Importance Sampling Integration with Truncated Gaussian Mixture Model

K. Sargsyan
Sandia National Laboratories, Livermore, CA
April 8, 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 Monte Carlo Integration

Monte-Carlo integration assumes we collect uniformly random samples $\{\boldsymbol{x}_i\}_{i=1}^N$ in domain Ω, and evaluate the integral in (1) with

$$\mathbb{I}_{MC}[f] \approx \frac{|\Omega|}{N} \sum_{i=1}^N f(\boldsymbol{x}_i), \tag{2}$$

where $|\Omega|$ is the volume of integration.

3 Importance Sampling

3.1 Default Setup

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

$$\mathbb{I}[f] = \int_\Omega f(\boldsymbol{x}) \pi(\boldsymbol{x}) d\boldsymbol{x}, \tag{3}$$

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

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

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

3.2 Importance Sampling for Unweighted Integration

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

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

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

$$\mathbb{I}_\pi[f] \approx \frac{|\Omega|}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i), \tag{6}$$

coinciding with (2).

However, the estimate in (5) 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 (5) is. The next subsections detail increasingly complex choices of $\pi(\boldsymbol{x})$ in order to match the shape of $f(\boldsymbol{x})$ as closely as possible.

3.2.1 Gaussian ISMC

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

$$\pi_{\mathrm{MVN}(\boldsymbol{\mu},\,\boldsymbol{C})}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{C}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{7}$$

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

$$f(\boldsymbol{x}) = e^{-\beta E(\boldsymbol{x})} \propto e^{-\frac{\beta}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\boldsymbol{H}(\boldsymbol{x}-\boldsymbol{x}_0)}, \tag{8}$$

where $\boldsymbol{x}_0$and $\boldsymbol{H}$ are the minimum and Hessian there, respectively.

Given Eqs. (7) and (8) 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, scaled by β.

3.2.2 Gaussian Mixture Model (GMM) ISMC

If the integrand is not unimodal, one should consider a Gaussian Mixture Model (GMM) PDF defined by weights w_k, means $\boldsymbol{\mu}_k$and covariances $\boldsymbol{C}_k$, such that $\sum_{k=1}^K w_k = 1$, and

$$\pi_{\mathrm{GMM}} = \sum_{k=1}^{K} w_k \pi_{\mathrm{MVN}(\boldsymbol{\mu}_k,\boldsymbol{C}_k)}(\boldsymbol{x}). \tag{9}$$

*we have seen this when doing uniform sampling, i.e. MC integration.

To make $\pi_{\rm GMM}$ as close to the integrand as possible, we should choose K to be the number of minima of $E(\boldsymbol{x})$, and

$$\boldsymbol{\mu}_k = \boldsymbol{x}_{0k} \tag{10}$$

are the locations of the minima, while $\boldsymbol{C}_k$'s are the β-scaled inverse-Hessians at the minima, i.e.

$$\boldsymbol{C}_k = (\beta \boldsymbol{H}_k)^{-1}. \tag{11}$$

Additionally, the weights w_k need to be chosen to keep

$$\frac{w_k}{f(\boldsymbol{\mu}_k)|\boldsymbol{C}_k|^{1/2}} = \text{const}, \tag{12}$$

so the peaks of $\pi_{\rm GMM}$ are best aligned for those of $f(\boldsymbol{x})$.

3.2.3 Truncated Gaussian Mixture Model (GMMT) ISMC

The Gaussian and GMM PDFs are defined on $\Omega = R^d$. If the integration domain is a hypercube instead, i.e. $\Omega = [a_1, b_1] \times \cdots \times [a_d, b_d]$, we just need to compute the appropriate scaling factors. For example, the truncated MVN (MVNT) PDF is

$$\pi_{\rm MVNT}(\boldsymbol{x}) = F_\Omega(\boldsymbol{\mu}, \boldsymbol{C}) \pi_{{\rm MVN}(\boldsymbol{\mu}, \boldsymbol{C})}(\boldsymbol{x}) \mathbb{I}_{\boldsymbol{x} \in \Omega}, \tag{13}$$

where the factor is defined as

$$F_\Omega(\boldsymbol{\mu}, \boldsymbol{C}) = \frac{1}{\int_\Omega \pi_{{\rm MVN}(\boldsymbol{\mu}, \boldsymbol{C})}(\boldsymbol{x}) d\boldsymbol{x}}, \tag{14}$$

and $\mathbb{I}_{\boldsymbol{x} \in \Omega}$ is the indicator function of Ω (i.e. it is 1 if $x \in \Omega$ and 0 otherwise). Closed-form formula for $F_\Omega(\boldsymbol{\mu}, \boldsymbol{C})$ is given in `https://www.cesarerobotti.com/wp-content/uploads/2019/04/JCGS-KR.pdf`.

Finally, the GMMT PDF is simply a linear combination of MVNT PDFs from (13), i.e.

$$\pi_{\rm GMMT}(\boldsymbol{x}) = \mathbb{I}_{\boldsymbol{x} \in \Omega} \sum_{k=1}^{K} w_k F_\Omega(\boldsymbol{\mu}_k, \boldsymbol{C}_k) \pi_{{\rm MVN}(\boldsymbol{\mu}_k, \boldsymbol{C}_k)}(\boldsymbol{x}). \tag{15}$$

To summarize, our final integration scheme is (5), where the PDF $\pi(\boldsymbol{x})$ is chosen to be (15), with parameters of GMMT, $w_k, \boldsymbol{\mu}_k, \boldsymbol{C}_k$ selected according to Eqs. (12), (10) and (11), respectively. The samples in the summation (5) are also drawn from GMMT (5), which is straightforward task, based on multinomial sampling (to pick the modality), MVN sampling (to sample Gaussian for selected modality) and rejection sampling (to reject samples outside Ω).