Information Theory

Francesco Mussin

2025-02-28

Table of contents

Preface

Hello! Welcome to the notes on Information Theory. These are based on the course taught at ETH by professor Lapidoth but they may include topics outside that course in the future.

1 Introduction

Mathematically, what is information?

2 Quantifying Information

Mathematically, what is information? Well, if you imagine mathematics as a "divine" computer that figures out everything that follows from logic, the only thing that it cannot predict is randomness; if you werre to tell someone a mathematical result, like $3 \cdot 3 = 9$ or $\dim \mathbb{R}^n = n$, you're technically not conveying any new information, since it is something that, on paper, your receiver could figure out. However, if you tell someone the outcome of a random variable, then you're conveying some "information" that is unattainable in any other way.

> Blockquote Mathematically, information is the outcome of a random variable

However, not all outcomes are the same: if I knew that $X \sim \delta_0$ had take outcome $X = 0$, I wouldn't be surprised at all, while if $X \sim \text{Pois}(1)$, we should be surprised when $X = 100.\backslash$ In other words, we need to measure the outcomes of random variables based on how much "surprise" they give use. Here's another way of seeing it. Remember that when you're in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an event $A \in \mathcal{F}$ happens, then you have to adjust the probability space to $(A, \mathcal{F}|_A, \mathbb{P}(\,\cdot\,|A))$. A simple way to measure how much this space changes from the original is to compute $\mathbb{P}(A)$. The bigger this is, the less we have to change our notion of probability, because it means that most of $\mathbb{P}$'s mass sits on A, but if $\mathbb{P}(A)$ is small, then it means that you have to cut off a lot of mass from $\mathbb{P}$, potentially making $\mathbb{P}(\,\cdot\,|A)$ behave a lot differently than $\mathbb{P}(\,\cdot\,)$ on most sets.

Another way of seeing it is, is that when $X \sim \text{Pois}(1)$ assumes value $X = 100$, you have to change your probability space from $(\Omega, \mathcal{F}, \mathbb{P})$ to the subspace $(\{X = 100\}, \mathcal{F}|_{\{X=100\}}, \mathbb{P}(\,\cdot\,|X = 100))$, you

3 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```