1 Simple Monte Carlo

$$\mathbb{E}_{p\in\mathcal{S}}[x(p)] \equiv \frac{1}{|\mathcal{S}|}\sum_{p\in\mathcal{S}} x(p) \tag{1}$$

$$\mathbb{E}_{p\in\mathcal{S}}[x(p)] \approx \frac{1}{S}\sum_{s=1}^{S} x\left(p^{(s)}\right) \tag{2}$$

$$\begin{aligned} p(x \mid \mathcal{D}) = \int p(x \mid \theta, \mathcal{D}) p(\theta \mid \mathcal{D}) \mathrm{d}\theta = \mathbb{E}_{p(\theta\mid\mathcal{D})}[p(x \mid \theta, \mathcal{D})] \\ \approx \frac{1}{S}\sum_{s=1}^{S} p\left(x \mid \theta^{(s)}, \mathcal{D}\right), \quad \theta^{(s)} \sim p(\theta \mid \mathcal{D}) \end{aligned} \tag{3}$$

Monte Carlo is usually simple and its $\frac{1}{\sqrt{R}}$scaling of error bars ”independent of dimensionality” may be good enough.

1.1 Rejection Sampling

Straightforward explanation: rejection sampling is a method for generating random samples from a simple distribution and accepting those that within that target distribution.

1.2 Importance Sampling

Sampling over the expectation of the target distribution.

$$\begin{aligned} \int f(x)p(x)\mathrm{d}x &= \int f(x)\frac{p(x)}{q(x)}q(x)\mathrm{d}x, \quad \text{if } q(x) > 0 \text{ wherever } p(x) > 0 \\ &= \int f(x)w(x)q(x)\mathrm{d}x, \quad w(x) = p(x)/q(x) \\ &\approx \frac{1}{S}\sum_{s=1}^{S} f\left(x^{(s)}\right) w\left(x^{(s)}\right), \quad x^{(s)} \sim q(x) \end{aligned} \tag{4}$$

2 Markov chain Monte Carlo

Markov chain Monte Carlo methods can be used to sample from $p(x)$ distributions that are complex and have unknown normalization.

Key Idea: We can construct an init Markov chain, then by getting the detailed balance state of markov chain to get the distribution approximation.

$$p\left(x'\right) = \sum_{x} T\left(x' \leftarrow x\right) p(x) \quad \text{for all } x' \tag{5}$$

$$\widetilde{T}\left(x \leftarrow x'\right) \propto T\left(x' \leftarrow x\right) p(x) = \frac{T\left(x' \leftarrow x\right) p(x)}{\sum_{x} T\left(x' \leftarrow x\right) p(x)} = \frac{T\left(x' \leftarrow x\right) p(x)}{p\left(x'\right)} \tag{6}$$

$$T\left(x' \leftarrow x\right) p(x) = \widetilde{T}\left(x \leftarrow x'\right) p\left(x'\right) \quad \text{for all } x, x' \tag{7}$$

2.1 Metropolis methods

The state transition matrix is related to the acceptance probability. If the sample is accepted, the current state update, otherwise using the previous state.

$$p_a\left(x' \leftarrow x\right) q\left(x' \leftarrow x\right) p(x) = p_a\left(x \leftarrow x'\right) q\left(x \leftarrow x'\right) p\left(x'\right), \quad \text{for all } x, x' \tag{8}$$

2.2 Gibbs sampling

Gibbs sampling resamples each dimension x_i of a multivariate quantity x from their conditional distributions $p(x_i|x_{j\neq i})$

2.3 Two Stage Acceptance

$$p_a = \min\left(1, \frac{q\left(x \leftarrow x'\right)\pi\left(x'\right)}{q\left(x' \leftarrow x\right)\pi(x)}\right) \min\left(1, \frac{L\left(x'\right)}{L(x)}\right) \tag{9}$$

2.4 Variants of MCMC

Auxiliary variable methods instantiate the auxiliary variables in a Markov chain that explores the joint distribution.

3 The Laplace Approximation

$$p(z) = \frac{1}{Z} f(z), \tag{10}$$

Z is untractable. In the Laplace method the goal is to find a Gaussian approximation q(z) which is centred on a mode of the distribution p(z).

$$\ln f(\mathbf{z}) \simeq \ln f\left(\mathbf{z}_0\right) - \frac{1}{2}\left(\mathbf{z}-\mathbf{z}_0\right)^{\mathrm{T}} \mathbf{A}\left(\mathbf{z}-\mathbf{z}_0\right) \tag{11}$$

$$q(\mathbf{z}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{M/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{z}-\mathbf{z}_0\right)^{\mathrm{T}} \mathbf{A}\left(\mathbf{z}-\mathbf{z}_0\right)\right\} = \mathcal{N}\left(\mathbf{z} \mid \mathbf{z}_0, \mathbf{A}^{-1}\right) \tag{12}$$