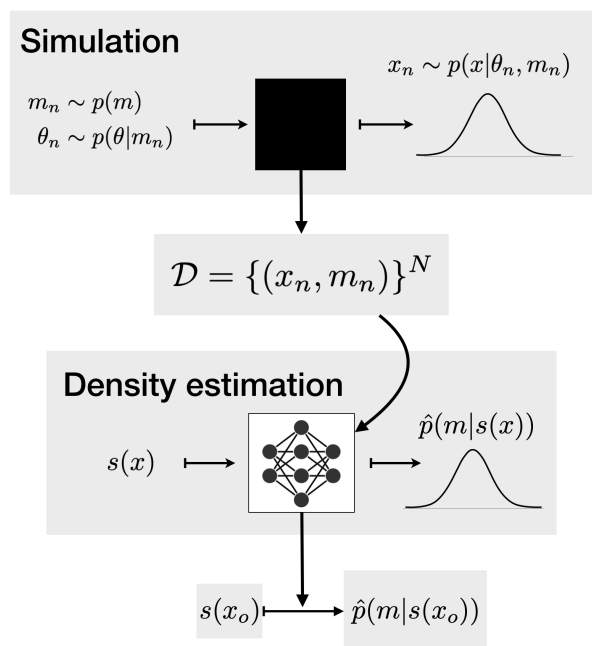


Figure 1: **Schematic overview of the approach:** In the simulation phase, training data are generated from the simulator model. The resulting training dataset \mathcal{D} is used in the density estimation phase to train a neural network which approximates the posterior over models. In practice, one often reduces the data x to a set of summary statistics $s(x)$.

First, we choose priors over models and parameters. Then we repeatedly sample from the priors to generate a large amount of simulated data $\mathcal{D} = \{(x_n, m_n)\}_{n=1}^N$ (Figure 1, Simulation). The simulated data are used to train a density estimation neural network $q_\Psi(m|x)$ with parameters Ψ to map from data to the model index, (Figure 1, Density estimation). Finally, the density estimator is used to predict the posterior over models given the empirically observed data, $\hat{p}(m_i|x_o)$.

Starting from the joint data generating distribution

$$p(m, \theta, x) = p(\theta|m, x)p(m|x)p(x) \quad (1)$$

one can derive a loss function for neural network training that minimizes the Kullback-Leibler divergence between the true

posterior $p(m|x)$ and the approximating parametrized posterior $q(m|x)$. This results in the objective of maximizing the expected log-probability of the model.

The derivation involves both the posterior approximation for the models and for the parameters, thus, in principle the network can be trained to predict both the parameters θ_n and the generating model m_n , effectively doing model comparison and inference simultaneously. Here, we are interested in the discrete posterior over models only, so that the objective corresponds to the cross-entropy loss

$$L(\Psi) = \left\langle \log q_\Psi(m|x) \right\rangle_{p(x, m, \theta)} \quad (2)$$

$$\approx \frac{1}{N} \sum_{n=1}^N \sum_{i=0}^{|\mathcal{M}|} m_n^i \log q_\Psi(m_n^i|x_n). \quad (3)$$

Experiments

We perform two experiments to demonstrate and test the accuracy of the approach, and one experiment to showcase the application in a computational neuroscience setting. In all numerical experiments, we used a two-layer neural network q_Ψ with hyperbolic tangent activation functions; and used ADAM (Kingma & Ba, 2014) to optimize the loss (eq. 3).

1. Gaussian toy example

We first provide an intuitive toy example – the problem of deciding between three Gaussian models with different means, given only a single data point as observation. Here, the training data consist of samples (m_n, x_n) from all three models, each with different Gaussian priors $p(\mu|m_i)$ on the mean parameter μ (Figure 2, top). After training with 50000 data points, the posterior probability of each model predicted for the whole range of observed test data points x_o closely matches the analytic posterior probability (Fig. 2, bottom).

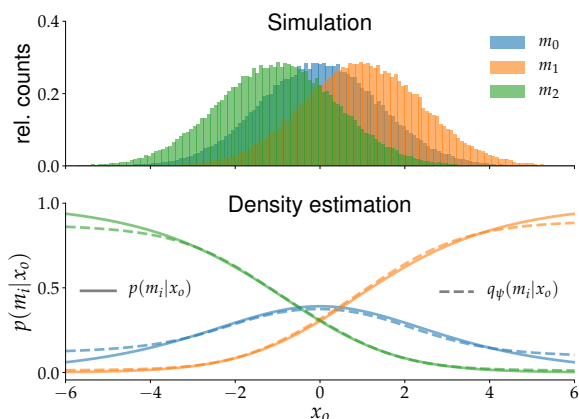


Figure 2: **Gaussian toy example:** Simulated training dataset (top) and predicted posterior probabilities of the three models (bottom).

2. Spike count models

Second, we present a model comparison problem from computational neuroscience for which the exact posteriors can be calculated: the comparison of the Poisson model vs. the negative-binomial model for spike count data. It relates to the question of whether spike counts generally have equal mean and variance (Poisson model) or whether the variance exceeds the mean, i.e., the counts are overdispersed (negative-binomial model, Taouali, Benvenuti, Wallisch, Chavane, & Perinet, 2015). Thus, a statistical model comparison procedure is expected to favour the overdispersed model whenever the variance is sufficiently big.

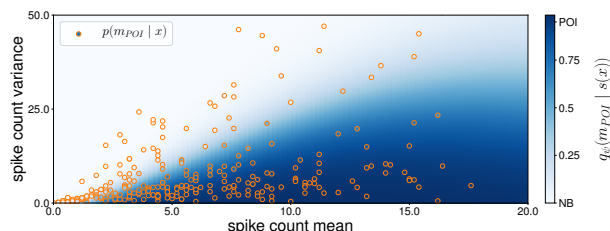


Figure 3: **Model comparison between overdispersed and Poisson spike count models:** Exact posterior probabilities (filled circles, orange edges) on top of predicted posterior probability for the Poisson model (background), same color code.

In this case, the training data consist of the mean and the variance of simulated spike counts from both models. We use a uniform prior $p(m)$ over the two models and Gamma priors over the parameters of the Poisson and the negative-binomial model to generate 50000 training data points and 300 test data points. To illustrate the performance, we compute the predicted posterior probability of the Poisson model $q_\psi(m_{POI}|s(x))$ for a fine grid of means and variance (Fig. 3, background) and plot the exact posterior probabilities $p(m_{POI}|x)$ of the test set on top using the same color code (Fig. 3, orange circles, foreground). The close match in color indicates accurate performance.

3. Ion channel models

Finally, we present a use case scenario from computational neuroscience for which no ground truth posterior is available: comparison of two ion channel models on the basis of observed current traces. Both ion channels models are taken from Pospischil et al. (2008), selected based on the ion channel genealogy (Podlaski et al., 2017). The output of the models consists in current traces in response to stereotypical voltage-clamp protocols applied to the membrane (Figure 4); summary statistics were calculated using PCA, as described in (Podlaski et al., 2017).

In this example, the priors over the two models and the priors over the model parameters are uniform. We train q_ψ with 100000 training data points to predict the probability of each

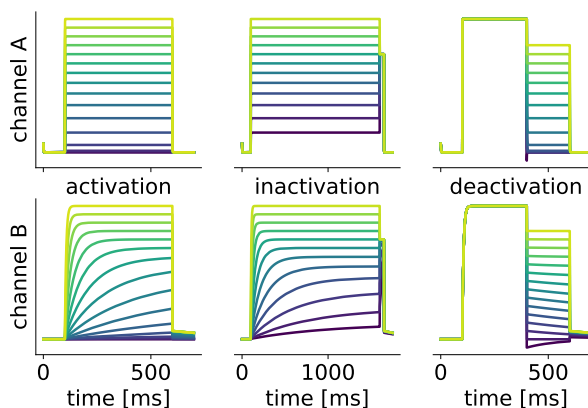


Figure 4: **Comparing models of ion channels:** Current traces of channel A (top) and channel B (bottom) in response to three voltage-clamp protocols.

channel model given observed summary statistics on 300 test data points. Because no ground truth posterior is available, no direct evaluation is possible here. However, we observe that q_ψ reliably assigns high posterior probability to the correct underlying channel model.

Evaluation

For a comparison to state-of-the-art model comparison methods, we apply ABC SMC (Toni et al., 2009) as provided in the pyABC toolbox (Klinger, Rickert, & Hasenauer, 2017) to our example models and compare it with the neural density estimation (NDE) method presented here. We base the comparison on the same number of training simulations. For the tractable examples, the absolute difference between predicted and exact posterior serves as an evaluation metric, while the performance on the ion channel example is assessed using the cross-entropy loss with respect to the true underlying model.

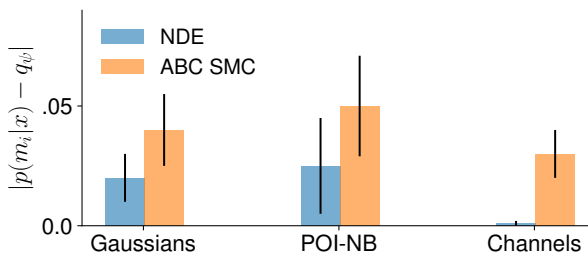


Figure 5: **Comparison of neural density estimation (NDE) and sequential Monte Carlo (ABC SMC) methods** on the presented example problems. Third column shows cross entropy loss, as ground-truth posteriors are not available for the ion-channel model.

Given the same number of simulations and based on a test set with 300 different observed data, NDE is competitive with ABC SMC on the analytically tractable examples. The comparison based on cross-entropy loss for the ion channel example shows that NDE clearly outperforms ABC SMC on this example.

Conclusions

Simulation-based models with intractable likelihoods play a central role in neuroscience and cognitive science (Palestro et al., 2018): recurrently connected spiking neurons (Ladenbauer, McKenzie, English, Hagens, & Ostojic, 2019), and models of decision making based on nonlinear diffusion mechanisms (Ratcliff & McKoon, 2008; Park, Lueckmann, von Kriegstein, Bitzer, & Kiebel, 2016) are examples of simulation-based models for which likelihoods are intractable or challenging to compute. The goal of this work is to provide statistical methods for performing Bayesian model comparison even on such complex mechanistic models. We show how neural density estimation for simulation-based inference can be extended to the task of Bayesian model comparison, and demonstrate its competitive performance on three example problems.

Our goal is to provide methods which will enable scientists to apply Bayesian model comparison to simulation-based models in computational neuroscience, cognitive science and beyond. Although we concentrated on training a neural network to do model comparison, one can perform both inference and model comparison simultaneously, which opens up new research possibilities in machine learning and neuroscience.

Acknowledgments

We thank CNE-members for discussions and feedback. Work was supported by the German Research Foundation (DFG) through SFB 1233 (276693517), SFB 1089 and SPP 2041 and the German Federal Ministry of Education and Research (BMBF, project 'ADMIMEM', FKZ 01IS18052 A).

References

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (Vol. 4) (No. 4). doi: 10.1117/1.2819119

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290.

Gerstner, W., Sprekeler, H., & Deco, G. (2012). Theory and simulation in neuroscience. *science*, *338*(6103), 60–65.

Greenberg, D., Nonnenmacher, M., & Macke, J. (2019, 09–15 Jun). Automatic posterior transformation for likelihood-free inference. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th international conference on machine learning* (Vol. 97, pp. 2404–2414). Long Beach, California, USA: PMLR.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*. Retrieved from <http://arxiv.org/abs/1412.6980>

Klinger, E., Rickert, D., & Hasenauer, J. (2017). pyabc: distributed, likelihood-free inference. *bioRxiv*, 162552.

Ladenbauer, J., McKenzie, S., English, D. F., Hagens, O., & Ostojic, S. (2019). Inferring and validating mechanistic models of neural microcircuits based on spike-train data. *bioRxiv*, 261016.

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., & Macke, J. H. (2017). Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in neural information processing systems* (pp. 1289–1299).

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., & Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition..

Papamakarios, G., & Murray, I. (2016). Fast ϵ -free inference of simulation models with bayesian conditional density estimation. In *Neurips* (pp. 1028–1036).

Park, H., Lueckmann, J.-M., von Kriegstein, K., Bitzer, S., & Kiebel, S. J. (2016). Spatiotemporal dynamics of random stimuli account for trial-to-trial variability in perceptual decision making. *Scientific Reports*, *6*(1). doi: 10.1038/srep18832

Podlaski, W. F., Seeholzer, A., Groschner, L. N., Miesenböck, G., Ranjan, R., & Vogels, T. P. (2017). Mapping the function of neuronal ion channels in model and experiment. *eLife*, *6*. doi: 10.7554/eLife.22152

Pospischil, M., Toledo-Rodriguez, M., Monier, C., Piwkowska, Z., Bal, T., Frégnac, Y., ... Destexhe, A. (2008). Minimal hodgkin-huxley type models for different classes of cortical and thalamic neurons. *Biological Cybernetics*, *99*(4-5), 427–441.

Ratcliff, R., & McKoon, G. (2008, April). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. doi: 10.1162/neco.2008.12-06-420

Rubin, D. B., et al. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, *12*(4), 1151–1172.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate bayesian computation. *PLoS Computational Biology*, *9*(1), e1002803.

Taouali, W., Benvenuti, G., Wallisch, P., Chavane, F., & Perrinet, L. U. (2015). Testing the odds of inherent vs. observed overdispersion in neural spike counts. *Journal of Neurophysiology*, *115*(1), 434–444.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, *6*(31), 187–202.