These notes are intended to give a summary of relevant concepts from the lectures which are helpful to complete the exercises. It is not intended to cover the lectures thoroughly. Learning this content is not a replacement for working through the lecture material and the exercises.

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

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

In importance sampling, we approximate the expected value via

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

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 our distribution $p_x(x)$ using uniform samples $y^{(i)} \sim \mathcal{U}(0,1)$,

$$F_x(\alpha) = \mathbb{P}(x \leq \alpha) = \int_{-\infty}^{\alpha} p_x(y)\mathrm{d}y \tag{3}$$

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

$$x^{(i)} = F_x^{-1}(y^{(i)}) \qquad\qquad y^{(i)} \sim \mathcal{U}(0,1) \tag{4}$$

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

$$f(\mathbf{x}) = \frac{1}{M}\frac{p(\mathbf{x})}{q(\mathbf{x})} \qquad\qquad M = \max_{\mathbf{x}} \frac{p(\mathbf{x})}{q(\mathbf{x})} \tag{5}$$

The acceptance probability then equals $1/M$.

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

$$\mathbf{x}^{(2)} = (x_1^{(1)}, \ldots, x_{i-1}^{(1)}, x_i^{(2)}, x_{i+1}^{(1)}, \ldots, x_d^{(1)}) \qquad\qquad i \sim \{0, \ldots, d\} \tag{6}$$

$$\vdots$$

$$\mathbf{x}^{(n)} = (x_1^{(n-1)}, \ldots, x_{j-1}^{(n-1)}, x_j^{(n)}, x_{j+1}^{(n-1)}, \ldots, x_d^{(n-1)}) \qquad\qquad j \sim \{0, \ldots, d\} \tag{7}$$

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.