

ICT2.0:2020-W116

THERMODYNAMICS AND THE EVOLUTION OF STOCHASTIC POPULATIONS

Themis Matsoukas

Department of Chemical Engineering,
 Pennsylvania State University,
 State College, PA 16802,
 Email: txm11@psu.edu

ABSTRACT

The appeal of thermodynamics to problems outside physics is undeniable, as is the growing recognition of its apparent universality, yet in the absence of a rigorous formalism divorced from the peculiarities of molecular systems all attempts to generalize thermodynamics remain qualitative and heuristic at best. In this paper we formulate a probabilistic theory of thermodynamics and set the basis for its application to generic stochastic populations.

1 WHAT IS THERMODYNAMICS?

There is growing recognition that thermodynamics is a universal science but what *is* thermodynamics? The typical answer is some variation of “thermodynamics is the study of the equilibrium properties of matter.” Undoubtedly the study of the equilibrium state of matter falls within the scope of thermodynamics, clearly though, such statement is far too narrow to serve as a definition of thermodynamics, far too narrow to justify its appeal in areas *outside* physics. The answer describes one particular problem (albeit a very important one) that thermodynamics can tackle, but tells little about what thermodynamics is capable of. *What is thermodynamics?*

Until Boltzmann and Gibbs thermodynamics was about heat and work and its central aim was to determine the maximum amount of work that can be extracted from heat in a cycle. This is a variational problem and its solution was found via a variational principle, the maximization of entropy. With Boltzmann and Gibbs [1], thermodynamics entered the realm of the probabilistic world –while still footed in the physical world of mechanical

particles, atoms and molecules. Entropy now has a probabilistic interpretation, it is a functional of the probability distribution p_i of microstates,

$$S[\{p\}] = - \sum_i p_i \log p_i. \quad (1)$$

The equilibrium distribution p_i^* is identified among all feasible distributions as the one that maximizes this functional:

$$S[\{p^*\}] \geq S[p]. \quad (2)$$

A feasible distribution is one that satisfies the macroscopic constraints that define the system. For a system with fixed mean energy $\bar{E}$, volume V and number of particles N this means any distribution with mean energy $\bar{E}$, i.e., any distribution that satisfies

$$\sum_i E_i p_i = \bar{E}. \quad (3)$$

The solution to this variational problem is the canonical probability of microstate,

$$p_i^* = \frac{e^{-\beta E_i}}{Q(\beta, V, N)}, \quad (4)$$

Q is the canonical partition function and β is the inverse temperature. With minimal additional calculus β and Q are expressed in terms of the microcanonical partition function $\Omega(E, V, N)$,

$$\beta = \frac{\partial \log \Omega}{\partial E}; \quad \log Q = \log \Omega - \bar{E} \frac{\partial \log \Omega}{\partial E}; \quad (5)$$

and similarly $\bar{E}$ is expressed in terms of the canonical partition

function

$$\bar{E} = -\frac{\partial \log Q}{\partial \beta}. \quad (6)$$

Equations (1) through (6) cast the variational problem in terms of a *probability space*, the space of all probability distributions that are feasible under the macroscopic specification of state. These equations bear only indirect relationship to the physical laws that govern that particles that produce the microstates in question via E_i . It should not be a surprise that thermodynamics survived the quantum revolution. Quantum mechanics changes the calculation of the energy E_i of microstate but leaves the entire network of Eqs. (2)–(6) intact. The hint is clear: thermodynamics is *not* a physical theory.

The door that allowed thermodynamics to escape the confines of physics was unlocked by Shannon, who arrived at the entropy functional, the same functional as that of Boltzmann and Gibbs, by methods that have nothing to do with physical particles or laws of motion [2]. Even so the door remained closed until Jaynes pushed it open. Jaynes conjectured that thermodynamics is about inference, not physics:

[...] we may have now reached a state where statistical mechanics is no longer dependent on physical hypotheses, but may become simply an example of statistical inference [3].

With this bold statement the toolbox of thermodynamics was made available to anyone who can use it. During the more than 60 years since Jaynes's pronouncement thermodynamics has been steadily moving closer to achieving the status of universal language with applications in ecology [4], epidemics [5], neuroscience [6], financial markets [7], and in the study of complexity in general. But what are the rules of this universal science? We have no comprehensive theory of generalized thermodynamics and no guidance on how to connect a particular system to thermodynamics. One is largely left to construct analogies to molecular systems and hope they hold.

Here we attempt to decipher the universal grammar of thermodynamics. First, we construct a mathematical formalism based on variational calculus that generates the entire network of mathematical thermodynamics, Eqs. (1) through (6), via a variational construction that assigns probabilities to a set of probability distributions and seeks the most probable distribution in the set. The most probable distribution is identified by maximizing its probability and this leads to the maximization of a functional analogous to but different from entropy, in fact, a generalized (negative) free energy. We show that any probability distribution can be expressed in the formalism of thermodynamics and illustrate how these ideas can be applied to a generic population that evolves under a stochastic process. The paper builds on work recently published in Refs. [8] and [9].

2 THERMODYNAMICS BEYOND MOLECULES

2.1 SAMPLING STATISTICS

We start with a discrete random variable $X = i$, $i = 1, 2, \dots$ with probability distribution $P(X = i) = h_i$. We collect a random sample of N points and count the number n_i of samples with $X = i$. The (intensive) empirical distribution of the sample is $p_i = n_i/N$ and represents a guess of the true distribution h_i . The probability to obtain a particular (extensive) distribution $\{n\} = N\{h\}$ is

$$P(\{n\}) = P(\{h\}) = N! \prod_{i=1}^N \frac{h_i^{n_i}}{n_i!}, \quad (7)$$

and its log is

$$\log P(\{h\}) = -N \sum_i p_i \log \frac{p_i}{h_i}. \quad (8)$$

We recognize the result as the Kullback-Leibler divergence [10] (or relative entropy). This functional, as is well known, is maximized by $p_i^* = h_i$, a result that makes intuitive sense: in a random sample taken from some distribution the most likely distribution to materialize in the sample is the distribution that is being sampled, even though other distributions are possible. When N is large (thermodynamic limit), p_i is overwhelmingly more probable than any other distribution.

We now bias the sampling process so that we may obtain a different distribution as the most probable distribution from the one that is being sampled. We bias the process by a functional $\{n\}$ such that the probability to obtain a sample with extensive distribution $\{n\}$ is

$$P(\{n\}) = \frac{W(\{n\})}{r^N} \left(N! \prod_{i=1}^N \frac{h_i^{n_i}}{n_i!} \right), \quad (9)$$

where r^N is a normalization constant. We require the bias to be of the form [9]

$$\log W(\{n\}) = \sum_i n_i \frac{\partial \log W(\{n\})}{\partial n_i} = \sum_i n_i \log w_i \quad (10)$$

with $\log w_i = \partial \log W(\{n\}) / \partial n_i$. This makes $\log W$ homogeneous functional of $\{n\}$ with degree 1 and ensures that sampling converges to a limiting distribution when $N \rightarrow \infty$ [9]. The log of probability to obtain the empirical distribution $p_i = n_i/N$ now is

$$\frac{\log P(\{p\})}{N} = - \sum_i p_i \log \frac{p_i}{h_i w_i} - \log r. \quad (11)$$

when $N \rightarrow \infty$, the set of feasible distributions contains all distributions that satisfy the normalization condition $\sum_i h_i = 1$. The most probable distribution is obtained by maximizing Eq. (11) with respect to p_i :

$$p_i^* = \frac{w_i}{r} h_i. \quad (12)$$

As we intended, the most probable distribution is different from the sampled distribution by a factor w_i/r . We can choose the partial bias w_i so as to obtain *any* distribution as the MPD of the biased sampling process: if we pick $w_i = rf_i/h_i$, where f_i is any distribution in the domain of h_i , then f_i will materialize as the most probable distribution by this biased sampling. In the limit $N \rightarrow \infty$ the MPD is overwhelmingly more probable than any other distribution.

The conclusion from this analysis is this: Sampling from a distribution establishes an entire space of distributions that contains every distributions that is defined on the domain of the sampled distribution. The bias functional establishes a probability measure on the distributions of this space. This bias can be constructed so as to select *any* of these distributions as the most probable distribution. For this reason we refer to W as *selection functional*.

2.2 MICROCANONICAL SAMPLING

We now sample from an exponential distribution with mean $\bar{x}$,

$$h_i = \frac{e^{-i/\bar{x}}}{\bar{x}}, \quad (13)$$

and consider the subspace of empirical distributions with the same mean. The probability of empirical distribution $\{p\}$ is given by Eq. (11) with a new normalization constant r' . We write the result in the form

$$\frac{\log P(\{p\})}{N} = -\sum_i p_i \log \frac{p_i}{w_i} - \log \omega. \quad (14)$$

with $\log \omega = 1 + \log \bar{x} + \log r'$. The MPD is obtained by maximizing this functional with respect to all h_i that satisfy the constraint and whose mean is $\bar{x}$. The result is

$$p_i^* = \frac{w_i e^{-\beta i}}{q}, \quad (15)$$

with β and q parameters associated with the Lagrange multipliers for the two constraints in this maximization. We can show [9]

$$\bar{x} = -\frac{d \log q}{d \beta}, \quad (16)$$

which establishes a relationship between the mean of the MPD and its parameters β and q .

2.3 UNIVERSAL THERMODYNAMICS

Define the microcanonical functional $\rho[\{p\}] = \log P(\{p\})/N$ use Eq. (14) to write it as

$$\rho[\{p\}] = -\sum_i p_i \log \frac{p_i}{w_i} - \log \omega. \quad (17)$$

The condition that defines the most probable distribution now is expressed in the form of the inequality $\rho[\{p^*\}] \geq \rho[\{p\}]$. For

large N the MPD is overwhelmingly more probable than any other distribution. Then $P(\{p^*\}) \rightarrow 1$ and $\rho[\{p^*\}] \rightarrow 0$. The condition that defines the MPD then becomes

$$\rho[\{p\}] \leq 0, \quad (18)$$

with the equal sign only for $\{p\} = \{p^*\}$. This inequality is the central variational condition of the theory. Apply this to Eq. (17)

$$\log \omega \geq -\sum_i h_i \log p_i + \log W[\{p\}], \quad (19)$$

or more compactly,

$$\log \omega \geq S[\{p\}] + \log W[\{p\}]. \quad (20)$$

With $\{p\} = \{p^*\}$ from Eq. (15) the above becomes an exact equality. The result is

$$\log \omega = \beta \bar{x} + \log q. \quad (21)$$

The simplest way to make contact with thermodynamics is to take i to refer to the energy E_i of microstate and p_i to refer to the probability of microstate. The selection functional in this case is uniform, $W = w_i = 1$, and expresses the postulate of equal a priori probabilities. With these substitutions (15) reverts to the familiar canonical probability distribution and Eq. (20) reads

$$\log \omega \geq S[\{p\}]. \quad (22)$$

This is the inequality of the second law: The MPD has the maximum entropy among all possible probability distributions that can be assigned to the set of microstates. With Eqs. (15), (16), (20) and (21) we have obtained the generalized forms of the thermodynamic set in Eqs. (1)–(6) via an argument that contains no physics. In this formulation the inequality of the second law has a simple non controversial and in fact trivial meaning: it states that the most probable distribution is more probable than any other in the set of feasible distributions.

3 EVOLUTION OF EXTENSIVE POPULATIONS

The implications for the universality of thermodynamics are profound. Any probability distribution in the domain of the exponential distribution ($x \geq 0$) can be obtained as the most probable distribution in this domain and therefore can be described in the formalism of thermodynamics. The key to this formulation is the selection bias that selects the MPD among all distributions, but what is W ? In the case of molecular systems it is fixed by the postulate of equal a priori probabilities, a *model assumptions* that assigns a probability to the elements of the feasible space. In the general case it is set by the rules of the stochastic process that governs the evolution of the random variable in question. Let us sketch how this works in the case of an extensive population $\mathbf{n} = (n_1, n_2, \dots)$. We suppose the elements of the population

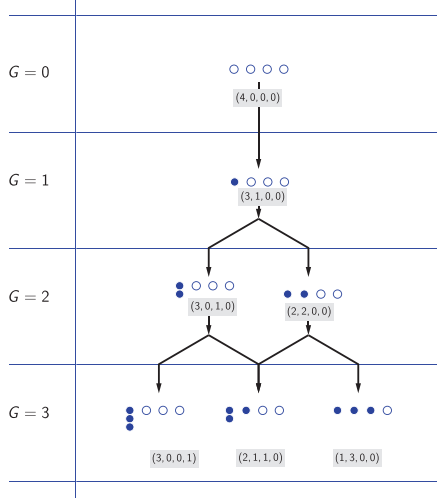


FIGURE 1. Trajectories of $N = 4$ walkers undergoing one transition at a time. Filled circles represent transitions accumulated by the walker. In this example n_i is the number of walkers with $g_i = i$ transitions. Distributions are represented pictorially and in vector form as (n_0, n_1, n_2, n_3) .

undergo transitions of the form

$$i \xrightarrow{p_{i \rightarrow j}} j \quad (23)$$

with transition probability $p_{i \rightarrow j}$. Accordingly, the distribution of walkers is transformed by these transitions and produces a Markov chain of the form

$$\cdots \mathbf{n}' \Big|_{G-1} \rightarrow \mathbf{n} \Big|_G \cdots \quad (24)$$

where $G = 0, 1, \dots$ is the generation of $\mathbf{n}$ and counts the number of transitions it has undergone from fixed initial state. We then have

$$G = \sum_i g_i n_i, \quad N = \sum_i n_i, \quad (25)$$

Here g_i is the number of transitions that produce element i and N is the number of elements in $\mathbf{n}$ and is constant in all generations.

The ensemble $\mathcal{E}_{G,N}$ is the set of all distributions that are formed by G sequential transitions starting from the same initial state (Fig. 1). The probability of distribution $\mathbf{n}$ in $\mathcal{E}_{G,N}$ is given by Eq. (9), which we now write as

$$P(\mathbf{n}) = \frac{\mathbf{n}!}{\Omega_{G,N}} W(\mathbf{n}), \quad (26)$$

with $\mathbf{n}! = N! / n_1! n_2! \cdots$ and $\Omega = \omega^N$. The distributions of the ensemble are sampled via the trajectories that lead to them and the selection functional W is determined by the transition probabilities along these trajectories. The most probable distribution, obtained by maximizing Eq. (26) under the two constraints in (21),

is given by Eq. (15) and obeys all thermodynamic relationships in Eqs. (16), (20) and (21). The entire formalism of thermodynamics is transferred to generic stochastic processes.

4 CONCLUSIONS

Stripped to its core statistical thermodynamics is the calculus of the most probable distribution. The theory outlined here establishes a space of distributions (the microcanonical ensemble) and a probability measure on it via a sampling process. This mathematical construction gives rise to the relationships we recognize as thermodynamics. A stochastic process may be viewed as a network of transitions that emanate from a known initial state and branch into the future. The ensemble is the set of distributions that can be reached in a given number of steps, and sampling refers to arriving to a distribution via a trajectory of transitions. The specifics of the process (transition probabilities) enter via the selection functional –this is the contact point between generalized thermodynamics and the particulars (“physics”) of system that is being studied. We suggest that this formalism can be applied to a variety of stochastic populations and plan to offer examples in the near future.

REFERENCES

- [1] Gibbs, J. W., 1981. *Elementary Principles in Statistical Mechanics*. Ox Bow Press, Woodbridge, CT. (reprint of the 1902 edition).
- [2] Shannon, C. E., 1948. “A mathematical theory of communication”. *Bell System Technical Journal*, **27**(3), pp. 379–423.
- [3] Jaynes, E. T., 1957. “Information theory and statistical mechanics”. *Phys. Rev.*, **106**(4), pp. 620–630.
- [4] Harte, J., Zillio, T., Conlisk, E., and Smith, A. B., 2008. “Maximum entropy and the state-variable approach to macroecology”. *Ecology*, **89**(10), pp. 2700–2711.
- [5] Durrett, R., 1999. “Stochastic spatial models”. *SIAM Review*, **41**(4), pp. 677–718.
- [6] Timme, N. M., and Lapish, C., 2018. “A tutorial for information theory in neuroscience”. *eNeuro*.
- [7] Voit, J., 2005. *The Statistical Mechanics of Financial Markets*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [8] Matsoukas, T., 2019. *Generalized Statistical Thermodynamics: Thermodynamics of Probability Distributions and Stochastic Processes*. Springer International Publishing.
- [9] Matsoukas, T., 2019. “Thermodynamics beyond molecules: Statistical thermodynamics of probability distributions”. *Entropy*, **21**(9).
- [10] Kullback, S., and Leibler, R. A., 1951. “On information and sufficiency”. *Ann. Math. Statist.*, **22**(1), 03, pp. 79–86.