A Probability Recap: Overview

This course assumes that you understand the sentence "$\epsilon_1, ..., \epsilon_n$ are independent and identically distributed normal random variables with mean 0." It's OK to understand this sentence intuitively: the ϵ_i don't depend on each other, and on average, each ϵ_i is zero. The fact that they have a normal distribution relates to the idea of a bell curve. Consider the normal distribution shown in Figure 11, and lets call the function $f(x)$. If the ϵ_i had $f(x)$ as its distribution, then the probability that ϵ_i is between a and b is the area under the curve from a to b, i.e., $\int_a^b f(x)dx$. A normal distribution with mean zero and is symmetric about the origin, so that the ϵ_i tend to be "close" to zero, and are just as likely to be positive as negative.

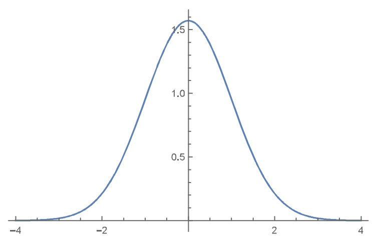

Figure 11: Standard Normal Distribution

If you'd like more rigor, you take A Crash Course in Axiomatic Probability. In the meantime, these notes prevent a quick "recap" that covers the prerequisites more formally. If you want to talk about how to define a probability, talk to me in Tau!.

B Random Variables

We define Ω as the set of all possible outcomes of experiment (usually Ω is called the **Sample Space**). For example, if we flip a coin three times, then

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

A subset $S \subset \Sigma$ is called an **event** and we typically chose subsets that encode meaningful events. For example, $S = \{HHH, HHT, HTH, THH\}$ is the event that we get at least two heads. Finally, a **random variable** X defined on Ω is a map $X : \Omega \to \mathbb{R}$.

I think of random variables as taking possible outcomes and mapping them to a number that communicates meaningful information. Continuing our example, we might have X be the function that counts the number of heads. Then, e.g., $X(\{HHH\}) = 3$.

C Independence

You may have seen the definition that events A, B are **independent** if

$$\mathbb{P}[A \text{ and } B] = \mathbb{P}[A]\mathbb{P}[B].$$

This means that knowing something about A tells you nothing about B.

Exercise: As a tedious but classic exercise, you might want to consider an experiment where you draw a card at random from a normal, 52-card deck. What is Ω in this case? Write the event A that you draw a heart and the even B that you draw a king in terms of Ω. Show that they are independent. What are two dependent events?

Exercise: If you've seen conditional probability before, show that $\mathbb{P}[A|B] = \mathbb{P}[A]$ if and only if A and B are independent random variables. This formalizes the intuition that knowing something about A tells you nothing about B.

Just as we define independence of events, we can define independent random variables. The intuition is as above: knowing something about one random variable tells us nothing about the other.

C.1 Finite Sample Spaces

In the simplest case, the sample space Ω is finite. Random variables X and Y defined on the same sample space Ω are **independent** if

$$\mathbb{P}[A = a \text{ and } B = b] = \mathbb{P}[A = a]\mathbb{P}[B = b]$$

for all possible $a, b \subset \Omega$.

Example C.1. Suppose you flip two coins, so that $\Omega = \{HH, HT, TH, TT\}$. Let A be the number of heads and let B be the number of tails. Then

$$0 = \mathbb{P}[A = 2 \text{ and } B = 2] \neq \mathbb{P}[A = 2]\mathbb{P}[B = 2] = \frac{1}{16}.$$

Hence, A and B are not independent.

Example C.2. Let Ω be the same sample space as in the previous example. Define the odd random variables by

$$A = \begin{cases} 1, & \text{if 1st coin is H} \\ 7, & \text{if 1st coin is T} \end{cases}, \qquad B = \begin{cases} 13, & \text{if 2nd coin is H} \\ 13579, & \text{if 2nd coin is T} \end{cases}.$$

You can show that A and B are independent by showing that

$$\mathbb{P}[A = 1 \text{ and } B = 13] = \mathbb{P}[A = 1]\mathbb{P}[B = 13], \quad \mathbb{P}[A = 1 \text{ and } B = 13579] = \mathbb{P}[A = 1]\mathbb{P}[B = 13579],$$

$$\mathbb{P}[A = 7 \text{ and } B = 13] = \mathbb{P}[A = 7]\mathbb{P}[B = 13], \text{ and } \mathbb{P}[A = 7 \text{ and } B = 13579] = \mathbb{P}[A = 7]\mathbb{P}[B = 13579].$$

This is tedious.

D Continuous Sample Spaces

Often (like in this course) we don't deal with finite, or even infinite, sample spaces. Instead, we deal with continuous sample spaces. For example, if were drawing a random real number between 0 and 1, we would have $\Omega = [0,1]$. The probability of any event is zero, like $\mathbb{P}[\text{ we draw } 0.421\ ] = 0$. Thus, the above definition of independent random variables holds trivially: all of the requisite probabilities are zero!

When Ω is continuous we instead define a **probability density function**, or pdf, and call the random variable continuous. A random variable X has pdf $f(x)$ means that

$$\mathbb{P}[a \le X \le b] = \int_a^b f(z)dz$$

for all $a, b \in \mathbb{R}$. For those who havn't taken calculus: all this means is that the probability that X is between a and b is the area under $f(x)$ between a and b.

Note: for f to be a pdf we must have $f(z) \ge 0$ for all z and $\int_{-\infty}^{\infty} f(z)dz = 1$.

Example D.1. Pick a number in [0, 1] uniformly at random, so that every number is "equally likely." Let X be the random variable that corresponds to what number we draw, and call the pdf $f(x)$. Then

$$f(z) = \begin{cases} 1, & 0 \le z \le 1 \\ 0, & \text{otherwise} \end{cases}.$$

Notice, e.g., that $\mathbb{P}[0 \le X \le \frac{1}{2}] = \int_0^{\frac{1}{2}} 1dz = \frac{1}{2}$.

Example D.2 (BIG Example)**.** X is **normally distributed** with mean 0 and variance σ^2 means that X has pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

If $\sigma = 1$, then X has a **standard normal distribution**, which is shown in Figure 11.

D.1 Independence, Take 3

When random variables are continuous, independence has the same intuition: knowing something about one random variable tells you nothing about the other one. Say that X, Y are random variables defined on the same continuous sample space Ω. Then X, Y are defined to be independent if

$$\mathbb{P}[-\infty \le X \le a \text{ and } -\infty \le Y \le b] = \mathbb{P}[-\infty \le X \le a]\mathbb{P}[-\infty \le Y \le b],$$

for all $a, b \in \mathbb{R}$.

E Expectation and Variance

E.1 Finite Sample Spaces

Let X be a random variable defined on finite sample space Ω. The **expectation** of X is

$$\mathbb{E}[X] := \sum_{s \in \Omega} \mathbb{P}[s] X(s),$$

remembering that X is a function on Ω. You might have seen this written $\sum_k \mathbb{P}[X = k]k$, where the sum ranges over all possible vales of the random variable. You can think of the expectation as the mean response.

Example E.1. Say that we flip a coin twice and let X be a random variable encoding the number of heads. We have that:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{s \in \Omega} \mathbb{P}[s] X(s) \\
&= \mathbb{P}[\{HH\}]X(\{HH\}] + \mathbb{P}[\{HT\}]X(\{HT\}] + \mathbb{P}[\{TH\}]X(\{TH\}] + \mathbb{P}[\{TT\}]X(\{TT\}] \\
&= \frac{1}{4}2 + \frac{1}{4}1 + \frac{1}{4}1 + \frac{1}{4}0 = 1.
\end{aligned}$$

Using the second formula, we notice that X can take on three values: 0, 1, and 2. So:

$$\mathbb{E}[X] = \sum_k \mathbb{P}[X = k]k = \mathbb{P}[0 \text{ heads }]0 + \mathbb{P}[1 \text{ heads}]1 + \mathbb{P}[2 \text{ heads}] = 0 + \frac{1}{2} + \frac{1}{2} = 1/$$

The **variance** measures how far the random variable tends to deviate from its mean: $Var(X) := \mathbb{E}[(X - \mu)^2]$, where $\mu = \mathbb{E}[X]$.

Exercise: Show that, in the above example, $Var(X) = \frac{1}{2}$.

E.2 Continuous Case

The ideas are basically the same, but we replace sums with integrals. Let X be a continuous random variable. Then

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} f(z) z dz,$$

and

$$Var(X) = \int_{-\infty}^{\infty} f(z)(z - \mu)^2 dz,$$

where $\mu = \mathbb{E}[X]$.