Refined Approximations for the Ewens Sampling Formula

A. D. Barbour*
Institut für Angewandte Mathematik, Universität Zürich, Rämistr. 74, CH-8001 Zurich, Switzerland

ABSTRACT

The Ewens sampling formula is a family of probability distributions over the space of cycle types of permutations of n objects, indexed by a real parameter θ. In the case $\theta = 1$, where the distribution reduces to that induced by the uniform distribution on all permutations, the joint distributions of the numbers of cycles of lengths less than $b = o(n)$ is extremely well approximated by a product of Poisson distributions, having mean $1/j$ for cycle length j: the error is super-exponentially small with nb^{-1}. For $\theta \neq 1$, the analogous approximation, with means adjusted to θ/j, is good, but with error only linear in $n^{-1}b$. In this article, it is shown that, by choosing the means of the Poisson distributions more carefully, an error quadratic in $n^{-1}b$ can be achieved, and that essentially nothing better is possible.

1. INTRODUCTION

The Ewens sampling formula is the probability distribution P_n on the set $S = \{r \in (\mathbb{Z}^+)^n; \Sigma_{j=1}^n jr_j = n\}$ defined by

$$P_n(r_1, \ldots, r_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{r_j} \frac{1}{r_j!}\,, \tag{1.1}$$

where $\theta > 0$ is a free parameter. The distribution has turned out to have many useful applications outside its original context of mathematical genetics; in the special case $\theta = 1$, it reduces to the distribution of the cycle type of a (uniform)

* This research was supported in part by Schweiz. Nationalfonds Projekt Nr 21-25579.88 and by NSF grant DMS90-05833.

Random Structures and Algorithms, Vol. 3, No. 3 (1992)

random permutation. Expression (1.1) is for many purposes inconvenient, and so it is useful to have good approximations to it for large n. One such is given in Arratia, Barbour, and Tavaré [2], and concerns the distribution of the random vector $C^{(nb)} = (C_1^{(n)}, \ldots, C_b^{(n)})$, where $(C_1^{(n)}, \ldots, C_n^{(n)})$ is a random element of S with distribution given by (1.1). It is shown there that, for any $1 \le b \le n$,

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)})) \le c_1 n^{-1} b \,, \tag{1.2}$$

for some constant $c_1 = c_1(\theta)$, where

$$\lambda_j^{(\infty b)} = \lambda_j^{(\infty)} = \theta/j \,, \qquad 1 \le j \le b \,,$$

and $\mathrm{Po}(\lambda^{(\infty b)})$ denotes the product of the b one-dimensional Poisson distributions $\mathrm{Po}(\lambda_j^{(\infty b)})$.

Although the bound is very simple and attractive, it is known to be far from sharp when $\theta = 1$; Arratia and Tavaré [1] show that the true order of approximation is super-exponentially small in $b^{-1}n$. It is thus of interest to know whether (1.2) is similarly inaccurate when $\theta \ne 1$. Now it is observed in [2] that $\lambda^{(nb)} = \mathbb{E}C^{(nb)}$, given by Watterson's [4] formula

$$\lambda_j^{(nb)} = \frac{\theta n(n-1)\cdots(n-j+1)}{j(\theta+n-1)\cdots(\theta+n-j)} \,,$$

differs from $\lambda^{(\infty b)}$ in each component by an amount of strict order n^{-1} when $\theta \ne 1$, as a consequence of which

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)})) \ge c_1' n^{-1} b / \log(nb^{-1}) \,, \tag{1.3}$$

for some $c_1' = c_1'(\theta)$. Thus no great improvement can be hoped for in this approximation. However, there remains the possibility that, if $\mathrm{Po}(\lambda^{(nb)})$ were used as an approximation instead of $\mathrm{Po}(\lambda^{(\infty b)})$, better bounds might be achieved.

In this article, it is shown that a Poisson approximation with mean $\lambda^{(nb)}$ is indeed an order of magnitude better, in that

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(nb)})) \le c_2 n^{-2} b^2 \,, \tag{1.4}$$

for some $c_2 = c_2(\theta)$. The bound is still not as spectacularly small as in the case $\theta = 1$, but it is shown that it is essentially correct, inasmuch as

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(nb)})) \ge c_2' n^{-2} b^2 / \log^2(nb^{-1})$$

for some $c_2' = c_2'(\theta)$, because of the discrepancies between the means and variances of the components of $C^{(nb)}$. As a result of (1.4), it is also possible to show that there is a $c_3 = c_3(\theta)$ such that

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)})) \ge c_3 n^{-1} b \,, \tag{1.5}$$

improving upon the lower bound (1.3) and completing the estimates of [2, Section 5].

I am indebted to Charles Stein for much helpful discussion of this work, and in particular for reminding me of the Neyman–Pearson lemma.

2. DETAILS

The argument is based on the following construction used in [2], which realizes random vectors U and V on the same probability space in such a way that $U \overset{\mathscr{D}}{=} \text{Po}(\lambda^{(\infty b)})$, $V \overset{\mathscr{D}}{=} C^{(nb)}$ and U and V are close. Let $(I_i, i \geq 1)$ be independent indicator random variables with $\mathbb{P}[I_i = 1] = \theta/(i + \theta - 1)$, and let U_j count the runs of $(j - 1)$ consecutive zeros in the sequence $(I_1, I_2, \ldots)$: then, as observed in [2],

$$U = (U_1, \ldots, U_b) \overset{\mathscr{D}}{=} \text{Po}(\lambda^{(\infty b)}) .$$

On the other hand, if I_{n+1} is forced to be one, and all other I_i, $i \geq n + 2$, are set equal to zero, the resulting b-vector V of runs of zeros has the same distribution as $C^{(nb)}$. Clearly, U and V have much in common, and only differ because of what happens to the I_i for large i, when there are only rarely short runs of zeros. The argument here involves identifying the main contribution to the difference between U and V.

In order to prove the bound (1.4), three approximation lemmas are combined. For the first, suppose that U is a random vector in $(\mathbb{Z}^+)^d$ with distribution $\text{Po}(\lambda)$, and that it is expressible as

$$U = W + Z + Y , \tag{2.1}$$

where W and Y are independent and $\mathbb{E}Y = \mu \leq \lambda$: set

$$\eta_1 = d_{TV}(\mathscr{L}(Y), \text{Po}(\mu)) ; \qquad \eta_2 = \mathbb{E}\|Z\|_1 , \tag{2.2}$$

where, here and subsequently, $\|z\|_1$ is used to denote $\Sigma_{j=1}^d |z_j|$.

Lemma 1. *Under the above circumstances, if* $\|\mu\|_1 \leq 1/10$, *then*

$$d_{TV}(\mathscr{L}(W), \text{Po}(\lambda - \mu)) \leq 2(\eta_1 + \eta_2) . \tag{2.3}$$

Proof. Let $h = h_{\lambda-\mu,A} : (\mathbb{Z}^+)^d \to \mathbb{R}$ denote the solution of the equation

$$\begin{aligned}(\mathscr{A}_{\lambda-\mu}h)(u) &= \sum_{j=1}^{d} \{(\lambda_j - \mu_j)[h(u + e_j) - h(u)] + u_j[h(u - e_j) - h(u)]\} \\ &= I[u \in A] - \text{Po}(\lambda - \mu)\{A\} , \end{aligned} \tag{2.4}$$

where e_j denotes the jth coordinate vector. Then, from Barbour, Holst, and Janson (1991) [3, Lemma X.1.3],

$$\max_{1 \leq j \leq d} \sup_u |h(u + e_j) - h(u)| \leq 1 . \tag{2.5}$$

Now, since $U \sim \mathrm{Po}(\lambda)$,

$$0 = \mathbb{E} \sum_{j=1}^{d} \{\lambda_j\{h(U + e_j) - h(U)\} + U_j\{h(U - e_j) - h(U)\}\}$$
$$= \mathbb{P}[U \in A] - \mathrm{Po}(\lambda - \mu)\{A\} + \sum_{j=1}^{d} \mu_j \mathbb{E}[h(U + e_j) - h(U)]\,,$$

by (2.4). Realize $Y' \sim \mathrm{Po}(\mu)$ in such a way that $\mathbb{P}[Y \neq Y'] = \eta_1$ and that Y' is independent of W: this can be done, on a possibly enlarged probability space, because Y is independent of W. Then it thus follows that

$$0 = \mathbb{P}[W + Y' \in A] - \mathrm{Po}(\lambda - \mu)\{A\} + \sum_{j=1}^{d} \mu_j \mathbb{E}[h(W + Y' + e_j) - h(W + Y')] + \epsilon_1\,, \tag{2.6}$$

where, from (2.2) and (2.5),

$$|\epsilon_1| \leq \{\mathbb{P}[Z \neq 0] + \mathbb{P}[Y \neq Y']\}(1 + 2\|\mu\|_1) \leq (1 + 2\|\mu\|_1)(\eta_1 + \eta_2)\,. \tag{2.7}$$

Thus, again using (2.4),

$$0 = \mathbb{E} \sum_{j=1}^{d} \{\lambda_j\{h(W + Y' + e_j) - h(W + Y')\} + (W_j + Y'_j)\{h(W + Y' - e_j) - h(W + Y')\}\} + \epsilon_1$$
$$= \mathbb{E} \sum_{j=1}^{d} \{(\lambda_j - \mu_j)\{h(W + Y' + e_j) - h(W + Y')\} + W_j\{h(W + Y' - e_j) - h(W + Y')\}\} + \epsilon_1\,,$$

by the independence of Y' and W, and because $Y' \sim \mathrm{Po}(\mu)$. Hence it follows that

$$0 = \mathbb{P}[Y' = 0]\mathbb{E}\{(\mathscr{A}_{\lambda-\mu} h)(W)\} + \sum_{y>0} \mathbb{P}[Y' = y]\mathbb{E}\{(\mathscr{A}_{\lambda-\mu} h^y)(W)\} + \epsilon_1\,, \tag{2.8}$$

where $h^y(w)$ denotes $h(w + y)$.

In order to complete the proof from (2.8), observe, once more from (2.4), that

$$(\mathscr{A}_{\lambda-\mu} h^y)(u) = I[u + y \in A] - \mathrm{Po}(\lambda - \mu)\{A\} + \sum_{j=1}^{d} y_j\{h(u + y - e_j) - h(u + y)\}\,, \tag{2.9}$$

and so, writing $\delta = d_{TV}(\mathscr{L}(W), \mathrm{Po}(\lambda - \mu))$ and using (2.5), that

$$|\mathbb{E}\{(\mathscr{A}_{\lambda-\mu} h^y)(W)\}| \leq (1 + 2\|y\|_1)\delta\,, \tag{2.10}$$

since $\mathrm{Po}(\lambda-\mu)\{\mathscr{A}_{\lambda-\mu}h\}=0$. Combining (2.10) with (2.8) thus yields the inequality

$$|\mathbb{E}\{(\mathscr{A}_{\lambda-\mu}h_{\lambda-\mu,A})(W)\}|\le(3\delta\mathbb{E}\|Y'\|_1+\epsilon_1)/\mathbb{P}[Y'=0]\,, \tag{2.11}$$

and taking the supremum of the left-hand side over A then gives

$$\delta\le(3\delta\|\mu\|_1+\epsilon_1)e^{\|\mu\|_1}\,,$$

in view of (2.4). For $\|\mu\|_1\le 1/10$, (2.3) now follows by direct calculation. ∎

The setting of the second lemma is somewhat similar. Suppose that $V=W+Z+Y$, where W and Y are independent, with

$$\begin{aligned}&\mathbb{P}[Y=e_j]=\nu_j\,,\quad 1\le j\le d\,; &&\mathbb{P}[Y=0]=1-\|\nu\|_1\,;\\ &d_{TV}(\mathscr{L}(W),\mathrm{Po}(\psi))=\eta_1\,; &&\mathbb{E}\|Z\|_1=\eta_2\,.\end{aligned} \tag{2.12}$$

Lemma 2. *Under the above circumstances,*

$$d_{TV}(\mathscr{L}(V),\mathrm{Po}(\psi+\nu))\le\eta_1+\eta_2+\|\nu\|_1^2\,.$$

Proof. Let $W'\sim\mathrm{Po}(\psi)$ be realized independently of Y, in such a way that $\mathbb{P}[W\ne W']=\eta_1$. Then it follows easily that

$$d_{TV}(\mathscr{L}(V),\mathscr{L}(W'+Y))\le\eta_1+\eta_2\,.$$

On the other hand, for instance by way of Theorem X.H of [3],

$$d_{TV}(\mathscr{L}(W'+Y),\mathrm{Po}(\psi))\le\|\nu\|_1^2\,.$$

The third lemma uses more of the structure of the problem under consideration. Let

$$J_{ij}=I_iI_{i+j}\prod_{l=1}^{j-1}(1-I_{i+l})\,,\quad i\ge 1\,,\qquad 1\le j\le b\,,$$

where the indicators I_i are independent, and

$$p_i=\mathbb{P}[I_i=1]=\theta/(i+\theta-1)$$

for some fixed $\theta>0$. Thus J_{ij} is the indicator of a run of exactly $(j-1)$ zeros, following from a one at index i. Define $Y=(Y_1,\ldots,Y_b)$ by setting

$$Y_j=\sum_{i=n+1-b}^{\infty}J_{ij}\,,$$

and write

$$\pi_{ij}=\mathbb{E}J_{ij}\le p_ip_{i+j}\,;\qquad \mu_j=\mathbb{E}Y_j\le\sum_{i=n+1-b}^{\infty}p_ip_{i+j}\,.$$

Lemma 3. *With these definitions, and if $b \le n/3$,*

$$d_{TV}(\mathscr{L}(Y), \mathrm{Po}(\mu)) \le cn^{-2}b^2$$

for some $c = c(\theta)$.

Proof. Let the indicators $(J_{kl}, k \ge n+1-b, 1 \le l \le b)$ be constructed as above from the $\{I_i, i \ge n+1-b\}$. For each (i, j), construct indicators $J'_{kl} = J^{(i,j)}_{kl}$ in exactly the same way on the basis of the indicators $\{I'_r, r \ge n+1-b\}$, where

$$I'_i = I'_{i+j} = 1\,; \quad I'_{i+l} = 0, 1 \le l \le j-1\,; \quad I'_r = I_r \text{ otherwise:}$$

then the J'_{kl} have the joint distribution of the J_{kl}, conditional on $J_{ij} = 1$. Furthermore, $0 = J'_{kl} \le J_{kl}$ for all $(k, l) \ne (i, j)$ such that $i - l < k < i + j$ and $1 \le l \le b$; $J'_{kl} \ge J_{kl}$ if $k + l = i$ or if $k = i + j$; and otherwise $J'_{kl} = J_{kl}$. Hence, by [3, Corollary X.J.1],

$$\begin{aligned} &d_{TV}(\mathscr{L}(Y), \mathrm{Po}(\mu)) \\ &\le \sum_{i=n+1-b}^{\infty} \sum_{j=1}^{b} \pi_{ij}\left\{\pi_{ij} + \sum_{l=1}^{b}\left(\sum_{k=i-l+1}^{i+j-1} \pi_{kl} + p_{i-l} + p_{i+j+l}\right)\right\} \\ &\le c(\theta)n^{-2}b^2\,, \end{aligned}$$

uniformly in $b \le n/3$, since $p_i \le \theta/(i-1)$. ∎

Returning to the setting of Section 1, all is prepared for a proof of the main result. The symbol c is used henceforth to denote a generic constant, depending only on θ, and not necessarily the same at each appearance.

Theorem 1. *There exists a constant $c_2 = c_2(\theta)$ such that, for all $b \le n$,*

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(nb)})) \le c_2 n^{-2}b^2\,.$$

Proof. It is obviously enough to prove the assertion for $b \le n/3$. Define the indicators I_i and J_{ij} as for Lemma 3, and write, for each $1 \le j \le b$,

$$U_j = \sum_{i \ge 1} J_{ij}\,; \quad Y_j = \sum_{i=n+1-b}^{\infty} J_{ij}\,; \quad W_j = \sum_{i=1}^{n-2b} J_{ij}\,;$$

$$Z_j = \sum_{i=n-2b+1}^{n-b} J_{ij}\,; \quad Y'_j = I_{n-j+1} \prod_{s=n-j+2}^{n} (1 - I_s)\,; \quad Z'_j = \sum_{i=n-2b+1}^{n-j} J_{ij}\,.$$

The reason for these definitions is that, from the construction of [2, Section 2], it follows that

$$W + Z + Y = U \overset{\mathscr{D}}{=} C^{(\infty b)} \sim \mathrm{Po}(\lambda^{(\infty b)}) \tag{2.13}$$

and that

$$V = W + Z' + Y' \overset{\mathscr{D}}{=} C^{(nb)}\,, \tag{2.14}$$

conveniently associating random variables with the distributions $\mathscr{L}(C^{(nb)})$ and $\mathrm{Po}(\lambda^{(\infty b)})$. The remainder of the proof consists of combining Lemmas 1–3 with (2.13) and (2.14) to obtain an approximation for $\mathscr{L}(C^{(nb)})$ by $\mathrm{Po}(\lambda^{(nb)})$. The idea is that Z and Z' are negligible, W contributes almost the whole of both U and V, and adding Y or Y' to W is like adding an independent Poisson-distributed random vector with the appropriate mean.

To apply Lemma 1, observe that W and Y are independent, and that, from Lemma 3, $d_{TV}(\mathscr{L}(Y), \mathrm{Po}(\mu)) \le cn^{-2}b^2$. The conditions of Lemma 1 are thus satisfied, with $\|\mu\|_1 = \|\mathbb{E}Y\|_1 \le cn^{-1}b$ and with $\eta_1 + \eta_2 \le cn^{-2}b^2$. Hence it follows that

$$d_{TV}(\mathscr{L}(W), \mathrm{Po}(\lambda^{(\infty b)} - \mu)) \le cn^{-2}b^2 . \tag{2.15}$$

On the other hand, W and Y' are independent,

$$\mathbb{P}[Y' = e_j] = p_{n-j+1} \prod_{s=n-j+2}^{n} (1 - p_s) = \nu_j \le p_{n-j+1} , \qquad 1 \le j \le b ,$$

and $\mathbb{P}[Y' = 0] = 1 - \|\nu\|_1$, and $\mathbb{E}\|Z'\|_1 \le cn^{-2}b^2$. Hence, from Lemma 2 and (2.14),

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)} - \mu + \nu)) \le cn^{-2}b^2 ,$$

with $\mu = \mathbb{E}Y$ and $\nu = \mathbb{E}Y'$. It also follows from (2.13) and (2.14) that

$$\|(\lambda^{(\infty b)} - \lambda^{(nb)}) - (\mu - \nu)\|_1 \le cn^{-2}b^2 ,$$

and the theorem follows. ∎

As a consequence of Theorem 1, it is possible to show that the error in total variation in approximating $\mathscr{L}(C^{(nb)})$ by $\mathrm{Po}(\lambda^{(\infty b)})$ is of strict order $n^{-1}b$. An upper bound has already been referred to in (1.2). A lower bound, stated in (1.5), follows from Theorem 1 together with the following comparison of multivariate Poisson distributions.

Suppose that $\lambda = \{\lambda_i, 1 \le i \le K\}$ and $\lambda^{\epsilon} = \{\lambda_i(1 + \epsilon_i), 1 \le i \le K\}$, where $0 \le \epsilon_i \le 1$ for all i. Define the quantities

$$\sigma^2 = \sum_{i=1}^{K} \lambda_i \epsilon_i^2 ; \quad \Lambda = \sum_{i=1}^{K} \lambda_i ; \quad \eta = \sigma^{-3} \sum_{i=1}^{K} \lambda_i^{3/2} \epsilon_i^3 .$$

Theorem 2. *With the above notation,*

$$\Delta^{\epsilon} = d_{TV}(\mathrm{Po}(\lambda), \mathrm{Po}(\lambda^{\epsilon})) \ge \max\{\psi_1, \psi_2, \psi_3\} ,$$

where

$$\psi_1 = c\{(1 \wedge \sigma) - k\eta\} ; \qquad \psi_2 = \sup_{\delta > 0} \psi_2(\delta) ;$$

$$\psi_2(\delta) = c\{(1 \wedge \Lambda^{-1/2}\delta) - k\eta\}\Big(1 \wedge \sum_{i \in I(\delta)} \lambda_i \epsilon_i\Big) ,$$

where $I(\delta) = \{i\colon \epsilon_i > \Lambda^{-1/2}\delta\sigma\}$;

$$\psi_3 = c(1 \wedge \Lambda^{-1/2})\left(1 \wedge \sum_{i=1}^{K} \lambda_i \epsilon_i\right),$$

and the constants k and $c > 0$ do not depend on K, δ, λ, or ε.

Remark. Each of the three estimates is, loosely speaking, of order $\Lambda^{1/2}\epsilon^*$, for some value ϵ^* typical of the ϵ_i's, if δ is taken to be 1/2, say, in the second. However, the first estimate is vacuous if $\sigma \le k\eta$, and the second becomes much weaker if $\Lambda^{1/2}\eta$ is big. In our application, neither of these restrictions can be ignored, and the crude order estimates turn out to be misleading: in particular, ψ_3 is only useful in a very restricted range of parameters.

Proof. Let $X \sim \mathrm{Po}(\lambda)$ and $Z \sim \mathrm{Po}(\lambda^\epsilon - \lambda)$ be independent, and write $Y = X + Z$, so that $Y \sim \mathrm{Po}(\lambda^\epsilon)$. Consideration of the likelihood ratio suggests comparing the distributions of X and Y by looking at

$$U = \sum_{i=1}^{K} \epsilon_i(X_i - \lambda_i) \text{ and } V = \sum_{i=1}^{K} \epsilon_i(Y_i - \lambda_i).$$

By the Berry–Esséen theorem,

$$|\mathbb{P}[U > 0] - 1/2| \le C\eta,$$

where C is the Berry–Esséen constant, and

$$\left|\mathbb{P}[V > 0] - \mathbb{P}\left[\mathcal{N}\left(0, \sum_{i=1}^{K} \lambda_i\epsilon_i^2(1+\epsilon_i)\right) > -\sigma^2\right]\right| \le 2^{3/2}C\eta.$$

It follows immediately that $\Delta^\epsilon \ge \psi_1$ for suitable choices of c and k.

For the second estimate, observe that, for any $\delta > 0$,

$$\mathbb{P}[V > 0] \ge \mathbb{P}[U > 0] + \mathbb{P}[\Lambda^{-1/2}\delta\sigma < U \le 0]\mathbb{P}\left[\bigcup_{i \in I(\delta)} \{Z_i \ge 1\}\right].$$

Now

$$\mathbb{P}\left[\bigcup_{i \in I(\delta)} \{Z_i \ge 1\}\right] = 1 - \exp\left\{-\sum_{i \in I(\delta)} \lambda_i\epsilon_i\right\} \ge \frac{1}{2}\left(1 \wedge \sum_{i \in I(\delta)} \lambda_i\epsilon_i\right),$$

and, by the Berry–Esséen theorem,

$$\mathbb{P}[\Lambda^{-1/2}\delta\sigma < U \le 0] \le \left(\frac{1}{2} \wedge \Lambda^{-1/2}\delta\right)e^{-1/8}(2\pi)^{-1/2} - 2C\eta,$$

which implies that $\Delta^\epsilon \ge \psi_2(\delta)$ for all $\delta > 0$, for suitably chosen c and k not depending on δ. To prove that $\Delta^\epsilon \ge \psi_3$, take $U' = \Sigma_{i=1}^K X_i$ and $V' = \Sigma_{i=1}^K Y_i$, and use the inequality

$$\mathbb{P}[V' > \Lambda] \geq \mathbb{P}[U' > \Lambda] + \mathbb{P}[U' = [\Lambda]]\mathbb{P}\left[\sum_{i=1}^{K} Z_i \geq 1\right]$$

in a similar fashion. ■

Corollary 2.1. *With notation as in Section 1, if $\theta \neq 1$ is fixed,*

$$d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)})) \geq c_3 n^{-1} b ,$$

for some $c_3 = c_3(\theta) > 0$.

Proof. We prove only the case $\theta > 1$: the case $\theta < 1$ can be treated similarly. Take $K = b$, $\lambda_i = \theta / i$ and

$$\epsilon_i = \left\{1 - \frac{n(n-1)\cdots(n-i+1)}{(n+\theta-1)\cdots(n+\theta-i)}\right\} = \frac{i(\theta-1)}{n} + O(n^{-2}i^2) .$$

It then follows that

$$\sigma \sim n^{-1} b \sqrt{\tfrac{1}{2}\theta(\theta-1)^2} ; \qquad \Lambda \sim \log b ;$$

$$\eta \asymp b^{-1/2} ; \qquad \sum_{i=1}^{b} \lambda_i \epsilon_i \sim n^{-1} b \theta(\theta - 1) ,$$

and, taking $\delta = \frac{1}{2} n^{-1} b \sigma^{-1} \Lambda^{1/2}$,

$$\sum_{i \in I(\delta)} \lambda_i \epsilon_i \sim \frac{1}{2} n^{-1} b \theta(\theta-1) .$$

Hence

$$\psi_2 \geq c\{(1 \wedge [2\theta(\theta-1)^2]^{-1/2}) - k' b^{-1/2}\} \cdot \tfrac{1}{2} n^{-1} b \theta(\theta-1) ,$$

which is of strict order $n^{-1}b$ for all $b \geq b_0(\theta)$. On the other hand,

$$\psi_3 \sim c(\log b)^{-1/2} \cdot \tfrac{1}{2} n^{-1} b \theta(\theta-1)$$

is of strict order $n^{-1}b$ for all $b \leq b_0(\theta)$, proving the corollary. Note that, in this example, $\psi_1 \asymp n^{-1} b$ only if $b \geq k' n^{2/3}$, for some constant $k' = k'(\theta)$. ■

The lower bound (1.5) for $d_{TV}(\mathscr{L}(C^{(nb)}), \mathrm{Po}(\lambda^{(\infty b)}))$ follows immediately from Theorem 1 and Corollary 2.1.

The conclusion to be drawn from the arguments of this section is that the error in approximating $\mathscr{L}(C^{(nb)})$ by $\mathrm{Po}(\lambda^{(\infty b)})$ is largely due to the difference in their means, and that correcting the mean of the approximating Poisson distribution to $\lambda^{(nb)}$ improves the error bound by an order of magnitude. To show that little more can in fact be achieved, consider the random variable $T = \Sigma_{j=m}^{b} C_j^{(nb)}$, where m is chosen to be $[b/e] + 1$. Elementary calculation shows that

$$\mathbb{E}T = \sum_{j=m}^{b} \lambda_j^{(nb)} \sim \theta$$

is of strict order 1, and that

$$\operatorname{Var} T - \mathbb{E}T = n^{-2}(b - m + 1)^2\theta^2(1 - \theta) + O(n^{-3}b^3)$$
$$\sim n^{-2}b^2\theta^2(1 - \theta)(1 - e^{-1})^2$$

is of strict order $n^{-2}b^2$ if $\theta \neq 1$. If $\theta > 1$, this is negative, and consequently Theorem III.A* of [3] now shows that no Poisson approximation to $\mathscr{L}(C^{(nb)})$, whatever the choice of means, can achieve better order than $n^{-2}b^2/\log^2(nb^{-1})$. To establish the same conclusion for $\theta < 1$, use the facts that $T \leq \Sigma_{j=m}^{b} C_j^{(\infty b)} + 1$ and that $\Sigma_{j=m}^{b} C_j^{(\infty b)} \sim \mathrm{Po}(\Sigma_{j=m}^{b} \lambda_j^{(\infty b)})$ to deduce that the exponential moment $\mathbb{E}\{\exp(t|T - \mathbb{E}T|)\}$ is uniformly bounded in n for some $t > 0$, and apply Theorem III.C* of [3].

REFERENCES

[1] R. Arratia and S. Tavaré, The cycle structure of random permutations, *Ann. Probab.* (1991).

[2] R. Arratia, A. D. Barbour, and S. Tavaré, Poisson process approximations for the Ewens sampling formula (submitted).

[3] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford University Press, Oxford, England, 1992.

[4] G. A. Watterson, The sampling theory of selectively neutral alleles, *Adv. Appl. Probab.* **6**, 463–488 (1974).

Received August 8, 1991