Tr. J. of Mathematics
22 (1998) , 47 – 52.

A BERRY-ESSEEN BOUND FOR EMPTY BOXES STATISTIC ON THE SCHEME AN ALLOCATIONS OF SEVERAL TYPE BALLS*

S.A. Mirakhmedov & O. Saidova

Abstract

A Berry-Esseen bound for the number of empty cells in the scheme of independent and random allocation of balls of s type into different cells is obtained.

Key words and phrases: Central limit theorem, empty cells, random allocations.

Introduction

Let n_1 balls be of first type and n_2 be balls of a second type, etc., and n_s balls of sTh.type be distributed independently and randomly into N different cells, in such a way that each ball of ith type has probability p_{ik} of landing into kth cell, $p_{i1} + \cdots + p_{iN} = 1, \quad i = 1, \dots, s$. Let $\mu_0(s) = \mu_0(s, N, n_1, \dots, n_s)$ be a number of empty cells after all $n_1, \dots, n_s$ losses. If $s = 1$ we deal with multinomial scheme of an allocation and $\mu_0(1)$ is a well-known empty box test statistic (see, for example, Koichin, Sevastjanov, Chistjacov (1976)). For example, the random variable (r.v.) $\mu_0(s)$ used as test statistic for verification of homogeneity hypothesis.

Here we get a bound for remainder in the central limit theorem for $\mu_0(s)$. Our theorem generalizes the result of Quine and Robinson (1982).

Result. We consider the case that s is fixed and $N = N(n_1, \dots, n_s)$ is growing as one of $n_1, \dots, n_s$ increases. Suppose that for all $j = 1, \dots, s$ and $k = 1, \dots, N$

$$Np_{jk} \leq C_0 \quad \text{and} \quad n_i \leq \exp\{C_1 N\}. \tag{1}$$

Here and in what follows, $C_j, C_j(\cdot)$ are positive constants not dependent on $N, n_1, \dots, n_\epsilon$.

Denote

$$\lambda_{jm} = n_j p_{jm}, \;\; \lambda_m^{(\epsilon)} = \lambda_{1m} + \cdots \lambda_{sm}, \;\; \alpha_i = n_i/N,$$

* Research supported by TUBITAK, TURKEY. We are grateful to Hacettepe University, whose hospitality we enjoyed while working on part of this paper.

$$A_N(s)=\sum_{m=1}^{N}\exp\{-\lambda_m\},\quad a_e=\frac{1}{n_e}\sum_{m=1}^{N}\lambda_{em}\exp\{-\lambda_m\},$$

$$\sigma_N^2=\sum_{m=1}^{N}\left[\exp\{-\lambda_m\}(1-\exp\{-\lambda_m\})-\sum_{j=1}^{s}\alpha_j a_j^2\right],$$

$$\omega_N^{(s)}(x)=\left|P\left\{\frac{\mu_0(s)-A_N(s)}{\sigma_N(s)}<x\right\}-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{u^2}{2}\right\}\right|.$$

Theorem. *Under condition (1) for arbitrary $\nu\geq 0$ there exist $C(s,C_0,\nu)$ such that*

$$\omega_N^{(s)}(x)\leq\frac{C(s,C_0,\nu)}{1+|x|^{\nu}}\left[\frac{1}{\sigma_N(s)}+\sum_{j=1}^{s}\frac{1}{\sqrt{n_j}}\right].$$

Corollary 1. *Under condition (1) there exist $C(s)>0$ such that*

$$\sup_x\omega_N^{(s)}(x)\leq C(s)\left[\frac{1}{\sigma_N(s)}+\sum_{j=1}^{s}\frac{1}{\sqrt{n_j}}\right].$$

Corollary 2. *Suppose that $N,n_1,\ldots,n_s\to\infty$ in such a way that $\sigma_N(s)\to\infty$ and (1) is hold true. Then $\mu_0(s)$ is asymptotically normal.*

The result of Quine and Robinson (1982) is $\omega_N^{(1)}(x)\leq C\sigma_N^{-1}(1)$, which also follows from our theorem since $\sigma_N^2(1)\leq n_1$: but in the general case we have $\sigma_N^2(s)\leq n_1+\cdots+n_s$.

Proof. It is clear that

$$\mu_0(s)=\sum_{m=1}^{N}f(\eta_{1m},\ldots,\eta_{sm}),$$

where η_{ik} is a number of balls of ith type in mth cell after $n_1,\ldots,n_\epsilon$ tosses, and $f(0,\ldots,0)=1$ and $f(y_1,\ldots,y_s)=0$ if $y_i>0$ for some $i=1,\ldots,s$.

Let ζ_{jm} be a Poisson with parameter $\lambda_{jm},\zeta_m^{(s)}=(\zeta_{1m},\ldots,\zeta_{sm})$, then

$$g_m(\zeta_m^{(s)})=f(\zeta_m^{(s)})-\exp\{-\lambda_m\}+\sum_{i=1}^{s}a_i(\zeta_{im}-\lambda_{im}).$$

From Corollary 2 of Mirakhmedov (1987) we get

$$\omega_N^{(s)}(x)\leq\frac{C(s,k)}{1+|x|^k}\left[\beta_{3N}+\beta_{k+2,N}+\sum_{i=1}^{s}\frac{1}{\sqrt{n_i}}\right] \tag{2}$$

for any integer $k > 0$, where

$$\beta_{kN} = \frac{1}{\sigma_N^k(s)} \sum_{m=1}^{N} E|g_m(\zeta_m^{(s)})|^k.$$

We remark that

$$\sigma_N^2(s) = \sum_{m=1}^{N} Dg_m(\zeta_m^{(s)}).$$

We rewrite $\sigma_N^2(s)$ and $g_m(\zeta_m^{(s)})$ as follows:

$$\sigma_N^2(s) = \sum_{m=1}^{N} [1-(1+\lambda_m)\exp\{-\lambda_m\}]\exp\{-\lambda_m\} + \sum_{m=1}^{N}\sum_{j=1}^{s} \lambda_{jm}(\exp\{-\lambda_m\} - a_j)^2, \quad (3)$$

$$g_m(\zeta_m^{(s)}) = f(\zeta_m^{(s)}) + \exp\{-\lambda_m\}\sum_{i=1}^{s}\zeta_{im} - (1+\lambda_m)\exp\{-\lambda_m\} + \sum_{i=1}^{s}(a_i - \exp\{-\lambda_m\})(\zeta_{im} - \lambda_{im}).$$

Then for arbitrary $b > 1$ we get

$$\begin{aligned} E|g_m(\zeta_m^{(s)})|^b &\le 2^{b-1} E\left| f(\zeta_m^{(s)}) + \exp\{-\lambda_m\}\sum_{i=1}^{s}\zeta_{im} - (1+\lambda_m)\exp\{-\lambda_m\}\right|^b \\ &\quad + (2s)^{b-1}\sum_{i=1}^{s} |a_i - \exp\{-\lambda_m\}|^b \, E|\zeta_{im} - \lambda_m|^b \\ &\equiv 2^{b-1}\Delta_{1m} + (2s)^{b-1}\Delta_{2m}. \end{aligned} \quad (4)$$

The r.v. $f(\zeta_m^{(s)})$ has the same distribution as r.v. $\varphi(\zeta_{1m} + \cdots + \zeta_{sm})$ where $\varphi(0) = 1$ and $\varphi(x) = 0$ if $x > 0$. Thus

$$\begin{aligned} \Delta_{1m} &= E\left|\varphi(\zeta_{1m} + \cdots + \zeta_{sm}) + \exp\{-\lambda_m\}\sum_{i=1}^{s} - (1+\lambda_m)\exp\{-\lambda_m\}\right|^b \\ &\le \exp\{-\lambda_m\}(1 - \exp\{-\lambda_m\}(1+\lambda_m))^b + \lambda_m^{b+1}\exp\{-(b-1)\lambda_m\} \\ &\quad + \sum_{j=2}^{\infty} |j - 1 - \lambda_m|^b \exp\{-(b+1)\lambda_m\}\frac{\lambda_m^j}{j!} \equiv \Delta'_{1m} + \Delta''_{1m} + \Delta'''_{1m} \end{aligned} \quad (5)$$

because $\zeta_{1m} + \cdots + \zeta_{sm}$ is Poisson with parameter λ_m. Since $(1+u)e^{-u} < 1$ for $u > 0$ and (3), we have

$$\sum_{m=1}^{N} \Delta'_{1m} \le \sum \exp\{-\lambda_m\}(1 - \exp\{-\lambda_m\}(1+\lambda_m)) \le \sigma_N^2(s). \quad (6)$$

From $u^2e^{-2u} \leq 1$ and $\frac{1}{2}u^2e^{-u} \leq 1-(1+u)e^{-u}$, we get

$$\sum_{m=1}^{N} \Delta''_{1m} \leq \sum_{m=1}^{N} \lambda_m^2 \exp\{-2\lambda_m\} \leq 2\sum_{m=1}^{N}(1-(1+\lambda_m))\exp\{-\lambda_m\} \leq 2\sigma_N^2(s). \quad (7)$$

Let b be odd, $\sum_{\lambda_m \leq 1}$ and $\sum_{\lambda_m \geq 1}$ be a sum on m such that $\lambda_m \leq 1$ and $\lambda_m \geq 1$, correspondingly. We have

$$\begin{aligned}\sum_{\lambda_m \leq 1} \Delta'''_{1m} &= \sum_{\lambda_m \leq 1} \exp\{-b\lambda_m\}[E(\zeta_{1m}+\cdots+\zeta_{sm}-1-\lambda_m)^b \\ &\quad +(-1)^b(1+\lambda_m)^b\exp\{-\lambda_m\}] + (-1)_m^{b+1}\exp\{-\lambda_m\},\end{aligned}$$

if $\lambda_m \leq 1$ then

$$E(\zeta_{1m}+\cdots+\zeta_{sm}-1-\lambda_m)^b = \sum_{i=1}^{b} C_b^i(-1)^i E(\zeta_{1m}+\cdots+\zeta_{sm})^{b-1} \leq C(b)\lambda_m^2 - 1-(b-1)\lambda_m.$$

Therefore we get

$$\begin{aligned}\sum_{\lambda_m \leq 1} \Delta'''_{1m} &\leq \sum_{\lambda_m \leq 1} \exp\{-b\lambda_m\}(C(b)\lambda_m^2-(1+(b-1)\lambda_m))(1-(1+\lambda_m)\exp\{-\lambda_m\}) \\ &\leq C(b)\sum_{m=1}^{N}\lambda_m^2\exp\{-2\lambda_m\} \leq C(b)\sigma_N^2(s). \quad (8)\end{aligned}$$

Since $(1+\lambda_m)^b \leq 1+(b-1)\lambda_m + C(b)\lambda_m^2$, if $\lambda_m \leq 1$. Using well known inequality between moments of r.v., we obtain:

$$\begin{aligned}\sum_{\lambda_m \geq 1} \Delta'''_{1m} &\leq \sum_{\lambda_m \geq 1} \exp\{-b\lambda_m\}E|\zeta_{1m}+\cdots+\zeta_{sm}|^b \\ &\leq \sum_{\lambda_m \geq 1}(E(\zeta_{1m}+\cdots+\zeta_{sm}-1-\lambda_m)^{b+1})^{b/b+1}\exp\{-b\lambda_m\} \\ &\leq C(b)\sum_{\lambda_m \geq 1}\lambda_m^{b/2}\exp\{-b\lambda_m\} \leq C(b)\sum_{\lambda_m \geq 1}\lambda_m^2\exp\{-2\lambda_m\} \leq C(b)\sigma_N^2(s). (9)\end{aligned}$$

From (5), (6), (7), (8) and (9) follows

$$\sum_{m=1}^{N} \Delta_{1m} \leq C(b)\sigma_N^2(s). \quad (10)$$

Let us estimate $\sum_{m=1}^{N} \Delta_{2m}$. We have

$$\sum_{m=1}^{N} |a_k \quad - \quad \exp\{-\lambda_m\}|^b\, E|\zeta_{km}-\lambda_{km}|^b \leq \sum_{\lambda_{km} \leq 1}(a_k - \exp\{-\lambda_m\})^b E|\zeta_{km}-\lambda_m|^b +$$

$$\sum_{\lambda_{km}\geq 1}(a_k - \lambda_{km})^b\, E[\zeta_{km}-\lambda_{km})^{b+1}]^{b/b+1} \leq C(b)\left[\sum_{\lambda_{km}\leq 1}(a_k-\exp\{-\lambda_m\})^2\lambda_{km} + \right.$$

$$\left.\sum_{\lambda_{km}>1,\alpha_k<1}(a_k - \exp\{-\lambda_m\})^2\lambda_{km}^{b/2} + \sum_{\lambda_{km}>1,\alpha_k\geq 1}|a_k-\exp\{-\lambda_m\}|^b\,\lambda_{km}^{b/2}\right] \leq C(b)\sigma_N^2(s) +$$

$$\sum_{\lambda_{km}>1,\alpha_k>1}|a_k - \exp\{-\lambda_m\}|^b\,\lambda_{km}^{b/2}. \tag{11}$$

Here, we used that $\lambda_{km}\leq C_0\alpha_k,\ldots,a_k\leq 1$ and $\lambda_{km}\leq C_0$ if $\alpha_k\leq 1$, $E|\zeta_{km}-\lambda_{km}|^i\leq C(i)\lambda_{km}$.

Let $\lambda_{km}>1$, $\alpha_k>1$. Since $a_k\leq\alpha_k^{-1}$ we have

$$|a_k-\exp\{-\lambda_m\}|\lambda_m^{1/2}\leq a_k\lambda_m^{1/2}+1\leq\sqrt{C_0/\alpha_k}+1\leq\sqrt{C_0}+1.$$

Therefore

$$\sum_{\lambda_{km}>1,\alpha_k>1}(|a_k-\exp\{-\lambda_m\}|\lambda_{km}^{1/2})^b\leq(\sqrt{C_0}+1)^{b-2}\Sigma(a_k-\exp\{-\lambda_m\})^2\lambda_{km}$$
$$\leq(\sqrt{C_0}+1)^{b-2}\sigma_N^2(s).$$

From this and (11)

$$\sum_{m=1}^{N}\Delta_{2m}\leq C(b)\sigma_N^2(s). \tag{12}$$

Thus if b is odd, then by (4), (10), (11) it follows that

$$\sum_{m=1}^{N}E|g_m(\zeta_m^{(s)})|^b\leq C(b)\sigma_N^2(s). \tag{13}$$

If b is odd then the theorem follows from (2) and (13). If b is even then the theorem follows from the well-known inequality between Ljapunov's ratio and (2), (13). Proof of theorem is complete. □

References

[1] Mirakhmedov, S.A., An approximations of multiple randomized divisible statistics by means of Normal distribution, *Theory Probabl. and Appl.*, **32** 761-771 (1987).

[2] Quine, M.P., Robinson, J., A Berry-Esseen bound for an occupancy problem, *Ann. Probabl.*, **10**, 663-671, 1982.

[3] Kolchin, V.F., Sevastjanov, B.V., Chistjakov, V.P., Random allocations, Nauka, Moskov, 1976.

Serzod A. MIRAKHMEDOV & Odila SAIDOVA
Department of Mathematics
Tashkent State University
700095, Tashkent - UZBEKISTAN

Received 13.05.1996