

NOTES ON ENTROPY

ARTHUR RYMAN

ABSTRACT. The concept of entropy crops up in many seemingly unrelated areas, ranging from physics to information theory to machine learning. This article summarizes the mathematical foundations of entropy, providing precise definitions and proofs.

1. INTRODUCTION

Carnot discovered the concept of entropy while analyzing the maximum theoretical efficiency of heat engines. Boltzman gave entropy a statistical definition in terms of the number of microscopic states available to a macroscopic state. Shannon showed how entropy was relevant to the rate at which information could be transmitted over a communications channel. Bekenstein proposed that black holes have entropy and Hawking defined their temperature and predicted that they would radiate. And today, machine learning algorithms regularly use the concepts of entropy and information gain to analyze large datasets.

This is certainly a fascinating and important circle of ideas! Unfortunately, it is difficult to find clear explanations of these concepts. Typical textbooks on thermodynamics do not lay out the material using precise mathematical language. My goal in this article is to collect together all the related definitions and give them a precise mathematical treatment.

The target audience for this article is me since I am personally dissatisfied with my grasp of the subject. However, if I manage to explain this material clearly and simply enough, then it may be of interest to the machine learning community.

2. MICROSCOPIC AND MACROSCOPIC STATES

When Carnot discovered entropy, the atomic theory of matter was not widely accepted as fact by the scientific community. Carnot analyzed heat engines using their macroscopic properties, such as temperature, volume, and pressure. When the atomic theory was being developed, physicists sought to derive the macroscopic properties of matter from the statistical properties of their microscopic constituents. This approach was called *statistical mechanics* or, more generally, *statistical physics*. In fact, the success of statistical mechanics in explaining thermodynamics was viewed as strong supporting evidence for the validity of the atomic theory of matter.

Date: February 26, 2017.

2.1. Boltzman Entropy. Although the thermodynamic definition of entropy came first, I am going to start with the statistical definition since it explains the thermodynamic definition and it applies more directly to machine learning.

The concept of entropy arises when we forget information about the states of a system. Consider a system and let A be the set of states that the system can be in. For example, consider a classical system that consists of N identical point particles each mass m moving in a box of fixed volume that has perfectly reflecting walls. The state of this system is determined by the position and momentum of each particle. We refer to A as the set of *microscopic states* of the system or, more briefly, its *microstates*.

Suppose further that B is a related set of more coarse-grained states referred to as the *macroscopic states* or *macrostates*. Every macrostate corresponds to one or more microstates. Continuing with the example of particles in a box, the macrostate might be defined simply by the system's temperature.

Let the correspondence between the microstates and macrostates be given by a *forgetful* mapping f of A onto B :

$$(1) \quad f : A \longrightarrow B$$

For particles in a box, the temperature is their average kinetic energy

The Boltzman definition of entropy applies directly to the case where for each macrostate b there is a finite number $\Omega_f(b)$ of microstates that map to b under f :

$$(2) \quad \Omega_f : B \longrightarrow \mathbb{N}_1$$

where

$$(3) \quad \Omega_f(b) = \#\{a : A \mid f(a) = b\}$$

In this case the entropy $S_f(b)$ of the macrostate b is defined to be:

$$(4) \quad S_f : B \longrightarrow \mathbb{R}$$

where

$$(5) \quad S_f(b) = k \ln \Omega_f(b)$$

In the system of particles in a box there are an infinite number of microstates for each macrostates so we'll need to adapt Boltzman's definition of entropy a little. Before doing that, let's examine the situation for finite system.

2.2. The Entropy of Finite Systems. Consider a particle that can exist in one of some finite number d of states. Let Φ denote the set these single-particle states:

$$(6) \quad \#\Phi = d$$

Let ϕ_1, \dots, ϕ_d denote the d states in Φ :

$$(7) \quad \Phi = \{\phi_1, \dots, \phi_d\}$$

Now consider a system of N of these particles. Each microstate of this system is given by an ordered N -tuple of single-particle states:

$$(8) \quad A = \Phi^N$$

Recall the definition of the Kronecker δ -function:

$$(9) \quad \delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

Given an N -particle microstate $a = (a_1, \dots, a_N)$ and a single-particle state $\phi \in \Phi$ let $n(a, \phi)$ be the number of particles in a that are in the single-particle state ϕ :

$$(10) \quad n : A \times \Phi \longrightarrow \mathbb{N}$$

where

$$(11) \quad n(a, \phi) = \sum_{i=1}^N \delta(a_i, \phi)$$

The number of particles in a single-particle state is referred to as the *occupancy number* of that state. Note that for any multi-particle state a we have:

$$(12) \quad \sum_{j=1}^d n(a, \phi_j) = N$$

Define the macrostates of the system to be the set of all possible N -particle occupancy number assignments b on Φ :

$$(13) \quad B = \{ b : \Phi \longrightarrow \mathbb{N} \mid \sum_{j=1}^d b(\phi_j) = N \}$$

The forgetful mapping $f : A \longrightarrow B$ sends an N -particle state to its occupancy numbers on Φ :

$$(14) \quad f(a)(\phi) = n(a, \phi)$$

TO DO: Compute the entropy of a state b in B . Use Stirling's formula to approximate it. Show how Shannon's entropy emerges.