ON THE UPPER BOUND OF THE RISK IN SELECTION OF THE T BEST ITEMS

E. Stoimenova

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
Sofia, BULGARIA
e-mail: jeni@math.bas.bg

Abstract

A fixed sample size procedure for selecting the t best system components is considered. The probability requirement is set to be satisfied under the indifference zone formulation. In order to minimize the average losses from misclassification, we use loss function which is sensitive to the number of misclassifications. The upper bound of the corresponding risk is derived for location parameter distributions. The risk function for the Least Favorable Configuration is derived in an integral form for a large class of distribution functions.

Keywords: data science, risk, least favorable configuration

1 Introduction

Consider the following replacement policy. Fixed number of system components should be replaced by new ones. There are given k competing items and the goal is to select the most reliable t out of them. The true values of the reliability parameters are not known but they are estimated by a singe test. On the basis of test estimates $X_1, \ldots, X_k$, we want to partition the set of item parameters $\theta_1, \ldots, \theta_k$ into two disjoint subsets such that the first subset contains the $k-t$ smallest θ_i, and the second subset contains the remaining t largest θ_i, where $1 \leq t < k$. We denote $X_{(i)}$ the observation corresponding to $\theta_{[i]}$. The parameters in the second subset are called "best".

The natural single stage procedure selects t items with largest parameters on the basis of parameter estimators [4]. The usual approach for this problem is in term of probability of correct selection (PCS). The procedure used should guarantee that the PCS is at least some specified value P^* whenever the true parameter configuration lies in some subset of parameter space. When the monotonicity property on PCS holds, the problem is to find the parameter configuration for which PCS reaches its minimum.

In a decision theoretic approach, the formulation is in term of a loss function and associated risk (average loss). Under a 0-1 loss function, the probability of a correct selection and the risk ρ are connected through, $PCS = 1-\rho$. Thus all decision theoretic formulations in terms of risk can be translated into the "PCS-language". More sensitive loss functions will be discussed in this paper. Their construction is based on a metric for partial rankings.

The optimum properties of the natural decision procedure for selecting the best single population are derived by Bahadur and Goodman [3]. Lehmann [10], Eaton [7], and Alam [1] have extended the results for more general problems and families of distributions. The problem is further discussed by Gupta and Miescke [8], Gupta and

Panchapakesan [9], Ng et al. [2]. The general partition problem stated by Bechhofer [4] is to classify the set of populations into a fixed number of categories. The general case is treated from a decision theoretic point of view by Sobel [12] and Sobel [13]. If the risk is a monotone function, the problem of its evaluating is reduced to a problem of determining the parameter configuration for which the risk is maximal.

2 Decision procedure

2.1 Natural Decision Procedure

Let $X = (X_1, \dots, X_k)$ be estimators of the items unknown reliability parameter vector $\theta = (\theta_1, \dots, \theta_k)$ in a set Θ. It is assumed that X_i has distribution function $F(x - \theta_i)$.

The Natural Decision Procedure divides the coordinate values of θ into two disjoint subsets according to the ranking of the observed vector X. Denote $\Lambda = \{\lambda\}$ the action space for the selection problem containing all partitions $\lambda = (\lambda_1, \lambda_2)$ of $\{1, \dots, k\}$ where λ_1 has $k - t$ elements and λ_2 has t elements.

For each $\lambda = (\lambda_1, \lambda_2) \in \Lambda$, let

$$B_\lambda = \{x \in \chi : \ x_i \le x_j, \ \text{for all } i \in \lambda_1, \ j \in \lambda_2\}.$$

For each $x \in \chi$, let $H(x) = \{\lambda \in \Lambda : x \in B_\lambda\}$ and $n(x)$ be the number of elements in the set $H(x)$ so that $n(x) \ge 1$. The decision rule $\varphi^*_\lambda(x)$ is defined by

$$\varphi^*_\lambda(x) = \begin{cases} 1/n(x) & \text{if} \quad \lambda \in H(x), \\ 0 & \text{if} \quad \lambda \notin H(x). \end{cases}$$

Thus $\varphi^* = \{\varphi^*_\lambda\}_{\lambda \in \Lambda}$.

This decision function φ^* divides the parameters $\theta_1, \dots, \theta_k$ into two ordered subsets. The first subset contains the $k - t$ smallest parameters, and the second subset contains the remaining t largest parameters. The procedure does not state any preferences among members of the same subset.

After the t best parameters have been selected, evaluation of the loss from incorrect partitioning can be made. Let $\theta_{[1]} \le \theta_{[2]} \le \ldots \le \theta_{[k]}$ denote the ordered $\theta_i, \ i = 1, \dots, k$. We assume that it is not known which parameter is associated with $\theta_{[i]}$.

The parameter $\theta_{[k-t]}$ divides the parameters into two ordered subsets so that the parameters $\theta_{[1]}, \dots, \theta_{[k-t]}$ form the first subset, and the parameters $\theta_{[k-t+1]}, \dots, \theta_{[k]}$ form the second subset. When the parameters are selected using their true (unknown) values, we say that a correct selection has been made. This partition of the parameters is called the true one.

We require the usual type of invariant assumptions regarding sample space, Θ, Λ and $F(x; \theta)$. For more explicit treatment of symmetry and invariance see Lehmann [11].

2.2 Loss function

Let $l(\theta, \lambda)$ denote the loss if we terminate selection with action $\lambda \in \Lambda$ when θ is the true value of the parameter vector. Calculate n_{ij} the number of items which are in the i^{th}

category according the true partition and in the j^{th} category according λ $(i, j = 1, 2)$. Define the loss function by

$$l(\theta, \lambda) = \sum_{i=1}^{2} \sum_{j=1}^{2} |c_i - c_j| n_{ij}, \tag{1}$$

where $c_1 = (k-t)(k-t+1)/2$ is the mean of the $k-t$ numbers in the first category and $c_2 = t(2k-t+1)/2$ is the mean of the t numbers in the second category of true partition. Thus the loss function counts the number of misclassified parameters and equals to two times the number of parameters which are among last t largest and are placed in the first subset by action λ.

The motivation for the use of the loss function (1) comes from a metric for partially ranked data induced by Spearman's footrule. Partial rankings from the same type correspond a set of partitions which is a coset space of the permutation group. Metrics on permutation group induce metrics on its cosets spaces which preserve the invariant properties. Several such metrics for partial rankings are constructed by Critchlow [5]. The idea of using metrics on permutations in the decision theoretic formulation has also been mentioned by Diaconis [6].

The function (1) computes the Spearman's footrule distance between two partial rankings using the "pseudo-ranks" c_i and c_j instead the ordinary ranks. The construction of $l(\theta, \lambda)$ implies that the loss function is a right invariant in the sense

$$l(\theta\tau, \lambda\tau) = l(\theta, \lambda).$$

2.3 Expected loss (risk)

Assuming that all partitions of the parameters are equally likely to observe, the risk function for $\varphi \in D$ is

$$\rho(\varphi^*, \theta) = \mathbf{E}[l(\theta, \lambda)] = \sum_{i=1}^{2} \sum_{j=1}^{2} |c_i - c_j| \mathbf{E} N_{ij}, \tag{2}$$

where N_{ij} is the random variable corresponding to n_{ij}.

Theorem 1. *The risk function $\rho(\varphi^*, \theta)$ defined by could be expressed by*

$$\rho(\varphi^*, \theta) = \sum_{m=k-t+1}^{k} \sum_{l=1}^{k-t} \mathbf{P}\{X_{(m)} = X_{[l]}\}. \tag{3}$$

2.4 Preference Zone and Least Favourable Configuration

The indifference Zone approach, proposed by Bechhofer [4], consists of dividing the parameter space into two regions, the so called Preference Zone (PZ) and its complement the Indifference Zone. We discuss distributions with location parameter.

Definition 1. For $0 < \delta < \infty$, the subset $PZ \in \Theta$ defined by

$$PZ = \{\theta \in \Theta : \theta_{[k-t+1]} - \theta_{[k-t]} \geq \delta\} \tag{4}$$

is called the Preference Zone.

The procedure used should guarantee that the risk of decision φ asserted from the observations is at most some specified value P^* whenever θ lies in PZ. So the Preference Zone represents a subset of parameter values where we have a strong preference for a correct selection. The Indifference Zone approach is directed towards the performance of Natural Decision Procedure for configurations in the PZ.

The Least Favourable Configuration (LFC) of the parameters is that one from PZ for which the risk reaches its maximum. For two-category problem with parameter of location we prove in Theorem 3 that

$$LFC : \left\{\theta_{[k]} - \theta_{[k-t+1]} = 0;\ \ \theta_{[k-t+1]} - \theta_{[k-t]} = \delta;\ \ \theta_{[k-t]} - \theta_{[1]} = 0\right\}. \tag{5}$$

The risk $\rho(\varphi^*, \theta)$ for LFC is expressed by (3) is

$$\rho(\varphi^*, LFC) = t \sum_{l=1}^{k-t} \mathbf{P}\{X_{(k)} = X_{[l]}\}.$$

3 Upper Bound of the Risk Function

Define

$$\gamma_{m,s} = \begin{cases} \theta_{[m]} - \theta_{[s]}, & s = 1, \ldots, m-1; \\ \theta_{[s]} - \theta_{[m]}, & s = m+1, \ldots, k. \end{cases}$$

Then the following Theorem holds.

Theorem 2. *For all $m = k-t+1, \ldots, k$, the risk function $\rho(\varphi^*, \theta)$ defined in (2) is a strictly decreasing function in $\gamma_{m,1}, \ldots, \gamma_{m,k-t}$ and nonincreasing in $\gamma_{m,k-t+1}, \ldots, \gamma_{m,k}$ for any parameter configuration from the Preference Zone (4), where*

Theorem 3. *With LFC defined in (5) we have*

$$\rho(\varphi^*, \theta) \leq \rho(\varphi^*, LFC)$$

for all parameter configurations from the Preference Zone (4).

Now, the risk function for LFC is stated for the case $k - t \leq t$.

Theorem 4. *The risk function $\rho(\varphi^*, LFC)$ for LFC defined in (5) is*

$$\sum_{y=0}^{k-t-1} \int \binom{t-1}{y} [F(x)]^y [1-F(x)]^{t-1-y} [$$
$$\sum_{m=0}^{k-t-y-1} \binom{k-t}{m} [F(x+\delta)]^m [1-F(x+\delta)]^{k-t-m}] dF(x).$$

The risk function $\rho(\varphi^*, \theta)$ is decreasing in δ. Thus we can choose δ^* to be the smallest $\delta > 0$ such that $\rho(\varphi^*, LFC) \leq P^*$. For all $\delta > \delta^*$, $\rho(\varphi^*, \theta)$ will be less than P^* for all parameter configurations (4) specified by δ.

4 Acknowledgements

This research was supported by the grant DH02-13 of the Bulgarian National Science Fund.

References

[1] Alam. K. (1973). On a multiple decision rule. *Annals of Statistics*, Vol. **1**, pp. 750-755.

[2] Ng H.K.T. , Balakrishnan N., Panchapakesan S. (2007). Selecting the best population using a test for equality based on minimal Wilcoxon rank-sum precedence statistic. *Methodology and Computing in Applied Probability*, Vol. **9**, pp. 263-305.

[3] Bahadur R., Goodman L. (1952). Impartial decision rules and sufficient statistics. *Annals of Mathematical Statistics*, Vol. **23**, pp. 553-562.

[4] Bechhofer. R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, Vol. **25**, pp. 16-39.

[5] Critchlow. D.E. (1985). *Metric methods for analyzing partially ranked data*. Lecture Notes in Statistics, Springer-Verlag, Berlin-New York.

[6] Diaconis P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics Lecture Notes – Monograph Series, Hayward.

[7] Eaton M.L. (1967). Some Optimum Properties of Ranking Procedures. *Annals of Mathematical Statistics*, Vol. **38**, pp. 124-137.

[8] Gupta S.S., Miescke K.J. (1984). Sequential selection procedures - a decision theoretic approach. *Annals of Statistics*, Vol. **12**, pp.336-350.

[9] Gupta S.S., Panchapakesan S. (2002). *Multiple decision procedures: theory and methodology of selecting and ranking populations*. Wiley, New York.

[10] Lehmann E.L.. (1966). On a theorem of Bahadur and Goodman. *Annals of Mathematical Statistics*, Vol. **37**, pp. 1-6.

[11] Lehmann, E.L., Romano J.P. (2008). *Testing Statistical Hypothesis*. Springer, New York.

[12] Sobel M.J. (1990). Nonparametric loss functions for partitioning problems. *Statistics & Decisions*, Vol. **8**, pp.71-83.

[13] Sobel M.J. (1990). Complete ranking procedures with appropriate loss functions. *Commun. Statist. - Theor. Meth.*, Vol. **19**, pp.4525-4544.