Partition of probability/statistical entropy

The formal information I (Equ. **Error! Reference source not found.**) gained through the acquisition of knowledge about which of any of the possible configurations is realized, represents the information that the totality of *nats* or *bits* of the Ω states can carry. This information is numerically equal to the number ($= \ell og_2 \Omega$) of questions that have to be posed in order to obtain certainty.

It is plausible that any asymmetry, any restriction in the probabilities for the various system configurations would imply a bias, which reduces the information content. For example, a 12-bit computer cell with two broken bits can not carry more information than a 10-bit word. Therefore, an ***equipartition of the total probability***

$$P = \sum_i p_i, \quad p_i = p = const. \tag{1}$$

among all system states (configurations) maximizes the information and the statistical entropy (cf. Equ. **Error! Reference source not found.**). In such a situation, where all configurations have an equal *a priori* probability, a many times repeated experiment is expected to exhibit every configuration with the same equal *a posteriori* (empirical) probability. For example, measuring 10000 times the same system with 3 equally probable configurations will show each one of them approximately 3300 times (1/3 of the total number of interrogations).

Mathematically the partition of the total probability among the system configurations that corresponds to maximum information (entropy) can be obtained by varying the probabilities p_n under the constraint of the normalization of all probabilities. This task is achieved efficiently with the ***method of Lagrange multipliers***. To illustrate the method, consider a 1-dimensional function *f(x)*. Here a "constrained maximum" of *f(x)*, under the constraint *g(x)=c=const.*, is found by searching for the maximum of the related function

$$\tilde{f}(x) = f(x) - \lambda\left[g(x) - c\right] \tag{2}$$

where λ is an arbitrary constant. Obviously, only along a path where the constraint $g(x)=c$ is fulfilled, are the functions $\tilde{f}$ and f identical, $\tilde{f}(x) = f(x)$.

In the case of interest here, the constraint is given by the normalization condition of the total probability for any of the Ω configurations,

$$g(p_1,\ldots,p_\Omega) = \sum_{i=1}^{\Omega} p_i = 1 \tag{3}$$

Then, in the usual way, the maximum of the information is found by setting to zero all first derivatives of $\tilde{f}$ with respect to individual probabilities p_n (for configuration n to be occupied, $n=1,\ldots,\Omega$,),

$$\frac{\partial}{\partial p_n}\left\{-\sum_{m=1}^{\Omega} p_m \ \ell n\left(p_m\right) + \lambda\left(\sum_{i=1}^{\Omega} p_i - 1\right)\right\} = 0 \tag{4}$$

The normalization condition of Equ. (3) has been multiplied by a yet undetermined but constant **Lagrange multiplier** λ and added to the function to be maximized. This procedure yields

$$-\ell n\left(p_n\right) - p_n \cdot \frac{1}{p_n} + \lambda = 0 \qquad n = 1,\ldots,\Omega \tag{5}$$

or

$$\ell n\left(p_n\right) = \lambda - 1 \tag{6}$$

and

$$p_n = e^{(\lambda-1)} = p = const. \tag{7}$$

implying ***an equal (a priori) probability for all n***. From the normalization condition $\sum_{n=1,..,\Omega} p_n = 1$ it follows immediately that

$$p_n = p = 1/\Omega = const \tag{8}$$

and

$$1 = \sum_{n=1}^{\Omega} p_n = \sum_{n=1}^{\Omega} e^{(\lambda-1)} = \Omega \cdot e^{(\lambda-1)} \tag{9}$$

which determines the Lagrange multiplier. This is the condition for ***maximum missing information*** concerning the locations of the particles of interest. It also signifies the situation for a ***maximum of information gained***, when the occupation of the single-particle states is disclosed.

In analogy to Equ. **Error! Reference source not found.**, the ***statistical entropy*** used in statistical mechanics for a system of N particles populating a set of Ω configurations ("micro-states") is defined as

$$S = -k_B \sum_{n=1}^{\Omega} p_n \ell n(p_n) \geq 0 \tag{10}$$

where $k_B = 1.38 \cdot 10^{-23}$ J/K is the universal ***Boltzmann constant***, which endows this information quantity with a non-trivial but unnecessary dimension that can obscure the real meaning of this important observable. If the system is interrogated N times and configuration n is observed N_n times out of N, the a ***posteriori (empirical) probabilities*** are determined by $p_n = N_n/N$. These probabilities fulfill the normalization condition

$$\sum_{n=1}^{\Omega} p_n = \sum_{n=1}^{\Omega} \frac{N_n}{N} = 1 \tag{11}$$

Following the same arguments as above for the information, one finds that the ***entropy is maximized for an equal population of all configurations (micro-states) with equal a priori probabilities p_n = 1/Ω***. In this case of independent individual objects that occupy ***with***

equal a priory probability all Ω available states, Equ. **Error! Reference source not found.** translates into the ***maximum statistical entropy***,

$$S = S_{max} = k_B \cdot \ell n\,\Omega = -k_B \sum_{n=1}^{\Omega} p_n\, \ell n\left(p_n\right) \geq 0 \qquad (12)$$

The left hand side of this equation is the famous *Boltzmann Equation* relating the number (Ω) of available micro-states to the phenomenological entropy, a state function defining macroscopic system states in phenomenological thermodynamics. The right hand side closes contact to the microscopic information content. The constant k_B, providing the information observable with an artificial dimension, makes more sense in phenomenological thermodynamics. The entropy is naturally bounded by the two fixed limits,

$$0 \leq S \leq S_{max} \qquad (13)$$

where S_{max} is attained for equal a priori probabilities, for any physical system it is itself a distribution that can be characterized by average expectation value, fluctuations and higher moments.

In a similar fashion, the method utilized above for evaluating the maximum information/entropy under the constraint of an equipartition of the total *a priori* probability can be extended to other constraints. For example, for a multi-particle system it is important how the system energy E is distributed among all Ω configurations, i.e., micro-states. For an isolated ("**micro-canonical**") system the total energy E is conserved, as are other observables. Then, all members of the set of possible equivalent configurations must have exactly the same energy,

$$E_i = E \qquad i = 1,....,\Omega \qquad (14)$$

Otherwise, the configurations would not be completely equivalent, i.e., not have the same *a priori probabilities*. In a statistical ensemble, this fixed energy E would be distributed over all members of the ensemble

with *equal a priori probabilities*. Systems that are not completely isolated and allow some exchange of energy with their surroundings are called "**canonical**." Here, the energies of the various configurations are subject to ***fluctuations about the averages***, *<E_i>*. ***Exactly equal a priori probabilities cannot be expected for these configurations.*** At most some energy averages can be well defined and assumed to have approximately the same value for every configuration:

$$\langle E_i \rangle = \sum_{i=1}^{\Omega} E_i \cdot p_i =: E \tag{15}$$

Here, $\langle E_i \rangle$ is the weighted average (mean) taken over all configurations. The condition that only small variations should exist between the mean energies E_i of the configurations implies that the ***mean square deviation (variance)*** of these energies

$$\sigma_E^2 = \sum_{i=1}^{\Omega} \left(E_i - \langle E_i \rangle \right)^2 \cdot p_i > 0 \tag{16}$$

be small compared to the average, i.e., $\sigma_E \ll E$. For an isolated, microcanonical system, there are no energy fluctuations, $\sigma_E = 0$.

Considerations of the ***maximum constrained information*** will reveal whether such a situation is possible and what the *a priori probabilities p_i* would look like. Certainly, ***because of an additional constraint, for a given energy E, the information/entropy cannot exceed the one obtained for a micro-canonical system*** with a minimum of constraints (total probability normalized).

A search for the maximum information/entropy has to take into account now two boundary conditions, Equ. (3) and Equ. (15). Therefore, the maximization condition (4) has to be extended to

$$\frac{\partial}{\partial p_n} \left\{ -k_B \sum_{m=1}^{\Omega} p_m \, \ell n\left(p_m\right) + \lambda_1 \left(\sum_{i=1}^{\Omega} p_i - 1 \right) + \lambda_2 \left(\sum_{j=1}^{\Omega} E_j p_j - E \right) \right\} = 0 \tag{17}$$

with two Lagrange multipliers, λ_1 and λ_2. For the constraint maximum information/entropy one now obtains the condition

$$-k_B\left(\ell n(p_n) + 1\right) + \lambda_1 + \lambda_2 E_n = 0 \tag{18}$$

This result implies that

$$p_n = exp\left\{\frac{\lambda_1 + \lambda_2 E_n}{k_B} - 1\right\} \qquad n = 1,...,\Omega \tag{19}$$

where λ_1 / k_B is dimensionless and λ_2 / k_B is an inverse energy. Obviously, the probabilities are not equal but depend on the energies E_n of the corresponding states. The normalization condition is now written as

$$1 = \sum_{n=1}^{\Omega} p_n = e^{\left(\frac{\lambda_1}{k_B} - 1\right)} \cdot \sum_{n=1}^{\Omega} e^{\frac{\lambda_2}{k_B} E_n} \tag{20}$$

a product of a constant and an energy sum. Obviously, the constant factor in Equ. (20) is equal to the inverse of the sum over the individual energy terms,

$$Z = e^{\left(1 - \frac{\lambda_1}{k_B}\right)} = \sum_{n=1}^{\Omega} e^{\frac{\lambda_2}{k_B} E_n} =: \sum_{n=1}^{\Omega} e^{-\beta \cdot E_n} \tag{21}$$

with the definition $\beta = -\lambda_2/k_B > 0$, the inverse of a characteristic energy. This function **Z=Z(β)** is also known as the **(canonical) partition sum**. According to Equ. (21) it can be cast both into a closed form (left) and as a sum over all configurations. Hence the normalization condition is recast as

$$\sum_{n=1}^{\Omega} p_n = \frac{1}{Z} \cdot \sum_{n=1}^{\Omega} e^{-\beta \cdot E_n} = 1 \tag{22}$$

This is a general result, valid for any number of configurations, their energy spectra (E_n) and varied system parameters λ_1 and λ_2.

Often groups of several (ϖ) states have the same energies, i.e., they are **energy degenerate** and bunched at some energy levels *E, E', etc.*

If the degeneracy (number of states at energy level E) is given by the function $\varpi(E)$, the partition function in Equ. (21) can be written

$$Z(\beta) = \underbrace{e^{-\beta \cdot E} + ... + e^{-\beta \cdot E}}_{\varpi(E)\ times} + \underbrace{e^{-\beta \cdot E'} + ... + e^{-\beta \cdot E'}}_{\varpi(E')\ times} + \underbrace{...}_{\varpi(...)} = \sum_E \varpi(E) \cdot e^{-\beta \cdot E} \quad (23)$$

According to Equ. (19), there is a term-by-term equivalence in Equ. (22). One therefore concludes that the **a priori probabilities for canonical system configurations are not equal but dependent on the energy.** Configuration by configuration, one has the normalized probability

$$p_n = \frac{1}{Z} \cdot e^{-\beta \cdot E_n} \quad (24)$$

Accordingly, the probabilities for the populations $p(E)$ of energy levels E are given by

$$p(E) = \frac{\varpi(E)}{Z} \cdot e^{-\beta \cdot E} \quad (25)$$

The requirements that the probabilities must be normalizable and that variations between the mean energies of equivalent (similar probabilities) be small suggests that the inverse-energy parameter β be positive, $\beta > 0$. Then, the populations for system configurations decrease exponentially with their energy, Configurations with extreme energies are simply not significantly populated. In fact, for thermodynamic systems independent considerations discussed further below show a relation of the parameter with the "canonical temperature" T, i.e., $\beta = 1/(k_B T)$. This implies that the information contained in such a system is incomplete at any temperature, reduced due to the decreased probability for energetic states.

The partition sum contains all relevant physical information on the system. Z **is a generating function for the system probability distribution.** This feature can be demonstrated by the following examples. The derivative of $\ln(Z)$ with respect to the energy E_i projects the probability of configuration i out of the partition sum,

$$-\frac{\partial}{\partial E_i} \ell n(Z) = \frac{-1}{Z}\frac{\partial}{\partial E_i}\sum_{n=1}^{\Omega} e^{-\beta \cdot E_n} = \frac{\beta}{Z}e^{-\beta \cdot E_i} = \beta p_i \qquad (26)$$

Here the chain rule

$$\frac{\partial}{\partial x}\ell n(Z) = \frac{1}{Z}\frac{\partial}{\partial x}Z \qquad (27)$$

has been used to generate the required normalization factor *1/Z*. Similarly, taking the derivative of $\ell n(Z)$ with respect to $-\beta$ produces the mean energy per configuration:

$$-\frac{\partial}{\partial \beta}\ell n(Z) = \frac{-1}{Z}\frac{\partial}{\partial \beta}\sum_{n=1}^{\Omega} e^{-\beta \cdot E_n} = \frac{1}{Z}\sum_{n=1}^{\Omega}E_n e^{-\beta \cdot E_n} = \sum_{n=1}^{\Omega}E_n \cdot p_n = \langle E_n \rangle = E \qquad (28)$$

Taking the result from Equ. (18), multiplying by p_n and summing over all configurations yields a connection between entropy, partition function and mean energy per configuration

$$0 = -k_B\sum_{n=1}^{\Omega}p_n\left\{(\ell n(p_n)+1)+\lambda_1/k_B+(\lambda_2/k_B)E_n\right\} =$$

$$= S - k_B\sum_{n=1}^{\Omega}p_n + \lambda_1\sum_{n=1}^{\Omega}p_n + \lambda_2\sum_{n=1}^{\Omega}p_n E_n \qquad (29)$$

$$= S - (k_B - \lambda_1) - k_B \cdot \beta \cdot E = S - k_B \cdot [\ell n(Z) - \beta \cdot E]$$

This results finally yields an expression for the macroscopic (mean) information/entropy in terms of the partition function *Z* and the expectation value of the energy,

$$S/k_B = \ell n\, Z + \beta \cdot E \qquad (30)$$

Equivalently, one can write for the partition function for a canonical system,

$$Z = e^{S/k_B} \cdot e^{-\beta \cdot E} \qquad (31)$$

This function replaces that for an isolated system, which according to Equ. (12), is simply the number of accessible states,

$$\Omega = e^{S/k_B} \qquad (32)$$

with the dimension-less statistical entropy S/k_B counting the number of "nats" measuring the size of the state space.

So far, the meaning and value of the parameter β appearing in the canonical partition function, have remained hidden. However, for any system obeying the Equ. (30), the parameter obeys the relation

$$\beta = \frac{\partial\left(S/k_B\right)}{\partial E} \quad or \quad \frac{\partial E}{\partial S} = \frac{1}{k_B \cdot \beta} \qquad (33)$$

It can therefore be evaluated for a system of interest, given specific relations between state energies and probabilities. Note that the derivatives in Equ. (33) are partial derivatives testing explicit dependencies, to be taken while keeping other coordinates constant.

1. *Illustrations of partition functions*

A detailed evaluation of the partition function is necessary for an interpretation of macroscopic observations in terms of the microscopic structure of a system, e.g., in terms of the internal energy spectrum. The task can be very demanding for quantal multi-particle systems with coupled degrees of freedom and correlated particles, e.g., for fermionic systems where the individual particles are indistinguishable and subject to the Pauli Exclusion Principle. Particle correlations are important for high particle densities in configuration or momentum space but loose efficiency at low densities and/or total internal energies. On the other hand, many classical d.o.f. such as molecular or nuclear rotations and vibrations are independent at low excitations but influence each other at higher energies. In the following, a few observations relevant to decoupled d.o.f. are made in order to illustrate basic structure of partition functions for various d.o.f.

For a system with multiple independent degrees of freedom, for example, a molecule with a set (*i=translational, rotational, vibrational, electronic, nuclear,...*), the total number of states Ω for is the product of the corresponding numbers Ω_i for the individual d.o.f. Its total energy $E = \langle E \rangle$ is the sum over the individual energies $E_i = \langle E_n(i) \rangle$. Therefore, the total partition function Z is the product of the individual functions Z_i corresponding to each of the d.o.f.,

$$Z(\beta) = \prod_i Z_i(\beta) = \prod_i \sum_n e^{-\beta \cdot E_n(i)} \tag{34}$$

Here, the energies $E_n(i)$ run over the entire energy spectrum associated with the i th d.o.f. In other words, as long as correlations can be neglected (quasi-classical, Boltzmann approximation), the partition function for such a system can be written as

$$Z = Z^{trans} \cdot Z^{rot} \cdot Z^{vib} \cdot Z^{electr} \cdot Z^{nucl} \cdots \tag{35}$$

Furthermore, for a quasi-classical *N*-particle molecular system, neglecting correlations, each partial partition function is a product of identical single-particle partition functions z_i. For example, for translational motion in 3D space $\{x, y, z\}$, the s.p. partition function can be approximated by

$$z^{trans} = \sum_{n_x, n_y, n_z} e^{-\beta \cdot \left(\varepsilon_{n_x} + \varepsilon_{n_y} + \varepsilon_{n_z} \right)} = \left(\sum_i e^{-\beta \cdot \varepsilon_i} \right)^3 \tag{36}$$

Here, the s.p. energy scheme of a particle in an infinite cubic 3D box of side length $a\tau\equiv$ is adopted. Such an infinite box accommodates the unrestricted translational motion of a free particle of mass *m*. For mathematical reasons, one adopts first a finite box but lets its dimension grow indefinitely in the final results.

For a finite side length the particle-in-a-box energy eigen values are given by a set of integer quantum numbers n_i

$$\varepsilon_i = \frac{h^2}{8ma^2} \cdot n_i^2 \tag{37}$$

Performing the transition to the infinite box, $a\delta\equiv$, the summation over discrete quantum numbers n_i in the partition sum can be replaced by an integral over continuous quantum numbers *n:*

$$z^{trans} = z^{trans}_{a \to \infty} = \lim_{a \to \infty} \left[\int_0^\infty dn\, e^{-\beta \cdot \frac{h^2}{8ma^2} \cdot n^2} \right]^3 = \left[\sqrt{\frac{2\pi}{\beta \cdot h^2}} m \cdot a \right]^3 \tag{38}$$

Since the volume is given by $V = a^3$, the translational partition function can also be written as

$$z^{trans} = \frac{V}{\lambda_{therm}^3} \propto \beta^{-3/2} \tag{39}$$

In this expression, λ_{therm} is the so-called thermal wave length,

$$\lambda_{therm} = \sqrt{\frac{\beta \cdot h^2}{2\pi m}} \tag{40}$$

With this detail knowledge of the s.p. translational partition function, one can now calculate the expectation (mean) value of a particle in a canonical system. From Equ. (28) one has

$$\langle \varepsilon \rangle = -\frac{\partial}{\partial \beta} \ell n\, z = -\frac{\partial}{\partial \beta} \ell n\, \beta^{-3/2} = \frac{3}{2} \cdot \frac{1}{\beta} \tag{41}$$

In Equ. (41) the fact has been used that there is only one factor in the function $z(\beta)$ that actually depends on the parameter β. Obviously, the mean energy of a free particle in a canonical system can, and has been, measured to be

$$\langle \varepsilon \rangle = \frac{3}{2} \cdot k_B \cdot T \tag{42}$$

In the development of thermodynamics, the mean kinetic energy of a free particle has been identified (by convention) with the product of Boltzmann constant k_B and **temperature T**. Therefore, one has to identify,

$$\beta = \frac{1}{k_B \cdot T} \tag{43}$$

The heretofore unknown model parameter β has now been linked to experimental observation. It is an inverse energy which at room temperature T=300K has the value,

$$\beta^{-1}(300K) = k_B \cdot 300K = 25\, meV = (1/40)\, eV \tag{44}$$

2. *Phase space probability distributions and H-Theorem*

Systems of N real particles occupy domains in *6N*-dimensional **phase space**, rather than cells of a CA. Phase space is a product space described by continuous *3N* spatial $\{\vec{q}_i, i = 1,...,N\}$ and *3N* momentum $\{\vec{p}_i, i = 1,...,N\}$ coordinates. Therefore, the probabilities p_i of discrete cells i discussed previously is replaced by continuous, time dependent (t) distribution functions $\{f(\vec{q}_i, \vec{p}_i, t), i = 1,...,N\}$ for the N particles. These functions are probability densities normalized to unity when integrated over the entire phase space,

$$\int d^3\vec{q}_i \int d^3 f(\vec{q}_i, \vec{p}_i, t) \equiv 1 \qquad \{i = 1,...,N\} \qquad (45)$$

Following the same line of arguments as before, the time dependent information content of an occupied multi-particle state is contained in the **Boltzmann H-function** (*eta*-function)

$$H(t) := \sum_{i=1}^{N} \int d^3\vec{q}_i \int d^3\vec{p}_i \ \{f(\vec{q}_i, \vec{p}_i, t) \cdot \ell n f(\vec{q}_i, \vec{p}_i, t)\} \leq 0 \qquad (46)$$

The H function is obviously equivalent to the negative of the information S given by the statistical entropy (cf. Equ.**Error! Reference source not found.**). It is negative since the distribution functions are probability densities.

Based on very general principles, predictions can be made as to the spontaneous time evolution of the H function, or the equivalent statistical entropy function $S.$ In the following, the entropy $S(t)$ is expressed as

$$S(t) = -H(t) = -k_B \sum_{n=1}^{\Omega} p_n(t) \ell n(p_n(t)) \geq 0 \qquad \sum_{n=1}^{\Omega} p_n(t) \equiv 1 \qquad (47)$$

in terms of time dependent (normalized) probabilities for discrete system states numbered by n.

This time dependence of the entropy function reflects an underlying dynamics, a transport process, which tends to redistribute the importance (or population) of the microscopic states and all of its attributes. The trend is equivalent to an entropy flux or current

$$j_s := \frac{dS}{dt} \tag{48}$$

If j_s has a finite magnitude, it defines a direction of increasing or decreasing diversity or spread in *a priori probabilities.*

The *a priori* probabilities p_n can be regarded as populations of these states which can be queried in experimental observations. If these populations are time dependent, there have to be microscopic transition probabilities w_{nm} connecting any state n and m. The transition probabilities describe the rate of change in the population of state n due to gain and loss from and to state m according to a balance "Master Equation,"

$$\frac{dp_n(t)}{dt} = \sum_m \left\{ \underbrace{w_{mn} \cdot p_m(t)}_{Gain} - \underbrace{w_{nm} \cdot p_n(t)}_{Loss} \right\} \tag{49}$$

For microscopic, quantal reasons, the transition probabilities are symmetric, $w_{nm} = w_{mn}$, which ensures time reversal invariance (detailed balance). Obviously, the Master Equation (49) is a classical approximation in that it neglects quantal interference terms involving transition amplitudes, rather than probabilities.

Now, the time derivative of the entropy function in Equ. (47), the entropy flux (Equ. (48)), can be calculated:

$$\frac{dS(t)}{dt} = -k_B \sum_{n=1}^{\Omega} \left\{ \left(\frac{dp_n(t)}{dt} \right) \ell n(p_n(t)) + p_n(t) \left(\frac{d\ell n\, p_n(t)}{dt} \right) \right\}; \quad \frac{d}{dt} \sum_{n=1}^{\Omega} p_n(t) \equiv 0 \tag{50}$$

Evaluating the derivatives one obtains

$$\frac{dS(t)}{dt} = -k_B \sum_{n=1}^{\Omega} \left\{ \left(\frac{dp_n(t)}{dt} \right) \ell n(p_n(t)) + p_n(t) \left(\frac{1}{p_n(t)} \frac{dp_n(t)}{dt} \right) \right\} =$$

$$\frac{dS(t)}{dt} = -k_B \sum_{n=1}^{\Omega} \left(\frac{dp_n(t)}{dt} \right) \ell n(p_n(t)) - k_B \underbrace{\sum_n \frac{dp_n(t)}{dt}}_{=0}$$

(51)

The last term drops out because of the conservation of total probability implied by Equ. (50). Now, inserting for dp_n/dt the expression given by the Master Equation (49), the second row in (51) reads,

$$\frac{dS(t)}{dt} = -k_B \sum_{n,m=1}^{\Omega} w_{mn} \cdot \{p_m(t) - p_n(t)\} \ell n(p_n(t))$$

(52)

Here, use has been made of the symmetry of the transition probabilities w_{mn}. Since the two indices n and m run over the same range, this expression can also be written as,

$$\frac{dS(t)}{dt} = -k_B \sum_{n,m=1}^{\Omega} w_{mn} \{p_n(t) - p_m(t)\} \ell n(p_m(t))$$

(53)

Taking the average of Equs. (52) and (53), a more symmetric expression is obtained fro the time rate of change of the entropy function:

$$\frac{dS(t)}{dt} = \frac{k_B}{2} \sum_{n,m=1}^{\Omega} w_{mn} \{p_n(t) - p_m(t)\} \left[\ell n(p_n(t)) - \ell n(p_m(t)) \right]$$

(54)

However, since $d\ell n(p)/dp > 0$, all terms in the sum are non-negative and therefore,

$$j_s = \frac{dS(t)}{dt} = -\frac{dH(t)}{dt} \geq 0$$

(55)

According to this derivation, the entropy S increases and the H function decreases in time, as long as the transition probabilities are finite, $w_{nm} = w_{mn} > 0$. The larger the differences between the populations p_i of different states are, the higher is the rate of entropy changes. When

$$p_n \approx const.; \quad n = 1, \ldots, \Omega \tag{56}$$

the S (or H) functions no longer change. The system described by such function has reached its asymptotic stationary state, also known as equilibrium state. ***This equilibrium state is characterized by maximum entropy corresponding to equal a priori probabilities p_n and chaotic dynamics***. While for a given theoretic model the expectation values of the functions S and H can be calculated exactly, there are also higher moments (fluctuations) to consider, since they depend on stochastic parameters, the probabilities p_n.

3. *Gibbs stability criterion for random states*

The situation of maximum entropy, where all accessible states are uniform and have equal *a priori* probabilities, is called "**equilibrium**." It will be shown further below how these information/entropy functions change in complex dynamical processes.

All systems where accessible states are not uniformly populated are in states of disequilibrium and have statistical entropies less than the maximum possible:

The equilibrium state is therefore defined by the variational condition

$$S(\vec{q}_i, \vec{p}_i) = S_{max}(\vec{q}_i, \vec{p}_i) =: S_{equ} \rightarrow \delta S(\vec{q}_i, \vec{p}_i) = 0 \qquad (57)$$

Here, δ stands for a variation with respect to the individual probability densities. Once a multi-particle system is in such an equilibrium state of maximum entropy, there is conceptionally **no net driving force** that would force it out of this state in one direction or another. However, such an equilibrium state can be either stable or unstable. Microscopically, there are always quantal fluctuations in all coordinates. Even systems presumably at rest show "zero-point fluctuations." In addition, physical particles move even classically from phase space cell to phase space cell, changing individual occupation probabilities (p_i or $f(\vec{q}_i, \vec{p}_i, t)$) instantaneously away from their respective equilibrium values. The magnitude of these fluctuations depend on their origin in classical or quantum dynamics. They may vary in size and follow a distribution in time or frequency (chance of occurrence). Therefore, the actual entropy at a given instant will reflect these fluctuations.

Connecting to discussions of stability in previous sections, one can obtain a stability criterion by studying the expansion of the entropy S of an actual system state about the equilibrium state ($\delta S = 0$),

$$S = S_{equ} + \delta S + \frac{1}{2} \delta^2 S + ... \approx S_{equ} + \frac{1}{2} \delta^2 S \qquad (58)$$

From this relation it simply follows that the state of maximum entropy is stable, only if fluctuations away from this state reduce the entropy,

$$\delta^2 S < 0 \qquad\qquad (59)$$

This "Gibbs" stability criterion has to be applied in specific cases to identify the stable equilibrium. Stable equilibrium states are attractors of complex system, as will be demonstrated in later sections.