Journal of Statistical Physics, Vol. 11, No. 4, 1974

The Initial Distribution in Classical Statistical Mechanics

J. M. Richardson[1,2]

Received May 29, 1974

This paper presents arguments proving that several kinds of experimental preparation procedures for classical systems lead in certain limits to initial distributions that are functions only of macroscopic variables.

KEY WORDS: Information theory; nonequilibrium statistical mechanics; initial distributions; adaptive control; state space.

1. INTRODUCTION

A basic conceptual problem of nonequilibrium classical statistical mechanics is the derivation of the initial distribution function corresponding to certain experimental preparation procedures, using a minimum of additional assumptions. Even in equilibrium classical statistical mechanics, the same problem exists because in practice measurements do not involve infinite time averages of properties of systems in arbitrary initial states—or arbitrary initial distributions in the context of repeated experiments. In practice, the system is prepared to be initially in equilibrium, after which measurements are made,

This research was supported by the U.S. Air Force Office of Scientific Research under Contract F44620-72-C-0072.

[1] Science Center, Rockwell International, Thousand Oaks, California.
[2] Present address: Electronics Research Division, Rockwell International, Anaheim, California.

© 1974 Plenum Publishing Corporation, 227 West 17th Street, New York, N.Y. 10011. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission of the publisher.

many of which can entail durations smaller than many of the characteristic relaxation times of the system.

In this paper we will address the problem of the initial distribution in a more general context than equilibrium or nonequilibrium statistical mechanics, namely that of adaptive control.[3] Since this context entails, after the initial preparation, observations alternating with control actions, certain kinds of arguments are not appropriate, e.g., the argument based upon the time average of the process of waiting for a suitable fluctuation and making a delayed observation. Another important generalization of the problem demanded by adaptive control is the necessity of considering distributions representing macroscopic statistical uncertainty as contrasted with the microscopic uncertainty exemplified by, say, the microcanonical or canonical distributions.

In this paper we prove that several kinds of experimental preparation procedures lead in certain limits to initial probability distribution functions that are functions only of a prescribed set of macroscopic variables. However, we do not presume to derive all such distributions. This treatment is limited to general autonomous Hamiltonian systems. There is no attempt to deal with the limit of infinite volume (see, e.g., Ref. 2). In a corresponding sense, there is no attention given to the derivation of canonical or hypercanonical[3] distributions. Two reasons for these limitations are that the generality of the Hamiltonian systems does not lend itself to definition of meaningful infinite sequences of systems of increasing size without special assumptions and also that our investigation will include macroscopic statistical uncertainty.

Our plan of action is the following. The first step is the derivation of the "prior" distribution at the time $t = -0$ based upon the random selection of microstates at random past times. If the system is ergodic, the limit of an infinitely broad distribution of past times (i.e., infinite variance) leads to a prior distribution that is a function only of the Hamiltonian and any other measurable dynamical invariants that may exist. On the other hand, if the system has the mixing property,[4] then the weaker assumption that the average past time recedes to $-\infty$ suffices to yield a prior distribution that is a function of the Hamiltonian and other measurable dynamic invariants. There are other kinds of system behavior[5] other than ergodicity and mixing (which implies ergodicity); however, our treatment is limited to these two cases.

[3] Adaptive control involves control decisions that depend upon the past history of observations and control actions. The latter data may be summarized by the distribution function in state (e.g., phase) space with conditioning on past observations. For further discussion see Aoki.[1]

[4] The concept of mixing was discussed by Gibbs[4] and was later generalized and rigorized by Hopf.[5]

[5] For reviews of modern investigations in ergodicity and allied topics the reader can consult Arnold and Avez.[6]

The only assumption underlying the derivation of the prior distribution is the rather weak one that the microstates and past times are statistically independent.

The next step in the overall procedure is the relaxation of a force, the removal of a constraint, or the conditioning on a set of observations, processes that lead to initial distributions (mostly nonequilibrium) that are functions only of macroscopic variables.

2. DEFINITIONS AND PRELIMINARIES

The following paragraphs are devoted to definitions, notational convention, and background material.

2.1. Microstates

We will consider only the conventional phase space composed of the set of canonical coordinates and momenta represented by the row vector

$$X = (p_i, \ldots, p_f, q_i, \ldots, q_f) \tag{1}$$

where f is the number of degrees of freedom and q_j and p_j are the coordinate and momentum, respectively, associated with the jth degree of freedom. It is assumed that the domain of definition of X space is $\mathscr{D}_x$.

An integral in microstate space will be written

$$\int dX\, g(X) \tag{2}$$

where $dX = \prod_{j=1}^{f} dq_j\, dp_j$. The integration spans the domain $\mathscr{D}_x$ of microstate space, a convention assumed for all subsequent integrations on X unless otherwise specified.

In some situations, it is necessary or desirable to employ microstate spaces that are more complex than ordinary phase space. A well-known example of this is the kind of microstate space involved in the grand canonical ensemble. In adaptive control it is usual to include in the microstate certain system parameters that define the dynamics of the system (e.g., masses, parameters in the interaction potential, etc.).

2.2. Equations of Motion

Since the present investigation is limited to Hamiltonian systems, the equations of motion of the coordinates and momenta are

$$\dot{q}_i = \partial H/\partial p_i, \qquad \dot{p}_i = -\partial H/\partial q_i, \qquad i = 1, \ldots, f \tag{3}$$

The Hamiltonian function H is assumed to depend on X alone and not explicitly on the time t.

With the initial condition $X(0) = X_0$, the solution of the above equations of motion can be expressed in the form

$$X(t) = U(t, X_0) \tag{4}$$

where

$$U(0, X_0) = X(0) = X_0 \tag{5}$$

This result is valid for negative as well as positive times. It follows from (3) that

$$U(t_1 + t_2; X_0) = U(t_1; U(t_2, X_0)) \tag{6}$$

a result that is valid whether t_1 and t_2 have the same sign or not.

2.3. Distribution Functions

At a given time, $t = 0$ say, we assume that a distribution function[6,7] $P(X; 0)$ can be defined such that $P(X^0; 0)\, dx^0$ is the probability that the microstate vector X lies in the volume element dX^0 at X^0 in X space. At another time $t \neq 0$, the distribution function can be defined by

$$P(X; t) \triangleq \int dX'\, P(X'; 0)\, \delta(X - U(t, X')) \tag{7}$$

which reduces to an identity as $t \to 0$. The preservation of normalization, i.e., $\int dX\, P(X; 0) = 1$ implies $\int dX\, P(X; t) = 1$, can be directly demonstrated.

It follows from (3) and (4) that $P(X; t)$ satisfies the Liouville equation

$$(\partial P/\partial t)(X; t) + \mathscr{L} P(X; t) = 0 \tag{8}$$

[6] In this paper we will use the term "distribution function" to be synonymous with the term "probability density" employed in the probability and statistics literature.

[7] In this paper we will use rather special notational conventions in relation to probability. The symbol $P(X, Y)$ denotes the distribution function with respect to X and Y which are in turn assumed to be continuous variables or sets of continuous variables. In the case where a variable may lose its description function by being substituted by another expression, we will introduce the original variable also as a subscript. For example, in Section 2.5 we will introduce a distribution function $P_\eta(\eta)$ in which the argument is replaced by $z - F(X)$, thereby obtaining $P_\eta(z - F(X))$, whereupon the subscript η indicates what distribution function it is. The preceding of a quantity by a | indicates that the distribution is conditioned on this quantity. The preceding of a quantity by a semicolon indicates that it is a parameter. For example, $P(X|z; -0)$ means that the distribution on X is conditioned by z [or $z = F(X) + \eta$] at the time $t = -0$. Sometimes the time is used to imply past conditioning.

where $\mathscr{L}$ is the Liouville operator defined by

$$\mathscr{L} \triangleq \sum_{i=1}^{f} \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i} \right) \tag{9}$$

without i multiplying the rhs as is frequently done. It can be shown that

$$\int dX\, g \mathscr{L} h = -\int dX\, h \mathscr{L} g \tag{10}$$

where $g(X)$ and $h(X)$ are functions that are arbitrary under the constraints that the integrals exist. It is implicitly assumed that the boundary conditions are such that

$$\int dX\, \mathscr{L}(gh) = 0 \tag{11}$$

Thus, from (9) and (10), it is evident that $\mathscr{L}$ is real and anti-Hermitian or alternatively $i\mathscr{L}$ is Hermitian.

The formal solution of (8) is

$$P(X; t) = e^{-t\mathscr{L}} P(X; 0) \tag{12}$$

In the subsequent discussion we will consider only those initial distributions $P(X; 0)$ for which the above solution exists.

Combining (12) with (7), we obtain

$$e^{-t\mathscr{L}} P(X; 0) = \int dX'\, \delta(X - U(t, X')) P(X'; 0) \tag{13}$$

Multiplication by the arbitrary function $f(X)$ and integration on X yields

$$\int dX [e^{t\mathscr{L}} f(X)] P(X; 0) = \int dX f(U(t, X)) P(X; 0) \tag{14}$$

from which it follows that

$$e^{t\mathscr{L}} f(X) = f(U(t, X)) \tag{15}$$

This implies the further relations

$$e^{t\mathscr{L}} X = U(t, X) \tag{16}$$

$$e^{t\mathscr{L}} f(X) = f(e^{t\mathscr{L}} X) \tag{17}$$

$$P(X; t) = P(U(-t, X); 0) \tag{18}$$

If the microstate vector contains system parameters in addition to coordinates and momenta, the above results are valid as long as the system parameters are dynamically invariant. Of course, with a microstate space of the kind occurring in the grand canonical ensemble, the integrations on X must be altered in a well-known way.

2.4. Macrostates

We define a macrostate to be the set of values of a set of macroscopic variables α_k, $k = 1, \ldots, m$, where m is assumed to be much smaller than the dimensionality of the microstate X, i.e., $2f$ in the case of ordinary phase space. Each macroscopic variable is assumed to be a function only of the microstate, i.e.,

$$\alpha_k = \alpha_k(X) \tag{19}$$

For convenience we introduce the vector function

$$\alpha(X) = (\alpha_1(X), \ldots, \alpha_m(X)) \tag{20}$$

We denote the range of $\alpha(X)$ by $\mathscr{R}_\alpha$, i.e., the domain of definition of α space.

It is convenient to introduce the concept of macroscopic manifold $\mathscr{M}$, which is defined here as the set of all functions of $\alpha(X)$. Alternatively, we can define $\mathscr{M}$ as the set of all linear combinations of

$$\delta(a - \alpha) \triangleq \prod_{k=1}^{m} \delta(a_k - \alpha_k) \tag{21}$$

where the constant vector

$$a = (a_1, \ldots, a_m) \tag{22}$$

assumes a continuous set of values in $\mathscr{R}_\alpha$, the range of α. It follows that $1 \in \mathscr{M}$. We will impose the requirement on $\mathscr{M}$ that the Hamiltonian H and all observables, i.e., functions of X that are actually observed, all belong to $\mathscr{M}$.

We next introduce the notion of various kinds of shells in X space. The most familiar is the H shell, or energy shell, defined as the set of points in X space consistent with the inequality

$$E \leqslant H(X) < E + dE \tag{23}$$

An α_k shell is correspondingly defined as the set of points consistent with the inequality

$$a_k \leqslant \alpha_k(X) < a_k + da_k \tag{24}$$

Of particular interest is the α shell defined as the set of points in X space consistent with the set of inequalities

$$a_k \leqslant \alpha_k(X) < a_k + da_k, \qquad k = 1, \ldots, m \tag{25}$$

It is clear that other shells can be defined by making obvious modifications of the above definitions.

The volume of an H shell is

$$d\Omega_H = \Lambda_H(E)\, dE \tag{26}$$

where $\Lambda_H(E)$ is the structure function of an H shell given by

$$\Lambda_H(E) = \int dX\, \delta(E - H) \tag{27}$$

Correspondingly, the volume of an α shell is

$$d\Omega_\alpha = \Lambda_\alpha(a)\, da \tag{28}$$

where

$$da = \prod_{k=1}^{m} da_k \tag{29}$$

and where the structure function for an α shell is given by

$$\Lambda_\alpha(a) = \int dX\, \delta(a - \alpha) = \int dX \prod_{k=1}^{m} \delta(a_k - \alpha_k) \tag{30}$$

The equations for the volume of an α_k shell and the corresponding structure function are obvious.

A distribution function $P(X)$ in X space will be called coarse-grained if it is uniform in each α shell. Thus it is a function only of α and we will write

$$P(X) = P^c(\alpha) \tag{31}$$

It is clear that

$$P^c(\alpha) = \mathscr{C}P(X) \tag{32}$$

where $\mathscr{C}$ is the coarse-graining operator (corresponding to $\mathscr{M}$) defined by

$$\begin{aligned} \mathscr{C}f(X) &= \left[\int dX' f(X')\, \delta(\alpha - \alpha')\right] \Big/ \int dX'\, \delta(\alpha - \alpha') \\ &= \Lambda_\alpha(\alpha)^{-1} \int dX' f(X')\, \delta(\alpha - \alpha') \end{aligned} \tag{33}$$

where $f(X)$ is an arbitrary function. If $P(X)$ is not uniform in each α shell, it can be made so by application of the coarse-graining operator $\mathscr{C}$.

The distribution function $P(a)$ in α space is defined by the statement that $P(a)\, da$ is the probability that α lies in the volume element da at a. It can be shown that

$$P(a) \triangleq \int dX\, \delta(a - \alpha) P(X) = \Lambda_\alpha(a) P^c(a) \tag{34}$$

where $\Lambda_\alpha(a)$ is defined by (30).

2.5. Observational Conditioning

We consider here the problem of calculating the distribution function after it has been conditioned[8] by the results of a set of observations. Let us assume that the prior distribution function is $P(X; -0)$ and that at the time $t = 0$, a set of observations is made. Let the observational process be defined by

$$z = F(X) + \eta \tag{35}$$

where z and η are n-vectors and $F(X)$ is an n-vector function of the microstate X. We require that $F(X)$ be actually a function only of α, i.e., $F_j(X) \in \mathscr{M}$, $j = 1, \ldots, n$. We further require that $n \leqslant m$. The observational errors are represented by η, a random m-vector which is *assumed to be independent* of X and whose probability distribution is $P_\eta(\eta)$. After observations are made, the joint distribution of X and z is

$$\int dX \int d\eta' \, \delta(X - X') \, \delta(z - F(X') - \eta') P(X'; -0) P_\eta(\eta')$$
$$= P(X; -0) P_\eta(z - F(X)) \tag{36}$$

The observationally conditioned distribution function of X is obtained directly from the last expression by normalizing on X as follows:

$$P(X; +0) \triangleq P(X \mid z; -0) = c(z) P(X; -0) P_\eta(z - F(X)) \tag{37}$$

where the normalization "constant" $c(z)$ is given by

$$c(\tilde{z})^{-1} = \int dX \, P(X; -0) P_\eta(\tilde{z} - F(X)) \tag{38}$$

In this expression we have used $\tilde{z}$, a particular value of z, instead of z itself to avoid the implicit dependence of z on X through (35). Henceforth, we will use $P(X; +0)$ for the distribution function of X observationally conditioned at time $t = +0$ instead of the more explicit form $P(X \mid z; -0)$.

2.6. Measurable Invariants

In general, one would expect in the case of ordinary phase space that there would exist $2f - 1$ dynamical invariants. However, it turns out that most of these invariants, particularly in large systems, have corresponding shells that fill all of microstate space or at least certain subspaces corresponding to the shells of well-behaved invariants. Such aberrant invariants are called nonuniform, nonintegrable, or nonmeasurable.

The determination of the set of measurable dynamical invariants for specified systems remains a largely unsolved problem. For special simple

[8] For a discussion of conditioning see van der Waerden.[7]

systems, e.g., a set of harmonic oscillators with irrational frequency ratios, hard spheres in a box, and a few others, this problem has been solved.[9]

In the case of a system of identical interacting particles in a rigid container or in a space with cyclic boundary conditions (toroidal topology), it appears that the only known measurable invariants are associated with symmetry. First of all, the Hamiltonian itself arises from time-translation invariance. The system with cyclic boundary conditions may be invariant to spatial translations in three independent directions, in which case the three components of total linear momentum are measurable dynamical invariants. In the case of a perfectly smooth, rigid, cylindrical container, the component of total angular momentum parallel to the cylindrical axis is a measurable dynamical invariant if certain additional conditions are satisfied.

For an isolated self-contained system in space, e.g., the solar system or a galaxy, there are at least seven measurable dynamical invariants: the Hamiltonian, three components of total linear momentum, and three components of total angular momentum. That these are the only measurable dynamical invariants in a system of point masses interacting with gravitational forces has been proved by Poincaré.(9) We hasten to remark that thermodynamic equilibrium does not exist in such a system.

The above discussion applied to state spaces composed only of coordinates and momenta. There are many situations in which it is expedient, and sometimes mandatory, to extend the state space to include other quantities which were formerly just time-independent parameters in the Hamiltonian. The most familiar example of this is the state space used in the grand canonical ensemble. In this example, the total numbers of particles of various kinds constitute additional microstate variables and hence additional dynamical invariants. In the earlier discussion these were implicit constants in the Hamiltonian. In the context of adaptive control, it is usually necessary to include in the state vector other kinds of parameters whose values are not known a priori. If these parameters are time independent, which is frequently true, then they must be treated as dynamical invariants. In an obvious sense, these dynamical invariants are always measurable.

In real physical systems, the question of the existence of measurable dynamical invariants is usually a matter of the relative rates of change of various functions of the state. One usually uses a Hamiltonian that is only approximate in that it ignores many slow processes which of course vitiate its strict invariance. With the true equations of motion, the approximate Hamiltonian will be slowly varying. If it varies sufficiently little during the course of an experiment, then it can be regarded as invariant in a practical sense. In a system of metastable, but slowly decaying molecules, an approxi-

[9] For further discussion of these matters see Ref. 8.

mate Hamiltonian that ignores the decay processes is an example of the general situation discussed above. As long as the duration of an experiment is short compared with the decay time, this Hamiltonian will be satisfactory in a practical sense.

For a more detailed discussion of dynamical invariants the reader is urged to consult the book by Farquhar.[10]

3. A PARTIALLY VALID DERIVATION OF THE INITIAL DISTRIBUTION FUNCTION

The simplest approach to the problem of the initial distribution function is to assume that a set of observations is made at the initial time $t = 0$ with a very flat distribution immediately prior to $t = 0$. The observationally conditioned distribution existing a moment later is then taken as the initial distribution.

In explicit terms, we assume that the prior distribution function is $P(X; -0)$ at the time $t = -0$. At the time $t = 0$, the macroscopic observables, defined by the n-vector function $F(X)$ contained in $\mathscr{M}$, are observed with errors defined by (35). The observationally conditioned distribution is, according to (38), given by

$$P(X; +0) = c(\tilde{z})P(X; -0)P_\eta(\tilde{z} - F(X)) \tag{39}$$

in which $\tilde{z}$ are the values of the observations.

If the prior distribution function $P(X; -0)$ is sufficiently flat over the region of X space in which $P_\eta(\tilde{z} - F(X))$ exceeds a certain threshold[10] θ, then we can as an approximation replace $P(X; -0)$ by a constant, thereby obtaining

$$P(X; 0) = d(\tilde{z})P_\eta(\tilde{z} - F(X)) \tag{40}$$

In this expression $d(\tilde{z})$ is a new normalization "constant" defined by

$$d(\tilde{z})^{-1} = \int dX\, P_\eta(\tilde{z} - F(X)) \tag{41}$$

Since $F(X) \in \mathscr{M}$, it follows that $P(X; +0) \in \mathscr{M}$ and hence by the definition of Section 2.4, it is a coarse-grained distribution.

This result corresponds to the old doctrine of uniform a priori distributions in phase space. This assumption is invariant to contact transformations and hence avoids within this context the paradox that what is "most random" in one coordinate system is not in another. However, some investigators are not willing to accept this assumption or even the somewhat weaker

[10] It is appropriate to choose θ such that $\int_{P_\eta < \theta} dX\, P_\eta \ll \int dX\, P_\eta$.

assumption that the distribution function is quite flat over the region in which $P_\eta > 0$.

We have not found a convincing way of justifying the assumption of a distribution function $P(X; -0)$ that is nearly constant in any extended domain. However, in the subsequent discussion we will derive by other arguments forms of $P(X; 0)$ that are functions of H and of any other measurable dynamical invariants that may exist.

4. THE PRIOR DISTRIBUTION FUNCTION

In this section we present a derivation of the prior distribution function based upon a limit of performable operations.[11]

4.1. Random Selection of Microstates

Let us consider the random selection of microstates X at random past times t_0. Such a random procedure is described by the joint distribution function $P(X, t_0)$ which, of course, vanishes for $t_0 \geqslant 0$. We now make the assumptions that X and t_0 are statistically independent, i.e.,

$$P(X, t_0) = P(X)P(t_0) \tag{42}$$

This in turn means that

$$P(X|t_0) \triangleq P(X, t_0)/P(t_0) = P(X) \tag{43}$$

The distribution function on X at any nonnegative time t is then given by

$$P(X; t) = \int_{-\infty}^{-0} dt_0\, P(t_0)\{\exp[-(t - t_0)\mathscr{L}]\}\, P(X) \tag{44}$$

in which $\mathscr{L}$ is the Liouville operator defined by (9).

It is to be stressed that the above assumptions are very weak indeed. That random microstates are chosen at random times is the inevitable microscopic description of a crude macroscopic selection process. The assumption of the statistical independence of microstates and times is perhaps not inevitable but is certainly reasonable. When compared with other assumptions often used in the foundations of statistical mechanics, it is clearly very innocuous.

At this point we consider two types of distribution functions of t_0: (i) very broad distribution functions, i.e., ones yielding very large variances of

[11] This approach is analogous to the method of arbitrary functions in probability theory. An example is the dropping of a coin with an arbitrary distribution function of initial conditions from increasing heights.

t_0; and (ii) distributions that are peaked in the very remote past, i.e., those corresponding to very large negative means of t_0. We will show that if the system is ergodic, most distribution functions of t_0 of type (i) will yield a $P(X; t)$, $t \geqslant 0$, that is a function only of the Hamiltonian H (and of any other measurable dynamical invariants). We will also show that any distributions of type (ii) will yield a distribution function $P(X; t)$ that is a function only of H and other measurable invariants if the system possesses the so-called mixing property.

4.2. The Ergodicity Assumption

Let us analyze in more explicit detail the assertion associated with type (i) distributions of t_0. Let us consider a distribution function of the form

$$\begin{aligned} P(t_0) &= 1/T, \qquad -T < t_0 < 0 \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{45}$$

In the limit $T \to \infty$, we obtain

$$P(X; t) = \operatorname*{Lim}_{T\to\infty} (1/T) \int_{-T}^{-0} dt_0 \{\exp[-(t - t_0)\mathscr{L}]\} P(X) \tag{46}$$

The theorems of von Neumann[11] and of Birkhoff[12] assert that the above limit exists in the mean or almost everywhere, respectively. It can be proved that for t finite and nonnegative, the distribution function $P(X; t)$ is independent of t. If H is the only measurable dynamical invariant, i.e., the system is ergodic, then clearly $P(X; t)$ is a function of H alone. Therefore we obtain

$$\operatorname*{Lim}_{T\to\infty} (1/T) \int_{-T}^{-0} dt_0 \{\exp[-(t - t_0)\mathscr{L}]\} P(X) = P^c(H) \tag{47}$$

almost everywhere and by multiplication by $\delta(H - E)$ and integration on X we deduce that

$$P^c(E) = \left[\int dX\, \delta(H - E) P(X)\right] \Big/ \int dX\, \delta(H - E) \tag{48}$$

Thus $P^c(H)$ is obtained from $P(X)$ by uniformizing $P(X)$ within each energy shell.

4.3. The Mixing Assumption

We turn now to the consideration of the type (ii) distribution function of t_0. Let us assume that

$$P(t_0) = \phi(t_0 - T)$$

with

$$\phi(t_0) = 0 \qquad \text{if} \quad t_0 \geqslant 0 \tag{49}$$

Then (44) becomes

$$P(X; t) = \int_{-\infty}^{-0} dt_0\, \phi(t_0 - T)\{\exp[-(t - t_0)\mathscr{L}]\}P(X) \tag{50}$$

We now make a stronger assumption, namely that the system possesses the mixing property. This property implies the ergodic property. Here, the mixing property can be defined by the statement that the limit

$$\lim_{t_0 \to -\infty} \int dX\, g(X)\{\exp[-(t - t_0)\mathscr{L}]\}P(X) \tag{51}$$

exists for an arbitrary function $g(X)$ of well-behaved character, i.e., $\int dX\, g^2(X)P(X) < \infty$. A direct corollary of the definition of the mixing property is that the limit is independent of t, i.e.,

$$\lim_{t_0 \to -\infty} \int dX\, g(X)\{\exp[-(t - t_0)\mathscr{L}]\}P(X)$$

$$= \lim_{t_0 \to -\infty} \int dX\, g(X)[\exp(t_0\mathscr{L})]P(X) \tag{52}$$

It should be emphasized that (51) is a weaker definition of mixing than is usually employed. The usual definition is concerned only with mixing in an infinitesimal energy shell (or conceivably a measurable, invariant subspace of this energy shell). Our weaker definition can admit situations in which there is *no* mixing within energy shells but in which there is mixing between energy shells (i.e., a process in which two points in neighboring shells become decorrelated).

It is to be emphasized that the limit (51) corresponds to the notion of weak convergence. The limit in the sense of strong convergence, i.e.,

$$\lim_{t_0 \to -\infty} \{\exp[-(t - t_0)\mathscr{L}]\}P(X) \tag{53}$$

does not exist except when $P(X)$ is a measurable invariant. As is well known, the volume of X space for which $\{\exp[-(t - t_0)\mathscr{L}]\}P(X) > \lambda \geqslant 0$ is independent of t_0, as is illustrated by the analogy of the stirring of coffee and cream (with interdiffusion neglected). Thus if $P(X)$ is noninvariant and if the system has the mixing property, the above limiting process corresponds to $\{\exp[-(t - t_0)\mathscr{L}]\}P(X)$ developing increasingly fine-grained striations oriented almost parallel to the trajectories in X space. In calculating the average of any well-behaved function $g(X)$, these striations are "smoothed out" when $t - t_0$ is sufficiently large and hence a limit can exist in this sense.

In the case where $P(t_0) = \phi(t_0 - T)$ we can show for an arbitrary, well-behaved function $g(X)$ that the following results hold:

$$\begin{aligned}
&\operatorname*{Lim}_{T\to-\infty} \int dX\, g(X)P(X;t) \\
&\quad = \operatorname*{Lim}_{T\to-\infty} \int dX\, g(X) \int_{-\infty}^{-0} dt_0\, \phi(t_0 - T)\{\exp[-(t - t_0)\mathscr{L}]\}\, P(X) \\
&\quad = \operatorname*{Lim}_{T\to-\infty} \int_{-\infty}^{-0} dt_0\, \phi(t_0 - T) \int dX\, g(X)\{\exp[-(t - t_0)\mathscr{L}]\}\, P(X) \\
&\quad = \operatorname*{Lim}_{T\to-\infty} \int_{-\infty}^{-0} du\, \phi(u) \int dX\, g(X)\{\exp[-(t - u - T)\mathscr{L}]\}\, P(X) \\
&\quad = \int_{-\infty}^{-0} du\, \phi(u) \operatorname*{Lim}_{T\to-\infty} \int dX\, g(X)\{\exp[-(t - u - T)\mathscr{L}]\}P(X) \\
&\quad = \int_{-\infty}^{-0} du\, \phi(u) \operatorname*{Lim}_{T+u\to-\infty} \int dX\, g(X)\{\exp[-(t - u - T)\mathscr{L}]\}P(X) \\
&\quad = \operatorname*{Lim}_{T\to-\infty} \int dX\, g(X)\{\exp[-(t - T)\mathscr{L}]\}P(X) \qquad (54)
\end{aligned}$$

Thus the case of a distribution of times t_0 we obtain the same limit as (51) when the center of the distribution recedes to $-\infty$. Again a direct corollary is that

$$\operatorname*{Lim}_{T\to-\infty} \int dX\, g(X)P(X, t) \qquad (55)$$

is independent of t.

If H is the only measurable dynamical invariant and if $g(X)$ is not in general invariant, then it is clear that

$$\operatorname*{Lim}_{T\to-\infty} \int dX\, g(X)P(X;t) = \int dX\, g(X)P^c(H) \qquad (56)$$

since the lhs is independent of t. As a special case let us assume that $g(X)$ is an invariant of the form

$$g(X) = \chi(E - H) \qquad (57)$$

where $\chi(u)$ is the unit step function defined by

$$\begin{aligned}
\chi(u) &= 1, \qquad u \geqslant 0 \\
&= 0, \qquad u < 0 \qquad (58)
\end{aligned}$$

we obtain

$$\int dX\,\chi(E-h)P^c(H) = \operatorname*{Lim}_{T\to-\infty}\int dx\,\chi(E-H)P(X|t)$$

$$= \operatorname*{Lim}_{T\to-\infty}\int dX\,\chi(E-H)\{\exp[-(t-T)\mathscr{L}]\}P(X)$$

$$= \int dX\,\chi(E-H)P(X) \tag{59}$$

Differentiation[12] on E yields

$$\int dX\,\delta(E-H)P^c(H) = P^c(E)\int dX\,\delta(E-H) = \int dX\,\delta(E-H)P(X) \tag{60}$$

or

$$P^c(E) = \left[\int dX\,\delta(E-H)P(X)\right]\Big/\int dX\,\delta(E-H) \tag{61}$$

which, of course, is identical to (48) obtained under the ergodicity assumption.

4.4. Several Measurable Dynamical Invariants

In the case in which there are additional measurable dynamical invariants, the previous results can be modified in a straightforward way. Let the set of measurable dynamical invariants be denoted by

$$I = (I_1(X),\ldots, I_q(X)) \tag{62}$$

where it is assumed that one of the I_s is the Hamiltonian or that the Hamiltonian can be expressed as a function of I. We require the manifold $\mathscr{M}$ to be such that it contains I, i.e., $I_s \in \mathscr{M}$, $s = 1,\ldots, q$. We must consider a more general dynamically invariant density, i.e., $P^c(I) \equiv P^c(I_1,\ldots, I_q)$. An appropriate generalization of the preceding analysis yields the result

$$P^c(I^0) = \left[\int dX\prod_s \delta(I_s^0 - I_s)P(X)\right]\Big/\int dX\prod_s \delta(I_s^0 - I_s) \tag{63}$$

in which the I^0 is a set of possible numerical values of $I(X)$.

5. THE INITIAL DISTRIBUTION

We have proved that the random selection of states at random past times will under certain assumptions yield at $t = -0$ a prior distribution that is a function only of the Hamiltonian and also of any other measurable

[12] It is of course understood that the limit $T\to-\infty$ is carried out before performing differentiation.

dynamical invariants that may exist. We turn next to the use of this result in various experimental preparation procedures. We discuss three possibilities.

5.1. Force Relaxation

Let us suppose that at the time $t = 0$, certain previously constant forces are suddenly relaxed. This means that there are different Hamiltonians for negative and positive times, i.e.,

$$\begin{aligned} H &= H_0 + H_1, \qquad t < 0 \\ H &= H_0, \qquad\quad t \geqslant 0 \end{aligned} \tag{64}$$

where it is assumed that H_0 and H_1 do not explicitly depend upon the time. Let us further assume that $H_0, H_1 \in \mathscr{M}$. Using the random selection of states at random past times with appropriate limiting processes discussed in Section 4, we obtain

$$P(X; -0) = P^c(H_0 + H_1) \tag{65}$$

in which

$$P^c(E) = \left[\int dX\, \delta(E - H_0 - H_1)P(X)\right] \Big/ \int dX\, \delta(E - H_0 - H_1) \tag{66}$$

If additional measurable dynamical invariants exist, this result must be extended in a manner similar to (63).

The initial distribution function is, by continuity, given by

$$P(X; 0) = P(X; -0) \tag{67}$$

However, it is a nonequilibrium distribution because $H_0 + H_1$ is no longer a dynamical invariant for $t \geqslant 0$ since H_0 is now the correct Hamiltonian. A possible exception is the case in which H_1 is also a measurable dynamical invariant for $t \geqslant 0$. Since $H_0, H_1 \in \mathscr{M}$, it follows that $P(X; 0) = P^c(H_0 + H_1)$ is a special distribution of the coarse-grained type.

5.2. Constraint Removal

Another type of process that may be regarded as a limiting form of force relaxation is the sudden removal of certain constraints. A well-known physical example is a sudden removal of a barrier that previously separated the gas in a container into two separate parts. If this barrier passes heat but not particles, then as a mild idealization we can assume that Hamiltonian H and the number of particles $N^{(1)}$ on one side are measurable dynamical invariants for negative times. For nonnegative times, after the barrier removal, the number of particles $N^{(1)}$ on one side of the surface where the barrier was is clearly no longer a dynamical invariant.

In the case in which the barrier is impervious to heat and particles, then during negative time, the Hamiltonians $H^{(1)}$ and $H^{(2)}$ of the volumes of gas on the two sides as well as the number of particles $N^{(1)}$ on one side are all measurable dynamical invariants. Clearly, after the removal of the barrier, the above quantities cease to be separately invariant and only the Hamiltonian H of the united system survives as a measurable invariant.

There are kinds of constraints that, while physically unrealizable, make possible constrained equilibria which are close approximations to real, almost quasistationary situations. The best example of this is the placement of a barrier along the reaction coordinate in a reacting collision complex—or, more precisely, in all possible reacting collision complexes of a specified type. This viewpoint implicitly underlies many theories of chemical reaction rates.

In most situations, the removal of a constraint causes the disappearance of one or more measurable dynamic invariants, e.g., the number of particles on one side of a barrier. It is not always true that measurable invariants only disappear; in some cases new ones (in fact, new types, not mere perturbations of old ones) may appear. For example, consider a one-component gas in a smooth cylindrical container with an internal barrier that destroys the cylindrical symmetry. After the removal of this barrier, the number of particles on one side of the barrier ceases to be an invariant but, because cylindrical symmetry is regained, the total angular momentum parallel to the cylindrical axis appears as a valid measurable invariant. However, regardless of how many measurable invariants disappear and appear at the instant the constraint is removed, the distribution function $P(X; 0)$ is a function of the previous measurable invariants and hence is in general noninvariant. If the previous invariants belong to $\mathscr{M}$, the initial distribution is coarse-grained.

5.3. Observation at $t = 0$

Let us assume, as before, that random microstates are selected at random past times, whereupon we obtain

$$P(X; -0) = P^c(H) \tag{68}$$

is the Hamiltonian H is the only measurable dynamical invariant for $t < 0$, or we obtain

$$P(X; -0) = P^c(I) \tag{69}$$

if there is more than one dynamical invariant for $t < 0$. Let us proceed on the assumption that H is the only measurable dynamical invariant.

Let us assume that an observation is made at $t = 0$ of the properties $F(X) = (F_1(X), ..., F_n(X)$ with the possible results

$$z = F(X) + \eta \tag{70}$$

where the m-vector η represents the observational errors and has the distribution $P_\eta(\eta)$. As shown in Section 2.5, the observationally conditioned distribution at $t = 0$ is then

$$P(X; +0) = c(\tilde{z})P_\eta(\tilde{z} - F(X))P(X; -0) = c(\tilde{z})P_\eta(\tilde{z} - F(X))P^b(H) \quad (71)$$

Therefore, the observational conditioning yields another coarse-grained distribution since $F_j(X) \in \mathcal{M}, j = 1, \ldots, n$, and of course this new distribution is noninvariant if some of the $F_j(X)$ are noninvariant.

6. COMMENTS

It has been proved that the analysis of several kinds of experimental preparation procedures leads to initial distributions that are functions only of a certain set of macroscopic observables. Our approach consisted of two parts: first the random selection of microstates at random past times, a process leading in certain limits to a prior (i.e., $t = -0$) distribution depending only on measurable invariants; and second an initial (i.e., $t = 0$) operation consisting of force relaxation, constraint removal, or observational conditioning, processes leading to distributions that depend only on a certain set of macroscopic variables which are in general noninvariant. The total procedure involves one ad hoc assumption, namely the statistical independence of past random microstates and past random times, a rather weak assumption which may be removable.

It is to be emphasized that the above procedure usually yields nonequilibrium distributions; however, not all distributions depending only on macroscopic variables are yielded by this procedure. Another point, which perhaps should be emphasized even more, is that situations involving *macroscopic uncertainty* are included in our preparation procedure. Thus we are not concerned only with the macroscopic behavior of a single system or, more loosely, ensembles of single systems with a relatively narrow spread of microstates.

It is of interest to note that distributions that are confined to a single shell (H shell, α shell, etc.) are special cases of our treatment. One of these can be obtained by choosing a random microstate distribution $P(X)$ that is confined to a shell corresponding to measurable invariants but is not necessarily uniform within this shell. Observational conditioning, if it is involved, should entail negligible error η.

REFERENCES

1. M. Aoki, *Optimization of Stochastic Systems*, Academic Press, New York (1967).
2. D. Ruelle, *Statistical Mechanics*, Benjamin, New York (1969).
3. J. M. Richardson, *J. Math. Anal. Appl.* 1:12–60 (1960).

4. J. W. Gibbs, *Elementary Principles in Statistical Mechanics*, Dover, New York (1960), pp. 144–145.
5. E. Hopf, *J. Math. Phys.* **13**:51 (1934).
6. V. I. Arnold and A. Avez, *Ergodic Problems in Statistical Mechanics*, Benjamin, New York (1968).
7. B. L. van der Waerden, *Mathematical Statistics*, Springer-Verlag, New York (1969).
8. A. S. Wightman, in *Statistical Mechanics at the Turn of the Decade* (E. G. D. Cohen, ed.), Marcel Dekker, New York (1971); also J. L. Lebowitz and O. Penrose, *Physics Today* **26**:23–29 (1973).
9. H. Poincaré, *Méthodes Nouvelle de la Mécanique Céleste*, Gauthiers-Villars, Paris, 1892; reprinted by Dover, New York.
10. I. E. Farquhar, *Ergodic Theory in Statistical Mechanics*, Interscience, New York (1964).
11. J. von Neumann, *Proc. Natl. Acad. Sci.* (*U.S.*) **18**:70–82 (1932).
12. G. D. Birkhoff, *Proc. Natl. Acad. Sci.* (*U.S.*) **17**:656–660 (1931).