Importance Sampling Integration with Truncated Gaussian Mixture Model

K. Sargsyan
Sandia National Laboratories, Livermore, CA
March 10, 2022

1 Goal

Consider a function $E(\boldsymbol{x})$ where $\boldsymbol{x} \in \mathbb{R}^d$, and an integral of a form

$$\mathbb{I}[f] = \int_\Omega f(\boldsymbol{x}) d\boldsymbol{x}, \tag{1}$$

where $f(\boldsymbol{x}) = e^{-\beta E(\boldsymbol{x})}$ and $\beta = 1/(kT)$.

2 Importance Sampling

2.1 Default Setup

Importance sampling Monte-Carlo (ISMC) is an integration technique to help find weighted integrals of a form

$$\mathbb{I}[f] = \int_\Omega f(\boldsymbol{x}) \pi(\boldsymbol{x}) d\boldsymbol{x}, \tag{2}$$

where $\pi(\boldsymbol{x})$ is a probability density function (PDF), i.e. $\pi(\boldsymbol{x})$ and $\int_\Omega \pi(\boldsymbol{x}) d\boldsymbol{x} = 1$.

The ISMC estimate of the integral is

$$\mathbb{I}_\pi[f] \approx \frac{1}{N} \sum_{i=1}^N f(\boldsymbol{x}_i) \tag{3}$$

in which the samples $\{\boldsymbol{x}_i\}_{i=1}^N$ are drawn from the PDF $\pi(\boldsymbol{x})$.

2.2 Importance Sampling for Unweighted Integration

For general $\pi(\boldsymbol{x})$, the unweighted integral (1), and consequently its ISMC estimate can be written as

$$\mathbb{I}_\pi[f] = \int_\Omega \frac{f(\boldsymbol{x})}{\pi(\boldsymbol{x})} \pi(\boldsymbol{x}) d\boldsymbol{x} \approx \frac{1}{N} \sum_{i=1}^N \frac{f(\boldsymbol{x}_i)}{\pi(\boldsymbol{x}_i)}, \tag{4}$$

where the samples $\{\boldsymbol{x}_i\}_{i=1}^N \sim \pi(\boldsymbol{x})$, i.e. are drawn from the PDF $\pi(\boldsymbol{x})$.

Note that we have a freedom to choose any PDF $\pi(\boldsymbol{x})$. For example, when $\pi(\boldsymbol{x})$ is uniform on the region Ω, i.e. $\pi(\boldsymbol{x}) = 1/|\Omega|$, we get a regular Monte-Carlo estimate

$$\mathbb{I}_\pi[f] \approx \frac{|\Omega|}{N} \sum_{i=1}^N f(\boldsymbol{x}_i). \tag{5}$$

However, the estimate in (4) can have very high variance if one is not careful in choosing $\pi(\boldsymbol{x})^*$. The closer $\pi(\boldsymbol{x})$ is to $f(\boldsymbol{x})$ (up to a normalizing constant), the better the integration scheme in (4) is.

3 Gaussian ISMC

The conventional Gaussian ISMC is when one chooses $\pi(\boldsymbol{x})$ to be a PDF of a multivariate normal (MVN)

$$\pi_{\mathrm{MVN}(\boldsymbol{\mu},\, \boldsymbol{C})}(\boldsymbol{x}) = \frac{1}{(2\pi|\boldsymbol{C}|)^{d/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{6}$$

Here clearly $\Omega = \mathbb{R}^d$ and the MVN samples drawn for the integration are easy to generate. This works well for unimodal functions $f(\boldsymbol{x}) = e^{-\beta E(\boldsymbol{x})}$ which have local behavior near minimum of $E(\boldsymbol{x})$ according to

$$f(\boldsymbol{x}) = e^{-\beta E(\boldsymbol{x})} \propto e^{-\frac{\beta}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T \boldsymbol{H}(\boldsymbol{x}-\boldsymbol{x}_0)}, \tag{7}$$

where $\boldsymbol{x}0$and $\boldsymbol{H}$ are the minimum and Hessian there, respectively.

Given Eqs. (6) and (7) we note that to make the integration as efficient as possible, $\boldsymbol{\mu} \approx \boldsymbol{x}_0$ (i.e. needs to be close to the minimum of $E(\boldsymbol{x})$) and $\boldsymbol{C} \approx (\beta \boldsymbol{H})^{-1}$, i.e. ideally should be close to the inverse-Hessian of $E(\boldsymbol{x})$ at the minimum, saled by β.

4 Gaussian Mixture Model ISMC

We do not have a unimodal integrand, hence we will consider a Gaussian Mixture Model (GMM) PDF with GMM centers coinciding with minima of $E(\boldsymbol{x})$ with the appropriate covariances that match with the Hessians at minima of $E(\boldsymbol{x})$.

$$\pi_{\mathrm{GMM}} = \sum_{k=1}^{K} w_k \pi_{\mathrm{MVN}(\boldsymbol{\mu}_k, \boldsymbol{C}_k)}(\boldsymbol{x}), \tag{8}$$

where K is the number of minima of $E(\boldsymbol{x})$, and $\boldsymbol{\mu}_k = \boldsymbol{x}_{0k}$ are the locations of the minima, while $\boldsymbol{C}_k$'s are the β-scaled inverse-Hessians at the minima (i.e. $\boldsymbol{C}_k = (\beta \boldsymbol{H}_k)^{-1}$). Additionally, the weights w_k need to be chosen ....

*we have seen this when doing uniform sampling.