Jagdish Saran; Kanwar Sen

Some distribution results on generalized ballot problems

Aplikace matematiky, Vol. 30 (1985), No. 3, 157–165

Persistent URL: http://dml.cz/dmlcz/104138

SOME DISTRIBUTION RESULTS ON GENERALIZED BALLOT PROBLEMS

JAGDISH SARAN, KANWAR SEN

(Received January 4, 1983)

1. INTRODUCTION

Suppose that in a ballot candidate A scores a votes and candidate B scores b votes. The votes are drawn one at a time and the $\binom{a+b}{a}$ different possible orders of drawing are assigned equal probabilities. Denote by α_r and β_r the number of votes registered for A and B respectively among the first r votes counted; $r = 1, 2, \ldots, a + b$ (notice that α_r and β_r are random variables and $\alpha_r + \beta_r = r$). Let c be a non-negative integer and $a \geqq \mu b$ where $\mu \geqq 0$. Denote by P_j the probability that the inequality $\alpha_r > \mu\beta_r$ holds for exactly j values among $r = 1, 2, \ldots, a + b$, by P_j^* the probability that $\alpha_r \geqq \mu\beta_r$ holds for exactly j values among $r = 1, 2, \ldots, a + b$, and by P_j^{**} the probability that $\alpha_r > \mu\beta_r - c$ holds for exactly j values among $r = 1, 2, \ldots, a + b$.

The ballot problem was first formulated by Bertrand [4] in 1887. He discovered that

$$P_{a+b} = \frac{a-b}{a+b} \tag{1}$$

if $\mu = 1$ and this was proved in the same year by André [2]. Also in 1887 Barbier [3] found that

$$P_{a+b} = \frac{a-\mu b}{a+b} \tag{2}$$

if $\mu(\geqq 0)$ is an integer and this was proved by Aeppli [1] in 1924. Similarly we have

$$P_{a+b}^* = \frac{a+1-b}{a+1} \tag{3}$$

if $\mu = 1$ and

$$P_{a+b}^* = \frac{a+1-\mu b}{a+1} \tag{4}$$

if $\mu(\geqq 0)$ is an integer. Formulae (3) and (4) are simple consequences of (1) and (2) respectively.

In later years several authors, viz. Dvoretzky-Motzkin [8], Grossman [12], Takács [14] and Mohanty-Narayana [13] have given the alternative proofs of the ballot problem.

In 1962 Takács [14] proved that for an arbitrary $\mu \geqq 0$,

$$(5) \qquad P_{a+b} = \frac{a}{a+b} \sum_{j=0}^{b} C_j \frac{\binom{b}{j}}{\binom{a+b-1}{j}}$$

where $C_0 = 1$ and the constants C_j $(j = 1, 2, \ldots, b)$ are given by the following recurrence formula:

$$(6) \qquad \sum_{j=0}^{k} C_j \frac{\binom{k}{j}}{\binom{[k\mu]+k-1}{j}} = 0 \quad (k = 1, 2, \ldots)$$

where $[k\mu]$ is the greatest integer $\leqq k\mu$. If μ is an integer, then $C_j = -\mu$ $(j = 1, 2, \ldots$ $\ldots, b)$ and (5) reduces to (2). Also in 1962 Takács [15] proved that

$$(7) \qquad P_j^* = \frac{1}{a+b} \quad (j = 1, 2, \ldots, a+b)$$

if a and b are relatively prime numbers and $\mu = a/b$.

It is interesting to mention here that Chung and Feller [7] have proved an analogue to (7). They have also found the distribution of the number of subscripts for which either

$$\alpha_r > \beta_r \quad \text{or} \quad \alpha_r = \beta_r \quad \text{but} \quad \alpha_{r-1} > \beta_{r-1}, \quad r = 1, 2, \ldots, a+b\,.$$

In 1963 Takács [16] derived the complete probability distribution $\{P_j\}$ of the number of strict lead positions, provided $a > \mu b$, where μ is a non-negative integer. He has also obtained an expression for P_{a+b}^{**} and deduced that for $\mu = 1$,

$$(8) \qquad P_{a+b}^{**} = P\{\alpha_r > \beta_r - c \quad \text{for} \quad r = 1, 2, \ldots, a+b\} = 1 - \frac{\binom{a+b}{a+c}}{\binom{a+b}{a}}\,.$$

The result (8) can directly be derived by using random walk model (see [18], p. 212).

In 1964 Engelberg [9] has derived the complete probability distribution of the number of strict and weak lead positions, viz. $\{P_j\}$ and $\{P_j^*\}$, for $\mu = 1$, in an 'unconditional' ballot problem (where only the total number of votes L is given and all 2^L possible arrangements of counting are equally probable).

In 1965 Engelberg [10] has determined the complete probability distribution $\{P_j^*\}$ of the number of weak lead positions, provided $a \geqq \mu b$, where μ is a positive integer. She has also obtained the distribution $\{P_j\}$ for $\mu = 1$ and $a = b$. In the general case both these complete probability distributions $\{P_j\}$ and $\{P_j^*\}$ have been determined by Takács [17].

Bizley [5] derived formulae for P_0 and P_{a+b-1} in case $\mu = a/b$. Bizley [6] also made a conjecture concerning the complete probability distribution P_j for $j = 0, 1, \ldots, a + b$ in the case $\mu = a/b$. Bizley's [6] formula was proved by Takács [19].

In this paper we shall give the probability distributions of the following random variables.

(i) $\delta_{a,b}^{(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$.

(ii) $\delta_{a,b}^{-(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c - 1$,

(iii) $\delta_{a,b}^{+(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c + 1$,

(iv) $\Delta_{a,b}^{-(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c - 1$ and $\alpha_{r+1} = \beta_{r+1} - c + 1$,

(v) $\Delta_{a,b}^{+(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c + 1$ and $\alpha_{r+1} = \beta_{r+1} - c - 1$,

(vi) $\sigma_{a,b}^{-(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c - 1$ and $\alpha_{r+1} = \beta_{r+1} - c - 1$,

(vii) $\sigma_{a,b}^{+(-c)}$ number of subscripts $r = 1, 2, \ldots, a + b$ for which $\alpha_r = \beta_r - c$ but $\alpha_{r-1} = \beta_{r-1} - c + 1$ and $\alpha_{r+1} = \beta_{r+1} - c + 1$,

where c is a non-negative integer. Similarly, we define the random variables $\delta_{a,b}^{(c)}$, $\delta_{a,b}^{-(c)}$, $\delta_{a,b}^{+(c)}$, $\Delta_{a,b}^{-(c)}$, $\Delta_{a,b}^{+(c)}$, $\sigma_{a,b}^{-(c)}$ and $\sigma_{a,b}^{+(c)}$ respectively by replacing c by $-c$ in the above definitions.

2. THE METHOD

Consider a sequence of auxiliary random variables $\theta_1, \theta_2, \ldots, \theta_{a+b}$ such that θ_k $(k = 1, 2, \ldots, a + b)$ can acquire only one of the values $+1$ or -1 in accordance with the rule:

$$\theta_k = \begin{cases} +1 & \text{if the } k\text{-th vote is cast for } A\,, \\ -1 & \text{if the } k\text{-th vote is cast for } B\,, \end{cases}$$

and let $S_0 = 0$, $S_k = \theta_1 + \theta_2 + \ldots + \theta_k$, Then S_k denotes the position at the moment $x = k$ of a moving particle performing a random walk on the y-axis. Thus each vote sequence represents a path from $(0, 0)$ to $(a + b, a - b)$. We assume that all $\binom{a+b}{a}$ possible different paths are equally probable. For convenience of writting we introduce the following symbols:

$F_{a,b}$	a path from $(0, 0)$ to $(a + b, a - b)$.
$R^{(t)}$ point	a point (k, S_k) for which $S_k = t$.
$R_+^{(t)}(R_-^{(t)})$	an $R^{(t)}$ point (k, S_k) such that $S_{k-1} = t + 1$ $(S_{k-1} = t - 1)$.
$W_-^{(t)}$	the segment of a path included between two consecutive $R^{(t)}$ points with $S_i < t$ at the intervening positions.
$F_{a,b,t}^j$	an $F_{a,b}$ with j $R^{(t)}$ points.
$F_{a,b,t}^{j-}$	an $F_{a,b}$ with j $R_-^{(t)}$ points
H_m^n	a path from $(0, 0)$ to (m, n) reaching the height n for the first time at the m-th step.
$N[\ldots]$	number of paths of the type ..., e.g., $N[H_m^n] = \frac{n}{m}\binom{m}{\frac{1}{2}(m+n)}$ (by [11, p. 89]).

3. DISTRIBUTION RESULTS

Theorem 1. *For* $a \geqq b - c, c > 0,$

$$(9) \quad \binom{a+b}{a} P\{\delta_{a,b}^{(-c)} = j\} = 2^{j-1} \frac{a - b + j + 2c - 1}{a + b - j + 1} \binom{a + b - j + 1}{a + c}.$$

Proof. To establish (9), consider a vote sequence with $\delta_{a,b}^{(-c)} = j$. The corresponding path is an $F_{a,b,-c}^j$. Let $OP_1P_2 \ldots P_jQ$ (Fig. 1) be an $F_{a,b,-c}^j$ path with j $R^{(-c)}$ points at $P_1(i_1, -c)$, $P_2(i_2, -c), \ldots,$ and $P_j(i_j, -c)$, say respectively; $0 < i_1 < i_2 < \ldots < i_j < a + b$. Consider this path as divided into three segments, viz. OP_1, P_1P_j, and P_jQ. Reflecting OP_1 about origin, i.e., replacing $(\theta_1, \theta_2, \ldots, \theta_{i_1})$ by $(-\theta_1, -\theta_2, \ldots, -\theta_{i_1})$ we get an $H_{i_1}^c$ path. In the segment P_1P_j change the signs of those θ's which lie above the line $y = -c$. Now P_1P_j consists of $(j - 1)$ $W_-^{(-c)}$. Removing $\theta_{i_1+1}, \theta_{i_2+1}, \ldots,$ and $\theta_{i_{j-1}+1}$ from the beginning of each $W_-^{(-c)}$ and then joining the remaining segments end-to-end, in order, the segment P_1P_j thus transforms to an $H_{i_j-i_1-j+1}^{j-1}$ path. Reversing the order of the last segment P_jQ, i.e., replacing $(\theta_{i_j+1}, \theta_{i_j+2}, \ldots, \theta_{a+b})$ by $(\theta_{a+b}, \theta_{a+b-1}, \ldots, \theta_{i_j+1})$ yields an $H_{a+b-i_j}^{a-b+c}$ path. On joining these three transformed segments end-to-end, in order, we get finally an $H_{a+b-j+1}^{a-b+2c+j-1}$ path.

Since any one of the $(j-1)$ $W_{-}^{(-c)}$ can be reflected in the line $y = -c$ to give an $F^{j}_{a,b,-c}$ path, there will be 2^{j-1} paths of the type $F^{j}_{a,b,-c}$ corresponding to a single $H^{a-b+2c+j-1}_{a+b-j+1}$ path. Thus the probability we seek for is

$$P\{\delta^{(-c)}_{a,b} = j\} = \binom{a+b}{a}^{-1} 2^{j-1} N[H^{a-b+2c+j-1}_{a+b-j+1}],$$

leading to (9) by [11, p. 89].

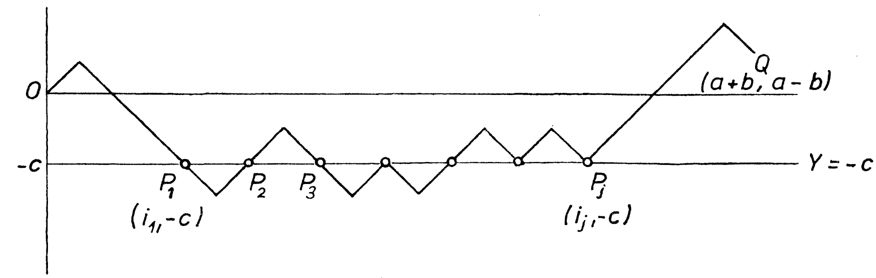

Fig. 1.

Theorem 2. *For $a \geqq b - c$, $c > 0$,*

$$(10) \qquad \binom{a+b}{a} P\{\delta^{-(-c)}_{a,b} \geqq j\} = \binom{a+b}{a+c+j}.$$

Proof. To derive (10), consider a vote sequence with $\delta^{-(-c)}_{a,b} = k$. The corresponding path is an $F^{k-}_{a,b,-c}$ path which can be shown to be in one-to-one correspondence with an $H^{a-b+2c+2k+1}_{a+b+1}$ path. Let $OPP_1P_2 \ldots P_kQ$ (fig. 2) be an $F^{k-}_{a,b,-c}$ path with

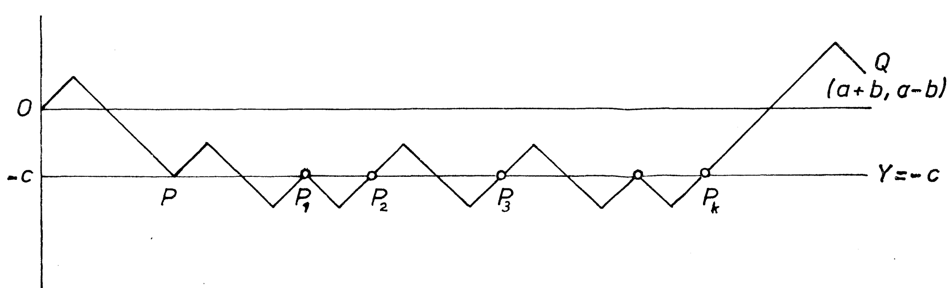

Fig. 2.

k $R_{-}^{(-c)}$ points at $P_1(i_1, -c)$, $P_2(i_2, -c), \ldots$, and P_k $(i_k, -c)$, say. Let P be the point (i, S_i) such that $S_i = -c$ for the first time; $0 < i < i_1 < i_2 < \ldots < i_k < a + b$. Consider this path as divided into three segments, viz. OP, PP_k, and P_kQ. Reflecting OP about origin, i.e., replacing $(\theta_1, \theta_2, \ldots, \theta_i)$ by $(-\theta_1, -\theta_2, \ldots, -\theta_i)$ we get an

H_i^c path. In the segment PP_k change the signs of those θ's which lie above the line $y = -c$. Now remove θ_{i+1}, θ_{i_1+1}, $\theta_{i_2+1}, \ldots,$ and $\theta_{i_{k-1}+1}$. Add $(+1)$ after each $\theta_{i_1}, \theta_{i_2}, \ldots,$ and θ_{i_k}. The segment PP_k thus transforms to an $H_{i_k-i}^{2k}$ path. Reversing the order of the last segment P_kQ, i.e., replacing $(\theta_{i_k+1}, \theta_{i_k+2}, \ldots, \theta_{a+b})$ by $(\theta_{a+b}, \theta_{a+b-1}, \ldots, \theta_{i_k+1})$ and adding $(+1)$ at the end we get an $H_{a+b+1-i_k}^{a-b+c+1}$ path. On joining these three transformed segments end-to-end, in order, we get finally an $H_{a+b+1}^{a-b+2c+2k+1}$ path. By reversing the above procedure it may be seen that this transformation is one-to-one. Thus the probability we seek for is

$$(11) \qquad P\{\delta_{a,b}^{-(-c)} = k\} = \binom{a+b}{a}^{-1} N[H_{a+b+1}^{a-b+2c+2k+1}]$$

$$= \binom{a+b}{a}^{-1} \left[\binom{a+b}{a+c+k} - \binom{a+b}{a+c+k+1}\right], \quad \text{(by [11, p. 89])}$$

whence (10) follows by using the equality

$$P\{\delta_{a,b}^{-(-c)} \geqq j\} = \sum_{k=j}^{b-c} P\{\delta_{a,b}^{-(-c)} = k\} .$$

Results (3) and (8) are now simple consequences of (10) as shown below.

Setting $k = 0$ in (11), replacing c by t and then summing the resulting expression over t from 0 to $c - 1$, we get the probability of the event that $\alpha_r > \beta_r - c$ for $r = 1, 2, \ldots, a + b$. The corresponding path from $(0, 0)$ to $(a + b, a - b)$ will be such that it does not touch the line $y = -c$. Thus

$$P\{\alpha_r > \beta_r - c \quad \text{for} \quad r = 1, \ldots, a + b\} = \sum_{t=0}^{c-1} \left[\binom{a+b}{a+t} - \binom{a+b}{a+t+1}\right]\binom{a+b}{a}^{-1},$$

which is in agreement with (8).

Setting $k = c = 0$ in (11), we get:

$$P\{\alpha_r \geqq \beta_r \quad \text{for} \quad r = 1, \ldots, a + b\} = \frac{a-b+1}{a+1},$$

which is in agreement with (3).

Now we quote below other results which can be derived similarly by using combinatorial arguments as used above.

(i) For $a \geqq b - c, c > 0$,

$$(12) \qquad \binom{a+b}{a} P\{\delta_{a,b}^{+(-c)} \geqq j\} = \binom{a+b}{a+c+j-1}$$

$$(13) \qquad \binom{a+b}{a} P\{\Delta_{a,b}^{-(-c)} = j\} = \binom{a+b}{a} P\{\Delta_{a,b}^{+(-c)} = j\} =$$

$$= \frac{a - b + 2c + 4j}{a + b + 2} \binom{a + b + 2}{a + c + 2j + 1}$$

(14) $$\binom{a+b}{a} P\{\sigma_{a,b}^{-(-c)} = j\} = \sum_{m=1}^{[(b-c-j+1)/2]} \frac{a - b + 2c + j + 3m}{a + b - j - m + 2} \cdot$$

$$\cdot \binom{j + m - 1}{m - 1} \binom{a + b - j - m + 2}{a + c + m + 1}$$

(15) $$\binom{a+b}{a} P\{\sigma_{a,b}^{+(-c)} = j\} = \sum_{m=0}^{[(b-c-j+1)/2]} \frac{a - b + 2c + j + 3m - 1}{a + b - j - m + 1} \cdot$$

$$\cdot \binom{j + m}{m} \binom{a + b - j - m + 1}{a + m + c}$$

(ii) For $a \geqq b + c, c > 0$,

(16) $$\binom{a+b}{a} P\{\delta_{a,b}^{(c)} = j\} = 2^{j-1} \frac{a - b + j - 1}{a + b - j + 1} \binom{a + b - j + 1}{a}$$

(17) $$\binom{a+b}{a} P\{\delta_{a,b}^{-(c)} \geqq j\} = \binom{a + b}{a + j - 1}$$

(18) $$\binom{a+b}{a} P\{\delta_{a,b}^{+(c)} \geqq j\} = \binom{a + b}{a + j}$$

(19) $$\binom{a+b}{a} P\{\Delta_{a,b}^{-(c)} = j\} = \frac{a - b + 4j - 2}{a + b + 2} \binom{a + b + 2}{a + 2j}$$

(20) $$\binom{a+b}{a} P\{\Delta_{a,b}^{+(c)} = j\} = \frac{a - b + 4j + 2}{a + b + 2} \binom{a + b + 2}{a + 2j + 2}$$

(21) $$\binom{a+b}{a} P\{\sigma_{a,b}^{-(c)} = j\} = \sum_{m=1}^{[(b-j+2)/2]} \frac{a - b + j + 3m - 2}{a + b - j - m + 2} \cdot$$

$$\cdot \binom{j + m - 1}{m - 1} \binom{a + b - j - m + 2}{a + m}$$

(22) $$\binom{a+b}{a} P\{\sigma_{a,b}^{+(c)} = j\} = \sum_{m=0}^{[(b-j)/2]} \frac{a - b + j + 3m + 1}{a + b - j - m + 1} \cdot$$

$$\cdot \binom{j + m}{m} \binom{a + b - j - m + 1}{a + m + 1}$$

We note that results (16)–(22) are independent of c.

(iii) For $a \geqq b, c = 0$,

$$(23)\qquad \binom{a+b}{a} P\{\delta_{a,b}^{(0)} = j\} = 2^j \frac{a-b+j}{a+b-j}\binom{a+b-j}{a}.$$

Relations (10), second part of (13), (14), (18), (20) and (22) hold also for $c = 0$.

Setting $j = 0$ in (23), we get

$$P\{\alpha_r > \beta_r \quad \text{for} \quad r = 1, 2, \ldots, a+b\} = \frac{a-b}{a+b},$$

verifying the classical ballot problem (1).

References

[1] *A. Aeppli:* Zur Theorie Verketteter Wahrscheinlichkeiten. Thèse, Zürich (1924).

[2] *D. André:* Solution directe du problème rèsolu par M. Bertrand. C. R. Acad. Sci. (Paris), 105 (1887), 436—437.

[3] *É. Barbier:* Généralisation du problème rèsolu par M. J. Bertrand. C. R. Acad. Sci. (Paris), 105 (1887), 407.

[4] *J. Bertrand:* Solution d'un problème. C. R. Acad. Sci. (Paris), 105 (1887), 369.

[5] *M. T. L. Bizley*: Derivation of a new formula for the number of minimal lattice paths from (0, 0) to (km, kn) having just t contacts with the line my = nx and having no points above this line; and a proof of Grossman's formula for the number of paths which may touch but do not rise above this line. J. Inst. Actuar., 80 (1954), 55—62.

[6] *M. T. L. Bizley*: Problem 5503. Amer. Math. Monthly, 74 (1967), 728.

[7] *K. L. Chung, W. Feller:* Fluctuations in coin tossing. Proc. Nat. Acad. Sci. U.S.A., 35 (1949), 605—608.

[8] *A. Dvoretzky, Th. Motzkin:* A problem of arrangements. Duke Math. Journal, 14 (1947), 305—313.

[9] *O. Engelberg:* Exact and limiting distributions of the number of lead positions in 'unconditional' ballot problems. J. Appl. Prob., 1 (1964), 168—172.

[10] *O. Engelberg:* Generalizations of the ballot problem. Z. Wahrscheinlichkeitstheorie, 3 (1965), 271—275.

[11] *W. Feller:* An introduction to probability theory and its Applications, Vol. I., Third Edition, John Wiley, New York (1968).

[12] *H. D. Grossman:* Another extension of the ballot problem. Scripta Math., 16 (1950), 120—124.

[13] *S. G. Mohanty, T. V. Narayana:* Some properties of compositions and their application to probability and statistics I. Biometrische Zeitschrift, 3 (1961), 252—258.

[14] *L. Takács:* A generalization of the ballot problem and its application in the theory of queues. J. Amer. Statist. Assoc., 57 (1962), 327—337.

[15] *L. Takács:* Ballot problems. Z. Wahrscheinlichkeitstheorie, 1 (1962), 154—158.

[16] *L. Takács:* The distribution of majority times in a ballot. Z. Wahrscheinlichkeitstheorie, 2 (1963), 118—121.

[17] *L. Takács:* Fluctuations in the ratio of scores in counting a ballot. J. Appl. Prob., 1 (1964), 393—396.

[18] *L. Takács:* Combinatorial methods in the theory of stochastic processes. John Wiley, New York (1967).

[19] *L. Takács:* On the fluctuations of election returns. J. Appl., Prob., 7 (1970), 114—123.

Souhrn

NĚKTERÉ VYSLEDKY O ROZLOŽENÍCH NA ZOBECNĚNÝCH PROBLÉMECH HLASOVÁNÍ

JAGDISH SARAN, KANWAR SEN

Předpokládejme, že při hlasování kandidát A dostal a hlasů, kandidát B dostal b hlasů, a že všech $\binom{a+b}{a}$ možných hlasovacích posloupností je stejně pravděpodobných. Označme α_r (resp. β_r) počet hlasů odevzdaných pro A (resp. B) mezi r prvními hlasy, $r = 1, \ldots, a + b$. Jednoduchými kombinatorickými metodami se v článku odvozují pro $a \geqq b - c$ pravděpodobnostní rozložení určitých náhodných veličin spojených s průběhem hlasování. Jde o náhodné veličiny rovnající se počtu indexů $r = 1, \ldots, a + b$, pro něž (i) $\alpha_r = \beta_r - c$, (ii) $\alpha_r = \beta_r - c$, ale $\alpha_{r-1} = \beta_{r-1} - c \pm 1$, (iii) $\alpha_r = \beta_r - c$, ale $\alpha_{r-1} = \beta_{r-1} - c \pm 1$ a $\alpha_{r+1} = \beta_{r+1} - c \pm 1$, kde $c = 0, \pm 1, \pm 2, \ldots$. Některé známé výsledky Andrého (1887), Takácse (1963) a jiných autorů jsou jednoduchými důsledky dokázaných vět.

Author's address: Prof. *Jagdish Saran*, Prof. *Kanwar Sen*, Department of Mathematical Statistics, Faculty of Mathematics, University of Delhi, Delhi-11CC07, Ir.dia.