J. Appl. Prob. **54,** 42–54 (2017)
*doi:*10.1017/jpr.2016.85

ESTIMATING THE LARGE MUTATION PARAMETER OF THE EWENS SAMPLING FORMULA

KOJI TSUKUDA,* *Kurume University*

Abstract

We derive some limit theorems associated with the Ewens sampling formula when its parameter is increasing together with a sample size. Moreover, the limit results are applied in order to investigate asymptotic properties of the maximum likelihood estimator.

Keywords: Asymptotic properties of estimators; maximum likelihood estimation; Ewens sampling formula

2010 Mathematics Subject Classification: Primary 62F12
Secondary 60F05

1. Introduction

Some data may have the form of a partition of a positive integer. For example, in population genetics, sample alleles are sometimes split in accordance with their sample frequencies of each allelic type if types of alleles are not connected. For a given positive integer n, which denotes a sample size, consider a random partition $n = C_1^n + 2C_2^n + \cdots + nC_n^n$, where $(C_1^n, \dots, C_n^n)$ is an n-dimensional nonnegative integer-valued random variable and its jth element denotes the number of alleles which appear j times. There is a well-known model of $(C_1^n, \dots, C_n^n)$ whose law is given by

$$\mathbb{P}[(C_1^n, \dots, C_n^n) = (c_1, \dots, c_n)] = \frac{n!}{(\theta)_n} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j!} \mathbf{1}\left\{\sum_{j=1}^{n} jc_j = n\right\}, \tag{1}$$

where θ is a positive constant called the mutation parameter and $(\theta)_n = \theta \times (\theta+1) \times \cdots \times (\theta + n - 1)$. This law is the celebrated Ewens sampling formula [1]; see also [7]. This distribution was derived from various probabilistic models and Hoppe's derivation using the Pólya-like urn; see, for example, [4] for a biological context. Consider the number $K_n = \sum_{j=1}^{n} C_j^n$ of alleles of an allelic partition whose law is given by (1). The probability mass function of K_n is

$$\mathbb{P}[K_n = k] = \bar{s}(n,k)\frac{\theta^k}{(\theta)_n},$$

where $\bar{s}(n,k)$ denotes the coefficient of θ^k in $(\theta)_n$. Then, it holds that

$$\mathbb{E}(K_n) = \theta L_n(\theta), \qquad \operatorname{var}(K_n) = \theta \ell_n(\theta),$$

where

$$L_n(\theta) = \sum_{j=1}^{n} \frac{1}{\theta + j - 1}, \qquad \ell_n(\theta) = \sum_{j=1}^{n} \frac{j-1}{(\theta + j - 1)^2}.$$

Received 2 July 2015; revision received 12 May 2016.
* Current address: Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, Japan. Email address: k.tsukuda@gmail.com

In this paper we present results which describe the asymptotic behavior of K_n and their statistical applications when θ tends to ∞ as n increases.

Consider the neutral infinite allele model of evolutions. Let N be the effective population size and u the mutation rate per individual. Then the mutation parameter θ is given by $4Nu$, so large θ corresponds to large N with fixed u. Since large N is of interest when selection exists and since asymptotic properties of K_n when θ tends to ∞ are different from asymptotic properties with fixed θ, various large θ settings have been considered; see Feng [2] and the references therein. Many works including Griffiths [3], Joyce *et al.* [5], and Feng [2] assumed that θ tends to ∞ with fixed n, and Feng [2] also argued the case where θ tends to ∞ as n increases. In Feng [2], the large deviation principles for the Poisson–Dirichlet distribution and the Ewens sampling formula were established and, as a corollary of the large deviation result, the weak law of large numbers for the Ewens sampling formula with the large mutation parameter were provided [2, Corollary 4.1]. The large deviation result is of interest in itself, but it is not necessary just for the purpose of proving the weak law of large numbers. In this paper we provide a direct proof of the weak law of large numbers (Proposition 2). Moreover, we present an extension of the weak law of large numbers when $n/\theta \to 0$ (Theorem 1) and the central limit theorem when $n^2/\theta \to \infty$ (Theorem 2). These results will be used to obtain asymptotic properties of estimators for θ. Furthermore, we derive another limit result when $n^2/\theta \to c$ (Theorem 3).

It is clear that the random variable K_n is a sufficient statistic for θ. To estimate θ, the maximum likelihood estimator, which is defined as the root of

$$K_n = \sum_{j=1}^{n} \frac{\hat{\theta}}{\hat{\theta} + j - 1} = \hat{\theta} L_n(\hat{\theta}), \tag{2}$$

is frequently used. It coincides with the moment estimator. On the other hand, there is a simple consistent estimator $K_n/\log n$ which is based on the law of large numbers; see [1] and [7]. When K_n is by far larger than $\log n$, there is a serious difference between these two estimators. In such a case, $K_n/\log n$ does not work well because the assumption for the consistency is considered violated, whereas the maximum likelihood estimator still works well. The main motivations of this paper are to show this phenomenon and to derive asymptotic properties of the maximum likelihood estimator by assuming that θ is an increasing function of n (Propositions 3–6 and Remark 9).

Let us explain some notation used in this paper. Consider functions $f(x)$ and $g(x)$ of x. If $|f(x)|$ is asymptotically bounded above by $|g(x)|$ up to a constant factor, we denote $f(x) = O(g(x))$. Moreover, if $|f(x)|$ is asymptotically bounded both below and above by a function $|g(x)|$ up to a constant factor, we write $f(x) = \Theta(g(x))$. If $f(x)/g(x) \to 1$ then we write $f(x) \sim g(x)$. We denote by '$\xrightarrow{\mathbb{P}}$' and '$\xrightarrow{\mathrm{D}}$' the convergence in probability and the convergence in distribution, respectively.

The paper is organized as follows. Section 2 is devoted to preliminaries. Section 3 includes limit theorems and Section 4 includes their statistical applications.

2. Preliminaries

2.1. Asymptotic settings

We consider the asymptotic situation $n \to \infty$, while $\theta \to c \neq 0$ or $\theta \to \infty$ as $n \to \infty$. Hereafter, this situation is denoted by $n \to \infty$, $\theta \not\to 0$. Moreover, we divide relations between $\theta$and n as follows:

(A) $n/\theta \to \infty$;

(B) $n/\theta \to c > 0$;

(C) $n/\theta \to 0$;

(C1) $n/\theta \to 0$ and $n^2/\theta \to \infty$;

(C2) $n^2/\theta \to c > 0$;

(C3) $n^2/\theta \to 0$.

Remark 1. Cases A, B, and C above correspond to cases D, C, and B of Feng's division in [2], respectively.

2.2. On evaluations of $L_n(\theta)$ and $\ell_n(\theta)$

To evaluate the mean and the variance of K_n, the next proposition will be used in the following sections.

Proposition 1. *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$.*

(i) *It holds that*

$$L_n(\theta) = \log\left(1 + \frac{n}{\theta}\right) + \Theta\left(\frac{n}{\theta(n+\theta)}\right). \tag{3}$$

(ii) *It holds that*

$$0 \le L_n(\theta) - \ell_n(\theta) - \frac{n}{n+\theta} \le \frac{1}{\theta}. \tag{4}$$

In particular, it holds that

$$\ell_n(\theta) = \log\left(1 + \frac{n}{\theta}\right) + O\left(\frac{n}{n+\theta}\right). \tag{5}$$

(iii) *It holds that*

$$\ell_n(\theta) = O\left(\frac{n^2}{\theta^2}\right).$$

In particular, in case C, it holds that

$$\ell_n(\theta) \sim \frac{n^2}{2\theta^2}. \tag{6}$$

Proof. (i) It holds that

$$L_n(\theta) - \log\left(1 + \frac{n}{\theta}\right) = \sum_{j=1}^{n}\left(\frac{1}{\theta+j-1} - \int_{\theta+j-1}^{\theta+j}\frac{\mathrm{d}x}{x}\right)$$

and that

$$\frac{1}{2}\left(\frac{1}{\theta+j-1} - \frac{1}{\theta+j}\right) < \frac{1}{\theta+j-1} - \int_{\theta+j-1}^{\theta+j}\frac{\mathrm{d}x}{x} < \frac{1}{\theta+j-1} - \frac{1}{\theta+j}$$

due to the convexity of $1/x$. These formulae and

$$\sum_{j=1}^{n}\left(\frac{1}{\theta+j-1} - \frac{1}{\theta+j}\right) = \frac{n}{\theta(\theta+n)}$$

imply the conclusion.

https://doi.org/10.1017/jpr.2016.85 Published online by Cambridge University Press

(ii) The former part follows from

$$\frac{n}{\theta(n+\theta)} = \int_{\theta}^{\theta+n} \frac{\mathrm{d}x}{x^2} \le \sum_{j=1}^{n} \frac{1}{(\theta+j-1)^2} \le \frac{1}{\theta^2} + \int_{\theta}^{\theta+n-1} \frac{\mathrm{d}x}{x^2} = \frac{1}{\theta^2} + \frac{n-1}{\theta(n-1+\theta)}$$

and

$$\ell_n(\theta) = \sum_{j=1}^{n} \frac{\theta+j-1}{(\theta+j-1)^2} - \sum_{j=1}^{n} \frac{\theta}{(\theta+j-1)^2} = L_n(\theta) - \sum_{j=1}^{n} \frac{\theta}{(\theta+j-1)^2}.$$

The latter part immediately follows from the former part and the result of (i).

(iii) The former part follows from

$$\ell_n(\theta) < \frac{1}{\theta^2} \sum_{j=1}^{n} (j-1) = \frac{n(n-1)}{2\theta^2}.$$

Moreover, the latter part follows from

$$\ell_n(\theta) > \frac{1}{(\theta+n)^2} \sum_{j=1}^{n} (j-1) = \frac{n(n-1)}{2\theta^2} \frac{1}{(1+n/\theta)^2} \sim \frac{n(n-1)}{2\theta^2} \quad \text{when } n/\theta \to 0.$$

This completes the proof. □

Remark 2. From (3), in case C, the Taylor expansion yields

$$L_n(\theta) = \frac{n}{\theta} - \frac{n^2}{2\theta^2} + \Theta\left(\frac{n^3}{\theta^3} \vee \frac{n}{\theta^2}\right),$$

where the notation $\vee$ in $\Theta(\cdot)$ is based on the magnitude relationship with large enough n. In case A, a similar argument yields

$$L_n(\theta) = \log\left(\frac{n}{\theta}\right) + \Theta\left(\frac{\theta}{n} \vee \frac{1}{\theta}\right).$$

3. Limit theorems

3.1. The weak law of large numbers and its extension

First, we provide a direct proof of the weak law of large numbers for the Ewens sampling formula when $\theta \to \infty$ as $n \to \infty$.

Proposition 2. (The weak law of large numbers.) *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. Then, it holds that*

$$\mathbb{E}\left[\left(\frac{K_n - \theta L_n(\theta)}{\theta L_n(\theta)}\right)^2\right] \to 0. \tag{7}$$

Proof. The left-hand side of (7) is equal to

$$\frac{\theta \ell_n(\theta)}{(\theta L_n(\theta))^2} = \frac{1}{\theta L_n(\theta)} \frac{\ell_n(\theta)}{L_n(\theta)}.$$

In cases A or B, (3) and (5) yield $\ell_n(\theta)/L_n(\theta) = O(1)$. So (7) follows from $\theta L_n(\theta) \to \infty$. In case C, (3) and (6) yield $L_n(\theta) \sim n/\theta$ and $\ell_n(\theta) = \Theta(n^2/\theta^2)$. So $\ell_n(\theta)/L_n(\theta) \to 0$ holds. Since (3) yields $\theta L_n(\theta) \sim n$, (7) holds. □

https://doi.org/10.1017/jpr.2016.85 Published online by Cambridge University Press

Remark 3. If the largest term of $L_n(\theta)$ is considered, this result corresponds to Corollary 4.1, cases B, C, and D of Feng [2]. The slight difference between the above result and Corollary 4.1 of [2] is that the convergence in (7) is in mean square.

Remark 4. In case C, the largest term of $\theta L_n(\theta)$ is n, which does not depend on θ. As a result, Proposition 2 is not sufficient in order to construct an estimator of θ.

In case C, consider an approximation $A_n^r(\theta)/\theta$ $(r = 0, 1, 2, \ldots)$ of $\log(1 + n/\theta)$ by polynomials of n/θ up to the order r, that is, $A_n^r(\theta) = 0$ for $r = 0$ and

$$A_n^r(\theta) = \sum_{k=1}^{r} \frac{(-1)^{k-1}\theta}{k}\left(\frac{n}{\theta}\right)^k \quad \text{for } r \geq 1.$$

It implies that

$$\begin{aligned}\theta L_n(\theta) &= \theta \log\left(1 + \frac{n}{\theta}\right) + \theta\left(L_n(\theta) - \log\left(1 + \frac{n}{\theta}\right)\right)\\ &= n - \frac{n^2}{2\theta} + \frac{n^3}{3\theta^2} - \cdots + \theta\left(L_n(\theta) - \log\left(1 + \frac{n}{\theta}\right)\right)\\ &= A_n^r(\theta) + R_n^{1,r}(\theta) + \theta\left(L_n(\theta) - \log\left(1 + \frac{n}{\theta}\right)\right),\end{aligned}$$

where $R_n^{1,r}(\theta)/\theta$ is the corresponding remainder term. Then, Proposition 1 yields

$$R_n^{1,r}(\theta) + \theta\left(L_n(\theta) - \log\left(1 + \frac{n}{\theta}\right)\right) = \Theta\left(\frac{n^{r+1}}{\theta^r} \vee \frac{n}{\theta}\right). \tag{8}$$

For example, it holds that

$$\begin{aligned}\theta L_n(\theta) &\overset{(r=0)}{=} \Theta(n),\\ &\overset{(r=1)}{=} n + \Theta\left(\frac{n^2}{\theta}\right),\\ &\overset{(r=2)}{=} n - \frac{n^2}{2\theta} + \Theta\left(\frac{n^3}{\theta^2} \vee \frac{n}{\theta}\right),\\ &\overset{(r=3)}{=} n - \frac{n^2}{2\theta} + \frac{n^3}{3\theta^2} + \Theta\left(\frac{n^4}{\theta^3} \vee \frac{n}{\theta}\right),\\ &= \cdots .\end{aligned}$$

Hereafter, we denote the left-hand side of (8) by $R_n^r(\theta)$. With this approximation, we extend the previous law of large numbers in case C.

Theorem 1. (An extension of the weak law of large numbers.) *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. Fix a nonnegative integer $r = 0, 1, 2, \ldots$. In case C,*

$$\frac{\theta^{2r-1}}{n^{2r}} \to 0 \tag{9}$$

is equivalent to

$$\mathbb{E}\left[\left(\frac{K_n - \theta L_n(\theta)}{R_n^r(\theta)}\right)^2\right] \to 0. \tag{10}$$

https://doi.org/10.1017/jpr.2016.85 Published online by Cambridge University Press

Proof. It holds that

$$\mathbb{E}\bigg[\bigg(\frac{K_n - \theta L_n(\theta)}{R_n^r(\theta)}\bigg)^2\bigg] = \frac{\theta \ell_n(\theta)}{(R_n^r(\theta))^2}.$$

The numerator is $\Theta(n^2/\theta)$ by (6). Firstly, we see that (9) $\Longrightarrow$ (10). Since

$$\frac{n^{r+1}}{\theta^r} > \frac{n}{\theta}$$

is equivalent to

$$\frac{\theta^{2r-1}}{n^{2r}} \frac{1}{\theta} < 1,$$

(8) and (9) imply that $R_n^r(\theta) = \Theta(n^{r+1}/\theta^r)$. Hence, the direct half follows. Next, we see that (9) $\Longleftarrow$ (10). If $R_n^r(\theta) = \Theta(n/\theta)$ holds then (10) should not hold, so (10) implies that $R_n^r(\theta) = \Theta(n^{r+1}/\theta^r)$ by contradiction. This yields

$$\frac{n^2/\theta}{n^{2(r+1)}/\theta^{2r}} = \frac{\theta^{2r-1}}{n^{2r}} \to 0,$$

which is (9). Hence, the converse half follows. □

Remark 5. Consider the case where θ is represented as $\theta = nf(n)$. If $f(n)$ is a slowly varying function of n then $n/\theta \to 0$ and (9) is satisfied for any r. If $f(n)$ is a power function of n then we have to consider the relationship of the power and r: setting $f(n) = n^{\beta}$ with $\beta > 0$, (9) is equivalent to $(2r-1)\beta < 1$. If θ is increasing faster than the power functions then (9) is not satisfied for any positive r.

In case C, Proposition 2 corresponds to the case of $r = 0$. Moreover, the following corollary follows from Theorem 1 with $r = 1$.

Corollary 1. *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. In case C1, it holds that*

$$\mathbb{E}\bigg[\bigg(\frac{K_n - n}{n^2/(2\theta)} + 1\bigg)^2\bigg] \to 0.$$

Proof. The conclusion follows from

$$\mathbb{E}\bigg[\bigg(\frac{K_n - \theta L_n(\theta)}{R_n^1(\theta)}\bigg)^2\bigg] = \mathbb{E}\bigg[\bigg(\frac{K_n - n}{R_n^1(\theta)} - 1\bigg)^2\bigg] \to 0$$

by Theorem 1 and $R_n^1(\theta) \sim -n^2/(2\theta)$. □

3.2. Limit distributions

For fixed θ, Watterson [7, Corollary 1 to Theorem 5] proved the central limit theorem for the Ewens sampling formula

$$\frac{K_n - \theta \log n}{\sqrt{\theta \log n}} \xrightarrow{\text{D}} N(0, 1), \tag{11}$$

where $N(0, 1)$ denotes a standard normal random variable. As an extension of this central limit theorem, we present the asymptotic normality of K_n when $n, \theta \to \infty$, $n^2/\theta \to \infty$. As we will see later, when $n^2/\theta \to c$, the following central limit theorem does not hold.

Theorem 2. (The central limit theorem.) *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. If $n^2/\theta \to \infty$ then it holds that*

$$Z_n = \frac{K_n - \mu}{\sigma} \xrightarrow{\mathrm{D}} N(0, 1), \tag{12}$$

where

$$\mu = \mu_n = \theta \log\left(1 + \frac{n}{\theta}\right), \qquad \sigma = \sigma_n = \sqrt{\theta\left(\log\left(1 + \frac{n}{\theta}\right) + \frac{\theta}{n+\theta} - 1\right)}.$$

Proof. When $\theta \to c \neq 0$, it is essentially the same as (11). Therefore, we will prove only the $\theta \to \infty$ case.

The moment generating function of Z_n is given by

$$\mathbb{E}(\mathrm{e}^{Z_n t}) = \mathrm{e}^{-(\mu/\sigma)t}\mathbb{E}(\mathrm{e}^{K_n t/\sigma}) = \mathrm{e}^{-(\mu/\sigma)t}\frac{(\theta \mathrm{e}^{t/\sigma})_n}{(\theta)_n} = \mathrm{e}^{-(\mu/\sigma)t}\frac{\Gamma(\theta \mathrm{e}^{t/\sigma} + n)}{\Gamma(\theta + n)}\frac{\Gamma(\theta)}{\Gamma(\theta \mathrm{e}^{t/\sigma})}$$

due to Watterson [7, Equation (2.12)]. Using Stirling's formula, this is asymptotically equal to

$$\begin{aligned}
&\mathrm{e}^{-(\mu/\sigma)t}\left(\frac{\theta \mathrm{e}^{t/\sigma} + n - 1}{\theta + n - 1}\right)^{1/2}\left(\frac{\theta \mathrm{e}^{t/\sigma} + n - 1}{\mathrm{e}}\right)^{\theta \mathrm{e}^{t/\sigma}+n-1}\left(\frac{\mathrm{e}}{\theta + n - 1}\right)^{\theta+n-1} \\
&\quad \times \left(\frac{\theta - 1}{\theta \mathrm{e}^{t/\sigma} - 1}\right)^{1/2}\left(\frac{\theta - 1}{\mathrm{e}}\right)^{\theta-1}\left(\frac{\mathrm{e}}{\theta \mathrm{e}^{t/\sigma} - 1}\right)^{\theta \mathrm{e}^{t/\sigma}-1} \\
&\quad \sim \mathrm{e}^{-(\mu/\sigma)t}(\theta \mathrm{e}^{t/\sigma} + n)^{\theta \mathrm{e}^{t/\sigma}+n-1}(\theta + n)^{-(\theta+n-1)}\theta^{\theta-1}(\theta \mathrm{e}^{t/\sigma})^{-(\theta \mathrm{e}^{t/\sigma}-1)} \\
&\quad = \mathrm{e}^{-(\mu/\sigma)t}\left(\frac{n}{\theta}\right)^{\theta(\mathrm{e}^{t/\sigma}-1)}\frac{(1 + \theta \mathrm{e}^{t/\sigma}/n)^{\theta \mathrm{e}^{t/\sigma}+n-1}}{(1 + \theta/n)^{\theta+n-1}}\mathrm{e}^{-t\theta \mathrm{e}^{t/\sigma}/\sigma + t/\sigma} \\
&\quad = \mathrm{e}^{-(\mu/\sigma)t}\left(\frac{n}{\theta}\right)^{\theta(\mathrm{e}^{t/\sigma}-1)}\left(1 + \frac{\theta}{n}\right)^{\theta(\mathrm{e}^{t/\sigma}-1)}\left(\frac{1 + \theta \mathrm{e}^{t/\sigma}/n}{1 + \theta/n}\right)^{\theta \mathrm{e}^{t/\sigma}+n-1}\mathrm{e}^{-t\theta \mathrm{e}^{t/\sigma}/\sigma + t/\sigma}.
\end{aligned} \tag{13}$$

So, it holds that

$$\begin{aligned}
\log \mathbb{E}(\mathrm{e}^{Z_n t}) &\sim -\frac{\mu}{\sigma}t + \theta(\mathrm{e}^{t/\sigma} - 1)\log\left(\frac{n}{\theta}\right) + \theta(\mathrm{e}^{t/\sigma} - 1)\log\left(1 + \frac{\theta}{n}\right) \\
&\qquad + (\theta \mathrm{e}^{t/\sigma} + n - 1)\log\left(1 + \frac{\theta/n(\mathrm{e}^{t/\sigma} - 1)}{1 + \theta/n}\right) - \frac{t}{\sigma}\theta \mathrm{e}^{t/\sigma} + \frac{t}{\sigma} \\
&= -\frac{\mu}{\sigma}t + \theta(\mathrm{e}^{t/\sigma} - 1)\log\left(\frac{n}{\theta} + 1\right) \\
&\qquad + (\theta \mathrm{e}^{t/\sigma} + n)\log\left(1 + \frac{\theta(\mathrm{e}^{t/\sigma} - 1)}{n + \theta}\right) - \frac{t}{\sigma}\theta \mathrm{e}^{t/\sigma} + o(1) \\
&= -\left(\sigma - \frac{\theta}{\sigma}\left(\frac{\theta}{n+\theta} - 1\right)\right)t + (\mathrm{e}^{t/\sigma} - 1)\left(\sigma^2 - \theta\left(\frac{\theta}{n+\theta} - 1\right)\right) \\
&\qquad + (\theta \mathrm{e}^{t/\sigma} + n)\log\left(1 + \frac{\theta(\mathrm{e}^{t/\sigma} - 1)}{n + \theta}\right) - \frac{t}{\sigma}\theta \mathrm{e}^{t/\sigma} + o(1)
\end{aligned}$$

$$
\begin{aligned}
&= -\sigma t + (\mathrm{e}^{t/\sigma} - 1)\sigma^2 + \theta\left(\frac{\theta}{n+\theta} - 1\right)\left(\frac{t}{\sigma} - (\mathrm{e}^{t/\sigma} - 1)\right) \\
&\quad + (\theta \mathrm{e}^{t/\sigma} + n)\log\left(1 + \frac{\theta(\mathrm{e}^{t/\sigma} - 1)}{n+\theta}\right) - \frac{t}{\sigma}\theta \mathrm{e}^{t/\sigma} + o(1). \qquad (14)
\end{aligned}
$$

The Taylor expansion yields

$$
\begin{aligned}
&(\theta \mathrm{e}^{t/\sigma} + n)\log\left(1 + \frac{\theta(\mathrm{e}^{t/\sigma} - 1)}{n+\theta}\right) \\
&= (\theta \mathrm{e}^{t/\sigma} + n)\frac{\theta}{n+\theta}(\mathrm{e}^{t/\sigma} - 1) - \frac{(\theta \mathrm{e}^{t/\sigma} + n)}{2}\left(\frac{\theta}{n+\theta}\right)^2 (\mathrm{e}^{t/\sigma} - 1)^2 + o(1) \\
&= \theta(\mathrm{e}^{t/\sigma} - 1) + \frac{\theta^2}{n+\theta}(\mathrm{e}^{t/\sigma} - 1)^2 - \frac{1}{2}\frac{\theta^2}{n+\theta}\left(1 + \frac{\theta(\mathrm{e}^{t/\sigma} - 1)}{n+\theta}\right)(\mathrm{e}^{t/\sigma} - 1)^2 + o(1) \\
&= \theta(\mathrm{e}^{t/\sigma} - 1) + \frac{1}{2}\frac{\theta^2}{n+\theta}(\mathrm{e}^{t/\sigma} - 1)^2 + o(1) \qquad (15)
\end{aligned}
$$

since $\mathrm{e}^{t/\sigma} - 1 \to 0$ and $\theta/\sigma^3 \to 0$ as $n \to \infty$ holds. Therefore, if

$$
\theta\left(\frac{\theta}{n+\theta} - 1\right)\left(\frac{t}{\sigma} - (\mathrm{e}^{t/\sigma} - 1)\right) + \theta(\mathrm{e}^{t/\sigma} - 1) + \frac{\theta^2}{2(n+\theta)}(\mathrm{e}^{t/\sigma} - 1)^2 - \frac{t}{\sigma}\theta \mathrm{e}^{t/\sigma} \qquad (16)
$$

converges to 0, then it holds that

$$
\log \mathbb{E}(\mathrm{e}^{Z_n t}) \sim -\sigma t + (\mathrm{e}^{t/\sigma} - 1)\sigma^2 = \tfrac{1}{2}t^2 + o(1). \qquad (17)
$$

The first term of (16) is

$$
-\frac{\theta}{2}\left(\frac{\theta}{n+\theta} - 1\right)\frac{t^2}{\sigma^2} + o(1)
$$

and the second, third, and fourth terms of (16) are

$$
\theta\left(\mathrm{e}^{t/\sigma} - 1 - \frac{t}{\sigma}\right) + \frac{\theta^2}{2(n+\theta)}(\mathrm{e}^{t/\sigma} - 1)^2 - \frac{t}{\sigma}\theta(\mathrm{e}^{t/\sigma} - 1) = \frac{1}{2}\left(\frac{\theta}{n+\theta} - 1\right)\frac{\theta t^2}{\sigma^2} + o(1).
$$

This completes the proof. □

Remark 6. It holds that

$$
\frac{K_n - \theta L_n(\theta)}{\sqrt{\theta \ell_n(\theta)}} - Z_n = \left(Z_n + \frac{\mu - \theta L_n(\theta)}{\sigma}\right)\left(\frac{\sigma}{\sqrt{\theta \ell_n(\theta)}} - 1\right) + \frac{\mu - \theta L_n(\theta)}{\sigma}.
$$

Consider the case when $n^2/\theta \to \infty$ holds. Proposition 1 yields $(\mu - \theta L_n)/\sigma \to 0$, and the proof of Proposition 1(i) and (4) yield $(\theta \ell_n(\theta) - n/(n+\theta))^{1/2} < \sigma < (\theta(\ell_n(\theta) + 1/\theta))^{1/2}$. Moreover, $Z_n = O_p(1)$ follows from Theorem 2. Hence, it holds that

$$
\left| Z_n - \frac{K_n - \theta L_n(\theta)}{\sqrt{\theta \ell_n(\theta)}} \right| \xrightarrow{\mathbb{P}} 0.
$$

Corollary 2. *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. In case A, it holds that*

$$\frac{K_n - \mu}{\sqrt{\mu}} \xrightarrow{\text{D}} N(0, 1).$$

Proof. Since it holds that

$$\frac{K_n - \mu}{\sqrt{\mu}} = \frac{\sigma}{\sqrt{\mu}} Z_n,$$

it is enough to show that $\sigma^2/\mu \to 1$, which follows from $n/\theta \to \infty$. □

Remark 7. In particular, when considering the power-form $\theta = \alpha n^{\beta}$, if $0 \le \beta \le \frac{2}{3}$ then it holds that

$$\frac{K_n - \theta \log(n/\theta)}{\sqrt{\theta \log(n/\theta)}} \xrightarrow{\text{D}} N(0, 1), \tag{18}$$

since it holds that $\log(1 + n/\theta) = \log(n/\theta) + \log(1 + \theta/n)$ and that $\theta^{1/2} \log(1 + \theta/n) = O(1)$ by using the Taylor expansion. Let $\psi(\cdot)$ be the digamma function and γ the Euler constant. Since

$$\psi(x) = -\gamma - \frac{1}{x} + \sum_{j=1}^{\infty} \frac{x}{j(j + x)}$$

holds, if θ is a large positive integer then $\log\theta$ is approximately equal to $\psi(\theta)$. In this case, the left-hand side of (18) is approximately equal to $(K_n - \theta(\log n - \psi(\theta)))/(\theta(\log n - \psi(\theta)))^{1/2}$. Yamato [8] showed that the approximation accuracy of this statistic is better than (11) when θ is a constant, which corresponds to $\beta = 0$.

In the proof of Theorem 2, the condition $n^2/\theta \to \infty$ yields $\sigma \to \infty$ which was used in (14), (15), and (17). When this condition is violated (cases C and C3), Z_n has a different limit. In case C2, Z_n converges in distribution to a standardized Poisson distribution and in case C3, Z_n converges in probability to 0.

Theorem 3. *Consider the asymptotic setting $n \to \infty$, $\theta \not\to 0$. Let Z_n be the left-hand side of (12). In case C2,*

$$Z_n \xrightarrow{\text{D}} -\sqrt{\frac{2}{c}}\left(\mathbb{P} - \frac{c}{2}\right) \tag{19}$$

holds, where $\mathbb{P}$ follows a Poisson distribution with $\mathbb{E}[\mathbb{P}] = c/2$ and $c = \lim_{n\to\infty} n^2/\theta$. In case C3, $Z_n \xrightarrow{\mathbb{P}} 0$ holds.

Proof. In case C2, from (13), it follows that

$$\begin{aligned}
\mathbb{E}(\mathrm{e}^{Z_n t}) &\sim \mathrm{e}^{-(\mu/\sigma)t}\left(\frac{\theta\mathrm{e}^{t/\sigma} + n}{\theta + n}\right)^{n}\left(\frac{\theta}{\theta + n}\right)^{\theta - 1}\left(\frac{\theta\mathrm{e}^{t/\sigma} + n}{\theta\mathrm{e}^{t/\sigma}}\right)^{\theta\mathrm{e}^{t/\sigma} - 1} \\
&= \mathrm{e}^{((n-\mu)/\sigma)t}\left(1 + \frac{n(\mathrm{e}^{-t/\sigma} - 1)}{\theta + n}\right)^{n}\left(1 + \frac{n}{\theta}\right)^{-(\theta-1)}\left(1 + \frac{n}{\theta\mathrm{e}^{t/\sigma}}\right)^{\theta\mathrm{e}^{t/\sigma} - 1} \\
&\sim \mathrm{e}^{(n^2/(2\theta))^{1/2} t} \exp\left(\frac{n^2}{\theta + n}(\mathrm{e}^{-t/\sigma} - 1)\right)\left(\mathrm{e}\left(1 - \frac{n}{2\theta}\right)\right)^{-n}\left(\mathrm{e}\left(1 - \frac{n}{2\theta\mathrm{e}^{t/\sigma}}\right)\right)^{n},
\end{aligned}$$

where we use $(1 + 1/x)^x = \mathrm{e}(1 - 1/(2x) + \Theta(1/x^2))$ as $x \to \infty$ for the last approximation.

The right-hand side converges to

$$e^{(c/2)^{1/2}t}\exp(c(e^{-(2/c)^{1/2}t}-1))\exp\big(\tfrac{1}{2}c(1-e^{-(2/c)^{1/2}t})\big)= e^{(c/2)^{1/2}t}\exp\big(\tfrac{1}{2}c(e^{-(2/c)^{1/2}t}-1)\big).$$

The right-hand side of the above formula is the moment generating function of the right-hand side of (19). In case C3, it similarly holds that $\mathbb{E}(e^{Z_n t}) \sim e^{(n^2/(2\theta))^{1/2}t}\cdot 1\cdot e^{-n}\cdot e^{n} \to 1$. $\square$

Remark 8. In case C3, Z_n converges in probability, but does not converge in mean square. Actually, $\mathrm{var}(Z_n) \to 1$ holds. That is, it holds that

$$\mathrm{var}(Z_n) = \frac{\mathrm{var}(K_n)}{\sigma^2} = \frac{\ell_n(\theta)}{\log(1+n/\theta)+\theta/(n+\theta)-1} \sim 1,$$

since both of the numerator and the denominator are asymptotically equal to $n^2/(2\theta)$.

4. Asymptotic properties of the maximum likelihood estimator

4.1. Consistency

In this section we argue asymptotic properties of the maximum likelihood estimator $\hat{\theta}$, which is the root of (2). First we see the consistency.

The (weak) consistency is a basic property which estimators should satisfy. However, we discuss the situation that the true value θ_0 of θ depends on n and becomes ∞ as $n\to\infty$, so the consistency is not defined in the usual sense. Thus, we will see that $\hat{\theta}/\theta_0 \xrightarrow{\mathbb{P}} 1$. Henceforth, when $\hat{\theta}/\theta_0 \xrightarrow{\mathbb{P}} 1$ holds, we consider that the consistency holds.

We can construct an intuitive consistent estimator based on Proposition 2 and Corollary 1 when $n^2/\theta \to \infty$. However, showing the consistency of Z-estimators such as the maximum likelihood estimators takes a little ingenuity, where Z-estimators are estimators which are defined as roots of some estimating equations. When a one-dimensional parameter does not depend on a sample size, we can directly use the following lemma by Van der Vaart [6] in order to show the consistency of Z-estimators.

Lemma 1. (Van der Vaart [6, Lemma 5.10].) *Let Θ be a subset of the real line and let Ψ_n be a random functions and Ψ a fixed function of θ such that $\Psi_n(\theta) \xrightarrow{\mathbb{P}} \Psi(\theta)$ for every θ. Assume that each map $\theta \mapsto \Psi_n(\theta)$ is continuous and has exactly one zero $\hat{\theta}_n$, or is nondecreasing with $\Psi_n(\hat{\theta}_n) = o_{\mathbb{P}}(1)$. Let θ_0 be a point such that $\Psi(\theta_0-\varepsilon) < 0 < \Psi(\theta_0+\varepsilon)$ for every $\varepsilon > 0$. Then $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.*

This lemma combined with the results in Section 3 implies the consistency of the maximum likelihood estimator.

Proposition 3. *Consider the asymptotic setting $n\to\infty$, $\theta_0 \not\to 0$, and $n/\theta_0 \to \infty$. Then, it holds that $\hat{\theta}/\theta_0 \xrightarrow{\mathbb{P}} 1$.*

Proof. For a variable $x > 0$, define the random function $\Psi_n(x;\theta_0)$ by

$$\begin{aligned}\Psi_n(x;\theta_0) &= \frac{xL_n(\theta_0 x)}{L_n(\theta_0)} - \frac{K_n}{\theta_0 L_n(\theta_0)}\\ &= x\bigg(1+\frac{1-x}{L_n(\theta_0)}\sum_{j=1}^{n}\frac{\theta_0}{(\theta_0 x+j-1)(\theta_0+j-1)}\bigg) - \frac{K_n}{\theta_0 L_n(\theta_0)}. \end{aligned} \tag{20}$$

https://doi.org/10.1017/jpr.2016.85 Published online by Cambridge University Press

As for the first term of the right-hand side of (20), it holds that

$$\begin{aligned}\sum_{j=1}^{n}\frac{\theta_0}{(\theta_0 x+j-1)(\theta_0+j-1)} &= \sum_{j=1}^{n}\frac{\theta_0}{(\theta_0+j-1)^2}\bigg(1+\frac{\theta_0(1-x)}{x\theta_0+j-1}\bigg)\\ &\le \sum_{j=1}^{n}\frac{\theta_0}{(\theta_0+j-1)^2}\bigg(1+\frac{|1-x|}{x}\bigg)\\ &= O\bigg(\frac{n}{n+\theta_0}\bigg).\end{aligned}$$

Moreover, the second term converges to 1 in probability due to (7). Hence, it holds that

$$\Psi_n(x;\theta_0)\xrightarrow{\mathbb{P}}\Psi(x;\theta_0)=x-1 \qquad \text{for every } x>0.$$

The function $x\mapsto\Psi_n(x;\theta_0)$ is continuous, so it follows from the definition of the maximum likelihood estimator that $\Psi_n(x;\theta_0)$ becomes (unique) 0 at $x=\hat{x}_n=\hat{\theta}/\theta_0$. Moreover, for every $\varepsilon>0$, $\Psi(1-\varepsilon;\theta_0)<0<\Psi(1+\varepsilon;\theta_0)$ holds. Hence, Lemma 1 yields the conclusion $\hat{x}_n\xrightarrow{\mathbb{P}}1$. This completes the proof. □

Proposition 4. *Consider the asymptotic setting $n\to\infty$, $\theta_0\not\to 0$, and $n/\theta_0\to 1/c>0$. Then, it holds that $\hat{\theta}/\theta_0\xrightarrow{\mathbb{P}}1$.*

Proof. As in the proof of Proposition 3, for a variable $x>0$, define the random function $\Psi_n(x;\theta_0)$ by

$$\Psi_n(x;\theta_0)=\frac{xL_n(\theta_0 x)}{L_n(\theta_0)}-\frac{K_n}{\theta_0 L_n(\theta_0)},$$

which converges to

$$\Psi(x;\theta_0)=\frac{cx\log(1+1/(cx))}{c\log(1+1/c)}-1$$

in probability for every $x>0$. For every $\varepsilon>0$, $\Psi(1-\varepsilon;\theta_0)<0<\Psi(1+\varepsilon;\theta_0)$ holds. This is because the function $f(x)=x\log(1+1/x)$ of $x>0$ is increasing since its derivative $\log(1+1/x)-1/(1+x)$ is positive and because $f(c)=c\log(1+1/c)$ holds. The remainder of the proof follows that of Proposition 3. □

Proposition 5. *Consider the asymptotic setting $n\to\infty$, $n/\theta_0\to 0$, and $n^2/\theta_0\to\infty$. Then, it holds that $\hat{\theta}/\theta_0\xrightarrow{\mathbb{P}}1$.*

Proof. For a variable $x>0$, define the random function $\Psi_n(x;\theta_0)$ by

$$\Psi_n(x;\theta_0)=\frac{\theta_0 xL_n(\theta_0 x)-n}{n^2/2\theta_0}-\frac{K_n-n}{n^2/(2\theta_0)}. \tag{21}$$

The first term of (21) converges to $-1/x$ since it holds that

$$\theta_0 xL_n(\theta_0 x)-n=-\frac{n^2}{2\theta_0 x}+O\bigg(\frac{n(n^2+\theta_0)}{\theta_0^2}\bigg),$$

and the second term converges to -1 in probability due to Corollary 1. Hence, it holds that

$$\Psi_n(x;\theta_0)\xrightarrow{\mathbb{P}}\Psi(x;\theta_0)=-\frac{1}{x}+1 \qquad \text{for every } x>0.$$

The remainder of the proof follows that of Proposition 3. □

Remark 9. From Proposition 2, when $n \to \infty$, $\theta_0 \not\to 0$, and $K_n/\log n$ is consistent if and only if θ_0 satisfies $\log\theta_0/\log n \to 0$. That is, when $n/\theta_0 \to \infty$,

$$\frac{K_n/\log n}{\theta_0} = \frac{K_n}{\theta_0\log(1+n/\theta_0)}\frac{\log(1+n/\theta_0)}{\log n} = \frac{K_n}{\theta_0\log(1+n/\theta_0)}\bigg(1-\frac{\log\theta_0}{\log n}+\frac{\log(1+\theta_0/n)}{\log n}\bigg)$$

holds and otherwise $(K_n/\log n)/\theta_0 \le n/(\theta_0\log n) \to 0$ almost surely holds. On the other hand, Propositions 3, 4, and 5 yield that the maximum likelihood estimator is consistent if $n^2/\theta_0 \to \infty$.

4.2. Asymptotic normality

We now show the asymptotic normality of the maximum likelihood estimator $\hat\theta$.

Proposition 6. *Consider the asymptotic setting $n \to \infty$, $\theta_0 \not\to 0$, and $n^2/\theta_0 \to \infty$. Then, it holds that*

$$\bigg(\frac{\ell_n(\theta_0)}{\theta_0}\bigg)^{1/2}(\hat\theta-\theta_0) \xrightarrow{\mathrm{D}} N(0,1).$$

Proof. The Taylor expansion yields that (2) is equal to

$$K_n = \sum_{j=1}^{n}\bigg(\frac{\theta_0}{\theta_0+j-1}+\frac{j-1}{(\tilde\theta+j-1)^2}(\hat\theta-\theta_0)\bigg)$$

if and only if

$$\bigg(\frac{\ell_n(\theta_0)}{\theta_0}\bigg)^{1/2}(\hat\theta-\theta_0) = \frac{K_n-\theta_0L_n(\theta_0)}{\sqrt{\theta_0\ell_n(\theta_0)}}\frac{\ell_n(\theta_0)}{\sum_{j=1}^{n}(j-1)/(\tilde\theta+j-1)^2}, \tag{22}$$

where $\tilde\theta$ is a value between θ_0 and $\hat\theta$. As for the denominator of the second factor in the right-hand side of (22), it holds that

$$\sum_{j=1}^{n}\frac{j-1}{(\tilde\theta+j-1)^2} = \ell_n(\theta_0)+\sum_{j=1}^{n}\bigg(\frac{j-1}{(\tilde\theta+j-1)^2}-\frac{j-1}{(\theta_0+j-1)^2}\bigg)$$

and the absolute value of this second term is bounded above by

$$\bigg|\sum_{j=1}^{n}\frac{j-1}{(\tilde\theta+j-1)(\theta_0+j-1)}\bigg(\frac{1}{\tilde\theta+j-1}+\frac{1}{\theta_0+j-1}\bigg)(\tilde\theta-\theta_0)\bigg|. \tag{23}$$

When $n/\theta_0 \to \infty$ or $n/\theta_0 \to c \ne 0$, (23) is bounded above by

$$\sum_{j=1}^{n}\frac{2}{(\underline{\theta}+j-1)^2}|\hat\theta-\theta_0| \le 2\bigg(\frac{\theta_0}{\underline{\theta}^2}+\frac{\theta_0(n-1)}{\underline{\theta}(n-1+\underline{\theta})}\bigg)\bigg|\frac{\hat\theta}{\theta_0}-1\bigg|, \tag{24}$$

where $\underline{\theta}$ denotes $\min(\hat\theta,\theta_0)$. When $n/\theta_0 \to 0$, (23) is bounded above by

$$\sum_{j=1}^{n}\frac{2(j-1)}{\underline{\theta}^3}|\hat\theta-\theta_0| \le \frac{n^2\theta_0}{\underline{\theta}^3}\bigg|\frac{\hat\theta}{\theta_0}-1\bigg|. \tag{25}$$

Owing to $\hat{\theta}/\theta_0 \xrightarrow{\mathbb{P}} 1$ (Propositions 3, 4, and 5), the right-hand sides of (24) and (25) converge to 0 in probability under each setting. Consequently, (12) and Slutsky's lemma [6, Lemma 2.8] yield the conclusion. □

Remark 10. The second-order derivative of the logarithm of the likelihood function is

$$\ddot{l}(\theta) := -\frac{K_n}{\theta^2} + \sum_{j=1}^{n} \frac{1}{(\theta + j - 1)^2},$$

so $-\mathbb{E}[\ddot{l}(\theta_0)] = \ell_n(\theta_0)/\theta_0$ holds. It coincides with the inverse of the asymptotic variance of the maximum likelihood estimator.

Acknowledgements

The author wishes to express his sincere gratitude to Professor Yoichi Nishiyama, Professor Takashi Yanagawa, the editor, and the anonymous referee for their helpful comments and suggestions which helped to improve the paper.

Part of the revision was carried out after the author had become a member of Graduate School of Arts and Sciences, The University of Tokyo.

References

[1] Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3,** 87–112.
[2] Feng, S. (2007). Large deviations associated with Poisson–Dirichlet distribution and Ewens sampling formula. *Ann. Appl. Prob.* **17,** 1570–1595.
[3] Griffiths, R. C. (1979). On the distribution of allele frequencies in a diffusion model. *Theoret. Pop. Biol.* **15,** 140–158.
[4] Hoppe, F. M. (1984). Pólya-like urns and the Ewens' sampling formula. *J. Math. Biol.* **20,** 91–94.
[5] Joyce, P., Krone, S. M. and Kurtz, T. G. (2002). Gaussian limits associated with the Poisson–Dirichlet distribution and the Ewens sampling formula. *Ann. Appl. Prob.* **12,** 101–124.
[6] Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
[7] Watterson, G. A. (1974). The sampling theory of selectively neutral alleles. *Adv. Appl. Prob.* **6,** 463–488.
[8] Yamato, H. (2013). Edgeworth expansions for the number of distinct components associated with the Ewens sampling formula. *J. Japan Statist. Soc.* **43,** 17–28.