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 by a sample average

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

The estimate is unbiased with variance $\mathbb{V}[g(\mathbf{x})]/n$. Always check variability of the estimate.

Importance sampling — We rewrite integrals and expectations as expectations under an importance (proposal) distribution $q(\mathbf{x})$, with $\mathbf{x}_i \overset{\text{iid}}{\sim} q(\mathbf{x})$:

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

q should be large where the integrand is large. If only $\tilde{p} \propto p$ is known (e.g. posteriors), *self-normalised* importance sampling uses $\mathbb{E}_{p(\mathbf{x})}[g(\mathbf{x})] \approx \sum_i W_i\, g(\mathbf{x}_i)$ with $W_i = w_i / \sum_k w_k$ and $w_i = \tilde{p}(\mathbf{x}_i)/\tilde{q}(\mathbf{x}_i)$.

Effective sample size— Assesses reliability of importance sampling estimates as

$$\text{ESS} = (\sum_i w_i)^2 / \sum_i w_i^2 = 1/\sum_i W_i^2 \in [1, n]$$

A small ESS means an unreliable estimate.

Inverse transform sampling — Given the cdf F_x, samples from p_x are $x^{(i)} = F_x^{-1}(u^{(i)})$ with $u^{(i)} \sim \mathcal{U}(0,1)$. Where F_x is steep (high density), the probability mass of u is compressed into a narrow interval of x; where it is flat, it is stretched. For discrete x, this amounts to partitioning $[0,1]$ into intervals of length $p(x)$.

Rejection sampling — Keeping samples $\mathbf{x}_i \sim q(\mathbf{x})$ with probability $f(\mathbf{x}_i) \in [0,1]$ gives samples from a pdf/pmf $\propto q(\mathbf{x})f(\mathbf{x})$. For a target $p(\mathbf{x}) = \tilde{p}(\mathbf{x})/Z$, choose $f(\mathbf{x}) = \tilde{p}(\mathbf{x})/(Mq(\mathbf{x}))$ with $M = \max_{\mathbf{x}} \tilde{p}(\mathbf{x})/q(\mathbf{x})$. The acceptance probability is $\int q(\mathbf{x})f(\mathbf{x})d\mathbf{x} = Z/M$;, it typically decays fast with increasing dimensionality of $\mathbf{x}$.

Ancestral sampling — For directed graphical models, sample each variable given its parents, in topological order; this gives iid samples from the joint.

Gibbs sampling — Repeatedly sample one variable $x_i \sim p(x_i \mid \mathbf{x}_{\setminus i}) = p(x_i \mid \text{MB}(x_i))$, keeping the others fixed (Z cancels). The samples form a Markov chain (MCMC) converging to $p(\mathbf{x})$; they are correlated, and mixing is slow for strongly correlated variables.