Thoughts on the Meeting: A Statistician's Perspective

Mikael Kuusela

Department of Statistics and Data Science,
Carnegie Mellon University

PhyStat-SBI 2024

Max Planck Institute for Physics,
Garching, Germany

May 17, 2024

Simulation-based inference

This meeting has been about making inference about a parameter of interest $\boldsymbol{\theta}$ given data $\boldsymbol{x}$ from a parametric model

$$\boldsymbol{x} \sim F_{\boldsymbol{\theta}}$$

when $F_{\boldsymbol{\theta}}$ is only available as a simulator

Ingredients:

- Sample of parameters: $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n \sim p(\boldsymbol{\theta})$
- Corresponding simulations from the model: $\boldsymbol{x}_i \sim F_{\boldsymbol{\theta}_i},\ i = 1, \ldots, n$
- Observed data: $\boldsymbol{x}_{\mathrm{obs}}$

Task: Infer $\boldsymbol{\theta}$ that generated $\boldsymbol{x}_{\mathrm{obs}}$ (i.e., produce point estimates, confidence sets, credible sets, posteriors, hypothesis tests, etc.)

Key insight: Machine learning enables us to do this with very high-dimensional $\boldsymbol{x}$

Simulation-based inference

"Simulation-based inference is a major evolution in the statistical capabilities for science, as it enables the analysis of complex models and data without simplifying assumptions."

— Gilles Louppe

The statistical fundamentals have not changed

While SBI has enabled inference in many previously intractable settings, it is important to keep in mind that it cannot circumvent fundamental limitations of statistics:

- Cramer–Rao lower bound: $\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$, where $I(\theta)$ is the Fisher information
- Uniformly most powerful tests do not exist in general
- Sufficient statistics only exist in exponential families
- Goodness-of-fit tests place power on specific alternatives
- ...

Two notions of coverage

When validating SBI techniques (or inferential procedures more generally), a common desideratum is the coverage of an interval estimator $[\underline{\theta}(\boldsymbol{x}), \overline{\theta}(\boldsymbol{x})]$ of θ (for simplicity, let's take θ to be scalar here)

It's good to keep in mind that there are two different notions of coverage that often get mixed up:

Marginal coverage: $\mathbb{P}_{\boldsymbol{x},\theta}(\theta \in [\underline{\theta}(\boldsymbol{x}), \overline{\theta}(\boldsymbol{x})]) = 1 - \alpha$, where both θ and $\boldsymbol{x}$ are random inside the probability statement

Conditional coverage: $\mathbb{P}_{\boldsymbol{x}|\theta}(\theta \in [\underline{\theta}(\boldsymbol{x}), \overline{\theta}(\boldsymbol{x})]) = 1 - \alpha$, for all θ, where $\boldsymbol{x}$ is random but θ is fixed inside the probability statement

Even though these look similar, these are fundamentally different notions of what we mean by "uncertainty"

- Marginal coverage is easier to achieve but is a weaker notion (in fact, conditional coverage implies marginal coverage but the reverse is not true)
- Marginal coverage only makes sense if it is sensible to think of θ as being random

SBI for purely statistical models

Purely statistical models ("spherical cows") vs. mechanistic simulators:

Domain	Purely statistical model	Mechanistic simulator
Oceanography	Gaussian process regression of irregularly sampled observations	Data assimilation with general circulation models
Epidemiology	ARIMA time series models	Compartmental models
Finance	Stochastic volatility models	??
Particle physics	??	MC generators

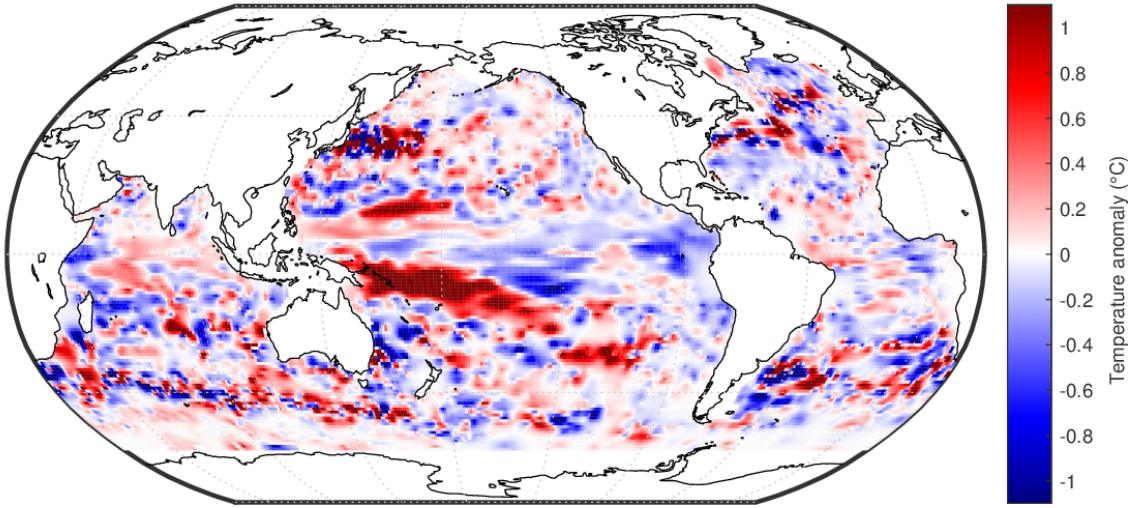

Figure: Spatio-temporal interpolation of subsurface ocean temperature anomalies using moving window-based locally stationary Gaussian processes (Kuusela and Stein, 2018)

SBI for spatial statistics

In recent years, there has been an explosion of interest in SBI for spatial statistics:

- Neural prediction for spatial models (Gerber and Nychka, 2021; Lenzi et al., 2023; Sainsbury-Dale et al., 2024)
- Neural likelihood for spatial models (Walchessen et al., 2024)
- Neural prediction with censored observations (Richards et al., 2023)
- Neural prediction with irregularly spaced observations (Sainsbury-Dale et al., 2023)

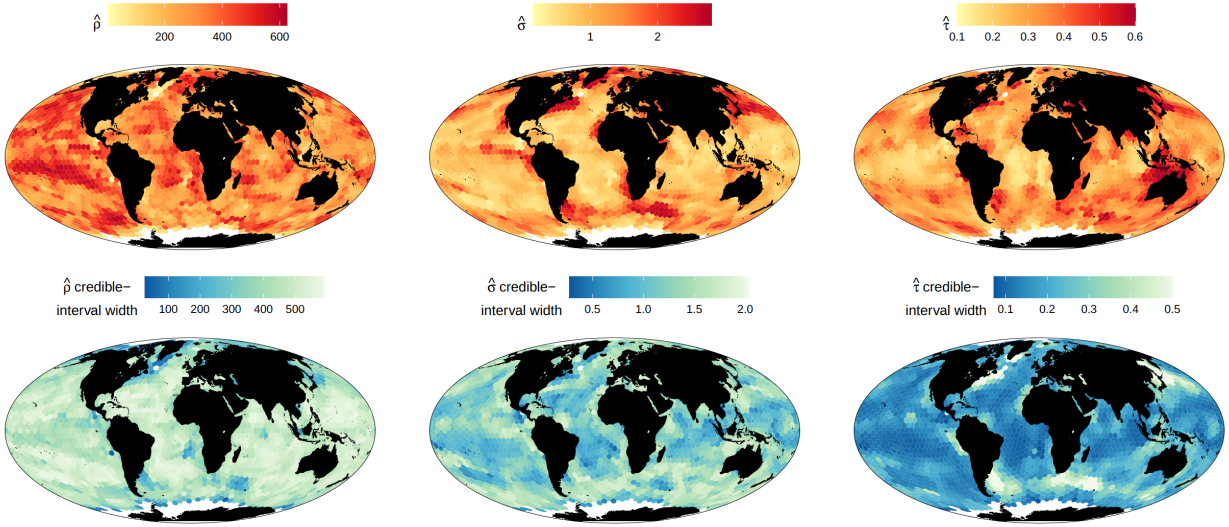

Figure: Neural prediction for satellite sea surface temperature using locally stationary Gaussian processes (Sainsbury-Dale et al., 2023)

Model misspecification

We have heard of at least the following approaches to model misspecification:

- Adding nuisance parameters (several talks)
- Optimal transport (talk by A. Wehenkel)
- Nonparametric Bayes + MMD (talk by H. Dellaporta)
- ...

One more approach to consider: adding a stochastic model discrepancy term (Kennedy and O'Hagan, 2001)

$$x_i = f(\boldsymbol{u}_i, \boldsymbol{\theta}) + \varepsilon_i \quad \hookrightarrow \quad x_i = f(\boldsymbol{u}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{u}_i) + \varepsilon_i,$$

where $\delta(\cdot)$ is a Gaussian process and $\boldsymbol{u}_i$ are control variables

Overall, there has been a lot of work in applied mathematics on variants of this approach to handling *model discrepancy*

Model misspecification

Climate science is another field that has had to think hard about misspecification of simulators

In that case, handling model misspecification is called *climate model bias correction*

A vast literature exits on this topic, including methods based on optimal transport

- Search Google Scholar for "climate model multivariate bias correction"

"Neyman inversion is slow"

Naive sampling-based Neyman inversion is very heavy computationally

However, with LF2I (see talks by A. Lee and L. Masserano), the computational complexity of the Neyman inversion step is comparable to the complexity of learning the test statistic

Huge opportunity in trying this out with LHC SBI analyses!

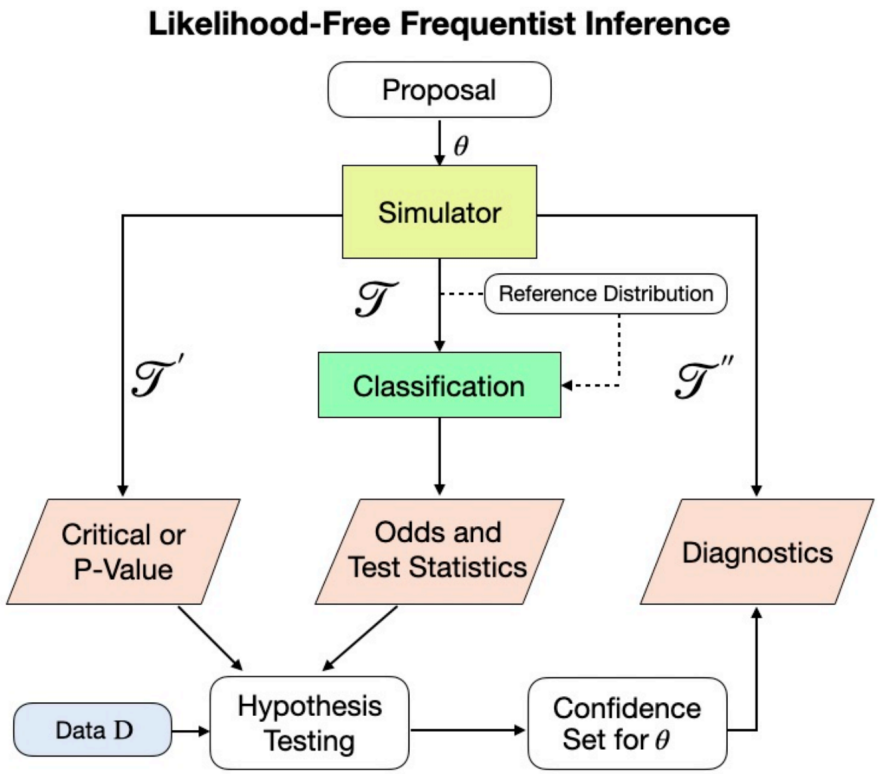

Figure: Learning critical values, test statistics and empirical coverage can all be performed efficiently using machine learning (Dalmasso et al., 2020, 2023)

How to best sample the parameters?

All SBI methods require us to produce a sample of parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$

If we want this sample to be uniform over the parameter space Θ, then Θ needs to be bounded

Alternatively, need to specify a reference distribution $p(\boldsymbol{\theta})$ from which $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ are sampled

- But this requires giving higher priority to some regions of Θ

Lots of discussion on adaptive / sequential / active / importance sampling of Θ

- Intuitively would like to sample more in regions of high likelihood / posterior
- But this seems to be fundamentally at odds with amortized inference...

For sampling-based Neyman inversion, would like to sample more where the distribution of the test statistic varies most

- In some ongoing work with constrained parameters, we're seeing that this makes a big difference because the test statistic distribution changes rapidly near the boundaries of the feasible set

How robust is neural SBI?

Any simulator $F_{\boldsymbol{\theta}}$ is misspecified

So when the neural likelihood / likelihood ratio / posterior is evaluated for $\boldsymbol{x}_{\mathrm{obs}}$, it is being evaluated out-of-training-distribution

How robust are these networks against this?

There are well-known adversarial attacks against neural networks

How much of a concern is this for neural SBI?

Does SBI scale to high-dimensional parameters?

SBI demonstrably scales to very high-dimensional $\boldsymbol{x}$

But do these methods also work for high-dimensional $\boldsymbol{\theta}$?

It seems to me that the message on this has been mixed

It's not clear to me if there should or shouldn't be an asymmetry in the dimension scaling of $\boldsymbol{x}$ and $\boldsymbol{\theta}$

Would appreciate further discussion on this and hearing about practical experience with high-dimensional $\boldsymbol{\theta}$

One thing that is clear is that frequentist inference with high-dimensional $\boldsymbol{\theta}$ has limitations that go beyond SBI

- High-dimensional posteriors are "easy" to work with because one can marginalize
- But projections of high-dimensional frequentist confidence sets give extremely conservative low-dimensional confidence sets

References I

N. Dalmasso, R. Izbicki, and A. Lee. Confidence sets and hypothesis testing in a likelihood-free inference setting. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2323–2334. PMLR, 13–18 Jul 2020.

N. Dalmasso, L. Masserano, D. Zhao, R. Izbicki, and A. B. Lee. Likelihood-free frequentist inference: Bridging classical statistics and machine learning for reliable simulator-based inference. arXiv:2107.03920 [stat.ML], 2023.

F. Gerber and D. Nychka. Fast Covariance Parameter Estimation of Spatial Gaussian Process Models using Neural Networks. *Stat*, 10(1):e382, 2021. doi: https://doi.org/10.1002/sta4.382.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001. doi: https://doi.org/10.1111/1467-9868.00294.

M. Kuusela and M. L. Stein. Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A*, 474:20180400, 2018.

A. Lenzi, J. Bessac, J. Rudi, and M. L. Stein. Neural Networks for Parameter Estimation in Intractable Models. *Computational Statistics & Data Analysis*, 185: 107762, 2023. doi: https://doi.org/10.1016/j.csda.2023.107762.

References II

J. Richards, M. Sainsbury-Dale, A. Zammit-Mangion, and R. Huser. Neural Bayes estimators for censored inference with peaks-over-threshold models. Preprint arXiv:2306.15642 [stat.ME], 2023.

M. Sainsbury-Dale, J. Richards, A. Zammit-Mangion, and R. Huser. Neural Bayes estimators for irregular spatial data using graph neural networks. Preprint arXiv:2310.02600 [stat.ME], 2023.

M. Sainsbury-Dale, A. Zammit-Mangion, and R. Huser. Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*, 78(1):1–14, 2024. doi: 10.1080/00031305.2023.2249522.

J. Walchessen, A. Lenzi, and M. Kuusela. Neural likelihood surfaces for spatial processes with computationally intensive or intractable likelihoods. Preprint arXiv:2305.04634 [stat.ME], 2024.

Backup