ON IMPROVEMENTS OF THE ORDER OF APPROXIMATION IN THE POISSON LIMIT THEOREM

K. BOROVKOV,* *University of Melbourne*
D. PFEIFER, ** *Universität Hamburg*

Abstract

In this paper we consider improvements in the rate of approximation for the distribution of sums of independent Bernoulli random variables via convolutions of Poisson measures with signed measures of specific type. As a special case, the distribution of the number of records in an i.i.d. sequence of length n is investigated. For this particular example, it is shown that the usual rate of Poisson approximation of order $\mathrm{O}(1/\log n)$ can be lowered to $\mathrm{O}(1/n^2)$. The general case is discussed in terms of operator semigroups.

SIGNED MEASURES; RECORDS; OPERATOR SEMIGROUPS
AMS CLASSIFICATION: PRIMARY: 60 F 05, 28 A 33; SECONDARY: 47 D 03

1. Introduction

Let $X_1, \ldots, X_n$ be independent random variables with a continuous cumulative distribution function. We say that $X_k,\ k \leq n$, is a record of this sequence if $X_k > \max\{X_1, \ldots, X_{k-1}\}$. By convention, X_1 is always a record value. The corresponding record indices $I_1, \ldots, I_n \in \{0, 1\}$ mark the times when new records occur, i.e. $I_k = 1$ iff X_k is a record. Due to the remarkable Dwass – Rényi theorem [Dwass (1960), Rényi (1962)] we know that the record indices are independent with success probabilitites p_k given by

$$p_k = \mathbf{P}(I_k = 1) = 1/k, \quad k = 1, \ldots, n.$$

In the first part of the paper we consider the behaviour of the distribution function

$$G_n(x) = \mathbf{P}(S_n \leq x), \quad S_n = \sum_{k=2}^{n} I_k \tag{1}$$

of the number S_n of "true" records. Apart from the natural motivation, related to the theory of records, the study of the distribution function G_n is of interest for other applications too, e.g. in connection with the secretary problem (cf. Pfeifer (1989)) or for the linear search problem of the maximum element in a field of n entries, where S_n denotes the number of re–storages during the procedure (for details see e.g. Kemp

---

* Postal address: Statistics Department, University of Melbourne, Australia.
This research was partially supported by a grant of the Alexander von Humboldt Foundation, Germany.
** Postal address: Universität Hamburg, Institut für Mathematische Stochastik, Bundesstr. 55, D–20146 Hamburg, Germany.

(1984) and Pfeifer (1991)). Another field of relevance is the average case analysis of the simplex method in linear programming (see Ross (1982) and Deheuvels and Pfeifer (1987)). Clearly,

$$\nu_n = \mathbf{E}S_n = \sum_{k=2}^{n} \frac{1}{k} = \log n + \gamma - 1 + O\left(\frac{1}{n}\right), \tag{2}$$

$$\mathrm{Var}(S_n) = \sum_{k=2}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) = \log n + \gamma - \frac{\pi^2}{6} + O\left(\frac{1}{n}\right) \tag{3}$$

as $n \to \infty$, where $\gamma = 0.5772\ldots$ is Euler's constant. Since S_n is the sum of independent bounded r.v.'s, we are in the range of normal approximation. However, as is easy to see, the rate of this approximation is only $O(\log^{-1/2} n)$. Likewise, S_n is the sum of $\{0,1\}$-valued r.v.'s, and although the success probabilities for the first summands of S_n are not small, Poisson approximation is also possible due to the fact that the variance–to–mean–ratio tends to one for $n \to \infty$. Moreover, the exact first term in an expansion of the rate of convergence is known: if Π_λ is the distribution function of the Poisson distribution with mean λ, then

$$\begin{aligned} \sup_x |G_n(x) - \Pi_{\nu_n}(x)| &= \frac{\pi^2/6 - 1}{2\nu_n\sqrt{2\pi \mathrm{e}}} + O(\log^{-3/2} n) = \\ &= \frac{\beta}{\log n + \gamma - 1} + O(\log^{-3/2} n), \quad \beta = 0.078\ldots, \end{aligned} \tag{4}$$

which is better than for the normal approximation (for a more thorough discussion of a comparison between normal and Poisson approximation in such cases, see Deheuvels and Pfeifer (1988)). However, the rate of convergence is still far from being sufficient, from a practical point of view. In such situations, it could therefore be fruitful to turn to a "second-order" approximation, obtainable usually via some correction of the main term (for second-order refinements of the Poisson approximation see e.g. Borovkov (1988) and references therein). In the particular situation under consideration, this procedure is indeed very promising: the error of the "corrected" approximation, which is just a convolution of Π_{ν_n} with some fixed signed measure and a two-point distribution, is of the order $O(n^{-2})$, compared with $O(\log^{-1} n)$ in (4).

The third section of the paper is concerned with a discussion of possible extensions of such correction techniques to more general situations. It will be shown there that indeed a general procedure of this kind exists if for the success probabilities p_k there holds

$$\sum_{k=2}^{\infty} p_k = \infty, \quad \sum_{k=2}^{\infty} p_k^2 < \infty, \tag{5}$$

which covers the case of records above. However, this general method which is based on the semigroup approach as originally developed in Deheuvels and Pfeifer (1986) is not as sharp as the particular approach possible in the very special situation outlined above. For records in particular, the general method gives a rate of approximation of $O\big((n \log n)^{-1}\big)$ only, which is of course still considerably better than what is achievable otherwise. Problems concerning record indicators of this general form occur e.g. in the generalized secretary problem (see Pfeifer (1989)) or in connection with search problems when non–uniform distributions for the position of entries are possible (see

Pfeifer (1991)). In extreme value theory, such indicators occur in a natural way also in Nevzorov's record model (cf. Nevzorov (1988) and Borovkov and Pfeifer (1993)).

2. Refined Poisson approximation for the i.i.d. record model

Here we shall give a detailed analysis of the second–order approach via signed measures for the Poisson approximation of the distribution of the number S_n of true records in a sequence of n i.i.d. random variables.

Let $L(x)$ be the distribution function of the signed measure on the integers with generating function $1/\Gamma(1+z)$, $\Gamma(z)$ being the gamma-function,

$$V_n(x) = \mathrm{e}^{-\mu_n}(E_0(x) + l_n E_2(x)), \quad \mu_n = \frac{1}{2(n+1)}, \quad l_n = \frac{1}{2}(\mathrm{e}^{\mu_n} - 1 + \mu_n), \tag{6}$$

$E_k(x)$ corresponding to the distribution function of the unit mass concentrated in the point k. Let further denote

$$M_n = \Pi_{\lambda_n} * L * V_n, \quad \lambda_n = \log(n+1) - \frac{3}{2(n+1)}. \tag{7}$$

This is the distribution function of a (possibly) signed measure, which in fact is easy to evaluate. As for the explicit form of L, we have $1/\Gamma(1+z) = 1/z\Gamma(z)$, and the first 26 terms c_k of the series expansion for $1/\Gamma(z)$ near zero (that is, the values of the measure L) can be found e.g. in Abramowitz and Stegun (1964) (they decrease in absolute value rather fast; say, $|c_9| \approx 10^{-3}$, $|c_{13}| \approx 10^{-6}$, $|c_{26}| \approx 10^{-15}$). On the other hand,

$$\frac{1}{\Gamma(1+z)} = (1+z)\exp\left((\gamma - 1)z - \sum_{k\geq 2}\frac{(-1)^k}{k}(\zeta(k) - 1)z^k\right), \tag{8}$$

where $\zeta(k)$ is the Riemann zeta function. Note also that the coefficients b_k in

$$\exp\left(\sum_{k\geq 1} a_k z^k\right) = \sum_{k\geq 0} b_k z^k$$

can be calculated easily using the recursive formula

$$b_{j+1} = \frac{1}{j+1}\sum_{m=0}^{j}(m+1)a_{m+1}b_{j-m}, \quad b_0 = 1,$$

see e.g. Sect. 10.4 in Johnson and Kotz (1969) (the coefficients a_k decrease exponentially fast in our case: $|a_k| \sim k^{-1}2^{-k}$).

Theorem 1. *For $n > 1$,*

$$\Delta = \sup_x |G_n(x) - M_n(x)| \leq C_\Gamma \varepsilon_n \mathrm{e}^{2\varepsilon_n} + \frac{1}{16n(n+1)},$$

where

$$\varepsilon_n = \frac{5}{4(n-1)^2}, \quad C_\Gamma = \max_{|z|=1} \left| \frac{1}{\Gamma(z)} \right| < 1.9615.$$

Proof. The generating function of S_n is

$$g_n(z) = \mathbf{E} z^{S_n} = \prod_{j=2}^{n} \left(1 - \frac{1}{j} + \frac{z}{j}\right) = \frac{1}{\Gamma(1+z)} \cdot \frac{\Gamma(n+z)}{\Gamma(n+1)}.$$

Now by the second Binet formula (see e.g. 6.1.50 in Abramowitz, Stegun (1964)), for Re $z > 0$,

$$\Gamma(z) = \exp\left[\left(z - \frac{1}{2}\right)\log z - z + \frac{1}{2}\log 2\pi + 2\int_0^\infty \frac{\arctan t/z}{\exp(2\pi t) - 1}\, dt\right].$$

Hence, for $|z| = 1 < n$,

$$\begin{aligned} \frac{\Gamma(n+z)}{\Gamma(n+1)} = \exp\Big[(n+z-\frac{1}{2})\log(n+z) - (n+\frac{1}{2})\log(n+1) - \\ - z + 1 + 2\int_0^\infty \frac{\alpha-\beta}{\exp(2\pi t)-1}\, dt\Big] \end{aligned} \tag{9}$$

where $\alpha = \arctan \dfrac{t}{n+z}$, $\beta = \arctan \dfrac{t}{n+1}$. Clearly,

$$\begin{aligned} \alpha - \beta = \arctan\, \tan(\alpha - \beta) &= \arctan \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta} \\ &= \arctan\left[-\frac{(z-1)t}{(n+z)(n+1)+t^2}\right] \equiv \arctan A, \end{aligned}$$

and $|A| \le 2t/(n^2 - 1 + t^2) \le (n^2-1)^{-1/2}$. Thus, using the fact that for complex z, $\arctan z = \frac{1}{2i}\log\big((1+iz)/(1-iz)\big)$ is bounded in modulus by $|z|/(1-|z|^2)$,

$$|\arctan A| \le \frac{|A|}{1-|A|^2} \le \frac{|z-1|t}{n^2-1}\left(1 - \frac{1}{n^2-1}\right)^{-1} = \frac{|z-1|t}{n^2-2},$$

and

$$\left|2\int_0^\infty \frac{\alpha-\beta}{\exp(2\pi t)-1}\, dt\right| \le \frac{2|z-1|}{n^2-2}\int_0^\infty \frac{t}{\exp(2\pi t)-1}\, dt = \frac{|z-1|}{12(n^2-2)}, \tag{10}$$

since

$$\int_0^\infty \frac{x^{\nu-1}}{\exp(\mu x)-1}\, dx = \mu^{-\nu}\Gamma(\nu)\zeta(\nu), \quad \mathrm{Re}\,\mu > 0, \quad \mathrm{Re}\,\nu > 1$$

(see e.g. 3.411 in Gradshteyn, Ryzhik (1980)). Now

$$(n+z-\frac{1}{2})\log(n+z)-(n+\frac{1}{2})\log(n+1)=$$
$$=(n+z-\frac{1}{2})\log\left(1+\frac{z-1}{n+1}\right)+(z-1)\log(n+1)=$$
$$=z-1+(n+1)\int\limits_0^{(z-1)/(n+1)}\log(1+x)\,dx-\frac{1}{2}\log\left(1+\frac{z-1}{n+1}\right)+$$
$$+(z-1)\log(n+1)=(z-1)\left(1+\log(n+1)-\frac{1}{2(n+1)}\right)+$$
$$+\frac{(z-1)^2}{2(n+1)}+\frac{(z-1)^3\theta_1}{6(n-1)^2}+\frac{(z-1)^2\theta_2}{4(n-1)^2},\quad |\theta_i|<1.$$

From this relation and from (9) and (10) it follows that

$$\frac{\Gamma(n+z)}{\Gamma(n+1)}=\exp\Big[\lambda_n(z-1)+\mu_n(z^2-1)+\frac{z-1}{(n-1)^2}\theta_3\Big],\tag{11}$$

where $|\theta_3|\le\frac{1}{4}|\theta_1||z-1|+\frac{1}{6}|\theta_2||z-1|^2+\frac{1}{12}\le\frac{5}{4}$ for $|z|\le 1$. Here the term $\exp\big(\lambda_n(z-1)\big)$ corresponds to $\Pi_{\lambda_n}(x)$, the term $\exp\big(\mu_n(z^2-1)\big)$ corresponds to $\Pi_{\mu_n,2}(x)=\Pi_{\mu_n}(x/2)$.

Let now $G_n^-=G_n*\Pi^{-1}_{\mu_n,2}$, where $\Pi^{-1}_{\mu_n,2}$, the inverse to $\Pi_{\mu_n,2}$ in the algebra of distributions with convolution operation, corresponds to the generating function $r_n(z)=\exp(-\mu_n(z^2-1))$; then

$$\Delta_0\equiv\sup_x|G_n(x)-G_n^-*V_n(x)|\le\sup_x|\Pi_{\mu_n,2}(x)-V_n(x)|\cdot\mathrm{var}\,(G_n^-)\le\frac{1}{16n(n+1)}.$$

Here we made use of the fact that for the total variation $\mathrm{var}(G_n^-)\le 1+2/(2n+1)$, since

$$\mathrm{var}(\Pi^{-1}_{\mu_n,2})=\exp(2\mu_n)\le\frac{1+\mu_n}{1-\mu_n}=1+\frac{2}{2n+1},$$

and also of the following bound implied by the choice of l_n:

$$\sup_x|\Pi_{\mu_n,2}(x)-V_n(x)|=\frac{1}{2}\big(1-\mathrm{e}^{-\mu_n}-\mu_n\mathrm{e}^{-\mu_n}\big)\le\frac{\mu_n^2}{4}=\frac{1}{16(n+1)^2}.\tag{12}$$

Clearly, $\Delta\le\Delta_0+\Delta_1$, where

$$\Delta_1\equiv\sup_x|G_n^-*V_n(x)-\Pi_{\lambda_n}*V_n*L(x)|\le\sup_x|G_n^-(x)-\Pi_{\lambda_n}*L(x)|.$$

By Tzaregradskii's (1958) inequality (cf. Kruopis (1986)),

$$\Delta_1\le\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\frac{|g_n(\mathrm{e}^{it})r_n(\mathrm{e}^{it})-\omega_n(\mathrm{e}^{it})|}{|\mathrm{e}^{it}-1|}\,dt,\tag{13}$$

where $\omega_n(z)=\exp(\lambda_n(z-1))/\Gamma(1+z)$. Now, for $|z|=1$, we have

$$|\Gamma(z+1)|=|\Gamma(z)|,\quad \mathrm{Re}\,(z-1)\le 0,\quad |z-1|\le 2,$$

and hence one obtains from (11)

$$|g_n(z)r_n(z) - \omega_n(z)| = \frac{1}{|\Gamma(z)|} \cdot \left| \frac{\Gamma(n+z)}{\Gamma(n+1)} \mathrm{e}^{-\mu_n(z^2-1)} - \mathrm{e}^{\lambda_n(z-1)} \right| \le$$
$$\le C_\Gamma \left| \exp\big((z-1)(n-1)^{-2}\theta_3\big) - 1 \right| \le C_\Gamma \varepsilon_n |z-1| \exp(2\varepsilon_n),$$

where, to get the last relation, we have made use of the inequalities $|\mathrm{e}^\delta - 1| \le |\delta| \mathrm{e}^{|\delta|}$ and $|\theta_3| \le 5/4$. Combining this with (13) it follows that $\Delta_1 \le C_\Gamma \varepsilon_n \mathrm{e}^{2\varepsilon_n}$. Hence we obtain from (12) the inequality stated in Theorem 1 for Δ. It remains to estimate C_Γ. Using Euler's infinite product formula

$$\frac{1}{\Gamma(z)} = z\mathrm{e}^{\gamma z} \prod_{n=1}^{\infty} \left(\left(1 + \frac{z}{n}\right) \mathrm{e}^{-z/n} \right) \tag{14}$$

and the representation $\gamma = \alpha \sum_{n \ge 1} n^{-2}$, $\alpha = 6\gamma\pi^{-2}$, we have, after term-wise estimating the product, the bound

$$C_\Gamma^2 \le \Gamma(1-\alpha) \exp(\gamma/\alpha - \alpha\zeta(3) - \gamma\alpha),$$

where ζ is the Riemann zeta function, as before. A simple numerical calculation shows that $C_\Gamma \le 1.9615$. Note that this is a rather sharp estimate, for $|1/\Gamma(i)| > 1.9173$. Hence Theorem 5 is proved.

Remark. In conclusion of this section, note that

$$g_n(z) = \frac{1}{zn!}[z]^n = \frac{1}{n!} \sum_{k=1}^{n} |s(n,k)| z^{k-1},$$

where $[x]^n$ is the rising factorial polynomial, and $s(n, k)$ are Stirling numbers of the first kind, so that

$$\mathbf{P}(S_n = k) = \frac{1}{n!} |s(n, k+1)|. \tag{15}$$

The numbers $s(n, k)$ are of great importance in many fields of mathematics: say, according to Jordan (1965), "they should be placed in the centre of the Calculus of Finite Differences" (see also Charalambides, Singh (1988) and Butzer et al. (1989)). One more interpretation of $|s(n, k)| = (-1)^{n+k} s(n, k)$ is the number of permutations of n elements, which have exactly k cycles. It is known (see e.g. Jordan (1965)) that, for a fixed k,

$$|s(n, k+1)| \sim (n-1)!(\log n + \gamma)^k / k! \quad \text{as } n \to \infty$$

(cf. also Kemp (1984)). The estimates for the Poisson approximation (4) together with (15) imply that

$$|s(n, k+1)| = n!\big(\mathrm{e}^{-\nu_n} \nu_n^k / k! + O(\log^{-1} n)\big) \quad \text{as } n \to \infty,$$

uniformly in $k \ge 0$. Our approximation shows that

$$|s(n, k+1)| = n!(m_n(k) + \delta_n(k)),$$

where $m_n(k) = M_n(k) - M_n(k-0)$ is the mass of the measure M_n in the point $k \geq 0$, and, using instead of (13) just the inversion formula for generating functions, we get the estimate

$$|\delta_n(k)| \leq \frac{5}{2(n-1)^2} + \frac{1}{16n(n+1)} \leq \frac{2.6}{(n-1)^2}.$$

3. A general case

In this section we shall assume that the indicators $I_1, \ldots, I_n$ are independent with success probabilitites fulfilling the condition (5)

$$\sum_{k=2}^{\infty} p_k = \infty, \quad \sum_{k=2}^{\infty} p_k^2 < \infty.$$

Let again $\nu_n = \sum_{k=2}^{n} p_k = \mathbf{E}(S_n)$. As was pointed out in Deheuvels and Pfeifer (1988), assumption (5) is sufficient for Poisson approximation of G_n by Π_{ν_n} at a rate of

$$\sup_x |G_n(x) - \Pi_{\nu_n}(x)| = \frac{1}{2\sqrt{2\pi \mathrm{e}}} \frac{\sum_{k=2}^{n} p_k^2}{\sum_{k=2}^{n} p_k} + O\left(\left(\sum_{k=2}^{n} p_k\right)^{-3/2}\right) \tag{16}$$

which means that the exact approximation rate is of order $O\left(\left(\sum_{k=2}^{n} p_k\right)^{-1}\right) = O(1/\nu_n)$. We show here that this rate can be lowered to $O\left(\sum_{k=n+1}^{\infty} p_k^2 \Big/ \sum_{k=2}^{n} p_k\right)$ by convolution of Π_{ν_n} with the fixed (i.e. independent of n) signed measure with distribution function U having

$$u(z) = \prod_{k=2}^{\infty} \big(1 + p_k(z-1)\big) \mathrm{e}^{-p_k(z-1)} = 1 - \frac{1}{2} \sum_{k=2}^{\infty} p_k^2 (z-1)^2 + \sum_{k=3}^{\infty} a_k (-1)^k (z-1)^k \tag{17}$$

as generating function where the a_k are recursively defined as

$$a_k = -\frac{1}{k}\left(\tau_k + \sum_{i=2}^{k-2} a_i \tau_{k-i}\right), \quad k \geq 2; \quad a_2 = -\frac{\tau_2}{2} \tag{18}$$

where

$$\tau_i = \sum_{k=2}^{\infty} p_k^i, \quad i \geq 2. \tag{19}$$

This can be seen from e.g. Deheuvels, Pfeifer and Puri (1989), relation (2.11) there, or Shorgin (1977), Lemma 5. Note that by (5), the product converges with

$$u(z) \leq \prod_{k=2}^{\infty} \mathrm{e}^{p_k(z-1)} \mathrm{e}^{-p_k(z-1)} = 1.$$

The relationship between U and L above (see (7)) is as follows. In case of $p_k = 1/k,\ k = 2, \ldots, n$, we have

$$u(z) = \prod_{k=2}^{\infty}\left(1 + \frac{z-1}{k}\right)\mathrm{e}^{-(z-1)/k} = \frac{1}{\Gamma(1+z)}\mathrm{e}^{(1-\gamma)(z-1)} \qquad (20),$$

in view of Euler's product formula (14), while the generating function of L is just $1/\Gamma(1+z)$. W.r.t. relation (8), this means

$$u(z) = \frac{1}{\Gamma(1+z)}\mathrm{e}^{(1-\gamma)(z-1)} = (1+z)\exp\left((\gamma-1) - \sum_{k\geq 2}\frac{(-1)^k}{k}(\zeta(k)-1)z^k\right), \quad (21)$$

which indicates that the linear term in the exponent of (8) is just replaced by a constant term instead.

Theorem 2. *For $n > 1$,*

$$\Delta' = \sup_x |G_n(x) - \Pi_{\nu_n} * U(x)| = \frac{1}{2\sqrt{2\pi\mathrm{e}}}\frac{\sum_{k=n+1}^{\infty} p_k^2}{\sum_{k=2}^{n} p_k} + o\left(\sum_{k=n+1}^{\infty} p_k^2 \Big/ \sum_{k=2}^{n} p_k\right).$$

Proof. The easiest way to prove Theorem 2 is to apply the operator semigroup technique as in Deheuvels and Pfeifer (1988) or Deheuvels, Pfeifer and Puri (1989). It follows from there that the distance Δ' can be represented as a norm

$$\Delta' = \left\|\Big(\prod_{k=2}^{n}(I + p_k A) - T \circ \exp(\nu_n A)\Big)g\right\| \qquad (22)$$

where I is the identity operator and A is the difference operator (infinitesimal generator) acting on the sequence space ℓ^∞ endowed with the usual sup–norm $\|\cdot\|$, defined as

$$Af(n) = \begin{cases} f(n-1) - f(n), & n \geq 1 \\ -f(0), & n = 0, \end{cases} \qquad (23)$$

for any sequence $f = \big(f(0), f(1), \ldots,\big) \in \ell^\infty$, and $\circ$ means operator product (corresponding to convolution; $\{\mathrm{e}^{tA} \mid t \geq 0\}$ is the Poisson convolution semigroup generated by A). g is the particular sequence $g = (1, 1, 1, \ldots) \in \ell^\infty$. T denotes the operator

$$T = \prod_{k=2}^{\infty}(I + p_k A)\mathrm{e}^{-p_k A}, \qquad (24)$$

which corresponds to the signed measure with distribution function U. Let T_m denote the operator

$$T_m = \prod_{k=2}^{m}(I + p_k A)\mathrm{e}^{-p_k A}, \quad m \geq 1. \qquad (25)$$

Then

$$\Delta' = \left\|\Big(\prod_{k=2}^{n}(I + p_k A) - T \circ \exp(\nu_n A)\Big)g\right\| = \left\|\big(T - T_n\big)\mathrm{e}^{\nu_n A}g\right\|$$

where $T - T_n$ has a series expansion

$$T - T_n = -\frac{1}{2} \sum_{k=n+1}^{\infty} p_k^2 A^2 + \sum_{m=3}^{\infty} (a_m - a'_m)(-A)^m \tag{26}$$

where the a_m are as above (see relations (18) and (19)), and the a'_m are defined similarly as

$$a'_m = -\frac{1}{m}\Big(1 + \sum_{i=2}^{m-2} a'_i \tau'_{m-i}\Big), \quad m \geq 2; \quad a'_2 = -\frac{\tau'_2}{2} \tag{27}$$

where

$$\tau'_i = \sum_{k=2}^{n} p_k^i, \quad i \geq 2. \tag{28}$$

So the a'_m are just the finite counterparts of a_m, approximating a_m for $n \to \infty$ pointwise. It follows that the leading term in the expansion of Δ' is thus given by $\frac{1}{2}\sum_{k=n+1}^{\infty} p_k^2 \big\|A^2 \mathrm{e}^{\nu_n A} g\big\|$ with

$$\big\|A^2 \mathrm{e}^{\nu_n A} g\big\| \approx \frac{1}{\nu_n \sqrt{2\pi \mathrm{e}}}$$

(see Deheuvels and Pfeifer (1988) or Deheuvels, Pfeifer and Puri (1989)). The remainder term estimation for $a_m - a'_m$ and thus for the estimation of the consecutive terms in the expansion of $T - T_n$ can be carried out correspondingly, giving the result stated above.

Theorem 2 is proved.

Remark. The foregoing analysis might suggest that instead of the correcting signed measure corresponding to U one could likewise use the first two terms in the expansion of T, which correspond to the signed measure with distribution function $W = E_0 - 1/2\sum_{k=2}^{\infty} p_k^2 E_2$, resembling the distribution function V_n above. However, an analoguous argument as above shows that we would then only obtain

$$\Delta'' = \sup_x |G_n(x) - \Pi_{\nu_n} * W(x)| = O\left(\Big(\sum_{k=2}^{\infty} p_k\Big)^{-3/2}\right)$$

which is in general much worse than the r.h.s. of the estimate in Theorem 2. For instance, in the i.i.d. record model of Section 1, Theorem 2 gives the expansion

$$\Delta' = \frac{0.1995}{n \log n} + o\Big((n \log n)^{-1}\Big) \tag{29}$$

for large n.

References

[1] Abramowitz, M., and Stegun, I.A., eds. (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, Washington, D.C.

[2] Borovkov, K.A. (1988) Refinement of Poisson approximation. *Theory Probab. Appl.* **33**, 343 – 347.

[3] Borovkov, K., and Pfeifer, D. (1993) On record indices and record times. To appear in: *J. Stat. Plann. Inference.*

[4] Butzer, P.L., Hauss, M., and Schmidt, M. (1989) Factorial functions and Stirling numbers of fractional orders. *Results in Math.* **16**, 16 – 48.

[5] Charalambides, Ch.A., and Singh, J. (1988) A review of the Stirling numbers, their generalizations and statistical applications. *Comm. Statist. Theory Methods.* **17**, 2533 – 2592.

[6] Deheuvels, P., and Pfeifer, D. (1986) A semigroup approach to Poisson approximation. *Ann. Probab.* **14**, 665 – 678.

[7] Deheuvels, P., and Pfeifer, D. (1987) Semigroups and Poisson approximation. In: *New Perspectives in Theoretical and Applied Statistics.* Wiley, New York, 439 – 448.

[8] Deheuvels, P., and Pfeifer, D. (1988) On a relationship between Uspensky's theorem and Poisson approximations. *Ann. Inst. Statist. Math.* **40**, 671 – 681.

[9] Deheuvels, P., Pfeifer, D., and Puri, M.L. (1989) A new semigroup technique in Poisson approximation. *Semigroup Forum* **38**, 198 – 201.

[10] Dwass, M. (1960) Some k-sample rank order tests. In: *Contributions to Probability and Statistics: Essays in Honor of H. Hotelling.* Eds. I. Olkin et al. Stanford Studies in Math. & Statist. 2. Stanford Univ. Press, 198 – 202.

[11] Gradshteyn, I.S., and Ryzhik, I.M. (1980) *Tables of Series, Products, and Integrals.* Academic Press, New York.

[12] Johnson, N.L., and Kotz, S. (1969) *Discrete Distributions.* Wiley, Ney York.

[13] Jordan, C. (1965) *Calculus of Finite Differences.* 3rd ed. Chelsea, New York.

[14] Kruopis, J. (1986) Precision of approximation of the generalized binomial distribution by convolutions of Poisson measures. *Lithuanian Math. J.* **26**, 37 – 49.

[15] Kemp, R. (1984). *Fundamentals of the Average Case Analysis of Particular Algorithms.* Wiley–Teubner, N.Y.

[16] Nevzorov, V.B. (1988) Records. *Theory Probab. Appl.* **32**, 201 – 228.

[17] Pfeifer, D. (1989) Extremal processes, secretary problems and the 1/e–law. *J. Appl. Prob.* **26**, 722 – 733.

[18] Pfeifer, D. (1991) Some remarks on Nevzorov's record model. *Adv. Appl. Prob.* **23**, 823 – 834.

[19] Rényi, A. (1962) Théorie des éléments saillants d'une suite d'observations. Colloquium on Combinatorial Methods in Probability Theory, Mathematisk Institut, Aarhus Universitet, Denmark, 104 – 115.

[20] Ross, S. M. (1982) A simple heuristic approach to simplex efficiency. *Eur. J. Operat. Res.* **9**, 344 – 346.

[21] Shorgin, S.Y. (1977) Approximation of a generalized binomial distribution. *Theory Probab .Appl.* **22**, 846 – 850.

[22] Tzaregradskii, I.P. (1958) On a uniform approximation to the binomial distribution by infinitely divisible distributions. *Theory Probab. Appl.* **3**, 434 – 438.