Probability and Statistics Fundamentals (Session 2)

Contents

Hello

Blabla this section introduces you to bla

Random Variables

Random variables are variables that take on numerical values and are determined by experiments. Mathematically speaking, probability experiments (or trials) are procedures which can be infinitely repeated and which have a well-defined set of possible outcomes. These outcomes are affected by chance. This shall be explained by an example. The simplest probability experiment is certainly the coin toss. In theory, it can be infinitely repeated and it also has a well-defined set of outcomes, namely heads or tails. Now lets assign the value 1 to heads and 0 to tails. So our variable, let's call it X can take on either 1 or 0 but unless we perform an experiment (a coin toss!), the value it takes on remains essentially *random*, i.e. it is determined by chance.

A few notational remarks

We normally denote a random variable by a capital letter such as X or Y. We use small letters x or y when we talk about realisations of X or Y, i.e. observations that we have made by performing an experiment. For example, when X is our variable for the possible outcomes of a coin toss, and we perform a coin toss, x denotes a particular outcome of X, for example, 1, or heads, respectively. Looking at different experiments of the same kind

- Notational Remarks (X or x_i, and what about Y?)
  - X is a random variable, x is a particular outcome.

Discrete random variables

Discrete random variables are variables which can only take a countable number of values. For example, throwing a die represents a random variable. For this example, let D be a random variable with outcomes $j = 1, 2, 3, 4, 5, 6$, then:

$$p_j = P(D = d_j), j = 1, 2, 3, 4, 5, 6$$

with $p_1 + p_2 + ... + p_6 = 1$.
This reads "The Probability of D to take d_j is p_j". Each p_j is between 0 and 1 and all of them add up to 1. Why? Because there is a certain probability for every side of a die that it will be thrown, and if you add those probabilities together, they have to add up to 100%! Obviously, the probability of throwing a 7 with a common 6-side die is 0%. If the die is fair, each p_j is $\frac{1}{6}$

Probability Density Functions and Cumulative Density Functions

Features of Probability Distributions

Probability distributions have many features and can be described in many ways, but for our purposes of statistics and econometrics a focus on three basic features is sufficient:

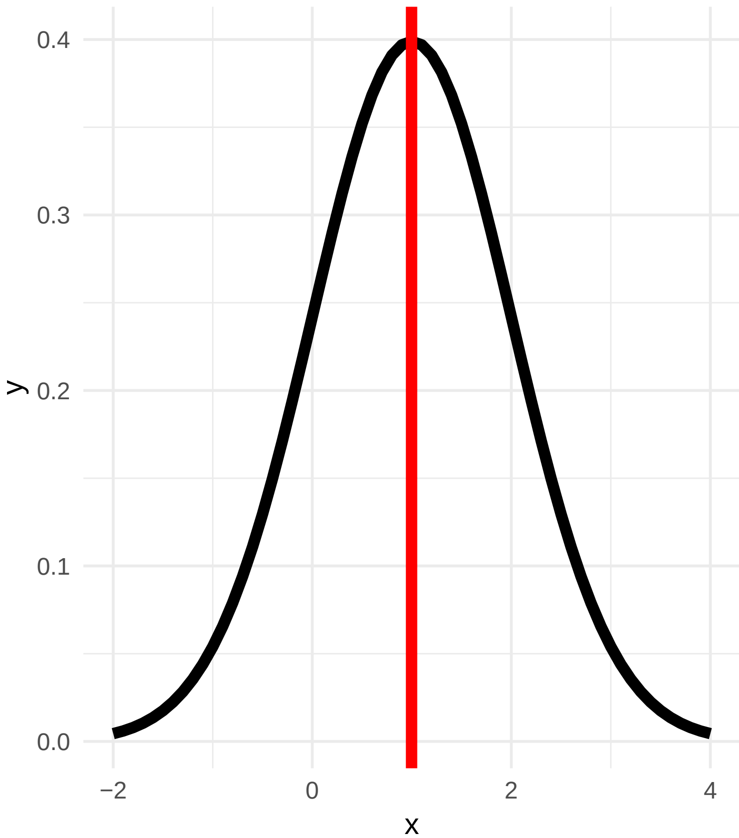
1. Measures of Central Tendency
2. Measures of Variability (or Spread)
3. Measures of Association between two random Variables

These are the building blocks of statistical theories and concepts An understanding on how they work and how they can be manipulated is required for adeqaute understanding on a beginner and also on an advanced level.

A Measure of Central Tendency: The Expected Value

The expected value is a measure of central tendency, meaning that the realisations of a random variable tend to cluster around it.

Speaking in probabilities, for a given probability distribution the expected value denotes the value for which it holds that the likelihood of another value being greater or smaller is the same. This can also be represented graphically: the expected value splits the probability under the pdf in half:



The expected value is usually denoted as μ or μ_X.

Rules for calculating expected values

For any constants a and b and a random variable X, the following rules apply:

- The expected value of a constant is the constant

$$\mathbb{E}[c] = c$$

- The expected value of a linear function of a random variable is the linear function of the expected value of the variable

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

- The expected value of a sum of random variables is the sum of the expected values of these random variables:

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} \mathbb{E}[a_i X_i] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

- For a variable Y that is independent of X it holds that:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[X]$$

For example:

$$\begin{aligned}\mathbb{E}[2X_1 + 4X_2 - 5] &= \mathbb{E}[2X_1] + \mathbb{E}[4X_2] - \mathbb{E}[5] \\ &= 2\mathbb{E}[X_1] + 4\mathbb{E}[X_2] + 5\end{aligned}$$

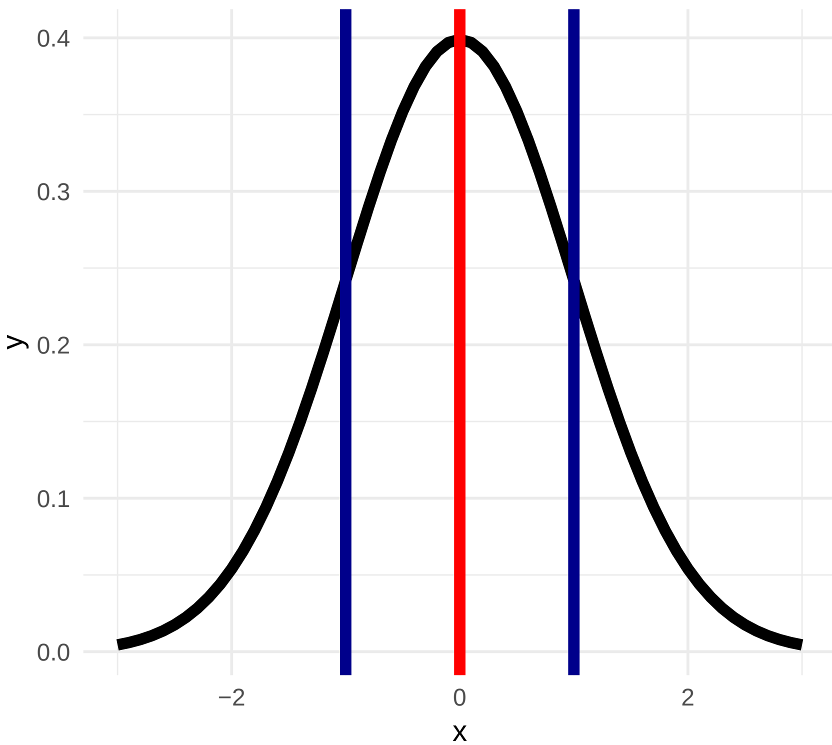
A Measure of Variability: Variance and Standard Deviation

Measures of variability measure the deviation (or distance) of a random variable from its expected value. The most popular is the variance, defined as the expected squared deviation of the variable from its expected value:

$$Var[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

The deviation from the expected value ($X - \mathbb{E}[X]$) is squared in order to prevent perfect cancellation of deviations, since

$$\mathbb{E}\left[(X - \mathbb{E}[X])\right] = \mathbb{E}[X] - \mathbb{E}[X] = 0$$



The variance is usually denoted by σ^2 or σ_X^2.

Rules for Calculating Variances

For any constants a and b and a random variable X, the following rules apply:

- Alternative representation (see Appendix 1 for proof)

$$Var[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

- Variance of a constant is zero:

$$Var[c] = 0$$

- Variance of a linear combination:

$$Var[aX + b] = a^2 Var[X]$$

The Standard Deviation

The standard deviation is yimply defined as the square root of the variance:

$$sd[X] = \sqrt{Var[X]}$$

We also denote it by σ or σ_X.

Standardising Random Variables

Using the properties of the expected value and variance, we can transform any variable to a standardised variable with mean 0 and standard deviation 1:

$$Z := \frac{X - \mu}{\sigma}$$

Proof: write $Z = aX + b$ and define $a := 1/\sigma$ and $b := -(\mu/\sigma)$. Then:

$$\mathbb{E}[Z] = a\mathbb{E}[X] + b = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0 \qquad \text{and}$$
$$Var[Z] = a^2 Var[X] = \frac{\sigma^2}{\sigma^2} = 1$$

This property is useful to compare different random variables and is utilised e.g. in t-tests.

Measures of Association: Covariance and Correlation

The covariance and correlation describe how two random variables vary together, indicating a relationship between them. We begin by defining the covariance:

$$Cov[X,Y] = \mathbb{E}\left[(X - \mathbb{E}[X]) - (Y - \mathbb{E}Y)\right]$$

Notice that the covariance of a variable with itself is the variance of the variable:

$$\begin{aligned} Cov[X,X] &= \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}X)\right] \\ &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= Var[X] \end{aligned}$$

The covariance is usually denoted by σ_{XY}

Rules for Calculating Covariances

- Alternative representations:

$$\begin{aligned} Cov[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] && \text{see Appendix 1.2 for proof} \\ &= \mathbb{E}[(X - \mathbb{E}[X])Y] \\ &= \mathbb{E}[X(Y - \mathbb{E}[Y])] && \text{see Appendix 1.3 for proof} \end{aligned}$$

- Covariance given independence of X and Y:

$$\begin{aligned} Cov[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= 0 && \text{since } \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[X] \end{aligned}$$

- Variance of linear combinations:

$$Cov[a_1X + b_1, a_2Y + b_2] = a_1a_2Cov[X,Y]$$

The Correlation Coefficient

The covariance captures the relationship between two random variables, but is dependent on their unit of measurement. However, it is clear that the strenght of the relationship is unrelated to the units of measurement. To capture this “pure” relationship, we use the correlation coefficent:

$$Corr[X,Y] = \frac{Cov[X,Y]}{sd[X]sd[Y]} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Appendix 1

Proof for the alternative representation of the Variance

$$\begin{aligned}
Var[X] &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[2X\mathbb{E}[X]\right] + \mathbb{E}\left[\mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}$$

Proof for the alternative representation of the covariance

$$\begin{aligned}
Cov[X,Y] &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathbb{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X\mu_Y] - \mathbb{E}[Y\mu_X] + \mathbb{E}[\mu_X\mu_Y] \\
&= \mathbb{E}[XY] - \mu_Y\mathbb{E}[X] - \mu_X\mathbb{E}[Y] + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\mu_Y
\end{aligned}$$

Proof for the second alternative representation of the covariance

$$\begin{aligned}
Cov[X,Y] &= \mathbb{E}[XY] - \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[\mu_X Y] \\
&= \mathbb{E}[XY - \mu_X Y] \\
&= \mathbb{E}[(X - \mu_X)Y]
\end{aligned}$$

The proof is analoguous for $\mathbb{E}[X(Y - \mathbb{E}[Y])]$