A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 255, 2011

Wiesław Wagner[*], *Andrzej Mantaj*[**]

REMARKS ON THE SCHEME OF SIMPLE SAMPLING WITH REPLACEMENT

Abstract. In the paper there were carried out deliberations on the formula for the probability of the second-row belonging of the set pair of units to the sample in the scheme of sampling with replacement. It was derived in various ways. Some formulas were illustrated with examples.

Key words: Representative method, scheme of simple sampling with replacement, probability of second-row belonging of set pair of elements to sample.

I. INTRODUCTION

The scheme of individual unlimited independent sampling from a limited parent population containing *N* units allow for the possibility of multiple occurrence of the same units in the sampled sample. It happens when sampling of units is done at invariable composition of the population, i.e. the probability of selection of units from the population is constant and remains equal in the whole process of sampling the sample.

In the paper there were given the mathematical model of the scheme of sampling with replacement, the formulas for the number of samples expressed by the number of variation with repetitions, the probability for sampling the samples with the set element and there was derived, in various ways, the formula for the probability of the second-row belonging of the set pair of units to the sample. Some formulas were illustrated with examples.

II. SCHEME OF SAMPLING WITH REPLACEMENT

Let *X* be the tested characteristic, and *n* the number of sampled units to the sample from the parent population having *N* units according to the scheme of simple sampling with replacement. The mathematical model of the mentioned

[*] Profesor, University of Information Technology and Menagement in Rzeszów.
[**] Doktor, University of Information Technology and Menagement in Rzeszów.

population, expressed by the set ***U*** composed of N units, tested in respect of the characteristic X is the N-element sequence $(X_1, X_2, ..., X_j, ..., X_N)$, where X_j expresses the value of the characteristic X for the j-th unit of the sampling. Each unit in the given model has equal chance to be selected to the sample (e.g. Sarndal et. al. 1992, Wywiał 1991, Steczkowski 1995, Bracha 1996). In the scheme of simple sampling with replacement we assume the following:

- there is taken into account the sequence of sampled units,
- the sampled unit is returned to the parent population,
- the sampled unit can appear in the sample repeatedly,
- at each stage of the sampling the parent population remains constant,
- pairs (j, k), (k, j) of units are deemed different.

In the scheme of simple sampling with replacement the sample of units ***s*** of the size n from the population containing N units expresses n-term variation with repetitions from N-element set. The set of all such samples ***S*** contains N^n of samples. The numbers of the units in the samples can repeat themselves or not, in particular, they all can be equal. The probability of sampling any sample $\boldsymbol{s} \in \boldsymbol{S}$ is constant and is $p_s = \frac{1}{N^n}$.

The probability of sampling the j-th unit from the indicated population equals $\frac{1}{N}$, and of not sampling it $1 - \frac{1}{N}$. Hence the probability that the sample composed of n units, sampled according to the scheme of simple sampling with replacement, will contain a given unit sampled repeatedly, let us say r times, is

$$\binom{n}{r}\left(\frac{1}{N}\right)^r\left(1-\frac{1}{N}\right)^{n-r} \tag{1}$$

Derivation of the formula (1). Let N be the size of the set of units ***U***. By j and $\sim j$ we denote sampled or not sampled unit of this number from the set ***U*** at the i-th stage of sampling. There are n of these stages, which corresponds to the size of the sampled sample. For the specific n and r, but at optional N, there are obtained as many possible r-element subsets from the n-element set as is the number of combinations $\binom{n}{r}$. It is illustrated by the following example.

Example 1. The sequentially created sample in subsequent stages of the sampling from the set ***U*** containing N units, for the singled out j-th unit sampled two times ($r = 2$), for example for $n = 5$, shows the scheme, where $b = 1/N$:

Stage of sampling	Sampled unit	Probability of sampling the unit	Composition of sample	Probability of composition of sample
1	$\sim j$	$1 - b$	$\sim j$	$1 - b$
2	j	b	$\sim j, j$	$b(1 - b)$
3	$\sim j$	$1 - b$	$\sim j, j, \sim j$	$b(1 - b)^2$
4	$\sim j$	$1 - b$	$\sim j, j, \sim j, \sim j$	$b(1 - b)^3$
5	j	b	$\sim j, j, \sim j, \sim j, j$	$b^2(1 - b)^3$

In the statement we have $r = 2$ and $n - r = 3$. The sample consists of two types of units which were denoted binary-wise: 0 (a unit different from the j-th unit) and 1 (the j-th unit), which gives the binary sequent 01001. There can be created as many such sets as many there are 2-term combinations from the 5-element set there is, i.e. $\binom{5}{2} = 10$. It is finally determined by the formula (1) for any N at $n = 5$ and $r = 2$.

III. FORMULA FOR PROBABILITY OF SECOND-ROW BELONGING OF PAIR OF UNITS

In monographs of Sarndal et. al. (1992, s. 50) and Brach (1996, s. 25) there is given the formula

$$\pi_{jk} = 1 - 2\left(1 - \frac{1}{N}\right)^n + \left(1 - \frac{2}{N}\right)^n \tag{2}$$

for the probability of the 2. row belonging of the pair of units (j, k) in the n-element sample sampled according to the scheme of simple sampling with replacement from the population having N units. In the mentioned monographs there are no analytical deliberations concerning the formula (2). Further on we will deal with various ways of its deriving and its properties.

In particular from the given formula (1) we obtain the probability $\left(1 - \frac{1}{N}\right)^n$ of non-occurrence of the singled out unit ($r = 0$) in the n-element sample, and hence by the use of the formula of the reverse probability we obtain the probability of the 1^{st} row belonging of the singled out j-th unit to the sample.

$\pi_j = 1 - \left(1 - \frac{1}{N}\right)^n$ From the given formula for π_j, the formula (2) can be written as $\pi_{jk} = 1 - 2\ (1 - \pi_j) + \left((1 - \pi_j)^{1/n} - \frac{1}{N}\right)^n$. At $n \to \infty$ and determined N there occurs $\pi_{jk} \to 1$ and at the set *n* and $N \to \infty$, we have $\pi_{jk} \to 0$, which is shown in the exemplifying statement of the determined probabilities π_{jk} for the set *n* and *N*:

n	*N*					
	20	50	100	200	500	1000
10	0,1512	0,0307	0,0083	0,0022	0,0004	0,0001
20	0,4046	0,1068	0,0318	0,0087	0,0015	0,0004
30	0,6131	0,2029	0,0661	0,0189	0,0033	0,0008
50	0,8513	0,4015	0,1542	0,0484	0,0089	0,0023
100	0,9882	0,7516	0,4006	0,1545	0,0326	0,0090
150	0,9991	0,9056	0,6054	0,2785	0,0670	0,0193
200	0,9999	0,9651	0,7496	0,4001	0,1085	0,0328

Further on we give a few ways of deriving the formula (2).

A. Derivations with the use of the classical definition of probability.

In this case the deliberations consist in specification of specific sets of elements occurring in the samples of the size *n* created from the elements (units) of the parent population of the size *N*. We assume that the population contains the elements *a, b, c, ...*

Let us examine the variants: (a) $N = 2$, $n = 2, 3, 4$, (b) $N = 3$, $n = 2, 3, 4$. For each variant we determine the probabilities of the 2. row belonging of the set pair (*a, b*) to the sample by calculating appropriate quotient of the number of favouring cases to the possible ones, i.e. we use the classical definition of probability. At examining individual cases we give compositions of elements of all possible *n*-element samples, their number and the number of favouring samples, i.e. containing the set pair (a, b):

(a) $N = 2$:

n	Samples	Number of possible samples	Number of favouring samples	Probabilities of belonging
2	a,a; a,b; b,a; b,b	$4 = 2^2$	2	1/ 2 = 2/4
3	a,a,a; a,a,b; a,b,a; b,a,a; a,b,b; b,a,b; b,b,a; b,b,b	$8 = 2^3$	6	3/4 = 6/8
4	a,a,a,a; a,a,a,b; a,a,b,a; a,b,a,a; b,a,a,a; a,a,b,b; a,b,a,b; a,b,b,a; b,a,b,a; b,a,a,b; b,b,a,a; a,b,b,b; b,a,b,b; b,b,a,b; b,b,b,a; b,b,b,b	$16 = 2^4$	14	7/8 = 14/16

The determined probabilities of belonging at $N = 2$ for $n = 2, 3, 4$ lead to the formulas:

n	2	3	4
Probabilities	1/2	3/4	7/8
Formulas	$1-2\left(1-\frac{1}{2}\right)^2$	$1-2\left(1-\frac{1}{2}\right)^3$	$1-2\left(1-\frac{1}{2}\right)^4$

hence the general formula for $N = 2$ and for any n in the form $1-2\left(1-\frac{1}{2}\right)^n$.

(b)N =3:

<table>
<tr><th>n</th><th>Samples</th><th>Number of possible samples</th><th>Number of favouring samples</th><th>Probabili-ties of belonging</th></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>2</td><td>a,a; a,b; a,c; b,b; b,a; b,c; c,a; c,b; c,c;</td><td>9 = 3²</td><td>2</td><td>2/9</td></tr>
<tr><td>3</td><td><table>
<tr><th>a,b</th><th>a,c</th><th>b,c</th><th>a,b,c</th></tr>
<tr><td>a,a,a</td><td>a,a,c</td><td>b,b,c</td><td>a,b,c</td></tr>
<tr><td>a,a,b</td><td>a,c,a</td><td>b,c,b</td><td>a,c,b</td></tr>
<tr><td>a,b,a</td><td>c,a,a</td><td>c,b,b</td><td>b,a,c</td></tr>
<tr><td>a,b,b</td><td>a.c.c</td><td>c,c,b</td><td>b,c,a</td></tr>
<tr><td>a,a,b</td><td>c.a.c</td><td>c,b,c</td><td>c,a,b</td></tr>
<tr><td>a,b,a</td><td>c,c,a</td><td>b,c,c</td><td>c,b,a</td></tr>
<tr><td>a,b,b</td><td>c,c,c</td><td></td><td></td></tr>
<tr><td>b,b,b</td><td></td><td></td><td></td></tr>
<tr><td>8</td><td>7</td><td>6</td><td>6</td></tr>
</table></td><td>27=3³</td><td>12 (= 6 + 6)</td><td>12/27 = 4/9</td></tr>
</table>

<table>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>4</td><td>
<table>
<tr><th>a,b</th><th>a,c</th><th>b,c</th><th>a,a,b,c</th><th>a,b,b,c</th><th>a,b,c,c</th></tr>
<tr><td>a,a,a,a</td><td>a,a,a,c</td><td>b,b,b,c</td><td>a,a,b,c</td><td>a,b,b,c</td><td>a,b,c,c</td></tr>
<tr><td>a,a,a,b</td><td>a,a,c,a</td><td>b,b,c,b</td><td>a,a,c,b</td><td>a,b,c,b</td><td>a,c,b,c</td></tr>
<tr><td>a,a,b,a</td><td>a,c,a,a</td><td>b,c,b,b</td><td>a.b,a,c</td><td>a,c,b,b</td><td>a,c,c,b</td></tr>
<tr><td>a,b,a,a</td><td>c,a,a,a</td><td>c,b,b,b</td><td>a,b,c,a</td><td>b,a,b,c</td><td>b,a.b,c</td></tr>
<tr><td>b,a,a,a</td><td>a,a.c.c</td><td>b,b,c,c</td><td>a,c,a,b</td><td>b,c,b,a</td><td>b,c,b,a</td></tr>
<tr><td>a,a,b,b</td><td>a,c,a,c</td><td>b,c,b,c</td><td>a,c,b,a</td><td>b,a,c,b</td><td>b,a,c,b</td></tr>
<tr><td>a,b,a,b</td><td>a,c,c,a</td><td>b,c,c,b</td><td>b,a,a,c</td><td>b,c,a,b</td><td>b,c,a,b</td></tr>
<tr><td>a,b,b,a</td><td>c,a,c,a</td><td>c,b,c,b</td><td>b,c,a,a</td><td>b,b,a,c</td><td>b,b,a,c</td></tr>
<tr><td>b,a,b,a</td><td>c,a,a,c</td><td>c,b,b,c</td><td>b,a,c,a</td><td>b,b,c,a</td><td>b,b,c,a</td></tr>
<tr><td>b,b,a,a</td><td>c,c,a,a</td><td>c,c,b,b</td><td>c,a,a,b</td><td>c,b,b,a</td><td>c,a,b,b</td></tr>
<tr><td>b,a,a,b</td><td>a,c,c,c</td><td>b,c,c,c</td><td>c,a,b,a</td><td>c,b,a,b</td><td>c,b,a,b</td></tr>
<tr><td>a,b,b,b</td><td>c,a,c,c</td><td>c,b,c,c</td><td>c,b,a,a</td><td>c,a,b,b</td><td>c.b.b.a</td></tr>
<tr><td>b,a,b,b</td><td>c,c,a,c</td><td>c,c,b,c</td><td></td><td></td><td></td></tr>
<tr><td>b,b,a,b</td><td>c,c,c,a</td><td>c,c,c,b</td><td></td><td></td><td></td></tr>
<tr><td>b,b,b,a</td><td>c,c,c,c</td><td></td><td></td><td></td><td></td></tr>
<tr><td>b,b,b,b</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>16</td><td>15</td><td>14</td><td>12</td><td>12</td><td>12</td></tr>
</table>
</td><td>$81=3^4$</td><td>50
(=14+3·12),</td><td>50/81</td></tr>
</table>

The given probabilities of belonging at $N = 3$ for $n = 2, 3, 4$ lead to the schemes:

n	Probabilities	Formulas
2	$\frac{2}{9}=\frac{1}{9}+\frac{1}{9}$	$1-2\left(1-\frac{1}{3}\right)^2+\left(1-\frac{2}{3}\right)^2$
3	$\frac{4}{9}=\frac{12}{27}=\frac{11}{27}+\frac{1}{27}$	$1-2\left(1-\frac{1}{3}\right)^3+\left(1-\frac{2}{3}\right)^3$
4	$\frac{50}{81}=\frac{49}{81}+\frac{1}{91}$	$1-2\left(1-\frac{1}{3}\right)^4+\left(1-\frac{2}{3}\right)^4$

hence there is obtained the general scheme for $N = 3$ and for any n in the form $1-2\left(1-\frac{1}{3}\right)^n+\left(1-\frac{2}{3}\right)^n$ or $1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n$, which finally gives the formula (2).

B. Derivation with the use of probability of random events.

We introduce the following random events:

- A, A' – sampled and not sampled the element j to the n-element sample,
- B, B' – sampled and not sampled the element k to the n-element sample,

- A∩B, A'∪B' – sampled and not sampled the pair of elements (*j, k*) to the *n*-element sample.

From the previous deliberations follow the formulas for probabilities:

$$P(A') = \left(1 - \frac{1}{N}\right)^n, P(B') = \left(1 - \frac{1}{N}\right)^n, \ P(A' \cap B') = \left(1 - \frac{2}{N}\right)^n \tag{3}$$

There should be determined the formula for the probability $P(A \cap B)$?. We use de Morgan's law for random events $(A' \cup B' = (A \cap B)'$ and the formula for the probability of the sum of two non-excluding each other random events $P(A' \cup B') = P(A') + P(B') - P(A' \cap B')$.

The appropriate derivation of the formula for $P(A \cap B)$ runs in the following way:

a) $P(A' \cup B') = P[(A \cap B)'] = 1 - P(A \cap B)$,

b) $1 - P(A \cap B) = P(A') + P(B') - P(A' \cap B')$,

c) $P(A \cap B) = 1 - P(A') - P(B') + P(A' \cap B')$.

After appropriate substitution (3) to c) we obtain (2). The formula c) is also given by Koronacki and Mielniczuk (2001, p. 403). Substituting to the formula given in d) to its right side of the probability (3), we get the formula (2).

C. *Derivation by the use of permutation with repetitions*.

Let us write the formula for π_{jk} for particular cases, when $n = 2, 3, 4$ with any N, basing on the formulas listed in point A. We have respectively:

$$n = 2, \ \pi_{jk} = \frac{2}{N^2}; \quad n = 3, \ \pi_{jk} = \frac{6(N-1)}{N^3}; \ n = 4,$$
$$\pi_{jk} = \frac{2(6N^2 - 12N + 7)}{N^4}. \tag{4}$$

The formulas (4) show that the probabilities π_{jk} can be also expressed by the quotient $\pi_{jk} = \frac{W_{n-2}(N)}{N^n}$, whose numerator is the multinomial of n-2 degree of the variable N, and the denominator is the power of the number N. To demonstrate the formulas (4) we will take into consideration the possible samples of the size n. For this purpose we will use the notion of the number of permutation with repetitions from the n-element set in which there occurs m various elements of the multiplication factors $n_1, n_2, ..., n_m$ so that $n = n_1 + n_2 + ... + n_m$. The number of such permutations is determined by the formula (e.g. Mizerski et. al. 1999, p. 324)

$$P(n; n_1, n_2, ..., n_m) = \frac{n!}{n_1! n_2! \cdot ... \cdot n_m!}. \quad (5)$$

For further deliberations we assume the denotations:

- $\boldsymbol{U} = \{1, 2, ..., N\}$ – the set of the numbers of the units in the parent population,
- (j, k) – the set pair of units in the sample,
- $\boldsymbol{U'} = \boldsymbol{U} - \{j, k\}$ - the set of the numbers of the units of the parent population not containing the pair of units (j, k),
- $\# \boldsymbol{U} = N$, $\# \boldsymbol{U'} = N - 2$ – the sizes of the sets $\boldsymbol{U}$ and $\boldsymbol{U'}$,
- $\mathbf{s} = (j_1, j_2, ..., j_n)$, at $j_1, j_2, ..., j_n \in \boldsymbol{U}$ – the sample of n units from the set $\boldsymbol{U}$,
- $\mathbf{s} = (j, k, j_3, j_4, ..., j_n)$, at $j_3, j_4, ..., j_n \in \boldsymbol{U'}$- the sample of n units with the singled out pair (j, k).

For the determined size n we determine various forms of the samples $\boldsymbol{s}$, the number of their occurrence and the total number of the samples L_{jk} with the singled out pair of the units (j, k), i.e. these will be the expressions given in the numerators of the formulas (4). For the cases $n = 3$ and 4 there are examined different variants of the possible combinations with repetitions and without repetitions of the sampled numbers of the units:

(a) $n = 2$, $\boldsymbol{s} = (j, k)$,

$$P(2;1,1) = \frac{2!}{1!1!} = 2, \quad L_{jk} = 2;$$

(b) $n = 3$, $\boldsymbol{s} = (j, k, j_3)$, $j_3 \in \boldsymbol{U}$,

i) $j_3 = j$, $\boldsymbol{s} = (j, k, j)$,

$$P(3;2,1) = \frac{3!}{2!1!} = 3, \text{ the same for } j_3 = k, \text{ i.e. } 6 \text{ samples},$$

ii) $j_3 \in \boldsymbol{U'}$, $N - 2$ cases,

$$P(3;1,1,1) = \frac{3!}{1!1!1!} = 6, \quad \text{i.e. } 6(N-2) \text{ samples},$$

which at $n = 3$ finally leads to the number of samples $L_{jk} = 6 + 6(N-2) = 6(N-1)$;

(c) $n = 4$, $\boldsymbol{s} = (j, k, j_3, j_4)$, $j_3, j_4 \in \boldsymbol{U}$.

i) $j_3 = j_4 = j$, $\boldsymbol{s} = (j, k, j, j)$,

$$P(4;3,1) = \frac{4!}{3!1!} = 4, \text{ the same for } j_3 = j_4 = k, \quad \text{in total 8 samples,}$$

ii) $j_3 = j, j_4 = k$, $\boldsymbol{s} = (j, k, j, k)$,

$$P(4;2,2) = \frac{4!}{2!2!} = 6, \text{ i.e. } 6 \text{ samples,}$$

iii) $j_3 = j$, $j_4 \in \boldsymbol{U}'$, $\boldsymbol{s} = (j, k, j, j_4)$, $N - 2$ cases,

$$P(4;2,1,1) = \frac{4!}{2!1!1!} = 12, \text{ the same for } j_3 = k, \text{ i.e. } 24(N-2) \text{ samples,}$$

iv) $\boldsymbol{s} = (j, k, j_3, j_4)$, $j_3, j_4 \in \boldsymbol{U}'$**:**

aa) $\boldsymbol{s} = (j, k, j_3, j_3)$, $N - 2$ cases,

$$P(4;1,1,2) = 12, \text{ i.e. } 12(N-2) \text{ samples,}$$

bb) $\boldsymbol{s} = (j, k, j_3, j_4)$, j_3 $j_3 \neq j_4 \in \boldsymbol{U}'$, $\binom{N-2}{2}$ cases,

$$P(4;1,1,1,1) = 24, \text{ i.e. } 24\binom{N-2}{2} = 12(N-2)(N-3) \text{ samples,}$$

which at $n = 4$ leads to the number of samples

$$L_{jk} = 14 + 36(N-2) + 12(N-2)(N-3) =$$
$$= 12N^2 - 24N + 14 = 2(6N^2 - 12N + 7) .$$

In all cases (a), (b) and (c) there occurs the compatibility with the given formulas of the numerators in (4), and at the same time there is derived the formula (2).

The formula (2) can also be given in another form by the use of Newton's binomial $(1-b)^n = \sum_{r=0}^{n} \binom{n}{r} (-1)^r b^r$, then we have $\pi_{jk} = 2\sum_{r=2}^{n} \frac{(-1)^r}{N^r} \cdot (2^{r-1} - 1)$.

Derivations of presented probability formulas were given, among other authors, by. W. Czerniak (1971) and Y. Tillé (2006).

To illustrate the formula (2) let us examine the case when $n > N$.

Example 2. Let us consider the case of $N = 3$ and $n = 4$, i.e. when the sizes of the samples are greater than the size of the parent population. All such samples contain the repeating elements of the population. There are listed in the columns 1, 2, 3 in the statement:

No.	1	2	3	4	W	No.	1	2	3	4	W	No.	1	2	3	4	W
1	1	1	1	1	0	28	2	1	1	1	1	55	3	1	1	1	0
2	1	1	1	2	1	29	2	1	1	2	1	56	3	1	1	2	1
3	1	1	1	3	0	30	2	1	1	3	1	57	3	1	1	3	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
25	1	3	3	1	0	52	2	3	3	1	1	79	3	3	3	1	0
26	1	3	3	2	1	53	2	3	3	2	0	80	3	3	3	2	0
27	1	3	3	3	0	54	2	3	3	3	0	81	3	3	3	3	0

In total there are $3^4 = 81$ samples. We assume here that the singled out pair of elements is constituted by the pair (1, 2). The column „W" shows the result of the binary condition indicating if the pair of elements of a given sample occurred (number 1) or did not occur (number 0). There are samples 50 with the result of the binary condition equal 1, which leads to the quotient $\pi_{12} = \frac{50}{81}$. In turn, from the formula (2) we have $\pi_{12} = 1 - 2\left(1-\frac{1}{3}\right)^4 + \left(1-\frac{2}{3}\right)^4 = \frac{50}{81}$, i.e. there occurs the required compatibility.

The formula (2) can be also examined simulation-wise at generating the adequate random samples by a computer. Such samples will be created at the use of the program EXCEL. The essence of the Monte Carlo experiment carried out here comes down to random creation of many random samples of the set size n, and the to performing an appropriate transformation of the sequence of numbers from the interval (0, 1) to the integers 1, 2, …, N.

We assume the denotations:

- ϵ – the set sufficiently low constant (e.g. 10^{-4}),
- M – the number of simulation repetitions.

Successive steps of the Monte Carlo experiment include the following actions:

a) we determine the quotient $h = 1/N$,

b) we create N intervals

$$(0,\ h); (h+\varepsilon,\ 2h); (2h+\varepsilon,\ 3h); ,...,\ \ ((N-1)h+\varepsilon,\ Nh);$$

c) we generate the sequence of n random numbers $r_1, r_2, ..., r_n$ from the interval (0, 1) by the RANDOM procedure of the program EXCEL,

d) we assign to the random numbers from c) the class intervals from the step b), and then we replace the random numbers of these ranges with the numbers of these ranges, which leads to determining the sequence of integers $j_1, j_2, ..., j_n$, such as $j_1, j_2, ..., j_n \in \{1,2,...,N\}$,

e) we check if in the sequence $j_1, j_2, ..., j_n$ there is the pair (1, 2); if so then we assign the number 1, otherwise 0,

f) we repeat the steps c), d) and e) M times,

g) we make calculation of the number of cases meeting the condition of belonging of the pair (1, 2) in the step e) assuming that it occurred m times,

h) the quotient m/M is the estimation of the probability (2).

We give the results of the carried out Monte Carlo experiment for $N = 5$, at $n = 3, 6$. For the assessment of the quality of the generator of random numbers there was also determined the arithmetic mean and the standard deviation for the set M. Their theoretical values are 0,5 and 0,2887. The simulations were carried out at M = 300, 500, 1000, 1500, 2000, 3000, 5000 and there was assumed $\epsilon = 0{,}0001$. Appropriate intervals from the step b) have the form:

1	2	3	4	5
(0, 0,2)	(0,2001, 0,4)	(0,4001, 0,6)	(0,6001, 0,8)	(0,8001, 1)

(a), value from the formula (2) $\pi_{12} = 0{,}192$

The procedure in EXCEL sheet at $N=5$, $n=3$ is shown by a fragment of the scheme:

	R1	R2	R3	J1	J2	J3	1	1	1	Sum	2	2	2	Sum	Value
1	0,6946	0,8391	0,8014	4	5	5	0	0	0	0	0	0	0	0	0
2	0,3355	0,4333	0,3467	2	3	2	0	0	0	0	2	0	2	4	0
3	0,5593	0,9056	0,6781	3	5	4	0	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

In columns 1, 1, 1 and 2, 2, 2 there is made simultaneous checking of occurrence of the elements of the pair (1, 2), i.e. there should occur at least one figure one and one figure two. The columns of the sums are determined from summing the elements in the marked three columns „1" and „2". In turn the column „value" gives the element 1, when at the same time both previously determined sums are non-zero. The values of the probability of the 2. row belonging for the assumed *N, n* obtained from the simulation are given in the statement:

	$n = 3,\ \pi_{12} = 0{,}192$			$n = 6,\ \pi_{12} = 0{,}5224$		
M	Mean	Standard deviation	Probability	Mean	Standard deviation	Probability
300	0,4877	0,2933	0,200	0,5053	0,2918	0,520
500	0,4912	0,2929	0,204	0,4996	0,2893	0,500
1000	0,4923	0,2886	0,200	0,5007	0,2903	0,518
1500	0,4981	0,2872	0,205	0,5012	0,2906	0,516
2000	0,4981	0,2878	0,201	0,4948	0,2871	0,533
3000	0,4965	0,2912	0,204	0,4990	0,2899	0,525
5000	0,4973	0,2887	0,198	0,5000	0,2889	0,517

At $n = 3$ the weighed arithmetic mean, the standard deviation and the probability of belonging and their percentage participations in comparison with the expected ones are 0,4960 (99,2%), 0,2894 (100,24%) and 0,2011 (104,74%), and for $n = 6$, the results were 0,4992 (99,85%), 0,2892 (100,17%) and 0,5204 (99,63%).

The presented results of the simulation indicate good approximations of the probabilities of the second row belongings of a set pair of elements to the sample. They are better at the increasing size of the sample n at the set N. The tests showed that at $N = 5$ for $n = 4, 6$ there were obtained the approximations differing by less than 0,5% from the real values given by the formula (2).

IV. SUMMARY

In the paper there were presented the formulas connected with the model of sampling the samples according to the scheme of sampling with replacement. There was derived, in various ways, the formula for the probability of the second row belonging of a set pair of units from the parent population in the sample. In the first way there was used the classical definition of the probability, in the second one – the permutation with repetitions, in the third one – Newton's binomial formula, and in the fourth one the simulation tests were used.

Let us notice that the formula (2) can be written in the form

$$1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n=\sum_{q=0}^{2}\binom{2}{q}(-1)^q\left(1-\frac{q}{N}\right)^n.$$

There immediately arises the analogy to give the formula for the probability of belonging of three singled out elements, i.e.

$$1-3\left(1-\frac{1}{N}\right)^n+3\left(1-\frac{2}{N}\right)^n-\left(1-\frac{3}{N}\right)^3=\sum_{q=0}^{3}\binom{3}{q}(-1)^q\left(1-\frac{q}{N}\right)^n,$$

which allows giving a more general form of the right side of the formula for the singled out sequence of r elements $\sum_{q=0}^{r}\binom{r}{q}(-1)^q\left(1-\frac{q}{N}\right)^n$. This formula also occurs for $r=1$.

The issue of the independent sampling can also be examined for finite populations with probabilities of selection of individual units, not necessarily equal ones. It allows determining the probabilities of the 2. row belonging of a set pair of units depending on the set probabilities.

BIBLIOGRAPHY

Bracha Cz., (1996): *Teoretyczne podstawy metody reprezentacyjnej (Theoretical Basics of Representative Method).* Wyd. Naukowe PWN, Warszawa.

Czermak W., (1971): *O losowaniu niezależnym ze zwracaniem, Wybrane problemy metodologiczne badań reprezentacyjnych (On Independent Sampling with Replacement. Selected Methodological Problems of Representative Research), Biblioteka Wiadomości Statystycznych, tom 15*, GUS, Warszawa.

Koronacki J., Mielniczuk J., (2001): *Statystyka dla studentów kierunków technicznych i przyrodniczych. (Statistics for Students of Technical and Natural Majors.)* WN-T, Warszawa.

Mizerski W., Sadowski W., Garbarczyk A., Tokarska B., Mazur K., (1999): *Tablice matematyczne. (Mathematical Tables.)* Wyd. Adamantan, Warszawa.

Steczkowski J. (1995): *Zastosowanie metody reprezentacyjnej w badaniach społweczno-ekonomicznych. (Application of Representative Method in Social-Economic Research.)* Wyd. Naukowe PWN, Warszawa.

Sarndal C. E., Swensson B., Wretman J. (1992): *Model Assisted Survey Sampling.* Springer-Verlag, New York.

Tille Y. (2006): Sampling Algorithms, Springer-Verlag, New York.

Wywiał J. (1999): *Elementy metody reprezentacyjnej z wykorzystaniem statystycznego pakietu SPSS. (Elements of Representative Method with the Use of Statistical Packet SPSS.)* Wyd. AE Katowice.

Wiesław Wagner, Andrzej Mantaj

UWAGI O SCHEMACIE LOSOWANIA PROSTEGO ZE ZWRACANIEM

Wyróżnia się losowanie proste bez zwracania (lpbz) oraz ze zwracaniem (lpzz). Oba schematy określa się jako indywidualne losowania nieograniczone zależne i niezależne. Schemat lpzz ze skończonej populacji generalnej zawierającej N jednostek jest stosowany w sytuacjach, gdy do-

puszcza się możliwość wielokrotnego występowania tych samych jednostek w wylosowanej próbie. Ma to miejsce wtedy, gdy pobieranie jednostek odbywa się przy niezmiennym jej składzie, czyli prawdopodobieństwo wyboru każdej jednostki z populacji generalnej jest stałe i pozostaje jednakowe w całym procesie losowania próby.

W pracy przedstawiono ogólny model matematyczny schematu losowania ze zwracaniem, wzory na liczbę prób wyrażoną wariacją z powtórzeniami, podano prawdopodobieństwo wylosowania prób z zadanym elementem oraz wyprowadzono różnymi sposobami wzór na prawdopodobieństwo przynależności rzędu drugiego zadanej pary jednostek. Niektóre wzory zilustrowano przykładami.