

These notes summarise selected lecture concepts and are not a substitute for working through the lecture slides, tutorials, and self-study exercises. Feel free to personalise and develop them into your own summary sheet.

Monte Carlo integration — We approximate an expectation via a sample average

$$\mathbb{E}[g(\mathbf{x})] = \int g(\mathbf{x})p(\mathbf{x})d\mathbf{x} \approx \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}_i), \quad \mathbf{x}_i \stackrel{\text{iid}}{\sim} p(\mathbf{x}) \quad (1)$$

In importance sampling, we approximate the expected value via

$$\mathbb{E}[g(\mathbf{x})] = \int g(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x} \approx \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}_i) \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i)}, \quad \mathbf{x}_i \stackrel{\text{iid}}{\sim} q(\mathbf{x}), \quad (2)$$

where $q(\mathbf{x})$ is the importance distribution. To avoid division by small values, $q(\mathbf{x})$ needs to be large when $g(\mathbf{x})p(\mathbf{x})$ is large.

Inverse transform sampling — Given we have a cdf $F_x(\alpha)$ which is invertible, we can generate samples $x^{(i)}$ from a distribution $p_x(x)$ using uniform samples $u^{(i)} \sim \mathcal{U}(0, 1)$:

$$x^{(i)} = F_x^{-1}(u^{(i)}) \quad u^{(i)} \sim \mathcal{U}(0, 1) \quad (3)$$

Rejection sampling — If we sample $\mathbf{x}_i \sim q(\mathbf{x})$ and only keep $\mathbf{x}_i$ with probability $f(\mathbf{x}_i) \in [0, 1]$, the retained samples follow a pdf/pmf proportional to $q(\mathbf{x})f(\mathbf{x})$. The normalising constant equals the acceptance probability $\int q(\mathbf{x})f(\mathbf{x})d\mathbf{x}$. The samples follow $p(\mathbf{x})$ if $f(\mathbf{x})$ is chosen as

$$f(\mathbf{x}) = \frac{1}{M} \frac{p(\mathbf{x})}{q(\mathbf{x})} \quad M = \max_{\mathbf{x}} \frac{p(\mathbf{x})}{q(\mathbf{x})} \quad (4)$$

The acceptance probability then equals $1/M$.

Gibbs sampling — Given a multivariate pdf $p(\mathbf{x})$ and an initial state $\mathbf{x}^{(1)} = (x_1^{(1)}, \dots, x_d^{(1)})$, we obtain multivariate samples $\mathbf{x}^{(k)}$ by sampling from a univariate distribution $p(x_i | \mathbf{x}_{\setminus i})$, and updating individual variables many times.

$$\mathbf{x}^{(2)} = (x_1^{(1)}, \dots, x_{i-1}^{(1)}, x_i^{(2)}, x_{i+1}^{(1)}, \dots, x_d^{(1)}) \quad i \sim \{0, \dots, d\} \quad (5)$$

$\vdots$

$$\mathbf{x}^{(n)} = (x_1^{(n-1)}, \dots, x_{j-1}^{(n-1)}, x_j^{(n)}, x_{j+1}^{(n-1)}, \dots, x_d^{(n-1)}) \quad j \sim \{0, \dots, d\} \quad (6)$$

In the multidimensional space of $\mathbf{x}$, the iterative Gibbs sampling process will appear as a path in orthogonal axes. Like other MCMC methods, Gibbs sampling typically exhibits a warm-up period, where the samples are not representative of the distribution $p(\mathbf{x})$ and the samples are not independent from each other. For multi-modal distributions Gibbs sampling may fail to sample from one or more modes, especially if the modes do not overlap when projected onto any of axes.