LIETUVOS MATEMATIKOS RINKINYS. LMD DARBAI ISSN 0132-2818
51 tomas, 2010, 470–474 www.mii.lt/LMR/

On lower bounds for Poisson approximation to 2-runs statistic

Jūratė Petrauskienė, Vydas Čekanavičius

Vilnius University
Naugarduko 24, LT-03225, Vilnius
E-mail: vydas.cekanavicius@mif.vu.lt; jurate.petrauskiene@zebra.lt

Abstract. *Two*-runs statistic is approximated by various compound Poisson distributions and second order asymptotic expansions. Estimates of lower bounds are obtained for the uniform Kolmogorov and local metrics.

Keywords: 2-runs, compound Poisson distribution, m-dependent variables, uniform Kolmogorov norm, local norm.

Runs statistic is one of the most thoroughly studied examples when dealing with Poisson and compound Poisson approximations for sums of dependent indicators, see [2, 3, 6]; and the references therein. However, to the best of our knowledge, all known estimates are upper bound estimates. In this paper, we obtain lower bound estimates. We consider the simplest (and probably most popular) case of *two*-runs statistic. Let ξ_j, $j = 0, 1, 2, \dots, n$ be independent identically distributed indicator variables, $P(\xi_1 = 1) = p$, $P(\xi_1 = 0) = 1 - p$. Let $\eta_j = \xi_j\xi_{j-1}$, $S = \eta_1 + \eta_2 + \dots + \eta_n$, $\varphi_1(t) = \mathrm{E}\exp\{it\eta_1\}$ and

$$\varphi_k(t) = \frac{\mathrm{E}\exp\{itS_k\}}{\mathrm{E}\exp\{itS_{k-1}\}} \quad (k = 2, 3, \dots, n).$$

It is obvious, that η_j are 1-dependent random variables. We denote the distribution and characteristic function of S by F and $\hat{F}(t)$, respectively.

Let I_a denote the distribution concentrated at real a and set $I = I_0$. To make expressions shorter we also set $U = I_1 - I$, $z = \hat{U}(t) = e^{it} - 1$. In what follows, let V and M be two finite signed measures concentrated on integers $\mathbb{Z}$. Products and powers of V and M are understood in the convolution sense, i.e, $VM\{A\} = \sum_{k=-\infty}^{\infty} V\{A-k\}M\{k\}$ for a set $A \subseteq \mathbb{Z}$; further $M^0 = I$. The total variation norm, uniform Kolmogorov norm and the local norm of M are denoted by

$$\|M\| = \sum_{k=-\infty}^{\infty} \big|M\{k\}\big|, \qquad |M| = \sup_{k\in\mathbb{Z}} \big|M\big\{(-\infty, k]\big\}\big|, \qquad \|M\|_\infty = \sup_{k\in\mathbb{Z}} \big|M\{k\}\big|,$$

respectively. Let $\hat{M}(t)$ $(t \in \mathbb{R})$ be the Fourier transform of M. The exponential of M is given by

$$e^M = \exp\{M\} = \sum_{k=0}^{\infty} \frac{1}{k!} M^k, \quad \widehat{\exp\{M\}}(t) = \exp\big\{\hat{M}(t)\big\}.$$

Let us define measures used for approximations of F:

$$G_1 := \exp\{\gamma_1 U\},$$
$$G_2 := \exp\bigl\{\gamma_1 U + \gamma_2 U^2\bigr\}, \qquad G_3 := \exp\bigl\{\gamma_1 U + \gamma_2 U^2 + \gamma_3 U^3\bigr\}.$$

Here

$$\gamma_1 = np^2, \qquad \gamma_2 = \frac{np^3(2-3p) - 2p^3(1-p)}{2},$$
$$\gamma_3 = \frac{np^4(3-12p+10p^2) - 6p^4(1-p)(1-2p)}{3}.$$

We denote by C positive absolute constants. The letter θ stands for any complex or real number satisfying $|\theta| \leqslant 1$. The values of C, θ can vary from line to line, or even within the same line. Sometimes to avoid possible ambiguity, the C are supplied with indices.

We can formulate our results.

Theorem 1. *Let $p \leqslant 1/5$, $n \geqslant 3$. Then*

$$|F - G_1| \geqslant C_1 \min\bigl(np^3, p\bigr),$$
$$\|F - G_1\|_\infty \geqslant C_2 \min\left(np^3, \frac{1}{\sqrt{n}}\right).$$

Theorem 2. *Let $p \leqslant 1/5$, $n \geqslant 3$. Then*

$$\bigl|F - G_1\bigl(I + \gamma_2 U^2\bigr)\bigr| \geqslant C_3 \min\bigl(np^4, p^2\bigr),$$
$$\bigl\|F - G_1\bigl(I + \gamma_2 U^2\bigr)\bigr\|_\infty \geqslant C_4 \min\left(np^4, \frac{p}{\sqrt{n}}\right).$$

Theorem 3. *Let $p \leqslant 1/5$, $n \geqslant 3$. Then*

$$|F - G_2| \geqslant C_5 \min\left(np^4, \frac{p}{\sqrt{n}}\right),$$
$$\|F - G_2\|_\infty \geqslant C_6 \min\left(np^4, \frac{1}{n}\right).$$

For better understanding of the accuracy of our results we formulate some upper bound estimates. If $p \leqslant 1/5$, $n \geqslant 3$, then

$$\|F - G_1\| \leqslant C_7 \min\bigl(np^3, p\bigr), \tag{1}$$
$$\bigl\|F - G_1\bigl(I + \gamma_2 U^2\bigr)\bigr\| \leqslant C_8 \min\bigl(np^4, p^2\bigr), \tag{2}$$
$$\|F - G_2\| \leqslant C_9 \min\left(np^4, \frac{p}{\sqrt{n}}\right). \tag{3}$$

The estimate (1) is valid for all p and n and follows from Theorem 1 in [1]. The estimate (2) was proved in [5]. In principle, the estimate (3) with explicit constants and under weaker assumptions was obtained in [2]. However, in [2], the edge effect was not taken into account (it was assumed that η_1 depends on η_n). The estimate with the edge effect taken into account was obtained in [5]. Since $|M| \leqslant \|M\|$, we see, that, in general, upper bound estimates are of the right order. Similar conclusion can be drawn for the estimates in the local metric.

For the proof of Theorems we need auxiliary results. Lemmas 1 and 2 were proved in [5]. For the proof Heinrich's [4] method was used.

Lemma 1. *Let $p \leqslant 1/5$, $k = 3, 4, \ldots$. Then*

$$\ln \varphi_1 = p^2 z - \frac{p^4}{2} z^2 + \frac{p^6}{3} z^3 + C\theta p^5 |z|^4,$$
$$\ln \varphi_2 = p^2 z + \frac{p^3(2-3p)}{2} z^2 + \frac{p^5(7p-6)}{3} z^3 + C\theta p^5 |z|^4,$$
$$\ln \varphi_k = p^2 z - \frac{p^3(2-3p)}{2} z^2 + \frac{p^4(3-12p+10p^2)}{3} z^3 + C\theta p^5 |z|^4.$$

Lemma 2. *Let $p \leqslant 1/5$. Then, for all t,*

$$\big|\hat{G}_2(t)\big| \leqslant C \exp\bigg\{-\frac{6np^2}{5} \sin^2 \frac{t}{2}\bigg\}, \qquad \big|\hat{G}_3(t)\big| \leqslant C \exp\bigg\{-np^2 \sin^2 \frac{t}{2}\bigg\}. \tag{4}$$

Lemma 3. *Let M be concentrated on $\mathbb{Z}$, $\alpha \in \mathbb{R}$, $b \geqslant 1$. Then,*

$$|M| \geqslant C \bigg| \int_{-\infty}^{\infty} e^{-t^2/2} \hat{M}\bigg(\frac{t}{b}\bigg) e^{-it\alpha}\, dt \bigg|, \tag{5}$$
$$\|M\|_\infty \geqslant \frac{C}{b} \bigg| \int_{-\infty}^{\infty} e^{-t^2/2} \hat{M}\bigg(\frac{t}{b}\bigg) e^{-it\alpha}\, dt \bigg|. \tag{6}$$

The estimates (5) *and* (6) *remain valid if $e^{-t^2/2}$ is replaced by $te^{-t^2/2}$.*

Lemma's proof can be found in [6]. The assumption $p \leqslant 1/5$ is determined by the method of proof.

All proofs are similar. Therefore, in a more detailed way, we prove Theorem 3 only.

Proof of Theorem 3. We assumed that $p \leqslant 1/5$ and $n \geqslant 3$. Therefore,

$$|\gamma_3| = \gamma_3 \geqslant \frac{np^4(3-12p+10p^2)}{3} \geqslant \frac{np^4}{3}. \tag{7}$$

Applying Lemmas 1 and 2 we obtain

$$\big|\hat{F}(t) - \hat{G}_3(t)\big| \leqslant \big|\ln \hat{F}(t) - \ln \hat{G}_3(t)\big| \leqslant Cnp^5|z|^4 \leqslant Cnp^5|t|^4, \tag{8}$$

We have

$$\begin{aligned}\big|\hat{G}_3(t) - \hat{G}_2(t)\big(1+\gamma_3 z^3\big)\big| &= \big|\hat{G}_2(t)\big(\exp\big\{\gamma_3 z^3\big\} - 1 - \gamma_3 z^3\big)\big| \\ &= \bigg|\hat{G}_2(t)\gamma_3^2 z^6 \int_0^1 (1-\tau)\exp\big\{\tau\gamma_3 z^3\big\}\, d\tau\bigg| \\ &\leqslant C\gamma_3^2 |z|^6 \leqslant Cn^2p^8t^6,\end{aligned} \tag{9}$$

and

$$\hat F(t)-\hat G_2(t)=\hat G_2(t)\gamma_3(it)^3+\hat G_2(t)\gamma_3\big(z^3-(it)^3\big)\\ +\big(\hat G_3(t)-\hat G_2(t)\big(1+\gamma_3z^3\big)\big)+\big(\hat F(t)-\hat G_3(t)\big). \tag{10}$$

Let $b=h\max(1,\sqrt{n}p)$, $h\geqslant 1$. Then applying (7) and (10) we obtain

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} te^{-t^2/2}\big(\hat F(t)-\hat G_2(t)\big)\left(\frac{t}{b}\right)e^{-it\alpha}\,dt\right|\\
&\geqslant\left|\int_{-\infty}^{\infty} te^{-t^2/2}\hat G_2(t/b)e^{-it\alpha}\frac{\gamma_3t^3}{b^3}\,dt\right|\\
&\quad -C_{11}\left|\int_{-\infty}^{\infty} te^{-t^2/2}\left(\frac{|\gamma_3||t|^4}{b^4}+\frac{n^2p^8t^6}{b^6}+\frac{np^5|t|^4}{b^4}\right)dt\right|\\
&\geqslant\frac{|\gamma_3|}{b^3}\left|\int_{-\infty}^{\infty} t^4e^{-t^2/2}\,dt\right|-C_{12}\left|\int_{-\infty}^{\infty} t^2\frac{|\gamma_3|np^3t^6}{b^6}e^{-t^2/2}\,dt\right|\\
&\quad -C_{13}\left(\frac{np^4}{b^4}+\frac{n^2p^8}{b^6}+\frac{np^5}{b^4}\right)\\
&\geqslant C_{14}\frac{np^4}{b^3}-C_{15}\frac{n^2p^6}{b^6}-C_{16}\frac{np^4}{b^4}\\
&\geqslant C_{14}\frac{np^4}{b^3}-\frac{C_{17}np^4}{h^4\max(1,\sqrt{n}p)^3}\\
&\geqslant C_{14}\frac{np^4}{h^3\max(1,n^{3/2}p^3)}\left(1-\frac{C_{17}}{h}\right)\\
&\geqslant\frac{C_{14}}{h^3}\min\left(\frac{p}{\sqrt{n}},np^4\right)\left(1-\frac{C_{17}}{h}\right).
\end{aligned}$$

It suffices to take $h=2C_{17}$ and apply Lemma 3 with $\alpha=np^2$.

Proof of Theorem 1 is similar to the proof of Theorem 3. One needs to replace G_3, G_2 and γ_3z^3 by G_2, G_1 and γ_2z^2, respectively.

Proof of Theorem 2 is also similar to the proof of Theorem 3. One needs to replace G_3, G_2 and γ_3z^3 by G_2, G_1 and $1+\gamma_2z^2$. Instead (10) we have to use

$$\begin{aligned}
&\big(\hat F(t)-\hat G_1(t)\big(1+\gamma_2z^2\big)\big)\\
&\quad=\left(\hat G_1(t)\frac{(\gamma_2z^2)^2}{2}\right)-\left(\hat G_2(t)-\hat G_1(t)\left(1+\gamma_2z^2+\frac{(\gamma_2z^2)^2}{2}\right)\right)-\big(\hat F(t)-\hat G_2(t)\big).
\end{aligned}$$

References

[1] R. Arratia, L. Goldstein and L. Gordon. Poisson approximation and the Chen–Stein method. *Statist. Sci.*, **5**:403434, 1990.

[2] A.D. Barbour and A. Xia. Poisson perturbations. *ESAIM: Probab. Statist.*, **3**:131–150, 1999.

[3] T.C. Brown and A. Xia. Stein's method and birth-death processes. *Ann. Probab.*, **29**:1373–1403, 2001.

[4] L. Heinrich. A method for the derivation of limit theorems for sums of m-dependent random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **60**:501–515, 1982.

[5] J. Petrauskienė and V. Čekanavičius. Compound poisson approximations for sums of 1-dependent random variables I, 2010. Submited *Lith. Math. J.*

[6] X. Wang and A. Xia. On negative binomial approximation to k-runs. *J. Appl. Probab.*, **45**:456–471, 2008.

REZIUMĖ

Puasono aproksimacijų apatiniai rėžiai dviejų narių serijų statistikoms

J. Petrauskienė, V. Čekanavičius

Dviejų narių serijų statistika aproksimuojama įvairiais sudėtiniais Puasono skirstiniais ir trumpais asimptotiniais skleidiniais. Gauti tolygus Kolmogorovo ir lokalus aproksimacijos tikslumo iverčiai iš apačios.

Raktiniai žodžiai: m-priklausomi atsitiktiniai dydžiai, dviejų narių serijų statistika, sudėtinis Puasono skirstinys, tolygi Kolmogorovo metrika, lokali metrika.