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

\[ \mathbb{E}[g(x)] = \int g(x)p(x)dx \approx \frac{1}{n} \sum_{i=1}^{n} g(x_i), \quad x_i \overset{iid}{\sim} p(x) \quad (1) \]

In importance sampling, we approximate the expected value via

\[ \mathbb{E}[g(x)] = \int g(x) \frac{p(x)}{q(x)} q(x)dx \approx \frac{1}{n} \sum_{i=1}^{n} g(x_i) \frac{p(x_i)}{q(x_i)}, \quad x_i \overset{iid}{\sim} q(x), \quad (2) \]

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

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

\[ F_x(\alpha) = \mathbb{P}(x \leq \alpha) = \int_{-\infty}^{\alpha} p_x(y)dy \quad (3) \]

Using the inverse cdf \( F_x^{-1}(y) \), a sample \( x^{(i)} \sim p_x(x) \) can be generated using

\[ x^{(i)} = F_x^{-1}(y^{(i)}) \quad y^{(i)} \sim U(0, 1) \quad (4) \]

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

\[ f(x) = \frac{1}{M} \frac{p(x)}{q(x)} \quad M = \max_{x} \frac{p(x)}{q(x)} \quad (5) \]

The acceptance probability then equals \( 1/M \).

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

\[ x^{(2)} = (x_1^{(1)}, \ldots, x_{i-1}^{(1)}, x_i^{(2)}, x_{i+1}^{(1)}, \ldots, x_d^{(1)}) \quad i \sim \{0, \ldots, d\} \quad (6) \]

\[ \vdots \]

\[ x^{(n)} = (x_1^{(n-1)}, \ldots, x_{j-1}^{(n-1)}, x_j^{(n)}, x_{j+1}^{(n-1)}, \ldots, x_d^{(n-1)}) \quad j \sim \{0, \ldots, d\} \quad (7) \]

In the multidimensional space of \( x \), 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 \( p(x) \) 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.