Journal of Theoretical Physics

Founded and Edited by M. Apostol **98** (2005)

ISSN 1453-4428

On probabilities
(Lecture four of the Course of Theoretical Physics)

M. Apostol
Department of Theoretical Physics, Institute of Atomic Physics,
Magurele-Bucharest MG-6, POBox MG-35, Romania
email: apoma@theory.nipne.ro

Introduction. Our world is statistical. It is motion and numbers. Mechanical or quantal motion, both non-relativistic and relativistic, are only approximations to the statistical motion. Statistical numbers are probabilities.

Theory of gases. Probabilities occurred perhaps for the first time in our reasoning about the natural world, *i.e.* in our *Philosophiae Naturalis*, with Maxwell, who, around 1870, had to admit that atoms in gases are distributed over velocities v according to the $\exp(-const\cdot v^2)$ law, *i.e.* their number$dN \sim \exp(-const\cdot v^2)d\mathbf{v}$. Obviously, $dN/d\mathbf{v}$ is a density of probability. This is shocking, because we do not see any reason for it to be so. Moreover, it is more fundamental than even our atomistic concept of the natural world, because from atoms we cannot derive probabilities, but from probabilities we can conceive atoms as parts, or possibilities, of the matter.

Probability and frequency. Much longer before, around 1700, one of the Bernoulli's discovered that if something is going to happen with probability p, then it appears for q and only q times in N draws with the probability

$$f(p) = C_N^q p^q (1-p)^{N-q} \ . \tag{1}$$

This function has a remarkable property: it is peaked on the frequency $p_0 = q/N$ and dispersed around by $\sqrt{p_0(1-p_0)/N}$, which tells that for N large enough the probability p is precisely the frequency of occurrence q/N. It is called the "law of large numbers", and it is the foundation of the empirical probability, or the empirical foundation of probability, which made probabilities the basis of the scientific method. The probability (1) is called the binomial distribution. For $p = \nu/N$ it becomes Poisson's distribution $\nu^q e^{-\nu}/q!$ of very improbable events (like Bonaparte's soldiers kicked to death by mules), while expanding it around $p_0 = q/N$ we get the "normal" distribution $(1/\sigma\sqrt{2\pi})\exp[-(p-p_0)^2/2\sigma^2]$, where $\sigma = \sqrt{p_0(1-p_0)/N}$ is the dispersion. The latter is said to have been shown by Gauss, around 1800, while suspecting his baker of fouling him with the weight of his daily breads.

Chances. It seems that there have been existed, once upon a time, a monk, by name of Bayes, who would have written an *Essay toward solving a Problem in the Doctrine of Chances*, published in 1764 after his death. Bayes looked at the probability $P(AB/C)$ to have both A and B under condition C, and noticed that

$$P(AB/C) = P(A/BC)P(B/C) \ . \tag{2}$$

On the other hand,

$$P(AB/C) = P(BA/C) = P(B/AC)P(A/C) \ , \tag{3}$$

so that

$$P(A/BC) = P(A/C) \cdot \frac{P(B/AC)}{P(B/C)} \ , \tag{4}$$

which means that we can know how to update the prior probability $P(A/C)$ to the posterior probability $P(A/BC)$ under the occurrence of a new condition B. It appears that this Bayesian logic of probabilities is indeed the scientific method. It has been embraced and developed by Laplace around 1800, and in modern times by Jeffreys and Jaynes. It tells us that probability is not necessarily a frequency, and the chances depend on conditions. Usually, if we have a frequency f then $p(f)df$ is the probability to get it within df.

More, in order to get something that might be reasonable, we need more than "let data speak for themselves". Data never "speak for themselves". We need a prior reasonable knowledge, to update it with plausibility. In order to get a theory, we need another apriori theory! This is another thing Bayes taught us!

Indeed, let C be unspecified, and let us denote A by some x_i and let B be some α, such that $p_i = p(x_i)$and $p(x_i/\alpha) = p_i(\alpha)$. We have then

$$p_i(\alpha) = p(\alpha/i)p_i \ . \tag{5}$$

Suppose that we know nothing apriori about p_i. Then, they must be equal, as given by the maximum of the entropy

$$S = -\sum_i p_i \ln p_i \ ; \tag{6}$$

our state of knowledge is minimal at this stage, as measured by this Shannon entropy given above (it is called the principle of "insufficient reason"!). Let us introduce then a constraint α (like, for instance, imposing the condition of having a fixed mean energy of a statistical ensemble). We know then that the probabilites $p_i(\alpha)$ get non-uniform with respect to i, which, of course, means that we know more. Indeed, on the other hand, Bayes equation (5) tells that $p_i(\alpha) < p_i$, so the Shannon entropy (6) decreases, and we got more knowledge, as expected.

The most likely state. A statistical ensemble of N particles has a (density of) probability $\rho \sim 1/\mathcal{N}$ over its states in number of $\mathcal{N}$. It is convenient to introduce the additive entropy $\mathcal{S} = \ln \mathcal{N} = -\ln \rho$ for the multiplicative $\mathcal{N}$. At equilibrium, the entropy must be maximal, and, therefore steady. Consequently, it must be proportional to the constants of motion, like energy, for instance, $\mathcal{S} \sim \beta \mathcal{E}$, where β is some constant. Then, $\rho \sim e^{-\beta \mathcal{E}}$. Let us normalize this distribution (which is called the canonical distribution):

$$Z \sim \int d\mathcal{E} d\mathcal{N} \cdot e^{-\beta \mathcal{E}} = \int d\mathcal{E} \cdot e^{\mathcal{S} - \beta \mathcal{E}} \ , \tag{7}$$

where $\mathcal{N}$ is the number of states correspoding to $\mathcal{E}$, and entropy is labelled now by $\mathcal{E}$. If

$$\partial \mathcal{S} / \partial \mathcal{E} = \beta \tag{8}$$

for some $\mathcal{E} = E$ (and $\mathcal{S} = S$), then $\mathcal{S} - \beta \mathcal{E} = S - \beta E - |S''/2| \, (\mathcal{E} - E)^2$... and Z becomes

$$Z \sim e^{S - \beta E} \int d\mathcal{E} \cdot e^{-|S""/2|(\mathcal{E}-E)^2} = \sqrt{2\pi / \, |S''|} e^{S - \beta E} \ . \tag{9}$$

The prefactor may be dropped out, and define the canonical partition function (sum of states) as $Z = e^{-\beta F}$ and the free energy $F = E - \beta^{-1} S$. What is more important is that the distribution

$\rho \sim \exp\{-|S''/2|\,(\mathcal{E}-E)^2\}$ is higly peaked on E (because $S'' \sim 1/N$); it is close to the "micro-canonical distribution" $\sqrt{2\pi/S''}\delta(\mathcal{E}-E)$. This is why a statistical ensemble has a rather definite energy E, in spite of its (statistical) moving over so many states (or, precisley due to that): it is the thermodynamical energy, it is also the mean energy, and the corresponding thermodynamical state, reached at equlibrium by so many (statistical) movements (and perhaps such a long time), is the most likely state. Equation (8) defines temperature $T = \beta^{-1}$, and the fluctuations in energy are $\delta E \sim 1/\sqrt{|S''|} = T\sqrt{c}$, where c is the heat capacity. A similar analysis holds for the grand-canonical ensemble, telling that the probability is peaked on the number of particles N, and $c \sim N$, and so does energy E. And the relative fluctuations go like $1/\sqrt{N}$.

If the statistical equilibrium exists, then it is the most likely; or it does exist precisely because it is the most likely. And this is true for large ensembles, with large energies, large volume, etc, *i.e.* with a large number of states. Because, if the number of states is large enough, we ask for the ensemble to be in one subset of an equally large number of states (as, for instance, to have energy E, which is proportional to this number of states for a given temperature), so it is not very surprising that equilibrium does exist. As it may equally be read from fluctuations. On the contrary, for low temperatures, or low energies, or small size, etc, the relative fluctuations grow indefinitely, and equlibrium is not attained anymore, and this non-thermodynamical limit is thermodynamically meaningless.

Apart from being an extensive function of number N of particles and volume V (as seen from the summation over states), the thermodynamical energy E is also a function of entropy S, as seen from (8), while the latter is obviously $S = -\sum \rho \ln \rho$; and TdS is heat. All this thermodynamics as the most likely behaviour of the statistical ensembles has clearly been made explicit by Boltzmann up to around 1900.

Moreover, the entropy as the logarithm of the number of states can be made explicit simply. For the classical occupation numbers $n = N/\mathcal{N} \ll 1$, the number of states is $\prod \mathcal{N}^N/N!$...,so that $S = -\sum n \ln(n/e)$. For fermions with $\mathcal{N}$ states out of which N are occupied and $\mathcal{N}-N$ are empty, the number of states is $\prod \mathcal{N}!/N!(\mathcal{N}-N)!$, so that $S = -\sum[n \ln n + (1-n)\ln(1-n)]$, for the mean occupation number $n = N/\mathcal{N}$. For bosons we distribute N particles and $\mathcal{N}-1$ "walls" among $\mathcal{N}$ states, so the number of states is $\prod(N+\mathcal{N}-1)!/(\mathcal{N}-1)!N!$, leading to $S = \sum[(n+1)\ln(n+1) - n\ln n]$ for the mean occupation number $n = N/\mathcal{N}$. Statistical distributions for the mean occupation number can be obtained from the maximum of these entropies under constraints of energy, number of particles, etc.

Motion through probabilities. Around 1905 Einstein realized that the statistical motion proceeds by probabilities, and this probabilistic motion is the origin of heat, as another component of energy. Indeed, from $\sum \exp(c - \beta\mathcal{E}) = 1$, where $Z = e^{-c}$, we have straightforwardly $dc - \beta(\partial E/\partial\lambda)d\lambda - Ed\beta = 0$, or $d(\beta E - c) = \beta dE - \beta(\partial E/\partial\lambda)d\lambda = \beta dQ$, so $dE = (\partial E/\partial\lambda)d\lambda + dQ$, where Q is heat and λ is any other parameter. Moreover, since $S = -\sum \rho \ln \rho = \beta E + \ln Z = \beta E - c$, it follows $dS = \beta dQ$, as a total differential. Similarly, $\sum(E-\mathcal{E})e^{-\beta\mathcal{E}} = 0$ leads to $\partial E/\partial\beta = E^2 - \bar{E}^2$, so that the energy fluctuations are $\delta E \simeq \sqrt{|\partial E/\partial\beta|}$, or $\delta E \simeq \sqrt{1/\,|\partial^2 S/\partial E^2|}$, since $dS = \beta dE$, for constant parameters. In general, since probability goes like e^S, the fluctuations go like $1/\sqrt{|S''|}$.

In addition, since the statistical probability is labelled by states, the statistical motion is assignable to every state of the ensemble, in particular to particles, or excitation quanta. Indeed, according to (8), temperature T is a scale energy, corresponding to equilibrium, and time $\tau_{th} \sim \hbar/T$ is time over which equilibrium is reached. It must be much shorter than the quantal time $\tau_q \sim \hbar/\hbar\omega$, where ω is a characteristic frequency of quantal states. For statistical ensembles, such frequencies are very low, so the condition $\tau_{th} \ll \tau_q$ is easily fulfilled. For non-ideal ensembles, the quantal

states are elementary excitations, and the condition reads $\tau_{th} \ll \tau_{lf}$, where τ_{lf} is the lifetime of these excitations. For instance, quasiparticle lifetime for a Fermi liquid is $\sim \hbar/(T^2/\mu)$, where μ is the Fermi energy. A special situation pertains to the classical gas, where the state energy corresponds to the particle energy $\hbar^2/ma^2$, where m is the particle mass and a represents the mean inter-particle distance. However, the condition $T \gg \hbar^2/ma^2$, *i.e.* $\tau_{th} \ll \tau_q$, is precisely the condition of equilibrium in this case. The role of the elementary excitations are played now by colliding particles, whose lifetime is longer by a factor a^2/σ than the "quantal" time, where σ is the cross-section. Therefore, inequalities $\tau_{th} \ll \tau_q \ll \tau_{lf}$ must be obeyed.

Fluctuations have their own time scale τ_{flcs}. Indeed, the fluctuating energy is $\delta e \sim T\sqrt{c}$, where c is the heat capacity (per particle), so the fluctuation time may be written as $\tau_{flcs} \sim n\hbar/\delta e$, where n is a undetermined number of quanta of action. It is easy to see that $\tau_{th} \ll \tau_{flcsq} \ll \tau_q \ll \tau_{lf}$. Similarly, the mean inter-particle distance fluctuates by some a, such that $a^2 \sim (1/9v^{4/3})(\partial^2 s/\partial v^2)^{-1}$, where s is the entropy per particle volume v. It is about the mean inter-particle distance. Momentum fluctuates also by δp, given by $(\delta p)^2 \sim mT$, etc.

Probability goes in time and space by fluctuations. In the simplest form, the particle density obeys

$$\partial n/\partial t = \frac{1}{\tau}[n(x-a) - n(x) - n(x) + n(x+a)] = (a^2/2\tau)\partial^2 n/\partial x^2 \ , \tag{10}$$

where a and τ are the fluctuating distance and time. This is the diffusion equation, whose solution is $n \simeq (N/\sqrt{4\pi Dt})e^{-x^2/4Dt}$, for an original peak $N\delta(x)$, where N is the number of particles per unit area and $D = a^2/2\tau$ is the diffusion coefficient. Since $D = T/6\pi a\eta$, whereη is the viscosity, and $a^2/2\tau \simeq a^2T\sqrt{c}/2\hbar n$, it is easy to see that the viscosity per unit mass density is $\eta a^3/m \sim n(h/m)$, so that h/m may be viewed as quanta of viscosity.

Equation (10) is, in fact, more general. Momentum can be added for completely characterizing classical states, and, in the presence of transport velocities and external forces, equation (10) becomes Boltzmann's kinetic equation with the right-hand side, writen as $\delta n/\tau$, the collision integral. Boltzmann's H-theorem for the increase of entropy $-\sum \rho \ln(\rho/e)$ is then easily proven. Similarly, the master equation for quantal evolution of the probability density under the action of transition probabilities per unit time is easily reducible to Boltzmann's equation written in this form, on the ground of the quasi-classical description. The approach to equilibrium and the transport (which proceeds over longer distance and time scales, those of the collision time, lifetime and mean free path) are thereby governed by such a diffusion equation of the form (10), which may be called Einstein's kinetic equation. Originally, Einstein employed it for the Brownian motion. Fluctuations and dissipation go by this equation.

Probability waves. Perhaps the most general motion for detemined quantities like $\mathbf{r}$ and t proceeds by

$$\mathbf{k}d\mathbf{r} - \omega dt = d\Phi \ . \tag{11}$$

Though not necessary, $\mathbf{r}$ may be position and t may denote time, so that the wavevector $\mathbf{k} = grad\Phi$ is perpendicular to $\Phi = const$ and $k = 2\pi/\lambda$ is then the inverse of a wavelength λ, while the frequency $\omega = 2\pi/T$ is the inverse of a period T. Obviously, Φ is the phase of a wavefunction $\psi \sim e^{i\Phi}$, and there must be a universal constant of action $\hbar$, such that the mechanical action is $dS = \hbar d\Phi$, and the momentum $p = \hbar k = h/\lambda$ is quantized, the energy $\varepsilon = \hbar\omega = h/T$ is quantized, since $S \sim 2\pi n\hbar$, where n is an integer. The wavefunction reads then $\psi \sim e^{iS/\hbar}$, and we may start the description of the quantal motion, with Planck's constant $\hbar$, de Broglie's quantization of the momentum and Einstein's quanta of energy.

The wavefunction ψ may be expanded in plane waves, $\psi = \sum c(\mathbf{k})e^{i\mathbf{kr}}$, or, for sets of determined coefficients $c(\mathbf{k})$, the wavefunction ψ may also be expanded in orthogonal wavefunctions φ_n, $\psi =$

$\sum(\varphi_n,\psi)\varphi_n$, where (φ_n,ψ) is the scalar product of ψ by φ_n. Obviously, the scalar product (φ_n,ψ) is the content of φ_n in ψ, it tells how much of $\varphi_n$is contained in ψ, and, by virtue of the normalization of the wavefunctions, $\sum|\psi|^2=\sum|(\varphi_n,\psi)|^2$, its square $|(\varphi_n,\psi)|^2$ may be viewed as the probability of getting φ_n in ψ. Obviously, (φ_n,ψ) is a wavefunction of φ_n, so the square of wavefunctions is density of probability. Thus, $\psi(\mathbf{r})=\int d\mathbf{r}'\cdot\psi(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')$, so $|\psi(\mathbf{r})|^2\,d\mathbf{r}$ is the probability of having $\mathbf{r}$ in this wavefunction $\psi(\mathbf{r})$. Because in $\psi(\mathbf{r})$ position $\mathbf{r}$ is not determined, only ψ is determined, and, in this sense, the quantal motion, or the waves motion, is not a complete description of reality. Indeed, in the plane wave $e^{i\mathbf{kr}}$ the wavevector $\mathbf{k}$ is determined, position $\mathbf{r}$ is not determined. Moreover, the variation $\delta\Phi=\mathbf{k}\delta\mathbf{r}+\mathbf{r}\delta\mathbf{k}+\delta\mathbf{k}\delta\mathbf{r}=d\Phi+\delta\mathbf{k}\delta\mathbf{r}$ becomes $\delta\Phi=\delta\mathbf{k}\delta\mathbf{r}$ on $\Phi=const$, so that $\delta\mathbf{k}\delta\mathbf{r}\sim 2\pi n$, which means , $\delta\mathbf{k}\delta\mathbf{r}>\pi$ at least, in order to have meaningful values for both $\mathbf{k}$ and $\mathbf{r}$. This holds also for frequency and time, $\delta\omega\delta t>\pi$, and may be called the uncertainty of the phase variables in waves. It is worth noting that if the wavelength is small enough, much smaller than the characteristic length of the movement, then we may have a reasonable accuracy for both wavevectors and positions, and say then that we are in the quasi-classical limit, or approximation, or description. Similarly, if the position is much sharply defined over a range which is much smaller than the wavelengths, then again we are in the quasi-classical limit. In both cases $\delta k\delta x\gg\pi$, *i.e.* the phase varies over a large range of cycles, and the mechanical action varies over a much larger range than Planck's constant, and we may let then $\hbar$ formally go to zero. It is also worth noting that this waves uncertainty acts independently upon each spatial coordinate, so we may have a sharp localization on two coordinates, and a small wavelength along the third, which is a ray, as obtained from a small aperture of size d, where $\lambda\ll d$. We may note that in such quasi-classical limit the wavevector, or wavelengths, vary slowly in space, while abrupt variations bring about big changes in phase, which can only be accommodated by letting the amplitude of the wavefunction changing, which amounts formally to let phase become imaginary.

In order to see how much of a quantity f is contained in a wavefunction ψ we should act with an operator f on that wavefunction. For instance, if $-i\partial/\partial\mathbf{r}$ acts upon $e^{i\mathbf{kr}}$ we get $\mathbf{k}e^{i\mathbf{kr}}$, and may say that the wavevector is determined in the wavefunction $e^{i\mathbf{kr}}$, having the value $\mathbf{k}$. It may happen that φ_n is an eigenfunction for the operator f, *i.e.* $f\varphi_n=f_n\varphi_n$, and then we say that f has a well-determined value f_n on that wavefunction, which is an eigenvalue of f. But usually $f\psi=\sum c_nf\varphi_n=\sum c_nf_n\varphi_n$, so that f is not determined, in the sense that it may be any f_n, with probability $|c_n|^2$, so that its average is $(\psi,f\psi)=\sum|c_n|^2f_n$, and during this measurement of f we may reduce the wavefunction to any φ_n, with a probability, thus disturbing the original wavefunction. Wavefunction ψ itself may be the eigenfunction for another quantity g, and these two quantities f and g are not simultaneously well-determined, as long as they do not commute. They are represented, in general, by matrices, of the form $f_{nm}=(\varphi_n,f\varphi_m)$, and they must be hermitian, *i.e.* f transposed and conjugate f^{t*} must be equal to f in order to have real eigenvalues. A principle of Heisenberg's uncertainty of the form $\delta f\delta g>finite$ holds for them, as for instance, $(\psi,(\delta f-i\lambda\delta g)(\delta f+i\lambda\delta g)\psi)>0$ for any λ, so that $\delta f\delta g>|C|/2$, where $C=[f,g]$ is their commutator. Momentum $\mathbf{p}=-i\hbar\partial/\partial r$ does not commute with its canonical-conjugate operator of position $\mathbf{r}$, *i.e.* $[p_x,x]=-i\hbar$, so that $\delta p_x\delta x>\hbar/2$, etc. The most complete description is attainable by wavefunctions which are eigenfunctions of the most complete set of commuting operators, and again such a description is not a complete one, in principle.

If energy is going to have determined values, then they must be eigenvalues of $i\hbar\partial/\partial t$, since, indeed, this operator gives energy $E=\hbar\omega$ when acting upon the plane wave. On the other hand, it may be represented as the hamiltonian $H=p^2/2m+V$, for instance, for a particle of mass m moving in the potential V, and Schroedinger's equation $H\psi=E\psi=i\hbar\partial\psi/\partial t$ leads to a wavevector dependence $\omega(\mathbf{k})$ of frequency, like any other equation.

Let us consider such a superposition of plane waves

$$\psi(\mathbf{r},t)=\int d\mathbf{k}\cdot c(\mathbf{k})e^{i(\mathbf{kr}-\omega t)} \tag{12}$$

where ω is a function of $\mathbf{k}$. Usually, the coefficients $c(\mathbf{k})$ are localized over a certain range around a certain wavevector, depending on the initial condition at $t=0$. We may assume that $c(\mathbf{k})$ are uniformly distributed, so that at $t=0$ the wavefunction $\psi(\mathbf{r},0)=\delta(\mathbf{r})$ is localized as a δ-peak on $\mathbf{r}=0$. If ω would be linear in $\mathbf{k}$, then ψ would move as a δ-peak with phase velocity ω/k. It follows that the main contribution to ψ comes from the linear part of ω, so that we expand ω around some $\mathbf{k}_0$, which may be taken zero, for simplicity. The expansion reads $\omega=\omega_0+vk+\omega"_{ij}k_ik_j/2...$, where

$$\mathbf{v}=\partial\omega/\partial\mathbf{k} \tag{13}$$

is a velocity. We further assume that the tensor $\omega"_{ij}$ is brought to the principal axes, so that we may estimate the contribution to the wavefunction for each coordinate separately. It reads

$$\psi(x,t)=\int dk\cdot e^{ik(x-vt)-i\omega"k^2t/2} \tag{14}$$

which is readily estimated as

$$\psi(x,t)=\sqrt{2\pi/i\omega"t}e^{-i\frac{(x-vt)^2}{2|\omega"|t}}\ . \tag{15}$$

Such an oscillating diffusion is related to Fresnel's stationary phase, or Debye's steepest descent. It follows that the wavefunction looks like wavepackets, which are localized (periodically) on $x=vt$ over a spread $\delta x\sim\sqrt{|\omega"|t}$, propagate with a group velocity v, which is dispersive, *i.e.* each wave packet propagates with its own group velocity, depending on its central wavevector, and the wave packets flatten in time and oscillate slower. Obviously, $|\omega"|t\sim 1/\delta k^2$, so that the larger δk the sharper the localization, *i.e.* the uncertainty $\delta x\delta k\sim 1$. Moreover, since $x=t\partial\omega/\partial k$, the frequency may be given a space dependence, and the wavevector may be given a time dependence, such that $\partial k/\partial t=\partial\omega/\partial x$, and, in general,

$$\partial\mathbf{k}/\partial t=-\partial\omega/\partial\mathbf{r}\ , \tag{16}$$

i.e. the frequency may act as a hamiltonian for the Hamilton-Jacobi equations of motion. Under such circumstances, waves behave like particles (or quasi-particles, their lifetime being $\tau\sim\omega"/v^2$), and particles are waves, or quasi-waves. Everything with probabilities, certain uncertainty, and, of course, incompleteness.

Power laws. Suppose that something occurs repeatedly with an average time t in a long duration T. The number of occurences is then $N=T/t$. Similarly, the total number may be taken as $N_0=T/t_o$, where t_o is a characteristic time (actually, it is $(T/t_0)\ln(T/t_0)$, which amounts to renormalize the threshold time to $\widetilde{t_0}=t_0/\ln(T/t_0)\to 0$ for $T\to\infty$, such that the results are in fact independent of arbitrary cutoff time). The frequency is then $N/N_0=t_0/t$, and the probability is $-d(N/N_0)=(t_0/t^2)dt$. Suppose further that each N set is characterized by some size S. For large scales of time and size, the logarithms $\ln S$ and $\ln t$ vary slowly, so it is reasonable to say that

$$d\ln S/d\ln t=1/r\ , \tag{17}$$

where r is a constant. It follows $t/t_0=(S/S_0)^r$, where $S_0$is a threshold size. The probability becomes $-d(N/N_0)=(rS_0^r/S^{1+r})dS$, or, if we denote $p=-dN/N_0dS$ and $\alpha=1+r$,

$$p=(\alpha-1)S_0^{\alpha-1}/S^{\alpha}\ , \tag{18}$$

or

$$\ln p = const - \alpha \ln S \ . \tag{19}$$

Such power-law distributions of probability seem to be ubiquitous. They seem to be present in word frequency ($\alpha \sim 2.2$), papers citations ($\alpha \sim 3$), web hits ($\alpha \sim= 2.4$), books sold ($\alpha \sim 3.5$), telephone calls ($\alpha \sim 2.2$), craters on the moon ($\alpha \sim 4$), solar flares ($\alpha \sim 1.8$), wars ($\alpha \sim 1.8$), worth of the people ($\alpha \sim 2$), family names ($\alpha \sim 2$), cities distributed over size ($\alpha \sim 2.3$). Sometimes $\ln S$ may be viewed as a magnitude M, $\ln S \sim M$, and, in this respect, the earthquakes are distributed with $\alpha \geq 1$, or $\alpha \sim 2$, for the biggest ones (size being in this case the released seismic energy). There is no scale probability in such power laws, simply the probability distributes over a large range. It is noticed that beta function $B(S,\alpha) = \Gamma(S)\Gamma(\alpha)/\Gamma(S+\alpha)$, where Γ is the gamma function, goes like $S^{-\alpha}$ for large α.

Power-law distributions may also be got by some specific mechanisms. For instance, let a random walk of step length a (like that of a drunken sailor). The position after N steps is $r_N = r_N - r_{N-1} + r_{N-1} - r_{N-2} + ... = \sum s_n$, where $\bar{s}_n = a$, $\overline{s_n^2} = a^2$ and $\overline{s_n s_m} = 0$. The distance after N steps is given by $\overline{r_N^2} = \overline{(\sum s_n)^2} = Na^2$, so it goes like square root $\sqrt{t}$ of time, since $N \sim t$. Let a random walk along an axis, by two distinct steps only (one upwards, other downwards), and let u_{2n} be the probability of crossing the axis after $2n$ steps. Let f_{2m} be the probability of crossing the axis for the first time after $2m$ steps, so that $u_{2n} = \sum_1^n f_{2m} u_{2n-2m}$, where $u_0 = 1$ and $f_0 = 0$. Generating functions $U(z) = \sum u_{2n} z^n$ and $F(z) = \sum f_{2n} z^n$ give immediately $F(z) = 1 - 1/U(z)$. On the other side, $u_{2n} = C_{2n}^n / 2^{2n}$, so that $U(z) = 1/\sqrt{1-z}$, and $F(z) = 1 - \sqrt{1-z}$, whose expansion coefficients are $f_{2n} = C_{2n}^n/(2n-1)2^{2n}$. For large n, with Stirling's $\ln n! = n \ln n - n + (1/2)\ln n$, we get $f_{2n} \sim \sqrt{2/n(2n-1)^2}$, which amounts to a probability $\sim 1/t^{3/2}$ of zeroing for the first time in time t. It gives the lifetime of a gambler's ruin process. A maximal winning streak, or losses streak, in N tries is $\sqrt{N}$.

Suppose an ensemble with a certain distribution of sizes s. Obviously, the probability for s is a function of s/a, where a is a scale size, like unit of measure. The ensemble is also characterized by, say, its mean size $\bar{s}$, so the probability is $p(s) = Cf(s/a, \bar{s}/a)$. Changing the unit a will not change the probability, so that $p(s) = C' f(s/\lambda a, \bar{s}/\lambda a)$, which, however, belongs to another ensemble characterized by $\bar{s}/\lambda$. The ensemble rests the same only for $\bar{s} \to \infty$, so that the probability reads then $p(s) = C' f(s/\lambda a, \infty) = (C"/C)p(s/\lambda)$, which amounts to

$$p(bs) = g(b)p(s) \ . \tag{20}$$

This scaling is specific to the phase transitions, where the mean size percolates to infinity at the critical point. The power law is easily obtained from (20), since $p(b) = g(b)p(1)$, *i.e.* $p(bs) = [p(b)/p(1)]p(s)$, and $sp'(s) = [p'(1)/p(1)]p(s)$, hence $p(s) = p(1)s^{p'(1)/p(1)}$. Formally, solution of (20) contains also a contribution which is periodic in $\ln s$. A large number of natural phenomena may be viewed as being always close to their critical point, oscillating around it, which is a self-organized criticality, their probability distribution following a power law. It is striking the similarity of the scaling equation (20) with Bayes' equations (4) or (5). It is also worth noting that a second-order term to (17) may produce a log-normal distribution, *i.e.* a normal distribution in logarithms. It seems that the bird species do obey such a distribution, for instance.

Finally, let us note one of the most convenient fitting method. Suppose that $f_\alpha(x)$ is a function that may fit a set of n data x_i. A probability may be defined as $p_\alpha(x) = f_\alpha(x)/I_\alpha$, where $I_\alpha = \int f_\alpha(x)dx$, so that the fit probability (or likelihood) is $\prod p_\alpha(x_i)$. The extrema of $\ln \prod p_\alpha(x_i) = \sum \ln p_\alpha(x_i) = \sum \ln f_\alpha(x_i) - n \ln I_\alpha$ give the most likely values of the fitting parameter α. Obviously, the method is based on Bayes theory.