Random Variables

1 Random Variables

Definition. A **random variable** X is a function from the set of outcomes to the real numbers; $X : \Omega \to \mathbb{R}$.

This definition is intimidating to students. Instead, think of random variables to be **anything** in this world that could correspond to something we don't know the outcome of, a priori to an observation.

Examples:

- The outcome of a coin toss. (Discrete)
- Whether or not the cat is dead in the box with poison. (Discrete)
- Your possible grade in this course. (Continuous, if we consider real numbers.)

Once we make a valid mapping between one of these outcomes to a real number, we then *formally* have what is called a random variable.

Example. Page 128, Worked example 5.1. Let X be a random variable describing the outcome of two consecutive coin tosses. Then we could assign the following outcomes to the following integers:

- TT: 0
- TH: 1
- HT: 2
- HH: 3

The book does this kind of mapping without being totally explicit about it.

Random variables can either be **discrete** or **continuous**. The biggest difference between a discrete and continuous random variable is, a discrete random variable has countably many outcomes, and a continuous random variable has uncountably many.[1]

We will see below that distinguishing the two is important in understanding some probability fundamentals and notation.

2 Probability Distributions

A **probability distribution** is an assignment (function) between states of the random variable with a value between $[0, 1]$. We'll use

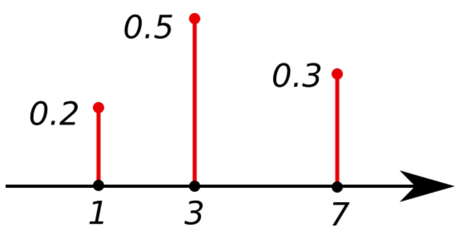
$$P(X)$$

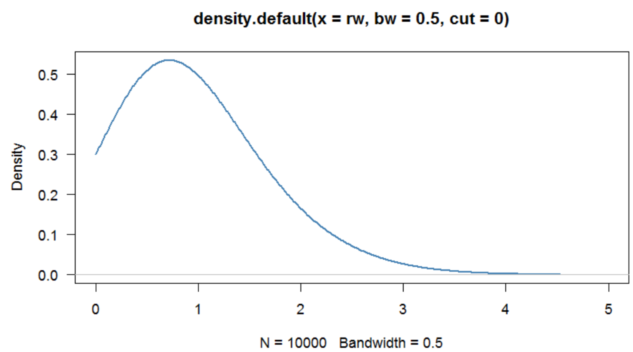
to denote the probability distribution itself, and not values from the distribution.

- When X is finite, the probability distribution is sometimes referred to as the **probability mass function** (pmf). This can be notated several ways, including $P(X = x)$ or $P(x)$. The book uses $P(\{X = x\})$.
- When X is continuous, the probability distribution is sometimes referred to as the **probability density function** (pdf). This can be notated several ways, usually $p(x)$. (*Notice the lowercase* p.)

[1] A set A is countable if it is finite, or there is a bijection between A and the set of natural numbers $\mathbb{N}$.

Observe the analogies mathematicians tried to make when distinguishing these two kinds of random variables. Discrete random variables put their probability on specific numbers (*masses*), while continuous random variables spread their probability along a number line like a spread of butter on bread.





The reason why continuous random variables are usually denoted as $p(x)$ rather than $P(x)$ is because mathematicians are used to continuous functions being lowercase (like $f(x)$, $g(x)$, ...); it's often the case that $p(x)$ has a "closed" function form.

For any valid probability distribution, we have these two properties.

- For any $x \in X$, $P(x) \geq 0$ for discrete random variables.[2] $p(x) \geq 0$ for continuous random variables.
- $\sum_{x \in X} P(X = x) = 1$ for a discrete random variable. $\int_{-\infty}^{\infty} p(x) = 1$ for a continuous random variable.

I'll make a very important distinction here:

- For a discrete random variable X, $P(X = x)$ is a valid probability.
- For a continuous random variable X, $P(X = x)$ is **not** a valid probability. To denote probabilities with continuous random variables, you need $P(a \leq X \leq b) = \int_a^b p(x)dx$. [a]

[a] Well okay. I guess in some notations you could use this as a placeholder for zero. But even *saying* this in the context of continuous random variables sounds weird.

3 Cumulative Distribution Functions

Skipped.

[2] Observe the abuse of notation here with "$x \in X$". X is not a set, but this is how mathematicians express a possible outcome from X.

4 Expected Value

Well, what do we expect X to be when we observe it? Expectation provides this very nice intuition.

The **expected value** $\mathbb{E}[X]$ of a random variable X is given by

- $\mathbb{E}[X] = \sum_{x \in X} xP(X = x)$ when X is discrete
- $\mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)$ when X is continuous

There are two intuitions to this definition.

- You expected value is equal to a *weighted average* of each x. If you interpret $P(X = x)$ to be a weight, and each x to be some kind of score, the expected value will tend to the x that is assigned the greatest weight.
- The expected value is the "balancing point" on a continuous distribution. (If you think of a disk that isn't uniform in density the fulcrum of the disk is where the expected value lies.)

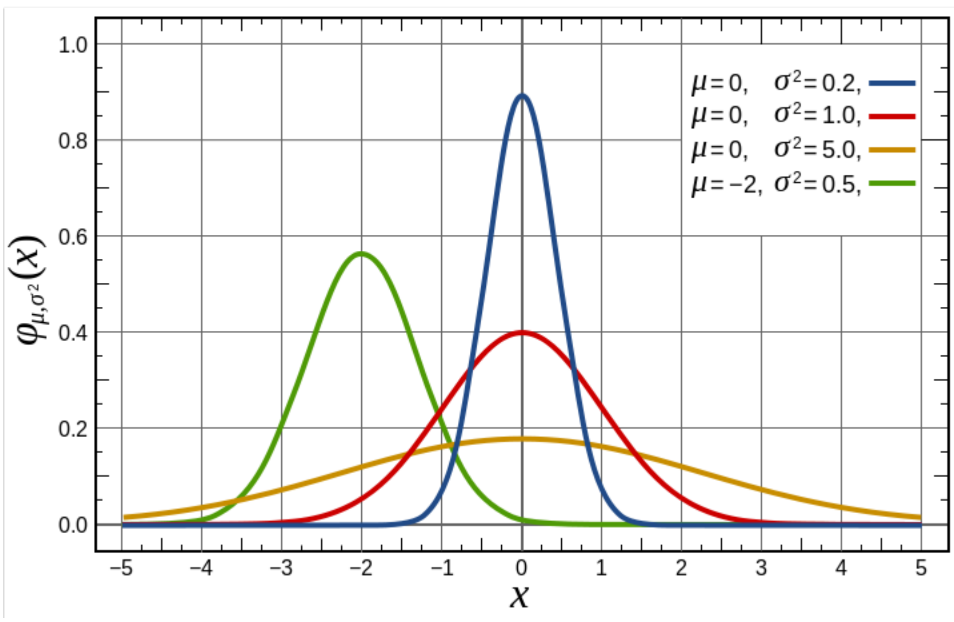


Figure 1: The expected value is denoted as μ in this figure.

If we interpret a dataset of exam scores, how do we expect the next student to do? The best guess is, probably average. This draws a very important point.

For a finite dataset $\{x\}$ of N items, let us interpret $freq(x)/N$, the frequency of an entry over the total number of entries, to represent the probability of x, $P(X = x)$.

Then $\mathbb{E}[X] = \text{mean}(\{x\})$, where $\text{mean}(\{x\}) = \frac{1}{N}\sum_x x$.