THE EDGEWORTH EXPANSION FOR THE NUMBER OF DISTINCT OBSERVATIONS WITH THE MIXTURE OF DIRICHLET PROCESSES

HAJIME YAMATO AND MASAO KONDO

Received August 17, 2014; revised September 22, 2014

Abstract. Let a random distribution $\mathcal{P}$ on the real line $\mathbb{R}$ have the mixture of Dirichlet processes. Let $S^{(n)} = (S_1, \cdots, S_n)$ be the random partition of the positive integer n based on a sample of size n from $\mathcal{P}$. For the number $K_n = S_1 + \cdots + S_n$ of distinct observations among the sample, Yamato (2012) gives the asymptotic distribution of K_n and the rate $O(1/\log^{1/3} n)$ of its convergence. In this pager we give the Edgeworth expansion for K_n with the rate $O(1/\log^{2/5} n)$ and the rate $O(1/\log^{3/7} n)$.

1 Introduction. Let H_0 be a continuous distribution on the real line $\mathbb{R}$ and $\mathcal{B}$ be the σ-field which consists of the subsets of $\mathbb{R}$. Let θ be a positive random variable having the distribution γ. Given θ, let the random distribution $\mathcal{P}$ have the Dirichlet process $\mathcal{D}(\theta H_0)$ on $(\mathbb{R}, \mathcal{B})$ with parameters θ and H_0. Then this random distribution $\mathcal{P}$ has the mixture of Dirichlet processes $\mathcal{D}(\theta H_0)$ with the mixing distribution γ (Antoniak (1974)). For a sample of size n from the random distribution $\mathcal{P}$, S_1 denotes the number of observations which occur only once, S_2 the number of observations which occur exactly twice, ... and so on. Then $K_n = S_1 + \cdots + S_n$ denotes the number of distinct observations among the sample. For the convergence of K_n, Yamato (2012) gives

$$\sup_{-\infty<x<\infty}\left|P\left(\frac{K_n}{\log n} \le x\right) - \gamma(x)\right| = O\left(\frac{1}{\log^{1/3} n}\right).$$

In case the distribution γ is degenerate at θ_0, that is the θ equals to a positive constant θ_0, the random distribution $\mathcal{P}$ has the Dirichlet process $\mathcal{D}(\theta_0 H_0)$. Then, K_n has the well-known Ewens sampling formula and the asymptotic normality, whose Edgeworth expansion is given by

$$P\left(\frac{K_n - \theta_0 \log n}{\sqrt{\theta_0 \log n}} \le x\right) = \Phi(x) - \frac{1}{6\sqrt{\theta_0 \log n}}\phi(x)\left[x^2 - 1 - 6\theta_0\psi(\theta_0)\right] + O\left(\frac{1}{\log n}\right),$$

which holds uniformly in $x \in \mathbb{R}$ (Yamato (2013)). Here Φ and ϕ are the distribution function and the density function of the standard normal distribution, respectively, and ψ is the digamma function defined by $\psi(x) = \Gamma'(x)/\Gamma(x)$, where $\Gamma(x)$ is the gamma function. The purpose of this paper is to give the Edgeworth expansion for K_n, in case $\mathcal{P}$ has the mixture of Dirichlet processes $\mathcal{D}(\theta H_0)$ with the mixing distribution γ which is not degenerate. We denote the distribution function (d.f) of the distribution γ by $G(x)$. Let g be the bounded density function of the d.f. G.

In the section 2 , we give the Edgeworth expansion for K_n with the rate $O(1/\log^{2/5} n)$. In the section 3, we give it with the rate $O(1/\log^{3/7} n)$. In the section 4, we show numerical examples.

2000 *Mathematics Subject Classification.* Primary 62E20; Secondary 60E99.

Key words and phrases. Edgeworth expansion, mixture of Dirichlet processes, random partition, rate of convergence, smoothing lemma.

2 The Edgeworth expansion with the rate $1/\log^{2/5} n$. We first note the random variable P_n^*, which has the Poisson distribution with the mean $\theta(\log n - \psi(\theta))$, given θ. By Lemma 2.1 of Yamato (2013), we have:

Lemma 2.1 *Under the condition that $E_\gamma \theta$ and $E_\gamma \theta^{-1}$ are finite ,*

$$\sup_{B \subset \mathbb{Z}_+} \big| P(K_n \in B) - P(P_n^* \in B) \big| = O\Big(\frac{1}{\log n}\Big), \quad n \to \infty, \tag{2.1}$$

where where $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ and E_γ denotes the expectation with respect to the distribution γ.

In this section 2, we suppose that $E_\gamma(\theta^{-1})$, $E_\gamma(\theta^2)$, and $E_\gamma[\theta^2\psi(\theta)^2]$ exist. The following conditions are necessary for the proof of the proposition 2.2 using the smoothing lemma (see, for example, Petrov (1995; Theorem 5.2)); (i) $g(x)$ is twice differentiable, (ii) $x\psi(x+1)g(x)$, $g(x)$ and $xg'(x)$ are the functions of bounded variation, (iii) $g'(x)$, $xg''(x)$ and $[x\psi(x+1)g(x)]'$ are bounded, and (iv) $g(x) \to 0$, $xg'(x) \to 0$ as $x \to 0$, and $x\psi(x+1)g(x) \to 0$, $xg'(x) \to 0$ as $x \to +\infty$. Note that for $x \geq 0$, $\psi(x+1)$ is monotone increasing and $\psi(x+1) \geq \psi(1)$, where $-\psi(1)$ equals Euler's constant $(= 0.57721 \cdots)$. Then we have:

Proposition 2.2 *For $n > 3$, we have*

$$\sup_{-\infty < x < \infty} \left| P\Big(\frac{K_n}{\log n} \leq x\Big) - \Big[G(x) + \frac{1}{2\log n}\Big\{\big[2x\psi(x+1) - 1\big]g(x) + xg'(x)\Big\}\Big] \right| = O\Big(\frac{1}{\log^{2/5} n}\Big). \tag{2.2}$$

In the following proof, we use the well-known relations for any complex number z such that

$$e^z = 1 + z + \frac{c_1}{2} \mid z \mid^2, \tag{2.3}$$

$$= 1 + z + \frac{1}{2}z^2 + \frac{c_2}{6} \mid z \mid^3, \tag{2.4}$$

where for $i = 1, 2$ c_i is a complex number satisfying $\mid c_i \mid \leq 1$.

Proof of Proposition 2.2. Given θ, the characteristic function of $P_n^*/\log n$ is given by the following; For $-\infty < t < \infty$,

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\} \mid \theta\Big] = \exp\big\{\theta[\log n - \psi(\theta)]\big[e^{it/\log n} - 1\big]\big\}, \tag{2.5}$$

which is written as

$$= \exp\theta\Big\{\big[\log n - \psi(\theta)\big]\Big[\frac{it}{\log n} - \frac{t^2}{2\log^2 n} + \frac{c_{1n}}{6}\frac{\mid t \mid^3}{\log^3 n}\Big]\Big\}$$

by (2.4), where c_{1n} is a complex number such that $\mid c_{1n} \mid \leq 1$. Thus we have

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\} \mid \theta\Big] = \exp\theta\{it + A_1\} = e^{it\theta} \times e^{\theta A_1} \tag{2.6}$$

where

$$A_1 = -\psi(\theta)\frac{it}{\log n} - \frac{t^2}{2\log n} + \psi(\theta)\frac{t^2}{2\log^2 n} + \frac{c_{1n}}{6}\frac{\mid t \mid^3}{\log^2 n} - \psi(\theta)\frac{c_{1n}}{6}\frac{\mid t \mid^3}{\log^3 n}.$$

By using (2.3) to the term $e^{\theta A_1}$ of the right-hand side of (2.6), we have

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\} \mid \theta\Big] = e^{it\theta}\Big\{1+\theta A_1 + \frac{c_{2n}(\theta)}{2}\theta^2 \mid A_1\mid^2\Big\},$$

where $c_{2n}(\theta)$ is a complex number such that $\mid c_{2n}(\theta)\mid\le 1$. This is written as

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\} \mid \theta\Big] = e^{it\theta}\Big\{1-\theta\psi(\theta)\frac{it}{\log n} - \frac{\theta t^2}{2\log n} + B_1\Big\},$$

where

$$B_1 = \theta B_{10} + \frac{c_{2n}(\theta)}{2}\theta^2 \mid A_1\mid^2,\quad B_{10} = \Big[\psi(\theta)\frac{t^2}{2\log^2 n} + \frac{c_{1n}}{6}\frac{\mid t\mid^3}{\log^2 n} - \psi(\theta)\frac{c_{1n}}{6}\frac{\mid t\mid^3}{\log^3 n}\Big]. \tag{2.7}$$

Thus we get

$$\left|E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\} \mid \theta\Big] - e^{it\theta}\Big\{1-\theta\psi(\theta)\frac{it}{\log n} - \frac{\theta t^2}{2\log n}\Big\}\right| \le \mid B_1\mid. \tag{2.8}$$

About B_{10}, for $\mid t\mid < \log^{2/5} n$ $(n>3)$ we have

$$\mid B_{10}\mid \le \frac{\mid t\mid}{\log^{4/5} n}\Big[\frac{1}{6}+\frac{2}{3}\mid\psi(\theta)\mid\Big], \tag{2.9}$$

because of the following relations,

$$\frac{\mid t\mid}{\log^2 n} < \frac{1}{\log^{8/5} n} < \frac{1}{\log^{4/5} n},\quad \frac{t^2}{\log^2 n} < \frac{1}{\log^{6/5} n} < \frac{1}{\log^{4/5} n}\quad \text{and}\quad \frac{t^2}{\log^3 n} < \frac{1}{\log^{4/5} n}.$$

The similar inequalities to these are used, hereafter. About $|A_1|$, for $|t| < \log^{2/5} n$ $(n>3)$, by $t^2 < \log n$ and $\log n > 1$ we have

$$\begin{aligned}
\mid A_1\mid^2 &\le \Big\{\mid\psi(\theta)\mid\frac{\mid t\mid}{\log n} + \frac{t^2}{2\log n} + \mid\psi(\theta)\mid\frac{t^2}{2\log n} + \frac{\mid t\mid}{6\log n} + \mid\psi(\theta)\mid\frac{\mid t\mid}{6\log n}\Big\}^2\\
&= \Big\{\Big(\frac{7}{6}\mid\psi(\theta)\mid + \frac{1}{6}\Big)\frac{\mid t\mid}{\log n} + \big(\mid\psi(\theta)\mid+1\big)\frac{t^2}{2\log n}\Big\}^2\\
&\le 2\Big\{\Big(\frac{7}{6}\mid\psi(\theta)\mid + \frac{1}{6}\Big)^2\frac{t^2}{\log^2 n} + \big(\mid\psi(\theta)\mid+1\big)^2\frac{t^4}{4\log^2 n}\Big\}\\
&\le 2\Big\{\Big(\frac{7}{6}\mid\psi(\theta)\mid + \frac{1}{6}\Big)^2 + \frac{1}{4}\big(\mid\psi(\theta)\mid+1\big)^2\Big\}\frac{\mid t\mid}{\log^{4/5} n}.
\end{aligned} \tag{2.10}$$

By applying (2.9) and (2.10) to (2.7), for $|t| < \log^{2/5} n$ $(n>3)$,

$$\mid B_1\mid \le \frac{\mid t\mid}{\log^{4/5} n}\Big[\frac{1}{6}\theta + \frac{2}{3}\theta\mid\psi(\theta)\mid + \Big\{\Big(\frac{7}{6}\theta\mid\psi(\theta)\mid + \frac{1}{6}\theta\Big)^2 + \frac{1}{4}\big(\theta\mid\psi(\theta)\mid+\theta\big)^2\Big\}\Big]. \tag{2.11}$$

Therefore, for $|t| < \log^{2/5} n$ $(n>3)$, by $E_\gamma(\theta^2)$ and $E_\gamma[\theta^2\psi(\theta)^2] < \infty$, (2.8) and (2.11) yield

$$\left|E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\}\Big] - \Big\{E_\gamma e^{it\theta} - \frac{it}{\log n}E_\gamma\big[\theta\psi(\theta)e^{it\theta}\big] - \frac{t^2}{2\log n}E_\gamma\big[\theta e^{it\theta}\big]\Big\}\right| \le d_{11}\frac{\mid t\mid}{\log^{4/5} n}, \tag{2.12}$$

where d_{11} is a positive constant. Since $E_\gamma e^{it\theta}$ is the characteristic function of the distribution function G, that is, it is the Fourier transform of the distribution function $G(x)$. Similarly, $-itE_\gamma[\theta\psi(\theta)e^{it\theta}]$ is the Fourier transform of the function $x\psi(x)g(x)$, and $-t^2E_\gamma[\theta e^{it\theta}]$ is the Fourier transform of the function $\{xg(x)\}' = g(x) + xg'(x)$. Therefore, by applying the smoothing lemma to (2.12), we have the following.

$$\sup_x \left| P\left(\frac{P_n^*}{\log n} \le x\right) - \left[G(x) + \frac{1}{\log n} x\psi(x)g(x) + \frac{1}{2\log n}\{g(x) + xg'(x)\}\right]\right| \le \frac{d_{11}}{\log^{4/5} n}\int_0^{\log^{2/5} n} dt \; + \; \frac{d_{12}}{\log^{2/5} n} = O\left(\frac{1}{\log^{2/5} n}\right), \tag{2.13}$$

where d_{12} is positive constant depending only on d_{11}. We get (2.2) by (2.1) and (2.13), using the relation $x\psi(x) = x\psi(x+1) - 1$.

3 The Edgeworth expansion with the rate $1/\log^{3/7} n$. In addition to the assumption of the section 2, we assume that $g(x)$ is differentiable four times, $\{x^2g(x)\}^3$ is the function of bounded variation and $\{x^2g(x)\}^4$ is bounded. Besides, we suppose $E_\gamma(\theta^3)$ and $E[\theta^3|\psi(\theta)|^3]$ exist. Then we have:

Proposition 3.1 *For $n > 3$, we have*

$$\sup_{-\infty<x<\infty} \left| P\left(\frac{K_n}{\log n} \le x\right) - \left[G(x) + \frac{1}{2\log n}\Big\{\big[2x\psi(x+1) - 1\big]g(x) + xg'(x)\Big\} + \frac{1}{8\log^2 n}\{x^2g(x)\}^{(3)}\right]\right| = O\left(\frac{1}{\log^{3/7} n}\right). \tag{3.1}$$

In the following proof, in addition to (2.4) we use the well-known relation for any complex number z such that

$$e^z = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{c_3}{24} \mid z \mid^4, \tag{3.2}$$

where c_3 is a complex number satisfying $\mid c_3 \mid \le 1$.

Proof of Proposition 3.1. Given θ, the characteristic function (2.5) of $P_n^*/\log n$ is written as

$$E\left[\exp\left\{it\frac{P_n^*}{\log n}\right\} \;\Big|\; \theta\right] = \exp\left\{\theta[\log n - \psi(\theta)]\left[\frac{it}{\log n} - \frac{t^2}{2\log^2 n} - \frac{it^3}{6\log^3 n} + \frac{c_{3n}}{24}\frac{t^4}{\log^4 n}\right]\right\},$$

by (3.2), where $-\infty < t < \infty$ and c_{3n} is a complex number satisfying $\mid c_{3n} \mid \le 1$. Thus we can write

$$E\left[\exp\left\{it\frac{P_n^*}{\log n}\right\} \;\Big|\; \theta\right] = \exp\theta\{it + A_2\} = e^{it\theta} \times e^{\theta A_2}, \tag{3.3}$$

where

$$A_2 = -\psi(\theta)\frac{it}{\log n} - \frac{t^2}{2\log n} + \psi(\theta)\frac{t^2}{2\log^2 n} - \frac{it^3}{6\log^2 n} + \psi(\theta)\frac{it^3}{6\log^3 n} + \frac{c_{3n}}{24}\frac{t^4}{\log^3 n} - \frac{c_{3n}}{24}\psi(\theta)\frac{t^4}{\log^4 n}.$$

By using (2.4) to the term $e^{\theta A_2}$ of the right-hand side of (3.3), we have

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\}\ \Big|\ \theta\Big] = e^{it\theta}\Big\{1+\theta A_2+\frac{1}{2}\theta^2 A_2^2+\frac{c_{4n}(\theta)}{6}\theta^3 \mid A_2\mid^3\Big\},$$

where $c_{4n}(\theta)$ is a complex number such that $\mid c_{4n}(\theta)\mid\leq 1$. This is written as

$$E\Big[\exp\Big\{it\frac{P_n^*}{\log n}\Big\}\ \Big|\ \theta\Big] = e^{it\theta}\Big\{1-\theta\psi(\theta)\frac{it}{\log n}-\frac{\theta t^2}{2\log n}+\frac{\theta^2 t^4}{8\log^2 n}+B_2\Big\}, \tag{3.4}$$

where

$$\begin{aligned}
B_2 &= \theta B_{21}+\frac{\theta^2}{2}B_{22}+\frac{c_{4n}(\theta)}{6}\theta^3\mid A_2\mid^3,\\
B_{21} &= \psi(\theta)\frac{t^2}{2\log^2 n}-\frac{it^3}{6\log^2 n}+\psi(\theta)\frac{it^3}{6\log^3 n}+\frac{c_{3n}}{24}\frac{t^4}{\log^3 n}-\frac{c_{3n}}{24}\psi(\theta)\frac{t^4}{\log^4 n},\\
B_{22} &= A_2^2-\frac{t^4}{4\log^2 n}.
\end{aligned}$$

For $|t|<\log^{3/7} n$ $(n>3)$, by $|t|<\log^{1/2} n$ we have

$$\mid B_{21}\mid\leq\Big(\frac{5}{24}+\frac{17}{24}\mid\psi(\theta)\mid\Big)\frac{\mid t\mid}{\log n} < \Big(\frac{5}{24}+\frac{17}{24}\mid\psi(\theta)\mid\Big)\frac{\mid t\mid}{\log^{6/7} n}. \tag{3.5}$$

About $\mid A_2\mid$, at first for $|t|<\log^{3/7} n$ $(n>3)$, by $|t|<\log^{1/2} n$ we have

$$\begin{aligned}
\mid A_2\mid &\leq \mid\psi(\theta)\mid\frac{\mid t\mid}{\log n}+\frac{t^2}{2\log n}+\mid\psi(\theta)\mid\frac{t^2}{2\log^2 n}+\frac{\mid t\mid^3}{6\log^2 n}\\
&\qquad +\mid\psi(\theta)\mid\frac{\mid t\mid^3}{6\log^3 n}+\frac{t^4}{24\log^3 n}+\mid\psi(\theta)\mid\frac{t^4}{24\log^4 n}\\
&\leq \mid\psi(\theta)\mid\frac{\mid t\mid}{\log n}+\frac{t^2}{2\log n}+\mid\psi(\theta)\mid\frac{\mid t\mid}{2\log^{3/2} n}+\frac{\mid t\mid}{6\log n}\\
&\qquad +\mid\psi(\theta)\mid\frac{\mid t\mid}{6\log^2 n}+\frac{\mid t\mid}{24\log^{3/2} n}+\mid\psi(\theta)\mid\frac{\mid t\mid}{24\log^{5/2} n}.
\end{aligned}$$

Thus, for $|t|<\log^{3/7} n$ $(n>3)$, we have

$$\mid A_2\mid\leq\frac{\mid t\mid}{\log n}\eta(\theta)+\frac{t^2}{2\log n}\quad\text{where}\quad \eta(\theta)=\frac{41}{24}\mid\psi(\theta)\mid+\frac{5}{24}. \tag{3.6}$$

Therefore, for $|t|<\log^{3/7} n$ $(n>3)$ we have

$$\mid A_2\mid^3\leq 4\Big\{\frac{\mid t\mid^3}{\log^3 n}\eta(\theta)^3+\frac{t^6}{8\log^3 n}\Big\}\leq 4\Big\{\eta(\theta)^3+\frac{1}{8}\Big\}\frac{\mid t\mid}{\log^{6/7} n}. \tag{3.7}$$

For the evaluation of B_{22}, at first we write A_2 as

$$A_2=-\frac{t^2}{2\log n}+A_{21}$$

where

$$A_{21} = -\psi(\theta)\frac{it}{\log n} + \psi(\theta)\frac{t^2}{2\log^2 n} - \frac{it^3}{6\log^2 n} + \psi(\theta)\frac{it^3}{6\log^3 n} + \frac{c_{3n}}{24}\frac{t^4}{\log^3 n} - \frac{c_{3n}}{24}\psi(\theta)\frac{t^4}{\log^4 n}.$$

Then we

$$| B_{22} | \le \frac{t^2}{\log n} | A_{21} | + | A_{21} |^2 . \tag{3.8}$$

We note that A_{21} is obtained by deleting $-t^2/(2\log n)$ from A_2. Similarly to (3.6), for $|t| < \log^{3/7} n$ $(n > 3)$, we have

$$| A_{21} | \le \frac{| t |}{\log n}\eta(\theta). \tag{3.9}$$

Applying (3.9) to (3.8), for $|t| < \log^{3/7} n$ $(n > 3)$, we have

$$| B_{22} | \le \frac{| t |^3}{\log^2 n}\eta(\theta) + \frac{t^2}{\log^2 n}\eta(\theta)^2 \le \{\eta(\theta) + \eta(\theta)^2\}\frac{| t |}{\log n}. \tag{3.10}$$

From (3.4) we get

$$\left| E\left[\exp\left\{it\frac{P_n^*}{\log n}\right\} \Big| \theta\right] - e^{it\theta}\left\{1 - \theta\psi(\theta)\frac{it}{\log n} - \frac{\theta t^2}{2\log n} + \frac{\theta^2 t^4}{8\log^2 n}\right\}\right| \le B_2, \tag{3.11}$$

and from (3.5), (3.7) and (3.10) we have

$$| B_2 | \le \theta\left(\frac{5}{24} + \frac{17}{24} | \psi(\theta) | \right)\frac{| t |}{\log^{6/7} n} + \frac{1}{2}\theta^2\{\eta(\theta) + \eta(\theta)^2\}\frac{| t |}{\log n} + \frac{2}{3}\theta^3\left\{\eta(\theta)^3 + \frac{1}{8}\right\}\frac{| t |}{\log^{6/7} n}. \tag{3.12}$$

Therefore, for $|t| < \log^{3/7} n$ $(n > 3)$, under the condition $E_\gamma(\theta^3),\ E_\gamma[\theta^3|\psi(\theta)|^3] < \infty$, (3.11) and (3.12) give

$$\left| E\left[\exp\left\{it\frac{P_n^*}{\log n}\right\}\right] - \left\{E_\gamma e^{it\theta} - \frac{it}{\log n}E_\gamma\left[\theta\psi(\theta)e^{it\theta}\right] - \frac{t^2}{2\log n}E_\gamma\left[\theta e^{it\theta}\right] + \frac{t^4}{8\log^2 n}E_\gamma\left[\theta^2 e^{it\theta}\right]\right\}\right| \le d_{21}\frac{|t|}{\log^{6/7} n},$$

where d_{21} is a positive constant. Since $t^4 E_\gamma[\theta^2 e^{it\theta}]$ is the Fourier transform of the function $\{xg(x)\}^{(3)}$, by the reason similar to (2.13) we have

$$\sup_x \left| P\left(\frac{P_n^*}{\log n} \le x\right) - \left[G(x) + \frac{1}{\log n}x\psi(x)g(x) + \frac{1}{2\log n}\{g(x) + xg'(x)\} + \frac{1}{8\log^2 n}\{x^2 g(x)\}^{(3)}\right]\right| = O\left(\frac{1}{\log^{3/7} n}\right). \tag{3.13}$$

Therefore we get (3.1) by (2.1) and (3.13), using the relation $x\psi(x) = x\psi(x+1) - 1$.

4 numerical examples. We examine the Propositions 2.2 and 3.1 graphically by using the gamma distribution as γ whose density is given by $g_c(x) = x^{c-1}e^{-x}/\Gamma(c)$. The distribution function of $K_n/\log n$ is obtained approximately by using the random numbers of R and described by the step function.

At first, for the examination of (2.2) by taking $c > 1$. Then, the conditions of the Propositions 2.2 are satisfied. The approroximate function $G_1(x) = G(x) + \big\{[2x\psi(x+1) - 1]g(x) + xg'(x)\big\}/(2\log n)$ is described by the broken curve. The distribution function G_c of $g_c(x)$ is described by the dotted curve. For $n = 50$, the Figure's 1, 2, 3 and 4 give the cases of $c = 1.1$, $c = 1.5$, $c = 2$ and $c = 3$. If c is small and near 1, then the function $G_1(x)$ is good approximation to the distribution function of $K_n/\log n$. Even if c increases, the function $G_1(x)$ is better than $G_c(x)$ as the approximation to the distribution function of $K_n/\log n$. But, the tail is not good approximation, similar to the usual Edgeworth expansion.

Figure 1: c=1.1, n=50

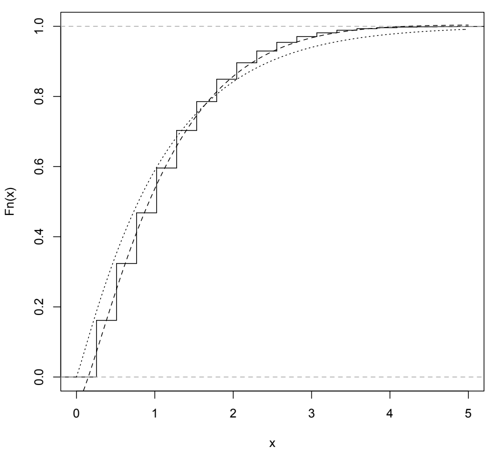


Figure 2: c=1.5, n=50

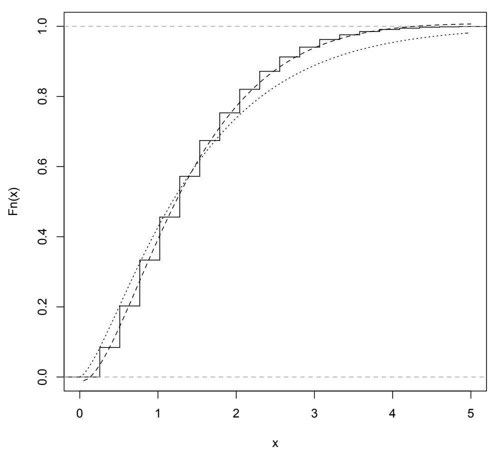


Figure 3: c=2.0, n=50

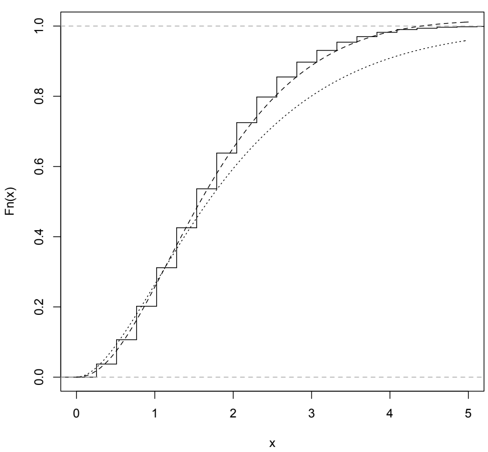
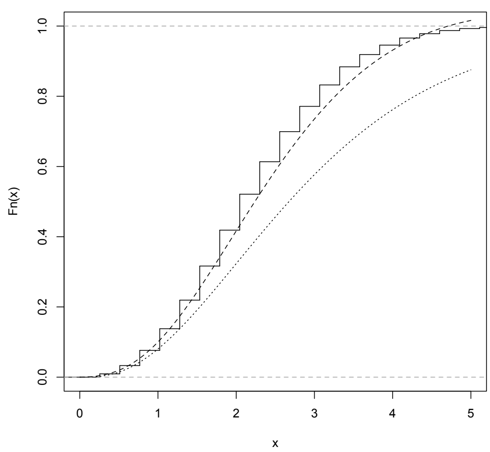


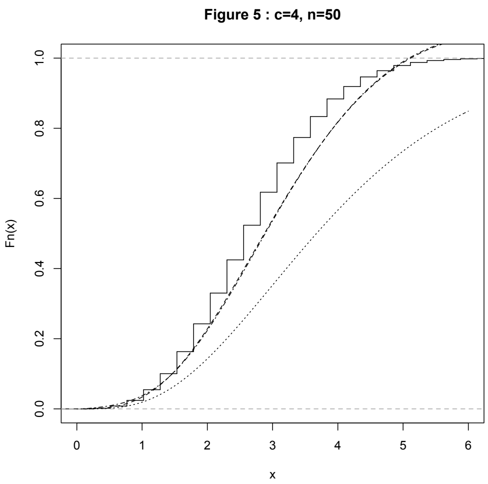
Figure 4: c=3.0, n=50

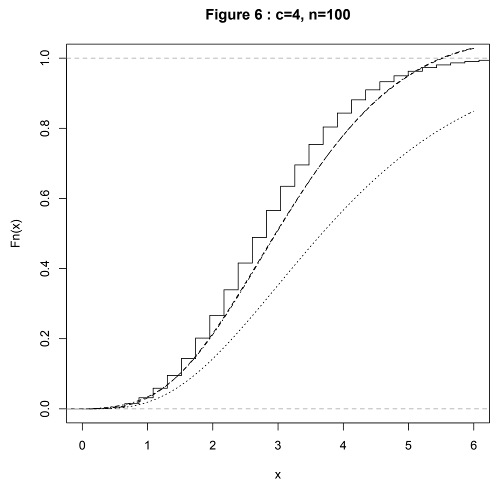


Next, we examine the relation (3.5) by taking $c = 4$. Then, the conditions of the Propositions 3.1 are satisfied.

The approximate distribution G_1 is described by the broken curve. The approximate function $G_2(x) = G(x) + \big\{[2x\psi(x+1) - 1]g(x) + xg'(x)\big\}/(2\log n) + \{x^2 g(x)\}^{(3)}/(8\log^2 n)$

is described by the dot-broken curve. The distribution function G_c of $g_c(x)$ is described by the dotted curves. For $c = 4$, the Figure's 5 and 6 give the cases of $n = 50$ and $n = 100$, respectively. Both the functions G_1 and G_2 give a little good approximate to the distribution function of $K_n/\log n$. But there are no obvious difference between G_1 and G_2, becuase the value of $G_2(x) - G_1(x) = \{x^2 g(x)\}^{(3)}/(8\log^2 n)$ is small. The little difference may be seen at the left tail.





Acknowledgements. The authors are grateful to the referee for his careful reading and useful comments.

References

Antoniak, C. E. (1774). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Ann. Statist.*, **2**, 1152–1174.

Petrov, V. (1995) *Limit theorems of probability theory.* New York : Oxford Univ. Press.

Yamato, H. (2012). Asymptotic distribution of number of distinct observations among a sample from mixture of Dirichlet processes. *Bull. Inform. Cyber.*, **44**, 41–47.

Yamato, H. (2013). Edgeworth expansion for the number of distinct components associated with Ewens Sampling Formula. *J. Japan Statist. Soc*, **43**, 17–28

Communicated by *Hisao Nagao*

First author:
Emeritus of Kagoshima University, Take 3-32-1-708 Kagoshima 890–0045, Japan
Second author:
Department of Mathematics and Computer Science, Kagoshima University, Kagosima 890-0065, Japan