Colloquium Biometricum 44
2014, 119–126

ON THE TYPE I AND II ERRORS IN THE SEQUENTIAL TESTING OF HYPOTHESIS ABOUT A MEAN IN NORMAL POPULATION

Dorota Domagała, Małgorzata Szczepanik

Department of Applied Mathematics and Computer Science
University of Life Sciences
Głęboka 28, 20-612 Lublin, Poland
e-mails: dorota.domagala@up.lublin.pl,
malgorzata.szczepanik@up.lublin.pl

Summary

In the sequential testing of hypotheses both the type I and II errors are under control. In this paper we focus on testing the hypothesis concerning mean value in normal population versus the simple alternative formulated in terms of standard deviation. The aim of this paper is to compare nominal errors type I and II with the empirical ones evaluated by means of simulation study.

Keywords and phrases: testing hypothesis, sequential test, type I and II errors, simulation

Classification AMS 2010: 62F03, 62L10

1. Introduction

Sequential analysis was developed by Abraham Wald in the early 1940s (Ghosh and Sen; 1991, Marek and Noworol, 1987) for military purposes. In sequential hypothesis testing three possible actions are taken: to accept the null hypothesis, to accept the alternative one or to take more observations, which can be collected one by one (Ghosh and Sen, 1991; Ghosh, 1970; Marek and

Noworol, 1987; Siegmund, 1985; Tweel and all., 1996) or as a whole group (Denne and Jennison, 2000; Jennison and Turnbull, 1990). Each additional observation, or group of observations, can lead to accepting the null hypothesis or the alternative one, or continuing the inspection (Marek and Noworol, 1987; Ghosh and Sen, 1991). Thus the sample size is not fixed in advance.

Sequential analysis can be used in almost every agricultural or biological experiments. Certain limitations of its applications are connected with experiment time; when experiment time is too long it can change the conditions of the examination.

2. Testing of hypothesis about a mean

Let us assume that X is normally distributed with mean μ and known variance σ^2. We are interested in testing the following hypothesis

$$H_0:\ \mu = 0 \tag{2.1}$$

against the simple alternative one formulated in terms of standard deviation

$$H_1:\ \mu = \pm d\sigma$$

where *d* is a positive constant. Hypothesis (2.1) can be verified by means of the test statistic in the following form:

$$i_n = lncosh\left(\frac{d\left|\sum_{i=1}^{n} x_i\right|}{\sigma}\right) - \frac{nd^2}{2} \tag{2.2}$$

This is the logarithm of sequential probability ratio test (SPRT) defined at the stage n by observed values of X, $x_1, \dots, x_n$ successively and independently. The concept of SPRT was developed by Wald as a cumulative product of likelihood ratio as a new data arrive. In practice the logarithm of SPRT is commonly used. Testing of hypothesis (2.1) was conducted by Marek and Noworol (1987) but the authors formulated the alternative hypothesis a little bit differently. They determined mean value as equal to some real constant without the reference to the standard deviation. Thus (2.2) is the modification of the test statistic from Marek and Noworol (1987). The sampling process continues as long as $b < i_n < a$ where *a*, *b* are the real numbers determined by the type I (α) and type II (β) errors as follows

$$a = ln(1-\beta) - ln\alpha,$$
$$b = ln\beta - ln(1-\alpha).$$

These numbers a and b do not depend on the outcome of the observations and, therefore, they can be computed before examination starts. The first time that i_n does not remain between a and b, inspection is terminated and the null hypothesis is accepted if $i_n \leq b$ or the alternative one is accepted if $i_n \geq a$.

The sequential approach for testing hypothesis enables to control both, I and II type errors. Taking into consideration a and b as above means that the type I error will not exceed assuming value of α and the type II error will not exceed assuming value of β.

In this paper the empirical type I and II errors were compared with the nominal ones. All simulation were done in the R program (R Development Core Team, 2008).

3. The type I error study

In order to evaluate the type I error for test statistic i_n 10,000 samples from $N(0,1)$ and $N(0,4)$ distribution were generated. In simulations β was determined as 0.01, 0.05 and 0.1. The nominal type I errors were taken as 0.01, 0.05, 0.1. The empirical type I error for chosen d was calculated as the proportion of accepted alternative hypotheses, i.e. rejected true null hypotheses, to 10,000 runs. The results are presented in tables 1 and 2 with additional information on sample sizes, namely, the minimum (n_{min}), maximum (n_{max}) and average (n_{avg}) sample sizes for which the examination was stopped. The average sample sizes were rounded up to the nearest whole number.

The results enclosed in Tables 1 and 2 show that calculated type I error is always smaller or equal than the nominal one. Moreover, a decrease in obtained α value coincides with an increase in d value. Analysing the sample sizes it is seen that the smaller value of d is, the bigger sample size is needed to stop the examination (in Table 1, for β=0.01, α=0.05, d=0.25, n_{min}=16, n_{max}=1087 and n_{avg}=200). For the bigger values of d (d=2, 2.5, 3) on average several (for $N(0,1)$) or more than ten (for $N(0,4)$) observations are enough.

Table 1. The type I error and chosen sample sizes for N(0,1)

β	α	d	0.25	0.5	0.75	1	1.5	2	2.5	3
		α_{emp}	0.009	0.007	0.007	0.006	0.004	0.004	0.003	0.002
	0.01	$n_{min};n_{max}$	38;972	8;245	4;127	3;62	1;33	1;19	1;14	1;8
		n_{avg}	208	54	25	15	7	5	3	3
		α_{emp}	0.039	0.040	0.031	0.029	0.022	0.016	0.014	0.008
0.01	0.05	$n_{min};n_{max}$	16;1087	4;213	3;103	2;69	1;30	1;15	1;10	1;7
		n_{avg}	200	51	24	14	7	5	3	3
		α_{emp}	0.086	0.079	0.063	0.060	0.044	0.036	0.024	0.017
	0.1	$n_{min};n_{max}$	13;819	3;202	2;108	1;45	1;24	1;17	1;10	1;8
		n_{avg}	188	49	23	14	7	4	3	3
		α_{emp}	0.008	0.009	0.007	0.006	0.004	0.004	0.002	0.002
	0.01	$n_{min};n_{max}$	49;815	11;173	5;90	4;62	1;27	1;17	1;10	1;7
		n_{avg}	144	38	18	11	5	3	3	2
		α_{emp}	0.044	0.032	0.031	0.028	0.021	0.016	0.012	0.010
0.05	0.05	$n_{min};n_{max}$	16;739	5;180	2;88	2;52	1;23	1;14	1;12	1;7
		n_{avg}	135	35	16	10	5	3	2	2
		α_{emp}	0.085	0.077	0.067	0.058	0.042	0.031	0.022	0.015
	0.1	$n_{min};n_{max}$	9;518	3;180	1;69	1;40	1;21	1;12	1;11	1;11
		n_{avg}	126	33	16	10	5	3	3	2
		α_{emp}	0.008	0.008	0.076	0.006	0.005	0.003	0.002	0.002
	0.01	$n_{min};n_{max}$	45;709	9;183	6;77	4;49	2;25	1;16	1;9	1;7
		n_{avg}	115	30	15	9	5	3	2	2
		α_{emp}	0.043	0.041	0.036	0.031	0.021	0.017	0.013	0.010
0.1	0.05	$n_{min};n_{max}$	16;474	5;206	2;72	2;37	1;21	1;10	1;9	1;7
		n_{avg}	107	29	14	8	5	3	2	2
		α_{emp}	0.087	0.072	0.068	0.057	0.048	0.034	0.026	0.017
	0.1	$n_{min};n_{max}$	10;498	3;177	2;52	1;34	1;17	1;14	1;10	1;5
		n_{avg}	100	27	13	8	4	3	2	2

Table 2. The type I error and chosen sample sizes for N(0,4)

β	α	d	0.25	0.5	0.75	1	1.5	2	2.5	3
0.01	0.01	α_{emp}	0.009	0.008	0.007	0.008	0.007	0.006	0.005	0.004
		$n_{min};n_{max}$	220;3644	55;945	25;404	9;272	3;103	3;49	2;42	2;27
		n_{avg}	825	209	95	54	25	15	10	7
	0.05	α_{emp}	0.047	0.044	0.040	0.032	0.029	0.030	0.026	0.020
		$n_{min};n_{max}$	66;3380	19;839	7;359	5;288	2;107	1;61	1;38	1;29
		n_{avg}	778	199	90	52	24	14	10	7
	0.1	α_{emp}	0.091	0.090	0.076	0.074	0.069	0.055	0.048	0.039
		$n_{min};n_{max}$	26;3404	10;766	5;362	3;210	2;94	1;64	1;36	1;30
		n_{avg}	736	189	86	49	23	14	9	7
0.05	0.01	α_{emp}	0.010	0.009	0.007	0.007	0.006	0.006	0.005	0.005
		$n_{min};n_{max}$	105;3329	28;788	17;354	11;197	5;89	4;53	2;44	2;29
		n_{avg}	566	142	65	38	17	11	7	5
	0.05	α_{emp}	0.044	0.045	0.040	0.038	0.032	0.030	0.028	0.022
		$n_{min};n_{max}$	77;2995	19;687	9;311	4;182	2;98	2;67	1;38	1;24
		n_{avg}	533	135	61	36	17	10	7	5
	0.1	α_{emp}	0.096	0.089	0.086	0.079	0.069	0.058	0.051	0.041
		$n_{min};n_{max}$	12;2273	5;594	3;260	2;158	1;68	1;43	1;27	1;19
		n_{avg}	494	127	58	34	16	10	7	5
0.1	0.01	α_{emp}	0.008	0.008	0.009	0.007	0.007	0.004	0.004	0.005
		$n_{min};n_{max}$	169;2624	35;695	14;279	12;164	4;80	3;60	2;35	2;21
		n_{avg}	449	114	52	30	14	9	6	5
	0.05	α_{emp}	0.048	0.044	0.041	0.040	0.033	0.030	0.026	0.024
		$n_{min};n_{max}$	71;2347	17;763	9;272	4;150	3;64	2;49	1;27	1;19
		n_{avg}	417	108	49	29	14	8	6	5
	0.1	α_{emp}	0.092	0.086	0.081	0.082	0.068	0.064	0.053	0.045
		$n_{min};n_{max}$	28;1939	9;473	5;262	3;123	2;55	1;35	1;23	1;24
		n_{avg}	385	99	46	27	13	8	5	4

4. The type II error study

In order to evaluate the type II error the value of α=0.05 was only considered. The following distributions were taken: *N*(1,1), *N*(2,1), *N*(2,4), *N*(4,4). The nominal type II errors were taken as 0.01, 0.05 and 0.1. Values of *d* were different for the data generated from *N*(1,1), *N*(2,4) and from *N*(2,1), *N*(2,4). This choice was connected with the alternative hypothesis formulation. The alternative hypotheses were true when *d* = 1 for *N*(1,1), *N*(2,4) and *d* = 2 for *N*(2,1), *N*(2,4). For each case 10,000 testing of the hypothesis (2.1) were carried out. The type II error was calculated as the ratio of accepted null hypotheses to 10,000 runs. The obtained results are enclosed in Tables 3 and 4.

Table 3. Type II error and chosen sample sizes for distributions N(1,1) and N(2,4); α=0.05

	β	d	0.5	0.75	1	1.25	1.5	2	2.5	3
N(1,1)	0.01	β_{emp}	0.000	0.000	0.005	0.049	0.159	0.478	0.726	0.857
		$n_{min};n_{max}$	7;43	3;48	1;46	1;61	1;59	1;44	1;30	1;21
		n_{avg}	18	11	9	9	9	7	5	4
	0.05	β_{emp}	0.000	0.002	0.037	0.124	0.267	0.570	0.763	0.872
		$n_{min};n_{max}$	2;37	2;43	1;43	1;57	1;44	1;44	1;25	1;16
		n_{avg}	11	9	9	8	7	5	4	3
	0.1	β_{emp}	0.000	0.006	0.067	0.187	0.333	0.607	0.782	0.873
		$n_{min};n_{max}$	2;41	2;41	1;45	1;58	1;35	1;28	1;19	1;12
		n_{avg}	11	9	8	7	6	4	3	2
N(2,4)	0.01	β_{emp}	0.000	0.000	0.006	0.052	0.159	0.479	0.732	0.855
		$n_{min};n_{max}$	3;35	2;44	1;61	1;61	1;70	1;45	1;30	1;19
		n_{avg}	11	10	9	9	9	7	5	4
	0.05	β_{emp}	0.000	0.002	0.031	0.125	0.268	0.563	0.774	0.875
		$n_{min};n_{max}$	3;39	2;45	1;43	1;52	1;46	1;31	1;25	1;15
		n_{avg}	11	9	9	8	7	5	4	3
	0.1	β_{emp}	0.000	0.007	0.062	0.191	0.332	0.611	0.780	0.874
		$n_{min};n_{max}$	3;37	2;42	1;43	1;38	1;39	1;33	1;21	1;14
		n_{avg}	11	9	8	7	6	5	3	2

In the case of *N*(1,1), *N*(2,4) if the parameter *d* is less or equal than 1, or in the case of *N*(2,1), *N*(4,4) is less or equal 2 the empirical type II errors do not exceed the nominal one. In situation when *d* is greater than 1 or 2 in these cases, the calculated β exceed the assuming value of β, e.g. for *N*(1,1), β =0.05 and *d*=3 obtained type II error is equal to 0.872 (Table 3).

Table 4. Type II error and chosen sample sizes for distributions N(2,1) and N(4,4); α=0.05

	β	d	1.5	1.75	2	2.25	2.5	3	3.5	4
N(2,1)	0.01	β_{emp}	0.000	0.001	0.003	0.014	0.031	0.115	0.292	0.491
		$n_{min};n_{max}$	1;15	1;13	1;17	1;15	1;17	1;22	1;20	1;20
		n_{avg}	3	3	3	3	3	3	3	3
	0.05	β_{emp}	0.001	0.009	0.021	0.041	0.087	0.222	0.389	0.566
		$n_{min};n_{max}$	1;12	1;15	1;13	1;13	1;13	1;15	1;18	1;15
		n_{avg}	3	3	3	3	3	3	3	3
	0.1	β_{emp}	0.005	0.017	0.036	0.088	0.149	0.281	0.427	0.578
		$n_{min};n_{max}$	1;12	1;13	1;14	1;15	1;11	1;13	1;14	1;12
		n_{avg}	3	3	3	3	3	3	2	2
N(4,4)	0.01	β_{emp}	0.000	0.001	0.003	0.015	0.035	0.121	0.299	0.491
		$n_{min};n_{max}$	1;11	1;15	1;15	1;15	1;24	1;20	1;25	1;17
		n_{avg}	3	3	3	3	3	3	3	3
	0.05	β_{emp}	0.001	0.009	0.019	0.041	0.085	0.219	0.391	0.554
		$n_{min};n_{max}$	1;16	1;11	1;14	1;15	1;20	1;15	1;17	1;14
		n_{avg}	3	3	3	3	3	3	3	3
	0.1	β_{emp}	0.004	0.017	0.034	0.086	0.143	0.275	0.428	0.599
		$n_{min};n_{max}$	1;11	1;11	1;12	1;16	1;19	1;15	1;12	1;15
		n_{avg}	3	3	3	3	3	3	3	2

This simulation study shows whether the type II error does not exceed the nominal one is depending on formulation of the alternative hypothesis. The alternative hypothesis is formulated as a simple one so it is possible to make a mistake evaluating the μ value. When the alternative hypothesis gives the bigger value of μ than the real one the type II error is greater than the nominal one. So, it is important to formulate the alternative hypothesis carefully and with the deliberation to be able to control the type II error.

5. Discussion

In the case of II type error the simulations conducted in the paper showed that the nominal type I error was not exceeded while testing hypothesis (2.1) by means of the test statistic (2.2). Moreover, in many cases the empirical type I error was much smaller than the nominal one.

The results obtained for the II type error showed that controlling this error is much more difficult than controlling the I type error. The error value depends on the precise formulation of the alternative hypothesis.

Comparing the type I and II errors studies it can be noticed that in the second one on average much smaller sample sizes are needed to stop the process, e.g. in many cases only 3 observations are enough to make the decision about the hypothesis (2.1).

References

Denne J. S., Jennison C. (2000). A group sequential *t*-test with updating of sample size. *Biometrika* 87, 125–134.

Ghosh B.K. (1970). *Sequential test of statistical hypotheses*. Addison-Wesley Publishing Company, Inc.

Ghosh B.K., Sen P.K. (1991). *Handbook of sequential analys*is. Marcel Dekker New York.

Jennison C., Turnbull B.W. (1990). Exact calculation for sequential t, χ^2 and F tests. *Technical Report No.* 888. Cornell University, Ithaca.

Marek T., Noworol C. (1987). *Analiza sekwencyjna w badaniach empirycznych*. PWN Warszawa.

Siegmund D. (1985). *Sequential analysis: tests and confidence intervals*. Springer-Verlag New York.

Tweel van der I., Kaaks R., Noord van P.A.H. (1996). Comparison of one-sample two-sided sequential *t*-tests for application in epidemiological studies. *Statistics in Medicine* 15, 2781–2795.

R Development Core Team (2008). R: *A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria, URL http://www.R-project.org.