Direct unfolded density estimation

Kyle Cranmer

June 5, 2018

Abstract

We describe an approach to high-dimensional, unbinned unfolding problems, when the stochastic folding process is implicitly defined by a simulator.

1 Introduction

Consider a random variable $X \sim p_X$ that can be regarded as the "folding" of a latent variable $Z \sim p_Z$ through some stochastic process described by $p(X|Z)$. We can write the density for X as

$$p_X(x) = \int dz \, p_{X|Z}(x|z) \, p_Z(z) \; . \tag{1}$$

We assume that we have a simulation of the stochastic process $p(X|Z)$ that provides training pairs $\{(x,z)_j\}_{j=1}^{M}$ that we can use to learn $\hat{p}_{X|Z}(x|z)$, our estimate $p_{X|Z}(x|z)$.

Given some observations $\{x_i\}_{i=1}^{N}$, our goal is to estimate the "unfolded" distribution p_Z. This is an inverse problem, often referred to as "unfolding" in particle physics. This inverse problem is *ill posed*, and is typically regularized via Tikhonov regularization. There is a connection to Tikhonov regularization and reproducing kernel Hilbert spaces (RKHS). Here we consider a similar setting. our focus is not on regularization, which can be seen as a penalty or prior on $\hat{p}_Z$, but instead on recent methods for direct density estimation of the unfolded distribution.

1.1 density estimation via a bijection

Our starting point is to consider the idealized case where we observe z. We can estimate the density p_Z with techniques that use a bijection $f : Z \to V$ (e.g. an invertible flow or autoregressive model) and a tractable density $p_V(v)$. In particular

$$p_Z(z) = p_V(f_\theta(z)) \, J_z \tag{2}$$

where

$$J_z = \left| \det \frac{\partial f_\theta(z)}{\partial z_T} \right| \tag{3}$$

and θ are the internal network parameters for the bijection f_θ. Learning proceeds via gradient ascent $\nabla_\theta \sum_i \log p_Z(z_i)$ with data z_i (i.e. maximum likelihood wrt. the internal parameters θ).

Of course, we don't estimate the latent variable Z, but we can still use this technique as described below.

1.2 learning the folding process with a conditional bijection

Similarly, we can estimate the stochastic folding process with a conditional bijection $g_z : X \to U$, a tractable density $p_U(u)$, and training data from the simulator. In particular,

$$p_{X|Z}(x|z) = p_U(g_{\phi;z}(x)) \, J_{x|z} \; , \tag{4}$$

where

$$J_{x|z} = \left| \det \frac{\partial g_{\phi,z}(x)}{\partial x_T} \right| \tag{5}$$

and ϕ are the internal network parameters for the bijection $g_{\phi;z}$. Learning of the stochastic folding process proceeds via gradient ascent $\nabla_\phi \sum_i \log p(x_j|z_j)$ with training data $(x,z)_j$.

1.3 learning the unfolded density

Once we have the estimate $\hat{p}_{X|Z}(x|z)$, with ϕ fixed, we can estimate $p_Z(z)$ from observations $\{x_i\}_{i=1}^N$ using Eq. 6. In particular we can approximate the integral $\int dz p(z)$ by efficiently sampling $\{v_k\}_{k=1}^K$ from p_V, which induces an efficient sampling of $z_k = f^{-1}(v_k)$. In particular

$$p_X(x) = \int dz\, p_{X|Z}(x|z)\, p_Z(z) \approx \sum_{k=1}^{K} p_U(g_{\phi;z_k}(x))\, J_{x|z_k)}\ J_{z_k}\ . \tag{6}$$

Learning proceeds via gradient ascent $\nabla_\theta \sum_i \log p(x_i)$ with data x_i (i.e. maximum likelihood wrt. the internal parameters θ). We envisage mini-batches of $\{v_k\}_{k=1}^K$. Critically, $\nabla_\theta \sum_i \log p(x_i)$ includes a term $\nabla_\theta \sum_k \log J_{z_k}$, where J_{z_k} depends on θ via Eq. 3.

1.4 Discussion

The approach described above corresponds to a maximum likelihood estimate in the ill-posed inverse problem, thus there is still need for regularization for $p_Z z$. However, the approach described above works for un-binned data, and situations where the stochastic folding process $p_{X|Z}$ is implicitly defined by a simulator. Furthermore, X and Z may be high dimensional. Thus, this provides an in-roads to high-dimensional, unbinned unfolding using machine learning.