Random Variables and Distributions

1 Random Variable Definitions

Suppose we know all the possible outcomes of an experiment, and their probabilities. What can we do with them? Not much, yet. What we need, are some tools. We will now introduce these tools.

1.1 Random variables

It is often convenient to attach a number to each event ω_i. This number is called a **random variable** and is denoted by $\underline{x}(\omega_i)$ or simply $\underline{x}$. You can see the random variable as a number, which can take different values. For example, when throwing a dice we can say that $\underline{x}(\text{head}) = 0$ and $\underline{x}(\text{tail}) = 1$. So $\underline{x}$ is now a number that can be either 0 or 1.

Random variables can generally be split up in two categories: discrete and continuous random variables. A random variable $\underline{x}$ is **discrete** if it takes a finite or countable infinite set of possible values. (With countable finite we mean the degree of infinity. The sets of natural numbers $\mathbb{N}$ and rational numbers $\mathbb{Q}$ are countable finite, while the set of real numbers $\mathbb{R}$ is not.)

Both types of random variables have fundamental differences, so in the coming chapters we will often explicitly mention whether a rule/definition applies to discrete or continuous random variables.

1.2 Probability mass function

Let's look at the probability that $\underline{x} = x$ for some number x. This probability depends on the random variable "function" $\underline{x}(\omega_i)$ and the number x. It is denoted by

$$P_{\underline{x}}(x) = P(\underline{x} = x). \tag{1.1}$$

The function $P_{\underline{x}}(k)$ is called the **probability mass function** (PMF). It, however, only exists for discrete random variables. For continuous random variables $P_{\underline{x}}(k) = 0$ (per definition).

1.3 Cumulative distribution function

Now let's take a look at the probability that $\underline{x} \leq x$ for some x. This is denoted by

$$F_{\underline{x}}(x) = P(\underline{x} \leq x). \tag{1.2}$$

The function $F_{\underline{x}}(x)$ is called the **cumulative distribution function** (CDF) of the random variable $\underline{x}$. The CDF has several properties. Let's name a few.

- The limits of $F_{\underline{x}}(x)$ are given by

$$\lim_{x \to -\infty} F_{\underline{x}}(x) = 0 \qquad \text{and} \qquad \lim_{x \to \infty} F_{\underline{x}}(x) = 1. \tag{1.3}$$

- $F_{\underline{x}}(x)$ is increasing. If $x_1 \leq x_2$, then $F_{\underline{x}}(x_1) \leq F_{\underline{x}}(x_2)$.
- $P(\underline{x} > x) = 1 - F_{\underline{x}}(x)$.
- $P(x_1 < \bar{x} \leq x_2) = F_{\underline{x}}(x_2) - F_{\underline{x}}(x_1)$.

The CDF exists for both discrete and continuous random variables. For discrete random variables, the function $F_{\underline{x}}(x)$ takes the form of a staircase function: its graph consists of a series of horizontal lines. For continuous random variables the function $F_{\underline{x}}(x)$ is continuous.

1.4 Probability density function

For continuous random variables there is a continuous CDF. From it, we can derive the **probability density function** (PDF), which is defined as

$$f_{\underline{x}}(x) = \frac{dF_{\underline{x}}(x)}{dx} \qquad \Leftrightarrow \qquad F_{\underline{x}}(x) = \int_{-\infty}^{x} f_{\underline{x}}(t)dt. \tag{1.4}$$

Since the CDF $F_{\underline{x}}(x)$ is always increasing, we know that $f_{\underline{x}}(x) \geq 0$. The PDF does not exist for discrete random variables.

2 Discrete Distribution types

There are many distribution types. We'll be looking at discrete distributions in this part, while continuous distributions will be examined in the next part. But before we even start examining any distributions, we have to increase our knowledge on combinations. We use the following paragraph for that.

2.1 Permutations and combinations

Suppose we have n elements and want to order them. In how many ways can we do that? The answer to that is

$$n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1. \tag{2.1}$$

Here $n!$ means n **factorial**. But what if we only want to order k items out of a set of n items? The amount of ways is called the amount of **permutations** and is

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1). \tag{2.2}$$

Sometimes the ordering doesn't matter. What if we just want to select k items out of a set of n items? In how many ways can we do that? This result is the amount of **combinations** and is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}. \tag{2.3}$$

2.2 The binomial distribution and related distributions

Now we will examine some types of discrete distributions. The most important parameter for discrete distributions is the probability mass function (PMF) $P_{\underline{x}}(k)$. So we will find it for several distribution types.

Suppose we have an experiment with two outcomes: success and failure. The chance for success is always just p. We do the experiment n times. The random variable $\underline{x}$ denotes the amount of successes. We now have

$$P_{\underline{x}}(k) = P(\underline{x} = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2.4}$$

This distribution is called the **binomial distribution**.

Sometimes we want to know the probability that we need exactly k trials to obtain r successes. In other words, the r^{th} success should occur in the k^{th} trial. The random variable $\underline{x}$ now denotes the amount of trials needed. In this case we have

$$P_{\underline{x}}(k) = P(\underline{x} = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}. \tag{2.5}$$

This distribution is called the **negative binomial distribution**.

We can also ask ourselves: how many trials do we need if we only want one success? This is simply the negative binomial distribution with $r = 1$. We thus have

$$P_{\underline{x}}(k) = P(\underline{x} = k) = p(1-p)^{k-1}. \tag{2.6}$$

This distribution is called the **geometric distribution**.

2.3 Other discrete distributions

Let's discuss some other discrete distributions. A random variable $\underline{x}$ follows a **Poisson distribution** with parameter $\lambda > 0$ if

$$P_{\underline{x}}(k) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{2.7}$$

This distribution is an approximation of the binomial distribution if $np = \lambda$, $p \to 0$ and $n \to \infty$.

A random variable $\underline{x}$ has a **uniform distribution** if

$$P_{\underline{x}}(k) = \frac{1}{n}, \tag{2.8}$$

where n is the amount of possible outcomes of the experiment. In this case every outcome is **equally likely**.

A random variable has a **Bernoulli distribution** (with parameter p) if

$$P_{\underline{x}}(k) = \begin{cases} p & \text{for } k = 1, \\ 1-p & \text{for } k = 0. \end{cases} \tag{2.9}$$

Finally there is the **hypergeometric distribution**, for which

$$P_{\underline{x}}(k) = \frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}}. \tag{2.10}$$

3 Continuous Distribution Types

It's time we switch to continuous distributions. The most important function for continuous distributions is the probability density function (PDF) $f_{\underline{x}}(k)$. We will find it for several distribution types.

3.1 The normal distribution

We start with the most important distribution type there is: the **normal distribution** (also called **Gaussian distribution**). A random variable $\underline{x}$ is a **normal random variable** (denoted by $\underline{x} \sim N(\overline{x}, \sigma_x^2)$) if

$$f_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma_x}\right)^2}. \tag{3.1}$$

Here $\overline{x}$ and σ_x are, respectively, the mean and the standard deviation. (We will discuss them in the next part.) It follows that the cumulative distribution function (CDF) is

$$F_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\overline{x}}{\sigma_x}\right)^2} dt. \tag{3.2}$$

The above integral doesn't have an analytical solution. To get a solution anyway, use is made of the **standard normal distribution**. This is simply the normal distribution with parameters $\overline{x} = 0$ and $\sigma_x = 1$. So,

$$\Phi(z) = P(\underline{z} < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt. \tag{3.3}$$

There are a lot of tables in which you can simply insert z and retrieve $\Phi(z)$. To get back to the variable x, you make use of the transformation

$$z = \frac{x - \overline{x}}{\sigma_x} \qquad \Leftrightarrow \qquad x = \sigma_x z + \overline{x}. \tag{3.4}$$

3.2 Other continuous distributions

There is also a continuous **uniform distribution**. A random variable $\underline{x}$ has a uniform distribution (denoted by $\underline{x} \sim U(a, b)$) on the interval (a, b) if

$$f_{\underline{x}}(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

A random variable has an **exponential distribution** if

$$f_{\underline{x}}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases} \tag{3.6}$$

Finally, a random variable has a **gamma distribution** if

$$f_{\underline{x}}(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0, \end{cases} \tag{3.7}$$

where Γ is the **gamma function**, given by

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x}\, dx. \tag{3.8}$$

4 Important parameters

Certain parameters apply to all distribution types. They say something about the distribution. Let's take a look at what parameters there are.

4.1 The mean

The **mean** is the expected (average) value of a random variable $\underline{x}$. It is denoted by $E(\underline{x}) = \overline{x}$. For discrete distributions we have

$$E(\underline{x}) = \overline{x} = \sum_{i=1}^{n} x_i P_{\underline{x}}(x_i), \tag{4.1}$$

with $x_1, \ldots x_n$ the possible outcomes. For continuous distributions we have

$$E(\underline{x}) = \overline{x} = \int_{-\infty}^{\infty} x f_{\underline{x}}(x)\, dx. \tag{4.2}$$

By the way, $E(\ldots)$ is the mathematical **expectation operator**. It is subject to the rules of linearity, so

$$E(a\underline{x} + b) = aE(\underline{x}) + b, \tag{4.3}$$

$$E(g_1(\underline{x}) + \ldots + g_n(\underline{x})) = E(g_1(\underline{x})) + \ldots + E(g_n(\underline{x})). \tag{4.4}$$

4.2 The variance

The **variance** or **dispersion** of a random variable is denoted by σ_x^2. Here σ_x is the **standard deviation**. If $\underline{x}$ is discrete, then the variance is given by

$$\sigma_x^2 = D(\underline{x}) = E\left((\underline{x} - \overline{x})^2\right) = \sum_{i=1}^{n} (x_i - \overline{x})^2 P_{\underline{x}}(x_i) \tag{4.5}$$

If $\underline{x}$ is continuous, then it is given by

$$\sigma_x^2 = D(\underline{x}) = E\left((\underline{x} - \overline{x})^2\right) = \int_{-\infty}^{\infty} (x - \overline{x})^2 f_{\underline{x}}(x)\, dx. \tag{4.6}$$

Here $D(\ldots)$ is the mathematical **dispersion operator**. It can be shown that σ_x^2 can also be found (for both discrete and continuous random variables) using

$$\sigma_x^2 = E(\underline{x}^2) - \overline{x}^2. \tag{4.7}$$

Note that in general $E(\underline{x}^2) \neq \overline{x}^2$. The value $E(\underline{x}) = \overline{x}$ is called the **first moment**, while $E(\underline{x}^2)$ is called the **second moment**. The variance σ_x^2 is called the **second central moment**.

This is all very nice to know, but what is it good for? Let's take a look at that. The variance σ_x^2 tells something about how far values are away from the mean $\overline{x}$. In fact, **Chebyshev's inequality** states that for every $\epsilon > 0$ we have

$$P(|\underline{x} - \overline{x}| \geq \epsilon) \leq \frac{\sigma_{\underline{x}}^2}{\epsilon^2}. \tag{4.8}$$

4.3 Other moments

After the first and the second moment, there is of course also the **third moment**, being

$$E\left((\underline{x} - \overline{x})^3\right). \tag{4.9}$$

The third moment is a measure of the symmetry around the center (the **skewness**). For symmetrical distributions this third moment is 0.

The **fourth moment** $E\left((\underline{x} - \overline{x})^4\right)$ is a measure of how peaked a distribution is (the **kurtosis**). The kurtosis of the normal distribution is 3. If the kurtosis of a distribution is less than 3 (so the distribution is less peaked than the normal distribution), then the distribution is **platykurtic**. Otherwise it is **leptokurtic**.

4.4 Median and mode

Finally there are the median and the mode. The **median** is the value x for which $F_{\underline{x}}(x) = 1/2$. So half of the possible outcomes has a value lower than x and the other half has values higher than x.

The **mode** is the value x for which (for discrete distributions) $P_{\underline{x}}(x)$ or (for continuous distributions) $f_{\underline{x}}(x)$ is at a maximum. So you can see the mode as the value x which is most likely to occur.