Bose-Einstein and Fermi-Dirac Distributions and Their Use in Biological Sciences

Monica G. HĂŞMĂŞANU1, Sorana D. BOLBOACĂ2, Lorentz JÄNTSCHI2,*

1 Iuliu Hațieganu University of Medicine and Pharmacy, Department of Neonatology, 3-5 Clinicilor Str., 400006 Cluj-Napoca, Romania; 2 Iuliu Hațieganu University of Medicine and Pharmacy, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania; 3 AcademicDirect, 103-105 Muncii Blvd., 400641 Cluj-Napoca, Romania. E-mail: lorentz.jantschi@gmail.com

Abstract
The aim of this paper was to review and summarize the usefulness of Bose-Einstein and Fermi-Dirac distributions in biological sciences. Starting with the introductory presentation of these distributions, the paper presents their rationale, formulas and applications. Furthermore, the increase interest of applying Bose-Einstein and Fermi-Dirac statistics in nowadays life science researches has also been highlighted.

Key words
Bose-Einstein distribution; Fermi-Dirac distribution; population statistics

INTRODUCTION
The Bose-Einstein statistic started to find its applications to biological sciences (Fröhlich, 1968a; Fröhlich, 1970.) since 1968 when Fröhlich (Fröhlich, 1968b) showed how a driven collection of vibrational oscillators could achieve a highly ordered non-equilibrium state, a property usually compared to phenomena involving macroscopic quantum coherence like as the Bose-Einstein condensation. Such a condensation would have an important influence on the dynamical properties of systems, and there has been considerable interest in finding its applications (Reimers et al., 2009). Following Fröhlich, the authors of (Lauck et al., 1992) considered a system that models a biological structure represented by a long chain of proteins possessing polar modes of vibration on which energy is pumped through metabolic processes. A theory of relaxation based on the non-equilibrium statistical operator method was used in the derivation of the kinetic equations to introduce non-linearity due to interactions of the polar vibrations with the carriers and with a thermal bath. Non-linearity arising from high order relaxation processes lead to the emergence of the Fröhlich effect in the polar modes, the occurrence of a (non-equilibrium) Bose-Einstein-like condensation.

Shortly after Fröhlich’s paper, a series of distributions associated to Bose-Einstein statistics (defined as one of two possible ways that a collection of non-interacting indistinguishable particles may occupy a set of available discrete energy states) were revised in a more general path - distributions giving the probability of a new occurrence of an event proportional to the number of times it had previously been occurred - and under this more general frame some application on modeling the city sizes and growth were identified (Ijiri and Simon, 1975). The main idea behind the use of the Bose-Einstein, Fermi-Dirac and Maxwell-Boltzmann statistics is that the cells can be seen as urns towards the molecules of a given substance and therefore the humoral laws deducted from the cellular laws may be used to interpret the evolution of molecular metabolism with time (Boutros-Toni and Duhamel, 1972). Dipolar elements contained in biological systems
that oscillate in a frequency of $10^{11} - 10^{12}$ s$^{-1}$ were found as possible channel for various life processes such as cell division and protein synthesis, driving the system under certain conditions as a Bose-Einstein condensation into the lowest energy state (Wu and Austin, 1978). Plant metabolism responses to a simple plant growth model with three resources (light, water and nutrients) were described with continuous-time Markov chains when the effects of varying the used resource and tolerance on optimum growth were successfully modelled with a Fermi-Dirac distribution function (Olson et al., 1985). Bose-Einstein and Maxwell-Boltzmann statistics were compared in efficiency of developing stochastic models of arthropod populations (empirical data for support are presented for many models by Young and Willson (1987)). Patterns of organelles within cells, for example, invertebrate smooth muscle filaments, microtubules in axons and micropinocytic vesicles in capillary endothelial cells, found to occur non-randomly within cells, were studied and has been suggested that Bose-Einstein statistics could be of significant value in this context (James, 1989). The origin of the chirality of protein amino acids from the point of view of a phase transition from a racemic mixture into an optically pure state were studied under assumption that Bose-Einstein condensation may act as an amplification mechanism and the results were previously presented by Chela-Flores (1994).

### THE RATIONALE OF FERMI-DIRAC AND BOSE-EINSTEIN DISTRIBUTIONS

Considering an enzymatic reaction at equilibrium (Eq.1):

\[
\begin{align*}
E + S \rightarrow & \quad ES, \, v = k_1[E][S] \\
E + S \leftarrow & \quad ES, \, u = k_2[ES] \\
E + S \leftrightarrow & \quad ES, \text{ equilibrium (} u = v \text{) } \rightarrow k_{12} = [ES]/(E)[S]), \text{ where } k_{12} = k_1/k_2
\end{align*}
\]

where $E$ is the enzyme; $S$ is the substrate; $ES$ is the enzyme-substrate complex, and $u$ and $v$ are the corresponding reaction rates with the coefficients $k_1$ and $k_2$.

The ratio of linked substrate/enzyme is (Eq.2):

\[
\begin{align*}
\frac{f_S}{[S]+[ES]} = & \quad \frac{1}{[S]} + 1, \quad \frac{[ES]}{[E]+[ES]} = \frac{1}{k_{12}[E]} + 1, \quad \frac{f_E}{[ES]} = \frac{[ES]}{[E]+[ES]} = \frac{1}{k_{12}[S]} + 1
\end{align*}
\]

By using the well known the relationship between equilibrium constant and Gibbs free energy ($\ln k_{12} = -\Delta G^*/RT$), result the Eq.3:

\[
\begin{align*}
\frac{f_S}{\frac{1}{k_{12}[E]} + 1} = \quad e^{\Delta G^*/RT-ln([E])}, \quad \frac{f_E}{\frac{1}{k_{12}[S]} + 1} = \quad e^{\Delta G^*/RT-ln([S])}
\end{align*}
\]

It is easy to be recognized now the Fermi-Dirac function. At a given temperature and a given concentration, the standard free enthalpy of binding ($\Delta G^*$) vary and this enthalpy for an enzyme depends on a series of environmental factors and therefore we may see the process of site-binding as a process shaped probabilistically by a Fermi-Dirac function.

Even simpler is to reveal the potential of Fermi-Dirac distribution to DNA and RNA sequences. Thus, if the chain of amino acids is seen as a carrier of the genetic information, and let $n$ be its
size, then it is easy to imagine that some amino acids (let abbreviate with \( r \) their number) are not carriers of the information, due to misplacement. The distribution of misprints corresponds to a distribution of \( r \) balls in \( n \) cells with no cell containing more than one ball. It is therefore reasonable to suppose that, approximately, the misprints obey the Fermi-Dirac statistic. If we look for a specific amino acid and we count its occurrences in a certain chromosome, we may extend this counting to the whole space of chromosomes. Then, the chromosomes are distinguishable one to each other and are filled with same (undistinguishable) specific amino acid under the observation. Thus, the distribution of the amino acids in chromosomes corresponds to a distribution of \( r \) balls in \( n \) cells. We must take into account that it is required about same energy of binding of a certain amino acid to a chromosome, and therefore is very likely that the distribution of amino acids among chromosomes to follow the Bose-Einstein statistic.

\[
A(n, b) = \frac{(n + b - 1)!}{n!(b - 1)!} \rightarrow A(n, f) = \frac{(n + n/f - 1)!}{n!(n/f - 1)!}
\]  

(4)

Checking of the above relation can be conducted noting that: \( A(n, 1) = 1 \) (all particles are in one box), \( A(n, 2) = n + 1 \) (the first box is from 0 to \( n \) particles, remainder in the second), \( A(0, 3) = \{000\} = 1, A(1, 3) = \{001, 010, 100\} = 3, A(2, 3) = \{002, 011, 020, 101, 110, 200\} = 6, A(3, 3) = \{003, 012, 021, 030, 102, 111, 120, 201, 210, 300\} = 10 \). A program to generate for \( \{'b'\} \) boxes given all the numbers represented in the \( \{'n'\} \) +1 (with ‘digits’ between 0 and \( n \)), represented in exactly the \( \{'b'\} \) memory units can be made and should count how many of these exact figures have the sum of digits ‘\( n'\).’ Considering that there are \( J (j = 1, 2, ..., J) \) energy states \((\varepsilon_1 < \varepsilon_2 < \varepsilon_j \leq \varepsilon_1)\), each containing a number of boxes \((b)\), and each populated by a number of molecules \((n)\), then the number of arrangements \(W\) is the product of the number of arrangements of each individual states (filling of two energy states \(\varepsilon_j < \varepsilon_{j'}\) each of \( n \) and \( n_j \) molecules are independent of one another) is given by Eq.5:

\[
W = \prod_{j=1}^{J} A(n_j, b_j) = \prod_{j=1}^{J} \frac{(n_j + n_j/f_j - 1)!}{n_j!(n_j/f_j - 1)!}, \quad \ln W = \sum_{j=1}^{J} \ln \left( \frac{(n_j + n_j/f_j - 1)!}{n_j!(n_j/f_j - 1)!} \right)
\]  

(5)

For the above relation observables are \( n_j \) - number of particles occupying each energy state \( (\varepsilon) \) as compared to the natural tendency at equilibrium function \( W \) reaches its maximum at the given number of particles \( N = \Sigma n_j \) and energy \( E = \Sigma n_j \varepsilon_j \). Function \( W = W (n_j, ..., n_j) \) reaches its maximum together with its logarithm. The reason for the switch from \( W \) to \( \ln W \) is that it is more convenient to work in logarithmic scale when operating with large numbers (the number of molecules in a system is a large number). Transforming the constrained maximum \( (N = \Sigma n_j = \text{constant}; E = \Sigma n_j \varepsilon_j = \text{constant}) \) on \( \ln W \) into unconstrained one (Lagrange’s method (Lagrange, 1811)) on \( \ln W1 = \ln W + \alpha (N-\Sigma n_j) + \beta (E-\Sigma n_j \varepsilon_j) \), results Eq.6:
\[ \ln W_1 = \ln W_1(n_1, \ldots, n_j) = \sum_{j=1}^{j} \ln \left( \frac{(n_1 + n_j / f_k - 1)!}{n_j!(n_j / f_k - 1)!} \right) + \alpha \left( N - \sum_{j=1}^{j} n_j \right) + \beta \left( E - \sum_{j=1}^{j} n_j \cdot e_j \right) = \max \quad (6) \]

Derivatives at the point of maximum it cancels and establishes a set of relationships between the number of states specific to each energy level and population of the levels with molecules (Eq.7):

\[ \frac{\partial \ln W_1(n_1, \ldots, n_j)}{\partial n_k} = \frac{\partial}{\partial n_k} \ln \left( \frac{(n_k + n_k / f_k - 1)!}{n_k!(n_k / f_k - 1)!} \right) + \alpha(0-1) + \beta(0 - e_k) = 0 \quad (7) \]

For large numbers, \( \log(n!) \approx n \cdot \log(n) - n \) is a good approximation (Stirling's approximation (Stirling, 1730)) and \( \partial(n \cdot \log(n) - n) / \partial n = \ln(n) \) simplifies the relationship given in Eq.7 as it is presented in Eq.8:

\[ \frac{\partial}{\partial n_k} \ln \left( \frac{(n_k + n_k / f_k - 1)!}{n_k!(n_k / f_k - 1)!} \right) = \frac{1 + f_k}{f_k} \ln(n_k + f_k - 1) - \ln(n_k) - \frac{1}{f_k} \ln(n_k + f_k) - f_k \ln(n_k / f_k) \to \]

\[ \frac{\partial \ln W_1(n_1, \ldots, n_j)}{\partial n_k} = \frac{1 + f_k}{f_k} \ln(n_k + f_k - 1) - \ln(n_k) - \frac{1}{f_k} \ln(n_k + f_k) - f_k \ln(n_k / f_k) \equiv 0 \quad (8) \]

For \( n_k(1+f_k)/f_k \gg 1 \) and \( n_k/f_k \gg 1 \) the relationship becomes (Eq.9):

\[ \alpha + \beta e_k = \frac{1 + f_k}{f_k} \ln(n_k + f_k) - \ln(n_k) - \frac{1}{f_k} \ln(n_k + f_k) \to \]

\[ \alpha + \beta e_k = \frac{1 + f_k}{f_k} \ln\left(\frac{1 + f_k}{f_k}\right) - \frac{1}{f_k} \ln\left(\frac{1}{f_k}\right) \quad (9) \]

Distribution of the energy states is therefore approximated by Eq.10:

\[ g(x) = \frac{1 + x}{x} \ln \frac{1 + x}{x} - \frac{1}{x} \ln \frac{1}{x} = \ln \left( (1 + x)^x \cdot x^{1-x} \right) = \ln \left( (1 + x)^x \cdot x^{-1} \right) = \ln \left( \frac{1 + x}{x} \right)^x \quad (10) \]

through the relationship presented in Eq.11:

\[ g(f_k) = \alpha + \beta \cdot e_k \quad (11) \]

As for Fermi-Dirac distribution (see below), \( f_k = n_k/b_k \) is a sub-unitary small number for low temperature. Bose-Einstein distribution is found when following limit approximations are used (Eq.12):
FERMI-DIRAC DISTRIBUTION OF ENERGY STATES

Number of possibilities to distribute 'N' particles 'A' levels with at most one particle per level (0 or 1) is equal to the number of possibilities to choose 'N' objects from a total of 'A' (the fraction \( f \) is denoted by a particle occupancy levels):

\[
w(N, A) = \frac{A!}{N!(A-N)!}; \quad f = \frac{N}{A} \rightarrow A = \frac{N}{f} \rightarrow w(N, f) = \frac{(N/f)!}{N!(N/f-N)!}
\]  

In the assumption of independence, namely a many-block of levels system ('B' blocks) that do not interact with each other levels (no particles are transferred from one group to another), the number of possibilities for the distribution of the \( N_1, ..., N_B \) particles is given in Eq.15:

\[
W(N_1, ..., N_B, f_1, ..., f_B) = \prod_{j=1}^{B} \left( \frac{f_j^{N_j}}{N_j!(f_j^{N_j} - N_j)!} \right); \quad U = \ln(W) = \sum_{j=1}^{B} \left( \ln\left( \frac{N_j}{f_j!} \right) - \ln(N_j!) - \ln(\frac{N_j}{f_j!} - N_j)! \right)
\]  

Extreme points of the function \( W \) are the same with the extreme points of the function \( U \). It will be observed that the value function \( W \) (hence the function \( U \)) that has the greatest chance of observation (relative to observables \( N_j \)) so in relation to observables \( N_j \), the functions \( W \) and \( U \) are in their extreme point. There are two conditions \( n = \Sigma N_j \) and \( E = \Sigma f_j N_j \) which makes this extreme to be conditioned. Applying the method of Lagrange multipliers, the extreme points of the function \( U \) (and the function \( W \)) subject to constraints are among the extreme points of the function \( V \) (Eq.16):
Expression of these extreme conditions is presented in Eq.17:

\[ 0 = \frac{\partial}{\partial N_i} \sum_{j=1}^{b_i} \left( \ln\left( \frac{N_{ij}}{f_{ij}} \right) - \ln\left( N_{ij} \right) - \ln\left( \frac{N_{ij} - N_{ij}}{f_{ij}} \right) \right) + \frac{\partial}{\partial N_i} \alpha(N - \sum_{j=1}^{b_i} N_{ij}) + \frac{\partial}{\partial N_i} \beta(E - \sum_{j=1}^{b_i} \varepsilon_{ij}) \]  

(17)

In order to obtain an analytic expression is necessary to employ an approximation in which the derivative factorial (Stirling's approximation) as given in Eq.18:

\[ \frac{\partial}{\partial x} \sum_{j=1}^{b_i} \ln((a \cdot x)!^j) = \frac{\partial}{\partial (a \cdot x)} \ln((a \cdot x)!^j) = \frac{\partial}{\partial (a \cdot x)} \sum_{j=1}^{b_i} \ln\left( \frac{a \cdot x}{e^x} \right) \cdot a = a \cdot \ln(a \cdot x) \]  

(18)

Substituting this approximation in all 3 expressions involving the derivative of the logarithm of a factorial, and grouping the terms that depend on \( N_i \), a reduction of the obtained relationship that is independent of the value of \( N_i \) is presented in Eq.19:

\[ \alpha + \beta \varepsilon_i = \frac{1}{f_i} \ln \frac{1}{f_i} \ln(N_i) - \ln(N_i) - (\frac{1}{f_i} - 1) \ln(\frac{1}{f_i} - 1) \cdot N_i + \alpha(0 - 1) + \beta(0 - \varepsilon_i) \]  

(19)

Eq.19 can be simplified by arranging the terms as is showed in Eq.20:

\[ \alpha + \beta \varepsilon_i = \ln \frac{1}{f_i} \ln \frac{1}{f_i} \ln(1 - f_i) \frac{1-f_i}{f_i} = \ln \frac{1}{f_i} \frac{1-f_i}{f_i} = \ln \frac{1}{f_i} \frac{1-f_i}{f_i} = \ln \frac{1}{f_i} \frac{1-f_i}{f_i} \]  

(20)

Let \( g(x) \) be the function presented in Eq.21:

\[ g(x) = \ln \frac{1}{x(1-x)^x} \]  

(21)
The solutions for energy distribution are obtained by the inverse function $g$ analytical problem and it is not solvable for the general case. A small value of $x$ has the meaning of a very low stocking levels of energy molecules ($f = N/A$), which is perfectly justified at very low temperatures. Analytical solution proposed (independently) by Fermi (Fermi, 1926) and Dirac (Dirac, 1926) is given in Eq. 22:

$$f_i = g^{-1}(\epsilon_i) \equiv \frac{1}{1 + e^{\alpha + \beta \epsilon_i}}$$  \hspace{1cm} (22)

The solution presented in Eq. 22 is an approximate supported by the relations given in Eq. 23:

$$\lim_{x \to 0^+} \left( \ln \frac{1}{x \cdot (1-x)^x} - \ln \frac{1-x}{x} - 1 \right) = 0; \lim_{x \to 0^+} \left( \ln \frac{1}{x \cdot (1-x)^x} - \ln \frac{1-x}{x} - 1 \right) \left/ x = \frac{1}{2} \to \infty \right. $$

$$\ln \frac{1}{x \cdot (1-x)^x} = 1 + \ln \frac{1-x}{x} + O(x)$$  \hspace{1cm} (23)

where $O(x)$ is the error in the approximation, linear in $x$. A closest the $x$ is to 0, the better the approximation is. Thereby approximating the function $g(x)$, the distribution in energy expression simplifies as presented in Eq. 24:

$$\ln \frac{1}{f_i (1-f_i)} \equiv 1 + \ln \frac{1-f_i}{f_i} = \alpha + \beta \epsilon_i \to \frac{1}{f_i} - 1 \equiv e^{\alpha + \beta \epsilon_i} \to f_i \equiv \frac{1}{1 + e^{\alpha + \beta \epsilon_i}}$$  \hspace{1cm} (24)

The distribution functions of the energy states according to Fermi-Dirac distribution model are presented in Fig. 1. It is noted that in the neighbourhood of 0, the proposed approximation distribution function overlaps the one from the exact model (see Fig. 1).

**Fig. 1** Fermi-Dirac distribution of energy states
The analysis of Fig. 1 led to the conclusion that the Fermi-Dirac distribution of molecules on energy has two vertical asymptotes.

**INCREASED INTEREST OF USING BOSE-EINSTEIN AND FERMI-DIRAC DISTRIBUTIONS**

Recent studies revealed an increased interest of using Bose-Einstein and Fermi-Dirac distributions. Thus, following a paper reporting successful results of Bose-Einstein model for predicting Piaget's cognitive-developmental variable, conceptualized as a quantitative construct, the maximum number of discrete pieces of information or schemes that can control or integrate in a single act, assumed to grow in an all-or-none manner as a function of age reported by Pascual-Leone (1970) other researches were more recent follows. Reimers *et al.* (2009) found evidences for weak incoherent Fröhlich condensates within individual proteins as being significant. They found also that coherent regimes or strong ones are not possible because no mechanical source of energy can produce such a condensate, and that although intense radiation could facilitate its formation, the energies required preclude their production in biological media. A sociological study was conducted on ciliate Spirostomum ambiguum's capacity to learn and store behavioural strategy advertising mating availability in (Clark, 2010).

<table>
<thead>
<tr>
<th>Tab. 1</th>
<th>Recent studies revealing an increased interest of using Bose-Einstein and Fermi-Dirac distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
<td><strong>Data</strong></td>
</tr>
<tr>
<td>Populations distribution</td>
<td>Seedling emergence of downy brome, johnsongrass, and round-leaved mallow</td>
</tr>
<tr>
<td></td>
<td>The frequency of siblings occurring in the same generation of a pedigree varies with the population size when applied to the population of Britain</td>
</tr>
<tr>
<td></td>
<td>Disease progression on commercial citrus groves</td>
</tr>
<tr>
<td></td>
<td>Intra-specific local abundances for a set North American breeding bird species</td>
</tr>
<tr>
<td>Structural biology and chemistry</td>
<td>A theoretical model for creating empirically measurable coherent states in vitro microtubules and DNA oligomers using a system of modulated tuneable laser interferometer in resonance with an applied Frohlich frequency</td>
</tr>
<tr>
<td></td>
<td>The binding of chloroform to bovine serum albumin with isothermal titration calorimeter</td>
</tr>
<tr>
<td></td>
<td>Thermodynamics, kinetics, and quantum mechanics of the primer/template duplex formation during DNA amplification by polymerase chain reaction: primer annealing process statistics</td>
</tr>
<tr>
<td></td>
<td>Atomic distributions as a function of the distance R from the molecular geometrical centre in a nonredundant set of compact globular proteins</td>
</tr>
<tr>
<td></td>
<td>The complete base sequence of HIV-1 virus and GP120 ENV gene: probability of runs of bases and No-bases</td>
</tr>
<tr>
<td></td>
<td>103 nucleotide sequences of the HIV-1 env gene, sampled from 35 countries: The expected random number of fixations per site</td>
</tr>
<tr>
<td></td>
<td>miRNA expression in the prediction of target occupancy improving the performance of two popular single miRNA target finders (rs17737058 disrupting estrogen receptor on NCOA1; miRISC protein IP independent datasets, ranging over three species, <em>D. melanogaster, C. elegans</em> and <em>H. sapiens</em>)</td>
</tr>
</tbody>
</table>
The authors of the study shown that when these ciliates switched from their first strategy choices, Bose-Einstein condensation of strategy use abruptly dissipated into a Maxwell-Boltzmann computational phase no longer dominated by a single fittest strategy. Recursive trial-and-error strategy searches annealed strategy use back into a condensed phase consistent with performance optimization. 'Social' decisions performed by ciliates showing no non-associative learning were largely governed by Fermi-Dirac statistics, resulting in degenerate distributions of strategy choices. Representative works reporting new applications of Bose-Einstein and Fermi-Dirac distributions are given in Tab.1.

REFERENCES


