On the deviation of statistical diffusion-exponent fluctuation from Poisson-like distribution

Yuichi Itto
Science Division, Center for General Education, Aichi Institute of Technology, Aichi 470-0392, Japan
E-mail: itto@aitech.ac.jp

Abstract. The deviation of the statistical distribution of the diffusion-exponent fluctuations from a Poisson-like distribution is discussed for virus in cytoplasm of a living cell. The deviation is shown to obey the multivariate Gaussian distribution for a class of small deviations. A brief discussion is also made about the behavior of such a deviation from a geometric viewpoint.

1. Introduction
An exotic diffusion has experimentally been observed in the infection pathway of adeno-associated viruses in living HeLa cells [1,2]. The virus fluorescently labeled with a dye molecule exhibits stochastic motions in cytoplasm of the cell in two different forms, which are the free form and the form being contained in the endosome. Analysis of each trajectory of such viruses offers the mean square displacement denoted here as

$$\overline{x^2} \sim t^\alpha,$$

(1)

where $t$ is elapsed time and $\alpha$ is termed the diffusion exponent. The diffusion property obtained from the experimental results is as follows [1]: normal diffusion with $\alpha = 1$, whereas the case with $0 < \alpha < 1$, which is referred to as anomalous diffusion [3,4]. A fact to be stressed is that $\alpha$ fluctuates between 0.5 and 0.9, which depends on localized areas of the cytoplasm. This presents an evidence for the existence of the diffusion-exponent fluctuations over the cytoplasm and, thus, makes the diffusion observed outstanding in anomalous diffusion under vital investigation in the literature. The fluctuations may reflect heterogeneous structure of the cytoplasm [1].

In a recent work [5], an attempt has been made to clarify the statistical property of the fluctuations mentioned above based on the experimental data. There, the following Poisson-like distribution of the fluctuations, which is purposely expressed here in the case of discrete values of $\alpha$, has been proposed:

$$P_{\alpha_i} \sim e^{\lambda \alpha_i} \quad (i = 1, 2, \ldots, A),$$

(2)

where $A$ is the total number of different values of the exponent, $\alpha_i$ the $i$th value of the exponent and $\lambda$ a positive constant. A basic premise for this distribution is that the time scale of variation of the fluctuations is much larger than that of stochastic motion of the virus in the localized area. That is, $\alpha$ slowly varies on a long time scale, but is assumed to be approximately constant.

If this assumption is relaxed, then the fluctuation distribution to be observed may deviate from the distribution in equation (2) only slightly. Here is a naive question: is it possible to determine the behavior of such a deviation?
In this article, we examine deviation of the fluctuation distribution from the Poisson-like distribution, which is henceforth simply referred to as deviation. In particular, we focus our attention on the behavior of small deviation. Such a deviation is found to explicitly obey the multivariate Gaussian distribution [6]. In addition, we briefly discuss the behavior of the deviation from a geometric viewpoint.

2. Entropy and its maximization

To develop our discussion, we introduce the entropy associated with the diffusion-exponent fluctuations, from which the Poisson-like distribution in equation (2) can be derived [5]. Such an entropy turns out to enable us to examine the deviation.

Suppose that the cytoplasm plays a role of a medium for stochastic motions of the virus in the two different forms, which is divided into many small virtual blocks. Each block is identified with the localized area of the cytoplasm, and the exponent in equation (1) fluctuates depending on these local blocks accordingly. The exponent is assumed to be approximately constant due to the slowness mentioned above. (This assumption will be relaxed later.)

The entropy is introduced as follows. There is no information about how the exponent locally distributes over the cytoplasm, and this is the situation we are considering here. Let us construct all of possible collections of the blocks, each of which corresponds to the medium and is distinguishable from each other with respect to the local fluctuations. Clearly, each collection is equivalent to the medium at the statistical level of the fluctuations. Denoting the total number of these collections by \( G \), the entropy is introduced as

\[
S = \ln \frac{G}{N}
\]

with \( N \) being the total number of blocks in the medium. This gives a measure of uncertainty about the fluctuations. It is necessary to evaluate \( G \), since otherwise equation (3) is formal. To do so, recall that the exponent in equation (1) is determined by average over each trajectory. It is therefore considered that the trajectory in a given localized area is included in the average, whereas the trajectories in other localized areas are not. This may imply that the local blocks are statistically independent from each other with respect to the exponent. Accordingly, for the medium with a set of exponents, \( \{\alpha_i\}_{i=1,2,...,A} \), \( G \) is given by

\[
G = N!/ \prod_{i=1}^{A} n_{\alpha_i}!
\]

where \( n_{\alpha_i} \) is the number of blocks with \( \alpha_i \) in the medium. It is clear that \( \sum_{i=1}^{A} n_{\alpha_i} = N \) is satisfied. Now, largeness of the number of blocks in the medium has been discussed in Ref. [6] by evaluating the volume of the block. The discussion given there suggests that \( n_{\alpha_i} \)'s are indeed large enough, allowing us to write \( S \) in equation (3) as the following form of the Shannon entropy:

\[
S \approx S[P] = - \sum_{i=1}^{A} P_{\alpha_i} \ln P_{\alpha_i},
\]

where \( P_{\alpha_i} = n_{\alpha_i}/N \) is the probability of finding \( \alpha_i \) in a given block of the medium. This symbol for the probability should not be mixed still at this stage with that for the distribution in equation (2).

Now, the Poisson-like distribution can be derived based on the entropy as shown in Ref. [5], where only information is supposed to be available about the statistical property of the fluctuations. In fact, with the constraints on the average of the exponent, \( \sum_{i=1}^{A} \alpha_i P_{\alpha_i} = \bar{\alpha} \), as well as the normalization condition, \( \sum_{i=1}^{A} P_{\alpha_i} = 1 \), the maximum entropy principle [7] for \( S[P] \) in equation (4) turns out to lead to the distribution in equation (2) as the stationary solution of the variational problem with respect to \( P_{\alpha_i} \), in which \( \lambda \) plays a role of a positive Lagrange multiplier for the constraint on the average.

3. Deviation of the fluctuation distribution

Up to the above stage, we have assumed that the exponent is approximately constant. In this section, we examine the deviation of the statistical fluctuation by relaxing this assumption. Our idea for it is based on Einstein’s theory of fluctuations of the thermodynamic quantities [8].
A basic observation is as follows. Let $G_{\text{max}}$ be the maximum value of $G$, for which the entropy in equation (3) takes the maximum value given by $S_{\text{max}} = (\ln G_{\text{max}})/N$. This situation is realized for the Poisson-like distribution in equation (2). Now, suppose that $G$ is decreased to a certain value given by $G' = G_{\text{max}}/e^k$, where $k$ is a positive constant. For the entropy, this corresponds to $S' = (\ln G')/N = S_{\text{max}} - k/N$. Accordingly, $S'$ is approximately equal to $S_{\text{max}}$, if $k/N$ is negligible compared to $S_{\text{max}}$, (recalling largeness of the total number of blocks, $N$), implying that the statistical fluctuation associated with $G'$ can be realized.

In what follows, we consider that the statistical fluctuation to be realized can deviate slightly from the Poisson-like fluctuation in equation (2) due to its slow variation. Then, from the above observation, we assume that such a statistical fluctuation is described by the fluctuation distribution associated with the Poisson-like fluctuation in equation (2) due to its slow variation. Then, from the above observation, the constraints not only on the normalization, $\sum_i P_{\alpha_i} = 1$, but also on the average of the exponent, $\sum_i\alpha_i P_{\alpha_i} = \bar{\alpha}$, which require the deviation, $\{\Delta P_{\alpha_i}\}_{i=1,2,\ldots,A}$, to fulfill the following two conditions

$$
\sum_{i=1}^A \Delta P_{\alpha_i} = 0, \quad \sum_{i=1}^A \alpha_i \Delta P_{\alpha_i}^{(n)} = 0,
$$

respectively. (It is noted that the present approach is applicable for $A > 2$, which seems to hold [1], otherwise the deviation identically vanishes.)

From equation (4), expanding $S[P^{(n)}]$ up to the second order of $\Delta P_{\alpha_i}$'s, $S^{(n)}$ is calculated to be

$$
S^{(n)} \simeq S_{\text{max}} - \frac{1}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2,
$$

where the conditions in equation (8) have been used. Substitution of equation (9) into equation (6) immediately reads

$$
W^{(n)} \sim W_{\text{max}} \exp \left[ -\frac{N}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2 \right],
$$

where $G_{\text{max}}$ stands for the value of $G$ for the $n$th set.

We here notice the following point. $W^{(n)}$ can be expressed in terms of the entropy for the $n$th set, $S^{(n)} = (\ln G^{(n)})/N$, as

$$
W^{(n)} \propto e^{N S^{(n)}}.
$$

This has a similarity with Einstein’s theory of fluctuations [8-10], i.e., the reversal of Boltzmann’s relation: $W \propto e^S$, where $S$ is the thermodynamic entropy and the Boltzmann constant is set equal to unity. [Note that $S$ appearing here is not the entropy in equation (3).]

Below, we evaluate the behavior of the deviation based on this similarity. Let $P_{\alpha_i}^{(n)}$ be the fluctuation distribution associated with the $n$th set, which is connected to $P_{\alpha_i}$ in equation (2):

$$
P_{\alpha_i}^{(n)} = P_{\alpha_i} + \Delta P_{\alpha_i}^{(n)} \quad (i = 1, 2, \ldots, A),
$$

where $\Delta P_{\alpha_i}^{(n)}$ denotes the deviation of the fluctuation distribution from the Poisson-like distribution and is supposed to be small. As mentioned above, the distribution in equation (7) should be examined under the constraints not only on the normalization, $\sum_{i=1}^A P_{\alpha_i}^{(n)} = 1$, but also on the average of the exponent, $\sum_{i=1}^A \alpha_i P_{\alpha_i}^{(n)} = \bar{\alpha}$, which require the deviation, $\{\Delta P_{\alpha_i}^{(n)}\}_{i=1,2,\ldots,A}$, to fulfill the following two conditions

$$
\sum_{i=1}^A \Delta P_{\alpha_i}^{(n)} = 0, \quad \sum_{i=1}^A \alpha_i \Delta P_{\alpha_i}^{(n)} = 0,
$$

where $G_{\text{max}}$ is decreased to a certain value of $S$ compared to $S_{\text{max}}$, we assume that such a statistical fluctuation is described by the fluctuation distribution associated with the Poisson-like fluctuation in equation (2) due to its slow variation. Then, from the above observation, the constraints not only on the normalization, $\sum_i P_{\alpha_i} = 1$, but also on the average of the exponent, $\sum_i\alpha_i P_{\alpha_i} = \bar{\alpha}$, which require the deviation, $\{\Delta P_{\alpha_i}\}_{i=1,2,\ldots,A}$, to fulfill the following two conditions

$$
\sum_{i=1}^A \Delta P_{\alpha_i} = 0, \quad \sum_{i=1}^A \alpha_i \Delta P_{\alpha_i} = 0,
$$

respectively. (It is noted that the present approach is applicable for $A > 2$, which seems to hold [1], otherwise the deviation identically vanishes.)

From equation (4), expanding $S[P^{(n)}]$ up to the second order of $\Delta P_{\alpha_i}$'s, $S^{(n)}$ is calculated to be

$$
S^{(n)} \simeq S_{\text{max}} - \frac{1}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2,
$$

where the conditions in equation (8) have been used. Substitution of equation (9) into equation (6) immediately reads

$$
W^{(n)} \sim W_{\text{max}} \exp \left[ -\frac{N}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2 \right],
$$

where $G_{\text{max}}$ stands for the value of $G$ for the $n$th set.

We here notice the following point. $W^{(n)}$ can be expressed in terms of the entropy for the $n$th set, $S^{(n)} = (\ln G^{(n)})/N$, as

$$
W^{(n)} \propto e^{N S^{(n)}}.
$$

This has a similarity with Einstein’s theory of fluctuations [8-10], i.e., the reversal of Boltzmann’s relation: $W \propto e^S$, where $S$ is the thermodynamic entropy and the Boltzmann constant is set equal to unity. [Note that $S$ appearing here is not the entropy in equation (3).]

Below, we evaluate the behavior of the deviation based on this similarity. Let $P_{\alpha_i}^{(n)}$ be the fluctuation distribution associated with the $n$th set, which is connected to $P_{\alpha_i}$ in equation (2):

$$
P_{\alpha_i}^{(n)} = P_{\alpha_i} + \Delta P_{\alpha_i}^{(n)} \quad (i = 1, 2, \ldots, A),
$$

where $\Delta P_{\alpha_i}^{(n)}$ denotes the deviation of the fluctuation distribution from the Poisson-like distribution and is supposed to be small. As mentioned above, the distribution in equation (7) should be examined under the constraints not only on the normalization, $\sum_{i=1}^A P_{\alpha_i}^{(n)} = 1$, but also on the average of the exponent, $\sum_{i=1}^A \alpha_i P_{\alpha_i}^{(n)} = \bar{\alpha}$, which require the deviation, $\{\Delta P_{\alpha_i}^{(n)}\}_{i=1,2,\ldots,A}$, to fulfill the following two conditions

$$
\sum_{i=1}^A \Delta P_{\alpha_i}^{(n)} = 0, \quad \sum_{i=1}^A \alpha_i \Delta P_{\alpha_i}^{(n)} = 0,
$$

respectively. (It is noted that the present approach is applicable for $A > 2$, which seems to hold [1], otherwise the deviation identically vanishes.)

From equation (4), expanding $S[P^{(n)}]$ up to the second order of $\Delta P_{\alpha_i}$'s, $S^{(n)}$ is calculated to be

$$
S^{(n)} \simeq S_{\text{max}} - \frac{1}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2,
$$

where the conditions in equation (8) have been used. Substitution of equation (9) into equation (6) immediately reads

$$
W^{(n)} \sim W_{\text{max}} \exp \left[ -\frac{N}{2} \sum_{i=1}^A \frac{1}{P_{\alpha_i}} \Delta P_{\alpha_i}^{(n)}^2 \right],
$$

where $G_{\text{max}}$ stands for the value of $G$ for the $n$th set.
where $W_{\text{max}} \equiv G_{\text{max}}/\sum_n G^{(n)}$ is the probability of finding the medium in the state with the Poisson-like fluctuation. Now, it is obvious from equation (8) that for instance, each of $\Delta P^{(n)}_{\alpha A^{-1}}$ and $\Delta P^{(n)}_{\alpha A}$ can be written in terms of the deviation, $\{\Delta P^{(n)}_{\alpha A}\}_{i=1,2,\ldots,A-2}$. Taking this into account, $W^{(n)}$ is also found to have the following form:

$$W^{(n)} \sim W_{\text{max}} \exp \left[ -\frac{N}{2} \sum_{i,j=1}^{A-2} h_{ij} \Delta P^{(n)}_{\alpha i} \Delta P^{(n)}_{\alpha j} \right],$$

(11)

where $H = (h_{ij})$ is the positive-definite symmetric matrix with the following elements:

$$h_{ij} = \frac{1}{P_{\alpha i}} \delta_{ij} + \frac{1}{P_{\alpha A^{-1}}} (\alpha_i - \alpha_A)(\alpha_j - \alpha_A) + \frac{1}{P_{\alpha A}} (\alpha_i - \alpha_{A-1})(\alpha_j - \alpha_{A-1}).$$

(12)

Thus, from equation (10) [or, equivalently (11)], we see that the deviation explicitly obeys the multivariate Gaussian distribution.

4. Deviation: A geometric viewpoint

The above results seem to give information about the scale of deviation and allow us to discuss the behavior of such a deviation from a geometric viewpoint.

To see this, we here suppose that such a scale is obtained through the extension of $W^{(n)}$ in equation (10) with respect to the deviation such as its half-width:

$$\sum_{i=1}^{A} \frac{1}{P_{\alpha i}} \Delta P^{(n)*2}_{\alpha i} = \frac{2 \ln 2}{N},$$

(13)

where $\Delta P^{(n)*}_{\alpha i}$ describes a typical scale of deviation for the $n$th set. Clearly, equation (13) can be rewritten as follows:

$$\frac{\Delta P^{(n)*2}_{\alpha 1}}{a_1^2} + \frac{\Delta P^{(n)*2}_{\alpha 2}}{a_2^2} + \cdots + \frac{\Delta P^{(n)*2}_{\alpha A}}{a_A^2} = 1$$

(14)

with $a_i = \sqrt{(2 \ln 2)P_{\alpha i}/N}$ ($i = 1, 2, \ldots, A$). Therefore, regarding $\{\Delta P^{(n)}_{\alpha i}\}_{i=1,2,\ldots,A}$ as a coordinate point in the $A$-dimensional space of the deviations, equation (14) shows that under the conditions in equation (8), the deviation, $\{\Delta P^{(n)*}_{\alpha i}\}_{i=1,2,\ldots,A}$, lies on an $A$-dimensional ellipsoid centered at the origin in the space. In other words, in this space, the conditions in equation (8) yield two different hyperplanes: one for $\Delta P^{(n)}_{\alpha 1} + \Delta P^{(n)}_{\alpha 2} + \cdots + \Delta P^{(n)}_{\alpha A} = 0$, and the other for $\alpha_1 \Delta P^{(n)}_{\alpha 1} + \alpha_2 \Delta P^{(n)}_{\alpha 2} + \cdots + \alpha_A \Delta P^{(n)}_{\alpha A} = 0$, each of which passes through the origin and intersects with each other. A set of points of the intersection then meets the above ellipsoid, on which the deviation lies. This may contribute to understanding a geometric interpretation of the behavior of the deviation.

5. Concluding remarks

We have discussed the deviation of the statistical distribution of the diffusion-exponent fluctuations for adeno-associated virus in cytoplasm of a HeLa cell. Examining the entropy associated with such fluctuations in the spirit of Einstein’s theory of fluctuations, we have shown that the small deviation from the Poisson-like fluctuation is governed by the multivariate Gaussian distribution. We have also mentioned the behavior of such a deviation in view of geometry.

It may be worth pointing out that a kinetic theory for diffusion of the virus has been presented in Ref. [5] (see Refs. [11,12] for a theoretical framework deriving it in a consistent manner) based on the Poisson-like fluctuation. (One might care about the continuum limit of entropy given there since a careful
treatment in measure theory is required for such a limit [7]. This point has been examined in a recent work [13].) In addition, it has been shown [14] that a fluctuation distribution observed in an experiment for virus capsid in a cell nucleus can be derived by the entropic approach. Thus, it is of interest to develop the present discussions for these studies.

Acknowledgement
The present article is based on the author’s talk at the Meeting on “Nonequilibrium thermodynamics and statistical physics: From rational modeling to its applications” (16–17 March, 2017, Fukuoka, Japan). He would like to thank the organizers of the meeting for providing him with opportunity to give a talk.

References
[1] Seisenberger G, Ried M U, Endreß T, Büning H, Hallek M and Bräuchle C 2001 Science 294 1929
[2] Bräuchle C, Seisenberger G, Endreß T, Ried M U, Büning H and Hallek M 2002 ChemPhysChem 3 299
[3] Bouchaud J -P and Georges A 1990 Phys. Rep. 195 127
[4] Metzler R, Jeon J -H, Cherstvy A G and Barkai E 2014 Phys. Chem. Chem. Phys. 16 24128
[5] Itto Y 2012 J. Biol. Phys. 38 673
[6] Itto Y 2016 Physica A 462 522
[7] Rosenkrantz R D (ed.) 1989 E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics (Dordrecht: Kluwer)
[8] Einstein A 1910 Ann. Phys. 33 1275
[9] Landau L D and Lifshitz E M 1980 Statistical Physics 3rd ed. (Oxford: Pergamon Press)
[10] Pais A 1982 ‘Subtle is the Lord . . . ’ (Oxford: Oxford University Press)
[11] Itto Y 2014 Phys. Lett. A 378 3037
[12] Itto Y 2017 Frontiers in Anti-Infective Drug Discovery vol 5, ed Atta-ur-Rahman and M Iqbal Choudhary (Sharjah: Bentham Science Publishers) p 3
[13] Itto Y 2018 Open Conf. Proc. J. 9 1
[14] Itto Y 2018 Phys. Lett. A 382 1238