Multivariate Probabilty and Statistics

Jussi Martin

October 24, 2016

Introduction

This presentation is a breef summary of some of the topics introduced in chapter 4 of the book *Natural Image Statistics* (reference at the last slide).
Since there won't be enough time to cover the whole chapter 4, I decide to go trough the things that are needed for understanding the *Baye's rule*, which I think is one of the core content in this chapter.
I will also introduce *expectation* and *variance* in these slides, since they are needed in the exercises.

Random Variables

Discrete random variable is a variable which can have any value from some given range with some given probabilty. Example: outcome of throwing a dice. The outcome can have any integer value from 1 to 6, each with the probability $\frac{1}{6}$.
Continuos random variable is a variable which can have any value from some given continuous range of values, with some given probabilities for the value been in any given interval. Example: uniform probability distribution on the interval $[0, 1]$. Now the outcome is in an interval $[a, b]$ with the probability $b - a$ for any given values $0 \leq a \leq b \leq 1$.

Random Vectors

Random vector is a n-tuple of random variables. Each of these can be either similar or different types of variables. Formally:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

where $\mathbf{z}$ is the random vector and the random variables z_i are its components.

Probability Density Function (pdf)

Let $P(z \text{ is in [a, b]})$ be the probability that the value of a random variable z is in the interval $[a, b]$, then the probabilty density function $p_z(a)$ can be defined approximately as follows:

$$p_z(a) \approx \frac{P(z \text{ is in } [a, a+\nu])}{\nu}$$

for very small values of ν.
Similarily, if $\mathbf{z}$ is a random vector:

$$p_{\mathbf{z}}(\mathbf{a}) \approx \frac{P(z_i \text{ is in } [a_i, a_i+\nu] \text{ for all } i)}{\nu^n}$$

for very small values of ν. Rigorous definitions are obtained by taking the one-sided limit $\nu \to 0^+$.

Joint and Marginal pdfs

The pdf $p(z_1, z_2, \ldots, z_n)$ of a random vector $\mathbf{z}$ is also called joint pdf, since it depends on all the components z_i.
The marginal pdfs $p_{z_i}(z_i)$ are obtained by integrating over the other variables (that is z_j, with $j \neq i$). Example in two dimensions:

$$p_{z_1}(z_1) = \int p(z_1, z_2) dz_2$$

or more rigorosly: (SOMETHING ELSE HERE!)

$$p_{z_1}(\nu_1) = \int p(\nu_1, \nu_2) d\nu_2$$

Conditional Probabilities

Conditional pdfs are obtained from any joint pdf by fixing the value of one (or several) of its components and multiplying the result with a normalization factor. Example: (conditional pdf of z_2 with z_1 given)

$$p(z_2|z_1 = a) = \frac{p_{\mathbf{z}}(a, z_2)}{\int p_{\mathbf{z}}(a, z_2) dz_2}.$$

or more concisely

$$p(z_2|z_1 = a) = \frac{p_{\mathbf{z}}(a, z_2)}{p_{z_1}(a)} = \text{e.t.c.}$$

FINISH THIS! (EXPLAIN THE NOTATION $p(z_1|z_2)$)

Conditional Probabilities (discrete case)

For discrete case the integral is replaced by sum, that is

$$P_{z_1}(z_1) = \sum_{z_2} P_{\mathbf{z}}(z_1, z_2)$$

FINISH THIS!

Independence

Two random variables z_1 and z_2 are said to be statistical independent if information about the value of one of them does not give any information about the value of the other. Formaly this can be writen as

$$p(z_2|z_1) = p(z_2) \text{ for every } z_1 \text{ and } z_2,$$

since conditional probability assumes that the value of the other variable is fixed (and hence known). Substituting the formula of $p(z_1|z_2)$ we obtain two alternative ways to define independence:

$$\frac{p(z_1, z_2)}{p(z_1)} = p(z_2) \text{ for every } z_1 \text{ and } z_2$$

or

$$p(z_1, z_2) = p(z_1)p(z_2) \text{ for every } z_1 \text{ and } z_2.$$

Baye's Rule

Baye's rule

$$p(\mathbf{s}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{s})p_{\mathbf{s}}(\mathbf{s})}{\int p(\mathbf{z}|\mathbf{s})p_{\mathbf{s}}(\mathbf{s})ds}$$

Or more rigorously WRITE HERE?

Expectation

Expectation $E\{z\}$ of a random variable z is a weighted average of the outcomes which the variable can obtain, with the weights been the individual propabilities (or values of the pdf in the continuous case).

$$\text{Discrete case:} \qquad E\{z\} = \sum_z P_z(z)z$$

$$\text{Continuous case:} \qquad E\{z\} = \int p_z(z)zdz$$

The expectation can be also defined for random vectors component wise:

$$E\{\mathbf{z}\} = \begin{pmatrix} E\{z_1\} \\ E\{z_2\} \\ \vdots \\ E\{z_n\} \end{pmatrix}$$

where $E\{z_i\}$ is defined as above, with z repleced by z_i for all i.

Variance

Variance in some sense tells how close the probability mass is concentrated near the ecpectation value. It is defined as follows:

$$var(z) = E\{z^2\} - (E\{z\})^2.$$

Alternatively it can defined as

$$var(z) = E\{(z - E\{z\})^2\}.$$

WRITE THIS IN A BETTER WAY!

References

Aapo Hyvärinen, Jarmo Hurri, and Patrik O. Hoyer
Natural Image Statistics: A Probabilistic Approach to Early Computational Vision
Springer-Verlag, 2009