

## Poisson-type approximation for sums of 1-dependent indicators

Jūratė PETRAUSKIENĖ, Vydas ČEKANAVIČIUS

Vilnius University

Naugarduko 24, 03225 Vilnius, Lithuania

e-mail: jurate.petrauskiene@zebra.lt; vydas.cekanavicius@mif.vu.lt

**Abstract.** The sum of 1-dependent indicators is approximated by compound Poisson type distribution. Estimates are obtained for the uniform Kolmogorov and local metrics.

*Keywords:*  $m$ -dependent random variables, compound Poisson distribution, uniform Kolmogorov metric, local metric.

Let  $X_j, j = 1, 2, \dots, n$  be a triangular array of 1-dependent identically distributed three-point random variables,  $P(X_j = 1) = p_1, P(X_j = -1) = p_{-1}, P(X_j = 0) = 1 - p_1 - p_{-1}$ . We denote the distribution and characteristic function of  $S_n = X_1 + X_2 + \dots + X_n$  by  $F_n$  and  $\widehat{F}_n(t)$ , respectively. Let  $x = e^{it} - 1, x_{-1} = \bar{x} = e^{-it} - 1, p = p_{-1} + p_1,$

$$a(j_1, j_2) = P(X_1 = j_1, X_2 = j_2) - P(X_1 = j_1)P(X_2 = j_2),$$

$$a(j_1, j_2, j_3) = P(X_1 = j_1, X_2 = j_2, X_3 = j_3) - P(X_1 = j_1)P(X_2 = j_2)P(X_3 = j_3),$$

$$b_j = E(e^{itX_1} - 1) \dots (e^{itX_j} - 1),$$

$$A_j = \widehat{E}(e^{itX_1} - 1) \dots (e^{itX_j} - 1) := b_j - \sum_{k=1}^{j-1} A_k b_{j-k}, \quad A_1 = p_1 x + p_{-1} \bar{x},$$

$$R_1 = |a(-1, -1) - a(-1, 1) - a(1, -1) + a(1, 1)| + p|p_1 - p_{-1}|,$$

$$R_2 = |a(-1, 1) - 2a(1, 1) + a(1, -1)| + \sum_{j,k \in \{-1, 1\}} |a(j, k, -1) - a(j, k, 1)|,$$

$$R_3 = \sum_{j,k \in \{-1, 1\}} |a(j, k)| + p^2.$$

We use  $C_1, C_2, \dots$  to denote positive absolute constants and symbol  $\theta$  for all (possibly different) quantities, satisfying  $|\theta| \leq 1$ .

Our goal is to investigate the closeness of  $F_n$  to its accompanying compound Poisson law. More precisely, let  $D^n$  be a compound Poisson distribution with the following

characteristic function:

$$\widehat{D}^n(t) = \exp\{nA_1\} = \exp\{np_1x + np_{-1}\bar{x}\}.$$

The closeness is estimated in the uniform Kolmogorov and local metrics.

THEOREM 1. *Let*

$$\sum_{j,k \in \{-1,1\}} |a(j,k)|/p + 90\sqrt{p} \leq 1/3. \quad (1)$$

Then, for all  $n = 1, 2, \dots$ ,

$$\begin{aligned} \sup_x |F_n\{(-\infty, x]\} - D^n\{(-\infty, x]\}| &\leq C_1 n R_1 \min\left(1, \frac{1}{np}\right) \\ &+ C_2 n R_2 \min\left(1, \frac{1}{np\sqrt{np}}\right) + C_3 n R_3 \min\left(1, \frac{1}{(np)^2}\right) \end{aligned} \quad (2)$$

and

$$\begin{aligned} \sup_x |F_n\{x\} - D^n\{x\}| &\leq C_4 n R_1 \min\left(1, \frac{1}{np\sqrt{np}}\right) \\ &+ C_5 n R_2 \min\left(1, \frac{1}{(np)^2}\right) + C_6 n R_3 \min\left(1, \frac{1}{(np)^2\sqrt{np}}\right). \end{aligned} \quad (3)$$

#### Remarks

1. We are unaware about any Poisson-type approximation result for dependent random variable, where symmetry is taken into account.
2. Numerous Poisson-type approximations are obtained via the Stein method. However, the Stein method is applicable to non-negative random variables only. Thus, it can not be applied in our case.
3. Condition (1) is a technical one and quite probably can be improved. It is only marginally better than  $p = o(1)$ ,  $a(j,k) = o(p)$ . Formally, it allows for  $p$  to be a (very) small absolute constant.
4. If  $X_1, X_2, \dots, X_n$  are symmetric independent random variables, then the right-hand-side in (2) becomes  $C_7 n^{-1}$ . This is consistent with known facts about Poisson approximation to symmetric three point distributions, see [2].
5. It is not difficult to construct an example of dependent array, which satisfies (1). Let  $\xi_1, \xi_2, \dots$  be symmetric i.i.d. r.v., having distribution  $P(\xi_1 = 1) = P(\xi_1 = -1) = \alpha$ ,  $P(\xi_1 = 0) = 1 - 2\alpha$ . Let  $X_1 = \xi_1 \xi_2$ ,  $X_2 = \xi_2 \xi_3$ , etc. If  $\alpha = o(1)$ , then (1) holds and the accuracy of approximation in (2) is  $O(n\alpha^3 \wedge (n\alpha)^{-1})$ .

For the proof of Theorem we need auxiliary results. Let  $z = it$  and  $\varphi_1 = \varphi_1(z) = \mathbb{E}e^{zX_1}$ ,

$$\varphi_k = \frac{\mathbb{E}e^{zS_k}}{\mathbb{E}e^{zS_{k-1}}}, \quad w_n(it) := \sqrt{\mathbb{E}|e^{itX_j} - 1|^2}, \quad k = 2, 3, \dots, n.$$

For a sequence of arbitrary complex-valued random variables  $Y_1, Y_2, \dots$  let  $\widehat{E}Y_1 = EY_1$  and

$$\widehat{E}Y_1 Y_2 \dots Y_k := EY_1 Y_2 \dots Y_k - \sum_{j=1}^{k-1} \widehat{E}Y_1 \dots Y_j EY_{j+1} \dots Y_k, \quad k \geq 2.$$

The following lemma follows from Lemma 3.1 and Lemma 3.2 in [1].

LEMMA 1. *Let (1) hold. Then, for  $k = 1, 2, \dots, n$ , and all real  $t$*

$$\begin{aligned} \varphi_k &= Ee^{zX_k} + \sum_{j=1}^{k-1} \frac{\widehat{E}(e^{zX_j} - 1)(e^{zX_{j+1}} - 1) \dots (e^{zX_k} - 1)}{\varphi_j \varphi_{j+1} \dots \varphi_{k-1}}, \\ |\varphi_k - 1| &\leq |Ee^{zX_k} - 1| + 2\sqrt{E|e^{zX_{k-1}} - 1|^2 E|e^{zX_k} - 1|^2} / (1 - 4w_n(z)) \\ &\leq 13p|e^z - 1| \leq 1/5, \\ |\ln \widehat{F}_n(t) - nA_1| &\leq np \left( \frac{\sum_{j,k \in \{-1,1\}} |a(j,k)|}{p} + 90\sqrt{p} \right) |x|^2. \end{aligned}$$

LEMMA 2. *Let (1) be satisfied. Then, for all  $|t| \leq \pi$ ,*

$$\max \{ |\widehat{F}_n(t)|, |\widehat{D}^n(t)| \} \leq \exp \{ -C_8 np \sin^2(t/2) \}.$$

*Proof.* Note that

$$|\widehat{F}_n(t)| \leq |\exp\{nA_1\}| \exp \{ |\ln \widehat{F}_n(t) - nA_1| \}$$

and apply Lemma 1. The estimate for  $\widehat{D}^n(t)$  follows directly from its definition and (1).

LEMMA 3 ([4]). *Let  $Y_1, Y_2, \dots, Y_k$  be 1-dependent random variables with  $E|Y_j|^2 < \infty, j = 1, \dots, k$ . Then*

$$|\widehat{E}Y_1 Y_2 \dots Y_j| \leq 2^{j-1} \prod_{k=1}^j \sqrt{E|Y_k|^2}.$$

LEMMA 4. *Let condition (1) be satisfied. Then, for  $k \geq 7$ , we have*

$$\begin{aligned} \varphi_k - 1 &= A_1 + A_2 + A_3 - A_2 A_1 + A_4 - 2A_3 A_1 + A_2(A_1^2 - A_2) \\ &\quad + A_5 - 3A_4 A_1 + A_3(3A_1^2 - 3A_2) + 3A_2^2 A_1 \\ &\quad + A_6 - 4A_5 A_1 + A_3(12A_2 A_1 - 2A_3) + A_2(2A_2^2 - 4A_4) \\ &\quad + A_7 + A_3(10A_2^2 - 5A_4) - 5A_2 A_5 + C_9 \theta p^4 |x|^4. \end{aligned} \tag{4}$$

Moreover, for  $k = 2, 3, 4, 5, 6$  the estimate (4) holds with  $A_3 = A_4 = A_5 = A_6 = A_7 = 0$ ,  $A_4 = A_5 = A_6 = A_7 = 0$ ,  $A_5 = A_6 = A_7 = 0$ ,  $A_6 = A_7 = 0$ ,  $A_7 = 0$ , respectively.

*Proof.* Applying Lemma 1 we obtain

$$\begin{aligned} \varphi_k - 1 &= A_1 + \frac{A_2}{\varphi_{k-1}} + \frac{A_3}{\varphi_{k-2}\varphi_{k-1}} + \frac{A_4}{\varphi_{k-3}\varphi_{k-2}\varphi_{k-1}} + \frac{A_5}{\varphi_{k-4}\varphi_{k-3}\varphi_{k-2}\varphi_{k-1}} \\ &+ \frac{A_6}{\varphi_{k-5}\varphi_{k-4}\varphi_{k-3}\varphi_{k-2}\varphi_{k-1}} + \frac{A_7}{\varphi_{k-6}\varphi_{k-5}\varphi_{k-4}\varphi_{k-3}\varphi_{k-2}\varphi_{k-1}} \\ &+ \sum_{j=1}^{k-7} \frac{\widehat{E}(e^{zX_j} - 1)(e^{zX_{j+1}} - 1) \dots (e^{zX_k} - 1)}{\varphi_j \varphi_{j+1} \dots \varphi_{k-1}}. \end{aligned} \tag{5}$$

and

$$\frac{1}{|\varphi_k|} \leq \frac{1}{1 - |1 - \varphi_k|} \leq \frac{5}{4}.$$

From Lemma 3 we obtain

$$\begin{aligned} |\widehat{E}(e^{zX_j} - 1)(e^{zX_{j+1}} - 1) \dots (e^{zX_k} - 1)| &\leq 2^{k-j} \prod_{m=j}^k \sqrt{E|e^{zX_m} - 1|^2} \\ &= 2^{k-j} |x|^{k-j+1} p^{(k-j+1)/2} \leq C_{10} p^4 |x|^4 (4\sqrt{p})^{k-j-7}. \end{aligned}$$

Noting that, due to (1), we can assume  $p$  to be small. Therefore,

$$\begin{aligned} \sum_{j=1}^{k-7} \frac{\widehat{E}(e^{zX_j} - 1)(e^{zX_{j+1}} - 1) \dots (e^{zX_k} - 1)}{\varphi_j \varphi_{j+1} \dots \varphi_{k-1}} \\ \leq C_{11} p^4 |x|^4 \sum_{j=1}^{k-7} (5\sqrt{p})^{k-j-7} \leq C_{12} p^4 |x|^4. \end{aligned} \tag{6}$$

From (5) and (6) we get

$$\begin{aligned} \frac{1}{\varphi_k} &= 1 + (1 - \varphi_k) + (1 - \varphi_k)^2 + C_{13} \theta p^3 |x|^4 = 1 - A_1 - \frac{A_2}{\varphi_{k-1}} - \frac{A_3}{\varphi_{k-1}\varphi_{k-2}} \\ &- \frac{A_4}{\varphi_{k-1}\varphi_{k-2}\varphi_{k-3}} - \frac{A_5 x^5}{\varphi_{k-1}\varphi_{k-2}\varphi_{k-3}\varphi_{k-4}} + (1 - \varphi_k)^2 + C_{14} \theta p^3 |x|^4 \\ &= 1 - A_1 + A_1^2 - A_2 + 3A_2 A_1 - A_3 + 2A_2^2 + 4A_3 A_1 \\ &- A_4 + 5A_2 A_3 - A_5 + C_{15} \theta p^3 |x|^4. \end{aligned}$$

Putting the last expression into (5) we complete the proof of Lemma 4, for  $k \geq 7$ . The cases  $k = 2, 3, 4, 5, 6$  are proved similarly.

LEMMA 5. Let condition (1) be satisfied. Then, for  $k \geq 7$ ,

$$\ln \varphi_k = A_1 + C_{16}\theta(R_1|x|^2 + R_2|x|^3 + R_3|x|^4).$$

*Proof.* Applying Lemma 1 and Lemma 4, it is not difficult to show that

$$\ln \varphi = A_1 + C_{17}\theta\left(\sum_{j=2}^7 |A_j| + |A_1|^2 + p^2|x|^4\right).$$

Since  $x = -\bar{x} - |x|^2$  one can easily obtain the following estimates

$$A_1 = (p_1 - p_{-1})x - p_{-1}|x|^2,$$

$$A_1^2 = C_{17}\theta(p|p_1 - p_{-1}||x|^2 + p^2|x|^4),$$

$$|A_2| \leq |a(-1, -1) - a(-1, 1) - a(1, -1) + a(1, 1)| |x|^2 \\ + |a(-1, 1) + a(1, -1) - 2a(1, 1)| |x|^3 + |a(1, 1)| |x|^4,$$

$$|A_3| \leq \sum_{j,k \in \{-1,1\}} |a(j, k, -1) - a(j, k, 1)| |x|^3 + \sum_{j,k \in \{-1,1\}} |a(j, k, -1)| |x|^4 + |A_2|.$$

Let  $a(j_1, j_2, \dots, j_k) = P(X_1 = j_1, X_2 = j_2, \dots, X_k = j_k) - P(X_1 = j_1)P(X_2 = j_2) \dots P(X_k = j_k)$  and let  $\sum_k^*$  denote the sum over all  $j_1, \dots, j_k \in \{-1, 1\}$ . Then, for  $k = 4, 5, 6, 7$ ,

$$|A_k| \leq \sum_k^* |a(j_1, j_2, \dots, j_k)| |x|^k + C_{18} \sum_{m=2}^{k-1} |A_m|.$$

Lemma's statement now follows from the following estimate

$$|a(j_1, j_2, \dots, j_k)| \leq C_{19}|a(j_1, j_2)| + C_{20}p^2.$$

LEMMA 6. Let (1) be satisfied and let  $|t| \leq \pi$ . Then

$$|\widehat{F}_n(t) - \widehat{D}^n(t)| \leq C_{21} \exp\{-C_{22}npt^2\} [R_1|t|^2 + R_2|t|^3 + R_3|t|^4].$$

*Proof.* Applying Lemma 2 we get

$$|\widehat{F}_n(t) - \widehat{D}^n(t)| \leq C_{23} \exp\{-C_{22}npt^2\} |\ln \widehat{F}_n(t) - \ln \widehat{D}^n(t)|.$$

Note that  $\ln \widehat{F}_n(t) = \sum_{k=1}^n \ln \varphi_k$ . Consequently, from Lemma 5 we get the required estimate.

Theorems proof now follows from Lemma 6, Tsaregradskii's inequality

$$\sup_x |F_n\{(-\infty, x]\} - D^n\{(-\infty, x]\}| \leq \frac{1}{4} \int_{-\pi}^{\pi} \frac{|\widehat{F}_n(t) - \widehat{D}^n(t)|}{|t|} dt$$

and formula of inversion for the local probability

$$\sup_x |F_n\{x\} - D^n\{x\}| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |\widehat{F}_n(t) - \widehat{D}^n(t)| dt.$$

### References

1. L. Heinrich. A method for the derivation of limit theorems for sums of  $m$ -dependent random variables. *Prob. Related Fields*, 60(4):501–515, 1982.
2. E.L. Presman. Approximation of binomial distributions by infinitely divisible ones. *Theory Prob. Appl.*, 28:393–403, 1983.

### REZIUOMĖ

#### **J. Petrauskienė, V. Čekanavičius. 1-priklausomų indikatorių sumos Puasono tipo aproksimacija**

1-priklausomų indikatorių suma aproksimuojama sudėtinio Puasono tipo skirstiniu. Gauti tolygūs Kolmogorovo ir lokalus aproksimacijos tikslumo iverčiai.

*Raktiniai žodžiai:*  $m$ -priklausomi atsitiktiniai dydžiai, sudėtinis Puasono skirstinys, tolygi Kolmogorovo metrika, lokali metrika.