Importance Sampling. Minimization of Renyi Divergence

Alena Shilova

Skoltech

August 02,2017

Importance Sampling

The problem of approximating the expected value (or integral):

$$I = \mathbb{E}\varphi(X) = \int \varphi(x) f(x)\, dx,$$

where $X \sim f$ is a random variable on $\mathbb{R}^k$ and $\varphi : \mathbb{R}^k \to \mathbb{R}$.

IS estimator

Sampling (importance) distribution $\tilde{f}$ satisfying $\tilde{f}(x) > 0$ whenever $\varphi(x)f(x) \neq 0$, take IID samples $X_1, X_2, \ldots, X_n \sim \tilde{f}$ (as opposed to sampling from f, the nominal distribution) and use

$$\hat{I}_n^{IS} = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) \frac{f(X_i)}{\tilde{f}(X_i)}.$$

Exponential family

From now on, one is going to use the family of exponential distributions and instead of $\tilde{f}(x)$, there will be

$$f_\theta(x) = \exp(\theta^T T(x) - A(\theta))h(x),$$

where $A : \mathbb{R}^p \to \mathbb{R} \cup \infty$, defined as

$$A(\theta) = \log \int \exp(\theta^T T(x))h(x)\, dx,$$

serves as a normalizing factor.

Rényi divergence. Definition

Rényi generalized divergence:

$$D_\alpha(g; f) = \begin{cases} \int \ln \frac{g(x)}{f(x)} g(x)\, dx & \alpha = 1 \\ \frac{1}{\alpha - 1} \ln \left(\int \left[\frac{g(x)}{f(x)} \right]^{\alpha - 1} g(x)\, dx \right) & \alpha > 0; \alpha \neq 1 \end{cases}$$

Further let's consider the case of minimizing the divergence with $\alpha = 1$, which is equivalent to

$$\max_\theta \int g(x) \ln f_\theta(x)\, dx$$

Optimization problem

In our case we have the following optimization problem:

$$\max_{\theta \in \Theta} \frac{1}{I} \int \varphi(x) f(x) \left[\theta^T T(x) - A(\theta)\right] \ln h(x)\, dx$$

It's concave as $A(\theta)$ is a convex function.
The first derivative:

$$\int \left[T(x) - \nabla A(\theta)\right] \varphi(x) f(x) \ln h(x)\, dx =$$

$$= \mathbb{E}_{X \sim f_\theta} \left[(T(x) - A(\theta)) \frac{\varphi(x) f(x) \ln h(x)}{f_\theta(x)} \right]$$