Appendix A

Probability Distributions and Bayes Theorem

This appendix introduces the basic principles of probability theory and Bayes Theorem. It is based on [27]. Other recommended texts introducing the reader to the matter of probabilities and random variables are [14, 69].

A.1 Probability distributions

A Probability Density Function (PDF) of a random variable x is denoted $p(x)$. It is defined on the elements of the set $\mathcal{X}$. This function maps an N-dimensional random vector $\mathbf{x}$ into one dimension, $\Re^N \mapsto \Re$. To be a valid PDF the following constraints have to be satisfied

$$p(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X} \tag{A.1}$$

The sum of all probabilities integrates to one if $\mathbf{x}$ is continuous-valued quantity

$$\int_{-\infty}^{\infty} p_x(\mathbf{x}) d\mathbf{x} = 1 \tag{A.2}$$

For the discrete case of $\mathbf{x}$

$$\sum_{x \in \mathcal{X}} p_x(\mathbf{x}) = 1 \tag{A.3}$$

The joint distribution or more precisely *joint probability density function* of two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined to satisfy the following properties

$$p_{xy}(\mathbf{x}, \mathbf{y}) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y} \tag{A.4}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{xy}(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = 1 \tag{A.5}$$

and for the discrete case of $\mathbf{x}$ and $\mathbf{y}$

$$\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{xy}(\mathbf{x},\mathbf{y}) = 1 \tag{A.6}$$

This function maps pairs of $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$ to a real number. $\mathcal{X} \times \mathcal{Y} \mapsto \Re$

Often it is of interest to calculate the distribution of a joint probability density function dependent on only one variable. This is achieved by integrating over the remaining variable, leaving us with a *marginal probability density function* which now depends only on the variable of interest.

$$p_x(\mathbf{x}) = \int_{-\infty}^{\infty} p_{xy}(\mathbf{x},\mathbf{y})d\mathbf{y} \tag{A.7}$$

$$p_y(\mathbf{y}) = \int_{-\infty}^{\infty} p_{xy}(\mathbf{x},\mathbf{y})d\mathbf{x} \tag{A.8}$$

and again for the discrete case of $\mathbf{x}$ and $\mathbf{y}$

$$p_x(\mathbf{x}) = \sum_{y\in\mathcal{Y}} p_{xy}(\mathbf{x},\mathbf{y}) \tag{A.9}$$

$$p_y(\mathbf{y}) = \sum_{x\in\mathcal{X}} p_{xy}(\mathbf{x},\mathbf{y}) \tag{A.10}$$

If, and only if, the joint PDF of two vectors equals the product of their marginals, the two vectors are said to be independent.

$$p(\mathbf{x},\mathbf{y}) = p(\mathbf{x})p(\mathbf{y}) \tag{A.11}$$

The conditional density for a vector $\mathbf{x}$ given a particular value of $\mathbf{y}$ is defined as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} \tag{A.12}$$

A.2 Bayes Theorem

Bayes theorem [11] as given by equation (A.13) is one of the key equations for probabilistic filtering. All probabilistic filtering methods are based on this theorem. It allows us to refine our knowledge about a state of interest x. Using Bayes theorem we can infer the (conditional) *posterior* probability of x given y, y being the observation, if we have a *prior* probability for x available and make an observation which is encoded in the conditional likelihood of y given x.

$$\mathcal{P}(x|y) = \frac{p(y|x)\mathcal{P}(x)}{p(y)} \tag{A.13}$$

Equation (A.14) expresses this in words

$$posterior = \frac{likelihood \times prior}{evidence} \tag{A.14}$$

By observing the state Bayes formula allows us to update a prior probability into a posterior probability given the observation we just made. The *evidence* factor just serves to normalize the posterior probabilities, so that they sum to one.