

Article

On the Numbers of Particles in Cells in an Allocation Scheme Having an Even Number of Particles in Each Cell

Alexey Nikolaevich Chuprunov ¹ and István Fazekas ^{2,*}

¹ Faculty of Applied Mathematics, Physics and Information Technology, Chuvash State University, Universitetskaia Str. 38, 428015 Cheboksary, Russia; achuprunov@mail.ru

² Faculty of Informatics, University of Debrecen, Egyetem Square 1, 4032 Debrecen, Hungary

* Correspondence: fazekas.istvan@inf.unideb.hu

Abstract: We consider the usual random allocation model of distinguishable particles into distinct cells in the case when there are an even number of particles in each cell. For inhomogeneous allocations, we study the numbers of particles in the first K cells. We prove that, under some conditions, this K -dimensional random vector with centralised and normalised coordinates converges in distribution to the K -dimensional standard Gaussian law. We obtain both local and integral versions of this limit theorem. The above limit theorem implies a χ^2 limit theorem which leads to a χ^2 -test. The parity bit method does not detect even numbers of errors in binary files; therefore, our model can be applied to describe the distribution of errors in those files. For the homogeneous allocation model, we obtain a limit theorem when both the number of particles and the number of cells tend to infinity. In that case, we prove convergence to the finite dimensional distributions of the Brownian bridge. This result also implies a χ^2 -test. To handle the mathematical problem, we insert our model into the framework of Kolchin's generalized allocation scheme.



Citation: Chuprunov, A.N.; Fazekas, I. On the Numbers of Particles in Cells in an Allocation Scheme Having an Even Number of Particles in Each Cell. *Mathematics* **2022**, *10*, 1099. <https://doi.org/10.3390/math10071099>

Academic Editors: Leonid V. Bogachev, Dirk Zeindler, Zhonggen Su and Yuri V. Yakubovich

Received: 25 January 2022

Accepted: 21 March 2022

Published: 29 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: random allocation; generalized allocation scheme; Poisson distribution; Gaussian distribution; limit theorem; local limit theorem; Brownian bridge; χ^2 -test

MSC: 0C05; 60F05; 62G10

1. Introduction and Notation

In this paper, we study the usual random allocation model.

The random variables η_1, \dots, η_N represent a non-homogeneous allocation scheme of n -distinguishable particles into N distinct cells if their joint distribution has the form

$$\mathbb{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} = \frac{n!}{k_1!k_2! \dots k_N!} (q_1)^{k_1} (q_2)^{k_2} \dots (q_N)^{k_N}, \quad (1)$$

where k_1, k_2, \dots, k_N are non-negative integers with $k_1 + k_2 + \dots + k_N = n$, $q_1 + q_2 + \dots + q_N = 1$, $0 \leq q_i \leq 1$ for $1 \leq i \leq N$. Here, q_i is the probability that the particle is inserted into the i^{th} cell, and the random variable η_i is the number of particles in the i^{th} cell after allocating n particles into the cells. When $q_1 = q_2 = \dots = q_N = \frac{1}{N}$, then scheme (1) is called a homogeneous allocation scheme of n distinguishable particles into N distinct cells. In [1], homogeneous and non-homogeneous allocation schemes of n distinguishable particles into N distinct cells were considered.

Our goal is to study allocations with an even number of particles in each cell. Thus, let A_2 be the set of even non-negative integers, i.e., $A_2 = \{2k : k = 0, 1, 2, \dots\}$; let η_1, \dots, η_N be the allocation scheme of $2n$ distinguishable particles into N different cells; and let η'_1, \dots, η'_N be the allocation scheme of $2n$ distinguishable particles into N different cells with an even number of particles in each cell. Then, η'_1, \dots, η'_N has the distribution

$$\begin{aligned} & \mathbb{P}\{\eta'_1 = 2k_1, \dots, \eta'_N = 2k_N\} \\ &= \mathbb{P}\{\eta_1 = 2k_1, \dots, \eta_N = 2k_N \mid \eta_i \in A_2, 1 \leq i \leq N\}, \end{aligned} \tag{2}$$

where k_1, k_2, \dots, k_N are non-negative integer numbers, such that $2k_1 + 2k_2 + \dots + 2k_N = 2n$.

To describe the results of the paper, we need the following notation. \xrightarrow{d} denotes the convergence in distribution. $\gamma, \gamma_i, i \in \mathbb{N}$ are independent, identically distributed Gaussian random variables with mean 0 and variance 1.

In [2], it was proved that

$$\left(\frac{\eta_1 - nq_1}{\sqrt{nq_1}}, \dots, \frac{\eta_K - nq_K}{\sqrt{nq_K}} \right) \xrightarrow{d} (\gamma_1, \dots, \gamma_K), \tag{3}$$

if K is a fixed number and $N, n \rightarrow \infty$ such that $q_i \rightarrow 0, nq_i \rightarrow \infty$ for $1 \leq i \leq N$.

The first aim of this paper is to obtain an analogue of the above result for an allocation scheme of distinguishable particles into distinct cells having an even number of particles in each cell. We shall prove that, under some conditions,

$$\left(\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}, \frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}, \dots, \frac{\eta'_K - 2nq_K}{\sqrt{2nq_K}} \right) \xrightarrow{d} (\gamma_1, \gamma_2, \dots, \gamma_K)$$

as $N, n \rightarrow \infty$, see Theorems 2 and 3.

A well-known fact is that the polynomial distribution (1) is asymptotically normal, when N is fixed and $n \rightarrow \infty$. This result serves as a basis of the proof that the limit of the empirical process is the Brownian bridge, see [3]. In this paper, we shall study this problem for allocations having an even number of particles in each cell. Here, we introduce the following two random processes:

$$X_{2n,N}(t) = \sum_{i=1}^{[tN]} \eta'_i, \quad 0 \leq t \leq 1,$$

and

$$Y_{2n,N}(t) = \frac{1}{\sqrt{2n}} \left(\sum_{i=1}^{[tN]} \eta'_i - [tN] \frac{2n}{N} \right) = \frac{1}{\sqrt{2n}} \left(X_{2n,N}(t) - [tN] \frac{2n}{N} \right), \quad 0 \leq t \leq 1. \tag{4}$$

Observe that $X_{2n,N}(0) = 0$ and $X_{2n,N}(1) = 2n, Y_{2n,N}(0) = Y_{2n,N}(1) = 0$.

The Gaussian random process, $W_0(t), 0 \leq t \leq 1$, is called a Brownian bridge if its mean value function is 0 and its correlation function is $f(t_1, t_2) = t_1(1 - t_2), 0 \leq t_1 \leq t_2 \leq 1$.

For the homogeneous allocation scheme, we shall prove in Theorem 4 that the finite dimensional distributions of $Y_{2n,N}$ converge to the finite dimensional distributions of W_0 , if $N, n \rightarrow \infty$, such that $\frac{2n}{N} \rightarrow \infty$, see Theorem 4.

Both Theorems 3 and 4 imply χ^2 -tests.

Our mathematical approach is based on the well-known notion of the generalized allocation scheme introduced by V. F. Kolchin in [4]. Thus, we recall the definition of the generalized allocation scheme. The random variables η_1, \dots, η_N obey the generalized allocation scheme of n particles into N cells, if their joint distribution has the form

$$\mathbb{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} = \mathbb{P}\left\{ \xi_1 = k_1, \dots, \xi_N = k_N \mid \sum_{i=1}^N \xi_i = n \right\}, \tag{5}$$

for non-negative integer numbers k_1, k_2, \dots, k_N , such that $k_1 + k_2 + \dots + k_N = n$ and for some independent non-negative integer valued random variables $\xi_1, \xi_2, \dots, \xi_N$.

The simplest particular case of the generalized allocation scheme is the usual allocation of particles into cells. Thus, let $\xi_1, \xi_2, \dots, \xi_N$ be independent Poisson random variables

with parameters $\alpha q_1, \alpha q_2, \dots, \alpha q_N$ for some $\alpha > 0$ and $\sum_{i=1}^N q_i = 1$, then the generalized allocation scheme is a usual allocation scheme of n distinguishable particles into N different cells. In other words, a generalized allocation scheme defined by independent Poisson random variables ζ_1, \dots, ζ_N with parameters $\alpha_1, \alpha_2, \dots, \alpha_N$, is the usual allocation scheme of n distinguishable particles into N different cells, such that $q_i = \frac{\alpha_i}{\sum_{j=1}^N \alpha_j}$, $1 \leq i \leq N$. Thus, in a certain general sense, we can consider the value ζ_i in Equation (5) as the number of particles in the i^{th} cell.

In the original paper [4], Kolchin obtained the basic properties of the generalized allocation scheme; moreover, he proved limit theorems for the number of cells containing precisely r particles. In Equation (5), the distribution of the random variable ζ_1 can be arbitrary. Fixing its distribution in various ways, several models of discrete probability theory, such as random forests, random permutations, random allocations, and urn schemes are obtained as particular cases of the generalized allocation scheme, see [5].

In our paper, we shall not use known limit theorems for the generalized allocation scheme, we shall just use the representation (7), which is a certain consequence of the generalized allocation scheme. To this end, we shall show that when there are an even number of particles in each cell, then the usual allocation can be described by a generalized allocation scheme in the following way. Let $\text{ch}(t) = \frac{e^t + e^{-t}}{2}$, $t \in \mathbb{R}$ be the hyperbolic cosine function.

Theorem 1. *Let η_1, \dots, η_N be a generalized allocation scheme of $2n$ particles into N cells defined by Poisson independent random variables ζ_1, \dots, ζ_N with parameters β_1, \dots, β_N . Then, η'_1, \dots, η'_N defined by (2) can be represented as a generalized allocation scheme of $2n$ particles into N cells defined by the independent random variables $\zeta'_1, \dots, \zeta'_N$, with distributions*

$$\mathbb{P}\{\zeta'_i = 2k\} = \frac{\beta_i^{2k}}{(2k)! \text{ch}(\beta_i)}, \quad k = 0, 1, 2, \dots, \quad 1 \leq i \leq N.$$

That is

$$\mathbb{P}\{\eta'_1 = 2k_1, \dots, \eta'_N = 2k_N\} = \mathbb{P}\left\{\zeta'_1 = 2k_1, \dots, \zeta'_N = 2k_N \mid \sum_{i=1}^N \zeta'_i = 2n\right\}, \quad (6)$$

for non-negative integer numbers k_1, k_2, \dots, k_N , such that $k_1 + k_2 + \dots + k_N = n$.

For identically distributed random variables ζ_1, \dots, ζ_N , Theorem 1 was proved in [6]. One can prove Theorem 1 using similar elementary calculations as in the proof given in [6]. From Theorem 1 and (5), it follows that

$$\begin{aligned} \mathbb{P}\{\eta'_1 = 2k_1, \dots, \eta'_K = 2k_K\} &= \left(\prod_{i=1}^K \mathbb{P}\{\zeta'_i = 2k_i\}\right) \frac{\mathbb{P}\{\sum_{i=K+1}^N \zeta'_i = 2n - 2k\}}{\mathbb{P}\{\sum_{i=1}^N \zeta'_i = 2n\}} \\ &= \left(\prod_{i=1}^K \mathbb{P}\{\zeta'_i = 2k_i\}\right) \frac{\mathbb{P}\{\sum_{i=K+1}^N \zeta_i^* = n - k\}}{\mathbb{P}\{\sum_{i=1}^N \zeta_i^* = n\}}, \end{aligned} \quad (7)$$

where $K \leq N$, $k = k_1 + k_2 + \dots + k_K$, and the independent random variables $\zeta_1^*, \dots, \zeta_N^*$ have the distributions

$$\mathbb{P}\{\zeta_i^* = k\} = \frac{\beta_i^{2k}}{(2k)! \text{ch}(\beta_i)}, \quad k = 0, 1, 2, \dots, \quad 1 \leq i \leq N.$$

Equation (7) plays a crucial role in our paper. The proof of Theorem 2 will be based on approximations of the fractional and the multipliers in (7).

The structure of our paper is as follows: In Section 2, further notation is given and the main results are presented. Theorem 2 is the integral version of the central limit theorem for the allocation scheme, when each cell contains an even number of particles. Theorem 2

is given in terms of the generalized allocation scheme, but the underlying distribution is the Poisson distribution, so the result concerns the usual allocation scheme. However, the general setting is important because in the proof, the general framework given in Theorem 1, is used. Corollary 1 is the local version of Theorem 2. Theorem 3 is a version of Theorem 2. For practical applications, Theorem 3 is more convenient than Theorem 2. Then, we turn to the homogeneous case and present Theorem 4, which states the convergence of the finite dimensional distributions to those of the Brownian bridge. In Section 3, two χ^2 -tests are proposed. The first one tests the probabilities q_1, \dots, q_N , when the sample comes from the random allocation with an even number of particles in each cell. Then, we give a proposal to apply the χ^2 -test to check binary files with parity bits. The second χ^2 -test can be applied when we have observations only for the numbers of particles in some unions of the cells. Examples 3 and 4 offer numerical evidence for our limit theorems. In Section 4, some auxiliary results are given. In Section 5, the proofs of the main results are presented. For the proofs, we use both known approximation theorems and direct calculations.

We shall apply the following usual notation. \mathbb{R} is the set of real numbers, \mathbb{N} is the set of positive integers, \mathbb{E} stands for the expectation, and \mathbb{D}^2 denotes the variance. $o(1)$ is a quantity converging to 0. $f(x) = O(h(x))$ if $f(x)/h(x)$ is bounded as $x \rightarrow 0$.

2. Main Results

First, we study the non-homogeneous allocation scheme. Consider the scheme (6) and representation (7). Consider the generic random variable $\zeta^*(\beta)$ with parameter $\beta > 0$, having the distribution

$$\mathbb{P}\{\zeta^*(\beta) = k\} = \frac{\beta^{2k}}{(2k)! \operatorname{ch}(\beta)}, \quad k = 0, 1, 2, \dots \tag{8}$$

Let

$$\zeta_1^* = \zeta^*(\beta_1), \zeta_2^* = \zeta^*(\beta_2), \dots, \zeta_N^* = \zeta^*(\beta_N) \tag{9}$$

be independent random variables so that for any i , $\zeta_i^* = \zeta^*(\beta_i)$ has the distribution (8) with parameter $\beta = \beta_i$.

The expectation and the variance of $\zeta^*(\beta)$ (see later in Equations (21) and (25)) are

$$m^*(\beta) = \mathbb{E}(\zeta^*(\beta)) = \frac{\beta}{2} \operatorname{th}(\beta), \quad \sigma^{*2}(\beta) = \mathbb{D}^2(\zeta^*(\beta)) = \frac{\beta}{4} \left(1 + \frac{\beta}{\operatorname{ch}^2(\beta)} + \frac{e^{-\beta}}{\operatorname{ch}(\beta)} \right), \tag{10}$$

where $\operatorname{th}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is the hyperbolic tangent function. Therefore, the expectation and the variance of

$$S_N^* = \sum_{i=1}^N \zeta_i^*$$

are

$$m_N^* = \mathbb{E}S_N^* = \frac{1}{2} \sum_{i=1}^N \beta_i \operatorname{th}(\beta_i), \quad \sigma_N^{*2} = \mathbb{D}^2 S_N^* = \frac{1}{4} \sum_{i=1}^N \beta_i \left(1 + \frac{\beta_i}{\operatorname{ch}^2(\beta_i)} + \frac{e^{-\beta_i}}{\operatorname{ch}(\beta_i)} \right). \tag{11}$$

In our main theorem, we will use the following condition: for some $C > 0$,

$$\frac{|n - m_N^*|}{\sigma_N^*} < C, \quad \min_{1 \leq i \leq N} \beta_i \rightarrow \infty, \quad \text{and} \quad \frac{\max_{1 \leq i \leq N} \beta_i}{\sum_{i=1}^N \beta_i} \rightarrow 0 \tag{12}$$

as $n, N \rightarrow \infty$.

Our first main results in this paper are the following theorems:

Theorem 2. Let η'_1, \dots, η'_N be defined by (2), where η_1, \dots, η_N are defined in (5), where ξ_1, \dots, ξ_N are independent Poisson random variables with the parameters β_1, \dots, β_N . Let condition (12) be valid. Then, we have

$$\left(\frac{\eta'_1 - \beta_1}{\sqrt{\beta_1}}, \frac{\eta'_2 - \beta_2}{\sqrt{\beta_2}}, \dots, \frac{\eta'_K - \beta_K}{\sqrt{\beta_K}} \right) \xrightarrow{d} (\gamma_1, \gamma_2, \dots, \gamma_K)$$

as $N, n \rightarrow \infty$.

During the proof of Theorem 2, we shall obtain the following local limit theorem.

Corollary 1. Under the conditions of Theorem 2, if $N, n \rightarrow \infty$, then, we have

$$\mathbb{P}(\eta'_1 = 2k_1, \dots, \eta'_K = 2k_K) = \left(\prod_{i=1}^K \frac{2}{\sqrt{2\pi\beta_i}} \left(\exp\left(-\frac{(2k_i - \beta_i)^2}{2\beta_i}\right) \right) \right) (1 + o(1)) \quad (13)$$

uniformly for the values of k_i , such that $C_{1i} < \frac{2k_i - \beta_i}{\sqrt{\beta_i}} < C_{2i}$, $1 \leq i \leq K$, for any fixed numbers $C_{1i} < C_{2i}$, $1 \leq i \leq K$.

In the following theorem, q_1, q_2, \dots, q_N will denote a discrete probability distribution depending on n and N .

Theorem 3. Let η'_1, \dots, η'_N be the usual allocation scheme of $2n$ distinguishable particles into N different cells with even number of particles in each cell. Assume that the allocation probabilities are q_1, q_2, \dots, q_N which depend on n and N . Suppose that, for some $C > 0$,

$$\sqrt{n} \sum_{i=1}^N q_i e^{-4nq_i} < C, \quad n \min_{1 \leq i \leq N} q_i \rightarrow \infty \quad \text{and} \quad \max_{1 \leq i \leq N} q_i \rightarrow 0 \quad (14)$$

as $n, N \rightarrow \infty$. Then, we have

$$\left(\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}, \frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}, \dots, \frac{\eta'_K - 2nq_K}{\sqrt{2nq_K}} \right) \xrightarrow{d} (\gamma_1, \gamma_2, \dots, \gamma_K)$$

as $N, n \rightarrow \infty$.

Theorem 3 can be obtained from Theorem 2 if we use $\beta_i = 2nq_i$, $1 \leq i \leq N$.

Now, we turn to the homogeneous allocation scheme; we assume that in (1), the parameters q_i are the same. If there are an even number of particles in each cell, then this allocation is described by Equation (6), and because of homogeneity, the random variables ξ'_1, \dots, ξ'_N , are independent and identically distributed with distribution

$$\mathbb{P}\{\xi'_i = 2k\} = \frac{\beta^{2k}}{(2k)! \text{ch}(\beta)}, \quad k = 0, 1, 2, \dots, \quad 1 \leq i \leq N, \quad (15)$$

where $\beta > 0$. From (6), it follows that

$$\begin{aligned} & \mathbb{P}\{X_{2n,N}(t_1) = 2k_1, X_{2n,N}(t_2) - X_{2n,N}(t_1) = 2k_2\} \\ &= \frac{\mathbb{P}\left\{\sum_{i=1}^{\lfloor t_1 N \rfloor} \xi'_i = 2k_1\right\} \cdot \mathbb{P}\left\{\sum_{i=\lfloor t_1 N \rfloor + 1}^{\lfloor t_2 N \rfloor} \xi'_i = 2k_2\right\} \cdot \mathbb{P}\left\{\sum_{i=\lfloor t_2 N \rfloor + 1}^N \xi'_i = 2k_3\right\}}{\mathbb{P}\{\xi'_1 + \dots + \xi'_N = 2n\}}, \end{aligned} \quad (16)$$

where k_1, k_2, k_3 are non-negative integer numbers, such that $k_1 + k_2 + k_3 = n$. We shall need this formula in the proof of our Theorem 4.

Theorem 4. Let $Y_{2n,N}$ be defined in (4). Assume that in (6), the random variables ξ'_1, \dots, ξ'_N are independent and identically distributed with distribution (15). Let $N, n \rightarrow \infty$, such that $\frac{2n}{N} \rightarrow \infty$. Then, the finite dimensional distributions of $Y_{2n,N}$ converge to the finite dimensional distributions of the Brownian bridge W_0 .

The idea of the proof for the particular case of two-dimensional distributions is the following. Let $0 < t_1 < t_2 < 1$. The vector of two increments of the Brownian bridge W_0

$$(W_0(t_1) - W_0(0), W_0(t_2) - W_0(t_1)) = (W_0(t_1), W_0(t_2) - W_0(t_1))$$

has the correlation matrix

$$\Sigma = \begin{pmatrix} t_1(1 - t_1) & t_1(t_1 - t_2) \\ t_1(t_1 - t_2) & (t_2 - t_1)(1 - (t_2 - t_1)) \end{pmatrix}. \tag{17}$$

The determinant of Σ is

$$|\Sigma| = t_1(1 - t_1)(t_2 - t_1)(1 - (t_2 - t_1)) - (t_1(t_1 - t_2))^2 = t_1(t_2 - t_1)(1 - t_2)$$

and its inverse is

$$\Sigma^{-1} = \begin{pmatrix} \frac{1-(t_2-t_1)}{t_1(1-t_2)} & \frac{1}{1-t_2} \\ \frac{1}{1-t_2} & \frac{1-t_1}{(t_2-t_1)(1-t_2)} \end{pmatrix}.$$

During the proof of Theorem 4, we shall show that the distribution of the vector $(Y_{n,N}(t_1) - Y_{n,N}(0), Y_{n,N}(t_2) - Y_{n,N}(t_1))$ converges to the two-dimensional Gaussian distribution with a mean of 0 and covariance matrix Σ .

3. Applications of the Main Results for χ^2 -Tests and Numerical Examples

Using our main results, we can construct some analogues of the well-known χ^2 -test. The first one is a consequence of Theorem 3, so we assume the conditions of that theorem.

Theorem 5. Let η'_1, \dots, η'_N be an allocation scheme of $2n$ distinguishable particles into N different cells with an even number of particles in each cell. Assume that the allocation probabilities are q_1, q_2, \dots, q_N which depend on n and N . Suppose that conditions (14) are valid. Then, we have

$$\sum_{i=1}^K \frac{(\eta'_i - 2nq_i)^2}{2nq_i} \xrightarrow{d} \chi^2(K)$$

as $N, n \rightarrow \infty$, where $\chi^2(K)$ denotes the χ^2 -distribution with degree of freedom K .

The proof of Theorem 5 is a simple application of Theorem 3 and the definition of the χ^2 -distribution.

Now, we turn to an application of the above χ^2 -test for a well-known method of coding, i.e., the parity checking.

Example 1. We can apply our χ^2 -test for testing a transmission channel for messages using parity bits. The well-known parity bits are used for error detection. First, we briefly describe the usage of parity bits in the case of the so-called even parity bit. Consider a binary message containing N blocks. If a fixed block contains an odd number of bits having value 1, then we add a parity bit having value 1. If the fixed block contains an even number of bits having value 1, then we set the value of the parity bit to 0. Thus, in the final block, the number of bits having value 1 should be always even. Sometimes this method is called control sum.

After transmission of the binary message through a noisy channel, one can check the parity of each block. If the parity is odd, it shows an error. More precisely, the parity check shows an odd number of errors. However, if a block contains an even number of errors, then this check does not

show an error. We are interested in finding the error rate of a transmission channel, assuming that the parity check does not show any error.

Our statistical model is as follows: Consider a file which contains N blocks. The m^{th} block, $1 \leq m \leq N$, is a sequence $i_1 i_2 \dots i_{l_m}$, where $i_j = 1$ or $i_j = 0$, $1 \leq j \leq l_m - 1$, and $i_{l_m} = \left(\sum_{j=1}^{l_m-1} i_j\right) \pmod{2}$. i_{l_m} represents the parity bit. An error in a block is a replacement of any element i_k of the block to its opposite value, that is the true value 1 is replaced by 0, or the true value 0 is replaced by 1.

We consider the following statistical model for the errors. The file contains a binary message. It is divided into N blocks. In each block, a parity bit is used. After the transmission of the file throughout a channel, the parity check does not show any error.

To check the quality of the channel, we should obtain the original file and compare it with the transmitted one to identify the errors. We can test the hypothesis H_0 : q_i is the probability that an error occurs in the i^{th} block, where $q_i > 0$, $\sum_{i=1}^N q_i = 1$. E.g., we can test that the probability that an error happens in the i^{th} block is proportional to the length of the block by using

$$q_i = \frac{l_i}{\sum_{k=1}^N l_k}, \quad 1 \leq i \leq N.$$

The numbers of errors in the N blocks are η'_1, \dots, η'_N with the following properties:

- (1) The number of errors in the whole file is $2n$ (i.e., $\eta'_1 + \dots + \eta'_N = 2n$);
- (2) Errors can occur in the blocks independently and the probability that an error occurs in the i^{th} block is q_i ;
- (3) The parity check does not find any block with error (that is each block has an even number of errors).

Then, the numbers of errors η'_1, \dots, η'_N can be considered as the allocation of $2n$ distinguishable particles into N different cells with an even number of particles in each cell.

We calculate the statistic

$$\sum_{i=1}^K \frac{(\eta'_i - 2nq_i)^2}{2nq_i}$$

from Theorem 5, and if its value is larger than a critical value, then we reject hypothesis H_0 .

Now, we turn to an application of Theorem 4 to mathematical statistics. Our next example is similar to Example 1. Consider again a binary file containing N blocks and any block that contains a parity bit. Assume that the parity check does not show any error in the blocks. So, in any block, there can be an even number of errors. We are not able to find the number of errors in the blocks, but we can find the number of errors in m super blocks (i.e., in some unions of the original N blocks). Using the following procedure, we can test either the sizes of the super blocks or when the super block sizes are known; then, we can test if the errors are uniformly distributed among the original N blocks. In the next example, we describe the statistical procedure in a general mathematical setting.

Example 2. Consider the homogeneous allocation model. Let $2k_1, \dots, 2k_N$ be the numbers of particles in the cells after allocating $2n$ distinguishable particles into N different cells having an even number of particles in each cell. However, the numbers $2k_1, \dots, 2k_N$ are not known for us, only the numbers of particles in some neighbouring cells are known.

Let $t'_0 = 0 < t'_1 < \dots < t'_m = 1$, where each t'_j has the form k/N . So we suppose that the numbers of particles in certain sets of the cells are known, more precisely $n_j = \sum_{i=[t'_{j-1}N]+1}^{[t'_jN]} 2k_i$, $1 \leq j \leq m$, are known. Let $t_0 = 0 < t_1 < \dots < t_m = 1$ be some fixed known numbers, where again each t_j has the form of k/N . We will check the null hypothesis $H_0 : t'_i = t_i$, $1 \leq i \leq m$, against the alternative hypothesis $H_1 : t'_i \neq t_i$ for some $1 \leq i \leq m$.

To this end, we propose the following χ^2 -test. Let

$$\chi_o^2 = \sum_{j=1}^m \frac{\left(n_j - 2n \frac{[t_j N] - [t_{j-1} N]}{N}\right)^2}{2n(t_j - t_{j-1})}$$

be the test statistic.

Let $0 < \alpha < 1$. Choose the critical value χ_c^2 , such that $\mathbb{P}\{\chi^2(m - 1) < \chi_c^2\} = 1 - \alpha$, where $\chi^2(m - 1)$ is a random variable having χ^2 -distribution with a degree of freedom of $m - 1$. The hypothesis H_0 is accepted if $\chi_o^2 < \chi_c^2$, and it is rejected if $\chi_o^2 \geq \chi_c^2$. By Theorem 4, if $\frac{n}{N} \rightarrow \infty$, then the probability of the type I error converges to

$$\mathbb{P}\{\chi^2(m - 1) \geq \chi_c^2\} = \alpha.$$

Above we used, besides Theorem 4, the following known fact from the statistical theory of χ^2 -tests. If $t_0 = 0 < t_1 < \dots < t_m = 1$, then, for the increments of the Brownian bridge W_0 , the distribution of

$$\sum_{j=1}^m \frac{(W_0(t_j) - W_0(t_{j-1}))^2}{t_j - t_{j-1}}$$

is $\chi^2(m - 1)$.

Example 3. We carried out computer experiments to show numerically the results of our theorems. We simulated the allocations using random numbers. We considered a homogeneous allocation, that is, when we allocate a particle, then we choose a cell uniformly at random from the N cells. We allocated $2n = 1000$ particles into $N = 20$ cells. We repeated this experiment several times and we only saved those results when there was an even number of particles in each cell. So we saved $s = 200$ times the results of the allocations. In this way, we obtained a sample of size $s = 200$ for our $N = 20$ -dimensional random vector η'_1, \dots, η'_N . Then, we constructed histograms for the first two coordinates of the above-mentioned $N = 20$ -dimensional sample. On the left-hand side of Figure 1, the histogram of the observations of $\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}$, together with the standard normal probability density function, are shown. On the right-hand side of Figure 1, the histogram for $\frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}$ and the standard normal probability density function can be seen. The fit to the normal distribution seems to be very good. On the left-hand side of Figure 2, the joint histogram of the sample for the variables $\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}$ and $\frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}$ is given. This figure supports the joint normality of the two coordinates. Therefore, we obtained numerical evidence for Theorems 2 and 3. Finally, we performed principal component analysis for the observations of the vector

$$\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}, \dots, \frac{\eta'_N - 2nq_N}{\sqrt{2nq_N}}.$$

The first 19 principal component variances were between 1.64 and 0.53, but the last one was zero; this result supports the theory that the degree of freedom of the χ^2 -statistic in Example 2 is $m - 1$.

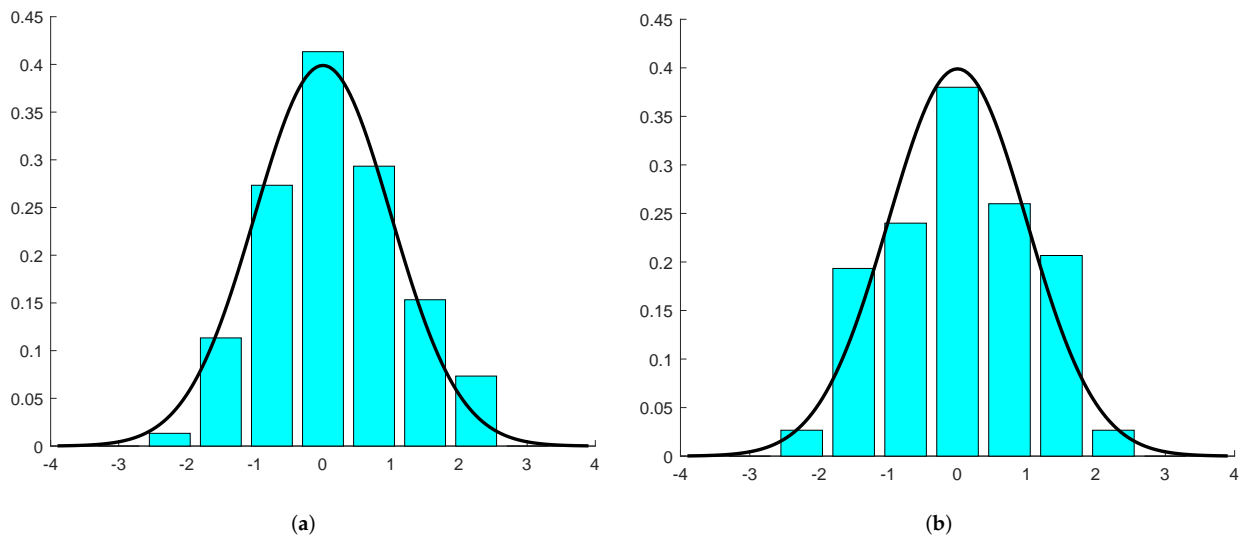


Figure 1. The histograms of the first and the second coordinates in Example 3. (a) First coordinate; (b) Second coordinate.

Example 4. We carried out the same computer experiment as in Example 3, but using other parameters. We allocated $2n = 2000$ particles into $N = 10$ cells. We saved $s = 1000$ times the results of those allocations when there was an even number of particles in each cell. In this way, we obtained a sample of size $s = 1000$ for the $N = 10$ -dimensional random vector η'_1, \dots, η'_N . On the left-hand side of Figure 3, the histogram of the observations of $\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}$, together with the standard normal probability density function, are presented. On the right-hand side of Figure 3, the histogram for $\frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}$ and the standard normal probability density function are given. The fit to the normal distribution is again, very good. On the right-hand side of Figure 2, the joint histogram of the sample for the variables $\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}$ and $\frac{\eta'_2 - 2nq_2}{\sqrt{2nq_2}}$ is given. This figure also supports the joint normality of the two coordinates. Therefore, we obtained another numerical confirmation for Theorems 2 and 3. Then, we performed principal component analysis for the observations of the vector

$$\frac{\eta'_1 - 2nq_1}{\sqrt{2nq_1}}, \dots, \frac{\eta'_N - 2nq_N}{\sqrt{2nq_N}}.$$

The first 9 principal component variances were between 1.17 and 0.87, but the last one was zero. It supports that the degree of freedom of the χ^2 -statistic in Example 2 is $m - 1$ and not m .

We mention that for relatively small values of n , e.g., for $2n = 500$ and $N = 25$, the numerical results show a pure fit to the normal distribution. It is also worth mentioning that we need a large sample size, i.e., $s > 100$, to numerically show the goodness of fit to the normal distribution

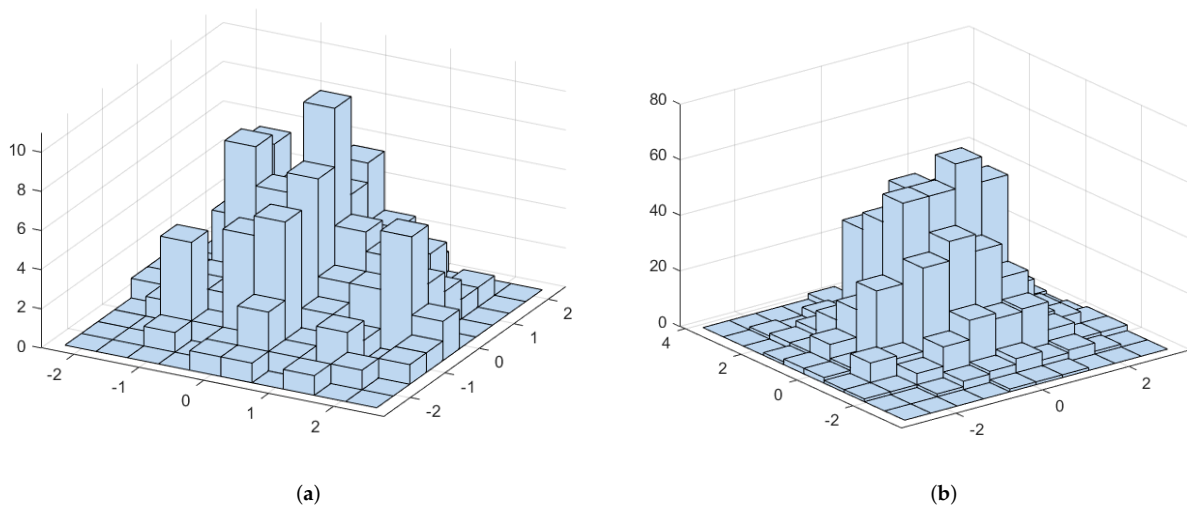


Figure 2. The joint histograms of the first two coordinates in Examples 3 and 4. (a) Histogram for Example 3; (b) Histogram for Example 4.

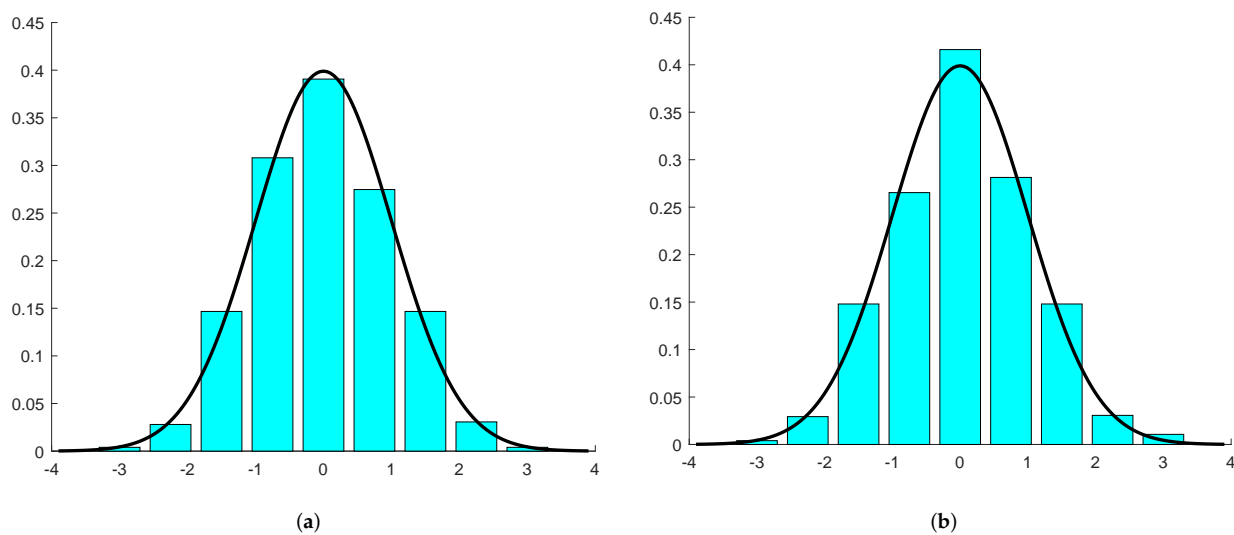


Figure 3. The histograms of the first and the second coordinates in Example 4. (a) First coordinate; (b) Second coordinate.

4. Auxiliary Results

We shall use the following notation. Let $\pi(\beta)$ denote a Poisson random variable with the parameter β , and let $\zeta'(\beta)$, where $\beta > 0$, be a random variable with the distribution

$$\mathbb{P}\{\zeta'(\beta) = 2k\} = \frac{\beta^{2k}}{(2k)!ch(\beta)}, \quad k = 0, 1, 2, \dots$$

Recall that this distribution appears in Theorem 1. We see that the distribution of $\zeta^*(\beta)$ is the same as the distribution of $\zeta'(\beta)/2$.

Lemma 1. *Let $C > 0$ be fixed. Then, we have*

$$\mathbb{P}\{\zeta'(\beta) = 2k\} = \frac{2}{\sqrt{2\pi\beta}} \left(\exp\left(-\frac{(2k - \beta)^2}{2\beta}\right) \right) (1 + o(1)) \tag{18}$$

as $\beta \rightarrow \infty$, uniformly for those values of k for which $\frac{|2k-\beta|}{\sqrt{\beta}} < C$.

Proof. We need the following approximation of the Poisson distribution by the normal density function, see p. 43 of [7]. Let τ have Poisson distribution $\mathbb{P}(\tau = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$ and let $k = \lambda + x\sqrt{\lambda}$. Then, as $\lambda \rightarrow \infty$,

$$\mathbb{P}(\tau = k) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{x^2}{2}} \left(1 + O\left(\frac{1}{\sqrt{\lambda}}\right) \right),$$

uniformly for $x \in [-c, c]$, where c is an arbitrary fixed positive number.

Using the above approximation for $\mathbb{P}\{\pi(\beta) = 2k\}$, we obtain

$$\begin{aligned} \mathbb{P}\{\zeta'(\beta) = 2k\} &= \frac{\beta^{2k}}{(2k)!e^\beta} \frac{1}{\frac{1}{2}(1 + e^{-2\beta})} = \mathbb{P}\{\pi(\beta) = 2k\} \frac{1 + o(1)}{\frac{1}{2}} \\ &= \frac{1}{\sqrt{2\pi\beta}} \left(\exp\left(-\frac{(2k - \beta)^2}{2\beta}\right) \right) (1 + o(1)) \frac{1 + o(1)}{\frac{1}{2}} \\ &= \frac{2}{\sqrt{2\pi\beta}} \left(\exp\left(-\frac{(2k - \beta)^2}{2\beta}\right) \right) (1 + o(1)) \end{aligned}$$

as $\beta \rightarrow \infty$ uniformly for k such that $\frac{|2k-\beta|}{\sqrt{\beta}} < C$. \square

Lemma 2. For the moments of $\zeta^*(\beta)$, we have

$$\begin{aligned} m^*(\beta) = \mathbb{E}\zeta^*(\beta) &= \frac{\beta}{2} \text{th}(\beta), \quad \sigma^{*2}(\beta) = \mathbb{D}^2\zeta^*(\beta) = \frac{\beta}{4}(1 + o(1)), \\ \mathbb{E}(\zeta^*(\beta) - \mathbb{E}\zeta^*(\beta))^4 &= \frac{3}{16}\beta^2(1 + o(1)) \end{aligned} \tag{19}$$

as $\beta \rightarrow \infty$.

Proof. By simple calculation, one can obtain that the characteristic function of $\zeta^*(\beta)$ is

$$\phi(t) = \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)}, \quad t \in \mathbb{R}, \tag{20}$$

where $i = \sqrt{-1}$. Using the hyperbolic sine function $\text{sh}(x) = \frac{e^x - e^{-x}}{2}$, we can obtain the derivatives of the characteristic function

$$\begin{aligned} \phi'(t) &= \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta e^{\frac{it}{2}} \frac{i}{2}, \\ \phi''(t) &= \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \left(\beta e^{\frac{it}{2}} \frac{i}{2} \right)^2 + \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta e^{\frac{it}{2}} \frac{(-1)}{4}, \\ \phi'''(t) &= \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \left(\beta e^{\frac{it}{2}} \frac{i}{2} \right)^3 + \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta^2 e^{it} \frac{(-i)}{4} + \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta^2 e^{it} \frac{(-i)}{8} + \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta e^{\frac{it}{2}} \frac{(-i)}{8} \\ &= \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \left(\beta e^{\frac{it}{2}} \frac{i}{2} \right)^3 + 3 \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta^2 e^{it} \frac{(-i)}{8} + \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta e^{\frac{it}{2}} \frac{(-i)}{8}, \end{aligned}$$

and

$$\phi^{iv}(t) = \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \left(\beta e^{\frac{it}{2}} \frac{i}{2} \right)^4 + 3 \frac{\text{sh}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta^3 e^{i\frac{3t}{2}} \frac{1}{16} + \frac{\text{ch}(\beta e^{\frac{it}{2}})}{\text{ch}(\beta)} \beta^2 e^{it} \frac{1}{16}$$

$$\begin{aligned}
 &+3\frac{\text{sh}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\left(\beta e^{\frac{i}{2}}\frac{i}{2}\right)^2\beta e^{\frac{i}{2}}\frac{(-1)}{4}+3\frac{\text{ch}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\beta^2 e^{it}\frac{1}{8}+\frac{\text{sh}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\beta e^{\frac{i}{2}}\frac{1}{16} \\
 &= \frac{\text{ch}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\left(\beta e^{\frac{i}{2}}\frac{i}{2}\right)^4+6\frac{\text{sh}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\beta^3 e^{i\frac{3t}{2}}\frac{1}{16}+7\frac{\text{ch}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\beta^2 e^{it}\frac{1}{16}+\frac{\text{sh}(\beta e^{\frac{i}{2}})}{\text{ch}(\beta)}\beta e^{\frac{i}{2}}\frac{1}{16}.
 \end{aligned}$$

Therefore, we obtain

$$m^*(\beta) = \mathbb{E}\zeta^*(\beta) = -i\phi'(0) = \frac{\beta}{2}\text{th}(\beta). \tag{21}$$

Moreover,

$$\mathbb{E}(\zeta^*(\beta))^2 = -\phi''(0) = \frac{\beta^2}{4} + \frac{\beta}{4}\text{th}(\beta), \tag{22}$$

$$\mathbb{E}(\zeta^*(\beta))^3 = i\phi'''(0) = \frac{\beta^3}{8}\text{th}(\beta) + \frac{3}{8}\beta^2 + \frac{\beta}{8}\text{th}(\beta), \tag{23}$$

and

$$\mathbb{E}(\zeta^*(\beta))^4 = \phi^{iv}(0) = \frac{1}{16}\beta^4 + \frac{6\beta^3}{16}\text{th}(\beta) + \frac{7}{16}\beta^2 + \frac{\beta}{16}\text{th}(\beta). \tag{24}$$

From Equations (21)–(24), it follows that

$$\begin{aligned}
 \sigma^{*2}(\beta) &= \mathbb{D}^2(\zeta^*(\beta)) = \mathbb{E}(\zeta^*(\beta))^2 - (\mathbb{E}\zeta^*(\beta))^2 \\
 &= \frac{\beta^2}{4} + \frac{\beta}{4}\text{th}(\beta) - \left(\frac{\beta}{2}\text{th}(\beta)\right)^2 = \frac{\beta^2}{4\text{ch}^2(\beta)} + \frac{\beta}{4}\text{th}(\beta) = \frac{\beta}{4}\left(1 + \frac{\beta}{\text{ch}^2(\beta)} - \frac{e^{-\beta}}{\text{ch}(\beta)}\right).
 \end{aligned} \tag{25}$$

It implies that $\sigma^{*2}(\beta) = \frac{\beta}{4}(1 + o(1))$. Then

$$\begin{aligned}
 &\mathbb{E}(\zeta^*(\beta) - \mathbb{E}\zeta^*(\beta))^4 \\
 &= \mathbb{E}(\zeta^*(\beta))^4 - 4(\mathbb{E}(\zeta^*(\beta))^3)\mathbb{E}\zeta^*(\beta) + 6(\mathbb{E}(\zeta^*(\beta))^2)(\mathbb{E}\zeta^*(\beta))^2 - 3(\mathbb{E}\zeta^*(\beta))^4 \\
 &= \frac{1}{16}\beta^4 + \frac{6\beta^3}{16}\text{th}(\beta) + \frac{7}{16}\beta^2 + \frac{\beta}{16}\text{th}(\beta) - 4\left(\frac{\beta^3}{8}\text{th}(\beta) + \frac{3}{8}\beta^2 + \frac{\beta}{8}\text{th}(\beta)\right)\frac{\beta}{2}\text{th}(\beta) \\
 &\quad + 6\left(\frac{\beta^2}{4} + \frac{\beta}{4}\text{th}(\beta)\right)\left(\frac{\beta}{2}\text{th}(\beta)\right)^2 - 3\left(\frac{\beta}{2}\text{th}(\beta)\right)^4 \\
 &= \frac{2\beta^4}{16}(\text{th}^2(\beta) - 1) - \frac{3\beta^4}{16}(\text{th}^4(\beta) - 1) - \frac{6\beta^3}{16}(\text{th}(\beta) - 1) + \frac{6\beta^3}{16}(\text{th}^3(\beta) - 1) \\
 &\quad + \frac{3}{16}\beta^2 - \frac{4\beta^2}{16}(\text{th}^2(\beta) - 1) + \frac{\beta}{16} + \frac{\beta}{16}(\text{th}(\beta) - 1).
 \end{aligned} \tag{26}$$

Since

$$\beta^k(\text{th}^l(\beta) - 1) \rightarrow 0 \quad \text{and} \quad \text{th}(\beta) \rightarrow 1 \quad \text{as} \quad \beta \rightarrow \infty,$$

for $k, l = 1, 2, 3, 4$, we obtain

$$\mathbb{E}(\zeta^*(\beta) - \mathbb{E}\zeta^*(\beta))^4 = \frac{3}{16}\beta^2 + \frac{1}{16}\beta + o(1) = \frac{3}{16}\beta^2(1 + o(1)).$$

Thus, (19) is proved. \square

We shall use the following general Berry–Esseen-type inequality. We should mention that in the following Lemma 3, there is no assumption on the distributions of the random variables $\zeta_i, 1 \leq i \leq N$.

Lemma 3. Let $\xi_i, 1 \leq i \leq N$, be independent random variables with variances σ_i^2 and expectations $m_i, i = 1, 2, \dots, N$. Let $S_N = \sum_{i=1}^N \xi_i$, and let $d_N^2 = \sum_{i=1}^N \sigma_i^2$ be its variance and let $\mu_N = \sum_{i=1}^N m_i$ be its expectation. Then, we have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{S_N - \mu_N}{d_N} < t \right\} - \Phi(t) \right| \leq 2c \left(\frac{\sum_{i=1}^N \mathbb{E}(\xi_i - m_i)^4}{d_N^4} \right)^{\frac{1}{2}}. \tag{27}$$

Here, Φ is the standard normal distribution function and c is the constant from the Berry–Esseen inequality.

Lemma 3 was proved in [2]. Now, we shall apply Lemma 3 to our model.

Lemma 4. Assume that for the random variables

$$\xi_1^*, \xi_2^*, \dots, \xi_N^*$$

from formula (9), condition (12) is valid. Then, for their standardized sum, we have

$$\frac{S_N^* - m_N^*}{\sigma_N^*} \xrightarrow{d} \gamma.$$

Proof. Using (19) for the right-hand side of (27), we obtain

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{S_N^* - m_N^*}{\sigma_N^*} < t \right\} - \Phi(t) \right| &\leq 2c \left(\frac{3 \left(\sum_{i=1}^N (\beta_i)^2 \right) (1 + o(1))}{\left(\sum_{i=1}^N \beta_i \right)^2} \right)^{\frac{1}{2}} \\ &\leq 2c \left(\frac{3(\max_{1 \leq i \leq N} \beta_i)(1 + o(1))}{\sum_{i=1}^N \beta_i} \right)^{\frac{1}{2}}. \end{aligned} \tag{28}$$

Now, (28) implies Lemma 4. \square

In the following lemma, we shall need the characteristic functions of the random variables in (9). Thus, let

$$\phi_j(t) = \frac{\text{ch}(\beta_j e^{i\frac{t}{2}})}{\text{ch}(\beta_j)} = \frac{e^{\beta_j e^{i\frac{t}{2}}} + e^{-\beta_j e^{i\frac{t}{2}}}}{e^{\beta_j} + e^{-\beta_j}}$$

be the characteristic function of ξ_j^* , let $\phi_j^c(t) = \phi_j(t)e^{-itm^*(\beta_j)}$ be the characteristic function of the centralized version of ξ_j^* , $1 \leq j \leq N$, and let $\phi_N(t)$ be the characteristic function of the standardized sum $\frac{S_N^* - m_N^*}{\sigma_N^*}$.

Lemma 5. Assume that for the random variables

$$\xi_1^*, \xi_2^*, \dots, \xi_N^*$$

from formula (9), condition (12) is valid. Let $C > 0$. Then, we have

$$\sigma_N^* \mathbb{P}\{S_N^* = n\} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(n - m_N^*)^2}{2\sigma_N^{*2}}\right) (1 + o(1)), \tag{29}$$

uniformly for those n , such that $\frac{|n - m_N^*|}{\sigma_N^*} < C$.

Proof. We shall need the notation

$$z = \frac{n - m_N^*}{\sigma_N^*}.$$

S_N^* is an integer valued random variable, so its distribution can be expressed by the following inverse Fourier transform

$$\mathbb{P}\{S_N^* = n\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} \phi_N^0(x) dx, \tag{30}$$

where ϕ_N^0 is the characteristic function of S_N^* . However,

$$\phi_N^0\left(\frac{t}{\sigma_N^*}\right) = \phi_N(t) e^{\frac{itm_N^*}{\sigma_N^*}} = \prod_{j=1}^N \phi_j^c\left(\frac{t}{\sigma_N^*}\right) e^{\frac{itm_N^*}{\sigma_N^*}}.$$

So, substituting $x = t/\sigma_N^*$ into the integral in (30), we obtain

$$\sigma_N^* \mathbb{P}\{S_N^* = n\} = \frac{1}{2\pi} \int_{-\pi\sigma_N^*}^{\pi\sigma_N^*} e^{-itz} \phi_N(t) dt = \frac{1}{2\pi} \int_{-\pi\sigma_N^*}^{\pi\sigma_N^*} e^{-itz} \prod_{i=1}^N \phi_i^c\left(\frac{t}{\sigma_N^*}\right) dt.$$

Let $0 < \varepsilon < 1$ and $B > 0$. Using the characteristic function of the standard normal law, we have, for any real z , that

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixz} e^{-\frac{x^2}{2}} dx.$$

So we can represent the difference of the two sides of (29) as the sum of four integrals

$$R_N = 2\pi \left(\sigma_N^* \mathbb{P}\{S_N^* = n\} - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right) = I_1 + I_2 + I_3 + I_4, \tag{31}$$

where

$$I_1 = \int_{|x| < B} e^{-ixz} \phi_N(x) dx - \int_{|x| < B} e^{-ixz} e^{-\frac{x^2}{2}} dx,$$

$$I_2 = - \int_{|x| > B} e^{-ixz} e^{-\frac{x^2}{2}} dx,$$

$$I_3 = \int_{B < |x| \leq \varepsilon\sigma_N^*} e^{-ixz} \left(\prod_{j=1}^N \phi_j^c\left(\frac{x}{\sigma_N^*}\right) \right) dx$$

and

$$I_4 = \int_{\varepsilon\sigma_N^* < |x| \leq \pi\sigma_N^*} e^{-ixz} \left(\prod_{j=1}^N \phi_j^c\left(\frac{x}{\sigma_N^*}\right) \right) dx.$$

Since

$$I_1 = \int_{|x| < B} e^{-ixz} \left(\phi_N(x) - e^{-\frac{x^2}{2}} \right) dx,$$

from Lemma 4, it follows that

$$I_1 \rightarrow 0 \tag{32}$$

for all fixed $B > 0$.

Since

$$|I_2| < \int_{|x| > B} e^{-\frac{x^2}{2}} dx,$$

we have

$$I_2 \rightarrow 0 \text{ as } B \rightarrow \infty. \tag{33}$$

Using formula (20) for the characteristic function of $\zeta^*(\beta)$, we obtain

$$\begin{aligned} |I_3| &\leq \int_{B < |x| \leq \varepsilon \sigma_N^*} \left| e^{-ixz} \left(\prod_{j=1}^N \phi_j^c \left(\frac{x}{\sigma_N^*} \right) \right) \right| dx \\ &= \int_{B < |t| \leq \varepsilon \sigma_N^*} \left(\prod_{j=1}^N e^{\beta_j \left(\cos \frac{t}{2\sigma_N^*} - 1 \right)} \right) \prod_{j=1}^N \left| \frac{1 + e^{-2\beta_j e^{i \frac{t}{2\sigma_N^*}}}}{1 + e^{-2\beta_j}} \right| dt. \end{aligned}$$

We know that

$$\cos(x) - 1 \leq -\frac{x^2}{2} + \frac{x^4}{24} \leq -\frac{11x^2}{24}, \quad \text{if } |x| \leq 1; \quad e^x - 1 \leq xe^x, \quad \text{if } x \geq 0.$$

Therefore, we obtain that

$$\prod_{j=1}^N e^{\beta_j \left(\cos \frac{t}{2\sigma_N^*} - 1 \right)} \leq e^{-(\sum_{j=1}^N \beta_j) \frac{11t^2}{24 \cdot 4\sigma_N^{*2}}} \leq e^{-\frac{11t^2}{24 \cdot 4(1+o(1))}}$$

as $B < |t| \leq \varepsilon \sigma_N^*$, where we applied that $\sigma^{*2}(\beta_j) = \frac{\beta_j}{4}(1 + o(1))$. Moreover,

$$\begin{aligned} &\prod_{j=1}^N \left| \frac{1 + e^{-2\beta_j e^{i \frac{t}{2\sigma_N^*}}}}{1 + e^{-2\beta_j}} \right| \leq \prod_{j=1}^N \frac{1 + e^{-2\beta_j \cos \left(\frac{t}{2\sigma_N^*} \right)}}{1 + e^{-2\beta_j}} \\ &= \prod_{j=1}^N \left(1 + \frac{e^{-2\beta_j \cos \left(\frac{t}{2\sigma_N^*} \right)} - e^{-2\beta_j}}{1 + e^{-2\beta_j}} \right) = \exp \left(\sum_{j=1}^N \ln \left(1 + \frac{e^{-2\beta_j \cos \left(\frac{t}{2\sigma_N^*} \right)} - e^{-2\beta_j}}{1 + e^{-2\beta_j}} \right) \right) \\ &\leq \exp \left(\sum_{j=1}^N e^{-2\beta_j} \frac{2\beta_j \left(1 - \cos \left(\frac{t}{2\sigma_N^*} \right) \right) - 1}{1 + e^{-2\beta_j}} \right) \leq \exp \left(\sum_{j=1}^N e^{-2\beta_j} \frac{2\beta_j \left(1 - \cos \left(\frac{t}{2\sigma_N^*} \right) \right)}{1 + e^{-2\beta_j}} \right) \\ &\leq \exp \left(\sum_{j=1}^N e^{-2\beta_j} \frac{2\beta_j \frac{11}{24} \left(\frac{t}{2\sigma_N^*} \right)^2}{1 + e^{-2\beta_j}} \right) \leq \exp \left(\frac{11}{12} \cdot \frac{\exp(-2 \min_{1 \leq j \leq N} \beta_j) t^2}{(1 + o(1))} \right) = e^{t^2 o(1)}, \end{aligned}$$

$B < |t| \leq \varepsilon \sigma_N^*$. Here, we used that $e^x \geq x + 1$, the shape of σ_N^* and condition (12). Therefore, we obtain

$$|I_3| \leq \int_{B < |t| \leq \varepsilon \sigma_N^*} \left(e^{-\frac{10t^2}{24}(1+o(1))} \right) dt \leq \int_{B < |t|} \left(e^{-\frac{10t^2}{24}(1+o(1))} \right) dt.$$

Consequently,

$$|I_3| \rightarrow 0 \quad \text{as } B \rightarrow \infty. \tag{34}$$

Now, we turn to I_4 . By (20), we have

$$\begin{aligned} |\phi_j(x)| &= \left| \frac{e^{\beta_j e^{i \frac{x}{2}}} + e^{-\beta_j e^{i \frac{x}{2}}}}{e^{\beta_j} + e^{-\beta_j}} \right| \\ &\leq \frac{e^{\beta_j \cos \left(\frac{x}{2} \right)} + e^{-\beta_j \cos \left(\frac{x}{2} \right)}}{e^{\beta_j} + e^{-\beta_j}} \leq \frac{e^{\beta_j \cos(\varepsilon)} + e^{-\beta_j \cos(\varepsilon)}}{e^{\beta_j} + e^{-\beta_j}} = \frac{\text{ch}(\beta_j \cos(\varepsilon))}{\text{ch}(\beta_j)} \end{aligned}$$

for $\varepsilon < |x| \leq \pi$. Thus,

$$|I_4| \leq 2\pi\sigma_N^* \prod_{j=1}^N \left(\frac{\text{ch}(\beta_j \cos(\varepsilon))}{\text{ch}(\beta_j)} \right) \leq 2\pi\sigma_N^* \prod_{j=1}^N \left(e^{-(1-\cos(\varepsilon))\beta_j} \frac{1 + e^{-2\beta_j \cos(\varepsilon)}}{1 + e^{-2\beta_j}} \right).$$

Since, by (12), $\min_{1 \leq j \leq N} \beta_j \rightarrow \infty$; therefore, we have

$$\frac{1 + e^{-2\beta_j \cos(\varepsilon)}}{1 + e^{-2\beta_j}} \rightarrow 1$$

as $N \rightarrow \infty$, uniformly for $1 \leq i \leq N$. Consequently, there exists $N_0 \in \mathbb{N}$, such that

$$\frac{1 + e^{-2\beta_j \cos(\varepsilon)}}{1 + e^{-2\beta_j}} < e^{\frac{1}{2}(1-\cos(\varepsilon))\beta_j},$$

$1 \leq j \leq N, N > N_0$. Therefore,

$$|I_4| \leq 2\pi\sigma_N^* \prod_{j=1}^N e^{-\frac{1}{2}(1-\cos(\varepsilon))\beta_j} \leq 2\pi\sigma_N^* e^{-\frac{1}{2}(1-\cos(\varepsilon))\sigma_N^{*2}(1+o(1))}$$

for $N > N_0$. Therefore, we obtain that

$$I_4 \rightarrow 0. \tag{35}$$

Finally, using formulae (32), (33), (34), and (35) to approximate the left-hand side of (31), we obtain (29). \square

5. Proofs of the Main Theorems

Proof of Theorem 2. During the proof, we represent $\eta'_1, \eta'_2, \dots, \eta'_N$ in the form of (7). First, we prove a local version of our limit theorem. To this end, we study the case when the standardized random variables are inside some bounded intervals. Therefore, we need the following notation. Let $C_{1i} < C_{2i}, 1 \leq i \leq K$. Let

$$k = \sum_{i=1}^K k_i, \quad C^* = \max\{|C_{ij}| : 1 \leq j \leq K, i = 1, 2, \dots\}.$$

Let

$$m_K^* = \sum_{i=K+1}^N m^*(\beta_i) \quad \text{and} \quad \sigma_K^{*2} = \sum_{i=K+1}^N \sigma^{*2}(\beta_i)$$

be the expectation and the variance of $\sum_{i=K+1}^N \zeta_i^*$. By Lemma 1, we have

$$\prod_{i=1}^K \mathbb{P}(\zeta'_i = 2k_i) = \prod_{i=1}^K \frac{2}{\sqrt{2\pi\beta_i}} \left(\exp\left(-\frac{(2k_i - \beta_i)^2}{2\beta_i}\right) \right) (1 + o(1)), \tag{36}$$

uniformly for $k_i, 1 \leq i \leq K$, such that

$$C_{1i} < \frac{2k_i - \beta_i}{\sqrt{\beta_i}} < C_{2i}, \quad 1 \leq i \leq K. \tag{37}$$

Since

$$\begin{aligned} \sigma_N^{*2} &\geq \sigma_K^{*2} = \sum_{i=K+1}^N \frac{\beta_i}{4} \left(1 + \frac{\beta_i}{\text{ch}^2(\beta_i)} - \frac{e^{-\beta_i}}{\text{ch}\beta_i} \right) \\ &= \frac{1 + o(1)}{4} \sum_{i=K+1}^N \beta_i = \frac{1 + o(1)}{4} \left(\sum_{i=1}^N \beta_i \right) \left(1 - \frac{\sum_{i=1}^K \beta_i}{\sum_{i=1}^N \beta_i} \right) \end{aligned}$$

$$\geq \frac{1 + o(1)}{4} \left(\sum_{i=1}^N \beta_i \right) \left(1 - \frac{K \max_{1 \leq i \leq K} \beta_i}{\sum_{i=1}^N \beta_i} \right) = (1 + o(1)) \sigma_N^{*2},$$

Therefore, we have

$$\sigma_K^{*2} = (1 + o(1)) \sigma_N^{*2} \quad \text{and} \quad \frac{\sigma_K^*}{\sigma_N^*} = o(1).$$

Let $k_i, 1 \leq i \leq K$, be such that $C_{1i} < \frac{2k_i - \beta_i}{\sqrt{\beta_i}} < C_{2i}$ for $1 \leq i \leq K$. Using (12) and the above calculation, we have

$$\frac{|n - k - m_K^*|}{\sigma_K^*} \leq \frac{|n - m_N^*|}{\sigma_K^*} + \sum_{i=1}^K \frac{|k_i - m^*(\beta_i)|}{\sigma_K^*} < C + KC^* + o(1). \tag{38}$$

Therefore, by Lemma 5, we obtain

$$\begin{aligned} \frac{\mathbb{P}\left\{\sum_{i=K+1}^N \zeta_i^* = n - k\right\}}{\mathbb{P}\left\{\sum_{i=1}^N \zeta_i^* = n\right\}} &= \frac{\frac{1}{\sqrt{2\pi\sigma_K^*}} \exp\left(-\frac{(n-k-m_K^*)^2}{2\sigma_K^{*2}}\right) (1 + o(1))}{\frac{1}{\sqrt{2\pi\sigma_N^*}} \exp\left(-\frac{(n-m_N^*)^2}{2\sigma_N^{*2}}\right) (1 + o(1))} \\ &= \exp\left(\frac{(n - m_N^*)^2}{2\sigma_N^{*2}} - \frac{(n - k - m_K^*)^2}{2\sigma_N^{*2}}\right) (1 + o(1)) \\ &= \exp\left(\frac{(k - (m_N^* - m_K^*)) (n - m_N^* + n - k - m_K^*)}{\sigma_N^* 2\sigma_N^*}\right) (1 + o(1)). \end{aligned}$$

Using (38), (37), and assumption (12), we obtain

$$\begin{aligned} \left| \frac{(n - m_N^* + n - k - m_K^*) (k - (m_N^* - m_K^*))}{\sigma_N^*} \right| &\leq (2C + KC^* + o(1)) \sum_{i=1}^K \frac{\sigma^*(\beta_i)}{\sigma_N^*} \frac{|k_i - m^*(\beta_i)|}{\sigma^*(\beta_i)} \\ &\leq (2C + KC^* + o(1))(C^* + o(1)) \sum_{i=1}^K \frac{\sigma^*(\beta_i)}{\sigma_N^*} = o(1). \end{aligned}$$

Using the above calculations, we have

$$\frac{\mathbb{P}\left\{\sum_{i=K+1}^N \zeta_i^* = n - k\right\}}{\mathbb{P}\left\{\sum_{i=1}^N \zeta_i^* = n\right\}} = 1 + o(1). \tag{39}$$

Now, using (36) and (39) in formula (7), we obtain

$$\mathbb{P}(\eta'_1 = 2k_1, \dots, \eta'_K = 2k_K) = \left(\prod_{i=1}^K \frac{2}{\sqrt{2\pi\beta_i}} \left(\exp\left(-\frac{(2k_i - \beta_i)^2}{2\beta_i}\right) \right) \right) (1 + o(1)) \tag{40}$$

uniformly for k_i , such that $C_{1i} < \frac{2k_i - \beta_i}{\sqrt{\beta_i}} < C_{2i}, 1 \leq i \leq K$. Thus, we obtained Corollary 1.

Now, we can apply the well-known method of obtaining the integral version of de Moivre–Laplace theorem from its local version. Thus, using the notation $t_{k_i} = \frac{2k_i - \beta_i}{\sqrt{\beta_i}}, \Delta t_{k_i} = \frac{2}{\sqrt{\beta_i}}$ for $1 \leq i \leq K$, we obtain

$$\mathbb{P}\left(\frac{\eta'_1 - \beta_1}{\sqrt{\beta_1}} = t_{k_1}, \dots, \frac{\eta'_K - \beta_K}{\sqrt{\beta_K}} = t_{k_K}\right) = \left(\prod_{i=1}^K \frac{\Delta t_{k_i}}{\sqrt{2\pi}} \left(\exp\left(-\frac{t_{k_i}^2}{2}\right) \right) \right) (1 + o(1)). \tag{41}$$

Here, on the right-hand side, there is a member of the approximating sum of the integral of the K dimensional standard normal probability density function; so, we obtain

$$\mathbb{P}\{C_{11} < \eta'_1 < C_{12}, \dots, C_{K1} < \eta'_K < C_{K2}\} \rightarrow \prod_{i=1}^K \mathbb{P}\{C_{1i} < \gamma_i < C_{2i}\}.$$

This implies Theorem 2. \square

Now, we turn to the proof of Theorem 4. Thus, we consider the homogeneous allocation scheme, and we assume that there are even numbers of particles in each cell. That is why we consider Equation (6) with independent and identically distributed random variables $\zeta'_1, \dots, \zeta'_N$, with distribution

$$\mathbb{P}\{\zeta'_i = 2k\} = \frac{\beta^{2k}}{(2k)! \text{ch}(\beta)}, \quad k = 0, 1, 2, \dots, \quad 1 \leq i \leq N.$$

As

$$\zeta'_i(\beta) = 2\tilde{\zeta}_i^*(\beta),$$

so

$$m(\beta) = \mathbb{E}\zeta'_i(\beta) = 2\mathbb{E}\tilde{\zeta}_i^*(\beta), \quad \sigma^2(\beta) = \mathbb{E}(\zeta'_i(\beta) - m(\beta))^2 = 4\sigma^{*2}(\beta) \tag{42}$$

are the expectation and the variance of $\zeta'_i(\beta)$. Let

$$S_N = \sum_{i=1}^N \zeta'_i$$

be the sum of our random variables. We need the following corollary of Lemma 5.

Corollary 2. Consider the homogeneous allocation scheme. Let $\beta \rightarrow \infty$. Then, we have

$$\mathbb{P}\{S_N = 2k\} = \frac{2}{\sqrt{2\pi N}\sigma(\beta)} \left(\exp\left(-\frac{(2k - Nm(\beta))^2}{2N\sigma^2(\beta)}\right) (1 + o(1)) \right) \tag{43}$$

as $N \rightarrow \infty$ uniformly for $\frac{|2k - Nm(\beta)|}{\sqrt{N}\sigma(\beta)} < C$ for any $C > 0$.

Proof of Theorem 4. First, we give a detailed proof for the two-dimensional distributions; then, we sketch the proof for the arbitrary finite dimensional distributions.

Let $-\infty < b_{i1} < b_{i2} < \infty, i \in \{1, 2\}$. Choose $C > 0$, such that $-C < b_{i1} < b_{i2} < C, i \in \{1, 2\}$. Let β be such that $\frac{2n}{N} = m(\beta)$. From (42) and Lemma 2, it follows that

$$m(\beta) = \beta \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} = \beta \text{th}(\beta) \tag{44}$$

and

$$\sigma^2(\beta) = \beta \left(1 + \frac{2e^{-\beta}}{e^\beta + e^{-\beta}} + \beta \frac{4}{(e^\beta + e^{-\beta})^2} \right).$$

Since $f(\beta) = \text{th}(\beta) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}, \beta \geq 0$ is a bounded function, from (44) and from condition $\frac{2n}{N} \rightarrow \infty$, we obtain that $\beta \rightarrow \infty$. Therefore,

$$\sigma^2(\beta) = \frac{2n}{N} (1 + o(1))$$

as $\frac{2n}{N} \rightarrow \infty$. Condition $\frac{2n}{N} \rightarrow \infty$ implies that

$$\frac{n}{[t_1 N]} \rightarrow \infty, \quad \frac{n}{[t_2 N] - [t_1 N]} \rightarrow \infty, \quad \frac{n}{N - [t_2 N]} \rightarrow \infty.$$

Consequently, from Corollary 2, it follows that

$$\begin{aligned} \mathbb{P}\{S_N = 2n\} &= \frac{2}{\sqrt{2\pi N}\sigma(\beta)} \exp\left(-\frac{(2n - Nm(\beta))^2}{2N\sigma^2(\beta)}\right) (1 + o(1)) \\ &= \frac{2}{\sqrt{4\pi n}} (1 + o(1)), \end{aligned} \tag{45}$$

uniformly for $\frac{|2n - Nm(\beta)|}{2\sqrt{N}\sigma(\beta)} < C$. Similarly,

$$\begin{aligned} \mathbb{P}\{S_{[Nt_1]} = 2k_1\} &= \frac{2 \exp\left(-\frac{(2k_1 - [Nt_1]m(\beta))^2}{2[Nt_1]\sigma^2(\beta)}\right)}{\sqrt{2\pi[Nt_1]}\sigma(\beta)} (1 + o(1)) \\ &= \frac{2 \exp\left(-\frac{(2k_1 - 2n\frac{[t_1N]}{N})^2}{4nt_1}\right)}{\sqrt{4\pi nt_1}} (1 + o(1)) \end{aligned} \tag{46}$$

uniformly for $\frac{|2k_1 - 2n\frac{[t_1N]}{N}|}{\sqrt{2nt_1}} < C$, and

$$\begin{aligned} \mathbb{P}\{S_{[Nt_2] - [Nt_1]} = 2k_2\} &= \frac{2 \exp\left(-\frac{(2k_2 - ([Nt_2] - [Nt_1])m(\beta))^2}{2([Nt_2] - [Nt_1])\sigma^2(\beta)}\right)}{\sqrt{2\pi([Nt_2] - [Nt_1])}\sigma(\beta)} (1 + o(1)) \\ &= \frac{2 \exp\left(-\frac{(2k_2 - 2n\frac{[t_2N] - [t_1N]}{N})^2}{4n(t_2 - t_1)}\right)}{\sqrt{4\pi n(t_2 - t_1)}} (1 + o(1)) \end{aligned} \tag{47}$$

uniformly for $\frac{|2k_2 - 2n\frac{[t_2N] - [t_1N]}{N}|}{\sqrt{2n(t_2 - t_1)}} < C$. Since

$$\frac{|2k_1 + 2k_2 - 2n\frac{[t_2N]}{N}|}{\sqrt{4n(1 - t_2)}} \leq \sqrt{\frac{t_1}{1 - t_2}} \frac{|2k_1 - 2n\frac{[t_1N]}{N}|}{\sqrt{4nt_1}} + \sqrt{\frac{t_2 - t_1}{1 - t_2}} \frac{|2k_2 - 2n\frac{[t_2N] - [t_1N]}{N}|}{\sqrt{4n(t_2 - t_1)}},$$

so we have

$$\begin{aligned} \mathbb{P}\{S_{N - [Nt_2]} = 2n - 2k_1 - 2k_2\} &= \\ &= \frac{2 \exp\left(-\frac{(2n - (2k_1 + 2k_2) - (N - [Nt_2])m(\beta))^2}{2(N - [Nt_2])\sigma^2(\beta)}\right)}{\sqrt{2\pi(N - [Nt_2])}\sigma(\beta)} (1 + o(1)) = \frac{2 \exp\left(-\frac{(2k_1 + 2k_2 - 2n\frac{[t_2N]}{N})^2}{4n(1 - t_2)}\right)}{\sqrt{4\pi n(1 - t_2)}} (1 + o(1)) \end{aligned} \tag{48}$$

uniformly for $\frac{|2k_1 - 2n\frac{[t_1N]}{N}|}{\sqrt{2nt_1}} < C$, $\frac{|2k_2 - 2n\frac{[t_2N] - [t_1N]}{N}|}{\sqrt{2n(t_2 - t_1)}} < C$.

For short, let $A = \frac{[t_1N]}{N}$, $B = \frac{[t_2N] - [t_1N]}{N}$. Using Equations (45)–(48) to approximate the probabilities in (16), and applying the definition of Σ from (17), we obtain

$$\begin{aligned} &\mathbb{P}\{X_{2n,N}(t_1) = 2k_1, X_{2n,N}(t_2) - X_{2n,N}(t_1) = 2k_2\} \\ &= \frac{\mathbb{P}\{S_{[Nt_1]} = 2k_1\} \mathbb{P}\{S_{[Nt_2] - [Nt_1]} = 2k_2\} \mathbb{P}\{S_{N - [Nt_2]} = 2n - 2k_1 - 2k_2\}}{\mathbb{P}\{S_N = 2n\}} \end{aligned}$$

$$\begin{aligned}
 & \frac{2 \exp\left(-\frac{(2k_1-2n\frac{[t_1N]}{N})^2}{4nt_1}\right)}{\sqrt{4\pi nt_1}} \frac{2 \exp\left(-\frac{(2k_2-2n\frac{[t_2N]-[t_1N]}{N})^2}{4n(t_2-t_1)}\right)}{\sqrt{4\pi n(t_2-t_1)}} \frac{2 \exp\left(-\frac{(2k_1+2k_2-2n\frac{[t_2N]}{N})^2}{4n(1-t_2)}\right)}{\sqrt{4\pi n(1-t_2)}} \\
 = & \frac{\frac{2}{\sqrt{4\pi n}}(1+o(1))}{\frac{2}{\sqrt{4\pi n}}(1+o(1))} \\
 = & \frac{4 \exp\left(-\frac{1}{2}\left(\frac{(2k_1-2nA)^2(1-(t_2-t_1))}{t_1(1-t_2)2n} + \frac{(2k_2-2nB)^2(1-t_1)}{(t_2-t_1)(1-t_2)2n} + 2\frac{(2k_1-2nA)(2k_2-2nB)}{(1-t_2)2n}\right)\right)}{2\pi 2n\sqrt{|\Sigma|}(1+o(1))} \\
 = & \frac{\frac{2}{n} \cdot \exp\left(-\frac{1}{2}\left(\frac{2k_1-2nA}{\sqrt{2n}}, \frac{2k_2-2nB}{\sqrt{2n}}\right)\Sigma^{-1}\left(\frac{2k_1-2nA}{\sqrt{2n}}, \frac{2k_2-2nB}{\sqrt{2n}}\right)\right)}{2\pi\sqrt{|\Sigma|}}(1+o(1)) \tag{49}
 \end{aligned}$$

uniformly for $\frac{|2k_1-2n\frac{[t_1N]}{N}|}{\sqrt{2nt_1}} < C, \frac{|2k_2-2n\frac{[t_2N]-[t_1N]}{N}|}{\sqrt{2n(t_2-t_1)}} < C.$

From (49) and (4), using the same argument as in the proof of the de Moivre–Laplace theorem, we obtain

$$\begin{aligned}
 & \mathbb{P}\{b_{11} < Y_{2n,N}(t_1) < b_{12}, b_{21} < Y_{2n,N}(t_2) - Y_{2n,N}(t_1) < b_{22}\} \\
 = & \sum_{\substack{b_{11} < \frac{2k_1-2nA}{\sqrt{2n}} < b_{12}, \\ b_{21} < \frac{2k_2-2nB}{\sqrt{2n}} < b_{22}}} \mathbb{P}\{X_{2n,N}(t_1) = 2k_1, X_{2n,N}(t_2) - X_{2n,N}(t_1) = 2k_2\} \\
 = & \sum_{\substack{b_{11} < \frac{2k_1-2nA}{\sqrt{2n}} < b_{12}, \\ b_{21} < \frac{2k_2-2nB}{\sqrt{2n}} < b_{22}}} \frac{\frac{2}{n} \cdot \exp\left(-\frac{1}{2}\left(\frac{2k_1-2nA}{\sqrt{2n}}, \frac{2k_2-2nB}{\sqrt{2n}}\right)\Sigma^{-1}\left(\frac{2k_1-2nA}{\sqrt{2n}}, \frac{2k_2-2nB}{\sqrt{2n}}\right)\right)}{2\pi\sqrt{|\Sigma|}(1+o(1))} \\
 = & \int_{b_{11}}^{b_{12}} \int_{b_{21}}^{b_{22}} \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x,y)\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right) dx dy (1+o(1)) \\
 = & \mathbb{P}\{b_{11} < W_0(t_1) < b_{12}, b_{11} < W_0(t_2) - W_0(t_1) < b_{12}\}(1+o(1)).
 \end{aligned}$$

Thus, the two-dimensional distributions of $Y_{2n,N}$ converge to the two-dimensional distributions of W_0 .

Now, we sketch the proof for the l -dimensional distributions. Let $0 < t_1 < t_2 < \dots < t_l < 1$. Then,

$$\begin{aligned}
 & \mathbb{P}\{X_{2n,N}(t_1) = 2k_1, X_{2n,N}(t_2) - X_{2n,N}(t_1) = 2k_2, \dots, X_{2n,N}(t_l) - X_{2n,N}(t_{l-1}) = 2k_l\} \\
 = & \frac{\mathbb{P}\{S_{[Nt_1]} = 2k_1\} \mathbb{P}\{S_{[Nt_2]-[Nt_1]} = 2k_2\} \dots \mathbb{P}\{S_{[Nt_l]-[Nt_{l-1}]} = 2k_l\} \mathbb{P}\{S_{N-[Nt_l]} = 2n - 2(k_1 + \dots + k_l)\}}{\mathbb{P}\{S_N = 2n\}} \\
 = & \frac{2 \exp\left(-\frac{(2k_1-2n\frac{[t_1N]}{N})^2}{4nt_1}\right)}{\sqrt{4\pi nt_1}} \prod_{j=2}^l \frac{2 \exp\left(-\frac{(2k_j-2n\frac{[t_jN]-[t_{j-1}N]}{N})^2}{4n(t_j-t_{j-1})}\right)}{\sqrt{4\pi n(t_j-t_{j-1})}} \frac{2 \exp\left(-\frac{(2k_1+\dots+2k_l-2n\frac{[t_lN]}{N})^2}{4n(1-t_l)}\right)}{\sqrt{4\pi n(1-t_l)}} \\
 = & \frac{\frac{2}{\sqrt{4\pi n}}(1+o(1))}{\frac{2}{\sqrt{4\pi n}}(1+o(1))} \\
 = & \left(\frac{2}{\sqrt{2n}}\right)^l \frac{1}{(2\pi)^{l/2}\sqrt{\prod_{r=1}^{l+1}(t_r-t_{r-1})}} \exp\left(\frac{-U(\mathbf{x})}{2}\right) (1+o(1)), \tag{50}
 \end{aligned}$$

where $t_0 = 0, t_{l+1} = 1$, and the quadratic form $U(\mathbf{x})$ has the following shape

$$U(\mathbf{x}) = \frac{x_1^2}{b_1} + \dots + \frac{x_l^2}{b_l} + \frac{(x_1 + \dots + x_l)^2}{1 - (b_1 + \dots + b_l)}.$$

Here, we used the notation

$$x_j = \frac{2k_j - 2n \frac{[t_j N] - [t_{j-1} N]}{N}}{\sqrt{2n}}, \quad b_j = t_j - t_{j-1}, \quad j = 1, 2, \dots, l. \tag{51}$$

Now, by some algebra, we see that

$$U(\mathbf{x}) = \frac{\sum_{j=1}^l \left(x_j^2 (1 - \sum_{i \neq j} b_i) \prod_{i \neq j} b_i \right) + \sum_{i \neq j} x_i x_j \prod_{r=1}^l b_r}{(1 - \sum_{r=1}^l b_r) \prod_{r=1}^l b_r}.$$

We need the $l \times l$ -type matrix

$$D = \begin{bmatrix} b_1(1 - b_1) & -b_1 b_2 & -b_1 b_3 & \dots & -b_1 b_l \\ -b_2 b_1 & b_2(1 - b_2) & -b_2 b_3 & \dots & -b_2 b_l \\ \vdots & \vdots & \ddots & \dots & \vdots \\ -b_l b_1 & -b_l b_2 & -b_l b_3 & \dots & b_l(1 - b_l) \end{bmatrix},$$

and its inverse

$$D^{-1} = \frac{1}{(1 - \sum_{r=1}^l b_r) \prod_{r=1}^l b_r} \begin{bmatrix} (1 - \sum_{r \neq 1} b_r) \prod_{r \neq 1} b_r & \prod_{r=1}^l b_r & \prod_{r=1}^l b_r & \dots & \prod_{r=1}^l b_r \\ \prod_{r=1}^l b_r & (1 - \sum_{r \neq 2} b_r) \prod_{r \neq 2} b_r & \prod_{r=1}^l b_r & \dots & \prod_{r=1}^l b_r \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \prod_{r=1}^l b_r & \prod_{r=1}^l b_r & \prod_{r=1}^l b_r & \dots & (1 - \sum_{r \neq l} b_r) \prod_{r \neq l} b_r \end{bmatrix}.$$

We can see that we obtain the covariance matrix of the increments

$$(W_0(t_1) - W_0(0), W_0(t_2) - W_0(t_1), \dots, W_0(t_l) - W_0(t_{l-1}))$$

of the Brownian bridge W_0 if we insert $b_j = t_j - t_{j-1}, j = 1, 2, \dots, l$, into the matrix D . Denote this matrix by Σ (for any fixed value of l). The determinant of D is $(1 - \sum_{r=1}^l b_r) \prod_{r=1}^l b_r$, so the determinant of Σ is $\prod_{r=1}^{l+1} (t_r - t_{r-1})$. We can also check that the matrix of the quadratic form $U(\mathbf{x})$ is D^{-1} .

So, by using the above considerations, Equation (50) gives

$$\begin{aligned} & \mathbb{P}\{X_{2n,N}(t_1) = 2k_1, X_{2n,N}(t_2) - X_{2n,N}(t_1) = 2k_2, \dots, X_{2n,N}(t_l) - X_{2n,N}(t_{l-1}) = 2k_l\} \\ &= \left(\frac{2}{\sqrt{2n}} \right)^l \frac{1}{(2\pi)^{l/2} \sqrt{\det \Sigma}} \exp\left(\frac{-\mathbf{x}^\top \Sigma^{-1} \mathbf{x}}{2} \right) (1 + o(1)), \end{aligned} \tag{52}$$

where

$$\mathbf{x} = (x_1, x_2, \dots, x_l)^\top$$

and x_j is defined in (51). It implies that the finite dimensional distributions of $Y_{2n,N}$ converge to the finite dimensional distributions of W_0 . \square

Remark 1. Relation (52) is a local limit theorem for the random allocation.

6. Discussion

The random allocation of particles into cells is a well-known model in probability theory. There are limit theorems when either the number of particles or the number of

cells or both of them tend to infinity, see [1]. The errors in the blocks of a binary file can be modelled as a random allocation. However, if the parity bits are used, then any odd number of errors in the blocks is always detected, but an even number of errors is never detected.

Therefore, describing the behaviour of the allocation model is an interesting problem when there is an even number of particles in each cell.

In this paper, we consider the numbers of particles in cells when we allocate $2n$ distinguishable particles into N distinct cells having an even number of particles in each cell. For the non-homogeneous case, we study the numbers of particles in the first K cells. We were able to prove the asymptotic normality of this K -dimensional random vector when $n, N \rightarrow \infty$. For the homogeneous allocation model, we proved a limit theorem to the finite dimensional distributions of the Brownian bridge, if $n, N \rightarrow \infty$. To handle the mathematical problem, we inserted our model into the framework of Kolchin's generalized allocation scheme. Using the above limit theorems, we obtained two χ^2 -tests. As the parity bit method does not detect any even number of errors in the blocks of a binary file, we suggest applying our model to study the distribution of errors in that file.

Author Contributions: Conceptualization, A.N.C.; methodology, A.N.C.; software, I.F.; formal analysis, A.N.C. and I.F.; investigation, A.N.C. and I.F.; writing—original draft preparation, A.N.C.; writing—review and editing, I.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank the referees for the helpful remarks.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kolchin, V.F.; Sevast'ynov, B.A.; Chistiakov, V.P. *Random Allocations*; Scripta Series in Mathematics; V. H. Winston & Sons: Washington, DC, USA, 1978.
2. Chikrin, D.E.; Chuprunov, A.N.; Kokunin, P.A. Gaussian limit theorems for the number of given value cells in the non-homogeneous generalized allocation scheme. *J. Math. Sci.* **2020**, *246*, 476–487.
3. Billingsley, P. *Convergence of Probability Measures*; Wiley: New York, NY, USA, 1968.
4. Kolchin, V.F. A class of limit theorems for conditional distributions. *Lith. Math. J.* **1968**, *8*, 53–63. (In Russian)
5. Kolchin, V.F. *Random Graphs*; Cambridge University Press: Cambridge, UK, 1998.
6. Abdushukurov, F.A.; Chuprunov, A.N. Poisson limit theorems in an allocation scheme with even number of particles in each cell. *Lobachevskii J. Math.* **2020**, *41*, 289–297.
7. Timashev, A.M. *Asymptotic Expansions in Probabilistic Combinatorics*; TVP Science Publishers: Moscow, Russia, 2011. (In Russian)