0 Probability Cheat Sheet

Here I will summarise the results from probability required for this course. There will be a problem sheet partnered with this sheet. If you struggle with these results you should see me so we can discuss how you may familiarise yourself with this material. If this course is optional for you and you are unfamiliar with these results **I strongly suggest you consider another course**.

0.1 Definitions and Basic Rules

Let Ω be the space of all outcomes and $A, B \subseteq \Omega$ (i.e. Ω is all possible outcomes of a fair dice $\Omega = \{1, 2, 3, 4, 5, 6\}$, A and B are things like the odd numbers $\{1, 3, 5\}$). A, B are called 'events', like the event of an odd number being rolled.

If outcomes are all equally likely (i.e. the roll of a fair dice) then we define the probability of set A to be

$$P(A) = \frac{|A|}{|\Omega|}$$

i.e. size of A divided by size of whole space.

More generally P is said to be a probability distribution (or measure) if the following three axioms hold

1. For every event $P(A) \geq 0$
2. $P(\Omega) = 1$
3. If $A_1, A_2, \ldots$ a sequence of pairwise disjoint events then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

0.1.1 Complement

The probability of the complement of a set A^c (i.e. the stuff in Ω but not in A) is given by

$$P(A^c) = 1 - P(A)$$

0.1.2 Sum Rule

The sum rule is defined as

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Typically we will consider the case where A and B are disjoint in which case this reduces to

$$P(A \cup B) = P(A) + P(B)$$

<u>NOTE:</u> The course book (PRML) treats the sum rule only for the case where A and B are disjoint.

0.2 Conditional probability

The probability of A occurring given that we observed $B \neq \emptyset$ is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

0.2.1 Independence

The sets A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

0.2.2 Product Rule

The product rule is defined as

$$P(A \cap B) = P(A|B)P(B)$$

0.2.3 Bayes' Rule

From the product rule one deduces Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

0.3 Random Variables

We now move on to discussing probability distributions, expectations and variances of a random variable. We define X and Y to be random variables (r.v.). A random variable is a function that assigns to every outcome a real value. Its distribution, also called probability mass function (pmf), is defined as

$$p(x) = P(\{X = x\})$$

The right hand side is the probability of the event that X takes the value x, i.e. the total probability of all outcomes where this is true. Once we have the distribution $p(x)$ we can essentially forget about the underlying space Ω as far as X is concerned.

If we generalise the setup appropriately to infinite Ω (the technical details are the subject of what is known as measure theory but won't concern us here), we can have random variables that can take a continuum of values, e.g. any real number. For these one defines a probability density function (pdf) via

$$p(x)\,dx = P(\{x < X < x + dx\})$$

which has to hold in the limit $dx \to 0$. The upshot of this is that formulas for discrete r.v.s translate to formulas for continuous r.v.s just by replacing sums with integrals.

0.3.1 Expectation and variance

We define the expectation (or average) of a random variable X by

$$\mathbb{E}[X] = \begin{cases} \sum_x xp(x) = \sum_x xP(\{X = x\}) & \text{if } X \text{ is discrete} \\ \int \mathrm{d}x\, xp(x) & \text{if } X \text{ continuous} \end{cases}$$

We define the expectation of a function f of a random variable by

$$\mathbb{E}[f(X)] = \begin{cases} \sum_x f(x)p(x) = \sum_x f(x)P(\{X = x\}) & \text{if } X \text{ is discrete} \\ \int \mathrm{d}x\, f(x)p(x) & \text{if } X \text{ continuous} \end{cases}$$

We define the variance of a random variable X to be

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

0.4 Joint distributions

We define the joint distribution of Y and X by

$$p(x, y) = P(\{X = x\} \cap \{Y = y\})$$

for discrete r.v.s. The right hand side is also sometimes written as $P(X = x, Y = y)$. For continuous r.v.s one defines analogously

$$p(x, y)\, dx\, dy = P(\{x < X < x + dx\} \cap \{y < Y < y + dx\})$$

0.4.1 Joint expectation

We define the expectation of a function of X and Y as

$$\mathbb{E}[f(X, Y)] = \begin{cases} \sum_{x,y} f(x, y)p(x, y) & \text{if } X \text{ and } Y \text{ discrete} \\ \int \mathrm{d}x \int \mathrm{d}y\, f(x, y)p(x, y) & \text{if } X \text{ and } Y \text{ continuous} \end{cases}$$

0.4.2 Marginal

For a joint distribution the marginal of X is given by

$$p(x) = \begin{cases} \sum_y p(x, y) & \text{if } Y \text{ discrete} \\ \int \mathrm{d}y\, p(x, y) & \text{if } Y \text{ continuous} \end{cases}$$

0.4.3 Conditional Expectation

The conditional expectation of X given $Y = y$ is defined as

$$\mathbb{E}[X|Y=y] = \begin{cases} \sum_x xp(x|y) & \text{if } X \text{ discrete} \\ \int \mathrm{d}x\, xp(x|y) & \text{if } Y \text{ continuous} \end{cases}$$

0.4.4 Covariance between X and Y

The covariance between X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(x - \mathbb{E}[X])(y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

0.5 Univariate Gaussian (Normal) distribution

The univariate (1D) Gaussian or Normal distribution is given by

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

It has an expected value or mean μ and variance σ^2. We call the inverse variance the precision.

0.6 Multivariate Gaussian distribution

The multivariate (ND) Gaussian or Normal distribution is given by

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

where $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ a positive definite matrix.

It has expected value $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.