

# Thermodynamics of Benford's First Digit Law

---

Don S. Lemons  
Santa Fe, New Mexico USA

---

Professor of Physics, Emeritus  
Bethel College  
North Newton, Kansas USA

1

Because I don't often give these kind of talks I am going to read the script that goes along with my slides. This will be good both for me and for you since I will make all my points and I will finish in plenty of time for questions and comments.

The 15 slides of my talk concerns makes only one point namely Benford's law emerges, in at least many cases, from the unbiased partitioning of a fixed quantity.

# Benford Quotations

- ❖ “I answer the question ‘Why are there more small things in the world than large things?’ in terms of a probabilistic model of partitioning a conserved quantity.” (Lemons, American Journal of Physics, 1986)
- ❖ “This derivation provides solid evidence for Lemons’ original argument” (Iafrate, Miller, and Strauch, Physical Review E, 2015)

argument

2

My interest in Benford’s law began about 50 years ago when as an undergraduate I read R. A. Raimi’s article, “The Peculiar Distribution of First Digits,” in Scientific American. Then in 1986 I made my own contribution to first digit scholarship in the form of a one-page paper in which I asked the question “Why are there more small things in the world than large things?” and attempted to answer it “in terms of a probabilistic model of partitioning a conserved quantity.”

A few years ago Joseph Iafrate, Steven Miller, and Frederich Strauch from Williams College in the United States constructed and numerically demonstrated an elegant statistical model based on the idea of the unbiased random partitioning of a conserved quantity — according to which (quoting) “This derivation provides solid evidence for Lemons’ original argument.” — In fact, I think my “original argument” was more suggestive than convincing.

It was the elegant Iafrate model and the urging of Alex Kossovsky that got me thinking about Benford’s law again after so many years.

Now I have an explicitly thermodynamic derivation of Benford’s first digit law. (I am, after all, a physicist.) In this model the thermal equilibrium among the subsystems into which a conserved quantity is partitioned produces the underlying distribution, the so-called Benford distribution, in which the number of things is inversely proportional to their size. The first digit law emerges from this distribution.

# What is Thermodynamics?

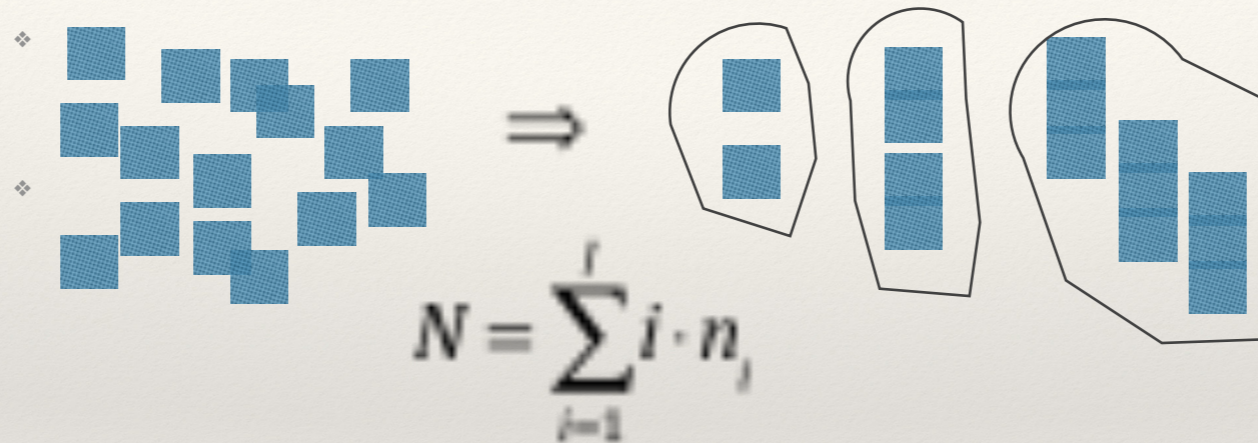
- ❖ Thermodynamics requires
- ❖ Additive Constraint or Constant
- ❖ Entropy
- ❖ Temperature

3

First, let's ask the question, "What is thermodynamics?" I claim that thermodynamics has at least three features . . . an additive constraint or constant (typically, but not necessarily, energy), from which can be determined an entropy, and from this entropy a temperature.

The application of thermodynamics I will present today is relatively simple because most applications feature more than one additive constraint or conserved quantity.

# Conserved Quantity



- ❖  $n_i$  is the number of “atoms” of size  $i$ , that is, the number of people, masses, volumes, surface areas . . . or energy in quantities of definite size.

4

Imagine that there are  $N$  “units” or “atoms” of a quantity whose total number is fixed — as illustrated on the left. These atoms are separated into partitions in which they are joined to one another in different combinations — as we see on the right. We might call these combinations “molecules.” The different-sized “molecules,” might be called singlets, doublets, triplets, . . . or, if you prefer chemical jargon, monomers, dimers, trimers, etc.

As indicated in the equation, each of the  $N$  atoms appears in one of  $l$  separate partitions of different sizes  $i$ . As we shall see, the numbers  $n_{sub\ i}$  in each partition of a different size are variable but their total quantity  $N$  is fixed.

The total number of atoms is the sum of the number of atoms in each partition. This summation plays very much the same role here as the conservation of energy plays in physical applications of thermodynamics. When  $N$  and  $l$  are large there are many different way the  $n_{sub\ i}$  may be chosen in accordance with this sum.

As we shall also see, it will be useful to think of the  $l$  partitions of molecular sizes as defining a set of orthogonal axes. In this view, the additive constraint defines a surface in an  $l$  dimensional space — a surface that contains the points, defined by the different sets of the numbers  $n_{sub\ i}$  that observe the constraint.

## Entropy $S_{N,J}$ and Multiplicity $\Omega_{N,J}$

$$S_{N,J} = \ln \Omega_{N,J}$$

- ❖ Multiplicity  $\Omega_{N,J}$  is the number of ways  $N$  “atoms” of a conserved quantity can be distributed into  $I$  categories of size  $i=1,2,\dots,I$
- ❖ The multiplicity  $\Omega_{N,J}$  may be determined!

5

The number of sets of numbers  $n_{sub\ i}$  that satisfy the additive constraint is called the multiplicity of the system and symbolized with an “Omega”. In general, the multiplicity of a system is the number of ways it may exist while maintaining the same gross features.

This multiplicity is a function of the total number of atoms  $N$  and the number  $I$  of different molecular sizes.

The logarithmic relation between entropy and multiplicity is observed as long as each of the ways of partitioning the conserved quantity is equally probable. Physicists call this hypothesis the fundamental postulate of statistical mechanics. In physics jargon the fundamental postulate tells us that “all the microstates of an isolated system are equally probable.”

The current system is, indeed, essentially isolated because the total number of its atoms  $N$  is fixed even while the number in each partition  $n_{sub\ i}$  can vary.

---

# Key Theorem:

---

## Volume of Thin Surface = Volume Enclosed

❖ Theorem:  $\Omega_{N,J} \Delta \rightarrow V_{N,J}$

❖ Recall,  $S_{N,J} = \ln \Omega_{N,J}$

6

The theorem we adopt in order to help us determine the multiplicity (and therefore the entropy) in terms of the system's gross quantities is one I found in a book by the American thermodynamicist Herbert Callen. It goes like this: The number of dimensions  $I$  can always be made large enough, so that the number of points within a surface having a small thickness "Delta" is approximately equal to the number of points contained in the volume enclosed by that surface. In brief, the volume of a thin, multi-dimensional surface is equivalent to the volume enclosed by that surface.

This theorem is, in our case, fairly easy to prove. For anyone interested, I have a short preprint in which it is worked out.

## Determining the Multiplicity and Entropy

- ❖ The volume  $V_{N,I} [= \Omega_{N,I}]$  is determined via the geometry of “high dimensionality” for which  $I \gg 1$ .
- ❖ See *Thermodynamics and Equilibrium Statistical Mechanics* by Herbert B. Callen (Wiley, 1985) or *Regular Polytopes* by H. S. M. Coxeter (Dover, 1973)

- ❖ Exact result:

$$V_{N,I} = \frac{N^I}{(I!)^2} \quad \therefore S_{N,I} = \ln \left[ \frac{N^I}{(I!)^2} \right]$$

7

The number of points contained by the volume enclosed by this surface is also easy to derive. Again, I borrow from Herbert Callen who may have taken his method and this result from the Canadian mathematician H. S. M. Coxeter who, in his book *Regular Polytopes*, codified and popularized results on multi-dimensional geometry first discovered in mid-nineteenth century.

Once we know how the multiplicity “Omega” depends upon N and I, as shown here, we know everything there is to know about the system as a whole — for in thermodynamics everything follows from the entropy as a function of its additive quantity (or quantities). In this case there is only one additive quantity N — and we see here how S depends upon N.

# Thermodynamical Relations

- ❖ The First and Second Law  
of Thermodynamics

$$dN = T_{N,I} dS_{N,I}$$

- ❖ The Definition of Temperature

$$\frac{1}{T_{N,I}} = \frac{\partial S_{N,I}}{\partial N}$$

$$\therefore T_{N,I} = \frac{N}{I}$$

8

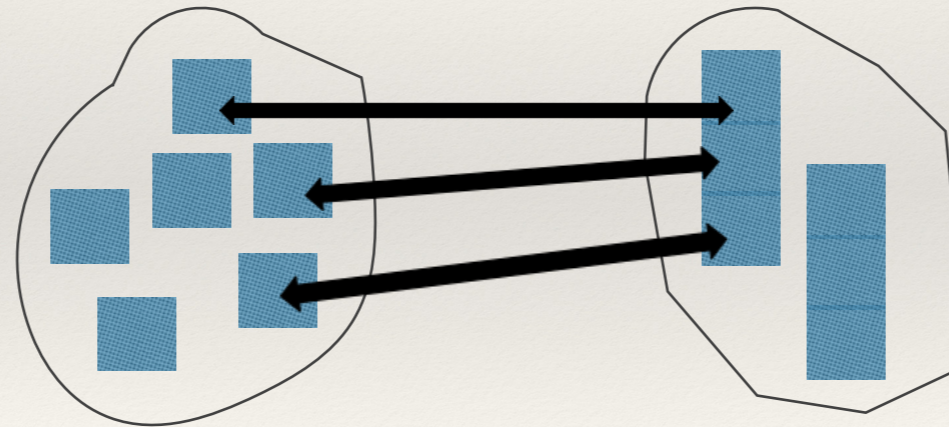
Conceptually, there is a lot of to this slide. So let's spend a little time with it.

Since the number of "atoms"  $N$ , like the energy of a thermodynamic system, is an additive quantity, there is no difference in the role these two quantities (Energy and  $N$ ) play in the structure of the thermodynamics of their relevant systems and subsystems. The first equation is an expression of the first and second laws of thermodynamics in which the  $TdS$  term plays the role of the heating (or cooling) of the system.

The second and third equations are merely mathematical deductions from the first. The temperature of this system works out to be  $N/I$  — the total number of atoms divided by the number of partitions of different-sized molecules.



# Temperature Regulates the Dynamic Equilibrium among Subsystems

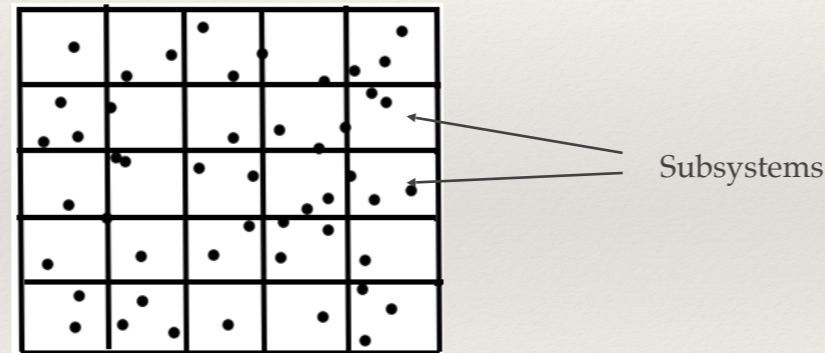


9

The temperature  $N/I$  accounts for and regulates the continuous fragmentation and combination of the pieces of the conserved quantity. Here we see how the units move between two different partitions, one containing only single atoms and the other containing only atoms in triplets. As we see three singlets may join to make one triplet or one triplet may fragment to make three singlets. The temperature of the system regulates this fragmentation and combination.

Each of these partitions and others are subsystems of the overall system.

# Physics Analog of Subsystems in Equilibrium: A Gas

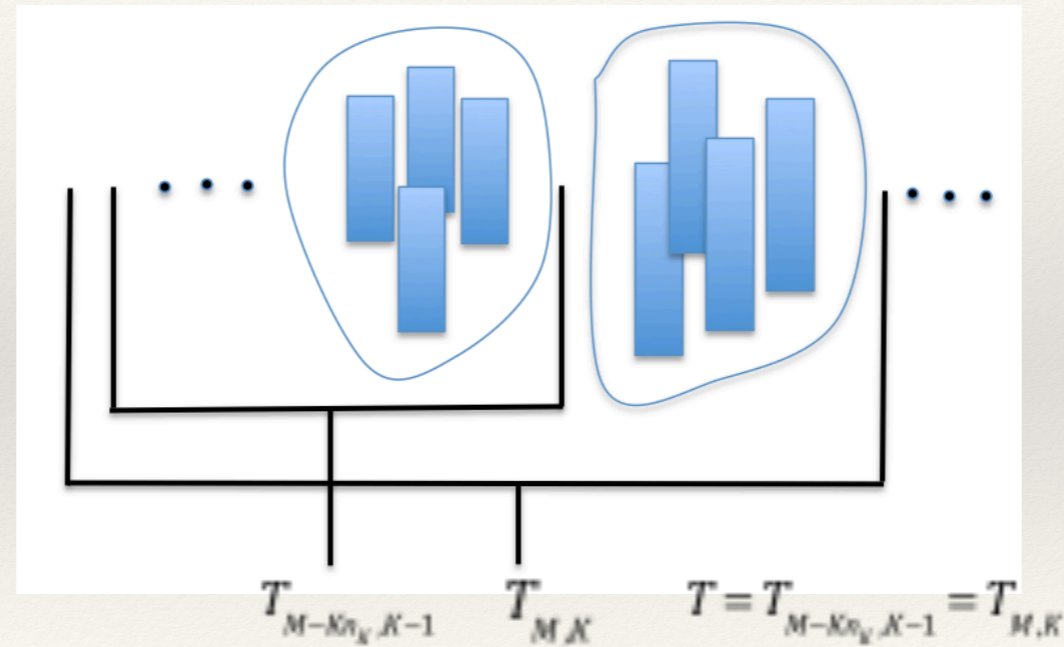


10

An (imperfect) physics analogy of this process exists. Here the various parts of a gas, as these parts exchange particles and energy, are the subsystems in equilibrium with one another. The analogy is imperfect because an isolated system of gas conserves two quantities: energy and particle number. The analogy is also imperfect because the subsystems of an ideal gas are simply smaller versions of the whole system, whereas the subsystems of the Benford system composed of pieces of different size are not smaller versions of the whole.

There is a technical word that describes this distinction: extensivity. The entropy of an ideal gas is an extensive function of its energy and particle number while the entropy of the current system of atoms separated into partitions of different-sized pieces is not an extensive function of its additive quantity.

## The Temperatures of Various Combinations of Subsystems are Equal to One Another



11

The subsystems of a system are in equilibrium with each other and with the whole system when they all have the same temperature.

Here we see that an arbitrary system containing, say, M units partitioned into K successive partitions must have the same temperature as one of its own subsets containing one less size category of sizes and so having only M - n sub K times K-1 units. This equilibrium is the key to deriving the Benford distribution and the first digit law.

# Deriving the Benford Distribution from Subsystem Equilibrium

$$T = \frac{M}{K} = \frac{M - K \cdot n_K}{K - 1}$$

$$\therefore n_K = \frac{T}{K}$$

12

Accordingly, the temperature of one sub-system characterized by M and K is the same as that of another subsystem characterized by M - n sub K and K-1, that is, by one less category of sizes. Solving the equation representing this equilibrium we find that the number of subsystems of size K is inversely proportional to K, the proportionality constant being the common temperature T.

This deduction shows that, from among those things that can combine and fragment and are restricted by (only) the conservation of a quantity, there emerges the Benford distribution in which the number of things is inversely proportional to their size.

## Benford's First Digit Law follows from the Benford Distribution

$$\frac{\int_{d \cdot 10^p}^{(d+1) \cdot 10^p} \frac{dx}{x}}{\int_{10^p}^{10^{p+1}} \frac{dx}{x}} = \log_{10} \left( 1 + \frac{1}{d} \right)$$

$d = 1, 2, \dots, 9$

13

When the set of successive integers describing the successive sizes is replaced with a continuous variable, the Benford distribution follows and from this distribution emerges the first digit law. On the left is the ratio of all the numbers within one decade starting with digit  $d$  to all the numbers in this decade. Benford's first digit law emerges whenever the integration is over an integer number of decades.

On the other hand, whenever the upper (or lower) limit of the integration falls short of completing a decade, the number of higher (or lower) digits falls short of what the first digit law predicts.

There are other ways the fragmentation and combining of a conserved quantity can fail to produce the Benford distribution and the first digit law. If, for instance, there are not enough individual units to populate the different partitions of different-sized quantities, then the partitions of larger molecules will be depressed below that expected by the first digit law.

In summary, in order for the current derivation to produce Benford's distribution and the first digit law, the dimensionality  $l$ , the multiplicity  $\Omega$ , and the quantity  $N$  must all be large numbers.

---

## Deviations from Benford's First Digit Law Signal *Science*

---

- ❖ Deviations from Benford's law signal not only possible fraud but also the operation of constraints other than that of mere conservation of a single quantity.
- ❖ Frank Benford: "the logarithmic law applies particularly to those outlaw numbers that are without known relationship."

Here is something I have only recently realized — in fact, while traveling to Stresa. If data does not observe the first digit law, this may be an indication, not only of fraud but also of missing science. For instance, deviations from the first digit law may indicate that more than one quantity is conserved.

Frank Benford's words suggest as much, that is, they suggest that the logarithmic law applies when numbers observe no other order than that imposed by the unbiased partitioning of a conserved quantity. Such random data is relatively "outlaw," in Benford's phrase, that is relatively unordered by law. Data that is structured or ordered by more than the unbiased partitioning of a conserved quantity will deviate from Benford's first digit law.

Conceivably, we could use these deviations to infer the missing science described by the extra conservation laws.