
Some Misconceptions about Entropy

S.F. Gull, 1989

Abstract

We explore some misconceptions about statistical mechanics that are, unfortunately, still current in undergraduate physics teaching. The power of the Gibbs ensemble is emphasized and explained, and the second law of thermodynamics is proved, following Jaynes (1965). We then study the crucial role of information in irreversible processes and demonstrate by means of a concrete example how time-dependent data enable the *equilibrium* Gibbs ensemble to predict time-varying fluxes during the return to equilibrium.

(Note: See also this [pdf of the original overheads](#) from the lecture, which includes the results of the numerical experiments at the end. Text presented without permission from "Maximum Entropy in Action", ed. Brian Buck and Vincent A. Macaulay (OUP, 1991). My apologies for the ascii-art html-ing of the equations. -- JH.)

7.1 Introduction

This contribution is the direct result of a discussion with second-year undergraduates at Cambridge that took place during a thermodynamics supervision. I asked them what they knew about entropy and about the statistical rationale for the second law. By way of answer they showed me their lecture notes, which reproduced *that awful H-theorem* given on page 39 of Waldram (1985). The main thrust of this chapter is to provide the antidote to that awful H-theorem and to draw attention to the beautiful proof of the second law given by Jaynes (1965). This short paper '*Gibbs vs Boltzmann entropies*' is a true masterpiece which states more clearly than anywhere else the key to the success of Gibbs' maximum entropy method.

The Gibbs entropy of the canonical ensemble is *numerically* equal to the experimental entropy defined by Clausius.

If more physicists knew this simple yet astonishing, fact the subject of thermodynamics (and plasma physics) would be far more advanced than it is today.

Jaynes's later work (1979, 1983) (see also Grandy (1987)) on non-equilibrium statistical mechanics is perhaps even more astounding, for it seems that Gibbs' maximum entropy algorithm is the complete answer for irreversible processes as well. Rather than discuss the general principles here, I illustrate this claim by an example: Brownian motion in one dimension. However, this little example is not nearly as trivial as it appears - there are some very serious lessons which can be learnt from it concerning non-equilibrium statistical mechanics.

Although it will not be mentioned again, it will become obvious that by championing the Gibbs/Jaynes/MaxEnt view of statistical mechanics, I am implicitly rejecting wholesale the approach of the 'Brussels' school (and indeed many other approaches). That implication is true: I do indeed feel that only the MaxEnt viewpoint distinguishes correctly the inferential content of statistical physics and separates it clearly from dynamical aspects. I believe that other approaches contain many misconceptions about the nature of inference in science. My title is, however, slightly ambiguous, and readers will have to

judge for themselves on whose side the misconceptions lie.

7.2 Entropy in thermodynamics and statistical mechanics

The science of thermodynamics came of age when the concept of entropy was defined as a state variable for systems in thermal equilibrium. Although more modern and more erudite definitions exist, for our present needs we can restrict ourselves to the definition of the experimental entropy S_E in the form given in 1850 by Clausius (for a detailed account of the history of thermodynamics see Grandy (1987, Vol. 1, Appendix A))

$$\Delta S_E = \int_{\text{reversible path}} \frac{dQ}{T} \quad (7.1)$$

where T is the absolute temperature and dQ is the amount of heat entering the system. In this way entropy is defined as a function of the macroscopic variables such as pressure and temperature, and its numerical value can be measured experimentally (up to a constant). This constant is provided for us by the third law, that S_E vanishes at the absolute zero of temperature.

Classical thermodynamics is the result of this macroscopic definition: it is conceptually clear and easy to apply in practice (I say this despite the effect it usually has on physics undergraduates!). Conceptual problems have arisen, though, when trying to give a microscopic, statistical interpretation. Statistical thermodynamics began in 1866 with Boltzmann's kinetic theory of gases. He considered a gas of N particles each in a 6-dimensional phase space of position and momentum, and studied how collisions led to an equilibrium distribution. He defined an H function, which we relate to a Boltzmann entropy $S_B = -k_B H$,

$$H = \int d^3x \int d^3p \rho \log \rho, \quad (7.2)$$

where $\rho(x, p, t)$ is the distribution of particles.

A little later, the statistical mechanics of Gibbs was developed. Gibbs focused attention on the $6N$ -dimensional joint phase space of the N particles, and to introduce statistical notions he employed the artifice of the *ensemble*, a large number of copies of the system. These (imagined) copies of the system provided insight into what the actual system might be doing.

Although many of the early results are due to Boltzmann, it was Gibbs who gave us the basic tool of statistical thermodynamics: the Gibbs algorithm. In order to set up the equilibrium ensemble, we maximise the Gibbs entropy

$$S_G = -k_B \int d\tau p_N \log p_N \quad (7.3)$$

under the available constraints (e.g., the ensemble average energy $\langle E \rangle = \int d\tau E p_N$), where p_N is the probability density function (p.d.f.) for the N -particle system. This method is successful to this day. Even the transition to quantum mechanics passed without incident: the quantum mechanical definition involves the density matrix $S_G = -k_B \text{Trace } \rho \log \rho$.

The situation in physics teaching today is still unsatisfactory despite the everyday success of statistical

mechanics with practical problems in physics and chemistry. We use the Gibbs algorithm in any detailed calculation, but teachers often try to justify this using the language of Boltzmann. As a result of this mixture of ideas, there are misconceptions about statistical mechanics. These misconceptions stem from a basic misunderstanding about the role of probability theory in physics, and it is there we must start.

7.3 Inference: the ground rules

I now give, without apology, a modern-day Bayesian viewpoint of the nature of inductive inference.

7.3.1 Bayes' theorem

In its simplest form this elementary theorem relates the probabilities of two events or hypotheses A and B . It states that the joint probability distribution function (p.d.f.) of A and B can be expressed in terms of the marginal and conditional distributions:

$$\text{pr}(A, B) = \text{pr}(A)\text{pr}(B|A) = \text{pr}(B)\text{pr}(A|B). \quad (7.4)$$

Bayes' theorem is merely a rearrangement of this decomposition, which itself follows from the requirement of consistency in the manipulations of probability (Cox 1946). Although anyone can prove this theorem, those who believe it and use it are called Bayesians. Before using it, however the joint p.d.f. has to be assigned. Because Bayes' theorem is simply a rule for manipulating probabilities, it cannot by itself help us to assign them in the first place, and for that we have to look elsewhere.

7.3.2 Maximum entropy

The maximum entropy principle (MaxEnt) is a variational principle for the assignment of probabilities under certain types of constraint called *testable information*. Such constraints refer to the probability distribution directly: e.g., for a discrete p.d.f. $\{p\}$, the ensemble average $\langle r \rangle = \sum_i r_i p_i$ of a quantity r constitutes testable information. MaxEnt states that the probabilities should be assigned by maximizing the entropy

$$S = - \sum_i p_i \log (p_i/m_i) \quad (7.5)$$

under the constraints $\sum_i p_i = 1$ and $\langle r \rangle = r_0$, where $\{m_i\}$ is a suitable measure over the space of possibilities (hypothesis space). The MaxEnt rule can be justified as the only consistent variational principle for the assignment of probability distributions (Shore and Johnson 1980; Gull and Skilling 1984). It can also be justified in numerous other ways (Jaynes 1982). In the simplest case there is no additional information other than normalization: MaxEnt then gives equal probabilities to all possible events, in accordance with Bernoulli's *principle of insufficient reason*. In fact, I believe MaxEnt to be the only logical method we have for the assignment of probabilities--it is so powerful that it is all we need. MaxEnt is, of course, a rule for assigning probabilities once the hypothesis space has been defined: to choose the hypothesis space we again have to look elsewhere.

7.3.3 Inference and statistical mechanics

The real art is to choose an appropriate space of possibilities, and this is our task as physicists. At this level we enumerate the possible states of the system and investigate its dynamics. Indeed, most physicists work entirely at this level, studying dynamics. One could even say that the process of building models for systems constitutes 'real physics', at a philosophical level that we call ontology:

models for reality == ontology.

Statistical mechanics, on the other hand, works almost entirely at the level of inference, where we are concerned with what we know about the state of the system:

knowledge about reality == epistemology.

Seen this way, we realize that the Gibbs ensemble represents the probability that our N -particle system is in a particular microstate. In Gibbs' statistical mechanics we are making inferences about the state of a system, given incomplete information. We know the values of the macroscopic variables, but there are many microstates compatible with this macrostate. We are not assuming that the system actually explores all the states accessible within the constraints, or indeed that it changes state at all. When this is realized, we see that ergodic assumptions are irrelevant to the rationale of statistical mechanics, even if such theorems could be proved. Rather, we set up a probability distribution (ensemble) using MaxEnt and whatever constraints are available, and see what predictions result. If this process leads to experimentally verified predictions, well and good; it then follows that our information that led to this ensemble was sufficiently accurate and detailed enough for our purposes. If our predictions are not verified, we conclude that there must be other, unknown influences which are relevant and which should be sought at the ontological level.

7.4 Gibbs versus Boltzmann entropies

For the present purposes a single specific example will suffice to illustrate the difference between Boltzmann's kinetic theory and Gibbs's statistical mechanics. Suppose we consider a system of N interacting particles in a box of volume V , with a purely classical Hamiltonian

$$H = \sum_{i=1}^N \frac{p_i^2}{2m} + U(x_1, x_2, \dots, x_N) \quad (7.6)$$

The Gibbs entropy S_G is defined in terms of the joint probability distribution p_N of the N particles,

$$S_G = -k_B \int \prod_{i=1}^N dx_i p_N \log p_N \quad (7.7)$$

The Boltzmann distribution function requires a little reinterpretation, but we can make sense of it in terms of the single particle distribution, defined as a marginal distribution over the N particles,

$$p_i = \int \prod_{j=2}^N dx_j p_N \quad (7.8)$$

where the integration is over all particle coordinates except the first. The Boltzmann entropy S_B then becomes

$$S_B = -k_B N \int dx_1 p_1 \log p_1 \quad (7.9)$$

Of these two expressions for the entropy, it should be immediately apparent that only Gibbs's definition is meaningful. No matter how big your system becomes, you always have one system with N particles in it, and not N systems each with one particle! However, the real power of Gibbs's definition lies in the following theorem proved by Jaynes (1965), which deserves to be more widely known. If the initial probability distribution of the system is that of maximum entropy S_G (the canonical ensemble), and the

state variables are then altered over a locus of equilibrium states (reversible paths), then

$$\Delta S_G = \int \frac{dQ}{T} \quad (7.10)$$

whereas

$$\Delta S_B = \int \frac{d\langle K \rangle + p_0 dV}{T} \quad (7.11)$$

where Q is the heat input, K the kinetic energy, T the temperature and $p_0 = Nk_B T / V$, the equivalent pressure of a perfect gas. Hence the Gibbs entropy when maximized (i.e., for the canonical ensemble) can be identified *numerically* with the thermodynamic entropy defined by Clausius. More generally, because S_G is defined for all probability distributions, not just the canonical ensemble, we have

$$S_G \leq S_E \quad (7.12)$$

with equality if and only if the distribution p_N is canonical.

On the other hand, the expression for the change of the Boltzmann entropy shows that it ignores both the internal energy and the effect of the inter-particle forces on the pressure. Because it is defined in terms of the single particle distribution, it is difficult to see how the situation could be otherwise. The Boltzmann entropy is the same as the Clausius entropy only for the case of a perfect gas, when it is equal to the maximised Gibbs entropy as well.

Our moral is simple: the Gibbs entropy is the correct theoretical concept because, when maximized, it is numerically equal to the experimental entropy. The Boltzmann entropy has no theoretical justification and is not equal to the experimental entropy.

7.4.1 That awful H-theorem

Versions of this theorem are found in many undergraduate texts (Lifshitz and Pitaevskii 1981; Waldram 1985, p. 39), purporting to show that the Boltzmann entropy always increases. In the 'quantum' form of the theorem one writes the change of Boltzmann entropy in terms of the microstates α of the 1-particle system as

$$\frac{dS_B}{dt} = -k_B N \sum_{\alpha} \left\{ \log p_1^{\alpha} \frac{dp_1^{\alpha}}{dt} \right\} \quad (7.13)$$

In fact, in the original example quoted to me by the undergraduate at Cambridge, the errors in this theorem were compounded by calling S the Boltzmann-Gibbs entropy!

One is then invited to consider the 1-particle system(s) making 'quantum jumps' between the 1-particle microstates. The master equation and the principle of detailed balance for the transition rates $v_{\alpha\beta}$ then imply

$$\frac{dS_B}{dt} = N k_B \sum_{\alpha\beta} v_{\alpha\beta} (\log p_{\beta} - \log p_{\alpha}) (p_{\beta} - p_{\alpha}) \geq 0 \quad (7.14)$$

What can one say about such a proof? There are several things wrong

1. The use of approximate quantum mechanics, which is not necessarily valid for large perturbations.
2. Worse, it is bad quantum mechanics. An N -particle system has N -particle states. An isolated system will presumably sit in one of its N -particle microstates and make no transitions at all.
3. Even if you could prove such a theorem it would not be useful unless the change of entropy integrated over time were *numerically* equal to the change of experimental entropy. From the discussion of the last section it is clear that this cannot be the case.
4. Last, but not least, there are counter-examples to the theorem! The free expansion of molecular oxygen at 45 atmospheres and 160 K provides such an example, found by Jaynes (1971).

7.4.2 The second law of thermodynamics

The psychological need for an H-theorem is related to another misconception, one that concerns the second law of thermodynamics. For an isolated system the experimental entropy can only increase, that is

$$\Delta S_E \geq 0, \quad (7.15)$$

with equality only if any changes are reversible.

The misconception this time is that, just because the experimental entropy has to increase, the theoretical entropy increases also. In fact, the Gibbs entropy S_G is actually a constant of the motion. This follows from Liouville's theorem for a classical system, or in the quantum case from the fact that the system will remain in an N -particle eigenstate. This dynamical constancy of the Gibbs entropy has sometimes been considered a weakness, but it is not. Remarkably, the constancy of the Gibbs theoretical entropy is exactly what one needs to *prove* the second law.

Once again, we return to the specific case of a gas of N particles, this time confined to one side of a box containing a removable partition. We suppose that the initial state is such that we can describe it using the canonical probability distribution. From our earlier discussion we can then say that the Gibbs entropy S_G is maximized and equal to the experimental entropy S_E .

We now suppose that the partition is opened and the atoms occupy the whole box. We wait until the state variables stop changing, so in that sense the system is in equilibrium and a new experimental entropy S'_E can be defined. Also, all the motions of the gas molecules are Hamiltonian, so that the Gibbs entropy S'_G has not changed: $S'_G = S_G$

The probability distribution of the N particles is no longer the canonical one, however, because of the (very subtle!) correlations it contains reflecting the fact that the molecules were originally on one side of the partition. This means that the Gibbs entropy S'_G is now in general less than the maximum attainable for the new values of the state variables, which is in turn equal to the new experimental entropy. So

$$S_E = S_G = S'_G \leq S'_E \quad (7.16)$$

This shows the fundamental result $S_E \leq S'_E$ and displays the second law of thermodynamics as a law concerning *experimental* quantities.

7.4.3 The theoretical second law

Another very important way of understanding the second law is to see it as a statement about phase volumes. Boltzmann's gravestone is engraved with the famous formula $S = k_B \log W$. The W in this formula is the number of microstates compatible with the macroscopic state. This epitaph was placed there by Planck and it is ironic that this (correct) formula leads at once to Gibbs' definition of the entropy rather than Boltzmann's own.

Imagine, as in Fig. 7.1, the set of microstates compatible with the initial macroscopic state variables (P_1 , T_1 , etc.). This phase volume describes our ability to reproduce the initial conditions: our system will be in a microstate somewhere inside this volume, but we do not know where. As the system evolves, the state variables change and finally reach new values (P_2 , T_2 , etc.). Our system has evolved dynamically and is now located somewhere inside the phase volume consistent with these new values. This simple picture reveals a fundamental requirement for a process to be reproducible, as follows.

The phase volume compatible with the final state cannot be less than the phase volume compatible with the initial state.

If the final phase volume were smaller there would be certain initial microstates that evolved to thermodynamic states other than (P_2 , T_2); i.e., the thermodynamic process would not be reproducible. We have also the condition for reversibility:

A process is reversible if the final phase volume is the same as the initial phase volume.

If the final phase volume were larger, then there would necessarily be some states in it that did not arise from states compatible with the initial state variables, and hence the process could not be reversible.

This completes our second, *theoretical* statement of the second law in terms of phase volumes.

7.5 Non-equilibrium phenomena

7.5.1 Time asymmetry in physics

There has been much debate about the nature of irreversibility and the 'arrow of time'. What has not been generally recognized is that temporal asymmetry enters because our knowledge of a system is not time-symmetric, and not because of any asymmetry inherent in its dynamics. It should be stressed that I am not claiming that all physical laws must necessarily be time-symmetric, but merely that the ones we know of and need to consider here happen to have this property. Nevertheless, it is certainly true that our knowledge of the state of a system is not symmetric; we usually have more knowledge of its past state than of its future. This asymmetry in our knowledge is then properly reflected in asymmetry of our inferences about its likely behaviour. Once again, the problem is one of epistemology, not ontology.

Another related question concerns the Gibbs algorithm. It is recognized as a fine way of setting up an equilibrium ensemble, but how must it be modified to cope with disequilibria? The astonishing answer to this is also the simplest: the Gibbs algorithm is already complete; just give the formalism some time-dependent information and it will predict how the system is likely to behave and approach equilibrium.

7.5.2 Brownian motion

Rather than discuss generalities, which can be found elsewhere (Jaynes 1983; Grandy 1987; Garrett (Chapter 6 of this volume)), I can best illustrate the claims made above by a case-study, namely Brownian motion. Suppose a particle moves in one dimension, having position $x(t)$, and experiences random

collisions from molecules at temperature T . We have previously considered a viewpoint that regards the microstate as a single point in an initial phase space, which thereafter moves deterministically as dictated by the Hamiltonian. For the present purposes we will abandon this Hamiltonian view and adopt a different approach which is able to cope with the outside influences. We consider instead the phase space to be described classically by $\text{pr}[x(t)]$. Our knowledge of the particle's position is now encoded by this much larger joint probability distribution, which has a dimension for each moment of time. Our (incomplete) knowledge of the dynamics of the particle has to enter via constraints on the joint p.d.f.. As in path integral methods, we restrict our attention to the positions $x_n = x(t_n)$ at a set of regularly spaced times $t_n = n\tau$. We define an average velocity $v_n = (x_{n+1} - x_n)/\tau$ and acceleration $a_n = (x_{n+1} - 2x_n + x_{n-1})/\tau^2$. These definitions provide linear operators in x -space corresponding to velocity and acceleration. The slightly asymmetric definition of velocity is of no consequence in what follows, since the results are identical for the alternative definition $(x_n - x_{n-1})/\tau$.

We now use the Gibbs algorithm to set up an equilibrium ensemble by maximizing the entropy

$$S(\text{pr}(\mathbf{x})) = - \int \text{pr}(\mathbf{x}) \log \text{pr}(\mathbf{x}) \, d\mathbf{x}. \quad (7.17)$$

In this definition we have dropped the dimensional factor of k_B and assumed a uniform measure over x -space.

We now introduce constraints suitable for the Brownian motion problem. Because the system is in equilibrium at temperature T we have, for all times t_n

$$\int d\mathbf{x} \, \text{pr}(\mathbf{x}) \, v_n^2 = \langle v_n^2 \rangle = k_B T / m \quad (7.18)$$

It is in fact only necessary, and certainly more convenient, to introduce a much weaker, single constraint, namely

$$\frac{1}{N_\tau} \sum_n \langle v_n^2 \rangle = k_B T / m \quad (7.19)$$

where N_τ is the number of time intervals considered.

We now introduce some knowledge of the dynamics. The colliding molecules can only provide a certain average impulse P to the particle in our time interval τ , so suppose in a similar way that

$$\frac{1}{N_\tau} \sum_n \langle a_n^2 \rangle = (P/m\tau)^2 \quad (7.20)$$

This specification of the average momentum transfer P certainly lies in the realm of dynamics, not inference. We suppose in what follows that it is sufficient to specify only one (P, τ) pair to describe all the reproducible features of Brownian motion. This may or may not be the case: only subsequent observation could tell us. If necessary, more information could be added in the form of further constraints. For example, the average impulse for time intervals of 2τ could be specified as well, in order to incorporate further details of the collision process.

I add one further constraint for the convenience of my computer program, which has the effect of

confining the particle on average to a box of size L :

$$\frac{1}{N_\tau} \int_{-n}^n \langle x_n^2 \rangle = L^2 \quad (7.21)$$

Maximizing the entropy under these constraints yields a joint p.d.f. that represents the equilibrium ensemble,

$$\text{pr}(\mathbf{x}) = \frac{1}{Z(\alpha, \beta, \gamma)} \exp(-\mathbf{x}^T (\alpha \mathbf{I} + \beta \mathbf{v}^T \mathbf{v} + \gamma \mathbf{a}^T \mathbf{a}) \mathbf{x} / 2) \quad (7.22)$$

In the above, $\mathbf{v}$ and $\mathbf{a}$ are the velocity and acceleration operators (matrices) implicitly defined earlier, $\mathbf{I}$ is the identity matrix and α , β and γ are Lagrange multipliers. These multipliers control the physical variables of position, velocity and acceleration respectively: for example, β is the inverse temperature and γ provides the particle with inertia. The partition function

$$Z(\alpha, \beta, \gamma) = \int d\mathbf{x} \exp(-\mathbf{x}^T (\alpha \mathbf{I} + \beta \mathbf{v}^T \mathbf{v} + \gamma \mathbf{a}^T \mathbf{a}) \mathbf{x} / 2) \quad (7.23)$$

provides the normalizing integral and can be evaluated using a z-transform. The multipliers α , β and γ can be found by the usual partition function manipulations.

It is interesting to note that this joint p.d.f. for $\mathbf{x}$ is of the form $\exp(-\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x} / 2)$, and can be recognized as a zero-mean, correlated, multivariate Gaussian time-series of a type well studied in digital signal processing: it is in fact an auto-regressive process of order 2. The covariance matrix of the time-series is given by $\langle \delta \mathbf{x} \delta \mathbf{x}^T \rangle = \mathbf{R}$.

However, the equilibrium ensemble stands ready to deliver all sorts of time-dependent predictions: we just have to give it more data. Suppose, for example, that we know the position of the particle at various times: $x(t_1) = x_1 \pm \delta_1$ and $x(t_2) = x_2 \pm \delta_2$. We now employ our formalism to manipulate the probability distribution, given these data D :

$$\text{pr}(\mathbf{x}, D) = \text{pr}(D) \text{pr}(\mathbf{x}|D) = \text{pr}(\mathbf{x}) \text{pr}(D|\mathbf{x}). \quad (7.24)$$

In the above, $\text{pr}(\mathbf{x}|D)$ is the answer we want, $\text{pr}(\mathbf{x})$ is the equilibrium ensemble and $\text{pr}(D|\mathbf{x})$ is the likelihood of the given data,

$$\text{pr}(D|\mathbf{x}) \sim \exp\left\{-\sum_{i=1,2} (x_i - X_i)^2 / 2\delta_i^2\right\} \quad (7.25)$$

For the present purposes, $\text{pr}(D)$ is an irrelevant normalizing constant. Bayes' theorem then gives us the answer $\text{pr}(\mathbf{x}|D)$, showing the evolution of $\langle x(t) \rangle$ and $\langle (\delta x)^2 \rangle(t)$ forwards and backwards in time.

In Fig. 7.2a (see [lecture notes pdf](#)), we have provided the information $x(84\tau) = 10 \pm 0.01$ to an ensemble with $\alpha \rightarrow 0$ and $\gamma \rightarrow 0$ (a system with no inertia). We know where the particle is at a certain time, so the display, which gives the average position and pointwise marginal uncertainty as a function of time, shows uncertainties which increase proportional to $t^{1/2}$ away from that time, in the manner of a random walk. The ensemble average position shows no net flux yet, because we only have this one piece of information. In particular, we do not yet know the velocity, although do we know that its magnitude is likely to have a value corresponding to thermal equilibrium.

In Fig. 7.2b we have added the extra information $x(42\tau) = 0 \pm 0.01$ to this same ensemble. The symmetrical random-walk behaviour persists for $t < 42\tau$ and $t > 84\tau$ but between these times, an average flux appears. The particle does, after all, have to go from $x = 0$ to $x = 10$ during this time interval and, in the absence of inertia, it is predicted to have travelled at constant velocity, though notice that our uncertainty about its position increases in the middle of this interval.

In our third example (Fig. 7.2c), we show the approach to equilibrium of a particle with inertia ($\gamma > 0$) projected with known velocity at $t = 0$. The average velocity decreases exponentially—indeed it satisfies a Langevin equation whilst the uncertainty increases. This time the retrodictions are not plotted; the reader is invited to ponder what they look like, and what, if anything, they represent.

In these examples we have displayed the overall marginal uncertainty in the particle positions. The posterior p.d.f. is, however, highly correlated and there is important, additional information contained in these correlations. A very good way of visualizing this information is to plot typical samples from the posterior p.d.f. as well as the ensemble average.

7.6 Uncertainty versus fluctuations

Suppose we use the Gibbs algorithm to set up an equilibrium ensemble, and calculate the ensemble average of a quantity of interest f , together with its variance $(\Delta f)^2 = \langle (f - \langle f \rangle)^2 \rangle$. Now Δf certainly represents our uncertainty about the quantity f but, according to most expositions of statistical mechanics, it is also supposed to indicate the level of temporal fluctuations of f . Here again, then, is a misconception—the fact that we are uncertain about the value of a quantity does not by itself mean that it must be fluctuating! Of course, it might be fluctuating and if that were the case, it would be a very good reason to be uncertain about its value. Without further analysis, however, we simply do not know whether it actually fluctuates. We have at last found a question in statistical mechanics where ergodic considerations are important. We can sketch a partial answer to this problem following Jaynes (1979).

We define

$$\bar{f} = \frac{1}{T} \int_0^T f(t) dt \quad (7.26)$$

as a long-term time average and

$$(\delta f)^2 = \frac{1}{T} \int_0^T (f(t) - \bar{f})^2 dt \quad (7.27)$$

as a long-term variance. Taking ensemble averages, we do indeed find that $\langle f \rangle = \langle \bar{f} \rangle$; however

$$\langle (\delta f)^2 \rangle = (\Delta f)^2 + \langle \bar{f}^2 \rangle - \langle \bar{f} \rangle^2 \quad (7.28)$$

and this second term is not necessarily zero.

The situation is as follows: if a time average is taken over too short a time interval, then the observed variation in f can of course be *less* than the Δf of the equilibrium ensemble. However, the long-term variation of f can actually be greater than Δf , depending on a particular property of the p.d.f. of the ensemble. Even then, although we can calculate $\langle \bar{f} \rangle$ and $\langle (\delta f)^2 \rangle$ as above, we still do not know that these estimates are reliable; to do that we have to examine higher-order correlations of the ensemble. The details are again in Jaynes (1979).

The moral is that the Gibbs algorithm gives the uncertainty of our predictions, not the observed temporal fluctuation. To say that a thermodynamic quantity actually fluctuates (which, of course, it may well do) requires further, decidedly non-trivial, analysis.

7.7 Conclusion

Most misconceptions about entropy result from misunderstandings about the role of probability theory and inference in physics. The epistemological content of statistical mechanics must be recognized and clearly separated from the ontological, dynamical aspects. Having done this, Gibbs' maximum entropy algorithm attains its full power, predicting the temporal evolution of *our state of knowledge* of a physical system.

Acknowledgements

I am extremely grateful to Geoff Daniell for helping me realize that I was a Bayesian and thereby introducing me to the work of Ed Jaynes.

References

- Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, **14**, 1-13.
- Grandy, W. T. (1987). *Foundations of statistical mechanics*, Vols 1 and 2. Reidel, Dordrecht.
- Gull, S. F. and Skilling, J. (1984). The maximum entropy method in image processing. *IEE Proceedings*, **131F**, 646-59.
- Jaynes, E. T. (1965). Gibbs vs Boltzmann entropies. *American Journal of Physics*, **33**, 391-8.
- Jaynes, E. T. (1971). Violation of Boltzmann's H theorem in real gases. *Physical Review A*, **4**, 747-50.
- Jaynes, E. T. (1979). Where do we stand on maximum entropy? In *The maximum entropy formalism* (ed. R. D. Levine and M. Tribus), pp. 15-118. MIT Press, Cambridge, Massachusetts.
- Jaynes, E. T. (1982). On the rationale of maximum entropy methods. *Proceedings of the IEEE*, **70**, 939-52.
- Jaynes, E. T. (1983). *Papers on probability, statistics and statistical physics*, Synthese Library, Vol. 158 (ed. R. D. Rosenkrantz). Reidel, Dordrecht.
- Lifshitz, E. M. and Pitaevskii, L. P. (1981). *Physical kinetics*. Pergamon, Oxford.
- Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, **IT-26**, 26-37.
- Waldram, J. R. (1985). *The theory of thermodynamics*. Cambridge University Press.
-