Random Variables and Distributions

Further Resources on Causal Inference

Kevin Li

This set of notes complements chapter 1 of *Uncovering the Causes: from Correlation to Causation*.

Here, I provide a more detailed and rigorous treatment of basic statistical concepts that are useful.

1 Introduction to Random Variables

A random variable (also called a distribution) is some variable with a set of potential outcomes Ω. We call Ω the sample space.

- For example, if we flip a coint, the potential outcomes are either heads or tails, so $\Omega = \{H, T\}$.

However, these outcomes are **potential**, but we do not know what outcome we will actually get.

- Before you flip a coin, you do not know what outcome you will get.

But, every outcome $\omega \in \Omega$ has some probability of being realised.

- The probability of heads is 50%, and the probability of tails is 50%.

2 Types of Random Variables

There are several types of random variables. The two we care about are discrete and continuous random variables.

Discrete random variables have a set of finite, distinct potential outcomes in set Ω.

- For example, rolling a die, you can only get outcomes 1, 2, 3, 4, 5, or 6. You cannot get 3.5.

Continuous random variables have an infinite number of outcomes Ω within a range.

- For example, the random variable *how long will it take for me to get to school tomorrow* is continuous - it could take 5 minutes, 6 minutes, or 5.342 minutes.

3 Probability Density Functions

So we know every potential outcome $\omega \in \Omega$ for a random variable has some probability of being realised.

We can define the probability of some outcome ω actually occuring with a **probability density function** (PDF).

Let us say we have a random variable called Y, which has several potential outcomes. y is one of these potential outcomes. The probability of y becoming the actual outcome of random variable Y is given by the probability density function $F_Y(y)$:

$$F_Y(y) = Pr(Y = y)$$

For a continuous variable, it often doesn't make sense to calculate the probability of one specific outcome. For example, why do I care about the probability that it will take exactly 5.342 minutes to get to school tomorrow?

Instead, for continuous random variables, we care about a range - what is the probability that the outcome is between a and b. This can also be calculated with the probability density function.

$$Pr(Y \in [a, b]) = \int_a^b f_Y(y)\ dy$$

4 Cumulative Density Functions

The **cumulative density function (CDF)** gives the probability of an outcome equal or less than some value y.

For example, a CDF would find the probability of getting to school tomorrow in 3.542 minutes <u>or less</u>.

A CDF is notated $F(y)$, and is defined as:

$$F(y) = Pr(Y \leq y) = \int_{-\infty}^{y} f_Y(y)\ dy$$

5 Expectations

The expectation of a random variable Y, notated $\mathbb{E}Y$ or μ_Y, is a meausre for the centre or average of a random variable.

- As the name implies, if we randomly draw an outcome from a random variable Y, the average outcome that we get is $\mathbb{E}Y$.

For discrete random variables, the expectation is defined as the sum of every possible outcome y multiplied by the probability of y occuring, $f_Y(y)$, as given by the probability density function:

$$\mathbb{E}Y = \mu_Y = \sum_i y_i\ f_Y(y_i)$$

For continuous random variables, the expectation is given by:

$$\mathbb{E}Y = \mu_Y = \int_{-\infty}^{\infty} y\ f_Y(y)\ dy$$

Expectations can be added together:

$$\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$$

They can be multiplied with constants (take a to be a constant):

$$\mathbb{E}(aX) = a \cdot \mathbb{E}X$$

And the expectation of a constant is itself:

$$\mathbb{E}(a) = a$$

6 Variance

Variance is the measure of how spread out a random variable is. Mathematically speaking, it is the average distance between each outcome y and the expectation $\mathbb{E}Y$, squared:

$$\mathbb{V}Y = \sigma_Y^2 = \mathbb{E}[(Y - \mathbb{E}Y)^2]$$

Variance has a very useful property that pops up in proofs. Suppose c and b are constants, and X is a random variable, then:

$$\mathbb{V}(c + bX) = b^2 \cdot \mathbb{V}(X)$$

We can also generalise this to linear algebra. If $\mathbf{u}$ is a n-dimensional vector of random variables, and $\mathbf{c}$ is an m–dimensional vector, and $\mathbf{B}$ is an $m \times n$ matrix with fixed constants, then:

$$\mathbb{V}(\mathbf{c} + \mathbf{B}\mathbf{u}) = \mathbf{B}\mathbb{V}(\mathbf{u})\mathbf{B}^\top$$

We will not prove this here, it is not that important to know the proof.

7 Extra Resources

For an even more in depth treatment with proofs, see the page I have written here.