A Bayesian Interpretation of First-Order Phase Transitions

Sergio Davis[1] · **Joaquín Peralta**[2] · **Yasmín Navarrete**[1] ·
Diego González[1] · **Gonzalo Gutiérrez**[1]

Abstract In this work we review the formalism used in describing the thermodynamics of first-order phase transitions from the point of view of maximum entropy inference. We present the concepts of transition temperature, latent heat and entropy difference between phases as emergent from the more fundamental concept of internal energy, after a statistical inference analysis. We explicitly demonstrate this point of view by making inferences on a simple game, resulting in the same formalism as in thermodynamical phase transitions. We show that analogous quantities will inevitably arise in any problem of inferring the result of a yes/no question, given two different states of knowledge and information in the form of expectation values. This exposition may help to clarify the role of these thermodynamical quantities in the context of different first-order phase transitions such as the case of magnetic Hamiltonians (e.g. the Potts model).

Keywords Maximum entropy · Bayesian inference · Phase transitions

1 Introduction

Jaynes' proposal of the principle of maximum entropy (PME) as a general tool of probabilistic inference [1,2] is remarkable in that it is both widely used [3] and somewhat controversial [4,5]. It asserts that the most unbiased probability distribution P given some fixed knowledge $\mathcal{I}$ is the one that maximizes Shannon's information entropy,

✉ Sergio Davis
sdavis@gnm.cl

1 Grupo de Nanomateriales, Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

2 Departamento de Física, Facultad de Ciencias Físicas, Universidad Andrés Bello, Santiago, Chile

$$S[P(x|\mathcal{I})] = -\sum_x P(x|\mathcal{I}) \log_2 P(x|\mathcal{I}) \qquad (1)$$

while being consistent with said knowledge. This is because $S[P]$ is a measure of uncertainty [6] or lack of knowledge about the degrees of freedom (represented collectively by x in the notation above) and, maximizing it leads to the probabilistic model containing the least amount of information, but nevertheless able to reproduce the features one demands of it. As this is a process of inference it cannot be deductive: predictions derived from the maximum entropy model may be proved wrong by subsequent measurements, and this reflects an incompleteness of the fixed knowledge used to constrain the maximization.

Jaynes' interpretation of the formalism of statistical mechanics sees it as just the application of this principle of maximum entropy, valid in all statistical inference, to the case of a macroscopic number of particles (and degrees of freedom). In this situation the predictions are almost perfectly sharp, with uncertainties vanishing as $1/\sqrt{N}$, with N the number of degrees of freedom. This is all well described for the case of thermodynamic equilibrium of a single phase. However, how this information-theoretical interpretation manifests itself in the case of the study of phase transitions, and what can we learn from this, is an issue which has not been so extensively clarified. For instance, in his book on probability theory [2] (p. 602), Jaynes wrote in a somewhat cryptic footnote, that

"... in statistical mechanics the relative probability P_j/P_k of two different phases, such as liquid and solid, is the ratio of their partition functions Z_j/Z_k, which are the normalization constants for the sub-problems of prediction within one phase. In Bayesian analysis, the data are indifferent between two models when their normalization constants become equal; in statistical mechanics the temperature of a phase transition is the one at which the two partition functions become equal..."

This suggests that the problem of liquid-solid phase transition, or in fact, any phase transition, can be posed as a model comparison problem, and therefore the transition temperature and the free energy can be given an information-theoretical meaning. In this work, we attempt to clarify ("demistify", one might even say) the interpretation of these quantities, showing that they are consequences of the maximum entropy inference rule under known internal energy averages. In this sense, only this internal energy is fundamental, whereas temperature, entropy and free energy are derived quantities, of a statistical nature. This statistical character of temperature is understood, for instance, in terms of the kinetic theory of gases; however, we will show that its meaning is much wider in the context of information theory.

In order to remove all particularities of thermodynamics from the treatment of first-order phase transitions, we present a parallel of the formalism used in first-order phase transitions based entirely on the application of the PME. We introduce a simple game, the "disc throwing" game, and answer two questions related to it by means of the PME. In the answers to these questions we will recover the concepts of transition temperature, Helmholtz free energy and the rule that imposes its equality for the two phases at the coexistence point.

The rest of the paper is organized as follows. In Sect. 2 we review the main features of the maximum entropy formalism. Section 3 shows an illustration of PME inference, while Sect. 4 describes and solves the disc throwing game problem. In Sect. 5 we expose the perfect parallel between the solution of this problem and that of the coexistence of two phases in thermodynamical equilibrium. Finally we conclude with some remarks.

2 Maximum Entropy Inference

Consider a system having N discrete degrees of freedom $\mathbf{x} = (x_1, \ldots, x_N)$ and being fully described in statistical terms by a function $f(\mathbf{x})$ with known expectation value f_0. Knowledge of f_0 is symbolically represented by $\mathcal{I}$. According to the PME, the most unbiased model is the one that maximizes the Gibbs-Shannon entropy functional

$$S = -\sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) \log_2 P(\mathbf{x}|\mathcal{I}) \tag{2}$$

subject to the constraint $\mathcal{I}$, i.e., to

$$\left\langle f(\mathbf{x}) \right\rangle = f_0. \tag{3}$$

Maximization under this constraint, and the always implicit constraint of proper normalization of the probability, is achieved by the inclusion of Lagrange multipliers λ and μ respectively, after which the problem reduces to the maximization of the augmented function

$$\tilde{S} = -\sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) \log_2 P(\mathbf{x}|\mathcal{I}) + \lambda(f_0 - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I}) f(\mathbf{x}))$$
$$+ \mu(1 - \sum_{\mathbf{x}} P(\mathbf{x}|\mathcal{I})). \tag{4}$$

This leads to the well-known maximum entropy (MaxEnt) model

$$P(\mathbf{x}|\lambda) = \frac{1}{Z(\lambda)} \exp(-\lambda f(\mathbf{x})) \tag{5}$$

in which we have changed the notation from the purely abstract $P(\mathbf{x}|\mathcal{I})$ to the mode concrete $P(\mathbf{x}|\lambda)$, given that the parameter λ distinguishes between all the possible states of knowledge compatible with the possible values of f_0. The function Z,

$$Z(\lambda) = \sum_{\mathbf{x}} \exp(-\lambda f(\mathbf{x})). \tag{6}$$

is known as the partition function. The Lagrange multiplier λ is usually determined as the unique solution of

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = f_0. \tag{7}$$

The procedure just outlined could in principle be performed using so-called generalized entropies in place of the Gibbs-Shannon entropy (Eq. 2), such as the Tsallis [7] or Rényi [8] entropies. Although these entropies may be valid tools in describing the complexity of non-extensive systems, their use in statistical inference has been shown to lead to inconsistencies [9–12]. If the degrees of freedom contained in **x** are continuous, Shannon entropy needs to be replaced with the relative entropy

$$S = - \int d\mathbf{x} P(\mathbf{x}|\mathcal{I} \wedge I_0) \log_2 \frac{P(\mathbf{x}|\mathcal{I} \wedge I_0)}{P(\mathbf{x}|I_0)} \tag{8}$$

where I_0 denotes an "initial" state of knowledge. The solution to the maximum entropy problem is now

$$P(\mathbf{x}|\mathcal{I} \wedge I_0) = \frac{1}{Z(\lambda)} P(\mathbf{x}|I_0) \exp(-\lambda f(\mathbf{x})) \tag{9}$$

with

$$Z(\lambda) = \int d\mathbf{x} P(\mathbf{x}|I_0) \exp(-\lambda f(\mathbf{x})). \tag{10}$$

In both cases (discrete and continuous degrees of freedom), the maximized entropy has a value

$$S = \ln Z(\lambda) + \lambda f_0. \tag{11}$$

Now, we have just described the formalism of the canonical ensemble if we think of the system as composed by n particles with position $\mathbf{r}_i$ and momentum $\mathbf{p}_i$ (with $i=1,...,n$) and the descriptor function as the Hamiltonian $f = \mathcal{H}(\mathbf{r}_1, \ldots, \mathbf{r}_n, \mathbf{p}_1, \ldots, \mathbf{p}_n)$. Then Eq. 5 is the canonical distribution where we identify $\lambda = \beta = 1/(k_B T)$.

In thermodynamic notation, Eq. 11 reads,

$$S(\beta)/k_B = \ln Z(\beta) + \beta E(\beta) \tag{12}$$

If we introduce the Helmholtz free energy $\beta F(\beta) = - \ln Z(\beta)$, Eq. 12 reduces to

$$S(\beta)/k_B = \beta(E(\beta) - F(\beta)) \tag{13}$$

i.e., using $\beta = 1/k_B T$,

$$F(T) = E(T) - T S(T). \tag{14}$$

3 An Illustration of the Maximum Entropy Formalism

Suppose we have a swimming pool full of plastic balls (all spherical) of different radii. The average volume of a ball is V. What is the average radius?

We have the constraint,

$$\left\langle \frac{4}{3}\pi r^3 \right\rangle = V, \tag{15}$$

which is equivalent to

$$\langle r^3 \rangle = \frac{3}{4\pi} V, \tag{16}$$

from which the most unbiased model for r is

$$P(r|\lambda) = \frac{1}{Z(\lambda)} \exp(-\lambda r^3) \Theta(r). \tag{17}$$

The partition function is given by

$$Z(\lambda) = \int_0^\infty dr \exp(-\lambda r^3) = \Gamma(4/3)\lambda^{-1/3}, \tag{18}$$

therefore, the value of λ is determined from

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = \frac{1}{3\lambda} = \frac{3}{4\pi} V, \tag{19}$$

i.e., $\lambda = 4\pi/(9V)$. Note that from inspection of Eq. 17 and the fact that λ is positive, the most probable radius is zero and the probability monotonically decreases with r. The expectation of r is then

$$\langle r \rangle = 7\frac{1}{Z(\lambda)} \int_0^\infty dr\, r \exp(-\lambda r^3) = \frac{3^{1/3}}{3} \frac{\Gamma(2/3)}{\Gamma(4/3)}$$
$$\cdot \left(\frac{3V}{4\pi}\right)^{1/3} \approx 0.729011 \sqrt[3]{\langle r^3 \rangle}. \tag{20}$$

From this example two interesting things emerge. First, the expected radius is less than the naïve estimate $r_0 = \sqrt[3]{\langle r^3 \rangle}$, valid in the case where all the balls have the same radius. Second, the Lagrange multiplier λ is larger for small V, and this is expected given that the smaller V is, the possible radii are more concentrated around zero and therefore there is less uncertainty about the value of the radius. This means the constraint of known V (Eq. 15) has greater "weight" for smaller V. As the distribution function $P(r|\lambda)$ decreases from $r = 0$ onward, there are more balls with $r \leq r_0$ than with $r > r_0$ and thus the estimate $\langle r \rangle$ is skewed towards zero.

4 A Simple Disc Throwing Game

Suppose a player can throw a disc into a surface A (with area Σ_A), containing within it a smaller surface B (with area $\Sigma_B < \Sigma_A$). We consider A and B to be disjoint regions, as shown in Fig. 1. A successful hit within B gives the player n_B points, whereas a hit inside A (outside B) gives n_A points to the player (as hitting B is more difficult, $n_B > n_A$). This is similar to the game "rayuela" as is known in some South American countries.

We can present two questions about this game:

Fig. 1 Schematic representation
of the disc throwing game
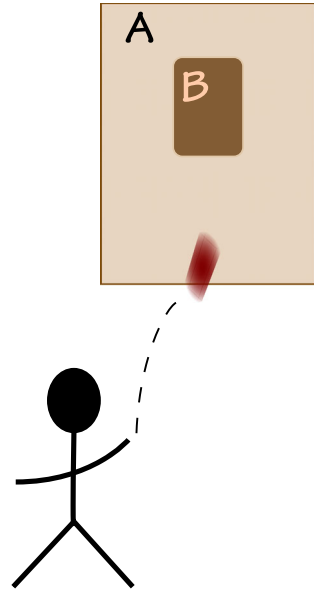
(a) With only the information laid out above, and particularly without knowing any-
thing about the performance of the player, what probability should one assign to
hitting B?

(b) Now consider the player has obtained an average score of $\bar{n}$ in the past (over
enough trials to be considered a reliable average). What probability should one
assign now to hitting B?

In (a) the intuitive answer is that the probabilities of hitting either A or B are com-
pletely determined by their areas. In fact, considering each landing point as a coordinate
inside A, and because such points are mutually exclusive, exhaustive alternatives and
there is symmetry under exchange, we can easily see that

$$\frac{P(A|\mathcal{I}_1)}{P(B|\mathcal{I}_1)} = \frac{\Sigma_A}{\Sigma_B} \tag{21}$$

From this, given that landing in A or B constitute mutually exclusive and exhaustive
propositions, $P(A|\mathcal{I}_1) + P(B|\mathcal{I}_1) = 1$. Therefore,

$$P(\alpha|\mathcal{I}_1) = \frac{\Sigma_\alpha}{\Sigma_A + \Sigma_B} \tag{22}$$

with $\alpha = A, B$. The predicted score of the player, with just the information we have
in (a), is then

$$\bar{n} = \frac{\Sigma_A n_A + \Sigma_B n_B}{\Sigma_A + \Sigma_B}. \tag{23}$$

We see that probabilities are governed only by the ratio Σ_A/Σ_B, and we can con-
clude that always $P(A|\mathcal{I}_1) > P(B|\mathcal{I}_1)$, given that the area of B is considerably

smaller. Now, what happens in (b) is that we have to constrain the inference to this new information, given in the form of an expectation value. We invoke the law of large numbers and assume $\langle n \rangle = \overline{n}$, then the most unbiased probability for either result given $\overline{n}$, according to the PME, is (using Eq. 9),

$$P(\alpha|\mathcal{I}_2) = \frac{1}{Z(\lambda)} \Sigma_\alpha \exp(-\lambda n_\alpha) \tag{24}$$

with

$$Z(\lambda) = \Sigma_A \exp(-\lambda n_A) + \Sigma_B \exp(-\lambda n_B), \tag{25}$$

and

$$-\frac{\partial}{\partial \lambda} \ln Z(\lambda) = \overline{n}. \tag{26}$$

After explicitly using the result of Eq. 25 in Eq. 26 and some algebra, we have that

$$\Sigma_A (\overline{n} - n_A) \exp(-\lambda n_A) = \Sigma_B (n_B - \overline{n}) \exp(-\lambda n_B) \tag{27}$$

from which it follows that λ is given by

$$\lambda(\overline{n}) = -\frac{1}{n_B - n_A} \left[\ln \Sigma_A - \ln \Sigma_B + \ln \frac{\overline{n} - n_A}{n_B - \overline{n}} \right]. \tag{28}$$

In order to simplify notation, let us introduce

$$\Delta n = n_B - n_A, \tag{29}$$
$$\Delta S = S_B - S_A = \ln \Sigma_B - \ln \Sigma_A. \tag{30}$$

Then Eq. 28 reads,

$$\lambda \Delta n - \Delta S = \ln \frac{n_B - \overline{n}}{\overline{n} - n_A} \tag{31}$$

It is clear that, when $\lambda = 0$, Eq. 31 implies

$$\frac{n_B - \overline{n}}{\overline{n} - n_A} = \frac{\Sigma_A}{\Sigma_B}. \tag{32}$$

which is nothing but the result of section (a), Eq. 23. This happens when the reported average score $\overline{n}$ is the same as predicted from the area information alone. This reflects a complete lack of ability from the player to control the hitting spot, because the results do not differ from pure "random" shots. However, if $\overline{n}$ is not consistent with Eq. 23, then $\lambda \neq 0$ and the ratio between probabilities is not simply the ratio of the respective areas, but it is given by

$$\frac{P(A|\mathcal{I}_2)}{P(B|\mathcal{I}_2)} = \frac{\Sigma_A}{\Sigma_B} \exp(-\lambda(n_A - n_B)) \tag{33}$$

i.e., defining $\Delta \ln P = \ln P(B|\mathcal{I}_2) - \ln P(A|\mathcal{I}_2)$,

$$\Delta \ln P = \Delta S - \lambda \Delta n, \tag{34}$$

or, if we define $F_\alpha = n_\alpha - S_\alpha/\lambda$, we have

$$\Delta \ln P = -\lambda \Delta F. \tag{35}$$

Therefore, the most probable outcome (A or B) would be the one with lowest value of F.

After comparing Eqs. 34 and 31, the ratio of probabilities is given by

$$\frac{P(A|\mathcal{I}_2)}{P(B|\mathcal{I}_2)} = \frac{n_B - \bar{n}}{\bar{n} - n_A}. \tag{36}$$

There will be an interesting value of $\bar{n}$, namely the average $(n_A + n_B)/2$, where $P(A|\mathcal{I}_2) = P(B|\mathcal{I}_2)$. In this case we are maximally uncertain with respect to which region the player will hit, i.e., we have "canceled out" all the information we had from the areas by using the average score. This situation corresponds to a "critical value" of the Lagrange multiplier,

$$\lambda_0 = \lambda\left(\frac{n_A + n_B}{2}\right) = \frac{\Delta S}{\Delta n}. \tag{37}$$

5 Bayesian Thermodynamics

Perhaps it will be striking to the reader (at first) to notice that we have replicated the formalism used to study first-order phase transitions in thermodynamical systems. Imagine the two regions A and B of the game introduced previously, as regions in phase space corresponding, for instance, to liquid and solid, respectively. We can relate the area of each region Σ to the volume in phase space occupied by each of the thermodynamic phases, and in this sense, the quantity

$$S = \ln \Sigma \tag{38}$$

is readily interpreted as the Boltzmann entropy (taking k_B=1). Therefore the most probable phase (i.e., the most stable phase in thermodynamical terms) is, in absence of any other information, the one with the largest value of entropy. This is the same situation as in the microcanonical ensemble [13].

When we have information about the expected (or average) score $\bar{n}$, analogous to the measured internal energy E of a thermodynamical system (n_A and n_B are then the internal energies for the liquid and solid phases, respectively), what decides the most probable phase is, according to Eq. 35, the difference in the quantity

$$F = n - S/\lambda \tag{39}$$

Table 1 Equivalences between concepts arising in the analysis of the throwing game and thermodynamical concepts

Throwing game	Thermodynamics
Logarithm of area ($\ln \Sigma$)	Entropy (S)
Game average score ($\overline{n}$)	Internal energy (E)
Score difference (Δn)	Latent heat (L)
Critical multiplier ($1/\lambda_0$)	Transition temperature (T_0)

which is precisely the Helmholtz free energy (under the identification $\lambda = \beta = 1/T$),

$$F = E - TS. \tag{40}$$

If we are given a low enough value of energy (close to the energy of the ideal solid) then, despite the fact that the liquid phase has a larger entropy, we are forced to conclude that the system is in one of the (relatively) few solid phase points. Because this reversal of our prediction after knowing $\overline{n}$ is strikingly unexpected, this situation is described by a large value of the Lagrange multiplier λ which, in the context of thermodynamics, corresponds to a low value of temperature T.

The limiting situation when we cannot claim to know the most probable phase happens when $\Delta F = 0$, which is the condition of thermodynamic phase coexistence. The Lagrange multiplier then is $\lambda_0 = \Delta S/\Delta n$, or, in thermodynamic notation,

$$T_0 = L/\Delta S(T_0), \tag{41}$$

where L is the latent heat associated with the first-order phase transition and $\Delta S(T_0)$ is the entropy difference at the transition temperature T_0.

All these equivalences are summed up in Table 1.

6 Concluding Remarks

We have shown that, because to every yes/no question we can associate a change in evidence introduced by a new fact, there exist analogous quantities to the free energy difference between phases and the transition temperature, that are closely connected to this change in evidence. When the evidence is strong enough to completely cancel out our initial judgments about the probability of one phase over another and leave us undecided, the "weight" of this evidence is proportional to the transition inverse temperature.

Thus, in this view, the problem of thermodynamic equilibrium between phases is seen as answering the question: *is the system in phase A if we know that its average energy is* $\overline{E}$? in a Bayesian/maximum entropy formalism. The concepts of transition temperature and free energy arise naturally as consequences of this inference framework, and therefore are not intrinsic properties of the systems or the phases.

References

1. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**, 620 (1957)
2. Jaynes, E.T.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge (2003)
3. Pressé, S., Ghosh, K., Lee, J., Dill, K.A.: Principles of maximum entropy and maximum caliber in statistical physics. Rev. Mod. Phys. **85**, 1115 (2013)
4. Uffink, J.: Can the maximum entropy principle be explained as a consistency requirement? Stud. Hist. Philos. Sci. Part B **26**, 223–261 (1995)
5. Dias, P.M.C., Shimony, A.: A critique of Jaynes' maximum entropy principle. Adv. Appl. Math. **2**, 172–211 (1981)
6. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
7. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. **52**, 479–487 (1988)
8. Rényi, A.: In: Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 547–561. University of California Press, Berkeley (1961)
9. Shore, J.E., Johnson, R.W.: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans. Inf. Theory **26**, 26–37 (1980)
10. Skilling, J.: The axioms of maximum entropy. In: Erickson, G.J., Smith, C.R. (eds.) Maximum Entropy and Bayesian methods in Science and Engineering, pp. 173–187. Kluwer Academic Publishers, Dordrecht (1988)
11. Caticha, A., Giffin, A.: Updating probabilities. AIP Conf. Proc. **872**, 31 (2006)
12. Pressé, S., Ghosh, K., Lee, J., Dill, K.A.: Nonadditive entropies yield probability distributions with biases not warranted by the data. Phys. Rev. Lett. **111**, 180604 (2013)
13. Callen, H.: Thermodynamics and an Introduction to Thermostatistics. Wiley, New York (1985)