Multiple Random Variables

1 Random Vectors

Previously we have only dealt with one random variable. Now suppose we have more random variables. What distribution functions can we then define?

1.1 Joint and marginal distribution functions

Let's suppose we have n random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. We can put them in a so-called **random vector** $\underline{\mathbf{x}} = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n]^T$. The **joint distribution function** (also called the **simultaneous distribution function**) $F_{\underline{\mathbf{x}}}(\mathbf{x})$ is then defined as

$$F_{\underline{\mathbf{x}}}(x_1, x_2, \ldots, x_n) = F_{\underline{\mathbf{x}}}(\mathbf{x}) = P(\underline{x}_1 \leq x_1, \underline{x}_2 \leq x_2, \ldots, \underline{x}_n \leq x_n). \tag{1.1}$$

(You should read the commas "," in the above equation as "and" or, equivalently, as the intersection operator $\cap$.) From this joint distribution function, we can derive the **marginal distribution function** $F_{\underline{x}_i}(x_i)$ for the random variable $\underline{x}_i$. It can be found by inserting ∞ in the joint distribution function for every x_j other than x_i. In an equation this becomes

$$F_{\underline{x}_i}(x_i) = F_{\underline{\mathbf{x}}}(\infty, \infty, \ldots, \infty, x_i, \infty, \ldots, \infty). \tag{1.2}$$

The marginal distribution function can always be derived from the joint distribution function using the above method. The opposite is, however, not always true. It often isn't possible to derive the joint distribution function from the marginal distribution functions.

1.2 Density functions

Just like for random variables, we can also distinguish discrete and continuous random vectors. A random vector is **discrete** if its random variables $\underline{x}_i$ are discrete. Similarly, it is continuous if its random variables are continuous.

For discrete random vectors the **joint (mass) distribution function** $P_{\underline{\mathbf{x}}}(\mathbf{x})$ is given by

$$P_{\underline{\mathbf{x}}}(\mathbf{x}) = P(\underline{x}_1 = x_1, \underline{x}_2 = x_2, \ldots, \underline{x}_n = x_n). \tag{1.3}$$

For continuous random vectors, there is the **joint density function** $f_{\underline{\mathbf{x}}}$. It can be derived from the joint distribution function $F_{\underline{\mathbf{x}}}(\mathbf{x})$ according to

$$f_{\underline{\mathbf{x}}}(x_1, x_2, \ldots, x_n) = f_{\underline{\mathbf{x}}}(\mathbf{x}) = \frac{\partial^n F_{\underline{\mathbf{x}}}(x_1, x_2, \ldots, x_n)}{\partial x_1 \, \partial x_2 \, \ldots \, \partial x_n}. \tag{1.4}$$

1.3 Independent random variables

In the first chapter of this summary, we learned how to check whether a series of events $A_1, \ldots, A_n$ are independent. We can also check whether a series of random variables are independent. This is the case if

$$P(\underline{x}_1 \leq x_1, \underline{x}_2 \leq x_2, \ldots, \underline{x}_n \leq x_n) = P(\underline{x}_1 \leq 1) P(\underline{x}_2 \leq 2) \ldots P(\underline{x}_n \leq n). \tag{1.5}$$

If this is, indeed the case, then we can derive the joint distribution function $F_{\underline{\mathbf{x}}}(\mathbf{x})$ from the marginal distribution functions $F_{\underline{x}_i}(x_i)$. This goes according to

$$F_{\underline{\mathbf{x}}}(\mathbf{x}) = F_{\underline{x}_1}(x_1) F_{\underline{x}_2}(x_2) \ldots F_{\underline{x}_n}(x_n) = \prod_{i=1}^{n} F_{\underline{x}_i}(x_i). \tag{1.6}$$

2 Covariance and Correlation

Sometimes it may look like there is a relation between two random variables. If this is the case, you might want to take a look at the covariance and the correlation of these random variables. We will now take a look at what they are.

2.1 Covariance

Let's suppose we have two random variables $\underline{x}_1$ and $\underline{x}_2$. We also know their joint distribution function $f_{\underline{x}_1,\underline{x}_2}(x_1, x_2)$. The **covariance** of $\underline{x}_1$ and $\underline{x}_2$ is defined as

$$C(\underline{x}_1, \underline{x}_2) = E\left((\underline{x}_1 - \overline{x}_1)(\underline{x}_2 - \overline{x}_2)\right) = \int_{-\infty}^{\infty} (\underline{x}_1 - \overline{x}_1)(\underline{x}_2 - \overline{x}_2) f_{\underline{x}_1,\underline{x}_2}(x_1, x_2) dx_1\, dx_2 = E(\underline{x}_1\underline{x}_2) - \overline{x}_1\overline{x}_2. \tag{2.1}$$

The operator $C(\ldots,\ldots)$ is called the **covariance operator**. Note that $C(\underline{x}_1, \underline{x}_2) = C(\underline{x}_2, \underline{x}_1)$. We also have $C(\underline{x}_1, \underline{x}_1) = D(\underline{x}_1) = \sigma_{x_1}^2$.

If the random variables $\underline{x}_1$ and $\underline{x}_2$ are independent, then it can be shown that $E(\underline{x}_1, \underline{x}_2) = E(\underline{x}_1)E(\underline{x}_2) = \overline{x}_1\overline{x}_2$. It directly follows that $C(\underline{x}_1, \underline{x}_2) = 0$. The opposite, however, isn't always true.

But the covariance operator has more uses. Suppose we have random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. Let's define a new random variable $\underline{z}$ as $\underline{z} = \underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_n$. How can we find the variance of $\underline{z}$? Perhaps we can add up all the variances of $\underline{x}_i$? Well, not exactly, but we are close. We can find σ_z^2 using

$$\sigma_z^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} C(\underline{x}_i, \underline{x}_j) = \sum_{i=1}^{n} \sigma_{x_i}^2 + 2 \sum\sum_{1\leq i<j\leq n} C(\underline{x}_i, \underline{x}_j). \tag{2.2}$$

We can distinguish a special case now. If the random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ are all independent, then $C(\underline{x}_i, \underline{x}_j) = 0$ for every i, j $(i \neq j)$. So then we actually are able to get the variance of $\underline{z}$ by adding up the variances of $\underline{x}_i$.

2.2 The correlation coefficient

Now let's make another definition. The **correlation coefficient** is defined as

$$\rho(\underline{x}_1, \underline{x}_2) = \frac{C(\underline{x}_1, \underline{x}_2)}{\sigma_{x_1}\sigma_{x_2}}. \tag{2.3}$$

This function has some special properties. Its value is always between -1 and 1. If $\rho(\underline{x}_1, \underline{x}_2) \approx \pm 1$, then $\underline{x}_2$ is (approximately) a linear function of $\underline{x}_1$. If, on the other hand, $\rho(\underline{x}_1, \underline{x}_2) = 0$, then we say that $\underline{x}_1$ and $\underline{x}_2$ are **uncorrelated**. This doesn't necessarily mean that they are independent. Two variables can be uncorrelated, but not independent. If two variables are, however, independent, then $C(\underline{x}_1, \underline{x}_2) = 0$, and they are therefore also uncorrelated.

3 Conditional Random Variables

In chapter 1 of this summary, we have seen conditional probability. We can combine this with functions like the cumulative distribution function, the probability density function, and so on. That is the subject of this part.

3.1 Conditional relations

Given an event B, let's define the conditional CDF as

$$F_{\underline{x}}(x|B) = P(\underline{x} \leq x|B) = \frac{P(\underline{x} \leq x, B)}{P(B)}. \tag{3.1}$$

Here the event B can be any event. Also, the comma once more indicates an intersection. The conditional PDF now follows as

$$f_{\underline{x}}(x|B) = \frac{dF_{\underline{x}}(x|B)}{dx}. \tag{3.2}$$

The nice thing is that conditional probability has all the properties of normal probability. So any rule that you've previously seen about probability can also be used now.

Let's see if we can derive some rules for these conditional functions. We can rewrite the **total probability rule** for the conditional CDF and the conditional PDF. Let $B_1, B_2, \ldots, B_n$ be a partition of Ω. We then have

$$F_{\underline{x}}(x) = \sum_{i=1}^{n} F_{\underline{x}}(x|B_i)P(B_i) \qquad \Rightarrow \qquad f_{\underline{x}}(x) = \sum_{i=1}^{n} f_{\underline{x}}(x|B_i)P(B_i). \tag{3.3}$$

From this we can derive an equivalent for **Bayes' rule**, being

$$f_{\underline{x}}(x|A) = \frac{P(A|x)f_{\underline{x}}(x)}{\int_{-\infty}^{\infty} P(A|x)f_{\underline{x}}(x)dx}. \tag{3.4}$$

Here the event A can be any event. The probability $P(A|x)$ in the above equation is short for $P(A|\underline{x} = x)$.

3.2 The conditional probability density function

In the previous paragraph, there always was some event A or B. It would be nice if we can replace that by a random variable as well. We can use the random variable $\underline{y}$ for that. By doing so, we can derive that

$$f_{\underline{x}}(x|y) = \frac{f_{\underline{x},\underline{y}}(x, y)}{f_{\underline{y}}(y)}, \tag{3.5}$$

where $f_{\underline{x},\underline{y}}(x, y)$ is the joint density function of $\underline{x}$ and $\underline{y}$. Note that if $\underline{x}$ and $\underline{y}$ are independent, then $f_{\underline{x},\underline{y}}(x, y) = f_{\underline{x}}(x)f_{\underline{y}}(y)$ and thus $f_{\underline{x}}(x|y) = f_{\underline{x}}(x)$.

We can also rewrite the **total probability rule**. We then get

$$f_{\underline{y}}(y) = \int_{-\infty}^{\infty} f_{\underline{y}}(y|x)f_{\underline{x}}(x)dx. \tag{3.6}$$

Similarly, we can rewrite **Bayes' rule** to

$$f_{\underline{x}}(x|y) = \frac{f_{\underline{y}}(y|x)f_{\underline{x}}(x)}{f_{\underline{y}}(y)} = \frac{f_{\underline{y}}(y|x)f_{\underline{x}}(x)}{\int_{-\infty}^{\infty} f_{\underline{y}}(y|x)f_{\underline{x}}(x)dx}. \tag{3.7}$$

3.3 The conditional mean

The **conditional mean** of $\underline{y}$, given $\underline{x} = x$, can be found using

$$E(\underline{y}|x) = \int_{-\infty}^{\infty} y\, f_{\underline{y}}(y|x)\, dy. \tag{3.8}$$

Note that this mean depends on x, and is therefore a function of x. Now let's look at $E(\underline{y}|\underline{x})$. We know that $\underline{x}$ is a random variable, and $E(\underline{y}|x)$ is a function of x. This implies that $E(\underline{y}|\underline{x})$ is a random variable. We may ask ourselves, what is the mean of this new random variable? In fact, it turns out that

$$E(E(\underline{y}|\underline{x})) = E(\underline{y}) = \overline{y}. \tag{3.9}$$

3.4 n random variables

Suppose we have n random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. We can then also have a conditional PDF, being

$$f(x_n, \ldots, x_{k+1}|x_k, \ldots, x_1) = \frac{f(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)}{f(x_1, \ldots, x_k)}. \tag{3.10}$$

From this, the so-called **chain rule** can be derived, being

$$f(x_1, \ldots, x_n) = f(x_n|x_{n-1}, \ldots, x_1) f(x_{n-1}|x_{n-2}, \ldots, x_1) \ldots f(x_2|x_1) f(x_1). \tag{3.11}$$