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.

Sum rule (marginalisation) — Given a joint distribution $p(\mathbf{x}, \mathbf{y})$, the marginal $p(\mathbf{y})$ is

$$p(\mathbf{y}) = \begin{cases} \sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}) & \text{for discrete r.v.} \\ \int p(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{x} & \text{for continuous r.v.} \end{cases} \tag{1}$$

If $p(\mathbf{x}, \mathbf{y})$ is represented by samples $(\mathbf{x}_i, \mathbf{y}_i)$, $p(\mathbf{y})$ is represented by the $\mathbf{y}_i$.

Product rule and conditioning — A joint distribution can be decomposed into marginal and conditional via

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{2}$$

We can obtain the conditional from the joint by division through the marginal. If $p(\mathbf{x}, \mathbf{y})$ is represented by samples $(\mathbf{x}_i, \mathbf{y}_i)$, $p(\mathbf{y}|\mathbf{x} = \mathbf{x}_o)$ is represented by the $\mathbf{y}_i$ for which $\mathbf{x} = \mathbf{x}_o$.

Bayes rule — From the product rule follows Bayes' rule

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})} \tag{3}$$

Chain rule — Given any ordering $x_1, \ldots, x_d$ of d random variables $\mathbf{x}$, their pdf/pmf can be factorised as

$$p(\mathbf{x}) = \prod_{i=1}^{d} p(x_i|\mathrm{pre}_i) \tag{4}$$

where $\mathrm{pre}_i = \{x_1, \ldots, x_{i-1}\}$, $\mathrm{pre}_1 = \varnothing$ and $p(x_1|\varnothing) = p(x_1)$.

Expectation — Let $\mathbf{x}$ be a random variable with pdf/pmf $p(\mathbf{x})$. Given a function $g(\mathbf{x})$, the expectation $\mathbb{E}\left[g(\mathbf{x})\right]$ is

$$\mathbb{E}\left[g(\mathbf{x})\right] = \begin{cases} \int g(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} & \text{for pdfs} \\ \sum_{\mathbf{x}} g(\mathbf{x})p(\mathbf{x}) & \text{for pmfs} \end{cases} \tag{5}$$

Conditional expectations $\mathbb{E}[g(\mathbf{x})|\mathbf{y}]$ use $p(\mathbf{x}|\mathbf{y})$ instead of $p(\mathbf{x})$.

Chain rule for the expectation — Given any ordering $x_1, \ldots, x_d$ of d random variables $\mathbf{x}$, the expectation over $\mathbf{x}$ can be decomposed as

$$\mathbb{E}_{p(\mathbf{x})} = \mathbb{E}_{p(x_1)}\mathbb{E}_{p(x_2|x_1)} \cdots \mathbb{E}_{p(x_d|x_1,\ldots,x_{d-1})} \tag{6}$$