

MINIMAL SAMPLE SIZE FOR TESTING TRINOMIAL PROPORTIONS FOR GIVEN PRECISION OF PROBABILITY OF TYPE I ERROR

Edyta Mazurek*
Katarzyna Ostasiewicz*

SUMMARY

The determination of sample size is a common task for many organizational researchers. Inappropriate, inadequate or excessive sample size continues to influence the quality, accuracy and costs of research. Sample size is one of the features of analysis that can influence the detection of significant differences for population so we can't ignore problem of sample size. This paper presents a procedure and a table for selecting sample size for simultaneously testing the parameters of a trinomial distribution. The results are obtained by examining the several possible value of a trinomial parameter vector and comparing the fixed first error type with the empirical one obtained by building the exact distribution through the code R.

Keywords: *Trinomial Distribution, Simultaneous Inference, Duncan-Pollastri Procedure, Sample Size.*

1. INTRODUCTION

Very often we have to face the problem of comparing many treatments or means of a variable in two different periods of time. For example, a frequent task is to compare consumers' preferences or to analyze a model of how consumers behave and how their preferences change during the time. The changes could be caused by, for example, a change of economic or political situation. A multinomial experiment is used for solving the above problems.

The multinomial experiment has the following characteristics:

- The experiment consists of n repeated trials (for example, we analyze n consumers).
- Each trial has a discrete number of possible outcomes (each consumer has a discrete number of possible outcomes for preferences).
- On any given trial (consumer) the probability that a particular outcome will occur is constant.
- The trials are independent, that is, the outcome on one trial does not affect the outcomes on other trials (preferences of any consumer are independent on preferences of other consumers).

* Department of Statistics - Wroclaw University of Economics - Komandorska 118/120, 53-352 WROCLAW (e-mail: edyta.mazurek@ue.wroc.pl ✉ katarzyna.ostasiewicz@ue.wroc.pl).

2. MULTINOMIAL DISTRIBUTION

A multinomial distribution is the probability distribution of outcomes of a multinomial experiment. The multinomial formula defines the probability of any outcome of a multinomial experiment. Suppose a multinomial experiment consists of n independent trials, and each trial can result in any of k possible outcomes: E_1, E_2, \dots, E_k . Suppose, further, that all possible outcomes can occur with probabilities

p_1, p_2, \dots, p_k $\left(p_i \geq 0 \text{ for } i = 1, 2, \dots, k \text{ and } \sum_{i=1}^k p_i = 1 \right)$. Then let the random variables (r.v.) X_i indicate the number of times outcome number i was observed over the n trials. The vector $\mathbf{X} = (X_1, X_2, \dots, X_k)$ follows a multinomial distribution with parameters n and \mathbf{p} , where $\mathbf{p} = (p_1, p_2, \dots, p_k)$. The probability function of this multinomial r.v. is:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, \sum_{i=1}^k x_i = n.$$

With a multinomial experiment, each trial can have two or more possible outcomes. Let us restrict to the special case, that is, trinomial experiment. It is defined by the possibility of three – and only three – possible outcomes of each trial of the experiment. In the further part of this paper we will focus on such a trinomial distribution of vector $\mathbf{X} = (X_1, X_2, X_3)$:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}.$$

Very often parameters of trinomial distribution: p_1, p_2 and p_3 are unknown. In such a situation the unknown parameters could be either estimated or tested. When we have more than one parameter very useful is simultaneous inference connected with estimation and testing. In the context of multinomial proportions, there appears a very important issue of choosing a sample size. Questions regarding simultaneous confidence intervals and sample size for such problems are discussed by Queensbury and Hurst (1964). Simultaneous estimation and testing are studied by Goodman (1965), Miller (1981), Thompson (1987), Sison and Glaz (1995).

The determination of sample size is common task for many organizational researchers. Inappropriate, inadequate or excessive sample size continue to influence the quality, the cost and accuracy of research. Sample size is one of the features of analysis that can influence the detection of significant differences for population so we can't ignore that problem. Small samples increase chance that a researcher obtains false positive results, or that important differences are missed. We should think about sample size while planning an analysis and generally we should minimize both errors: alpha – finding a difference that does not actually exist in the population and beta – failing to find a difference that actually exists in the population.

This paper presents a procedure based on the real distribution and a table reporting the sample size for simultaneously testing the parameters of a trinomial distribution, when we require that the fixed and the real first type error do not differ more than a required percentage.

3. THE DUNAN-POLLASTRI PROCEDURE OF SIMULTANEOUS TESTING THE PARAMETERS OF A TRINOMIAL DISTRIBUTION

Simultaneous inference is a common problem in many areas of application. If multiple null hypotheses are tested simultaneously, the probability of rejecting erroneously at least one of them increases beyond the pre-specified significance level. Simultaneous inference procedures have to be used which adjust for multiplicity and thus control the overall type I error rate.

Pollastri procedure (Pollastri, 2008; Pollastri and Riva, 2012) is concerned with two stage test of hypotheses about the trinomial distribution with parameters (p_1, p_2, p_3, n) and it is based on the proposal of Duncan procedure (Miller, 1980), which tests the null hypothesis against a series of mutually exclusive and exhaustive alternatives. While applied to results changing in time, this test allows to assess not only whether a parameter has changed over time, but also the direction of the change. Therefore, the test is particularly suitable for those situations in which the researcher must decide if the preferences for different products or proportions of population remained unchanged or have increased or decreased after some events.

Let (X_1, X_2, X_3) have a trinomial distribution with parameters n and p_1, p_2, p_3 . For trinomial distribution the following relationships exist: $p_3 = 1 - p_1 - p_2$, $X_3 = n - X_1 - X_2$. In what follows, the considered parameters shall be only p_1 and p_2 and the random variables X_1 and X_2 . This paper brings up problem of testing the null hypotheses:

$$H_0 : (p_1 = p_1^*) \cap (p_2 = p_2^*) \tag{1}$$

against the series of mutually exclusive and exhaustive alternatives:

- $H_{11} : (p_1 = p_1^*) \cap (p_2 > p_2^*)$
- $H_{12} : (p_1 = p_1^*) \cap (p_2 < p_2^*)$
- $H_{13} : (p_1 > p_1^*) \cap (p_2 = p_2^*)$
- $H_{14} : (p_1 < p_1^*) \cap (p_2 = p_2^*)$
- $H_{15} : (p_1 > p_1^*) \cap (p_2 > p_2^*)$
- $H_{16} : (p_1 < p_1^*) \cap (p_2 < p_2^*)$
- $H_{17} : (p_1 < p_1^*) \cap (p_2 > p_2^*)$
- $H_{18} : (p_1 > p_1^*) \cap (p_2 < p_2^*),$

where p_1^*, p_2^* represent the fixed values between 0 and 1, with $p_1^* + p_2^* < 1$.

For such problem of testing, Duncan proposed a two stages procedure based on the following statistics:

$$T = \max\{|Z_1|, |Z_2|\} \quad \text{and} \quad V = \min\{|Z_1|, |Z_2|\},$$

where

$$|Z_1| = \left| \frac{X_1 - n \cdot p_1^*}{\sqrt{n \cdot p_1^* \cdot (1 - p_1^*)}} \right| \quad \text{and} \quad |Z_2| = \left| \frac{X_2 - n \cdot p_2^*}{\sqrt{n \cdot p_2^* \cdot (1 - p_2^*)}} \right|.$$

Stage 1:

In the first stage of Duncan procedure statistics $T = \max\{|Z_1|, |Z_2|\}$ is compared with critical value c' and

- a) if $T = \max\{|Z_1|, |Z_2|\} \leq c'$ decide in favor of H_0 ($p_1 = p_1^*$ and $p_2 = p_2^*$) and the procedure stops
- b) if $T = \max\{|Z_1|, |Z_2|\} = |Z_i| > c'$ ($i = 1, 2$) decide $p_i \neq p_i^*$ (if the $Z_i > 0$ then $p_i > p_i^*$ and if the $Z_i < 0$ then $p_i < p_i^*$) and proceed to *stage 2*.

For Duncan procedure the critical value c' is decided by using Bonferroni inequality. Note that Duncan has not determined the exact distributions of T and V because he has not considered the correlation between Z_1 and Z_2 .

Stage 2:

Compare $V = \min\{|Z_1|, |Z_2|\} = |Z_j|$ ($j = 1, 2, j \neq i$) with the $(1 - \frac{\alpha}{2})$ th quantile of the standardized Normal distribution- $z_{1-\frac{\alpha}{2}}$.

- a) If the $|Z_j| \leq z_{1-\frac{\alpha}{2}}$, decide $p_j = p_j^*$.
- b) If the $|Z_j| > z_{1-\frac{\alpha}{2}}$, decide $p_j \neq p_j^*$ (if the $Z_j > 0$ then $p_j > p_j^*$ and if the $Z_j < 0$ then $p_j < p_j^*$).

In Stage 2 Duncan procedure also uses, as before underlined, only approximately distribution of statistic $V = \min\{|Z_1|, |Z_2|\}$.

The Pollastri's improvement of Duncan procedure relies on using the exact distribution of the statistics $T = \max\{|Z_1|, |Z_2|\}$ and $V = \min\{|Z_1|, |Z_2|\}$.

The density function of statistic $T = \max\{|Z_1|, |Z_2|\}$ is a mixture of two Arctangent density functions

$$f_T(t) = g(t; a_1) \cdot \pi_1 + g(t; a_2) \cdot \pi_2$$

with parameters $a_1 = \sqrt{\frac{1+\rho}{1-\rho}}$ and $a_2 = \sqrt{\frac{1-\rho}{1+\rho}}$, where ρ is the correlation coefficient between X_1 and X_2 and given by $\rho = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}$, and with proportions $\pi_1 = \frac{2}{\pi} \arctan(a_1)$ and $\pi_2 = \frac{2}{\pi} \arctan(a_2)$ (see Pollastri, 1979; Pollastri and Tornaghi, 2004; Zenga, 1979). The density function $g(t; a)$ for Arctangent random variable is defined as follows:

$$g(x; a) = \begin{cases} \frac{e^{-\frac{1}{2}x^2}}{\arctan(a)} \int_0^{ax} e^{-\frac{1}{2}y^2} dy & \text{for } x \geq 0, a > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

The density function of statistic V (Pollastri and Tornaghi, 2004) is given by $f_V(x) = 2(2\varphi(x)) - f_T(x)$, $x \geq 0$,

which is a linear combination of the density function of a Standard Normal random variable and of the density of the r.v. T .

Let $h(\alpha, |\rho|)$ be the $(1 - \alpha)100^{\text{th}}$ percentile of the random variable

$$TP\{\max[|X_1|, |X_2|] < h(\alpha, |\rho|)\} = 1 - \alpha.$$

Let $k(\alpha, |\rho|)$ the $(1 - \alpha)100^{\text{th}}$ percentile of the random variable V

$$P\{\min[|X_1|, |X_2|] > k(\alpha, |\rho|)\} = 1 - \alpha.$$

Using the exact distributions of statistics T and V , the percentiles of the two random variables considered have been computed and the values of $k(\alpha, |\rho|)$ for some α and $|\rho|$ are reported by Pollastri and Riva (2012).

The modified testing procedure by Pollastri is also two stage procedure and it runs as follows:

Stage 1:

If $T \leq h(\alpha, |\rho|)$, H_0 is accepted and the procedure stops. If $T = |Z_i| > h(\alpha, |\rho|)$ ($i = 1$ or 2), one concludes that $p_i > p_i^*$ or $p_i < p_i^*$ if $Z_i > 0$ or $Z_i < 0$, respectively, and proceeds to stage 2.

Stage 2:

If $V = |Z_j| \leq k(\alpha, |\rho|)$ ($i = 1$ or 2), one concludes that $p_j = p_j^*$. If $V = |Z_j| > k(\alpha, |\rho|)$ then one concludes that $p_j > p_j^*$ or $p_j < p_j^*$ if $Z_j > 0$ or $Z_j < 0$, respectively.

4. DETERMINATION OF SAMPLE SIZE FOR PROBLEM OF TESTING THE PARAMETERS OF A TRINOMIAL DISTRIBUTION

In theory and practice of testing hypothesis, determination of sample size is a very important problem. Here we face the problem of the sample size for testing the parameters of trinomial distribution with Duncan-Pollastri procedure. Sample size must be big enough, in order that obtained results be also statistically significant. It is just as important, however, that the study is not too big, where an effect of little scientific importance is nevertheless statistically detectable. Sample size is important also for economic reasons. An undersized study can be a waste of resources for not having the capability to produce useful results, while an oversized one uses more resources than are necessary.

One of the most popular approaches to sample size determination involves studying the power of a test of hypothesis. In this paper for size determination we propose to use methods which control only the first type error.

In the context of above discussion we propose to choose the minimum sample size (n) for which true value of first type error is very close to assumed by researcher value of first type error (nominal α). As we restrict to type I error, our study concerns only the *Stage 1* of the Pollastri procedure.

For the problem of testing defined by (1) for trinomial distribution with para-

meters: $p_1^*, p_2^*, 1 - (p_1^* + p_2^*)$, n type I error as the incorrect rejection of a true null hypothesis is easy to determine. For each pair (x_1, x_2) of values of the trinomial distribution we can calculate statistic T and the probability:

$$P(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} \cdot (p_1^*)^{x_1} \cdot (p_2^*)^{x_2} (1 - p_1^* - p_2^*)^{n-x_1-x_2}$$

Using definition of type I error the value of the first type error (α^*) for the case considered here is the following:

$$\alpha^* = \sum_{\wedge(x_1, x_2) \text{ in which } T > h(\alpha, |\rho|)} P(X_1 = x_1, X_2 = x_2) \quad (2)$$

Using in formula (2) the exact distribution we are able to obtain the real value of type I error α^* .

In the beginning let's consider the situation when we test all probabilities for trinomial distribution being equal, so the null hypothesis can be written down in the following way:

$$H_0 : \left(p_1 = \frac{1}{3} \right) \cap \left(p_2 = \frac{1}{3} \right).$$

Assuming, for example, $\alpha = 0,05$, calculating $|\rho| = \sqrt{\frac{1}{3} \cdot \frac{1}{3} / \frac{2}{3} \cdot \frac{2}{3}} = \sqrt{\frac{1}{4}} = 0.5$

and using table reported by Pollastri and Riva (2012), the value of $h(\alpha, |\rho|)$ is equal to 2.2111. Than we can determine the real value of type I error (α^*) found according to formula (2).

The Figure 1 presents the relationship between the value of the first type error (α^*) and sample size (n). The value are computed through a R code which consider all the possible outcomes of the trinomial r.v. having fixed parameters. We can observe that with increasing n the values of first type error approach the value fixed by researcher that in the present example is $\alpha = 0.05$.

The problem consists in determining the sample size (n) for which we can accept that the true value of the first type error is close to α assumed by researcher.

We propose to choose the minimum value of n_0 for which:

$$\bigwedge_{n \geq n_0} \frac{|\alpha_n^* - \alpha|}{\alpha} < \delta \quad (3)$$

Parameter δ , that is the acceptable error of precision, is fixed as some appropriate value. Despite the fact that α_n^* fluctuates with increasing n it is still possible to find fulfilling condition (3), as the envelope function of $|\alpha_n^* - \alpha|$ is decreasing to zero.

In the next section we study inequality (3) for different null hypothesis.

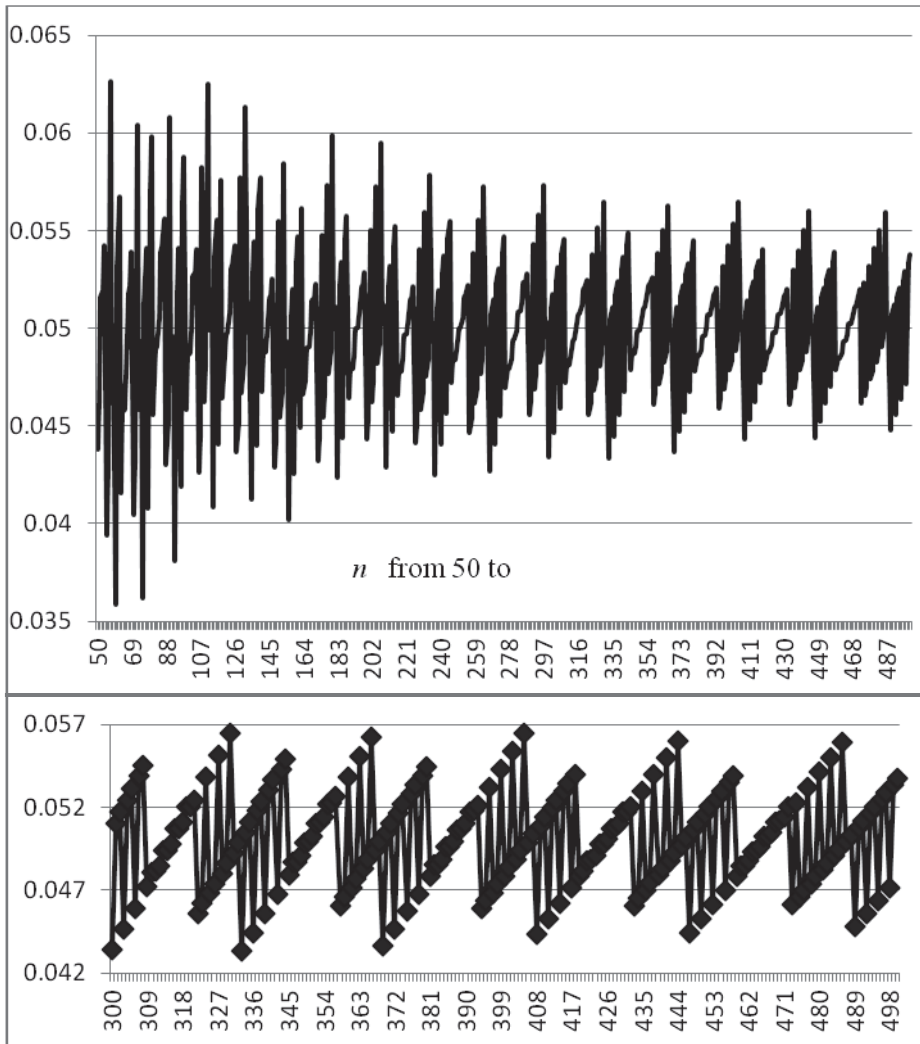


FIGURE 1. - True value of the first type error for,
 $H_0 : (p_1 = 1/3) \cap (p_2 = 1/3), \alpha = 0.05$

5. EMPIRICAL ANALYSIS

In this section we will examine the relationship between type I error, the acceptable error of precision and sample size. Four different cases of null hypothesis will be considered. The first assumes that all probabilities for trinomial distribution are equal, $p_1 = p_2 = p_3$. In subsequent hypotheses we gradually reduce the value of p_1 :

H01: $H_0 : (p_1 = 1/3) \cap (p_2 = 1/3)$

H02: $H_0 : (p_1 = 0.3) \cap (p_2 = 0.6)$ – equivalent to $H_0 : (p_1 = 0.6) \cap (p_2 = 0.3)$

H03: $H_0 : (p_1 = 0.2) \cap (p_2 = 0.7)$ – equivalent to $H_0 : (p_1 = 0.7) \cap (p_2 = 0.2)$

H04: $H_0 : (p_1 = 0.1) \cap (p_2 = 0.8)$ – equivalent to $H_0 : (p_1 = 0.8) \cap (p_2 = 0.1)$

We also take into consideration five different most sensible values of first type errors α : 0.01; 0.02; 0.05; 0.1; 0.15, and seven different values of the acceptable error of precision δ : 0.01; 0.02; 0.04; 0.05; 0.08; 0.1; 0.15.

For each null hypothesis, $H_0 : (p_1 = p_1^*) \cap (p_2 = p_2^*)$, and α the correlation coefficient is calculated using $\rho = -\sqrt{\frac{p_1^* \cdot p_2^*}{(1 - p_1^*) \cdot (1 - p_2^*)}}$ and the corresponding critical value is determined. The results obtained by these calculations are summarized in Table 1, where the critical value $h(\alpha, |\rho|)$ is also reported.

TABLE 1. - *Critical values for different null hypothesis and different values of first type errors*

H_0	ρ	Critical values $h(\alpha, \rho)$				
		$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$
H01	0.5	0.7898	2.5554	2.2111	1.9157	1.7237
H02	0.802	2.7479	2.5068	2.1514	1.8455	1.6461
H03	0.764	2.7541	2.5137	2.1594	1.8544	1.6558
H04	0.667	2.7724	2.5345	2.1842	1.8830	1.6870

The correlation coefficients ρ are essential for finding the values $h(\alpha, |\rho|)$, which are taken from Pollastri and Riva (2012). In what follows, for parameters collected in Table 1 we have made analysis for determining the minimal sample sizes for different cases. The results are reported in Table 2.

First of all we have confirmed what might be expected: with increasing n the first type error approaches near the value fixed by researcher. That can be seen in Table 2: for any fixed (fixed column) the larger n the smaller δ . Although this conclusion is not very surprising and thus much interesting, Table 2 gives one the exact values of minimal sample sizes, what is valuable of practical point of view and might be very useful for planning experiments. The second conclusion, much less obvious than the first one, concerns relationship between value of α and sample size: for any fixed (fixed row) the larger the value of α is, the smaller is the minimal value of n required to obtain assumed degree of accuracy (assumed δ). Note, that this does not hold for each single pair of α (compare, e.g., $\alpha = 0.02$ and $\alpha = 0.05$ for $\delta = 0.10$, H02 case), but there is a very clear negative relationship between value of α and minimal sample size, for fixed δ . Thus, having small sample, we have to accept either large value of I type error (to be more accurate about its actual value, that is, to have small δ) or large degree of uncertainty according to the actual value of α (that is, large δ - large relative difference between assumed and actual I type error). As for comparison between different null hypothesis, the relationship is not so clear.

However, it generally appears, that the larger difference between p_1 and p_2 , the larger sample is required to obtained the same value of type I error and the same level of accuracy of determining its actual value.

TABLE 2. - *Minimal sample size for different null hypothesis, type I errors and acceptable error δ*

H_0	δ	Sample size				
		$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$
H01	0.01	> 3000	> 3000	> 3000	2989	2981
	0.02	> 3000	> 3000	> 3000	2985	2977
	0.04	> 3000	> 3000	> 3000	1947	2932
	0.05	> 3000	> 3000	2887	2062	1750
	0.08	2152	1588	1003	829	667
	0.10	1333	844	622	541	388
	0.15	487	397	241	238	184
H02	0.01	> 3000	> 3000	> 3000	> 3000	> 3000
	0.02	> 3000	> 3000	> 3000	> 3000	> 3000
	0.04	> 3000	2956	2616	2591	1691
	0.05	> 3000	2631	1901	1261	1481
	0.08	1606	1006	751	646	451
	0.10	601	456	476	311	261
	0.15	266	191	231	93	56
H03	0.01	> 3000	> 3000	> 3000	> 3000	> 3000
	0.02	> 3000	> 3000	> 3000	> 3000	> 3000
	0.04	> 3000	> 3000	> 3000	2901	2771
	0.05	> 3000	> 3000	2961	2616	2186
	0.08	2491	1661	1286	1046	656
	0.10	1666	946	436	526	316
	0.15	601	286	231	216	141
H04	0.01	> 3000	> 3000	> 3000	> 3000	> 3000
	0.02	> 3000	> 3000	> 3000	> 3000	> 3000
	0.04	> 3000	> 3000	> 3000	> 3000	> 3000
	0.05	> 3000	> 3000	> 3000	2821	2841
	0.08	2921	2836	1691	1586	1261
	0.10	2736	1716	1341	701	561
	0.15	831	656	426	346	216

6. CONCLUSIONS

In theory of testing hypothesis determination of sample size is a very important problem and the main goal of this paper was to determine the sample size for testing the parameters of a trinomial distribution by using Pollastri procedure – improvement of Duncan procedure. In order to obtain the sample size we propose to use methods which will control the first type error.

The results obtained show (see Table 2) that minimal sample size is very dispersed, from 56 for the largest investigated type I error, the largest investigated acceptable error and H_0 null hypotheses when the probabilities are very close the case $p_1 = p_2 = p_3$, to more than 3000 for a lot of cases. The sample size depends on the null hypothesis. The more hypothesis H_0 differs from $H_0 : (p_1 = 1/3) \cap (p_2 = 1/3)$, the larger is the sample size (c.f., for example, the number of cells in Table 2, in which – for different H_0 hypotheses – value “>3000” appears).

Note that the results obtained are absolutely precise because they are obtained using the exact distribution of all possible outcomes.

ACKNOWLEDGEMENTS

This paper is prepared within a joint research, done in collaboration with prof. Angiola Pollastri, University of Milano-Bicocca. In this place we would like to acknowledge her remarks and suggestions that helped us a lot. Of course, all errors and omissions are ours. We also wish to thank the Referees for the helpful comments.

REFERENCES

- Goodman L.A. (1965). On Simultaneous Confidence Intervals for Multinomial Proportion. *Technometrics*, **7**, 247-254.
- Miller R.G. (1981). *Simultaneous Statistical Inference*. Springer-Verlag, New York.
- Olkin I., Sobel M. (1979). Admissible and Minimax Estimation for the Multinomial Distribution and for K Independent Binomial Distributions. *Annals of Statistics*, **7**(2), 284-290.
- Pollastri A. (2008). Test of hypothesis on the means of a bivariate correlated normal. *Statistics in Transition*, **9**(1), 129-138.
- Pollastri A. (1979). Intervalli di confidenza simultanei asintotici per le probabilità marginali in una tabella 2x2, *Quaderni di Statistica e Matematica applicata alle Scienze Economico – Sociali*, **2**, 45-58.
- Pollastri A., Riva D. (2012). Test for preference change in Marketing or opinion research. *Statistica & Applicazioni*, **X**, 115-130.

Pollastri A., Tornaghi F. (2004). Some properties of the Arctangent Distribution. *Statistica & Applicazioni*, **II**, 3-18.

Sison C.P., Glaz J. (1995). Simultaneous Confidence Intervals and Sample size Determination for Multinomial Proportions. *Journal of the American Statistical Association*, **90**(429), 366-369.

Thompson S.K. (1987). Sample Size for Estimating Multinomial Proportions. *The American Statistician*, **41**(1), 42-46.

Queensbury C.P., Hurst D.C. (1964). Large Sample Simultaneous Confidence Intervals for Multinomial Proportions. *Technometrics*, **6**, 191-195.

Zenga M. (1979). L'impiego della funzione arcotangente incompleta nello studio della distribuzione asintotica dello scarto standardizzato assoluto massimo di una trinomia. *Statistica*, **XXXIX**(2), 269-286.