

On lower bounds for Poisson approximation to 2-runs statistic

Jūratė Petrauskienė, Vydas Čekanavičius

Vilnius University

Naugarduko 24, LT-03225, Vilnius

E-mail: vydas.ceknavicius@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 $\eta_j = \xi_j \xi_{j-1}$, $S = \eta_1 + \eta_2 + \dots + \eta_n$, $\varphi_1(t) = E \exp\{it\eta_1\}$ and

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

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

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

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

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

Let us define measures used for approximations of F :

$$G_1 := \exp\{\gamma_1 U\},$$

$$G_2 := \exp\{\gamma_1 U + \gamma_2 U^2\}, \quad G_3 := \exp\{\gamma_1 U + \gamma_2 U^2 + \gamma_3 U^3\}.$$

Here

$$\gamma_1 = np^2, \quad \gamma_2 = \frac{np^3(2-3p) - 2p^3(1-p)}{2},$$

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

We denote by C positive absolute constants. The letter θ stands for any complex or real number satisfying $|\theta| \leq 1$. 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 $p \leq 1/5$, $n \geq 3$. Then*

$$|F - G_1| \geq C_1 \min(np^3, p),$$

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

Theorem 2. *Let $p \leq 1/5$, $n \geq 3$. Then*

$$|F - G_1(I + \gamma_2 U^2)| \geq C_3 \min(np^4, p^2),$$

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

Theorem 3. *Let $p \leq 1/5$, $n \geq 3$. Then*

$$|F - G_2| \geq C_5 \min\left(np^4, \frac{p}{\sqrt{n}}\right),$$

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

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

$$\|F - G_1\| \leq C_7 \min(np^3, p), \quad (1)$$

$$\|F - G_1(I + \gamma_2 U^2)\| \leq C_8 \min(np^4, p^2), \quad (2)$$

$$\|F - G_2\| \leq C_9 \min\left(np^4, \frac{p}{\sqrt{n}}\right). \quad (3)$$

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 $|M| \leq \|M\|$, 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 $p \leq 1/5$, $k = 3, 4, \dots$. Then*

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

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

$$|\hat{G}_2(t)| \leq C \exp\left\{-\frac{6np^2}{5} \sin^2 \frac{t}{2}\right\}, \quad |\hat{G}_3(t)| \leq C \exp\left\{-np^2 \sin^2 \frac{t}{2}\right\}. \quad (4)$$

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

$$|M| \geq C \left| \int_{-\infty}^{\infty} e^{-t^2/2} \hat{M}\left(\frac{t}{b}\right) e^{-it\alpha} dt \right|, \quad (5)$$

$$\|M\|_{\infty} \geq \frac{C}{b} \left| \int_{-\infty}^{\infty} e^{-t^2/2} \hat{M}\left(\frac{t}{b}\right) e^{-it\alpha} dt \right|. \quad (6)$$

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 $p \leq 1/5$ 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 $p \leq 1/5$ and $n \geq 3$. Therefore,

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

Applying Lemmas 1 and 2 we obtain

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

We have

$$\begin{aligned}|\hat{G}_3(t) - \hat{G}_2(t)(1 + \gamma_3 z^3)| &= |\hat{G}_2(t)(\exp\{\gamma_3 z^3\} - 1 - \gamma_3 z^3)| \\ &= \left| \hat{G}_2(t) \gamma_3^2 z^6 \int_0^1 (1-\tau) \exp\{\tau \gamma_3 z^3\} d\tau \right| \\ &\leq C \gamma_3^2 |z|^6 \leq C n^2 p^8 t^6, \quad (9)\end{aligned}$$

and

$$\begin{aligned}\hat{F}(t) - \hat{G}_2(t) &= \hat{G}_2(t)\gamma_3(it)^3 + \hat{G}_2(t)\gamma_3(z^3 - (it)^3) \\ &\quad + (\hat{G}_3(t) - \hat{G}_2(t)(1 + \gamma_3 z^3)) + (\hat{F}(t) - \hat{G}_3(t)).\end{aligned}\quad (10)$$

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

$$\begin{aligned}&\left| \int_{-\infty}^{\infty} te^{-t^2/2} (\hat{F}(t) - \hat{G}_2(t)) \left(\frac{t}{b}\right) e^{-it\alpha} dt \right| \\ &\geq \left| \int_{-\infty}^{\infty} te^{-t^2/2} \hat{G}_2(t/b) e^{-it\alpha} \frac{\gamma_3 t^3}{b^3} dt \right| \\ &\quad - C_{11} \left| \int_{-\infty}^{\infty} te^{-t^2/2} \left(\frac{|\gamma_3| |t|^4}{b^4} + \frac{n^2 p^8 t^6}{b^6} + \frac{np^5 |t|^4}{b^4} \right) dt \right| \\ &\geq \frac{|\gamma_3|}{b^3} \left| \int_{-\infty}^{\infty} t^4 e^{-t^2/2} dt \right| - C_{12} \left| \int_{-\infty}^{\infty} t^2 \frac{|\gamma_3| np^3 t^6}{b^6} e^{-t^2/2} dt \right| \\ &\quad - C_{13} \left(\frac{np^4}{b^4} + \frac{n^2 p^8}{b^6} + \frac{np^5}{b^4} \right) \\ &\geq C_{14} \frac{np^4}{b^3} - C_{15} \frac{n^2 p^6}{b^6} - C_{16} \frac{np^4}{b^4} \\ &\geq C_{14} \frac{np^4}{b^3} - \frac{C_{17} np^4}{h^4 \max(1, \sqrt{np})^3} \\ &\geq C_{14} \frac{np^4}{h^3 \max(1, n^{3/2} p^3)} \left(1 - \frac{C_{17}}{h} \right) \\ &\geq \frac{C_{14}}{h^3} \min \left(\frac{p}{\sqrt{n}}, np^4 \right) \left(1 - \frac{C_{17}}{h} \right).\end{aligned}$$

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 $\gamma_3 z^3$ by G_2 , G_1 and $\gamma_2 z^2$, respectively.

Proof of Theorem 2 is also similar to the proof of Theorem 3. One needs to replace G_3 , G_2 and $\gamma_3 z^3$ by G_2 , G_1 and $1 + \gamma_2 z^2$. Instead (10) we have to use

$$\begin{aligned}&(\hat{F}(t) - \hat{G}_1(t)(1 + \gamma_2 z^2)) \\ &= \left(\hat{G}_1(t) \frac{(\gamma_2 z^2)^2}{2} \right) - \left(\hat{G}_2(t) - \hat{G}_1(t) \left(1 + \gamma_2 z^2 + \frac{(\gamma_2 z^2)^2}{2} \right) \right) - (\hat{F}(t) - \hat{G}_2(t)).\end{aligned}$$

References

- [1] R. Arratia, L. Goldstein and L. Gordon. Poisson approximation and the Chen–Stein method. *Statist. Sci.*, **5**:403–434, 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. Submitted *Lith. Math. J.*
- [6] X. Wang and A. Xia. On negative binomial approximation to k -runs. *J. Appl. Probab.*, **45**:456–471, 2008.

REZIUOMĖ

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.