



HAL
open science

Central limit theorem for the prefix exchange distance under Ewens sampling formula

Simona Grusea, Anthony Labarre

► **To cite this version:**

Simona Grusea, Anthony Labarre. Central limit theorem for the prefix exchange distance under Ewens sampling formula. *Discrete Mathematics*, 2021, 344 (2), pp.Article 112206. 10.1016/j.disc.2020.112206 . hal-02494137

HAL Id: hal-02494137

<https://hal.science/hal-02494137>

Submitted on 28 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Central limit theorem for the prefix exchange distance under Ewens sampling formula

Simona Grusea

*Institut de Mathématiques de Toulouse, INSA de Toulouse, Université de Toulouse
135 avenue de Rangueil, 31400 Toulouse, France*

Anthony Labarre

*Université Paris-Est, LIGM UMR CNRS 8049, France
Université Paris-Est Marne-la-Vallée
5 boulevard Descartes
77420 Champs-sur-Marne, France*

Abstract

The prefix exchange distance of a permutation is the length of its shortest factorisation into transpositions that all contain 1. Using a probabilistic approach, we obtain expressions for the mean and the variance, and prove the asymptotic normality of the distribution of this distance for a random permutation verifying the Ewens sampling formula. Analogous results in the uniform setting follow as simple corollaries.

Keywords: n -star graph, permutation, Ewens sampling formula, prefix exchange distance, Feller coupling, central limit theorem

1. Introduction

A *prefix exchange* is a restricted kind of transposition that swaps any element of a permutation with its first element. The *prefix exchange distance* of a permutation was introduced by Akers and Krishnamurthy (1989) in the context of interconnection network design, and corresponds to the minimum number of prefix exchanges required to transform a given permutation into

Email addresses: `Simona.Grusea@insa-toulouse.fr` (Simona Grusea),
`Anthony.Labarre@u-pem.fr` (Anthony Labarre)

the identity permutation. Many work has been devoted to the study of this distance: Portier and Vaughan (1990) first gave expressions, recurrences and generating functions for computing its distribution (with later corrections by Shen and Qiu (2008)). Alternative explicit formulas, recurrence formulas and other properties of that distance and of its distribution have since appeared in various other works Imani et al. (2009); Cheng et al. (2009b); Qiu and Akl (1995); Grusea and Labarre (2016).

In our previous work Grusea and Labarre (2016), we proved the asymptotic normality of the prefix exchange distribution for a *uniform* random permutation. In the present work, we consider a more general family of probability distributions on the set of permutations, which includes the uniform distribution as a special case. This distribution is related to the *Multivariate Ewens distribution*, also known as the Ewens sampling formula. In this setting, a given permutation with k cycles is chosen with probability proportional to θ^k , where $\theta > 0$ is a fixed parameter.

2. Background and known results

2.1. Permutations, cycles, and prefix exchanges

For $n \geq 1$, we let \mathfrak{S}_n denote the *symmetric group*, i.e. the set of all permutations of $\{1, 2, \dots, n\}$, together with the usual function composition operation \circ applied from right to left. We view permutations as sequences, i.e. $\pi = \langle \pi_1 \ \pi_2 \ \dots \ \pi_n \rangle$, where $\pi_i = \pi(i)$ for $1 \leq i \leq n$. As is well-known, every permutation decomposes in a single way into disjoint cycles (up to the ordering of cycles and of elements within each cycle). We say that a permutation is written in *canonical cycle notation* if the cycles are ordered by their smallest element, which appears first in each cycle. For instance, when $\pi = \langle 7 \ 5 \ 3 \ 2 \ 4 \ 8 \ 1 \ 6 \rangle$, its *disjoint cycle decomposition* written in canonical notation is $\pi = (1, 7)(2, 5, 4)(3)(6, 8)$. We use $c_i(\pi)$ (resp. $c_{\geq i}(\pi)$) to denote the number of cycles of length i (resp. at least i) of π for $1 \leq i \leq n$.

For every $i = 2, 3, \dots, n$, the *prefix exchange* $(1, i)$ applied to a permutation π in \mathfrak{S}_n transforms π into $\pi \circ (1, i)$ by swapping elements π_1 and π_i . The *prefix exchange distance* of π , denoted by $\text{pexc}(\pi)$, is the minimum number of prefix exchanges needed to sort π , i.e. to transform it into the identity permutation $\iota = \langle 1 \ 2 \ \dots \ n \rangle$. Akers et al. (1987) proved the following formula for computing the prefix exchange distance:

Theorem 2.1 (Akers et al. (1987)). *The prefix exchange distance of π in \mathfrak{S}_n is*

$$pexc(\pi) = n + c_{\geq 2}(\pi) - c_1(\pi) - \begin{cases} 0 & \text{if } \pi_1 = 1, \\ 2 & \text{otherwise.} \end{cases} \quad (1)$$

2.2. The joint distribution of cycle counts and the Ewens sampling formula

Let n be fixed and let $\theta > 0$ be a fixed parameter. We let π denote a random permutation in \mathfrak{S}_n chosen according to the following probabilities: any permutation with k cycles is chosen with probability $\theta^k/\theta^{\bar{n}}$, where $\theta^{\bar{n}} = \theta(\theta + 1) \cdots (\theta + n - 1)$ denotes the rising factorial. Note that the uniform distribution on \mathfrak{S}_n is obtained in the particular case $\theta = 1$. Under this probability distribution, the joint distribution of the cycle counts is given by the *Ewens sampling formula* of parameter θ : if we denote $C_i = c_i(\pi)$ the number of cycles of length i in a random permutation π following this distribution, then for $1 \leq i \leq n$ we have

$$\mathbb{P}(C_1 = a_1, \dots, C_n = a_n) = \begin{cases} \frac{n!}{\theta^{\bar{n}}} \prod_{j=1}^n \frac{1}{a_j!} \left(\frac{\theta}{j}\right)^{a_j} & \text{if } \sum_{j=1}^n j a_j = n, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Ewens (1972) introduced this sampling formula in the context of population genetics, in order to describe the probability distribution of the partition of a sample of n selectively equivalent genes into a number of different gene types (alleles). The same formula was discovered independently by Antoniak (1974) in the context of Bayesian statistics.

For $k \geq 1$ and x in \mathbb{R} , let $x^{\underline{k}} = x(x - 1) \cdots (x - k + 1)$ denote the *falling factorial*, with the convention $x^{\underline{0}} = 1$. The following result of Watterson (1974) gives the *joint factorial moments* of the cycle counts. For every $k_1, \dots, k_n \in \mathbb{N}$, denoting $m = n - \sum_{j=1}^n j k_j$, we have the following formula:

$$\mathbb{E} \left(C_1^{k_1} \cdots C_n^{k_n} \right) = \begin{cases} \frac{n!}{m!} \frac{\theta^{\bar{m}}}{\theta^{\bar{n}}} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{k_j} & \text{if } m \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

2.3. Limit distribution of cycle counts and Feller coupling

Let us now assume that $n \rightarrow \infty$ and denote $C_j^{(n)}$ the number of cycles of length j in a random permutation in \mathfrak{S}_n chosen according to the Ewens sampling formula of parameter θ . Then we have the following result.

Proposition 2.1 (Arratia et al. (1992)). *The process of cycle counts $(C_j^{(n)})_{j \geq 1}$ converges in distribution, when $n \rightarrow \infty$, to (Z_1, Z_2, \dots) , where $Z_j, j \geq 1$ are independent Poisson random variables, with $\mathbb{E}(Z_j) = \theta/j$.*

In the case $\theta = 1$, which corresponds to the uniform distribution, this convergence result is due to Goncharov (1944) and Kolchin (1971).

We next describe the following construction, due to Arratia et al. (1992) for arbitrary θ and to Feller (1945) in the case $\theta = 1$, which results in a coupling of $(C_1^{(n)}, C_2^{(n)}, \dots)$ and (Z_1, Z_2, \dots) for all n simultaneously — the so-called *Feller coupling*. This coupling will prove very useful in the sequel. Let $\xi_i, i \geq 1$ be independent Bernoulli random variables with $\mathbb{P}(\xi_j = 1) = \frac{\theta}{\theta + j - 1}$. Note that $\mathbb{P}(\xi_1 = 1) = 1$. We use the random sequence $(\xi_n, \xi_{n-1}, \dots, \xi_1)$ to construct a random permutation in \mathfrak{S}_n satisfying the Ewens sampling formula, the permutation being written in canonical cycle notation.

The construction proceeds as follows (see Arratia et al. (1992) for more details). We start with 1 in the first cycle. If $\xi_n = 1$, then we close that cycle and start the next cycle with the smallest available integer, which in this case is 2; if $\xi_n = 0$, then we choose at random one of the $n - 1$ available integers and place it to the right of 1 in the same cycle. If $\xi_{n-1} = 1$, then we close the current cycle and start the next cycle with the smallest available integer; if $\xi_{n-1} = 0$, then we choose at random one of the $n - 2$ available integers and place it to the right in the current cycle. Continuing in this manner yields a random permutation in \mathfrak{S}_n written in canonical cycle notation and satisfying the Ewens sampling formula.

Note that the number of cycles of the resulting permutation is:

$$\sum_{j=1}^n C_j^{(n)} = \sum_{j=1}^n \xi_j. \quad (4)$$

Moreover, the lengths of the cycles correspond exactly to the lengths of the spacings between consecutive ones in the sequence $(1, \xi_n, \xi_{n-1}, \dots, \xi_1)$. The number $C_j^{(n)}$ of cycles of length j is the number of *j-spacings*, that is, the number of patterns of two ones separated by $j - 1$ zeros:

$$C_j^{(n)} = \sum_{i=1}^{n-j} \xi_i (1 - \xi_{i+1}) \cdots (1 - \xi_{i+j-1}) \xi_{i+j} + \xi_{n-j+1} (1 - \xi_{n-j+2}) \cdots (1 - \xi_n). \quad (5)$$

For $j \geq 1$, let us now define $C_j^{(\infty)}$ to be the number of *j-spacings* in the infinite sequence (ξ_1, ξ_2, \dots) , that is $C_j^{(\infty)} = \sum_{i=1}^{\infty} \xi_i (1 - \xi_{i+1}) \cdots (1 - \xi_{i+j-1}) \xi_{i+j}$.

As $n \rightarrow \infty$, the sequence of cycle counts $(C_1^{(n)}, C_2^{(n)}, \dots)$ converges in distribution to $(C_1^{(\infty)}, C_2^{(\infty)}, \dots)$, where $C_j^{(\infty)}, j \geq 1$ are independent Poisson random variables, with $\mathbb{E}(C_j^{(\infty)}) = \theta/j$.

The following result (see Arratia et al. (1992)) will also prove useful:

$$\sum_{j=1}^n \mathbb{E}(|C_j^{(n)} - C_j^{(\infty)}|) \leq \frac{n\theta(\theta + 1)}{n + \theta}. \quad (6)$$

3. Mean and variance of the prefix exchange distance distribution

Let both $n \geq 1$ and $\theta > 0$ be fixed, and let π be a random permutation in \mathfrak{S}_n verifying the Ewens sampling formula. We will drop the superscripts and simply denote C_j the number of cycles of length j in the permutation π , for $j = 1, \dots, n$. Theorem 2.1 and the definition of ξ_n allow us to write:

$$pexc(\pi) = n + \sum_{i=2}^n C_i - C_1 - 2(1 - \xi_n) = n - 2 + \sum_{i=1}^n C_i - 2C_1 + 2\xi_n. \quad (7)$$

Theorem 3.1. *The expected value $\mu_n = \mathbb{E}(pexc(\pi))$ of the prefix exchange distance for a random permutation $\pi \in \mathfrak{S}_n$ verifying the Ewens sampling formula of parameter θ is:*

$$\mu_n = n - 2 + \theta \sum_{i=0}^{n-1} \frac{1}{\theta + i} - \frac{2\theta(n-1)}{\theta + n - 1}. \quad (8)$$

Proof. For every $i \geq 1$, ξ_i is Bernoulli distributed with mean $\mathbb{E}(\xi_i) = \frac{\theta}{\theta+i-1}$. Moreover, Watterson's formula in Equation (3) gives:

$$\mathbb{E}(C_1) = \frac{n\theta}{\theta + n - 1}. \quad (9)$$

The stated formula easily follows, using Equation (4) and Equation (7). \square

In the particular case $\theta = 1$, corresponding to the uniform distribution on \mathfrak{S}_n , Equation (8) yields

Corollary 3.1. *The expected value μ_n of the prefix exchange distance for a uniform random permutation in \mathfrak{S}_n equals $\mu_{n,unif} = n + H_n - 4 + \frac{2}{n}$, where $H_n = \sum_{i=1}^n \frac{1}{i}$ denotes the n -th harmonic number.*

Note that the same expression for $\mu_{n,unif}$ can be obtained as a particular case of Theorem 6.1, page 203 of Cheng et al. (2009a) by setting $k = n - 1$ in the formula they derive. In our previous paper Grusea and Labarre (2016), we gave an analytic proof of the same expression using the probability generating function, known in the case of a uniform random permutation, and first obtained by Portier and Vaughan (1990).

We now turn to proving an expression for the variance σ_n^2 of the prefix exchange distance. The following equations will be useful in that regard: taking $j = 1$ in Equation (5) gives

$$C_1 = \sum_{i=1}^{n-1} \xi_i \xi_{i+1} + \xi_n. \quad (10)$$

Equation (4) and Equation (7) yield

$$pexc(\pi) = n - 2 + \sum_{i=1}^{n-1} \xi_i (1 - 2\xi_{i+1}) + \xi_n. \quad (11)$$

Theorem 3.2. *Let $n \geq 2$ and $\theta > 0$. The variance $\sigma_n^2 = \text{Var}(pexc(\pi))$ of the prefix exchange distance for a random permutation $\pi \in \mathfrak{S}_n$ verifying the Ewens sampling formula of parameter θ , equals*

$$\begin{aligned} \sigma_n^2 &= \sum_{i=1}^n \frac{\theta}{\theta + i - 1} - \sum_{i=1}^n \frac{\theta^2}{(\theta + i - 1)^2} - \frac{4\theta^2(n-1)}{(\theta + n - 2)(\theta + n - 1)^2} \\ &\quad + 4\theta^2 \sum_{i=0}^{n-2} \frac{\theta(i+1)(\theta+i-1) - i(i+1-\theta)(\theta+i+1)}{(\theta+i-1)(\theta+i)^2(\theta+i+1)^2}. \end{aligned}$$

Proof. Let $Y_i = \xi_i(1 - 2\xi_{i+1})$, for $i = 1, \dots, n-1$. Equation (11) reads

$$pexc(\pi) = n - 2 + \sum_{i=1}^{n-1} Y_i + \xi_n.$$

Note that ξ_n is independent of Y_1, Y_2, \dots, Y_{n-2} . Moreover, Y_i and Y_j are independent for $|i - j| > 1$. Therefore:

$$\sigma_n^2 = \sum_{i=1}^{n-1} \text{Var}(Y_i) + \text{Var}(\xi_n) + 2 \sum_{i=1}^{n-2} \text{Cov}(Y_i, Y_{i+1}) + 2\text{Cov}(Y_{n-1}, \xi_n). \quad (12)$$

Using the independence of the ξ_i 's and the fact that $\mathbb{E}(\xi_i) = \frac{\theta}{\theta+i-1}$, we obtain:

$$\mathbb{E}(Y_i) = \mathbb{E}(\xi_i)(1 - 2\mathbb{E}(\xi_{i+1})) = \frac{\theta}{\theta+i-1} \left(1 - \frac{2\theta}{\theta+i}\right) = \frac{\theta(i-\theta)}{(\theta+i-1)(\theta+i)}.$$

Since $\xi_{i+1} \in \{0, 1\}$, we have $1 - 2\xi_{i+1} \in \{-1, 1\}$, and therefore

$$\mathbb{E}(Y_i^2) = \mathbb{E}(\xi_i^2)\mathbb{E}((1 - 2\xi_{i+1})^2) = \mathbb{E}(\xi_i) = \frac{\theta}{\theta+i-1}.$$

Using further the relation $(i-\theta)^2 = (\theta+i)^2 - 4\theta i$, we obtain

$$\text{Var}(Y_i) = \frac{\theta}{\theta+i-1} - \frac{\theta^2}{(\theta+i-1)^2} + \frac{4\theta^3 i}{(\theta+i-1)^2(\theta+i)^2}. \quad (13)$$

Since ξ_n is Bernoulli distributed with parameter $\frac{\theta}{\theta+n-1}$, we have:

$$\text{Var}(\xi_n) = \frac{\theta(n-1)}{(\theta+n-1)^2}. \quad (14)$$

We now compute $\text{Cov}(Y_i, Y_{i+1})$, for $i = 1, \dots, n-2$. The independence of the ξ_i 's and the relation $\xi_{i+1}(1 - 2\xi_{i+1}) = -\xi_{i+1}$ yield

$$\mathbb{E}(Y_i Y_{i+1}) = \mathbb{E}(\xi_i)\mathbb{E}[\xi_{i+1}(1-2\xi_{i+1})](1-2\mathbb{E}(\xi_{i+2})) = -\frac{\theta^2(i+1-\theta)}{(\theta+i-1)(\theta+i)(\theta+i+1)}.$$

We deduce the following expression for the covariance between Y_i and Y_{i+1} :

$$\text{Cov}(Y_i, Y_{i+1}) = -\frac{2\theta^2 i(i+1-\theta)}{(\theta+i-1)(\theta+i)^2(\theta+i+1)}. \quad (15)$$

Concerning the covariance between Y_{n-1} and ξ_n , similar arguments yield:

$$\mathbb{E}(Y_{n-1}\xi_n) = \mathbb{E}(\xi_{n-1})\mathbb{E}[(1-2\xi_n)\xi_n] = -\mathbb{E}(\xi_{n-1})\mathbb{E}(\xi_n) = -\frac{\theta^2}{(\theta+n-2)(\theta+n-1)},$$

from which we easily deduce

$$\text{Cov}(Y_{n-1}, \xi_n) = -\frac{2\theta^2(n-1)}{(\theta+n-2)(\theta+n-1)^2}. \quad (16)$$

The formula in the statement follows from Equations (12) to (16). \square

In the case of a uniform random permutation, corresponding to $\theta = 1$, Theorem 3.2 implies the following.

Corollary 3.2. *Let $n \geq 2$. The variance of the prefix exchange distance for a uniform random permutation in \mathfrak{S}_n equals*

$$\sigma_{n,unif}^2 = H_n + \frac{4}{n} - \frac{8}{n^2} - \sum_{j=1}^n \frac{1}{j^2}.$$

In Grusea and Labarre (2016) we obtained the same expression for $\sigma_{n,unif}^2$ in a more involved way, using the probability generating function of $perc(\pi)$.

4. Central limit theorem for the distance distribution

Proposition 4.1. *We have the following asymptotics for the mean and the variance of the prefix exchange distribution as $n \rightarrow \infty$:*

$$\begin{aligned} \mu_n &= n + \theta \log n + O(1), \\ \sigma_n^2 &= \theta \log n + O(1), \end{aligned}$$

where $O(1)$ denotes a bounded sequence.

Proof. Immediate from Equation (8) and Theorem 3.2, using the well-known fact that $H_n = \log(n) + O(1)$ when $n \rightarrow \infty$. \square

We further prove a central limit theorem for $perc(\pi)$. Note that in Grusea and Labarre (2016) we proved a similar result in the case $\theta = 1$.

Theorem 4.1. *We have the following convergence in distribution*

$$\frac{perc(\pi) - n - \theta \log n}{\sqrt{\theta \log n}} \rightarrow_d \mathcal{N}(0, 1) \text{ when } n \rightarrow \infty,$$

which means that the distribution of the prefix exchange distance for a random permutation $\pi \in \mathfrak{S}_n$ verifying the Ewens sampling formula of parameter θ , is asymptotically normal, with mean $n + \theta \log n$ and variance $\theta \log n$.

Proof. Let $(C_1^{(n)}, C_2^{(n)}, \dots)$ and $(C_1^{(\infty)}, C_2^{(\infty)}, \dots)$ be the Feller coupling defined in Section 2.3. Let us denote

$$D_n := \frac{perc(\pi) - n - \theta \log n}{\sqrt{\theta \log n}} = \frac{\sum_{j=2}^n C_j^{(\infty)} - \theta(H_n - 1)}{\sqrt{\theta \log n}} + R_n,$$

using Equation (7) and denoting

$$R_n := \frac{\theta(H_n - 1 - \log n) + \sum_{j=2}^n (C_j^{(n)} - C_j^{(\infty)}) - C_1^{(n)} - 2(1 - \xi_n)}{\sqrt{\theta \log n}}. \quad (17)$$

The random variable $\sum_{j=2}^n C_j^{(\infty)}$ is a sum of independent Poisson random variables with parameters $\mathbb{E}(C_j^{(\infty)}) = \frac{\theta}{j}$, therefore it is also Poisson distributed with parameter $\sum_{j=2}^n \frac{\theta}{j} = \theta(H_n - 1)$. Using the normal approximation for the Poisson distribution (see e.g. Billingsley (1999)), we obtain

$$\frac{\sum_{j=2}^n C_j^{(\infty)} - \theta(H_n - 1)}{\sqrt{\theta \log n}} \rightarrow_d \mathcal{N}(0, 1).$$

It now suffices to note that R_n converges to 0 in probability. Indeed, using Proposition 4.1, Equation (6) and the fact that $\mathbb{E}(C_1^{(\infty)}) = \theta$ and $\mathbb{E}(1 - \xi_n) = \frac{n-1}{\theta+n-1} \leq 1$, we have $\mathbb{E}(|R_n|) \rightarrow 0$ as $n \rightarrow \infty$, and hence R_n converges to 0 in probability. This ends the proof. \square

References

- Akers, S.B., Krishnamurthy, B., 1989. A group-theoretic model for symmetric interconnection networks. *IEEE Transactions on Computers* 38, 555–566.
- Akers, S.B., Krishnamurthy, B., Harel, D., 1987. The star graph: An attractive alternative to the n -cube, in: *Proceedings of the Fourth International Conference on Parallel Processing (ICPP)*, Pennsylvania State University Press. pp. 393–400.
- Antoniak, C., 1974. Mixtures of Dirichlet processes with applications to Bayesian parametric problems. *The Annals of Statistics* 2, 1152–1174.
- Arratia, R., Barbour, A., Tavaré, S., 1992. Poisson process approximations for the Ewens sampling formula. *The Annals of Applied Probability* 2, 519–535.
- Billingsley, P., 1999. *Convergence of Probability Measures*. 2nd ed., Wiley Series in Probability and Statistics.
- Cheng, E., Qiu, K., Shen, Z., 2009a. A generating function approach to the surface areas of some interconnection networks. *Journal of Interconnection Networks* 10, 189–204.

- Cheng, E., Qiu, K., Shen, Z., 2009b. A short note on the surface areas of star graphs. *Parallel Processing Letters* 19, 19–22.
- Ewens, W., 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* 3, 87–112.
- Feller, W., 1945. The fundamental limit theorems in probability. *Bulletin of the American Mathematical Society* 51, 800–832.
- Goncharov, V., 1944. Some facts from combinatorics. *Izvestia Akademii Nauk SSSR, Ser. Mat.* 8, 3–48.
- Grusea, S., Labarre, A., 2016. Asymptotic normality and combinatorial aspects of the prefix exchange distance distribution. *Advances in Applied Mathematics* 78, 94–113.
- Imani, N., Sarbazi-Azad, H., Akl, S.G., 2009. Some topological properties of star graphs: The surface area and volume. *Discrete Mathematics* 1, 560–569.
- Kolchin, V., 1971. A problem of the allocation of particles in cells and cycles of random permutations. *Theory of Probability and its Applications* 16, 74–90.
- Portier, F.J., Vaughan, T.P., 1990. Whitney numbers of the second kind for the star poset. *European Journal of Combinatorics* 11, 277–288.
- Qiu, K., Akl, S.G., 1995. On some properties of the star graph. *VLSI Design* 2, 389–396. doi:10.1155/1995/61390.
- Shen, Z., Qiu, K., 2008. On the Whitney numbers of the second kind for the star poset. *European Journal of Combinatorics* 29, 1585–1586.
- Watterson, G., 1974. The sampling theory of selectively neutral alleles. *Advances in Applied Probability* 6, 463–488.