

Importance Sampling Integration with Truncated Gaussian Mixture Model

K. Sargsyan

Sandia National Laboratories, Livermore, CA

April 8, 2022

1 Goal

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

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

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

2 Monte Carlo Integration

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

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

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

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

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

The ISMC estimate of the integral is

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

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

3.2 Importance Sampling for Unweighted Integration

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

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

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

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

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

coinciding with (2).

However, the estimate in (5) can have very high variance if one is not careful in choosing $\pi(\mathbf{x})^*$. The closer $\pi(\mathbf{x})$ is to $f(\mathbf{x})$ (up to a normalizing constant), the better the integration scheme in (5) is. The next subsections detail increasingly complex choices of $\pi(\mathbf{x})$ in order to match the shape of $f(\mathbf{x})$ as closely as possible.

3.2.1 Gaussian ISMC

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

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

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

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

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

Given Eqs. (7) and (8) we note that to make the integration as efficient as possible, $\boldsymbol{\mu} \approx \mathbf{x}_0$ (i.e. needs to be close to the minimum of $E(\mathbf{x})$) and $\mathbf{C} \approx (\beta \mathbf{H})^{-1}$, i.e. ideally should be close to the inverse-Hessian of $E(\mathbf{x})$ 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 \mathbf{C}_k , such that $\sum_{k=1}^K w_k = 1$, and

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

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

To make π_{GMM} as close to the integrand as possible, we should choose K to be the number of minima of $E(\mathbf{x})$, and $\boldsymbol{\mu}_k = \mathbf{x}_{0k}$ are the locations of the minima, while \mathbf{C}_k 's are the β -scaled inverse-Hessians at the minima (i.e. $\mathbf{C}_k = (\beta \mathbf{H}_k)^{-1}$). Additionally, the weights w_k need to be chosen to keep

$$\frac{w_k}{f(\boldsymbol{\mu}_k)|\mathbf{C}_k|^{1/2}} = \text{const}, \quad (10)$$

so the peaks of π_{GMM} are best aligned for those of $f(\mathbf{x})$.

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

$$\pi_{\text{MVNT}}(\mathbf{x}) = F_{\Omega}(\boldsymbol{\mu}, \mathbf{C}) \pi_{\text{MVN}}(\boldsymbol{\mu}, \mathbf{C})(\mathbf{x}) \mathbb{I}(\mathbf{x} \in \Omega), \quad (11)$$

where the factor

$$F_{\Omega}(\boldsymbol{\mu}, \mathbf{C}) = \frac{1}{\int_{\Omega} \pi_{\text{MVN}}(\boldsymbol{\mu}, \mathbf{C})(\mathbf{x}) d\mathbf{x}}. \quad (12)$$

Closed-form formula for $F_{\Omega}(\boldsymbol{\mu}, \mathbf{C})$ 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 (11), i.e.

$$\pi_{\text{GMMT}}(\mathbf{x}) = \mathbb{I}(\mathbf{x} \in \Omega) \sum_{k=1}^K w_k F_{\Omega}(\boldsymbol{\mu}_k, \mathbf{C}_k) \pi_{\text{MVN}}(\boldsymbol{\mu}_k, \mathbf{C}_k)(\mathbf{x}). \quad (13)$$