Information Sciences

Louis Merlin

February 25, 2016

Contents

Chapter 1

Random variables, sources and entropy

1.1 Random variables

Single random variable A **random variable** X is a function $X : \Omega \to \mathbb{R}$. In this course, the **sample space** Ω is finite.

- all random variables are discrete
- the image $X(\Omega)$ is a finite subset $A \subset \mathbb{R}$
- we are able to assign a probability to every subset (event) of Ω

To each $x \in A$, we assign a probability $p(x)$.
The function $p : A \to [0, 1]$ is called the **probability mass function** (or **probability distribution**) of (the random variable) X.

Alphabet The **alphabet** of an event is the set of all of its outcomes.

Law of total probability Given the discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$:

$$\forall x \in \mathcal{X}, \quad p_X(x) = \sum_{y \in \mathcal{Y}} P_{X,Y}(x, y)$$

p_X (and p_Y) is called **marginal distribution**.

Definition Two random variables $S_1 \in \mathcal{A}_1$, $S_2 \in \mathcal{A}_2$ are **independent** if for all $s_1 \in \mathcal{A}_1$ and all $s_2 \in \mathcal{A}_2$,

$$p_{S_1,S_2}(s_1, s_2) = p_{S_1}(s_1) p_{S_2}(s_2)$$

Definition The **conditional probability** of event $\mathcal{S}$ conditioned on event $\mathcal{F}$ is

$$P(\mathcal{S}|\mathcal{F}) \stackrel{\text{def}}{=} \frac{P(\mathcal{F} \cap \mathcal{S})}{P(\mathcal{F})}$$

1.2 Source Model

A source is modeled as a random vector, like $(S_1, \ldots), S_n)$. By definition, an **information source** outputs symbols/letters that cannot be predicted with certitude.

1.3 Source Coding: The Idea, As a start

Source coding is about efficient representation using codewords from a given alphabet (e.g. binary). We assign a codeword to each valid sequence. The length of the codeword is inversely proportional to the probability of the corresponding sequence.

1.4 Entropy

Definition Let $S \in \mathcal{A}$ be a discrete random variable of probability distribution p_S. The entropy of S is

$$H(S) := -\sum_{s \in \mathcal{A}} p_S(s) log_2 p_S(s)$$

A few facts

- The unit of $H(S)$ is the *bit*.
- The entropy of S depends only on p_S.
- By convention, if $p_S(s) = 0$ for some $s \in \mathcal{A}$ then
$$p_S(s) log_2 \, p_S(s) = 0$$
- Equivalent form:
$$H(S) = \sum_{s \in \mathcal{A}} = p_S(s) log \frac{1}{p_S(s)}$$

Theorem (Entropy Bounds) Let S be a discrete random variable taking values in $\mathcal{A}$, and let $D \geq 2$ be a positive integer. Then

$$0 \leq H_D(S) \leq log_D |\mathcal{A}|$$

where $|\mathcal{A}|$ stands for the cardinality of $\mathcal{A}$ and

- the first $\leq$ holds with equality iff there exists an $s \in \mathcal{A}$ for which $p_S(s) = 1$
- the second $\leq$ holds with equality iff S is uniformly distributed over $\mathcal{A}$

Theorem Let $S_1, \ldots, S_n$ be discrete random variables. Then

$$H(S_1, S_2, \ldots, S_n) \leq H(S_1) + H(S_2) + \ldots + H(S_n)$$

with equality iff $S_1, \ldots, S_n$ are independent.