Reports of Enlarged Session of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 28, 2014

PROBABILITY OF ERRORS IN SEQUENTIAL METHOD OF BAYESIAN TYPE

Kachiashvili K.

Abstract. Formulae for computation of probability of errors in sequential method of Bayesian type are offered. In particular, some relations between the errors of the first and the second kinds in constrained Bayesian task and in sequential method of Bayesian type depending on the divergence between the tested hypotheses are given. Dependencies of the Lagrange multiplier and the risk function on the probability of incorrectly accepted hypotheses are also presented. Theses results are necessary for computation of errors of made decisions at testing multiple hypotheses using the offered new sequential methods of testing hypotheses. Computation results of some examples confirm the rightness of theoretical analysis.

Keywords and phrases: Constrained Bayesian problem, decision rule, errors type I and type II, hypotheses testing, sequential analysis.

AMS subject classification: 62F15, 62F03.

1. Introduction. A short review of the works devoted to the classical problem of sequential analysis beginning with the fundamental work of Wald is given in (Kachiashvili, 2014). It must be noted that there are many other works where different concepts of multiple sequential comparisons are considered. For example, in (Ghosh et al., 1997) sequential methods about multivariate parameters are described. To test a composite null hypothesis concerning a set of parameters against two tailed alternative hypothesis on the basis of sequentially obtained observations is considered in the works (Betensky, 1996; Jennison and Turnbull, 2000; Wilcox, 2004; Zacks, 2009). More general problem is considered in (De and Baron, 2012; Glimm et al., 2010; Tamhane et al., 2010; Maurer et al., 2011; Bartroff and Lai, 2010). In particular, individual hypotheses are considered about examined set of parameters of sequentially observed random vectors. In the present paper we sample some results for computation of probability of errors in sequential method of Bayesian type offered in (Kachiashvili, 2014; Kachiashvili and Hashmi, 2010) which is quite universal approach allowing to test hypotheses of any types considered in the above mentioned works.

2. The method of sequential analysis of Bayesian type. The sequential analysis method of Bayesian type is offered in (Kachiashvili, 2014; Kachiashvili and Hashmi, 2010). In particular, the following designations are used: R_m^n is the sampling space of all possible samples of m independent n-dimensional observation vectors $\mathbf{x}=(x_1,...,x_n)$; $R_{m,1}^n, R_{m,2}^n, ..., R_{m,S}^n, R_{m,S+1}^n$ are the splitting of R_m^n into $S+1$ disjoint sub-regions such that $R_m^n=\bigcup_{i=1}^{S+1} R_{m,j}^n$. Let $p(\mathbf{x}^1,...,\mathbf{x}^m|H_i)$ be the total probability distribution density of m independent n-dimensional observation vectors; m is sample size. Then $p(\mathbf{x}^1,...,\mathbf{x}^m|H_i)=p(\mathbf{x}^1|H_i)...p(\mathbf{x}^m|H_i)$.

The following decision rule is determined. If the matrix of observation results $\mathbf{x}=(\mathbf{x}^1,...,\mathbf{x}^m)$ belongs to the sub-region $R_{m,i}^n$, $i=1,...,S$, then hypothesis H_i is accepted, and, if $\mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^m)$ belongs to the sub-region $R^n_{m,S+1}$, the decision is not made, and the observations go on until one of the tested hypotheses is accepted.

Let us designate the population of sub-regions of intersections of acceptance regions Γ_i^m of hypotheses H_i $(i = 1, \dots, S)$ in constrained Bayesian task of hypotheses testing with the regions of acceptance of other hypotheses $H_j,\ j = 1, \dots, S;\ j \neq i$, by I_i^m. By $E^n_m = R^n_m - \bigcup_{i=1}^S \Gamma_i^m$, we designate the population of regions of space R^n_m which do not belong to any of hypotheses acceptance regions. Then the hypotheses acceptance regions in the method of sequential analysis of Bayesian type are determined in the following way:

$$R^n_{m,i} = \Gamma_i^m / I_i^m, \quad i = 1, \dots, S; \ \ R^n_{m,S+1} = \left(\bigcup\nolimits_{i=1}^S I_i^m \right) \bigcup E^n_m. \tag{1}$$

Here regions $\Gamma_i^m, I_i^m, E^n_m, i = 1, \dots, S$, are defined on the basis of hypotheses acceptance regions

$$\Gamma_j = \{\mathbf{x} : \sum\nolimits_{i=1, i \neq j}^S p(H_i) p(\mathbf{x}|H_i) < \lambda p(H_j) p(\mathbf{x}|H_j)\}, \quad j = 1, \dots, S, \tag{2}$$

where λ, the same scalar value for all regions, is determined so that the averaged value of incorrectly accepted hypotheses was less than the maximum allowed value α (Kachiashvili, Hashmi and Mueed, 2012a; Kachiashvili et al., 2012b; Kachiashvili and Mueed, 2013).

3. Relationship between the probability of errors of the first and the second kinds in constrained Bayesian task and in sequential method of Bayesian type. Let us designate $\lambda = {}^*\lambda(\mathbf{x}) \in [\lambda_*(\mathbf{x});\ \lambda^*(\mathbf{x})]$ for which there takes place: $\Gamma_i \bigcap \Gamma_j = \emptyset,\ i, j = 1 \dots, S, i \neq j, \bigcup_{i=1}^S \Gamma_i = R^n$. Taking into account this fact, at $\lambda > {}^*\lambda(\mathbf{x})$, the following transformations are true:

$$\begin{aligned} r_\delta &= \sum\nolimits_{i=1}^S p(H_i) \sum\nolimits_{i=1, j \neq i}^S \int_{\Gamma_j} p(\mathbf{x}|H_i) d\mathbf{x} \leq (S-1) \\ &- \sum\nolimits_{i=1}^S p(H_i) \sum\nolimits_{i=1, j \neq i}^S \left(\int_{R^n_{m,i}} p(\mathbf{x}|H_i) d\mathbf{x} + \sum\nolimits_{l=1, l \neq j, l \neq i}^S \int_{R^n_{m,l}} p(\mathbf{x}|H_i) d\mathbf{x} \right). \end{aligned} \tag{3}$$

Let us designate: $\alpha'_{m,i}$ is the probability of no acceptance of hypothesis H_i at its validity in sequential method of Bayesian type after m observations; $\beta'_{m,li}$ is the probability of acceptance of hypothesis H_l at validity of H_i in sequential method of Bayesian type after m observations; α_m is the average probability of the first kind error and $r_{\delta,m}$ is the value of the risk function in constrained Bayesian task obtained on the basis of m sequential observation results.

Then, on the basis of formula (3), we write:

$$r_\delta = \sum\nolimits_{i=1}^S p(H_i) \sum\nolimits_{j=1, j \neq i}^S \int_{\Gamma_j} p(\mathbf{x}|H_i) d\mathbf{x} \geq \sum\nolimits_{i=1}^S p(H_i) \alpha'_{m,i}, \tag{4}$$

and

$$1 - \alpha_m = \sum\nolimits_{i=1}^S p(H_i) \int_{\Gamma_i} p(\mathbf{x}|H_i) d\mathbf{x} \geq 1 - \sum\nolimits_{i=1}^S p(H_i) \alpha'_{m,i}. \tag{5}$$

Let us introduce the designations:

$$\bar{\alpha}_m = \sum\nolimits_{i=1}^S p(H_i) \alpha'_{m,i}, \ \ \bar{\beta}_m = \sum\nolimits_{i=1}^S p(H_i) \sum\nolimits_{j=1, j \neq i}^S \beta'_{m,ji},$$

The quantities $\bar{\alpha}_m$ and $\bar{\beta}_m$are the average probabilities of errors of the first and the second kinds, respectively, in the sequential method of Bayesian type.

Then, on the basis of (4) and (5), the following is true: $r_{\delta,m} \geq \bar{\alpha}_m$ and $\alpha_m \leq \bar{\alpha}_m$; i.e. finally we have: $\alpha_m \leq \bar{\alpha}_m \leq r_{\delta,m}$.

Hence, for calculation of the average probabilities of errors of the first and second kinds in sequential method of Bayesian type, we obtain:

$$\bar{\beta}_m \leq \bar{\alpha}_m, \quad \max\left\{\alpha_m; \frac{r_{\delta,m} - (S-2)\bar{\beta}_m}{(S-1)}\right\} \leq \bar{\alpha}_m \leq r_{\delta,m}. \tag{6}$$

At $\lambda \leq {}^*\lambda(\mathbf{x})$, we have:

$$r_{\delta,m} = \bar{\beta}_m, \ \ \alpha_m = \bar{\alpha}_m. \tag{7}$$

In formulae (6) and (7), $r_{\delta,m}$ and α_m are the values of average risk and significance level of the criterion, respectively, in constrained Bayesian task as a result of its solution after obtaining the next, m th observation result. The ratio between $r_{\delta,m}$ and α_m indicates which formulae (6) or (7) must be used for the estimation $\bar{\alpha}_m$ of $\bar{\beta}_m$ and for the sequential method of Bayesian type.

If it is necessary to know not average but all probability of errors of the first and the second kinds, we can act as follows. After testing hypotheses in the sequential method of Bayesian type on the basis of m sequential observation results, for already determined value λ, we can calculate probabilities $\alpha'_{m,i}$ and $\beta'_{m,li}, i, \ l = 1, \ldots, S, l \neq i$, for example, by the Monte-Carlo method.

4. Relations between the probability of errors of the first and the second kinds in constrained Bayesian task after obtaining sequential observation results. In constrained Bayesian task, for any sample size m , after testing hypotheses, the average probability of errors of the first and the second kinds are calculated as follows:

$$r_{\delta,m} = \sum\nolimits_{i=1}^{S} p(H_i) \sum\nolimits_{j=1,j\neq i}^{S} \int_{\Gamma_j} p(x|H_i) d\mathbf{x}, \ \ \sum\nolimits_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x|H_i) d = 1 - \alpha_m.$$

At $\lambda = {}^*\lambda(\mathbf{x})$ the following is true:

$$r_{\delta,m} = \sum\nolimits_{i=1}^{S} p(H_i)(1 - \int_{\Gamma_i} p(x|H_i) d\mathbf{x}) = 1 - \sum\nolimits_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x|H_i) d\mathbf{x}) = \alpha_m. \tag{8}$$

At $\lambda > {}^*\lambda(\mathbf{x})$, the following takes place

$$r_{\delta,m} > \bar{\beta}_m. \tag{9}$$

At $\lambda < {}^*\lambda(\mathbf{x})$, we have

$$r_{\delta,m} < \bar{\alpha}_m. \tag{10}$$

As in sequential method of Bayesian type after obtaining every next observation result, constrained Bayesian task is solved for all observation results having obtained by the current moment, depending on the value of λ, ratios (8), (9) or (10) between $r_{\delta,m}$ and α_m, i.e. between the values of the risk function and the average probability of rejection of true hypothesis, in constrained Bayesian task solved on the basis of m sequential observation results, remain true.

Using these values of $r_{\delta,m}$ and α_m, with the help of ratios (6) or (9), depending on the value of λ, there the values of average probability of errors of the first and the second kinds in sequential method of Bayesian type are calculated.

To show the practical applicability and usefulness of the results given above, the computation results for two examples are considered in the work. Computation results of these examples completely confirm the rightness of theoretical analysis.

REFERENCES

1. Bartroff J., Lai T.L. Multistage tests of multiple hypotheses. *Communications in Statistics - Theory and Methods*, **39**, (2010), 1597-1607.

2. Betensky R.A. An O'Brien-Fleming sequential trial for comparing three treatments. *Annals of Statistics*, **24**, (4) (1996) 1765-1791.

3. De S., Baron M. Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *Journal of Statistical Planning and Inference*, **142**, (2012), 2059-2070.

4. Ghosh M., Mukhopadhyay N., Sen P.K. Sequential Estimation. *Wiley, New York*, 1997.

5. Glimm, E., Maurer W., Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. *Statistics in Medicine*, **29**, (2010), 219-228.

6. Jennison C., Turnbull B.W. Group Sequential Methods with Applications to Clinical Trials. *Chapman & Hall, Boca Raton, FL.*, 2000.

7. Kachiashvili K.J. The methods of sequential analysis of Bayesian type for the multiple testing problem. *Sequential Analysis*, **33**, (2014), 23-38. DOI: 10.1080/07474946.2013.843318.

8. Kachiashvili K.J., Hashmi M.A. About using sequential analysis approach for testing many hypotheses. *Bulletin of the Georgian Academy of Sciences*, **4**, 2 (2010), 20-25.

9. Kachiashvili K.J., Hashmi M. A., Mueed A. Sensitivity analysis of classical and conditional Bayesian problems of many hypotheses testing. *Communications in Statistics-Theory and Methods*, **41**, 4 (2012), 591-605.

10. Kachiashvili G.K., Kachiashvili K.J., Mueed A. Specific Features of regions of acceptance of hypotheses in conditional Bayesian problems of statistical hypotheses testing. *Sankhya, The Indian Journal of Statistics*, 2012, Volume **74-A, Part 1**, (2012) 112-125.

11. Kachiashvili K.J., Mueed A. Conditional Bayesian task of testing many hypotheses. *Statistics*, **47**, (2013), 274-293. DOI:10.1080/02331888.2011.602681.

12. Maurer W., Glimm E., Bretz F. Multiple and repeated testing of primary, coprimary, and secondary hypotheses. *Statistics in Biopharmaceutical Research*, **3**, 2 (2011), 336-352.

13. Tamhane A.C., Mehta C.R., Liu L. Testing a primary and a secondary endpoint in a group sequential design. *Biometrics*, **66**, (2010), 1174-1184.

14. Wilcox R.R. 2004. Extension of Hochberg's two-stage multiple comparison method. In: Mukhopadhyay, N., Datta, S., Chattopadhyay, S. (Eds.), *Applications of Sequential Methodologies, Marcel Dekker, Inc., New York,* (2004), 371-380.

15. Zacks S. Stage-Wise Adaptive Designs. *Wiley, Hoboken, NJ.*, 2009.

Received 22.05.2014; revised 05.12.2014; accepted 29.12.2014.

Author's addresses:

K. Kachiashvili
Georgian Technical University
77, M. Kostava St., Tbilisi 0175, Georgia

I. Vekua Institute of Applied Mathematics of
Iv. Javakhishvili Tbilisi State University
2, University St., Tbilisi 0186, Georgia
E-mail: k.kachiashvili@gtu.ge