Liet. matem. rink, **47**, *spec. nr.*, 2007, 531–536

Normal approximation for sum of random number of summands

Leonas SAULIS, Dovilė DELTUVIENĖ (VGTU)
e-mail: lsaulis@fm.vtu.lt

Abstract. Normal aproximation of sum $Z_t = \sum_{i=1}^{N_t} X_i$ of i.i.d. random variables (r.v.) $X_i,\ i = 1, 2, \ldots$ with mean $\mathbf{E}X_i = \mu$ and variance $\mathbf{D}X_i = \sigma^2 > 0$ is analyzed taking into consideration large deviations. Here N_t is non-negative integer-valued random variable, which depends on t, but not depends at X_i, $i = 1, 2, \ldots$.

Keywords: distribution function, characteristic function, cumulant, large deviations, exponential inequality.

1. Introduction

Let $\{X_i,\ i = 1, 2, \ldots\}$ be a family of independent and identically distributed random variables (r.v.) with means $\mathbf{E}X_i = \mu$ and dispersions $\mathbf{D}X_i = \sigma^2 > 0$. Denote

$$Z_t = \sum_{i=1}^{N_t} X_i, \tag{1}$$

where integer-valued r.v. N_t, depending on parameter $t,\ t \geqslant 0$, is independent of X_i, $i = 1, 2, \ldots$.

For instance, in the continuous time dynamic model of insurance stock the surplus R_t at the moment t we can express by equation (see [2])

$$R_t = R_0 + P_t - Z_t, \quad t \geqslant 0.$$

Here R_0 – initial reserves, P_t – premiums obtained up to time t and r.v. Z_t defined by (1), where X_i is the ith claim and N_t denotes the number of claims by time t, express *the total claims in the time interval* $[0,\ t]$.

If N_t, $t \geqslant 0$ is a Poisson process with intensity λ, then Z_t defined by (1) is *compound Poisson process*. The mean and the variance of Z_t are

$$\mathbf{E}Z_t = \mu\lambda t, \quad \mathbf{D}Z_t = \big(\mu^2 + \sigma^2\big)\lambda t.$$

Analysis of the distribution of Z_t is of great significance not only in insurance and finance mathematics, but also in other fields of mathematics. This problem becomes more complicated if r.v. N_t is not a homogenious Poisson process.

Further, denote $\alpha_t = \mathbf{E}N_t, \quad \beta_t^2 = \mathbf{D}N_t$. Obviously,

$$\begin{aligned}\mathbf{E}Z_t &= \mathbf{E}X_1 \cdot \mathbf{E}N_t = \mu\alpha_t,\\ \mathbf{D}Z_t &= \mathbf{D}X_1\mathbf{E}N_t + (\mathbf{E}X_1)^2\mathbf{D}X_t = \sigma^2\alpha_t + \mu^2\beta_t^2.\end{aligned} \tag{2}$$

We denote

$$\widetilde{Z}_t = \left(\sqrt{\mathbf{D}Z_t}\right)^{-1}(Z_t - \mathbf{E}Z_t), \quad F_{\widetilde{Z}_t}(x) = \mathbf{P}\left(\widetilde{Z}_t < x\right),$$
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}y^2\right\} \mathrm{d}y. \tag{3}$$

The characteristic function and kth order cumulant of a r.v. X will be denote by

$$f_X(u) := \mathbf{E}\exp\left\{iuX\right\}, \quad -\infty < u < \infty,$$
$$\Gamma_k(X) := \frac{1}{i^k}\frac{\mathrm{d}^k}{\mathrm{d}u^k} \ln f_X(u)\bigg|_{u=0}, \quad k = 1, 2, \ldots . \tag{4}$$

Note that $\Gamma_1(X) = \mathbf{E}X$ and $\Gamma_2(X) = \mathbf{D}X$.

In what follows, we assume that the r.v. N_t satisfies the condition: exists constants $K_2 > 0$ and $q \geqslant 0$ such that

$$\left|\Gamma_l(N_t)\right| \leqslant \frac{1}{2}\, l!\, K_2^{l-2}\left(\beta_t^2\right)^{1+(l-2)q/(1+q)}, \quad l = 2, 3, \ldots . \tag{L}$$

It is easy to show, that by virtue of condition (L) with $0 \leqslant q < 1$ the cumulants $\Gamma_k\left(\widetilde{N}_t\right)$, $k = 3, 4, \ldots$ of the normalized r.v. $\widetilde{N}_t = \left(\sqrt{\mathbf{D}N_t}\right)^{-1} N_t$ decrease gradually if $\beta_t = \sqrt{\mathbf{D}N_t} \to \infty$.

In the case of homogenious Poisson process with intensity λ, $\Gamma_k(N_t) = \lambda t$, $k = 1, 2, \ldots$. So, condition (L) holds with $q = 0$ and $K_2 = 1$.

In this paper, the accuracy of approximation of distribution function $F_{\widetilde{Z}_t}(x)$ by a normal law $\Phi(x)$ is evaluated, a large deviation theorem is proved and exponential inequalities for $\mathbf{P}\left(\pm\widetilde{Z}_t \geqslant x\right)$ are obtained.

2. Large deviations theorems and exponential inequalities

All the results of this work have been obtained by the cumulant methods proposed by V. Statulevičius [4] and developed by R. Rudzkis, L. Saulis and V. Statulevičius in [3]. To use the lemmas obtained in these papers, we estimate the decreasing of the cumulant of r.v. $\widetilde{Z}_t$.

We assume that for i.i.d. variables X_i, $i = 1, 2, \ldots$, with mean $\mu = \mathbf{E}X_1$ and dispersion $\sigma^2 = \mathbf{D}X_1 > 0$ the following inequalities hold:

$$\left|\mathbf{E}X_1^k\right| \leqslant k!\, K^{k-2}\mathbf{E}X_1^2, \quad k = 3, 4, \ldots . \tag{B}$$

First, we estimate the kth order cumulant of the r.v. $Z_t = \sum_{i=1}^{N_t} X_i$. Denote $S_l = X_1 + \ldots + X_l$, $l \geqslant 1$. Since r.v. $X_1, X_2, \ldots$ and N_t are independent, and X_i, $i = 1, 2, \ldots$ are i.i.d., then

$$\mathbf{E}Z_t = \sum_l \mathbf{E}S_l\, \mathbf{P}(N_t = l) = \mathbf{E}X_1 \sum_l l\, \mathbf{P}(N_t = l) = \mathbf{E}X_1 \cdot \mathbf{E}N_t = \mu\alpha_t \tag{5}$$

and

$$\mathbf{E}Z_t^2 = \sum_l \mathbf{E}S_l^2\,\mathbf{P}(N_t = l)$$
$$= \sum_l \mathbf{E}\big(X_1^2 + \ldots + X_l^2 + \sum_{\substack{i,j;\\ i\neq j=\overline{1,l}}} X_i X_j\big)\mathbf{P}(N_t = l)$$
$$= \mathbf{E}X_1^2 \sum_l l\,\mathbf{P}(N_t = l) + \big(\mathbf{E}X_1\big)^2 \sum_l l(l-1)\,\mathbf{P}(N_t = l)$$
$$= \mathbf{E}X_1^2\mathbf{E}N_t + \big(\mathbf{E}X_1\big)^2\big(\mathbf{E}N_t^2 - \mathbf{E}N_t\big) = \mathbf{D}X_1\mathbf{E}N_t + \big(\mathbf{E}X_1\big)^2\mathbf{E}N_t^2.$$

In view of (5), we obtain

$$\mathbf{D}Z_t = \mathbf{D}X_1\mathbf{E}N_t + \big(\mathbf{E}X_1\big)^2\mathbf{D}N_t = \sigma^2\alpha_t + \mu^2\beta_t^2. \tag{6}$$

If N_t is homogenious Poisson process, then $\alpha_t = \mathbf{E}N_t = \mathbf{D}N_t = \beta_t^2 = \lambda t$. Then, we have

$$\mathbf{D}Z_t = \big(\mu^2 + \sigma^2\big)\lambda t, \quad \Gamma_k(Z_t) = \lambda t\mathbf{E}X_1^k, \quad k = 1, 2, \ldots . \tag{7}$$

The moments generating function of Z_t

$$\varphi_{Z_t}(z) := \mathbf{E}e^{zZ_t} = \sum_{l=1}^{\infty}\mathbf{E}e^{zS_l}\mathbf{P}\big(N_t = l\big) = \sum_{l=1}^{\infty}\big(\mathbf{E}e^{zX_1}\big)^l\mathbf{P}\big(N_t = l\big)$$
$$= \sum_{l=1}^{\infty}\exp\big\{l\cdot\ln\varphi_{X_1}(z)\big\}\mathbf{P}\big(N_t = l\big) = \mathbf{E}\exp\big\{N_t\ln\varphi_{X_1}(z)\big\}. \tag{8}$$

Let $\overline{X}_1 = X_1 - \mu$, and $\overline{Z}_t = Z_t - \mu\alpha_t$. It is clear that $\ln\varphi_{X_1}(z) = \mu z + \ln\varphi_{\overline{X}_1}(z)$. Note that in applications usually $\mu = \mathbf{E}X_i > 0$, $i = 1, 2, \ldots$. Therefore, we can assume that $\mu \neq 0$.

Let θ denote a quantity that does not exceed one in absolute value.

Using Lemma 3.1 [3] and condition (B), we get

$$\big|\Gamma_k(\overline{X}_1)\big| \leqslant k!M^{k-2}\sigma^2, \quad k = 3, 4, \ldots, \tag{9}$$

with $M = 2\max\big\{K,\, \sigma\big\} = 2\big(K \vee \sigma\big)$.

In view of (9), we obtain

$$\ln\varphi_{\overline{X}_1}(z) = \sum_{k=2}^{\infty}\frac{1}{k!}\,\Gamma_k\big(\overline{X}_1\big)z^k = \frac{1}{2}\sigma^2z^2 + \theta(\sigma z)^2\sum_{k=3}^{\infty}\big(M|z|\big)^{k-2}$$
$$= \frac{3}{2}\theta\big(\sigma|z|\big)^2, \quad |z| \leqslant (2M)^{-1}. \tag{10}$$

This equality, in view of (8) and condition (L), we get

$$\begin{aligned}
\ln\varphi_{\overline{Z}_t}(z) &= -\alpha_t\mu z + \ln\varphi_{Z_t}(z) = -\alpha_t\mu z + \ln \mathbf{E}\exp\bigl\{N_t\bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr)\bigr\} \\
&= -\alpha_t\mu z + \sum_{l=1}^{\infty}\frac{1}{l!}\,\Gamma_l(N_t)\bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr)^l \\
&= -\alpha_t\mu z + \alpha_t\bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr) \\
&\quad + \bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr)^2\sum_{l=2}^{\infty}\frac{1}{l!}\,\Gamma_l(N_t)\bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr)^{l-2} \\
&= \alpha_t\ln\varphi_{\overline{X}_1}(z) + \frac{1}{2}\beta_t^2\theta\bigl(\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr)^2\sum_{l=2}^{\infty}\Bigl(K_2\bigl|\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr|\beta_t^{\frac{2q}{1+q}}\Bigr)^{l-2} \\
&= \alpha_t\ln\varphi_{\overline{X}_1}(z) + \theta\bigl(1 + \bigl|(\mu z)^{-1}\ln\varphi_{\overline{X}_1}(z)\bigr|\bigr)^2\beta_t^2(\mu z)^2 \qquad (11)
\end{aligned}$$

for $K_2|\,\mu z + \ln\varphi_{\overline{X}_1}(z)|\,\beta_t^{2q(1+q)^{-1}} \leqslant 1/2$. Clearly that

$$\bigl|(\mu z)^{-1}\ln\varphi_{\overline{X}_1}(z)\bigr| \leqslant 1/5, \quad |z| \leqslant \bigl((1\vee|\mu|^{-1}\sigma)\,4M\bigr)^{-1},$$

and $K_2\bigl|\,\mu z + \ln\varphi_{\overline{X}_1}(z)\bigr|\,\beta_t^{2q(1+q)^{-1}} \leqslant 1/2$, $|z| \leqslant A$, where

$$A = \Bigl(\max\bigl\{(1\vee|\mu|^{-1}\sigma)4M,\ 3|\mu|K_2\,\beta_t^{2q/(1+q)}\bigr\}\Bigr)^{-1}. \qquad (12)$$

By virtue of (10)–(12), we obtain

$$\ln\varphi_{\overline{Z}_t}(z) = \theta\,\frac{3}{2}\bigl(\alpha_t\sigma^2 + \beta_t^2\mu^2\bigr)|z|^2, \quad |z| \leqslant A. \qquad (13)$$

Using the Cauchy formula, (13) implies that

$$\bigl|\Gamma_k\bigl(\overline{Z}_t\bigr)\bigr| \leqslant \frac{3}{2}\,k!\bigl(\alpha_t\sigma^2 + \beta_t^2\mu^2\bigr)\,A^{2-k}, \quad k = 3, 4, \ldots\ . \qquad (14)$$

Since $\widetilde{Z}_t$, defined by (3), we have $\Gamma_k\bigl(\widetilde{Z}_t\bigr) = \bigl(\sqrt{\mathbf{D}Z_t}\bigr)^{-k}\Gamma_k\bigl(\overline{Z}_t\bigr)$. There for

$$\Bigl|\Gamma_k\bigl(\widetilde{Z}_t\bigr)\Bigr| \leqslant \frac{(3/2)\,k!}{\Delta_t^{k-2}}, \quad k = 3, 4, \ldots, \qquad (15)$$

where

$$\Delta_t = A\sqrt{\mathbf{D}Z_t} = \frac{\bigl(\alpha_t\sigma^2 + \beta_t^2\mu^2\bigr)^{1/2}}{\max\bigl\{(1\vee|\mu|^{-1}\sigma)4M,\ 3|\mu|K_2\,\beta_t^{2q/(1+q)}\bigr\}}. \qquad (16)$$

It is evident, that

$$\Delta_t \geqslant C\bigl(\mu, \sigma, K, K_2\bigr)\,\beta_t\bigl(1\vee\ \beta_t^{2q/(1+q)}\bigr)^{-1}.$$

where $C(\mu, \sigma, K, K_2) = |\mu|\Big(\max\Big\{(1 \vee |\mu|^{-1}\sigma)4M,\ 3|\mu|K_2\Big\}\Big)^{-1}$. If $\beta_t \to \infty$, then $\Delta_t \geqslant C(\mu, \sigma, K, K_2) \cdot \beta_t^{(1-q)/(1+q)}$. Thus, $0 \leqslant q < 1$, $\Delta_t \to \infty$, of $\beta_t \to \infty$.

Using the estimate (15) of the kth order cumulant of r.v. $\widetilde{Z}_t$, we can get the results of the distribution function $F_{\widetilde{Z}_t}(x)$.

THEOREM 1. *Let i.i.d. random variables X_i, $i = 1, 2, \ldots$ with mean $\mu = \mathbf{E}X_i$ and variance $\sigma^2 = \mathbf{D}X_i$ satisfy condition (B), and non-negative integer-valued r.v. N_t with mean $\alpha_t = \mathbf{E}N_t$ and variance $\beta_t^2 = \mathbf{D}N_t$ satisfy condition (L). Random variables N_t and X_i, $i = 1, 2, \ldots$, are independent. Then, for the distribution function $F_{\widetilde{Z}_t}(x)$ of r.v. $\widetilde{Z}_t$, defined by* (3), *inequality*

$$\sup_x \big|F_{\widetilde{Z}_t}(x) - \Phi(x)\big| \leqslant \frac{7}{\Delta_t},$$

hold, where Δ_t defined by equality (16).

THEOREM 2. *Let r.v. X_i, $i = 0, 1, 2, \ldots$ satisfy the condition (B) and the r.v. N_t, is independent of X_i, satisfy the condition (L). Then for x, $x \geqslant 0$, $x = \mathrm{o}(\Delta_t^{1/3})$ the qualities*

$$\lim_{\beta_t \to \infty} \frac{1 - F_{\widetilde{Z}_t}(x)}{1 - \Phi(x)} = 1, \qquad \lim_{\beta_t \to \infty} \frac{F_{\widetilde{Z}_t}(-x)}{\Phi(-x)} = 1$$

holds.

THEOREM 3. *If the r.v. X_i, $i = 1, 2, \ldots$ and N_t satisfy the condition of the Theorem* 1, *then*

$$\mathbf{P}\big(\pm \widetilde{Z}_t \geqslant x\big) \leqslant \begin{cases} \exp\big\{-\frac{1}{12}x^2\big\}, & 0 \leqslant x \leqslant 3\Delta_t, \\ \exp\big\{-\frac{1}{4}(x\Delta_t)\big\}, & x \geqslant 3\Delta_t. \end{cases}$$

Proof of Theorem 1, on the strength of (16), follows from Corollary 2.1 in [3], using the estimate (15).

The proposition of the**Theorem 2** follows from the Lemma 2.3 at [3] on large deviations and kth order estimate $\Gamma_k(\widetilde{Z}_t)$ (15).

The exponential inequalities of probability $\mathbf{P}\big(\pm \widetilde{Z}_t \geqslant x\big)$ of the r.v. $\widetilde{Z}_t$, defined by (3), can be obtained using Lemma 2.4 at [3] or [1] and (15).

References

1. R. Bentkus, R. Rudzkis, On exponential estimates of the distribution of random variables, *Lith. Math. J.*, **20**, 15–30 (1980).
2. H. Pragarauskas, *Draudos matematika*, TEV, Vilnius (2007).
3. L. Saulis, V. Statulevičius, *Limit Theorems for Large Deviations*, Kluwer Academic Publisher, Dordrecht, Boston, London (1991).
4. V. Statulevičius, On large deviations, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **6**, 133–144 (1966).

REZIUMĖ

L. Saulis, D. Deltuvienė. Atsitiktinio dėmenų skaičiaus sumos skirstinio aproksimacija normaliuoju dėsniu

Darbe nagrinėjama nepriklausomų vienodai pasiskirsčiusių atsitiktinių dydžių X_i, $i = 1, 2, \ldots$ su vidurkiais $\mathbf{E}X_i = \mu$ ir dispersijomis $\mathbf{D}X_i = \sigma^2 > 0$ sumos $Z_t = \sum_{i=1}^{N_t} X_i$ skirstinio aproksimacija normaliuoju dėsniu didžiųjų nuokrypių zonoje. Laikoma, kad neneigiamas sveikareikšmis atsitiktinis dydis N_t, priklauso nuo parametro t, bet nepriklauso nuo at. d. X_i, $i = 1, 2, \ldots$.