PLISKA
STUDIA MATHEMATICA
BULGARICA

ПЛИСКА
БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

For further information on
Pliska Studia Mathematica Bulgarica
visit the website of the journal http://www.math.bas.bg/~pliska/
or contact: Editorial Office
Pliska Studia Mathematica Bulgarica
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: pliska@math.bas.bg

Pliska Stud. Math. Bulgar. 13 (2000), 195-198

PLISKA
STUDIA MATHEMATICA
BULGARICA

EVALUATING THE RISK IN SELECTION PROBLEMS

Eugenia Stoimenova

Let $X_{1i},\ldots,X_{ki}$ be the scores of the $1^{st},\ldots,k^{th}$ student on test i, and $\overline{X}_{1n},\ldots,\overline{X}_{kn}$ be the sample mean scores of the $1^{st},\ldots,k^{th}$ student corresponding to the first n tests given. It is assumed that the variableX_{ji} is normally distributed with mean μ_j and common known variance 1 for $1 \leq j \leq k$ and $i = 1,2,\ldots$ independently of all other observations. Let $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}$ denote the ordered μ_i, $i = 1,\ldots,k$. We assume that it is not known which parameter is associated with $\mu_{[i]}$. We denote $\overline{X}_{(i)n}$ the sample mean score corresponding to $\mu_{[i]}$ and also $\overline{X}_{[1]n} < \ldots < \overline{X}_{[k]n}$ the ordered sample means.

On the basis of $\overline{X}_{1n},\ldots,\overline{X}_{kn}$, we wish to partition the students into two disjoint subsets such that the first subset contains $k-t$ students with smallest μ_i, and the second subset contains the remaining t students with largest μ_i, where $1 \leq t < k$. The students in the second subset are called "best".

An *action space* can be taken to be the set $\Lambda = \{\lambda\}$ containing all partitions $\lambda = (\lambda_1,\lambda_2)$ of $\{1,\ldots,k\}$ where λ_1 has $k-t$ elements and λ_2 has t elements.

A *decision function* φ is a measurable function on $\mathcal{X}$ such that $\varphi = \{\varphi_\lambda : \lambda \in \Lambda\}$ where $0 \leq \varphi_\lambda \leq 1$ and $\sum_{\lambda\in\Lambda}\varphi_\lambda = 1$. Let $\mathcal{D}$ be the class of decision functions. For $\varphi \in \mathcal{D}$ and $\tau \in S_k$, define $\varphi\tau$ by $(\varphi\tau)_\lambda = \varphi_{\lambda\tau^{-1}}$.

We require the usual type of invariant assumptions regarding $\mathcal{X}$, Θ and Λ. A decision function is right invariant if $\varphi\tau(x) = \varphi(x\tau)$. Denote $\mathcal{D}_I$ the set of right invariant decision functions. For more explicit treatment of symmetry and invariance see Lehmann (1997).

The *Natural Decision Procedure* (Eaton, 1967) divides the students into two disjoint subsets according to the ranking of the observed mean vector $\overline{X}_n$. If n tests are given to each if the k studentsq then we use the procedure of classifying student j (for $1 \leq j \leq k$) into first group iff $\overline{X}_{jn} \leq \overline{X}_{[k-t]n}$

The decision function φ^* divides the students $\mu_1,\ldots,\mu_k$ into two ordered subsets. The first subset contains $k-t$ students corresponding to the $k-t$ smallest components of x, the observed value of $\overline{X}_n$, and the second subset contains the remaining t students. The procedure does not state any preferences among members of the same subset.

The optimum properties of the natural decision procedure for selecting the best single population are derived by Bahadur (1950), and Bahadur and Goodman (1952).

Lehmann (1966), Eaton (1967), and Alam (1973) have extended the results for more general problems and families of distributions. The problem is further discussed by Gupta and Miescke (1984). They give a general proof that the natural terminal decisions, i.e. decisions which are made in terms of largest sufficient statistics, are optimal in terms of the risk, uniformly in parameters under fairly general loss structure.

Loss function. Let $l(\mu, \lambda)$ denote the *loss* incurred if we terminate experimentation with action $\lambda \in \Lambda$ when μ is the true value of the parameter vector. Loss function $l(\mu, \lambda)$ is assumed to be right invariant, that is $l(\mu\tau, \lambda\tau) = l(\mu, \lambda)$. Moreover, $l(\mu, \lambda)$ is assumed to favour actions with large values:

$$l(\mu, \lambda) \leq l(\mu, \eta) \tag{1}$$

for $\mu_i \leq \mu_j$ and $(i, j)\eta = \lambda$.

The optimal properties of Natural Decision Procedure are derived for invariant loss functions satisfying monotonicity property (1).

The parameter $\mu_{[k-t]}$ divides the students into two ordered subsets so that the parameters $\mu_{[1]}, \ldots, \mu_{[k-t]}$ form the first subset, and the parameters $\mu_{[k-t+1]}, \ldots, \mu_{[k]}$ form the second subset. When the students are selected using theoretical means instead of the sample means, we say that a correct selection has been made.

For $\mu \in \Theta$ and $\lambda \in \Lambda$, we define a loss function by

$$l(\mu, \lambda) = \sum_{i=1}^{k} \left[I_{\{\overline{X}_{in} \leq \overline{X}_{[k-t]n}, \mu_i > \mu_{[k-t]}\}} + I_{\{\overline{X}_{in} > \overline{X}_{[k-t]n}, \mu_i \leq \mu_{[k-t]}\}} \right]. \tag{2}$$

The loss function counts the number of misclassified students and equals two times the number of students which are among thet t-best and are placed in the first subset by action λ.

The proposed loss function is an invariant metric on the action space Λ. The analogous loss functions for the general partitioning problem are discussed in Stoimenova (1995) and they are sensitive to the magnitude of misclassification as well.

Expected loss (risk).

Assuming no ties in X and using loss function (2) the risk for φ^* is

$$\rho(\varphi^*, \mu) = \frac{1}{\binom{k}{t}} \sum_{\lambda \in \Lambda} \sum_{i=1}^{k} \left[\mathbf{P}\{\overline{X}_{in} \leq \overline{X}_{[k-t]n}, \mu_i > \mu_{[k-t]}\} + \mathbf{P}\{\overline{X}_{in} > \overline{X}_{[k-t]n}, \mu_i \leq \mu_{[k-t]}\} \right]. \tag{3}$$

Preference Zone and Least Favourable Configuration. The indifference Zone approach, proposed by Bechhofer (1954), consists of dividing the parameter space into two regions, the so called Preference Zone (PZ) and its complement the Indifference Zone.

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

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

is called the Preference Zone of location parameters.

The procedure used should guarantee that the risk of decision φ asserted from the observations is at most some specified value P^* whenever μ lies in PZ. 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

$$LFC: \begin{cases} \mu_{[k]} - \mu_{[k-t+1]} &= 0 \\ \mu_{[k-t+1]} - \mu_{[k-t]} &= \delta \\ \mu_{[k-t]} - \mu_{[1]} &= 0. \end{cases} \tag{5}$$

Theorem 1 (Monotonicity property.)

For fixed $\lambda = (\lambda_1, \lambda_2) \in H(x)$ and $\mu_{[m]}$ $(1 \leq m \leq k)$, the probability $\mathbf{P}\{i \in \lambda_2, \mu_i = \mu_{[m]}\}$ is non-decreasing in $\gamma_s = \mu_{[m]} - \mu_{[s]}$ for $s = 1, \ldots, m-1, m+1, \ldots, k$.

Proof: If invariant property for density function holds, then derivatives of $\mathbf{P}\{i \in \lambda_2, \mu_i = \mu_{[m]}\}$ are positive over $\gamma_{m,1}, \ldots, \gamma_{m,k-t}$ and negative over $\gamma_{m,k-t+1}, \ldots, \gamma_{m,k}$.

Corollary 1 *The risk function $\rho(\varphi^*, \mu)$ defined in (3) is a strictly decreasing function in $\gamma_1, \ldots, \gamma_{k-t}$ and nonincreasing in $\gamma_{k-t+1}, \ldots, \gamma_k$ for any parameter configuration from the Preference Zone (4), where*

$$\gamma_s = \begin{cases} \mu_{[k-t+1]} - \mu_{[s]}, & s = 1, \ldots, k-t; \\ \mu_{[s]} - \mu_{[k-t+1]}, & s = k-t+1, \ldots, k. \end{cases}$$

The result follows by using that all derivatives in the proof of Theorem 1 are strictly positive over $\gamma_1, \ldots, \gamma_{k-t}$ when $\mu \in PZ$ defined by (4).

From Corollary 1 it follows that $\rho(\varphi^*, \mu)$ reaches its maximum for $\gamma_1 = \cdots = \gamma_{k-t} = \delta$, $\gamma_{k-t+1} = \cdots = \gamma_k = 0$, where $\gamma_s = \mu_{[k-t+1]} - \mu_{[s]}$, for $s = 1, \ldots, k-t$, and $\gamma_s = \mu_{[s]} - \mu_{[k-t+1]}$, for $s = k-t+1, \ldots, k$. Thus the upper bound of the risk function (3) is

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

for all parameter configurations from the Preference Zone (4), where LFC is defined in (5).

The risk function for LFC is now derived for the case $k - t \leq t$.

$$t \sum_{y=0}^{k-t-1} \int \binom{t-1}{y} [\Phi(x)]^y [1-\Phi(x)]^{t-1-y} \times \left\{ \sum_{m=0}^{k-t-y-1} \binom{k-t}{m} \left(\Phi(x+\frac{\delta}{\sqrt{n}})\right)^m \left(1-\Phi(x+\frac{\delta}{\sqrt{n}})\right)^{k-t-m} \right\} d\Phi(x).$$

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

We could drop the assumption that the X's are normally distributed and substitute any location family of densitites associated with a location parameter μ. (Stoimenova, 1998). Then a new distribution function shoud be substitute instead of Φ. For some distributions the integrals in risk function for LFC can be solved analytically. The case $t = 2$ is considered for uniform and logistic distributions by Stoimenova (1995).

REFERENCES

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

[2] R. BAHADUR. On a problem in the theory of k populations. *Annals of Mathematical Statistics*, **21** (1950), 362-375.

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

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

[5] R. E. BECHHOFER, J, KIEFER, M. SOBEL. Sequential Identification and Ranking Procedures (with special reference to Koopman-Darmois populations). Chicago University Press, Chicago, 1968.

[6] M. L. EATON. Some Optimum Properties of Ranking Procedures. *Annals of Mathematical Statistics*, **38** (1967), 124-137.

[7] S. S. GUPTA, K, J. MIESCKE. Sequential selection procedures – a decision theoretic approach. *Annals of Statistics*, **12** (1984)., 336-350.

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

[9] E L. LEHMANN. Testing Statistical Hypothesis. Second edition, Chapman & Hall, New York, 1997.

[10] M. SOBEL, M. J. HUYETT. Selecting the best one of several binomial populations. *Bell System Tech. J.*, **36** (1957), 537-576.

[11] E. STOIMENOVA. On the LFC of the natural classification rule under a new loss function. *Statistics & Decisions*, **13** (1995), 39-51.

[12] E. STOIMENOVA. Evaluating the Risk in Selection of the t Best Populations. *Statistics & Decisions*, (submitted).

Institute of Mathematics
Bulgarian Academy of Sciences
Acad. G.Bontchev str., bl. 8
1113 Sofia, Bulgaria
e-mail: jeni@math.bas.bg