The Concept of Probability in Statistical Physics, Y. M. Guttmann, Cambridge University Press, 1999.

Unlike in quantum mechanics, the use of probabilistic reasoning in classical physics is not based on the existence of indeterminacy or objective lawlessness. In the very foundation of classical physics we find the assumption that, given the precise state of the world in one instance, the laws of physics determine its future states completely. Then how are we to interpret statements from statistical physics whose abstract form is "The probability of *A* is *p*."? Do probabilistic statements form an integral part of the description of the physical world? Do they merely reflect our ignorance with respect to the precise state of the world? Can they be deduced from non-probabilistic statements? Can they be understood in terms of frequencies?

(Maxwell's demon) Maxwell, using the demon argument, concluded that the truth of the second law is a statistical and not a mathematical truth, for it depends on the fact that the bodies we deal with consists of millions of molecules and that we never can get a hold of single molecules.

Quantum statistical mechanics contains two types of probabilities that cannot be given a unified presentation. The first type of probabilities is related to the Schrödinger wave function. Probabilities of the second type are defined exactly the same way as in the classical theory.

1. The Neo-Laplacian Approach to Statistical Mechanics

In 1859 J. C. Maxwell applied the theory of errors to the problem of finding "the average number of particles whose velocities lie between given limits, after a great number of collisions among a great number of particles" in his paper, "Illustrations of the Theory of Gases." [1] His conclusion was that the velocities are distributed among the particles according to the same law as the errors are distributed and that if the collisions are very frequent, then the law of distribution of molecular velocities will be reestablished in an appreciably short time. This conclusion shows why "Illustrations" is considered to be the beginning of modern statistical mechanics.

The discussions on the theory of errors started more than 100 years earlier and it was Laplace who systematized the theory and showed how to apply it in various areas. He regarded the theory of probability in general, and the theory of observational errors in particular, as complementary theories to mechanics. Mechanics, for Laplace, provided an objective description of the way bodies move, while the theory of errors is only a heuristic that we use to derive useful predictions even when observations contain errors. Contrary to this classical view, Maxwell's application of the theory of errors had a novel character. With his application, probabilities became part and parcel of a science whose aim was to provide an objective description of the properties of systems in motion.

Maxwell arrived at his derivation after reading an article by Sir John Herschel in which he reviewed the theories of Quetelet.[2] What is interesting in the review is that the deviation from some mark can be regarded as errors, but they can also be regarded as the outcomes of a process that

---

[1] *Phil. Mag.* 4 (19) (1860). Reprinted in *The Scientific Papers of James Clerk Maxwell*, ed., W. D. Niven, Dover, 1963. p.380.

[2] See Gillespie in *Scientific Change*, ed., A. C. Crombie, Basic Books, 1963.

contains a stochastic element. Herschel's presentation indicates that the theory of errors may be applied even in cases where the notion of an error does not seem to arise at all.

Maxwell concluded that the mean velocity of the particles is $2\alpha/\sqrt{p}$ and the mean square velocity is $3\alpha^2/2$. But he was not satisfied with this derivation. In the derivation the assumption that the knowledge of v_x should not affect the probablities of v_y and v_z was used. Hence in a second article ("The Dynamic Theory of Gases") in 1866 he attempted a second derivation of the velocity distribution.

Maxwell attempted to derive the equilibrium distribution from the following two assumptions:

1. The velocities of the colliding molecules are uncorrelated. Hence, if $P(v_1, v_2)$ is the probability that a pair of colliding molecules will have, respectively, the velocities v_1 and v_2, then
$P(v_1, v_2) = P(v_1)P(v_2)$,
Where $P(v_1)$, $P(v_2)$ are the probabilities of choosing at random a molecule with the respective velocities.
2. Let v_1, v_2 be the velocities of a pair of colliding molecules before the collision and v'_1, v'_2 their velocities after the collision. Maxwell's second assumption was that the probability that a colliding pair will undergo the velocity change $v_1, v_2 \rightarrow v'_1, v'_2$ because of the collision will be the same as the probability of the velocity change $v'_1, v'_2 \rightarrow v_1, v_2$. That is, the "reverse" motion is as probable as the original motion.[3]

From these assumptions he derived the equation

$$P(v_1) = (N/\alpha^3 p^{3/2})e^{-(v2/\alpha 2)}.$$

What is the status of Maxwell's two assumptions? Behind the assumption is an idea that the positions and velocities can be assumed to be randomized. Moreover it is assumed that the molecules will remain randomized. But the question remains: Why are we justified in making these assumptions?

The next step in the development of the kinetic theory of gases was Boltzmann's 1868 paper.[4] It is important to mention that Boltzmann's derivation started from the same assumptions that Maxwell made. In his later writings, he became aware of the problematic status of the assumptions to which he introduced the term *Stosszahlansatz*, or the assumption of molecular chaos. Throughout his life Boltzmann tried to find a purely mechanical justification to the Stosszahlansatz.

In the same paper of 1868, Boltzmann derived the equilibrium velocity distribution with the assumption that the gas molecules had a fixed amount of energy and looked for an expression for the number of different ways of dividing this amount between the different molecules. The old assumptions on the nature of the collisions were no longer necessary. They were replaced by a single purely probabilistic assumption that all of the different possibilities were equally probable. This

---

[3] See Maxwell's "On the Dynamic Theory of Gases," *Phil. Trans.* 157(49) (1867).

[4] "Studien ueber das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten," *Sitzungberiche, K Akademie der Wissenschaften, Wien, Mathematisch-Naturwissenschaftliche Klasse* 58(1868), pp. 517-560.

assumption was sufficient for a mathematical proof that the most probable distribution of energy corresponds to the Maxwell-Boltzmann distribution.

Even after this combinatorial derivation, Boltzmann kept looking for a more mechanical version of the kinetic theory of gases with the atomistic models. He defended the use of the molecular models on the basis of the productive ideas to which it gave rise. For this form of atomism Boltzmann was criticized by a group of physicists. They believed that science should not depend on special speculative assumptions or on fanciful constructions of models. Another group criticized his derivation of irreversibility. (Loschmidt, Zermelo, and Poincaré)

Gibbs reminds us that, when we calculate the value of any variable that depends on the position and momenta of the particles of a system with many degrees of freedom, there will be a certain interval of values $[A - a, A + a]$ such that it is vastly probable that the system will assume values in this interval. Very often the most probable value is the mean value. Gibbs emphasized the abstract character of this point and presented it as an outcome of the theory of errors. He employed the theory of errors to simplify the computations of the values of various dynamic parameters. The simplification process consisted of a demonstration that, although in reality a variable may have infinitely many values, we may behave as if the values were identical to the mean value. Gibbs defined the notion of an "ensamble"(a notion introduced by Boltzmann, called "Ergode") for the more precise setting. An *ensemble* is a collection of systems of a similar kind that differ from one another in the particular values which their parameters assume. The "relative proportion" of systems belonging to the ensemble whose parameters lie within the various intervals determines a density function. A normalized density function is a probability function. It assigns numerical probabilities to the event that a randomly chosen system will assume a particular value that is located within certain intervals for a given parameter. Maxwell calculated the velocity distribution for systems with a fixed number of particles and a fixed energy. Boltzmann proved that the same type of distribution is obtained in the presence of forces. Gibbs thought two additional assumptions. He argued that (1) no real system is completely isolated, and hence we cannot expect systems to have a completely fixed energy; and (2) in many contexts the assumption that the number of particles is fixed can be limiting. His goal was to obtain the equilibrium distribution even when these considerations were taken into account. Gibbs's first step was to consider an ensemble of systems whose energy is allowed to vary. The probability that the system will have the energy level E was taken to be proportional to e^{-bE}, where $b = 1/kT$. Such an ensemble was called a *canonical ensemble*.

Gibbs proved that, as the number of particles grows to infinity, the canonical distribution coincides with the Maxwell-Boltzmann distribution. (He introduced the term *microcanonical distribution* for this distribution.) Then he considered an ensemble of systems with a varying number of particles. The number of systems having N particles was assumed to be proportional to e^{-mN} (m is the chemical potential). Such ensembles are known as *grand canonical ensembles*. Again, the grand canonical distribution coincides with the microcanonical one.

If you want to prove Gibbs's results rigorously, then the central limit theorem is necessary. Using the central limit theorem, Khinchin demonstrated that the details of the initial distribution of the energy between the particles need not be taken into account.

Laplace's view on probabilities was motivated by his determinism. The view that the state of the world at one instance determines its state at any other moment does not leave any room for randomness, stochastic processes, or any other objective statistical concept. Many modern thinkers begin with Laplace's conclusion, that is, they maintain that probabilities are subjective attitudes made because of ignorance and are devoid of any objective meaning. One of these thinkers is E. T. Jaynes, who developed a subjective view of statistical mechanics in his article entitled "Information Theory and Statistical Mechanics."[5] He attempted to show that the predictive capacity of statistical mechanics is largely independent of the deterministic character of its dynamics. To be sure, Tolman realized long before Jaynes that the rivalry between the ergodic approach and his own had a broad methodological background. Ter Haar was even firmer on this point when he insisted that statistical mechanics should be regarded as an instrument for correct predictions. Jaynes, however, was the first to regard the a priori probabilities that Tolman introduced into statistical mechanics as subjective probabilities, and he was also the first to call the general prediction process that ter Haar described a statistical inference.[6]

(The predicative approach to statistical mechanics)

Tolman says:

> We are now ready to begin our consideration of the statistical methods that can be employed when we need to treat the behavior of a system concerning which condition we have some knowledge but enough for a complete specification of the precise state. For this purpose we shall wish to consider the average behavior of a collection of systems of the same structure as the one of actual interest but distributed over a range of possible states. Using the original terminology of Gibbs we may speak of such a collection as an ensemble of systems.[7]

Each member of the ensemble is represented as a phase point. Such a representation encodes for each particle three coordinates of position and its three coordinates of momentum. Tolman says further:

> It is evident that the condition of an ensemble at any time can be regarded as appropriately specified by the density r with which representative points are distributed over the phase space.[8]

When we normalize r we obtain a probability measure over the phase space. This probability function is to be interpreted as assigning a "relative weight" to those members of the ensemble that lay in a given infinitesmal region of the phase space at a given time. As time progresses, each

---

5 See E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, ed., R. D. Rosenkrantz, Reidel, pp.4-39.

6 See R. C. Tolman, *The Principle of Statistical Mechanics*, Clarendon, 1938. See also ter Haar, "Foundation of Statistical Mechanics," *Rev. Mod. Phys*. 27 (3) (1955)

7 Tolman, *ibid*. p.41.

8 Tolman, *ibid*. p.46.

member of the ensemble follows its deterministic route and traces a trajectory in the phase space. The Gibbsian point of view concentrates only on the way the density r changes in time, that is, on the quantity $\partial r/\partial t$. This quantity summarizes all of the relevant dynamic factors that we need to describe how the ensemble average changes in time.

There are two approaches that aspire to replace Gibbs's pragmatist approach to statistical mechanics – the ergodic approach and the predictive approach. The former is a traditional approach that attempts to give dynamic explanation to the puzzles of statistical mechanics, and the latter is a revisionist approach, which provides explanations that depend only on very general dynamic considerations. The main tenet of the predictive approach is that the large number of degrees of freedom is responsible for equilibrium phenomena. Because the change in number does not seem to bring any new physical considerations, it is strange that similar systems that differ only in the number of their degrees of freedom have to be distinguished by physics. Hence many physicists felt the need to present the predictive approach in a broader methodological light. Jaynes chose subjectivism as his general methodology. He was generally interested only in the question of prediction, not of dynamic states. Therefore, when the probabilities are interpreted as subjective conditional probabilities, the shift from the objective dynamic questions concerning the behavior of the system to issues concerning what an agent may predict about the system can acquire a clearer motivation. This shift can be seen as an instance of the general subjectivist tendency to recast scientific concepts in epistemic terms. Once the subjectivist framework is introduced, the predictive approach seems more natural because the large number of degrees of freedom certainly affects our ability to make predictions.

According to Jaynes, the deterministic assumptions do not play an essential role for the purposes of prediction. This type of elimination program was advocated explicitly by de Finetti, who believed that questions about determinism should be divorce from those concerning predictions.[9]

2. Subjectivism and the Ergodic Approach

One of the consequences of the second law of thermodynamics is the irreversible trend toward a maximally disordered state. From the statistical-atomistic approach, this irreversible trend is rather puzzling. The mechanical motion of the atoms is presumed reversible; therefore, where does the irreversibility come from? It seems that there is an essential tension between the reversible dynamics of the molecules and the irreversible dynamics of the second law of thermodynamics. Boltzmann solved the puzzle when he argued that the irreversible trend toward disorder, while being the most

---

[9] de Finetti, "Probabilism," *Erkenntnis* 31 (1989), pp. 169-223.

probable course of events, is not without exceptions. And Perrin showed that the Brownian motion provides an experimental confirmation of it, namely, the possibility of visible fluctuations from the state of maximal disorder.[10] One of the consequences of the modern view of atomism was a reformulation of the aims and methods of statistical mechanics. According to the modern view, the excellent agreement between the predictions and the results of experiments is not in itself sufficient justification of the methods used in statistical mechanics. A satisfactory foundation of statistical mechanics has been sought whereby its methods may be seen to be extensions of exact mechanics classical or quantal. Physicists began to attempt to demonstrate that statistical mechanics actually follows from purely mechanical principles. In other words, the issue of reduction was introduced into the discussion.

(1) The conceptual foundations of the statistical approach to mechanics by P. and T. Ehrenfest

The modern conception of statistical mechanics received its first definitive formulation in an article written by Paul and Tatiana Ehrenfest in 1912.[11] The book remains one of the most influential contributions to the literature on statistical mechanics. The first part contains a discussion of the main contribution to statistical mechanics before the controversies of the 1880s and the 1890s. In the second part they discuss the more sophisticated statistical formulations that ark the modern treatment of the subject. Finally, in the third part they discuss the writings of Gibbs critically. *The Conceptual Foundations* is remarkable in the painstaking attention that is paid to methodological issues, the lengthy discussions in contains on purely mathematical points, and the deep and informed treatment of various statistical topics. The Ehrenfest proposed a list of questions that needed answers to them.

(Questions)

1. Some physicists, for example Einstein, regarded physical probabilities as observables. This view requires a new physical definition of probability. Einstein tried to solve the problem by defining probabilities as time averages.
2. The second problem was the justification of the principle that equal areas in the phase space should be regarded as equally probable. (the justification of the indifference principle)
3. Stosszahlansatz is the third problem. Even more crucial is the question of the consistency of the Stosszahlansatz with the underlying mechanical laws.
4. The fourth problem concerns the consequences of Poincaré's recurrence theorem or the "Wiederkehreinwand," as Zermelo referred to it. The theorem states that in the infinite long run it overwhelmingly probable that mechanical systems will return arbitrarily close to their initial states. It seems to contradict the second law of thermodynamics.
5. Loschmidt's "Umkehreinwand" is the fifth problem. He argued that, for every mechanical path where the entropy of the system increases, there exists a precisely opposite path where the entropy has to decrease. This is an immediate consequence of the reversibility of the equations of motion with respect to the direction of time. This fact seems to contradict the second law of thermodynamics.

---

10 See Brush, *The Kind of Motion We Call Heat*, North Holland, 1976.

11 See P. and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics*, Cornell University Press, 1959.

6. The sixth problem is a generalization of the forth and the fifth ones. It concerns the question of whether it is possible to deduce thermodynamic laws from purely mechanical principles or to show that there is no inconsistency between the two theories. One may generalize the problem even further and ask whether all of the laws that express macroscopic regularities can be reduced to purely mechanical laws.
7. The seventh problem is so-called the ergodic hypothesis.
8. Final problem is the justification of the astounding empirical success of Gibbs's methods.

(An introduction to ergodic theory)

A state of an n-particle system M depends on $6n$ parameters. Each of the particles must be assigned three position and three momentum coordinates. When the parameters $p_1,\ldots, p_{3n}, q_1,\ldots, q_{3n}$ are assigned, the state of the system is fixed. Hence it is possible to represent each state as a point in a $6n$-dimensional space that is isomorphic to a subspace of the Euclidean E^{6n}. This space is the phase space of M. One of the way of describing the evolution of M is by stating its Hamiltonian $H(p_1,\ldots, p_{3n}, q_1,\ldots, q_{3n})$ using the equations

$$\partial H/\partial p_i = dq_i/dt, \quad \partial H/\partial q_i = -dp_i/dt.$$

The Hamiltonian can be described as a transformation from E^{6n} to itself, which determines for every state its dynamic evolution after an infinitesmal time interval. Every state lies on a unique trajectory that is determined by its Hamiltonian. And because the solutions are required to be reversible in time, every state has a unique past trajectory (backward motion) as well. Every possible state of M lies on exactly one such trajectory. The exclusiveness is the result of the uniqueness of the solutions of H, and the exhaustiveness is the result of the existence of solutions for H.

One of the results concerning Hamiltonian systems is Liouville's theorem. Let A_i be a set of states, and let $A_{t+\Delta t}$ be the set of the future evolutions of the members of A_t after Δt. Liouville's theorem says that the volumes of A_t and $A_{t+\Delta t}$ are the same. This property of Hamiltonian systems is known as stationary or incompressibility.

Lets abstract from Hamiltonian only its trajectory that it assigns to every state. First, we look at the Hamiltonian as a topological group of operations $\{U_t\}_{t\in R}$; for every $t\in R$, U_t: $E^{6n}\rightarrow E^{6n}$ assigns to every state $s\in E^{6n}$ its future at t. The next step is to replace the continuous structure of $\{U_t\}_{t\in R}$ with a discrete $\{U_t\}_{t\in Z}$ (Z is the set of integers). We may do so by defining a transformation T: $E^{6n}\rightarrow E^{6n}$ such that, for all $s\in E^{6n}$, $T(s)$ is the future of s after a short interval of time; we then identify $\{U_t\}_{t\in Z}$ with the group of iteration T^i.

Let $s\in E^{6n}$ be an arbitrary state. The trajectory of s under T is $O_{T(s)} = \{s, T(s), T^2(s),\ldots\}$. Let A be an arbitrary set of states. The time average that M stays in A given its present state s is $O_T(s, A) = \lim_{n\to\infty} 1/n\sum_{i=1} X_A(T^i(s))$. ($X_A$ is the characteristic function of A.) $O_T(s, A)$ can be thought of as the frequency in which $O_T(s)$ assumes states that are located in A.

Let A_E be the set of states that are compatible with macroscopic equilibrium. We may ask the following questions:

1. Given that M is in the state s now, how often will M assume states that are compatible with macroscopic equilibrium, or what is $O_T(s, A_E)$?
2. Under which circumstances will $O_T(s, A_E)$ be close to 1?
3. Under which circumstances will $O_T(s, A_E)$ be independent of the choice of s?

Recall that: When M is made of many particles, the volume of A_E is overwhelmingly large. In this case we are well on our way to justifying our expectations that M will reach a persisting equilibrium independent of its initial conditions. Ergodic systems are those for which the above-mentioned conditions obtain assumption.

In ergodic theory, the basic mathematical structure is a quadruple $<S, B, m, T>$, which is called a dynamic system. S is a set of possible states of M, B is a σ-algebra of subsets of S, $m: B\rightarrow R$ is a measure, and $T: S\rightarrow S$ is the evolution transformation.

Definition. T is ergodic iff, for all $A\in B$, $T^{-1}(A) = A$ implies that $m(A) = 0$ or $m(A) = 1$.

Theorem 1 (Birkhoff 1931) If T is ergodic, then for all $f\in L^2$, and for m-almost every $s\in S$,

$$lim_{\ n\rightarrow\infty} 1/n \sum f(T^{-1}(s)) = f^*,$$

where $f^*\in L^2$. (L^2 is the space of the square-integrable functions. # A random variable $f:\Omega\rightarrow R$ is square-integrable if $\int |f|^2 \, dP<\infty$.)

Corollary 2. If T is ergodic, for every $A\in B$ and for m-almost every $s\in S$,

$$lim_{\ n\rightarrow\infty} 1/n \sum X_A(T^{-i}(s)) = m(A).$$

If T is ergodic, the time averages $O_T(s, A)$ are the same as the measure of A for every $A\in B$ and for almost every $s\in S$. In particular, if $m(A_E) = 1$ and T is ergodic, $O_T(s, A) = 1$ almost always.

Theorem 2. If T is ergodic, for every $A, C\in B$,

$$lim_{\ n\rightarrow\infty} 1/n \sum f(T^{-1}(C)\cap A) = m(A)m(C).$$

Corollary 1.

$$1/n \sum X_A(T^i(S))X_C \ \rightarrow\ m(A)X_C \text{ for almost every } s\in S.$$

Theorem 2 states that the ergodicity of T implies that, when we calculate $O_T(s, C)$ using only s that are members of A, the result will not be biased, on the average, because the C are distributed among the A in the same proportion that they are distributed among the general population. This is precisely why we can discard additional information when we predict the future state of an ergodic system. In particular, this explains why any extra knowledge concerning M will not affect our predictions concerning whether and how often M will assume an equilibrium state.

One example of a stochastic process is a coin that is tossed repeatedly. Such a process can be described as a dynamic system in the following way:

1. As the space of states S we shall take all of the infinite sequences of heads and tails. Each sequence $\ldots w_{-1}, w_0, w_1, \ldots$ is an exhaustive description of one of the possible outcomes of the infinite coin-tossing process.
2. We shall take the σ-algebra generated by the cylindrical sets. Cylindrical sets define sets of sequences by stipulating that their *i*th coordinate should have a particular value. For example, "the set of all sequences that have heads in their 103rd place" is a cylindrical set.
3. As T we may take the shift transformation. The shift is defined as $T(\ldots w_{-1}, w_0, w_1, \ldots) = \ldots w_0, w_1, w_2, \ldots$ The shift assigns to every sequence a different sequence that has the same elements ordered in the same way but parametrized differently. T can be thought of as the "unfolding in time" of a particular sequence. The shift reveals the identity of the result of the next toss at each moment.
4. Let $A \in B$. For example, take
   $A_{103} = \{\ldots w_{-1}, w_0, w_1, \ldots | w_{103} = \text{heads}\}$.
   The quantity $m(A_{103})$ measures the "relative proportion" of sequences that have heads in the 103th coordinate. We say that m is stationary if it id shift-invariant, that is, if m assigns the same probability to the event that heads appears in the 103rd coordinate as it assigns to the appearance of heads in any other coordinate. (This rationale for stationarity is unacceptable for subjectivists. For them, the fact that the tossing mechanism remains the same is besides the point.)
5. The conditional measure $m(A|A')$ is interpreted as the "relative proportion" of the sequences that have the property A among those that have the property A'. Measures for which for all $A, A' \in B$, $m(A|A') = m(A)$ are called Bernoulli measures.
6. For the definition of an arbitrary stochastic process, replace the set {heads, tails} with a fixed partition $p_1, \ldots, p_n$. The assumption is that at each moment exactly one of the p_i will happen.
7. The trajectory $O_T(s) = \ldots T^{-1}(s), s, T(s), \ldots$ of an element of a stochastic process that is shifted is the infinite sequence infinite sequences $(\ldots w_{-1}, w_0, w_1, \ldots)$, $(\ldots w_0, w_1, w_2, \ldots), \ldots$, $(\ldots w_n, w_{n+1}, w_{n+2}, \ldots), \ldots$ The time average $O_T(s, A_{103})$ is the measure of the relative frequency of the sequences in $O_T(s)$ that have heads in the 103th coordinate. Note that $O_T(s, A_{103}) = O_T(s, A_{102})$ because $O_T(s)$ has the same set of sequences but whose coordinates were shifted to the left.
8. Ergodicity in the shift space means that the "future" coordinates of a sequence are asymptotic independent of the present coordinates. Therefore, Bernoulli measures are clearly ergodic, because independence implies asymptotic independence.

3. The Haar Measure

A. Haar(1885 – 1933)

Haar worked in analysis studying orthogonal systems of functions, partial differential equations, Chebyshev approximations and linear inequalities. He is best remembered for his work on analysis on groups, introducing a measure on groups, now called the Haar measure.

Haar worked in analysis . His doctoral thesis studied orthogonal systems of functions. Later he went on to study partial differential equations. He also wrote on Chebyshev approximations and linear inequalities. Between 1917 and 1919 he worked on the variational calculus.

Haar is best remembered for his work of the Haar measure, which allows an analogue of Lebesgue integrals to be defined on locally compact topological groups. It was used by von Neumann, by Pontryagin in 1934 and Weil in 1940 to set up an abstract theory of commutative harmonic analysis.

4. Measure and Topology in Statistical Mechanics

5. Three Solutions