Information Theory

Manan Shah
manan.shah.777@gmail.com
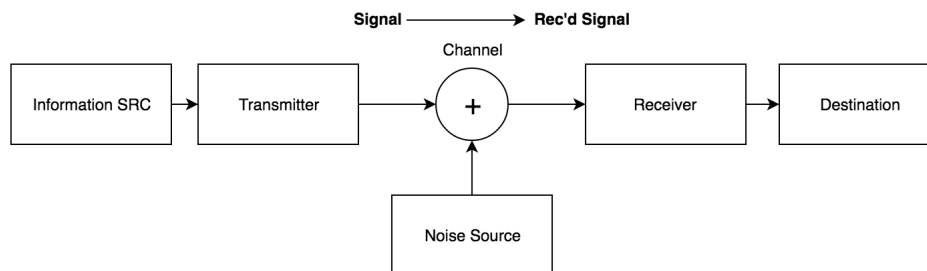The Harker School

August 24, 2016

This document contains lecture notes from Harker's Advanced Topics in Mathematics class in Information Theory, Parts I and II. These notes were taken using TeXShop and LaTeX2ϵ and will be updated for each class.

Contents

1 August 22, 2016

1.1 Class Overview

Signal ⟶ Rec'd Signal

Channel

| Information SRC | Transmitter | + | Receiver | Destination |

Noise Source

Information may be conveyed in multiple ways; for example, by a person talking, a video, or pictures. Our goal is to model this kind of information as a system with one output and no input. The information is therefore a box that emits output probabilistically, and we will therefore be spending 1-2 weeks on probability, specifically regarding distribution functions and expectation. This is illustrated in the "information source" box—the job of the information source is to **provide the signal**.

We will next define the compression problem; having obtained numbers as a stream, the problem is to remove the redundancy in a stream. In particular, we will discuss both lossless and lossy compression. One of the most critical ideas in this course is that of entropy, and we will define the concept (in general, a notion of how much information is originating from the source). We will also discuss the sensibility of such a definition and potentially expand upon it. An overarching concept here will be our definition of the performance of a compression algorithm via the source compression theorem and the boundedness of compression algorithms. This is illustrated in the "transmitter" box—the job of the transmitter is to **remove all redundancy**.

We will subsequently discuss the channel circle, which we will model with conditional probability distribution functions (PDFs). The output of this box, the receiver signal, is input to the receiver box, which has the job of **reconstruction of the signal and the noise**; this is known as the coding problem. Much like the compression problem of the source, the channel has the capacity problem (how much information can be transmitted?). And similar to the compression theorem, we will discuss the channel coding theorem, which will help us determine which parts of a code will require more redundancy for transmission.

We will finally discuss Shannon's theorem, in which he states that a communication with no loss and arbitrary cost (due to a constraint on the amount of bandwith, power, etc.) is possible if the entropy of the source H is less than the capacity of the channel C. In our course, we will first discuss the pipeline for lossless compression and then move to lossy compression.

2 August 24, 2016

Knowing what a random variable is and what it means is important for the source to transmitter relationship.

2.1 Random Variables

Definition 1. A random variable is a real-valued function of the outcome of an experiment. By definition, it is not deterministic.

Notation 2. We will represent random variables with capital letters (for example, X, Y, Z).

Example 3 (What are Random Variables?). (a) A coin is tossed ten times. Determine the number of heads obtained from the coin toss. The number of heads is the random variable in this case. (b) Given two rolls of die, identify their sum (a random variable) and the second roll raised to the fifth power (another random variable). (c) The time to transmit a message, the delay with which a message is received, and the number of errors in a message are both themselves random variables.

Two types of random variables exist; discrete variables and continuous variables.

Definition 4 (Discrete RV). A discrete random variable is one that has a finite set of outcomes (or is countably infinite), and has a probability associated with each outcome.

Definition 5 (Continuous RV). A continuous random variable is one that has a non-finite, uncountable set of outcomes.

Example 6. Define the function

$$sgn(a) = \begin{cases} 1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}$$

The variable a is a continuous random variable, while $sgn(a)$ is a discrete function (with outputs $\in \{1, 0, -1\}$)

2.1.1 Probability Mass Function (pmf)

Probability mass functions are functions of a discrete random variable, describing all instances of a random variable occurring and its associated probabilities. Define X as a random variable, and write $P(X = x)$ as $p_X(x)$. Then $\sum_x p_X(x) = 1$. Consider two tosses of a coin, letting X equal the number of heads obtained. We therefore have

$$p_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \end{cases}$$

note that $p_X(x)$ is clearly a discrete function that takes on values $x \in \{0, 1, 2\}$ with respective output probabilities.

2.1.2 Probability Densify Function (pdf)

Probability density functions are functions of a continuous random variable. Define X as a random variable. In this case, the expression $P(X = x)$ is meaningless (in fact, it is defined as 0). The probability of a range of values from $x = a$ to $x = b$ is more relevant; in particular, $\int_{-\infty}^{\infty} p_X(x) = 1$.

Example 7. Edgar's driving time to school is between 15 and 20 minutes if it is sunny and between 20 and 25 minutes if it is a rainy day. Assume a day is sunny with 2/3 probability and rainy with 1/3 probability. Construct the respective PDF.

The PDF construction is quite simple; remember that $\int_{-\infty}^{\infty} p_X(x) = 1$.

$$p_X(x) = \begin{cases} 2/15 & 15 \leq x \leq 20 \\ 1/15 & 20 \leq x \leq 25 \\ 0 & x \geq 25 \end{cases}$$

2.1.3 Types of Random Variables

We will define these distributions in terms of a coin toss.

Bernoulli (Discrete). Toss a coin, and it define the probability of heads as p. The respective PMF is

$$p_X(x) = \begin{cases} p & x = 1 \\ (1-p) & x = 0 \end{cases}$$

Binomial (Discrete). Toss a coin N times, and define x as the number of heads obtained in N tosses. Let $N = 10$, and define the probability of heads as p. Then the respective PMF is

$$p_X(x) = \begin{cases} (1-p)^{10} & x = 0 \\ \binom{10}{2}p^2(1-p)^8 & x = 2 \\ \ldots & \text{other values of } x \in [0, 10] \cup \mathbb{N} \end{cases}$$

Geometric (Discrete). Repeatedly toss a coin until the first "success." This means that we may theoretically have countably infinite $P(X = 1), P(X = 2), P(X = 3), \ldots$ output values. Define success to be a head, and the probability of heads as p. Then, we have

$$p_X(x) = \begin{cases} p & x = 1 \\ (1-p)p & x = 2 \\ (1-p)^2 p & x = 3 \\ \ldots & \text{other values of } x \end{cases}$$

Poisson (Discrete, $\sim$ Binomial). Instead of tossing a biased coin and counting the number of heads (a binomial distribution), count the number of times one must replace a biased lightbulb in a time t (with the lightbulb being on or off). This makes the poisson distribution a discrete one. Our distribution function is

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where λ is an arbitrary constant. The number of decay events in some radioactive element, for example, may be modeled using a Poisson distribution. Anything that involves units of time and something happening within a period of time t may include a Poisson distribution.

Exponential (Continuous, $\sim$ Geometric). Given a biased lightbulb, identify the amount of time until the bulb burns out. Examples of the use of this function include rates (as opposed to numbers defined with the Poisson distribution). The PDF is defined as

$$p_X(x) = \lambda e^{-\lambda x}$$

where λ is an arbitrary constant.