

# A Chinese restaurant process for multiset permutations

Dudley Stark

*Queen Mary, University of London, Mile End, London E1 4NS, United Kingdom*

[d.s.stark@qmul.ac.uk](mailto:d.s.stark@qmul.ac.uk) (D. Stark)

Received: 17 September 2025, Revised: 11 February 2026, Accepted: 13 February 2026,  
Published online: 3 March 2026

**Abstract** Multisets are like sets, except that they can contain multiple copies of their elements. If there are  $n_i$  copies of  $i$ ,  $1 \leq i \leq t$ , in multiset  $M_t$ , then there are  $\binom{n_1 + \dots + n_t}{n_1, \dots, n_t}$  possible permutations of  $M_t$ . Knuth showed how to factor any multiset permutation into cycles. For fixed  $n_i$ ,  $i \geq 1$ , we show how to adapt the Chinese restaurant process, which generates random permutations on  $n$  elements with weighting  $\theta^{\#\text{cycles}}$ ,  $\theta > 0$ , sequentially for  $n = 1, 2, \dots$ , to the multiset case, where we fix the  $n_i$  and build permutations on  $M_t$  sequentially for  $t = 1, 2, \dots$ . The number of cycles of a multiset permutation chosen uniformly at random, i.e.  $\theta = 1$ , has distribution given by the sum of independent negative hypergeometric distributed random variables. For all  $\theta > 0$ , and under the assumption that  $n_i = O(1)$ , we show a central limit theorem as  $t \rightarrow \infty$  for the number of cycles.

**Keywords** Chinese restaurant process, random permutations, multiset

**2020 MSC** 60C05, 60J10, 60F05

## 1 Introduction

A growth rule which received considerable attention in the literature is the following preferential attachment algorithm on permutations written as the product of cycles known as the Chinese restaurant process (CRP) [1, 5, 9, 12–14, 16, 20, 23]. For  $n \geq 1$ , given a permutation of  $[n-1] := \{1, 2, \dots, n-1\}$  which has been constructed at step  $n-1$ , element  $n$  is either appended to the permutation as a new singleton cycle with probability  $\frac{\theta}{\theta+n-1}$ , or inserted into a random position within the existing permutation with probability equal to  $\frac{n-1}{\theta+n-1}$ . Let  $S_n$  be the set of permutations on  $[n]$ . The permutation obtained at step  $n$  has the Ewens distribution, which is the uniform distribution on  $n!$  permutations in the case  $\theta = 1$ , and, in general, has probability

© 2026 The Author(s). Published by VTeX. Open access article under the [CC BY](#) license.

distribution  $\theta^{|\pi|}/\theta_{(n)}$ ,  $\pi \in S_n$ , where  $\theta_{(n)} := \theta(\theta + 1) \cdots (\theta + n - 1)$  and  $|\pi|$  is the number of cycles in  $\pi$ .

Let  $\Pi_n$  be the random permutation of  $[n]$  at the  $n$ th step of the CRP. The distribution of  $\Pi_n$  is invariant under relabelling. For different orders the permutations are consistent, in the sense that  $\Pi_m$ , for  $m < n$ , can be derived from  $\Pi_n$  by removing elements  $m + 1, \dots, n$  from their cycles and deleting empty cycles if necessary.

The CRP expresses every permutation  $\pi$  of  $[n]$  uniquely as the product

$$\pi = (x_{1,1} \dots x_{1,m_1} y_1)(x_{2,1} \dots x_{2,m_2} y_2) \cdots (x_{\tau,1} \dots x_{\tau,m_\tau} y_\tau), \quad \tau \geq 1,$$

where the following two conditions are satisfied:

1.  $y_1 < y_2 < \cdots < y_\tau$
2.  $y_i < x_{ij}$  for  $1 \leq j \leq m_i, 1 \leq i \leq \tau$ .

We will see in Section 2 that there is a similar unique expression for permutations of multisets.

The number of cycles in  $\Pi_n$ , denoted by  $K_n$ , has probability generating function (p.g.f.)

$$\mathbb{E}_z^{K_n} = \frac{(\theta z)_n}{(\theta)_n},$$

which corresponds to the distribution of the sum of  $n$  independent Bernoulli variables  $K_n = \sum_{i=1}^n I_i$ ,  $I_i \sim \text{Bernoulli}(\theta/(i + \theta - 1))$ ; see [1]. We have  $I_i = 1$  exactly when the element  $i$  starts a new cycle. As  $n \rightarrow \infty$ ,

$$K_n \sim \theta \log n \text{ a.s.}, \quad \frac{K_n - \theta \log n}{\sqrt{\theta \log n}} \xrightarrow{d} N(0, 1),$$

where  $\xrightarrow{d}$  denotes convergence in distribution.

In this paper, a CRP will be defined for multiset permutations and derived using the methods in [23]. As motivation, we note that multiset permutations are of fundamental interest in combinatorics [22] and random multiset permutations have been studied previously [10].

## 2 Multiset permutations and their factorisations into cycles

There are at least two different ways of defining cycles of multiset permutations [19, 22]. Knuth's way [19] will be used in this paper. In this section, we summarise the discussion and results on multiset permutation factorisation taken from Section 5.1.2 of [19] that is needed in this paper. In particular, we are going to use Theorem 1 below to factor multiset permutations. A multiset is like a set except that it can have repetitions of identical elements, for example

$$M = \{1, 1, 1, 2, 2, 3, 4, 4, 4, 4\},$$

where the set  $\{1, 2, 3, 4\}$  is given the usual ordering. A permutation of a multiset is an arrangement of its elements in a row, for example

$$3 \ 1 \ 2 \ 4 \ 4 \ 1 \ 2 \ 4 \ 1 \ 4.$$

We may write this in two-line notation as

$$\left( \begin{array}{cccccccccccc} 1 & 1 & 1 & 2 & 2 & 3 & 4 & 4 & 4 & 4 \\ 3 & 1 & 2 & 4 & 4 & 1 & 2 & 4 & 1 & 4 \end{array} \right). \quad (1)$$

The *intercalation product*  $\pi_1 \perp \pi_2$  of two multiset permutations  $\pi_1$  and  $\pi_2$  is defined in [6, 11] and is obtained by

1. Expressing  $\pi_1$  and  $\pi_2$  in the two-line notation.
2. Juxtaposing these two-line notations.
3. Sorting the columns into nondecreasing order of the top line, in such a way that the sorting is stable, in the sense that left-to-right order in the bottom line is preserved when the corresponding top line elements are equal.

For example,

$$\begin{aligned} & \left( \begin{array}{cccccc} 1 & 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 1 & 2 \end{array} \right) \top \left( \begin{array}{cccccc} 1 & 2 & 4 & 4 & 4 \\ 2 & 4 & 4 & 1 & 4 \end{array} \right) \\ &= \left( \begin{array}{cccccccccccc} 1 & 1 & 1 & 2 & 2 & 3 & 4 & 4 & 4 & 4 \\ 3 & 1 & 2 & 4 & 4 & 1 & 2 & 4 & 1 & 4 \end{array} \right) \end{aligned}$$

For possibly repeated elements  $x_1, x_2, \dots, x_m$ , the *cycle*  $(x_1, \dots, x_n)$  stands for the permutation obtained in two line form by sorting the columns of

$$\left( \begin{array}{cccc} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_1 \end{array} \right)$$

in a stable manner. For example, we have

$$\begin{aligned} & (4\ 2\ 4\ 4\ 1\ 3\ 1\ 1\ 2\ 4) \\ &= \left( \begin{array}{cccccccccc} 4 & 2 & 4 & 4 & 1 & 3 & 1 & 1 & 2 & 4 \\ 2 & 4 & 4 & 1 & 3 & 1 & 1 & 2 & 4 & 4 \end{array} \right) \\ &= \left( \begin{array}{cccccccccccc} 1 & 1 & 1 & 2 & 2 & 3 & 4 & 4 & 4 & 4 \\ 3 & 1 & 2 & 4 & 4 & 1 & 2 & 4 & 1 & 4 \end{array} \right) \quad (2) \end{aligned}$$

and so (1) is a cycle. For these general cycles,  $(x_1x_2 \dots x_n)$  is not always the same as  $(x_2 \dots x_nx_1)$  and it can happen that a cycle can be written as the intercalation product of other cycles, as we shall see at the end of this section.

Knuth [19] proves the following factorisation of multiset permutations into cycles.

**Theorem 1 ([19]).** *Let the elements of the multiset  $M$  be linearly ordered by the relation  $<$ . Every permutation  $\pi$  of  $M$  has a unique representation as the intercalation*

$$\pi = (x_{1,1} \dots x_{1,m_1}y_1) \top (x_{2,1} \dots x_{2,m_2}y_2) \top \dots \top (x_{\tau,1} \dots x_{\tau,n_\tau}y_\tau), \quad \tau \geq 1, \quad (3)$$

where the following two conditions are satisfied:

1.  $y_1 \leq y_2 \leq \dots \leq y_\tau$

2.  $y_i < x_{i,j}$  for  $1 \leq j \leq m_i$ ,  $1 \leq i \leq \tau$ .

Note that (2) is not in the form (3). The factorisation of (1) corresponding to Theorem 1 is

$$\begin{aligned}
 & (3\ 1) \top (1) \top (2\ 4\ 2\ 4\ 4\ 1) \top (4) \\
 = & \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \top \begin{pmatrix} 1 \\ 1 \end{pmatrix} \top \begin{pmatrix} 2 & 4 & 2 & 4 & 4 & 1 \\ 4 & 2 & 4 & 4 & 1 & 2 \end{pmatrix} \top \begin{pmatrix} 4 \\ 4 \end{pmatrix} \\
 = & \begin{pmatrix} 3 & 1 & 1 & 2 & 4 & 2 & 4 & 4 & 1 & 4 \\ 1 & 3 & 1 & 4 & 2 & 4 & 4 & 1 & 2 & 4 \end{pmatrix} \\
 = & \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 3 & 4 & 4 & 4 & 4 \\ 3 & 1 & 2 & 4 & 4 & 1 & 2 & 4 & 1 & 4 \end{pmatrix}.
 \end{aligned}$$

### 3 The CRP for multiset permutations

Given a fixed sequence of positive integers  $n_i \geq 1$ ,  $i \geq 1$ , we will construct a CRP process on permutations of multisets indexed by integers  $t \geq 0$ . We define  $M(n_1, \dots, n_t)$  to be the multiset which contains  $n_i$  copies of  $i$  for  $i \in [t]$ ,  $t \geq 1$ , and  $\emptyset$  for  $t = 0$ . We give the set of elements  $[t]$  of  $M(n_1, \dots, n_t)$  the usual ordering. The set of permutations of  $M(n_1, \dots, n_t)$  is denoted by  $S(n_1, \dots, n_t)$  and is of size

$$|S(n_1, \dots, n_t)| = \binom{N_t}{n_1, \dots, n_t},$$

where  $N_t = n_1 + \dots + n_t$ . Note that  $S_n = S(\underbrace{1, 1, \dots, 1}_{n \text{ times}})$ . The random multiset permu-

tation  $\tilde{\Pi}_t$  of our process at the  $t$ th step is an element of  $S(n_1, \dots, n_t)$ , represented in accordance with (3). At step  $t \geq 1$ , there are  $n_t$  copies of  $t$  either to be inserted in the existing permutation  $\tilde{\Pi}_{t-1}$  or to be put into singleton cycles. (It is not possible to put more than one copy of  $t$  in a new cycle because of condition 2 of Theorem 1.) Suppose that  $k_t \in \{0\} \cup [n_t]$  of the  $n_t$  copies of  $t$  start  $k_t$  new singleton cycles containing  $t$ , and the  $n_t - k_t$  other copies of  $t$  are inserted into the existing permutation. The  $n_t - k_t$  copies of  $t$  can be inserted to the left of any of the  $y_i$  or  $x_{ij}$  in (3). A ‘stars and bars’ argument shows that the number of ways of inserting  $n_t - k_t$  copies of  $t$  to the left of an element in the multiset  $S(n_1, \dots, n_{t-1})$  formatted like (3) is

$$\binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t}, \quad (4)$$

recalling that

$$\binom{-1}{k} = \begin{cases} 1 & \text{if } k = 0; \\ 0 & \text{if } k > 1. \end{cases}$$

We now define the multiset CRP. Given  $\theta > 0$ , define

$$F_t(\theta) = \sum_{k_t=0}^{n_t} \binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t} \theta^{k_t}$$

The special case

$$F_t(1) = \sum_{k_t=0}^{n_t} \binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t} = \binom{N_t}{n_t}. \quad (5)$$

results from the well known identity

$$\sum_{\rho=0}^{\nu} \binom{\mu + \rho}{\rho} = \binom{\mu + \nu + 1}{\nu}, \quad \nu, \mu \geq 0 \text{ integers}; \quad (6)$$

see page 161 of [15]. Note that (5) is the total number of ways of extending the permutation from time  $t - 1$  to time  $t$ . At time  $t \geq 1$ ,  $k_t$  new singleton cycles containing  $t$  are created and the other  $n_t - k_t$  copies are inserted into the existing permutation in a specified way with probability  $\theta^{k_t}/F_t(\theta)$ . If  $n_t = 1$  for all  $t$ , then  $F_t(\theta) = \theta + t - 1$  and we recover the usual CRP. An induction argument on  $t$  shows that all possible  $\pi \in S(n_1, \dots, n_t)$  can be constructed in this manner. This can also be seen from the identity

$$\binom{N_t}{n_1, \dots, n_t} = \prod_{s=1}^t \binom{N_s}{n_s}$$

Next we will find the distribution of  $\tilde{\Pi}_t$ .

**Theorem 2.** *Given  $\theta > 0$ , for all  $\pi \in S(n_1, \dots, n_t)$  the CRP generates  $\pi$  at step  $t$  with probability equal to  $\theta^{|\pi|}/S$ , where*

$$S = \sum_{\pi \in S(n_1, \dots, n_t)} \theta^{|\pi|} = \prod_{s=1}^t F_s(\theta). \quad (7)$$

**Proof.** We will use the method in [23]. An *arborescence* is a directed, rooted tree with edges oriented in agreement with paths from the root to the leaves. Our arborescence has vertex set  $V = \bigcup_{s=0}^t \Pi(n_1, \dots, n_s)$  and the root is  $\emptyset$  (when  $s = 0$ ). The leaves of the arborescence are  $\Pi(n_1, \dots, n_t)$ . Give each leaf  $\pi \in \Pi(n_1, \dots, n_t)$  the weight  $\theta^{|\pi|}$ , where  $|\pi|$  is the number of cycles in  $\pi$ . There is a directed edge from  $\pi_1 \in \Pi(n_1, \dots, n_{s-1})$  to  $\pi_2 \in \Pi(n_1, \dots, n_s)$ ,  $s \in [t]$ , if and only if, for some  $k_s \in \{0\} \cup [n_s]$ ,  $\pi_2$  is obtained from  $\pi_1$  by inserting  $n_s - k_s$  copies of  $s$  to the left of elements of  $\pi_1$ , when  $\pi_1$  is written as (1), and creating  $k_s$  singleton cycles each containing  $s$ . For each  $\pi \in V$ , define  $Q(\pi)$  to be the sum of the weights of all leaves  $\pi' \in \Pi(n_1, \dots, n_t)$  for which there is a directed path from  $\pi$  to  $\pi'$ . For  $\pi \in \Pi(n_1, \dots, n_s)$ , we have

$$Q(\pi) = \theta^{|\pi|} \prod_{u=s+1}^t F_u(\theta), \quad (8)$$

which follows immediately from the interpretation of the binomial coefficient preceding (4). Taking  $\pi = \emptyset$  and  $s = 0$  in (8) gives (7). With  $\pi_1$  and  $\pi_2$  taken as above, we have  $|\pi_2| = |\pi_1| + k_s$ , and therefore

$$p_{\pi_1, \pi_2} := \frac{Q(\pi_2)}{Q(\pi_1)} = \frac{\theta^{k_s}}{F_s(\theta)}. \quad (9)$$

These are the transition probabilities of the CRP. Theorem 1 of [23] states that if we put transition probabilities (9) on the edges of the arborescence, and let the leaves be absorbing states, then we obtain a Markov chain whose probability of absorption on a given leaf is proportional to the weight of the leaf. By our choice of weight we are done.  $\square$

For different steps the permutations are consistent, in the sense that  $\tilde{\Pi}_s$ , for  $s < t$ , can be derived from  $\tilde{\Pi}_t$  by removing elements  $s+1, \dots, t$  from their cycles and deleting empty cycles if necessary.

#### 4 The distribution of the number of cycles

Let  $X_t$  denote the number of new cycles added in moving from time  $t-1$  to time  $t$ ,  $t \geq 1$ , which is the choice of  $k_t$  above. The  $X_t$  are independent with distributions given by

$$\mathbf{P}(X_t = k_t) = \frac{\theta^{k_t}}{F_t(\theta)} \binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t}, \quad k_t \in \{0\} \cup [n_t]. \quad (10)$$

If  $\theta = 1$ , then, by (5), we have

$$\mathbf{P}(X_t = k_t) = \frac{1}{\binom{N_t}{n_t}} \binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t} = \frac{\binom{n_t}{k_t}}{\binom{N_t}{k_t}} \frac{N_t - n_t}{N_t - k_t} \quad (11)$$

This equals the probability that starting from an urn with  $n_t$  balls representing failures and  $N_t - n_t$  balls representing successes, that the first  $k_t$  balls pulled from the urn without replacement are failures, and the  $(k_t + 1)$ th ball pulled from an urn is a success. This description shows that  $Y_t = X_t + 1$  has a negative hypergeometric distribution, which we now describe. The negative hypergeometric distribution [17, 18, 21] with parameters  $N$ ,  $M$  and  $r$  is like the negative binomial distribution, but without replacement. There are  $N$  balls in total,  $M$  balls representing success and  $N - M$  balls representing failures. The hypergeometric distribution gives the probability that the  $r$ th success happens on the  $\kappa$ th draw,  $\kappa = r, r+1, \dots, N$ . A negative hypergeometric random variable  $Y$  with these parameters has probability mass function

$$\mathbb{P}(Y = \kappa) = \frac{\binom{M}{r-1} \binom{N-M}{\kappa-r}}{\binom{N}{\kappa-1}} \cdot \frac{M-r+1}{N-\kappa+1} \quad (12)$$

It is easily checked that (11) is (12) with  $N = N_t$ ,  $M = N_t - n_t = N_{t-1}$ ,  $r = 1$  and  $\kappa = k + 1$ . The expectation of  $Y$  is

$$\mathbb{E}(Y) = r \frac{N+1}{M+1},$$

the variance of  $Y$  is

$$\text{Var}(Y) = \frac{r(N-M)(N+1)(M+1-r)}{(M+1)^2(M+2)}$$

and the third centred moment (see [17, 18]) is

$$\mathbb{E}([Y - \mathbb{E}(Y)]^3) = \text{Var}(Y) \cdot \frac{(2N - M + 1)(M + 1 - 2r)}{(M + 1)(M + 3)}.$$

Therefore,

$$\mathbb{E}(X_t) = \mathbb{E}(Y_t) - 1 = \frac{N_t + 1}{N_{t-1} + 1} - 1 = \frac{n_t}{N_{t-1} + 1}, \quad (13)$$

$$\text{Var}(X_t) = \text{Var}(Y_t) = \frac{n_t(N_t + 1)N_{t-1}}{(N_{t-1} + 1)^2(N_{t-1} + 2)} \quad (14)$$

and

$$\begin{aligned} \mathbb{E}([X_t - \mathbb{E}(X_t)]^3) &= \mathbb{E}([Y_t - \mathbb{E}(Y_t)]^3) \\ &= \text{Var}(Y_t) \cdot \frac{(n_t + N_t + 1)(N_{t-1} - 1)}{(N_{t-1} + 1)(N_{t-1} + 3)}. \end{aligned} \quad (15)$$

It follows that the number of cycles  $K_t = \sum_{s=1}^t X_s$  has expectation

$$\mathbb{E}(K_t) = \sum_{s=1}^t \frac{n_s}{N_{s-1} + 1}$$

and variance

$$\text{Var}(K_t) = \sum_{s=1}^t \frac{n_s(N_s + 1)N_{s-1}}{(N_{s-1} + 1)^2(N_{s-1} + 2)}.$$

Given sequences  $a_t$  and  $b_t$ , we use the notation  $a_t = O(b_t)$  to mean  $a_t \leq Cb_t$  for a constant  $C > 0$ ;  $a_t \asymp b_t$  to mean  $C_1b_t \leq a_t \leq C_2b_t$  for constants  $C_1 > 0$ ,  $C_2 > 0$ ; and  $a_t \sim b_t$  to mean  $\lim_{t \rightarrow \infty} a_t/b_t = 1$ . We also define  $a_t = o(b_t)$  to mean  $\lim_{t \rightarrow \infty} a_t/b_t = 0$ . We will put the bound the  $n_t = O(1)$  to give a central limit theorem for  $K_t$  for all  $\theta$ .

**Theorem 3.** *If  $\theta = 1$  and  $n_i = n \geq 1$  for all  $i \geq 1$ , then*

$$\mathbb{E}(K_t) \sim \log t, \quad \text{Var}(K_t) \sim \log t. \quad (16)$$

*For general  $\theta$ , let  $n_1, n_2, \dots$ , be given,  $n_i \geq 1$  for all  $i$ , and suppose*

$$n_i = O(1). \quad (17)$$

*Then,*

$$\mathbb{E}(K_t) \asymp \log t, \quad \text{Var}(K_t) \asymp \log t, \quad (18)$$

*and*

$$K_n \sim \mathbb{E}(K_n) \text{ a.s.}, \quad \frac{K_t - \mathbb{E}(K_t)}{\sqrt{\text{Var}(K_t)}} \xrightarrow{d} N(0, 1). \quad (19)$$

**Proof.** First we prove the theorem for  $\theta = 1$ . From (13), (14) and (17) we get  $\mathbb{E}(X_t) \asymp 1/t$  and  $\text{Var}(X_t) \asymp 1/t$ , which result in (18), after which the Lindeberg-Feller Theorem easily shows the central limit theorem in (19), because  $X_t \leq n_t = O(1)$ . If we impose the condition  $n_i = n$ , for  $i \geq 1$ , then we immediately obtain (16). For  $\theta$  general, we reduce to the case  $\theta = 1$ . This follows from noting that in (10),  $\theta^{k_t} \asymp 1$  and  $F_t(\theta) \asymp F_t(1)$  uniformly over  $t \geq 1$ , by (17), and so (18) still holds. For the almost sure convergence, modify the proof of (6.6) Theorem on page 44 of [8].  $\square$

By the following theorem a Central Limit Theorem can be obtained even when  $n_t$  is unbounded, which is not the case in Theorem 3.

**Theorem 4.** *Suppose  $\theta = 1$ ,  $n_t = O(N_{t-1})$ . Then, (19) holds.*

**Proof.** In order to prove limiting normality, we will verify Lyapounov's Condition in [3] with  $\delta = 1$ , by showing that

$$\lim_{t \rightarrow \infty} \frac{1}{\text{Var}(K_t)^{3/2}} \sum_{s=1}^t \mathbb{E}([X_s - \mathbb{E}(X_s)]^3) = 0. \quad (20)$$

The assumption  $n_t = O(N_{t-1})$ , (14) and (15) produces

$$\mathbb{E}([X_t - \mathbb{E}(X_t)]^3) \asymp \text{Var}(X_t) \asymp n_t/N_{t-1}, \quad t \geq 2.$$

We have the lower bound

$$\begin{aligned} \sum_{s=2}^t \frac{n_s}{N_{s-1}} &\geq \sum_{s=2}^t \ln \left( 1 + \frac{n_s}{N_{s-1}} \right) \\ &= \sum_{s=2}^t (\ln(N_s) - \ln(N_{s-1})) \\ &= \ln(N_t) - \ln(N_1) \rightarrow \infty, \end{aligned}$$

where the convergence to infinity follows from the assumption  $n_t \geq 1$  for all  $t \geq 1$ . Hence,

$$\text{Var}(K_t) \asymp \sum_{s=1}^t \mathbb{E}([X_s - \mathbb{E}(X_s)]^3) \asymp \sum_{s=2}^t \frac{n_s}{N_{s-1}} \rightarrow \infty$$

and (20) holds.

For the almost sure convergence, modify the proof of (6.6) Theorem on page 44 of [8] as in the proof of Theorem 3.  $\square$

An example where Theorem 4 holds is obtained by setting  $\theta = 1$ ,  $n_t = \lceil Ct^\alpha \rceil$ , for constants  $C > 0$ ,  $\alpha > 0$ , where  $\lceil x \rceil$  is the smallest integer greater or equal to  $x$ . Then,  $n_t = Ct^\alpha + O(1)$  and  $N_t = Ct^{\alpha+1}/(\alpha+1) + O(t^\alpha)$  and the assumptions of Theorem 4 hold. The expectation and variance are asymptotically

$$\mathbb{E}(K_t) \sim (\alpha+1) \log t, \quad \text{Var}(K_t) \sim (\alpha+1) \log t.$$

Another example for  $\theta = 1$  is obtained through regular variation [4]. A positive, measurable function  $\ell(x)$ , defined on  $[\eta, \infty)$  for  $\eta$  real, is *slowly varying* if

$$\lim_{x \rightarrow \infty} \frac{\ell(\lambda x)}{\ell(x)} = 1, \quad \forall \lambda > 0. \quad (21)$$

The Uniform Convergence Theorem [4] provides that the convergence in (21) is uniform on compact  $\lambda$ -sets in  $(0, \infty)$ . A positive, measurable function defined on  $[\eta, \infty)$ , is *regularly varying* with real index  $\alpha$  if

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha, \quad \forall \lambda > 0.$$

A regularly varying function with index  $\alpha$  can be written as

$$f(x) = x^\alpha \ell(x),$$

where  $\ell(x)$  is a slowly varying function. Suppose that  $\ell(x)$  is locally bounded in  $[\eta, \infty)$ , i.e. bounded on compact subsets of  $[\eta, \infty)$ , and  $\alpha > -1$ . Karamata's theorem gives

$$\int_{\eta}^t x^\alpha \ell(x) dx \sim t^{\alpha+1} \ell(t) / (\alpha + 1), \quad t \rightarrow \infty.$$

We further suppose that  $\eta \leq 1$ ,  $\alpha \geq 0$ , with  $\lim_{x \rightarrow \infty} \ell(x) = \infty$  in the case  $\alpha = 0$ , and define

$$n_t = \left\lceil \int_t^{t+1} x^\alpha \ell(x) dx \right\rceil, \quad t \geq 1.$$

The Uniform Convergence Theorem implies that

$$\sup_{t \leq x \leq t+1} \left| \frac{\ell(x)}{\ell(t)} - 1 \right| \leq \sup_{1 \leq \lambda \leq 2} \left| \frac{\ell(\lambda t)}{\ell(t)} - 1 \right| = o(1),$$

from which

$$\begin{aligned} \int_t^{t+1} x^\alpha \ell(x) dx &= (1 + o(1)) \ell(t) \int_t^{t+1} x^\alpha dx \\ &= (1 + o(1)) \ell(t) (t^\alpha + o(t^\alpha)) \end{aligned}$$

and

$$n_t \sim t^\alpha \ell(t).$$

Moreover, Karamata's theorem gives us

$$N_{t-1} \sim t^{\alpha+1} \ell(t) / (\alpha + 1).$$

Hence, the conditions of Theorem 4 hold.

We will now show that  $n_t$  may be chosen in such a way that the conclusion of the central limit theorem does not hold for  $K_t$ . Suppose that  $\theta = 1$  and the  $n_t$  are chosen such that

$$N_{t-1} = o(n_t). \tag{22}$$

By (13) and (14), we have

$$\mathbb{E}(X_t) \sim \frac{n_t}{N_{t-1}}, \quad \text{Var}(X_t) \sim \left( \frac{n_t}{N_{t-1}} \right)^2.$$

We further choose the  $n_t$  such that

$$\sum_{s=1}^{t-1} \text{Var}(X_s) = o(\text{Var}(X_t)), \tag{23}$$

A particular example of such a sequence is

$$n_t = \lceil e^{t^3} \rceil = e^{t^3} + O(1),$$

from which

$$\begin{aligned}
 N_t &= \sum_{s=1}^t e^{s^3} + O(t) \\
 &= e^{t^3} \left( 1 + \sum_{s=1}^{t-1} e^{s^3 - t^3} \right) + O(t) \\
 &= e^{t^3} \left( 1 + O\left( t e^{(t-1)^3 - t^3} \right) \right) + O(t) \\
 &= e^{t^3} (1 + o(1)).
 \end{aligned}$$

Clearly, (22) holds and we also have

$$\begin{aligned}
 \sum_{s=1}^{t-1} \text{Var}(X_s) &= O\left( \sum_{s=1}^{t-1} e^{2s^3 - 2(s-1)^3} \right) \\
 &= O\left( \sum_{s=1}^{t-1} e^{6s^2 - 6s} \right) \\
 &= O\left( t e^{6(t-1)^2 - 6(t-1)} \right) \\
 &= o\left( e^{6t^2 - 6t} \right)
 \end{aligned}$$

and so (23) holds, as well, because  $\text{Var}(X_t) \sim e^{6t^2 - 6t + 2}$ .

Assuming (22) and (23), we have

$$\frac{K_t - \mathbb{E}(K_t)}{\sqrt{\text{Var}(K_t)}} = \frac{X_t - \mathbb{E}(X_t)}{\sqrt{\text{Var}(X_t)}} (1 + o(1)) + \frac{\sum_{s=1}^{t-1} (X_s - \mathbb{E}(X_s))}{\sqrt{\text{Var}(K_t)}}. \quad (24)$$

Suppose that

$$\frac{K_t - \mathbb{E}(K_t)}{\sqrt{\text{Var}(K_t)}} \xrightarrow{d} \mathbf{N}(0, 1).$$

By our assumptions, the variance of the last term of (24) is  $o(1)$  and so by Slutsky's theorem [8]

$$\lim_{t \rightarrow \infty} \mathbb{P}\left( \frac{X_t - \mathbb{E}(X_t)}{\sqrt{\text{Var}(K_t)}} \leq 0 \right) = \lim_{t \rightarrow \infty} \mathbb{P}\left( \frac{K_t - \mathbb{E}(K_t)}{\sqrt{\text{Var}(X_t)}} \leq 0 \right) = 1/2, \quad (25)$$

We will show that (25) is impossible. The c.d.f. of  $X_t$  is given by

$$\begin{aligned}
 P(X_t \leq x_t) &= \sum_{k_t=0}^{x_t} \mathbb{P}(X_t = k_t) \\
 &= \frac{1}{\binom{N_t}{n_t}} \sum_{k_t=0}^{x_t} \binom{N_{t-1} - 1 + n_t - k_t}{n_t - k_t}
 \end{aligned} \quad (26)$$

$$\begin{aligned}
 &= \frac{1}{\binom{N_t}{n_t}} \left( \sum_{j=0}^{n_t} \binom{N_{t-1} - 1 + j}{j} - \sum_{j=0}^{n_t - x_t - 1} \binom{N_{t-1} - 1 + j}{j} \right) \\
 &= \frac{1}{\binom{N_t}{n_t}} \left( \binom{N_t}{n_t} - \binom{N_t - x_t - 1}{n_t - x_t - 1} \right) \tag{27}
 \end{aligned}$$

$$\begin{aligned}
 &= 1 - \frac{(N_t - x_t - 1)! / (n_t - x_t - 1)!}{N_t! / n_t!} \\
 &\geq 1 - \left( \frac{N_t - x_t - 1}{N_t} \right)^{N_{t-1}}, \tag{28}
 \end{aligned}$$

where we have used (11) at (26), (6) at (27), and  $(N_t - x_t - 1 - i) / (N_t - i) \leq (N_t - x_t - 1) / N_t$  for all  $0 \leq i \leq N_{t-1} - 1$  at (28). We therefore have

$$\begin{aligned}
 \liminf_{t \rightarrow \infty} P(X_t \leq \mathbb{E}(X_t)) &\geq 1 - \lim_{t \rightarrow \infty} \left( \frac{N_t - \mathbb{E}(X_t) - 1}{N_t} \right)^{N_{t-1}} \\
 &= 1 - \lim_{t \rightarrow \infty} \left( \frac{N_t - n_t(1 + o(1)) / N_{t-1} - 1}{N_t} \right)^{N_{t-1}} \\
 &= 1 - \lim_{t \rightarrow \infty} (1 - (1 + o(1)) / N_{t-1})^{N_{t-1}} \\
 &= 1 - e^{-1} > 1/2,
 \end{aligned}$$

contradicting (25).

## 5 Discussion

Potentially, improvements might be made on these results. The CRP we have defined places all  $n_t$  elements labelled  $t$  at the same time. Placing them sequentially seems like a difficult problem. It would be interesting to get better estimates for  $\mathbb{E}(K_t)$  and  $\text{Var}(K_t)$ , especially when  $\theta \neq 1$ . Perhaps the hypotheses of Theorem 4 could be weakened.

In the case of random permutations, there is a Poisson approximation of  $K_n$ . The total variation distance between the law  $\mathcal{L}(K_n)$  of  $K_n$  and the  $\text{Poisson}(\mathbb{E}(K_n))$  distribution is of order

$$\begin{aligned}
 d_{\text{TV}}(\mathcal{L}(K_n), \text{Poisson}(\mathbb{E}(K_n))) &:= \frac{1}{2} \sum_{k=1}^n \left| \mathbb{P}(K_n = k) - \exp(-\mathbb{E}(K_n)) \frac{\mathbb{E}(K_n)^k}{k!} \right| \\
 &\asymp \frac{1}{\log n};
 \end{aligned}$$

see [2]. By considering the process of cycle counts  $(C_1(n), C_2(n), \dots, C_n(n))$ , where  $C_i(n)$  is the number of cycles of size  $i$ , we may write  $K_n = \sum_{i=1}^n C_i(n)$ . Moreover, [7] shows that  $B_n(\cdot) \xrightarrow{d} B(\cdot)$ , where

$$B_n(t) := \frac{\sum_{i=1}^{\lfloor n^t \rfloor} C_i(n) - t \log n}{\sqrt{\log n}}, \quad 0 \leq t \leq 1,$$

and  $B(t)$  is standard Brownian motion. If suitable approximations could be made to the process of cycle counts for multiset permutations, then similar progress might be made for the results obtained in this paper when  $\theta = 1$ .

## Acknowledgments

The author thanks the anonymous referee and the Associated Editor for a careful reading of this paper and for helpful suggestions.

## References

- [1] Arratia, R., Barbour, A.D., Tavaré, S.: Logarithmic Combinatorial Structures: a Probabilistic Approach. EMS Monographs in Mathematics (2003) [MR2032426](#). <https://doi.org/10.4171/000>
- [2] Barbour, A.D., Hall, P.G.: On the rate of Poisson convergence. *Math. Proc. Camb. Phil. Soc.* **95**, 473–480 (1984) [MR0755837](#). <https://doi.org/10.1017/S0305004100061806>
- [3] Billingsley, P.: Probability and Measure. John Wiley & Sons (1995) [MR1324786](#)
- [4] Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press (1987) [MR0898871](#). <https://doi.org/10.1017/CBO9780511721434>
- [5] Björnberg, J.E., Mailler, C., Mörders, P., Ueltschi, D.: A two-table theorem for a disordered chinese restaurant process. *Ann. Appl. Probab.* **34**, 5809–5841 (2024) [MR4839639](#). <https://doi.org/10.1214/24-aap2108>
- [6] Cartier, P., Foata, D.: Problèmes combinatoires de commutation et réarrangements. In: *Lecture Notes in Math*, 85. Springer, Berlin (1969) [MR0239978](#)
- [7] DeLaurentis, J.M., Pittel, B.G.: Random permutations and Brownian motion. *Pacific J. Math.* **119**, 287–301 (1985) [MR0803120](#)
- [8] Durrett, R.: Probability: Theory and Examples. Wadsworth & Brooks/Cole (1991) [MR1068527](#)
- [9] Feng, S.: The Poisson-Dirichlet Distribution and Related Topics. Springer (2010) [MR2663265](#). <https://doi.org/10.1007/978-3-642-11194-5>
- [10] Féray, V.: Central limit theorems for patterns in multiset permutations and set partitions. *Ann. Appl. Probab.* **30**, 287–323 (2020) [MR4068312](#). <https://doi.org/10.1214/19-AAP1502>
- [11] Foata, D.: Étude algébrique de certains problèmes d’analyse combinatoire et du calcul des probabilités. *Publ.Inst. Statist. Univ Paris* **14**, 81–241 (1965) [MR0220327](#)
- [12] Galganov, G., Iliencko, A.: Scaling limit for small blocks in the chinese restaurant process. *Stoch. Process. Appl.* **192**, 104793 (2026) [MR4972882](#). <https://doi.org/10.1016/j.spa.2025.104793>
- [13] Garza, J., Wang, Y.: Limit theorems for random permutations induced by Chinese restaurant processes. <https://arxiv.org/abs/2412.02162> (2024)
- [14] Gnedin, A., Stark, D.: Random partitions and queues. *Adv. Appl. Math.* **149**, 102549 (2023) [MR4587907](#). <https://doi.org/10.1016/j.aam.2023.102549>
- [15] Graham, R.L., Knuth, D.E., Patashnik, O.: Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley (1994) [MR1397498](#)
- [16] Iksanov, A., Kabluchko, Z., Marynych, A., Wachtel, V.: Multinomial random combinatorial structures and  $r$ -versions of Stirling, Eulerian and Lah numbers. *Disc. Math.* **349**, 114865 (2026) [MR4980340](#). <https://doi.org/10.1016/j.disc.2025.114865>

- [17] Johnson, N.L., Kemp, A.W., Kotz, S.: Univariate Discrete Distributions. Wiley (2005) [MR2163227](#). <https://doi.org/10.1002/0471715816>
- [18] Kemp, C.D., Kemp, A.W.: Generalized hypergeometric distributions. *J. R. Stat. Soc. Ser. B, Methodol.* **18**, 202–211 (1956) [MR0083837](#)
- [19] Knuth, D.: *The Art of Computer Programming, Volume 3*. Addison-Wesley (1988) [MR0445948](#)
- [20] Pitman, J.: Combinatorial stochastic processes. In: *Lecture Notes in Math, 1875*. Springer (2006) [MR2245368](#)
- [21] Rohatgi, V.K., Saleh, A.K.M.: *An Introduction to Probability and Statistics*. John Wiley & Sons (2001) [MR1789794](#)
- [22] Stanley, R.: *Enumerative Combinatorics, Volume 1*. Cambridge University Press (2012) [MR2868112](#)
- [23] Stark, D.: Markov chains generating random permutations and set partitions. *Stoch. Process. Appl.*, 104483 (2024) [MR4797244](#). <https://doi.org/10.1016/j.spa.2024.104483>