Approximating likelihood ratios with calibrated classifiers

Gilles Louppe
DIANA meeting

February 22, 2016

Likelihood ratio

We want to evaluate the likelihood-ratio statistic

$$\lambda(\mathcal{D}; \theta_0, \theta_1) = \prod_{\mathbf{x} \in \mathcal{D}} \frac{p_{\mathbf{X}}(\mathbf{x}|\theta_0)}{p_{\mathbf{X}}(\mathbf{x}|\theta_1)} \tag{1}$$

in the likelihood-free setting, i.e. when $p_{\mathbf{X}}(\mathbf{x}|\theta_0)$ cannot be evaluated but samples $\mathbf{x}$ can be drawn from p_{θ_0} (resp. for θ_1).

Issue. The input space $\mathcal{X}$ may be high dimensional, making it very difficult to build an approximate of $p_{\mathbf{X}}(\mathbf{x}|\theta_0)$.

Equivalent statistic

Theorem.

$$r(\mathbf{x};\theta_0,\theta_1) = \frac{p_{\mathbf{X}}(\mathbf{x}|\theta_0)}{p_{\mathbf{X}}(\mathbf{x}|\theta_1)} = \frac{p_{\mathbf{U}}(u = s(\mathbf{x})|\theta_0)}{p_{\mathbf{U}}(u = s(\mathbf{x})|\theta_1)} \tag{2}$$

provided the change of variable $\mathbf{U} = s(\mathbf{X})$ is monotonic with $r(\mathbf{x};\theta_0,\theta_1)$.

Idea. $s(\mathbf{x})$ projects x into a 1D space in which only the informative content of $r(\mathbf{x};\theta_0,\theta_1)$ is preserved. Building an approximate of $p_{\mathbf{U}}(u = s(\mathbf{x})|\theta_0)$ is now easy!

Approximating likelihood ratios with classifiers

A classifier trained to distinguish samples $\mathbf{x} \sim p_{\theta_0}$ from samples $\mathbf{x} \sim p_{\theta_1}$ eventually models

$$s^*(\mathbf{x}) = \frac{p_{\mathbf{X}}(\mathbf{x}|\theta_1)}{p_{\mathbf{X}}(\mathbf{x}|\theta_0) + p_{\mathbf{X}}(\mathbf{x}|\theta_1)}, \tag{3}$$

which satisfies conditions of the theorem.

Idea. For building an equivalent likelihood-ratio statistic, combine supervised learning for learning $s(x)$ with calibration for learning $p(s(x))$.

Likelihood-free inference

$$
\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} p(\mathcal{D}|\theta) \\
&= \arg\max_{\theta} \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta_1)} \\
&= \arg\max_{\theta} \prod_{\mathbf{x}\in\mathcal{D}} \frac{p(s(\mathbf{x};\theta,\theta_1)|\theta)}{p(s(\mathbf{x};\theta,\theta_1)|\theta_1)} \ , \qquad (4)
\end{aligned}
$$

where, for computational efficiency, $s(\mathbf{x};\theta,\theta_1)$ can be a single classifier parameterized by θ, θ_1.

Note. This can then be used for computing profile likelihood ratio tests, taking into account nuisance parameters.

For more details...

Approximating Likelihood Ratios with Calibrated Discriminative Classifiers

Kyle Cranmer, Juan Pavez and Gilles Louppe
New York University

February 16, 2016

Abstract

In many fields of science, generalized likelihood ratio tests are established tools for statistical inference. In practice, these tests are often complicated by the fact that computer simulators are used as a generative model for the data, which does not provide a way to evaluate the likelihood function. In this paper, we demonstrate that likelihood ratios are invariant under dimensionality reductions $\mathbb{R}^p \mapsto \mathbb{R}$, provided the transformation is itself monotonic with the likelihood ratio. As a direct consequence, we show that discriminative classifiers can be used to approximate the generalized likelihood ratio statistic when only a generative model for the data is available. In particular, the proposed method offers a machine learning-based approach to statistical inference that is complementary to likelihood-free Bayesian inference algorithms, such as Approximate Bayesian Computation, as it does not require the definition of a prior over model parameters. Experimental results on artificial problems illustrate the potential of the proposed method.

Keywords: likelihood ratio, likelihood-free inference, classification, particle physics

New version of http://arxiv.org/abs/1506.02169, in preparation for submission to JASA.

Carl, a likelihood-free inference toolbox for Python

- Approximation of likelihood ratios with classifiers
  - Supervised learning via Scikit-Learn ✓
  - Calibration (histograms, KDE, isotonic regression) ✓
  - Automatic decomposition of mixtures ✓
  - Parameterized approximated ratios for inference (in progress)
  - See toy example
- Canonical inference examples (in progress)
- (Minimal) Composition and fitting of PDFs, à la RooFit ✓
  - See API