FOREWORD OF THE REVISED REPRINT (1962)

In this reprint, several misprints have been corrected and a host of side-remarks and references have been eliminated. The reader is, however, asked to substitute by himself, all through, "Gibbs' distribution" for "Maxwell-Boltzmann distribution".

Most of the technical contents of this paper have been superseded by the following articles by the same author:

"The role of sufficiency and of estimation in thermodynamics", The Annals of Mathematical Statistics, Vol. 33, September 1962.

"Derivation of statistical thermodynamics from purely phenomenological principles", IBM Research Report NC-106.

"A critical Note on Information Theory and statistical mechanics", IBM Research Report NC-107.

However, none of these new publications touches the methodological problems which are discussed in the present older work.

The present paper had been summarized in a short Note in French, in the Comptes Rendus de l'Académie des Sciences, Vol. 243, 1956, pp 1835-1838.

The Parts II and III, announced in the original of this paper, have not been written as such. A summary of Part III was published as a short Note in French, in the Comptes Rendus de l'Académie des Sciences, Vol. 249, 1959, pp 1464-1466.

AN OUTLINE OF A PURELY PHENOMENOLOGICAL THEORY OF STATISTICAL THERMODYNAMICS : CANONICAL ENSEMBLES

by

Benoit MANDELBROT
Faculté des Sciences de l'Université de Genève
Genève, Suisse

Summary. "Boltzmann's problem" of statistical thermodynamics is that of eliminating the paradoxical incompatibility of structure, existing between the irreversibility of classical phenomenological thermodynamics, and the reversibility of kinetic models.

One finds that, in order to construct kinetic "analogs" to the laws of phenomenological thermodynamics, the dynamics of large assemblies of molecules (Liouville theorem, etc. ...) must be completed by some hypotheses of randomness. Once established, this randomness can be followed up in its development with no new conceptual paradox,

but the introduction of randomness still raises entirely uncleared problems. Since, therefore, the kinetic foundations of thermodynamics are not sufficient in the absence of further hypotheses of randomness, are they still quite necessary in the presence of such hypotheses? Or else, could not one "short-circuit" the atoms, by centering upon randomness introduced by the process of observation? Our aim is to show (partly after Szilard) that substantial results, usually obtained through kinetic arguments, could be obtained by postulating from the outset a statistical distribution for the properties of a system, and following up with a purely phenomenological argument. The spirit of the theory is extremely close to that of the Copenhagen approach to quantum theory,

Randomness is introduced by following the modern statistical theory of the estimation of non directly observable intensive variables of state, such as the temperature. The discussion of the methodological foundations of modern statistics can thus be translated into a full-fledged, and possibly significant, counterpart of the discussion of the kinetic foundations of thermodynamics. Statistics is thus provided with a particularly concrete example for some of its more involved methods; thermodynamics appears clarified in its classical aspects.

1. Introduction

1.1. The nature of the problem.

The aim of this paper differs in one essential respect from that of most investigations on statistical problems.

This difference should be stated at the outset, in order to avoid certain misunderstandings. Most authors in statistics are concerned with broadly engineering problems, and they

assume that a sufficiently complete understanding and description of the necessary laws of nature has already been acquired elsewhere.

Our problem is precisely the opposite: we wish to improve the knowledge and presentation of physics,

Any degree of success, we could achieve, would be new proof of the fact, which is of course quite familiar, that well-chosen engineering problems often bring out the essentials of a physical situation, in a way that is useful in a far wider context. Such used to be the role of heat engines in thermodynamics. Later on, one tried to use the problems of coding for the same purpose : the starting point of those attempts was the misleadingly simple-looking problem, raised by the fact that the definition of Shannon's information [6] is mathematically identical to a classical definition of entropy by Boltzmann. The study of this problem was brought together with that of Maxwell's Demon. But

none of the now numerous attempts, to clarify the relationship of information to entropy, is generally felt to have brought much to either communication or thermodynamics.

However, the more general problem, of the role of observation and of the observer in thermodynamics, could now be studied in detail with the help of the statistical theory of estimation, which has become a methodological model of inductive behaviour in the face of the unknown.[7]

It will be attempted to build a theory of thermodynamics, which will be statistical as well as phenomenological, around the problem of the statistical estimation of state variables; the problem of entropy and information will constitute an application of the theory. (A previous attempt by the author in chap 4 of [3] is now quite obsolete, but the philosophy of models given in other chap. of that reference stands fortified by the present work).

Szilard's [8] previous approach is of the greatest relevance for the problem.

1.2. The nature of the result

One aim of communication theory is to find ways and means satisfying certain criteria of quality, by which a signal could be detected through a background of noise. It has been recognized for some time that this problem is, in principle, simply one of estimating "at best" the emitted signal, S_e, knowing the received signal S_r. For that, S_r is considered to be an observation from a random (because perturbed by noise) population of signals, and S_e is the parameter of this population.

The distribution $p(S_r/S_e)$ is considered to be a part of the engineer's scientific knowledge of nature, that is : at worst, it must be determined by special observation; at best,

it is given by physical laws of general validity: those of thermodynamics and of quantum theory. We shall show that, conversely, the conventional laws of thermostatistics, can be obtained by caracterizing thermal noise as being the "least disturbing for the physicist." A certain amount of imprecision in what is the "least disturbing" will be shown to be allowed by a corresponding familiar imprecision of thermodynamics. The main criterion involved, that of "sufficiency of certain valuations of observation", is not even a variational criterion : it postulates the impossibility of a certain inference, and is an authentic counterpart of the exclusion of certain heat engines by the Carnot principle. Anyway, the arbitrary anthropo-centered character, remaining in the theory, will be a counterpart of the arbitrary assumptions about molecules, required in the kinetic models which aim to explain why noise is the least disturbing for the observer which is considered.

Further methodological discussion of the approach will be made in § 4, when the systems sutdied are defined (§2), and the Maxwell Botzmann distribution is derived in several ways (§3). Final discussion will be included in §6.

2. States of a statistical physical system

2.1 Restriction to one-parameter systems

In any truly stochastic physical theory, the relationships between the various state variables of a system are assumed, from the outset, to be ruled by probability laws. The degree of complication of a theory is then determined by the most complicated family of probability distributions considered in that theory. We shall start by a restricted problem, in which the only families to appear will depend upon a single real parameter.

2.2 Definitions

Consider a set of methods of measurement, defining a certain level of refinement of the physical analysis. Physical variables can then be of several kinds, distinguished as being, on one side, extensive or intensive; on the other side, as being observable or estimable.

Observable variables are those which can be considered as random variables (r.v.)before measurement (and expressed by capital letters, such as U); and can be actually and directly measured by real numbers (expressed by lower case letters, such as u). The precision of results of measurement will always be taken as infinite, which means that we shall consider that, by measuring an observable, one actually and physically puts the system in a "state" described by the value which has been found. The measurement may of necessity be infinitely slow. Of course, even then, the results of measurement are usually given only within a margin of possible error, relative to a "truer" value; the latter would be a r.v., depending upon the system itself, whereas the error would be a r.v. depending upon the process of observation. But in fact, there is no criterion for sharing the contributions of the two sources of randomness; at best, the "true" value could be considered as an estimable.

Consider now a system before measurement. Its "state" is a probability distribution function (p.d.f.); it can be considered as a "mixture" or "superposition" of "states" after measurement. If necessary, the state will be called "mixed", as opposed to "pure" states after measurement. This is a relative concept, since a state which is pure with respect to one observable may also be "mixed" with respect to another. A pure and a mixed state are complementary uncompatible descriptions of a system. Measurement thus involves an unpredictable sudden jump from state to state, or rather from being partly in several states into being in one: this has nothing shocking in detection theory.

Estimable variables are those parameters which express the dependence of the probability distributions of the observables, upon the properties of the physical system studied. One must therefore consider populations of systems having the same value for some estimable. When such an equality can be considered as realised by a physical interaction, the estimable is called "intensive". However, much of the theory holds for any estimables.

The obvious example of an observable is the energy U; the corresponding estimable variable is the inverse temperature B (the "observable" character of the energy is an universal axiom of physics, although sometimes hard to justify). A measure of energy is, for example, obtained with a so-called "thermometer", by observing the change in its geometrical shape when it absorbs U. On the other side, the uniformity of the B's of a set of systems can be ascertained by letting them a sufficiently long time in thermal contact.

Let the probability distribution and density of U be given by $P(u|B)$ and $p(u|B)$:

$$P(u|B) = \Pr\left[U \leq u \,|\, B\right] \;;\; p(u|B) = dP/dB$$

Both kinds of variables have been defined within a one-to-one transformation, only. Sometimes, there exists a particularly intrinsic determination of an observable, which is additive, i.e. such that the u of any union of disjoint sets is the sum of the separate u_i. The $p(u|B)$ is then given, knowing the $p_i(u_i|B)$, by "convolution", iterating the integral

$$\int_0^u p_1\,(u_1|B)\; p_2\,(u - u_1|B)\; du_1 \,.$$

B is still an estimable variable of the sum; it is also an estimable variable of any subdivision of the system, if possible.

2.3 The estimation of estimable variables .

Consider now a system of known energy u. Let the prior density $p(u|B)$ be positive for all positive u, as is usual. Since the energy is known, the very definition of estimable variables, as parameters, fails. In fact, a single system can be considered as being in "thermal equilibrium" with any thermostat (i.e. an infinite set of other systems in thermal equilibrium with each other), since the addition of a single system, having a possible value for u, does not perturb the probability distribution of an infinite set.

From the viewpoint of prediction of future events happening to the system, the population from which it is drawn should have, intuitiveley, little importance.

Strictly, to know B is therefore a problem of "retro-diction". But one may also wish to predict a prior distribution, for the energy of <u>other</u> systems of the set from which the first was drawn, and, for that, some reasonable guessing of the initial B would be useful. This $\hat{B}$, although unknown, is not a random variable. Except under very rare conditions, where one has a prior distribution for it (the "Bayes case"), there is no limit-of-frequency sense, to be attributed to the feeling that "some values of B are more 'probable' than others". However, there is a very clear meaning to statements of probability of making an error by guessing a certain value for B. One can then try to construct estimators, or estimating intervals, so as to minimize certain chosen probabilities of error. Mathematical statistics is a technique for doing so, but it cannot ever justify any chosen criterion of good guessing.

Any estimator is a single-valued function of u; therefore, by applying it to parts of a whole of uniform temperature, one obtains a classical probability distribution of estimated "local temperatures": the parameter of this distribution is the true B : therefore estimates of B are <u>not</u> intensive variables. Many estimators are consistent, i.e. the distribution of the local temperatures gets more and more concentrated around the true value, when the size of the samples used increases. This makes it possible to measure the temperature of an infinite set with any precision. But for any sample size, the estimation of B is an essentially irreversible procedure;

Under these conditions, the replacement of the knowledge of u, and of $p(u|B)$, by a single estimator $\hat{B}$ [or by upper and lower bounds for B] is an operation entirely different from the measure of the random variable U. In the Bayes case, however, it is very close to the operation of prevision of the most probable or average future evolution of a system ruled by probability laws. The predicted evolution is of course different from the actual one. The point has lead to great discussion in quantum theory [see von Neumann[13], §V.1.]. But a simpler case of the difference between spontaneous random evolution, and noisy estimation is given in information theory by a comparison of Shannon's definitions of information of a Markovian message and of a noise-perturbed one. Both are specified by matrices such that line sums are one. But the ideas are entirely different,

In view of this irreversibility of estimation, we shall delay the consideration of specific methods to §5, and first consider the case where there exist intermediate steps of the estimation, which are such that if one takes them, one loses "no information", in appropriate senses of "information" (more general than Shannon's). Our definition of reversibility will fully determine the Maxwell-Boltzmann distribution, that is : it will characterize thermodynamics.

3. The reversible part of the estimation of B. Derivations of the Maxwell Boltzmann distribution.

3.1 First derivation, based upon the concept of sufficiency.

<u>Criterion of sufficiency.</u> Suppose that the extensive observable u can be chosen to be <u>additive</u> : this is indeed a very strong hypothesis, since it excludes any interactions between neighbouring systems, and a fortiori quantum interactions between distant systems. Let us take several sample systems from a population believed of uniform B. Let u_i be their energies, and $u = \sum_i u_i$. One could estimate B, through $\hat{B}$, either starting from u or from all the u_i. Independence of the result with respect to shape or disposition of the systems, implies that $\hat{B}$ should be a symmetric function of the u_i. If moreover any $p_i(u_i|B)$ can be considered as the distribution of a sum of still smaller variables, one can estimate $\hat{B}$ from finer and finer subdivisions of the system. But B is considered as a "<u>macroscopic</u>" intensive variable, <u>which</u> roughly <u>implied</u> that from a certain level down, no improvement of the estimation can be obtained by measuring finer subdivisions,

Postulate this independence of the estimate, relative to the knowledge of individual u_i, to be strict, and no more asymptotic. This means that, for example :

$$(S) \quad Pr(u_1 | u, B) = Pr(u_1 | u)$$

independently of B. Nothing could then be drawn from the observation of the sample distribution of the energy among the different parts ; no Maxwell's Demon could beat the macroscopic observer in the estimation of B, by measuring the partition of energy among molecules. In R.A. Fisher's[14] statistical terminology, u is then said to be a sufficient statistic for the estimation of the original B.

It will turn out that sufficiency is implicit in the usual thermodynamics.

To sum up, the principle of sufficiency can be considered as established by all the experiments leading to the belief in the possibility of macroscopic description: It is thus established only below a certain level. On the contrary, the principle of additivity is established only above a certain level. We postulate that both hold exactly in a certain strip of sizes.

Maxwell Boltzmann (M.B.) distribution. Under certain regularity conditions, the only distribution for which u is a sufficient statistic for B is the M.B. probability density :

$$p(u|B) = G^{-1}(B)\, S(u) \exp(-Bu)$$

The detailed statistics can be thus derived from a single overall principle, a qualitative one, in the sense that it answers to a "yes or no" alternative: this cannot fail to recall Carnot's principle;

the scale of B is fixed by $p(u|B)$; that of u is fixed by additivity.

One can prove from sufficiency that the best estimate of any function of B is that function of the best estimate of B ; in the case of sufficiency, estimation and functional transformation are commutable.

The above result was proved independently, in 1936, by G. Darmois[16], B. O. Koopman[17] and E.J.G. Pitman[18]; before them, it was implicitly proved by L.Szilard[8] in 1925. Under slightly weaker conditions of regularity, the distribution need not have a density, and may be of the form :

$$dP(u/B) = G^{-1}(B)\, dV(u) \exp(-Bu)$$

S(u) and V(u) are the "structure function" and the "integral structure function" of the system. The "structure generating function" G is, because of the requirement that $\int p\,du = 1$, the Laplace transform

$$G(B) = \int S(u) \exp(-Bu)\, du = \int \exp(-Bu)\, dV(u)$$

$$EU = E(U/B) = \int_{-\infty}^{+\infty} u\, p(u/B)\, du = -\frac{d \log G(B)}{dB}$$

$$DU = D(U/B) = \int_{-\infty}^{\infty} (u - EU)^2 p(u/B)\, du = \frac{d^2 \log G(B)}{dB^2}$$

Note that G(B) may be defined only for $B \geq B_d > 0$, and diverge for $B < B_d$. It is impossible in ordinary thermodynamics of matter, that $B_d > 0$; but, clearly, the present theory is much more general than is required by matter, and there are exceptional applications where a positive abscissa of convergence of the Laplace transform is the main feature; see [1,2,3,4,5]

$\log G(B)/B$ is identified to the "free energy" of a system . It is seen that the free energy cannot be any function of B : its exponential must have a positive inverse Laplace transform (it is also called "completely monotone"; the signs of successive derivatives alternate). This fact has, surprisingly, never been mentioned, to our knowledge, in papers on statistical thermodynamics : it is clearly because for very large systems, it is irrelevant, cf. §6, so that there is no need to justify it in large scale thermodynamics.

Distribution of a sum of MB systems. Consider the sum of two systems following MB distributions with respectively structure and generating functions S_1, S_2, G_1, G_2 and same B. The distribution of the energy of the sum of these systems will be

$$p(u/B) = \int_0^u S_1(u_1) G_1^{-1}(B) \exp(-Bu_1) S_2(u-u_1) G_2^{-1}(B) \exp[-B(u-u_1)] \, du_1$$
$$= \left[\int_0^u S_1(u_1) S_2(u-u_1) \, du_1\right] [G_1(B) G_2(B)]^{-1} \exp(-Bu)$$

Thus, the distribution of the sum of MB systems of same B is still MB : in the addition, structure functions transform by convolution, generating functions simply multiply, and $\log G$ (B) functions simply add.

The same results clearly hold for finite sums of systems, and also for denumerable sums.

3.2. Szilard's derivation of the M.B.distribution

It turns out that the above derivation of the M.B. distribution could have been an inverted paraphrase of a derivation due to Szilard[8],

often quoted (see von Neumann[13]) but seldom analysed. Similar considerations can be found in a paper by G.N. Lewis[19].
Both authors wished to show how fluctuation phenomena can be introduced into classical thermodynamics, without destroying its structures and spirit. "The second principle loses nothing in rigour because of fluctuations, and in no way becomes an approximate principle : it melts into a higher harmony containing the laws of fluctuation."

Szilard considers systems in random evolution, i.e. such that their properties at one time determine only probabilities at a later time; in particular, energy exchanges between systems in contact are only random. He assumes that the probability distribution of a system at an instant of time depends on one parameter only, the temperature, that systems which have long been in "thermal" contact and have exchanged energy have equal temperatures and may be considered to be in thermal equilibrium; finally, that systems in thermal equilibrium at one time remain in thermal equilibrium. The existence of a single number T, characterizing thermal equilibrium, is the "Zeroth principle" of thermodynamics[21].

(The fact that energy does not vary in the interactions, is the First Principle). Temperature is not however assumed to determine a single energy, but only some superposition of states of different energy;

this distribution is realized among systems having long been in long thermal contact with a "thermostat".

The infinity of possible energy distributions will now be reduced, and the fact that $p(u|T)$ cannot be modified without compensation

will be translated into postulates about the more detailed nature of thermal equilibrium, which are a priori only "reasonable", but a posteriori experimentally correct, in the sense that they lead to the M.B. distribution.

Let two systems, of energies u_1 and u_2, initially in contact with a thermostat, be separated and brought in very long thermal contact with each other. Their initial energies U_1 and U_2 were independent random variables, of same parameter T. Szilard assumes that the nature of thermal equilibrium is such, that the energies U'_1 and U'_2, after long contact, are random variables having distributions independent of the initial temperature and of initial values u_1 and u_2, and are conditioned only by the constant total energy u.

Similarly, Lewis considers that, when a quantity is shared between two systems, the ratio of specific probabilities of any couple of partitions depends only upon the nature of the two systems, and in no way upon their method of connection, or upon the existence, nature or mode of connection of other systems. He then postulates that one should expect that the probabilities of various partitions of a quantity U between two systems depend only upon the total u, and in no way upon the reservoir with which the two parts are in connection. Thus, one would expect the same partition, whether the two systems are in very imperfect contact with one reservoir, or whether they are in contact with another reservoir, of very much higher temperature, but have the same total energy through a very rare fluctuation.

It is seen, with great pleasure, that both authors have rediscovered the principle of sufficiency, exactly in the form (S) of § 3.1, but have interpreted it as a property of equilibrium, not of observation. As already mentioned, Szilard has even anticipated the derivation of the M.B. formula, Through a discussion of fluctuation sizes for different systems, Szilard shows that T is a universal temperature;

Principle of Onsager-Casimir. In the theory of the irreversible decay of deviations from equilibrium[22], one has to assume that the same laws apply to "small but macroscopic" deviations, and to fluctuations. This amounts to postulating that the way a deviation was reached is irrelevant, and only its amplitude imports. This markovian hypothesis[23], a probabilistic form of Huygens' s principle,

leads to Onsager' s relations of reciprocity. In the theory of equilibrium, one makes a weaker hypothesis that independence from the past is attained after a long time only. Thus, Onsager' s theory is a quite proper "interpolation" of the present theory, with a full markovian hypothesis. It is curious to note that a weakened probabilistic Huygens' s principle leads to the same results as Carnot' s principle.

3.3. Second derivation, based upon the concept of efficiency.

Criterion of efficiency. It was mentioned in §3.1, that, since the estimate of B is a non-random function of u, it is a random variable, when referred to the ensemble of constant B, from which the system is drawn. Assume that the distribution of the u is known, but not necessarily M.B., and compare different possible methods of estimation, all assumed to be unbiased, that is such that the mean of the estimator is equal to the true value. For that, compare their variances, i.e. mean square deviations from the mean.

It can be shown that, whichever the distribution $p(u|B)$, the variance of an estimate of a function $f(B)$ is necessarily bounded below by the expression :

$$D[f(B)] = \frac{\left(\frac{df}{dB}\right)^2}{F}, \text{ where } F = E\left(\frac{d \log p(u/B)}{dB}\right)^2 = E\left(\frac{d^2 \log p(u/B)}{dB^2}\right).$$

F is called "Fisher' s information". This limit means not only that one does not know how to perform a more precise estimation, but that under certain conditions of regularity, one could not conceive of any such estimation.

Maxwell-Boltzmann' s distribution. Usually, there are still closer lower limits to variance, but Fisher' s limit can be exceptionally attained.

Under certain conditions of regularity, the lower limit to variance can be attained only with the M.B. distribution. Thus the overall statistics can also be derived from a variational principle.

Then, Fisher' s information takes the especially simple form

$$F = \frac{d^2}{dB^2} \log G(B) \qquad = D(E)$$

Uncertainty relationship. Let $f(B) = -\partial \log G/\partial$ Then

$$\boxed{DU.D\{(B) = 1}$$

The less well-known is the energy of a system, when its true temperature is known, the "larger" the system [in terms of its log $G(B)$] and the better the temperature can be estimated And conversely.

Or else, disregard the true temperature. Any estimation gives B together with DB ; identify the estimator of B with the true B, but only for the purpose of estimating the DU of U before the measurement. Then, the larger DU, the smaller DB.

One cannot fail to note the formal identity of this with Heisenberg's relationship, or its equivalent in communication: Gabor's relation "Df. Dt = 1/2" in the spectral analysis of signals. In fact, all three result from equality cases of Schwartz's inequality for dual variables. It does not matter whether the duality is Fourier, (quantum theory and spectral analysis) or Laplace (here). But, in Heisenberg's relationship, the observer can choose which variable he wants to know with greater precision; here, B is known exactly only for the infinite thermostat, and U exactly only when it is certainly zero.

Proper scale. It would be pleasing to have a new scale of B, such that DB be independent of the estimated B, that is, from the measured u. (This seems to be what MacKay calls the "proper scale"). Clearly :

$$f_p(B) = \int^B \sqrt{\frac{d^2 \log G(B)}{dB^2}}\, dB$$

Note that $f(B) = -d \log G(B)/dB = E(U|B)$. In other terms, the estimation variance of the "true" value EU, is equal to the fluctation of u around its mean value. This is not obvious, but is a theorem which should be brought together with the fact, mentioned above, that when a sufficient statistic exists, the estimate of a function of B is the function of the estimate. The fact, that even when u is assumed strictly known, there is a sense to be attributed to the fluctuation of something very close to U, is very comforting, in view of the paradox in considering the measure as infinitely precise. But even this new noise has nothing to do with "hidden variables".

3.4 Other derivations of the MB distribution

The concept of sufficiency has many other aspects; sufficient valuations of observations preserve "information" in many different senses of the word, Fisher's, and also Shannon's and in a generalisation of both these senses due to Schutzenberger.

4. Discussion on the methodology of §3 : Two complementary types of statistical thermodynamics.

Let us interupt here the construction of the theory, to comment upon what is being done.

We have succeeded (partly after Szilard) in deriving, from purely phenomenological criteria, some results of statistical thermodynamics, a science considered to be so deeply related to kinetic and such models, as to be also called statistical mechanics. The M.B. distribution is no longer characterized as being the one which would be steady under collision phenomena, but as having certain good properties under observation.

How does this fit into the classical scheme?

4.1 The two classical methods of thermodynamics.

At least since Clausius, one recognizes two methods of mathematical structuration for thermodynamics: the phenomenological, also called macroscopic, pure, classical, axiomatic, etc., and the kinetic, also called microscopic, statistical, etc... The latter contains more results than the former, in particular it includes the statistics, and is also the older of the two (the Greeks, Gassendi, the Bernoullis), and closer to intuition. Despite this, it is considered as conceptually subordinate to the other, since it is required to derive the principles of the other as theorems, whereas the reciprocal is usually not considered.

How well is this explanation achieved? From the viewpoint of rigor, notoriously poorly: There is in fact a complete paradox in any kinetic explanation: the fundamental incompatibility of structure between the irreversibility of some principles of phenomenological thermodynamics, and the mechanical reversibility of any purely kinetic model of these principles, (e.g. the paradoxes of Loschmidt and of Zermelo[25]). These objections gradually forced Boltzmann to add to the mechanical assumptions. Following Uhlenbeck, let us call the problem of reconciling the two viewpoints, Boltzmann's problem. The ideal would have been to derive the macroscopic results from microscopic assumptions; but in fact one[27] is more properly looking for logical structures analogous to thermodynamics, by adding to the kinetic models some rather arbitrary randomness assumptions, often introduced through the necessary inprecision of "coarse observation": Any initial unevenness of probability distribution, although strictly preserved, because of Liouville's theorem, is supposed to dissipate itself into thinner and thinner streamlines; so that any coarse density becomes uniform, after an average time which increases when the coarseness decreases.

A very old system tends towards the situation, in which there is a sharp discontinuity of phenomena near infinitely sharp definition. (This recalls the singular phenomena observed when the viscosity of a fluid tends to zero, which also relate to irreversibility and dissipation). Once established, this randomness can be followed in its development, with no new conceptual paradox, although with great technical difficulty,

But the introduction of randomness still raises entirely uncleared problems.

4.2 Statistical thermodynamics without hidden variables.

Since, therefore, the kinetic foundations of thermodynamics are not sufficient, without a further hypothesis of randomness, are they still necessary in the presence of such a hypothesis? After all, although statistical thermodynamics owes its origin and development to the theory of atomism, it need not always be so. In fact, partly after Szilard[8], we are in the process of showing that one can "short-circuit" the atoms by centering upon any element of randomness, for example introduced through necessarily unprecise observation, and we are deriving a substantial part of the fundamental results, usually obtained through kinetic arguments, by following up the introduction of randomness in a purely phenomenological way.

The fact would have been of course more striking, before the times when atomic theory ceased to be a doubtful conjecture. It is also a pity, that Szilard's original paper was not more often quoted in contemporrary (around 1925) discussions about the "causal interpretation" of quantum theory through "hidden variables". In fact, a rapid decline of interest in the phenomenological and "energetist" approach to thermodynamics was contemporary of the Copenhagen approach to quantum theory.

There is however a renewed interest now in causal reinterpretations. Take for example von Neumann's[13] proof of the impossibility of introducing hidden variables, into the Copenhagen approach to quantum theory, This proof is part of a reduction of a large part of the quantum rules to a set of phenomenological and axiomatic rules for observation, starting from a purely stochastic viewpoint very similar to that used here. The incompatibility between a quantum structure and hidden variables, may therefore be compared to the incompatibility between thermodynamics and kinetic theory

. Any attempt to go around those paradoxes, such as current work of Bohm[30], de Broglie[31] and Vigier[32], can be compared to the attempts to "solve" Boltzmann's problem. This current work does not actually attempt to build a theory of which the quantum theory would be the "thermodynamics", but simply to disprove the impossibility of building such a theory, by introducing randomness not fundamentally, but through a chaos hypothesis. It may be that the possibility of a twin set of quantum theories would now appear less shocking in view of the existence of a statistical thermodynamics without hidden variables, apparently just as "closed" as the conventional quantum theory. However, even if there some pleasure in the great similarity of the words used to describe the two situations, there is no question of mathematical identity.

This methodological discussion will be continued through §6 and in the conclusion

5. The irreversible and final part of the estimation of B.

5.1 First method of estimation: maximum likelihood B and Boltzmann's most likely state. Degeneracy.

To derive the M.B. distribution [in §3] we did not need to specify the actual arbitrary method of estimation,

The function linking B to u must now be specified. In principle, rules of estimation should be derived from the properties desired from the estimate. In fact, however, the usual rules of statistics were introduced for no conscious reasons, except simplicity; or they were first motivated by considerations of "degree of reasonable belief" quite foreign to classical probability, and only later justified by their properties[7]. The estimation theory of thermostatistics is only implicit, though quite real; it will turn out, curiously enough, that it has used exactly the same procedures as statistics.

Start by R.A. Fisher's[14] theory of maximum likelihood estimation. To find a measure of rational belief in a value of B, when we are reasoning from the sample to the population, Fisher inverts the function $p(u|B)$, taking now u as a parameter, and B as a variable. Of course, $p(u|B)$ ceases then to be a probability distribution, and B is no random variable. For example $\int p\,dB \neq 1$;

therefore one cannot speak of the likelihood of the set of all values of B, but only compare likelihoods. Taking now the M.B. distribution as having been derived in §3, maximize

$$\log p(u|B) = -\log G(B) + \log S(u) - Bu$$

This requires that

$$u = -d \log G(B)/dB.$$

B will be obtained by inverting this implicit equation.

One cannot fail to note the identity of this result with a classical formula by Boltzmann . Let us review the proof of that result. To define the temperature of a system of given u, one takes that only EU = u is known, then one derives the "most likely" distribution of this energy between the available (discrete) states. For that, one maximizes the logarithm of the likelihood, given EU = u, and E1 = 1. The derivation, using Lagrange multipliers, is too classical to repeat. The reason for saying "likelihood" instead of probability, will appear soon. This gives the distribution

$$p(\text{any state of energy } u | B) = e^{-Bu}/G(B)$$

as being the most likely (M.L.) This seems to differ from the M.B. distribution, but in fact one introduces the further assumption of "degeneracy": that there may be $S(u)$ different states of same energy. Then ,the most likely distribution of energy turns out to be the MB pdf; without further arbitrariness, one gets a relation between B and EU :

$$EU = -d \log G(B)/dB.$$

Then one replaces herein E(U) by the known u, and inverts to get B. Altogether, Boltzmann's argument amounts to an implicit and improvised theory of estimation, anticipating Fisher's theory. Note that the assumption of degeneracy would not have changed the result of taking the M.L. value of B, in Fisher's approach,

since the log $S(u)$ term drops out in the maximation of log p, relative to B.

Therefore, taking degeneracy for granted, our only technical progress comes from the fact that maximum likelihood estimates have been rather more carefully studied than most likely states. It is quite true that the M.B. distribution is now obtained without the arbitrariness of the "maximum" specification (which needs no amplification in thermostatistics, see Darwin and Fowler and Khinchin[33]) but the same arbitrariness is found immediately at the next step.

p. 200 is omitted.

§6. The arbitrariness in estimation vanishes asymptotically.

6.1. Passage to the limit of very large systems, i.e. very large log G(B).

Since there is no motivation a priori, in classical probability theory, for the Boltzmann - Fisher estimates of "rational Belief", these estimates should be judged by the properties they turn out to have. In fact, they have extremely satisfactory asymptotic properties.

In thermostatistics, they converge to the Darwin Fowler estimates; when the size of the system tends to infinity.

Therefore, since the physicist is not interested in small bodies, the Boltzmann procedure is justified by its simplicity, and by its convergence to theories considered as correct.

In statistics, they behave well for large samples. Chiefly, they are consistent (they converge to the true value, with probability approaching unity, as the size tends to infinity); they are normally distributed around the true value, and their variance is the smallest possible (efficiency). Given the fundamental arbitrariness of estimation criteria, one could not say, that any other procedure is more correct than this one; so, whenever it is the simplest, (for example when a sufficient statistic exists) one can use this procedure for large samples.

6.2. Small systems

As their main drawback, from the viewpoint of present day problems of statistics, maximum likelihood estimates lack reasonable small sample properties:

The statistician cannot any more justify the definition of an estimation procedure by its asymptotic properties.

However, the physicist is not troubled by this, and continues to take the best from the facilities, permitted by the "fortunate insensibility of thermodynamic functions", stressed by Lorentz.

6.3 Possibility of constructing observation models of thermostatistics.

This fortunate insensibility is the very key to the possibility of constructing models in physics, based upon conditions of optimality of observation, Naturally,

the criterion of an engineering problem can be changed at will, without its ceasing to be an engineering problem. But the state of nature, most favourable for one problem, may not be so for any other, Hence, being most favourable is not in general a way of describing a real physical situation. One may say that each time one describes (and can fully characterize) outside circumstances, as being the most favourable for somebody, one builds a new anthropocentered physics.

To invert methodologically an engineering problem is to identify this physics with an actual one. But how to defend the model so obtained, if the criterion of the engineering problem is irreducibly arbitrary?

The whole point is that arbitrariness exists anyway in thermostatistics, and that is exactly the same than in statistics. This is in no way surprising, in view of the way in which probabilities were assigned to hypothetical ensembles in thermostatistics: to "agree with our partial knowledge" (a fact linking irreversibility not to incompleteness or inexactness of mechanics, but to incompleteness of specification). Criteria of equipartition are only "reasonable" criteria, never considered as anything but mathematical abstractions of reality, or in other words as "fascinating" mathematical tricks, but still tricks; and they may become very dangerous conceptually when "asymptotically slight" differences go undetected altogether, and entirely different concepts get exchanged because they lead to the same numerical results. Entropy is the worst sinner in that respect,

It is a pity, of course, that the phenomenological and statistical approach can only be franker about doubtful points, can help circumscribe and better understand them, but cannot help eliminate them. Statistics may be sorry for it too, because it makes hopeless any belief in an improvement of the choice among statistical criteria, on the basis of their appropriateness for physics.

Conclusion

Thermostatistics has followed a development parallel to mathematical statistics. Recall how late (Khinchin[33]) the parallel development of probability theory and of termostatistics have converged; statistics and thermostatistics have been even slower,

In the comparison, statistics appears as by far the more advanced discipline. Thus mathematical and methodological help turns out to go the "wrong" way, up the scale of sciences of Auguste Comte, since the modern small-sample statistics was made necessary in biology, agriculture, social science and industrial acceptance (where asymptotics are of no help, and the process of observation of so great importance that te there is no difficulty in accepting theories centered around the observer). Statistics became necessary for the physicists only when the problem of noise reduction had to be faced: this is the historical reason for the context of this reasarch.

It is in a sense very discouraging to have found on this occasion, that the wish for conceptual coherence of the theoretical physicist had been weaker than practical needs

The moral is that the foundations of thermostatistics should be reviewed critically, after each important progress in the mathematics or methodology of other statistical disciplines.

Acknowledgements. The author wishes to thank the Rockefeller Foundation and the Centre National de la Recherche Scientifique (Paris), for research grants while this paper was written

References

1 B. Mandelbrot : Simple games of startegy, occuring in communication through natural languages; Trans. I. R. E. Inf. Theory 3 (1954) 124.

2 B. Mandelbrot : Statistical structure of language; Communication theory; Academic Press (1953) 486.

3 B. Mandelbrot : Jeux de Communication; Publ. Inst. Stat. Paris 2 (1953) 3.

4 B. Mandelbrot : Structure formelle des textes; Word (New York) 10 (1954) 1.

5 B. Mandelbrot : Thermostatistics of Willis systems; Information Theory; Academic Press, (1956).

6 C. E. Shannon : Mathematical theory of communication; Bell S. T. J. (1948)

7 J. Neyman : Lectures and conferences on mathematical statistics and probability. U.S. Department of Agriculture (1952), chap. 4

8 Leo Szilard : Uber die ausdehnung der phänomenologischen thermodynamik auf der schwankungserscheinungen; Z. Physik 32 (1925) 753

9 J. L. Lawson, G. E. Uhlenbeck : Threshold signals; Mac Graw Hill (1950)

13 J. von Neumann : Mathematical principles of quantum mechanics, Princeton (1954)

14 R. A. Fisher : Contributions to mathematical statistics, Wiley (1950)

16 G. Darmois : XXII Session de l'Inst. Int. Stat. Athènes (1936)

17 B. O. Koopman : Trans. Am. Math. Soc. 39 (1936) 399

18 E. J. G. Pitman : Proc. Camb. Philo. Soc. 32 (1936) 567

19 G. N. Lewis : J. Am Chem Soc. 53 (1931) 25

21 E. A. Guggenheim : Thermodynamics; North Holland (1949)

22 de Groot : Thermodynamics of irreversible processes; North Holland

23 L. Onsager, S. Machlup : Phys. Rev. 91 (1953) 1505

25 P. T. Ehrenfest; Encyclopedie Sc. Math. IV 1, 6 (1915) 188

27 J. W. Gibbs : Statistical Mechanics : Yale (1903)

30 D. Bohm : Phys. Rev. 85 (1952) 166

31 L. de Broglie : Comptes-rendus, since 1952

32 J. L. Vigier : Interprétation causale de la mécanique quantique; Gauthier Villars (1956)

33 A. I. Khinchin : Statistical Mechanics; Dover (1949)

35 D. Blackwell, M. A. Girshick : Theory of Games and Statistical Decisions, Wiley (1954)