Multinomial Distribution, Quantum Statistics and Einstein-Podolsky-Rosen Like Phenomena

Ratan Dasgupta · Sisir Roy

Abstract Bose-Einstein statistics may be characterized in terms of multinomial distribution. From this characterization, an information theoretic analysis is made for Einstein-Podolsky-Rosen like situation; using Shannon's measure of entropy.

Keywords Bose-Einstein statistics · Einstein-Podolsky-Rosen phenomena · Shannon's measure of entropy · Prior probability

1 Introduction

The Bose-Einstein statistics for indistinguishable micro particles may be explained and interpreted [1] within the framework of traditional probability theory. Bose-Einstein statistics may be characterized by considering it as a compound distribution of a multinomial distribution with a Dirichlet distribution. This characterization via prior probability [2, 3]; gives us a method of putting one marble at a time in cells such that the probability of resulting random arrangements of marbles in different cells follow Bose-Einstein statistics. The aim of this note is to investigate Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics using Shannon's measure of entropy and to reanalyze the Einstein-Podolsky-Rosen like situation within the information theoretic framework in the light of prior probability distribution. For reader's convenience let us briefly recapitulate the results of Bose-Einstein statistics.

R. Dasgupta (✉) · S. Roy
Indian Statistical Institute, Calcutta 700 108, India
e-mail: rdgupta@isical.ac.in

S. Roy
e-mail: sisir@isical.ac.in

Consider $\mathbf{W} = (W_1, W_2, \dots, W_k)$ to be a random vector uniformly distributed in the region:

$$\Delta = \left\{ (W_1, W_2, \dots, W_k) : W_i \geq 0,\ \sum_{i=1}^{k} W_i = 1 \right\} \tag{1}$$

The following is a Dirichlet integral

$$\int \cdots \int_{\Delta} w_1^{n_1 - 1} \cdots w_k^{n_k - 1} d\mathbf{w} = \frac{\Gamma(n_1) \cdots \Gamma(n_k)}{\Gamma(n_1 + n_2 + \cdots + n_k)} \tag{2}$$

where $n_1, \dots, n_k > 0$, $\mathbf{w} = (w_1, \dots, w_k)$; $d\mathbf{w} = dw_1 \cdots dw_k$.

For $n_1 = \cdots = n_k = 1$ the r.h.s. of (2) is $\frac{1}{(k-1)!}$, providing the volume of the region of integration Δ. Thus, the joint probability density of $\mathbf{W}$, when the density is constant, is given by

$$f(\mathbf{w}) = \begin{cases} (k-1)!, & \text{if } w_i \geq 0 \text{ and } \sum_{i=1}^{k} w_i = 1, \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

This is uniform distribution on the probability simplex: $w_1 + \cdots + w_k = 1$, where $w_i \geq 0, \forall i$.

Let $\mathbf{N} = (N_1, N_2, \dots, N_k)$ be a random vector with non-negative integer valued coordinates such that given $\mathbf{W} = \mathbf{w}$, the vector $\mathbf{N}$ has multinomial distribution with parameters $n, w_1, w_2, \dots, w_k$. That is,

$$P\left(N_1 = n_1, \dots, N_k = n_k \middle| \mathbf{W} = \mathbf{w}, \sum_{i=1}^{k} n_i = n \right) = \frac{n!}{n_1! n_2! \cdots n_k!} w_1^{n_1} w_2^{n_2} \cdots w_k^{n_k} \tag{4}$$

The r.h.s. of (4) when integrated over Δ by uniform *a priori* distribution (3), yields Bose-Einstein statistics; i.e.,

$$\begin{aligned} P\left(\mathbf{N} = \mathbf{n} \middle| \sum_{i=1}^{k} n_i = n \right) &= \int_{\Delta} P\left(\mathbf{N} = \mathbf{n} \middle| \mathbf{W} = \mathbf{w}, \sum_{i=1}^{k} n_i = n \right) f(\mathbf{w}) d\mathbf{w} \\ &= \binom{n+k-1}{k-1} \end{aligned} \tag{5}$$

vide (2), where $\mathbf{n} = (n_1, \dots, n_k)$; see Tersoff and Bayer [4] for a similar argument leading to (5). The Maxwell-Boltzmann statistics refer to distinguishable particles. Here probabilities for a marble (particle) to be assigned in any cell (state) are equal; whereas for Bose-Einstein statistics we assume *equal expected probability* with uniform *a priori* distribution (3) on the region (1). The coordinates of the vector $\mathbf{w} = (w_1, \dots, w_k)$ have an exchangeable distribution in (3). Here w_i represents the probability of a particle to be assigned in the i-th cell; $i = 1, \dots, k$. The continuous random variables w_i's are of similar stochastic magnitude and the above integral (5) represents the probability of particle-arrangement if the *cell probabilities* were allowed to be randomly distributed in an uniform manner, specified in (3). The terminology 'arbitrary weighting' used in [4] is different in concept from that of 'random uniform prior' used here. Bose-Einstein statistics of indistinguishable particles is obtained via integration of multinomial probabilities corresponding to distinguishable particles with respect to uniform prior on Δ. From Bayesian point of view, this result may be interpreted as follows: *distinguishable particles lose the property of distinguishability, after their probabilities are averaged by uniform prior on* Δ. This is similar to the situation where individual observations cannot be recovered from their average value.

It is worth mentioning [5–8] that the Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac statistics are special cases of the degenerate statistics for which the weights i.e., the number of ways in which each identifiable realization (macrostate) of the system can occur is, respectively given by the following:

$$W_{\mathrm{MB}}^{\mathrm{degen}} = n! \prod_{i=1}^{k} \frac{g_i^{n_i}}{n_i!} \tag{6}$$

$$W_{\mathrm{BE}}^{\mathrm{degen}} = \prod_{i=1}^{k} \frac{(g_i + n_i - 1)!}{(g_i - 1)! n_i!} \tag{7}$$

$$W_{\mathrm{FD}}^{\mathrm{degen}} = \prod_{i=1}^{k} \frac{g_i!}{n_i!(g_i - n_i)!} \tag{8}$$

where g_i is the degeneracy (multiplicity) of each level i.

2 Correlation and Information

Consider the uniform *a priori* distribution (3) used in the characterization (5). One can show that the correlation structure of W_1 and W_2 is given by

$$\mathrm{Corr.}(W_1, W_2) = -\frac{1}{k-1} \tag{9}$$

where k is the number of cells; see Proposition 1 in the Appendix. The random prior probabilities W_i add up to one. Increasing the value of one W_i is likely to reduce the values of the other W_j's, inducing a negative correlation structure amongst these.

This correlation structure is absent in the case of distinguishable particles following Maxwell-Boltzmann statistics. This statistical dependence might be related to the wave aspect of the particle. Wootters and Zurek [9] used Shannon entropy to give a quantitative formulation of wave-particle duality in the double-slit experiment. The general definition is the following. If a system can be in one of the several possible states in S, but if we know only the probabilities p_i of its being in each state i in S, then the amount of information about the system is

$$I = \sum_{i \in S} p_i \ln p_i \tag{10}$$

This is the negative of the Shannon entropy.

If $p_i = p$ for each micro-state i, then $I = \ln p$ as $\sum_{i \in S} p_i = 1$.

To apply Shannon's information measure, one needs to introduce the concept of canonical ensemble where the index i does not indicate the state of the particle but rather refers to configuration of the system or the microstate of the system. This is nothing but the entire set of allocations of particles to particle-states. The information $I = \sum_{i \in S} p_i \ln p_i$ then refers to the allocation of distinguishable systems to a set of distinguishable microstates. Within each such microstate, we then consider the allocation of distinguishable (MB), indistinguishable (BE) or restricted indistinguishable (FD) particles to the set of non-degenerate distinguishable (and non-interacting) particle states.

To apply "maximum entropy" principle, the weight or probability of each realization (i.e., macrostate) is considered instead of each configuration (microstate), since the microstates are generally grouped into macrostates. For example, in the case of MB statistic we consider the probability as [6, 7, 10];

$$P_{\mathrm{MB}} = \frac{n!}{k^n} \prod_{i=1}^{k} \frac{1}{n_i!}$$

the probability of each microstate being k^{-n}.

For BE and FD statistics each weight is 1. It is well known [7] that

$$W_{\mathrm{FD}}^{\mathrm{degen}} \leq \frac{W_{\mathrm{MB}}^{\mathrm{degen}}}{n!} \leq W_{\mathrm{BE}}^{\mathrm{degen}}$$

So, I_{MB} lies between I_{BE} and I_{FD}. The equality sign holds for $n = 1, k = 2$. This indicates that it is not possible to make any difference between the behaviors of 'one distinguishable' and 'one indistinguishable' particle within the present framework of information theory. It is possible only for the particles $n \geq 2$ and $k \geq 2$ states. It gives rise to an interesting possibility to reanalyze the EPR like situation for two particles.

Let us then calculate information for $k = 2, n = 2$;

$$I_{\mathrm{FD}} = 0, \quad I_{\mathrm{BE}} = -\ln 3$$

But for individual particles,

$$I_{\mathrm{FD}} = I_{\mathrm{BE}} = -\ln 2$$

Now, if the two states of the two particles be treated as independent, then total information is sum of I^{I} and I^{II} (for particles I and II, say)

$$I^{\mathrm{I}} + I^{\mathrm{II}} = -2\ln 2 \tag{11}$$

So,

$$I_{\mathrm{BE/FD}}^{\mathrm{I+II}} > (I^{\mathrm{I}} + I^{\mathrm{II}})$$

It seems that some information is lost during the preparation of the *independent* states by the process of separation of the particles. This can be explained by the fact that whenever we form a two particle state, the particles become correlated. As a result, by measuring information on any of the subsystems we may not be able to predict all the information for the remaining system unambiguously. It is also clear from the following analysis.

Consider a particle in any of the two states; probability of the first and second state are p and $(1-p)$ respectively, i.e. the probabilities are $p^x(1-p)^{1-x}$ where $x=0,1$ represents the number of particle in first state. Now, assume p to have a uniform distribution on $(0,1)$. Then, from the characterization of Bose-Einstein statistics via multinomial distribution,

$$\begin{aligned} P_{\mathrm{BE}} &= \int_0^1 p^x(1-p)^{1-x} u_1(0,1)dp, \quad x=0,1 \\ &= \frac{1}{2} \end{aligned} \tag{12}$$

For 2 particles going in any of the two states (vide (2)),

$$P_{\mathrm{BE}}^{\mathrm{I+II}} = \int_0^1 \int_0^1 \frac{2!}{1!1!} p^x(1-p)^{2-x} u_2(0,1)dp, \quad x=0,1,2$$

where u_1 is uniform *a priori* distribution of p over [0, 1] and u_2 is uniform *a priori* distribution on $\{(w_1,w_2): w_1+w_2=1\}$. Here $w_1=p$ and $w_2=1-p$.

Observe that, one can not reconstruct the joint *a priori* distribution $u_2(0,1)$ from the marginal *a priori* distribution $u_1(0,1)$ of the two particles, unless we know the dependence structure; i.e. past history of w_1 and w_2 before the particles are separated. In other words, the prior $u_2(0,1)$ of the 2 particles in their joint state contains more information than the total of two individual priors $u_1(0,1)$, considered after separation. Here, the non-separability of the wave function of the joint state is a manifestation of the peculiar quantum characteristics of the ensemble of indistinguishable particles, which is absent for that of distinguishable particles following Maxwell-Boltzmann statistics. Here no information is assumed to be instantaneously propagated from one particle to another particle, since the correlation arises due to the ensemble property of the Bose-Einstein or Fermi-Dirac particles. If we want to measure on the particle I or particle II, without disturbing the other i.e., we like to prepare the states of I and II as *independent states* then the very preparation of these states may destroy the correlation structure.

This lack of additive ness is due to non-separability of a pair of EPR correlated objects, classical or micro. Consider an unbiased coin with probability of turning head as 1/2. Let the coin be tossed and without looking at whether head or tail has turned up, suppose one sets apart the two sides by splitting it perfectly in the middle. Now, if two persons A and B take the upper and lower sides respectively and note the outcome at a very distant place from each other, then by observing the outcome of a person, other person's outcome is known. From information theoretic point of view, we may analyze the situation as follows.

Since A may observe head or tail with equal probability 1/2, ignoring the existence of the observer B, the Shannon's information for the observer A is $I_A = \frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2} = -\ln 2$.

Similarly, the marginal information of observer B is $I_B = -\ln 2$. But the total information before isolating the coin into two parts is different. Denote HT as head turning up and tail turning down. Then the outcomes are HT and TH, each with probability 1/2. The total information before segregation is seen to be greater than the sum of two *marginal information* of the two observers A and B.

$$I_{AB} = \frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2} = -\ln 2 > -2\ln 2 = I_A + I_B \tag{13}$$

Again, the conditional information of B given A is zero. Since, probability that B observes tail, given that A observes head is one, i.e., $P(B \simeq T \mid A \simeq H) = 1$ and $P(B \simeq H \mid A \simeq T) = 1$. So, $I_{B|A} = \ln 1 = 0$. Similarly, $I_{A|B} = 0$ and

$$\begin{aligned} I_{AB} &= -\ln 2 = I_A + I_{B|A} \\ &= I_B + I_{A|B} \end{aligned} \tag{14}$$

Here, to regain the total information from that of one observer (say A) of two EPR correlated observations of A and B, the past history that the observation of A *was* coupled with that of B in a particular way is essential. Such a past history reduces the variability of B's observation given that of A. This in turn increases the conditional information, leading to the recovery of total information.

3 Partial Indistinguishability and Smooth Correlation

The correlation due to Bose-Einstein and Fermi-Dirac type indistinguishability is of discrete nature depending on n and k. In general it need not be so. Note that, the multinomial distribution (4) of distinguishable particles, when integrated by a uniform *a priori* distribution over Δ gives Bose-Einstein statistics. The Bose-Einstein statistics for quantum micro particles of mass m ($\simeq 0$) may be seen in relation to a class of smoothly changing priors on Δ indexed by a parameter L depending on decreasing mass m ($\downarrow 0$) of distinguishable objects. One may then consider prior distributions which may converge to the uniform prior corresponding to Bose-Einstein statistics in a *continuous* manner.

Instead of uniform distribution over Δ which leads to Bose-Einstein statistic, consider the following *a priori* distribution for $\mathbf{W}$, where immediate two cells are highly correlated. Let,

$$f_L(\mathbf{w}) = (k-1)!\left(1 + \frac{w_1 - w_2}{Lk}\right) \tag{15}$$

on Δ, where $L = L(m) > 1$ is a large constant, thus causing a little perturbation on the uniform prior. This distribution, when compounded with the multinomial distribution (4), gives

$$P_L = P(N_1 = n_1, \ldots, N_k = n_k) = \frac{n!(k-1)!}{(n+k-1)!}\left\{1 + \frac{n_1 - n_2}{(n+k)Lk}\right\} \tag{16}$$

vide (2). Then in the limiting case, as $k \to \infty$ or, $L \to \infty$, we regain the Bose-Einstein statistics. $L \to \infty$ implies that the perturbation over uniform *a priori* distribution (3) is negligible. In (15), we considered a small perturbation over uniform prior affecting only the *first two cell probabilities. Such intermediate situations may arise when occupation of a state by a particle has an influence on the occupation of other states in a special manner; related to it is the screening type effect where a cluster of nearby cells are noticeably correlated.*

It may be mentioned that, although the elementary particles in nature are either bosons or fermions, one can always generate a special mechanism of selection such that the resultant probability distribution is of the above type. The intermediate statistics can be applied to explain composite-particle systems; e.g., the Cooper pair in the theory of superconductivity, the Fermi gas super fluid, the exciton, etc. Sometimes composite particles, composed of several fermions, may behave like bosons, obeying Bose-Einstein statistics, when they are far distant from each other. However, when they come closer, the fermions in different composite bosons start to 'feel' each other, and the statistics of the composite particles deviate from ideal Bose-Einstein statistics. Intermediate statistics may then be used as an effective tool for studying these systems; see [11], for relevant discussions and references.

In the second part of the prior (15), there is an odd function of $\mathbf{w}$ and some other odd function $g(\mathbf{w})$ of $\mathbf{w}$ may also be considered. One may interpret (15) as follows: on certain restricted sets of Δ, restriction being on first two coordinates; the particles are indistinguishable e.g., when $w_1 = w_2$ i.e., first two cells are of equal *random* probability, then $f_L(\mathbf{w}) = (k-1)!$; i.e., f is uniform on Δ and (16) becomes Bose-Einstein statistic with $n_1 = n_2$. One may interpret $L = L(m)$ ($\uparrow \infty$ for $m \downarrow 0$) of (15) as a degree of indistinguishability; since Bose-Einstein statistics of indistinguishable particles is regained when $L \to \infty$.

From (16), observe that

$$(P_L - P_{\mathrm{BE}})/P_{\mathrm{BE}} = \frac{n_1 - n_2}{(n+k)Lk} \to 0$$

as $L \to \infty$. The Shannon's information I_L of the probability distribution (16) is given in (17). It turns out to be sum of two components; the first component is Shannon's information for Bose-Einstein statistics and the second component is a remainder with diminishing effect, as $L \to \infty$. See Proposition 2 in the Appendix for a proof

$$\sum_i p_i \ln p_i = -\ln \binom{n+k-1}{k-1} + \sum_{A_n} \binom{n+k-1}{k-1}^{-1} \left\{1 + \frac{n_1 - n_2}{(n+k)Lk}\right\} \times \ln\left\{1 + \frac{n_1 - n_2}{(n+k)Lk}\right\} \tag{17}$$

where $A_n = \{(n_1, \ldots, n_k) : \sum_{i=1}^k n_i = n, n_i \geq 0\ \forall i = 1, \ldots, k\}$, $L \geq 1$. Note that (17) is a *continuously differentiable function of* L. Expanding the logarithm and using the variance covariance results of occupancy vectors of Bose-Einstein statistics [2], one obtains for large L, vide Proposition 2 in the Appendix, the following

$$I_L = \sum_i p_i \ln p_i = -\ln \binom{n+k-1}{k-1} + \frac{1}{L^2} \frac{2n(nk+2k-1)}{k^4(k+1)(n+k)^2}\{1+o(1)\} \tag{18}$$

where $o(1)$ term goes to zero, as $L \to \infty$.

From (3), (15), (16) and (18), it is interesting to observe that

$$\| f_L - f\| = O(L^{-1}) = \| P_L - P_{\text{BE}}\|, \quad \text{although} \quad |I_L - I_{\text{BE}}| = O(L^{-2})$$

The correlation function for the partial type of indistinguishability (15) is

$$I_L - I_{\text{MB}} = n \ln k - \ln \binom{n+k-1}{k-1} + \frac{1}{L^2} \frac{2n(nk+2k-1)}{k^4(k+1)(n+k)^2}\{1+o(1)\} \tag{19}$$

Thus the probability of the arrangement of particles, may *smoothly* change to Bose-Einstein type indistinguishability; e.g., when the distinguishable particles are of $m \to 0$ and the *a priori* probabilities of the particles going to different cells are random variables of similar magnitude, and e.g., when the first two cells have *equal* random probability.

4 Homogeneously Perturbed Uniform-prior and BE Statistics

In the above, we considered perturbation of uniform prior over first two cells. Next consider the following prior where *all the cell probabilities* are homogeneously perturbed over uniform-prior distribution. Let,

$$f_{\alpha, L}(\mathbf{w}) = c[1 + (w_1 w_2 \cdots w_k)^{\alpha-1}/L] \tag{20}$$

$\alpha > 0$, $L > 0$. Then from (2), $c^{-1} = [\Gamma(k)]^{-1} + \frac{(\Gamma(\alpha))^k}{\Gamma(\alpha k)L}$, so that the total probability is 1. Note that $\alpha \to 1$ and / or $L \to \infty$ imply $c \to \Gamma(k)$, then (20) provides uniform prior (3) on Δ, in the limit. The multinomial distribution (4), when integrated with prior (20) gives the following, vide (2)

$$\begin{aligned} &P(N_1 = n_1, \ldots, N_k = n_k) \\ &\quad = c\left[\frac{\Gamma(n+1)}{\Gamma(n+k)} + \frac{\Gamma(n+1)}{\Gamma(n+k\alpha)} \frac{\Gamma(n_1+\alpha)}{\Gamma(n_1+1)} \cdots \frac{\Gamma(n_k+\alpha)}{\Gamma(n_k+1)} \frac{1}{L}\right] \end{aligned} \tag{21}$$

Instead of perturbation over *all* the cell probabilities, one may consider a perturbation of uniform prior over r *consecutive cells in a circular manner*, $2 \le r \le k$. Such a choice may be relevant especially when the state space is circular or spherical. Let,

$$\begin{aligned} f(\mathbf{w}) = c\Big[1 + \frac{1}{kL}\{&(w_1 w_2 \cdots w_r)^{\alpha-1} + (w_2 w_3 \cdots w_{r+1})^{\alpha-1} + \cdots \\ &+ (w_k w_1 \cdots w_{r-1})^{\alpha-1}\}\Big] \end{aligned} \tag{22}$$

$\alpha > 0, L > 0$. From (2), one has $c^{-1} = [\Gamma(k)]^{-1} + \frac{(\Gamma(\alpha))^r}{\Gamma(k+r\alpha-r)L}$, so that the total probability is 1. As before, $c \to \Gamma(k)$, as $\alpha \to 1$ and / or $L \to \infty$. Thus (22) provides uniform prior (3) on Δ, in the limit. The multinomial distribution (4), when integrated with the prior probability distribution (22) gives the following probabilities of different arrangements over k cells

$$\begin{aligned} P(N_1 = n_1, \dots, N_k = n_k) = c\Bigg[\frac{\Gamma(n+1)}{\Gamma(n+k)} + \frac{1}{kL}\frac{\Gamma(n+1)}{\Gamma(n+r(\alpha-1)+k)} \\ \times \left\{\frac{\Gamma(n_1+\alpha)}{\Gamma(n_1+1)}\cdots\frac{\Gamma(n_r+\alpha)}{\Gamma(n_r+1)} + \cdots \right. \\ \left. + \frac{\Gamma(n_k+\alpha)}{\Gamma(n_k+1)}\cdots\frac{\Gamma(n_{r-1}+\alpha)}{\Gamma(n_{r-1}+1)}\right\}\Bigg] \end{aligned} \tag{23}$$

Observe that the first terms in the r.h.s. of (21) and (23) are of the form $c\frac{\Gamma(n+1)}{\Gamma(n+k)}$. These in the limit when $\alpha \to 1$ and / or $L \to \infty$, provide the Bose-Einstein statistic. The remainders in the r.h.s. of (21) and (23) has diminishing effect as $L \to \infty$. Thus (21) and (23), in the limit, provide the Bose-Einstein statistic. The calculations of Shannon's information and correlation function for the type of partial distinguishability (21) and (23) are similar to that of (17)–(19).

Since (17) is a continuously differentiable function of $L = L(m)$, (19) is so in L. The transition $m \to 0$, $(L(m) \to \infty)$ may be interpreted as particles are more and more indistinguishable. Unlike the correlation explained by local theories [12], (17) indicates that the l.h.s. of (19) does not have a *kink* anywhere. *Similar observations hold for the type of partial indistinguishability* (21) *and* (23). The presence of *kink* in correlation structure was observed in explaining quantum correlation via hidden variable theory as calculated by Bell. Hess and Philipp [13] discussed breakdown of Bell's theorem for local parameter spaces. The local theory was not in conformity with quantum predictions as seen from violation of Bell's inequality. The quantum correlation is a smooth function. The l.h.s. of (19) is also smooth. But the correlations calculated from local theories of hidden variable are not so.

Acknowledgements We thank the anonymous reviewers for a very detailed and constructive suggestions.

Appendix

Proposition 1 *Let* $\mathbf{W} = (W_1, W_2, \dots, W_k)$ *be a random vector with uniform distribution on the region* $\Delta = \{(W_1, W_2, \dots, W_k) : W_i \geq 0, \sum_{i=1}^k W_i = 1\}$. *Then,* $\text{Corr.}(W_1, W_2) = -\frac{1}{k-1}$.

Proof Observe that W_i's have same marginal distribution and $\sum_{i=1}^k W_i = 1$. Thus, $k\text{var}(W_1) + k(k-1)\text{cov}(W_1, W_2) = 0$. This states, $\text{cov}(W_1, W_2) = -\frac{\text{var}(W_1)}{k-1}$.

Since W_1 and W_2 has same marginal distribution, the result follows. □

Proposition 2 *The Shannon's information for the probability distribution* (16) *is*

$$\sum_i p_i \ln p_i = -\ln\binom{n+k-1}{k-1} + \frac{1}{L^2}\frac{2n(nk+2k-1)}{k^4(k+1)(n+k)^2}\{1+o(1)\}$$

Proof With $A_n = \{(n_1,\ldots,n_k) : \sum_{i=1}^k n_i = n, n_i \geq 0 \; \forall i = 1,\ldots,k\}$ note that,

$$\begin{aligned}\sum_i p_i \ln p_i &= \sum_{A_n}\binom{n+k-1}{k-1}^{-1}\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\ln\left[\binom{n+k-1}{k-1}^{-1}\right.\\ &\quad\times\left.\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\right]\\ &= \sum_{A_n}\binom{n+k-1}{k-1}^{-1}\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\left[\ln\binom{n+k-1}{k-1}^{-1}\right.\\ &\quad+\left.\ln\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\right]\\ &= -\ln\binom{n+k-1}{k-1} + \sum_{A_n}\binom{n+k-1}{k-1}^{-1}\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\\ &\quad\times\ln\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\end{aligned}$$

as $\sum_i p_i = 1$. Now for large L, use the approximation $\ln(1+x) = x(1+o(1))$, for small x, to obtain the following

$$\begin{aligned}\sum_i p_i \ln p_i &= -\ln\binom{n+k-1}{k-1} + \sum_{A_n}\binom{n+k-1}{k-1}^{-1}\left\{1+\frac{n_1-n_2}{(n+k)Lk}\right\}\\ &\quad\times\left\{\frac{n_1-n_2}{(n+k)Lk}\right\}(1+o(1))\\ &= -\ln\binom{n+k-1}{k-1} + \sum_{A_n}\binom{n+k-1}{k-1}^{-1}\\ &\quad\times\left\{\frac{n_1-n_2}{(n+k)Lk}+\frac{(n_1-n_2)^2}{(n+k)^2(Lk)^2}\right\}(1+o(1))\end{aligned}$$

Now use the results on moments of occupancy numbers of BE statistic [2], to obtain

$$\sum_{A_n}\binom{n+k-1}{k-1}^{-1} n_1 = E(N_1) = n/k$$

$$\sum_{A_n}\binom{n+k-1}{k-1}^{-1} n_1^2 = E(N_1)^2 = V(N_1) + E(N_1)^2 = \frac{n+1}{k+1}\frac{n(k-1)}{k^2} + \frac{n^2}{k^2}$$

$$\sum_{A_n} \binom{n+k-1}{k-1}^{-1} n_1 n_2 = \operatorname{cov}(N_1, N_2) + E(N_1)E(N_2) = -\frac{n+k}{k+1}\frac{n}{k^2} + \frac{n^2}{k^2}$$

Using the above, we finally obtain

$$\sum_i p_i \ln p_i = -\ln \binom{n+k-1}{k-1} + \frac{1}{L^2}\frac{2n(nk+2k-1)}{k^4(k+1)(n+k)^2}\{1+o(1)\}. \qquad \square$$

References

1. Dasgupta, R., Roy, S.: Quantum statistics, distinguishability and random trajectory. Phys. Lett. A **149**, 63–66 (1990)
2. Kunte, S.: Sankhya **39A**, 305 (1977)
3. Rao, C.R.: Linear Statistical Inferences and Its Applications, p. 336. Wiley Eastern, New Delhi (1974)
4. Tersoff, J., Bayer, D.: Phys. Rev. Lett. **50**(8), 553 (1983)
5. Niven, R.K.: Phys. Lett. A **342**, 286 (2005)
6. Tolman, R.C.: The Principles of Statistical Mechanics. Oxford Univ. Press, London (1938), Chap. X
7. Davidson, N.: Statistical Mechanics. McGraw-Hill, New York (1962), Chap. 6
8. Tanaka, T.: Methods of Statistical Physics. Cambridge University Press, Cambridge (2002), Chap. 3
9. Wootters, W.K., Zurek, W.H.: Phys. Rev. D **19**, 473 (1979)
10. Feller, W.: An Introduction to Probability Theory and Its Applications, 2nd edn. Wiley, New York
11. Shen, Y., Dai, W.-S., Xie, M.: Intermediate-statistics quantum bracket, coherent state, oscillator, and representation of angular momentum [su(2)] algebra. Phys. Rev. A **75**, 042111 (2007)
12. Bell, J.S.: Speakable and Unspeakable in Quantum Mechanics, p. 85. Cambridge University Press, Cambridge (1989)
13. Hess, K., Philipp, W.: Breakdown of Bell's theorem for certain objective local parameter spaces. Proc. Natl. Acad. Sci. **101**(7), 1799–1805 (2004)