Statistica Sinica **7**(1997), 595-606

AN INVARIANT SELECTION RULE FOR MULTI-TREATMENT TRIAL WITH LINEAR PRIOR PREFERENCE

T. Timothy Chen and Cun-Hui Zhang

National Cancer Institute and Rutgers University

Abstract: In clinical trials, asymmetric designs are often used to reflect prior preference of treatments based on factors other than efficacy, such as toxicity and cost. We consider the case where treatments have a linear order of prior preference, and derive likelihood-based invariant procedures which select the most preferred treatment among the equally most effective ones with a preassigned error probability for normal errors, when the prior preference is solely reflected through a set of hypotheses. Extensions are given for the case where different levels of error probabilities are preassigned to the hypotheses. Application to binomial or exponential data with random censoring is through large sample approximation.

Key words and phrases: Invariance, likelihood, selection.

1. Introduction

In clinical trials, asymmetric designs are often used to reflect prior preference of treatments based on factors other than efficacy, such as side effects, costs, and convenience to patients. In a hypothesis-test setting to select between two treatments, this can be done by careful specification of levels of error probabilities and parameter subspaces for hypotheses. In a multi-armed trial, a simultaneous comparison procedure of Dunnett (1955) provides an overall protection for a standard treatment. Without prior preference among the experimental treatments, the procedures of Paulson (1952) and Dunnett (1984) can be used to select the best treatment, while those of Gupta and Sobel (1958) can be used to select a subset of treatments containing the best one, when the standard treatment is rejected. Recently, Chen and Simon (1993, 1994) considered the case where treatments have a linear order of prior preference and derived multi-step procedures which select the most preferred treatment among equally most effective ones with preassigned probabilities for normal errors. Chen and Simon (1993) developed an extension of Dunnett's (1955) one-step many-one test that will give protection to the preferred treatments. Chen and Simon (1994) proposed two multiple-step decision procedures that are similar to the bubble sorting algorithm. For various other multiple-selection procedures, see Bartholomew (1961), Bechhofer and Turnbull (1978), and Thall, Simon and Ellenberg (1988), among others.

In this paper we also consider a selection rule for treatments with linearly ordered prior preference in the normal case via simultaneous comparison which is more efficient than the procedures in Chen and Simon (1993, 1994). The efficiency is measured by the sample size requirement. We shall first derive an invariant likelihood selection rule in Section 2, when the prior preference is solely reflected through a set of hypotheses. Each treatment is given a score based on data, and our procedure chooses the treatment with the highest score. Extensions are given in Section 3 for the case where different levels of error probabilities are preassigned to the hypotheses. Section 4 contains additional discussion which shows the connection of our problem to a change point problem. Section 5 indicates application in cancer clinical trials, where the normal distribution is a good approximation for binomial or exponential data with random censoring due to large sample size.

2. An Invariant Selection Procedure

Let $T_1, \ldots, T_k$ be the treatments of concern. Throughout the paper, we assume that the observations from patients allocated to T_i are summarized by a normal random variable X_i with mean μ_i and a known common variance σ_n^2, and that $X_1, \ldots, X_k$ are independent. These assumptions reflect good approximations to the true distributions of summary statistics in many applications with moderate or large sample sizes and balanced design. Here the subscript in σ_n^2 represents a sample size n for each treatment or simply a design which depends on an index n. For example, if X_i is the average of n independent observations $X_{i1}, \ldots, X_{in}$ with $EX_{ij} = \mu_i$ and $\mathrm{Var}(X_{ij}) = \sigma^2$, then $\sigma_n^2 = \sigma^2/n$. The parameter μ_i represents the treatment effect of T_i, and the variance σ_n^2 can be adjusted according to error probability constraints by choosing an appropriate sample size for each treatment.

Unless otherwise stated, we use the following notation in the sequel: $\mathbf{X} = (X_1, \ldots, X_k)$; $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_k)$; $T_i \prec T_j$ if T_i is preferred to T_j; $a^+ = \max(a, 0)$ for all real numbers a; $a_{i-} = \max_{1 \le j < i} a_j$ and $a_{i+} = \max_{i < j \le k} a_j$ with $\max \emptyset = -\infty$ for all vectors $\mathbf{a} = (a_1, \ldots, a_k)$. For example, $X_{0+} = \max_{1 \le j \le k} X_j$. Also, we use $Z_1, \ldots, Z_k$ to denote k independent standard normal random variables.

For comparison of a standard treatment T_1 with an experimental treatment T_2, the selection is usually done by testing $H_1 : \mu_2 = \mu_1$ against $H_2 : \mu_2 = \mu_1 + \delta$ with error probabilities α and β. The Neyman-Pearson test selects T_2 if $(X_2 - X_1)/(\sigma_n \sqrt{2}) > z_\alpha$ and T_1 otherwise, and the sample size is determined by $\delta/(\sigma_n \sqrt{2}) = z_\alpha + z_\beta$ to satisfy the error probability constraints. We can choose $\delta > 0$ or $\beta \ge \alpha$ to reflect our preference to T_1. When $\alpha = \beta$, we simply select T_2 if and only if (iff) $X_2 - X_1 > \delta/2$. Here we are interested in a one-sided alternative because T_2 is more toxic or costly than T_1.

Consider the problem of selecting among k treatments with a decreasing order of preference, $T_1 \prec \cdots \prec T_k$. A natural extension of the above design for $k = 2$ is a selection rule such that

$$P\{\text{select } T_i \; |H_i\} = 1 - \alpha_i, \quad 1 \leq i \leq k, \tag{2.1}$$

where α_i are preassigned levels, and for $1 \leq i \leq k$ and a given $\delta > 0$

$$H_i : \quad \mu_1 = \cdots = \mu_{i-1} = \mu_i - \delta, \quad \mu_i = \mu_{i+1} = \cdots = \mu_k. \tag{2.2}$$

We consider the simplest case $\alpha_1 = \cdots = \alpha_k = \alpha$ here and the general case in Section 3. Here δ is assumed to be known; this is the amount as a trade-off before a less preferred treatment is selected, or a clinically significant difference we would like to detect in a comparative clinical trial. Our objective is that if the mean of a less preferred treatment is better than means of all more preferred treatments by δ, then we want to select this treatment with a high probability.

2.1. Selection based on likelihood

Consider a general mean vector $(\mu_1, \ldots, \mu_k)$. For $k = 2$, the Neyman-Pearson test is also the most powerful one for $H_1^* : \mu_1 \geq \mu_2$ against the alternative $K_1 : \mu_2 \geq \mu_1 + \delta$ with preassigned error probabilities α and β. The set $\{(\mu_1, \mu_2) : 0 < \mu_2 - \mu_1 < \delta\}$ is considered as the indifference zone. The treatment T_1 is the only correct selection in H_1^* and the incorrect one in K_1. A natural extension to the case $k > 2$ is to assert that T_i is the only correct selection when $\boldsymbol{\theta}$ is in the subspace

$$H_i^* : \quad \mu_{i-} \leq \mu_i - \delta, \quad \mu_{i+} \leq \mu_i. \tag{2.3}$$

and that T_i is an incorrect selection when $\boldsymbol{\theta}$ is in

$$K_i : \quad \max(\mu_{i-}, \mu_{i+} - \delta) \geq \mu_i. \tag{2.4}$$

The subspaces K_i are instrumental in our derivation of selection rules, which can be viewed as alternatives for H_i or H_i^*, since $\cup_{j \neq i} H_j \subset \cup_{j \neq i} H_j^* \subset K_i$. In fact, (2.3) is a consequence of (2.4) as $H_i^* = \cap_{j \neq i} K_j$. The spaces $\{K_1, \ldots, K_k\}$ also provide more information about the performance of the treatments than $\{H_1, \ldots, H_k\}$ or $\{H_1^*, \ldots, H_k^*\}$. For example, if $\mu_j = \mu_{j-1} + \delta/2$ for $j = 2, 3$, and $k = 3$, then T_1 is an incorrect selection as $\boldsymbol{\theta} \in K_1$ $(\mu_1 \leq \mu_3 - \delta)$, but there is no unique correct selection as $\boldsymbol{\theta} \notin \cup_{i=1}^k H_i^*$.

Let $f(\mathbf{X}, \boldsymbol{\theta})$ be the joint density of $\mathbf{X}$. Then the maximum likelihood for T_i to be an incorrect selection is

$$\lambda_i = \lambda_i(\mathbf{X}) = \max\{f(\mathbf{X}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in K_i\}. \tag{2.5}$$

This motivates a simple selection rule

$$d(\mathbf{X}) = i \qquad \text{iff} \qquad \lambda_i = \min_{1\le j\le k} \lambda_j, \tag{2.6}$$

where $d(\mathbf{X}) = i$ means selecting T_i. Rule (2.6) essentially chooses the safest action in the sense that $T_{d(\mathbf{X})}$ is least likely to be an incorrect selection given the data.

Since $f(\mathbf{X}, \boldsymbol{\theta})$ is a decreasing function of $\|\mathbf{X} - \boldsymbol{\theta}\| = \{\sum_{j=1}^{k}(X_j - \mu_j)^2\}^{1/2}$, the maximum in (2.5) is reached at $\boldsymbol{\theta} = \mathbf{X}$ for $\mathbf{X} \in K_i$, and at

$$\mu_i = (X_i + X_i^*)/2,\ \mu_{j_i} = X_{j_i} + (X_i - X_i^*)/2,\ \mu_j = X_j \ \ \forall j \notin \{i, j_i\},$$

for $\mathbf{X} \notin K_i$, where $X_i^* = \max(X_{i-}, X_{i+} - \delta)$, j_i is such that $X_{j_i} = X_{i-}$ if $X_{i-} \ge X_{i+} - \delta$ and $X_{j_i} = X_{i+}$ otherwise. This is clear when $i = k$, and we may simply shift each of $X_{i+1}, \ldots, X_k$ by δ for $1 \le i < k$. The minimum of $\|\mathbf{X} - \boldsymbol{\theta}\|$ over $\boldsymbol{\theta} \in K_i$ is $(X_i - X_i^*)^+/\sqrt{2}$. Thus, the selection rule (2.6) can be written as

$$d(\mathbf{X}) = i \qquad \text{iff} \qquad S_i = \max_j S_j \tag{2.7}$$

with

$$S_i = S_i(\mathbf{X}) = X_i - X_i^* = X_i - \max(X_{i-}, X_{i+} - \delta). \tag{2.8}$$

We call S_i the score of T_i.

2.2. Invariance

We discuss the invariance properties of our selection rule (2.7). The family of density functions $f(\mathbf{X}, \boldsymbol{\theta})$ is invariant under the groups of transformations

$$g_{c,j} : \mathbf{X} \to (X_{k-j+1} - \delta + c, \ldots, X_k - \delta + c, X_1 + c, \ldots, X_{k-j} + c),$$

$$\bar{g}_{c,j} : \boldsymbol{\theta} \to (\mu_{k-j+1} - \delta + c, \ldots, \mu_k - \delta + c, \mu_1 + c, \ldots, \mu_{k-j} + c),$$

for $-\infty < c < \infty$ and $0 \le j < k$. Define

$$\tilde{g}_j : i \to \tilde{g}_j(i) = \begin{cases} i + j, & \text{if } i + j \le k, \\ i + j - k, & \text{otherwise.} \end{cases}$$

Then, $\bar{g}_{c,j}(H_i) = H_{\tilde{g}_j(i)}$ and $\bar{g}_{c,j}(K_i) = K_{\tilde{g}_j(i)}$. The selection problem is invariant if we define the loss function $L(\boldsymbol{\theta}, d) = I\{\boldsymbol{\theta} \in K_d\}$, as $L(\bar{g}_{c,j}(\boldsymbol{\theta}), \tilde{g}_j(d)) = L(\boldsymbol{\theta}, d)$ for all $\boldsymbol{\theta}$ and $1 \le d \le k$. A selection rule $d(\mathbf{X})$ is invariant if $d(g_{c,j}(\mathbf{X})) = \tilde{g}_j(d(\mathbf{X}))$ (cf. Ferguson (1967), pages 143-159).

Proposition 1. *The selection rule* (2.6)*, or equivalently* (2.7)*, is invariant under the above loss function and groups of transformations.*

Proof. By (2.8) $S_i(g_{c,j}(\mathbf{X})) = S_{\tilde{g}_j(i)}(\mathbf{X})$, so that $d(g_{c,j}(\mathbf{X})) = \tilde{g}_j(d(\mathbf{X}))$.

Corollary. *For the selection rule* $d(\mathbf{X})$ *in* (2.7), $P\{d(\mathbf{X}) = i\}$ *is a constant on* $\boldsymbol{\theta} \in H_j$ *for every* $1 \le j \le k$, *and* $P\{d(\mathbf{X}) = i \mid H_1\} = P\{d(\mathbf{X}) = \tilde{g}_j(i) \mid H_{1+j}\}$, $\forall 1 \le j < k$. *In particular,* $P\{\textit{select } T_i \mid H_i\}$, $1 \le i \le k$, *are all equal.*

Proof. Since $d(\mathbf{X})$ is invariant, it is a function of statistics $\mathbf{Y} = \mathbf{Y}(\mathbf{X}) = (Y_2, \ldots, Y_k)$, $Y_i = Y_i(\mathbf{X}) = X_i - X_1$, which have a fixed joint distribution under H_j.

The rule (2.7) has preassigned probability of correct selection (2.1) with constant $\alpha_i = \alpha$, if σ_n^2 is small enough such that

$$P\{\text{select } T_1 | H_1\} = 1 - \alpha. \tag{2.9}$$

2.3. Determination of sample size

Since the variance σ_n^2 can be adjusted by choosing a right sample size for each treatment, our problem is to determine the value of σ_n^2 for which (2.9) holds. We present a table for this purpose and compare our selection rule with the multi-step rule of Chen and Simon (1993).

Proposition 2. *The error probability* α *in* (2.9) *is a decreasing function of* $\tau = \delta/(\sigma_n\sqrt{2})$.

Proof. Let $Z_i = (X_i - \mu_i)/\sigma_n$, $1 \le i \le k$. If H_1 is true, then by (2.8)

$$S_i(\mathbf{X})/\sigma_n = Z_i + \tau\sqrt{2} - \max\left(Z_1 + \tau\sqrt{2}, \ldots, Z_{i-1} + \tau\sqrt{2}, Z_{i+1}, \ldots, Z_k\right),$$

so that $\{S_1(\mathbf{X}) - S_i(\mathbf{X})\}/\sigma_n$ are all increasing functions of τ for fixed realizations of $Z_1, \ldots, Z_k$.

Since $\alpha = \alpha(\tau)$ is a decreasing function, its inverse $\tau = \tau(\alpha)$ exists and can be used to determine σ_n^2 by setting $\delta/(\sigma_n\sqrt{2}) = \tau(\alpha)$. If $\sigma_n^2 = \sigma^2/n$, then the sample size is $n = 2\tau^2\sigma^2/\delta^2$. Some values of $\tau(\alpha)$ are listed in Table 1. The algorithms of Schervish (1984) and Genz (1992) were used in the computations.

Table 1. Selected values of $\tau = \tau(\alpha)$.

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
$k = 3$	2.9901	3.6279	4.1885	4.8508
$k = 4$	3.1779	3.7766	4.3072	4.9391
$k = 5$	3.2809	3.8569	4.3596	4.9432

It appears that the selection rule (2.7) is more efficient than that of Chen and Simon (1993) in the sense that the τ values here are smaller than theirs. For k =3, 4, and 5, their values of $\tau(0.1)$ are respectively 3.004, 3.220, and 3.360.

Note that our δ is their δ/σ and our k is their $k+1$, but our τ is comparable with theirs.

The following alternative description of the selection rule (2.7) is useful for calculation and approximation of error probabilities. Define $i_1 = \min(i \geq 1 : X_i > X_{0+} - \delta)$ and $i_{j+1} = \min(i \geq i_j : X_i > X_{i_j})$ for $j \geq 1$ until the maximum $X_{0+} = X_{i_m}$ is reached at some $j = m$. Then, by (2.8) $S_i > 0$ iff $i \in \{i_1, \ldots, i_m\}$. Furthermore, for $m \geq 2$

$$S_{i_1} = X_{i_1} - X_{0+} + \delta, \quad S_{i_j} = X_{i_j} - X_{i_{j-1}},\ 1 < j \leq m. \tag{2.10}$$

As a consequence, we have

$$\sum_{i=1}^{k} S_i^+ = \sum_{j=1}^{m} S_{i_j} = \delta \tag{2.11}$$

for $m \geq 2$, and $S_{i_1} \geq \delta$ for $m = 1$. Thus, T_i is selected if $S_i > \delta/2$, and for $m \leq 2$, T_i is selected iff $S_i > \delta/2$. For $k = 2$, (2.7) is the Neyman-Pearson test, which selects T_2 iff $S_2 = X_2 - X_1 > \delta/2$, as $m \leq 2$. The following lower bound for the probability of correct selection is an immediate consequence of (2.11) and Proposition 1.

Proposition 3. *Let the selection rule be defined by* (2.7)*. Then* $P\{d(\mathbf{X}) = i\} \geq P\{S_i > \delta/2\}$*. In particular,* $P\{d(\mathbf{X}) = i | H_i\} \geq P\{Z_1 - \max_{2\leq j\leq k} Z_j > -\delta/(2\sigma_n)\}$*, where* $Z_1, \ldots, Z_k$ *are independent standard normal random variables.*

This proposition implies the asymptotic consistency of our selection rule (2.7) in the sense of

Corollary. *Let* H_i^* *and* K_i *be given by* (2.3) *and* (2.4) *respectively. Then,* $P\{d(\mathbf{X}) = i\} \to 1$ *as* $\sigma_n^2 \to 0$ *for all* $\boldsymbol{\theta} \in H_i^*$*, and* $P\{d(\mathbf{X}) = i\} \to 0$ *as* $\sigma_n^2 \to 0$ *for all* $\boldsymbol{\theta} \in K_i$*.*

Let $\tau^* = \tau^*(\alpha)$ be the solution of $P\{Z_1 - \max_{2\leq j\leq k} Z_j > -\tau^*/\sqrt{2}\} = 1-\alpha$. Then, we have an upper bound $\tau(\alpha) \leq \tau^*(\alpha)$. In Section 4, we find a lower bound for $\tau(\alpha)$ via a change-point problem. Table A.1 of Gibbons, Olkin, and Sobel (1977), page 400 can be used to find the values of $\tau^*(\alpha)$ for various α and k. Note that the τ in their Table A.1 is our $\tau^*/\sqrt{2}$ and $z_{\alpha,k}\sqrt{2}$ defined in (3.3) below. For example, when $k = 2$ and $\alpha = 0.05$, $z_{\alpha,k} = 1.645$ and $\tau^* = 2z_{\alpha,k} = 3.290$, while the value in their Table A.1 is $2.326 = 1.645\sqrt{2}$.

3. Extensions

In this section, we consider error probability constraint (2.1) with unequal α_i. For example, when T_{i_0} is a standard treatment for some $1 \leq i_0 \leq k$ (e.g. $i_0 = 1$), we may want α_{i_0} to be smaller than other preassigned levels in (2.1).

Suppose $T_1 \prec \cdots \prec T_k$. We construct selection rules satisfying

$$P\{ \text{ select } \ T_i \ |H_i\} \approx 1 - \alpha_i, \quad 1 \le i \le k, \tag{3.1}$$

for some preassigned levels α_i, where H_i are given in (2.2). Our primary interest is the case where for some α, β, and $1 \le i_0 \le k$

$$\alpha_{i_0} = \alpha, \quad \alpha_i = \beta > \alpha \ \text{ for } \ i \ne i_0. \tag{3.2}$$

A simple modification of (2.7) is to add some constants c_i to the scores defined by (2.8). Let $z_{\alpha,k}$ be the solution of

$$P\left\{Z_1 - \max_{2\le j\le k} Z_j \ge -z_{\alpha,k}\sqrt{2}\right\} = 1-\alpha, \tag{3.3}$$

where $Z_j = (X_j - \mu_j)/\sigma_n$ are independent standard normal random variables. By (2.8) and Proposition 1, we can easily see that

$$P\left\{S_i + z_{\alpha_i,k}\sqrt{2}\sigma_n \ge \delta|H_i\right\} = 1-\alpha_i. \tag{3.4}$$

This leads to the selection rule

$$d(\mathbf{X}) = d(\mathbf{X}, \alpha_1, \ldots, \alpha_k, \tau) = i \qquad \text{iff} \qquad \hat{S}_i = \max_j \hat{S}_j \tag{3.5}$$

with $\tau = \delta/(\sigma_n\sqrt{2})$ and the scores

$$\hat{S}_i = \hat{S}_i(\mathbf{X}) = S_i(\mathbf{X}) + z_{\alpha_i,k}\sqrt{2}\sigma_n, \quad 1 \le i \le k, \tag{3.6}$$

where S_i and $z_{\alpha_i,k}$ are given by (2.8) and (3.3) respectively. If $\max_{j\ne i}\hat{S}_j$ is close to δ with high probability for suitable sample sizes, (3.1) is a consequence of (3.4) when (3.5) is used. For $k = 2$, $z_{\alpha,2}$ is the same as the usual z_α, while Table A.1 of Gibbons, Olkin, and Sobel (1977), page 400 can be used to find the value of $z_{\alpha,k}$. For details, see the discussion at the end of Section 2.3.

Consider the case

$$\delta \ge \max_{1\le i<j\le k}\left|z_{\alpha_i,k} - z_{\alpha_j,k}\right|\sqrt{2}\sigma_n. \tag{3.7}$$

By (3.6), this condition holds if and only if the selection rule (3.5) has the property that $d(\mathbf{X}) = i$ implies $S_i + \delta > \max_{j\ne i} S_j$ for all $\mathbf{X}$. For example, T_i is selected in this case if $S_i > 0$ and $m = 1$ by (2.11). A great part of the probability of correct selection under H_i is captured in the event

$$2\hat{S}_i \ge \tau_i^*\sqrt{2}\sigma_n + \delta, \tag{3.8}$$

where $\tau_i^* = z_{\alpha_i,k} + \max_{j\neq i} z_{\alpha_j,k}$. Clearly, (3.7) and (3.8) imply $S_i > 0$ by (3.6). This gives $d(\mathbf{X}) = i$ under (3.8) if $m = 1$, where m is as in (2.11). For $m \geq 2$, $\delta = \sum_j S_j^+$ by (2.11), so that (3.7) and (3.8) further imply

$$\begin{aligned}\hat{S}_i &\geq \tau_i^* \sqrt{2}\sigma_n + \delta - \hat{S}_i \\ z_{\alpha_i,k}\sqrt{2}\sigma_n &= \max_{j\neq i} z_{\alpha_j,k}\sqrt{2}\sigma_n + \sum_{j\neq i} S_j^+ \\ &\geq \max_{j\neq i} \hat{S}_j.\end{aligned}$$

Thus, (3.8) implies $d(\mathbf{X}) = i$ in all the cases under the assumption (3.7). This fact and (3.4) give immediately

Proposition 4. *Let $d(\mathbf{X})$ be given by* (3.5) *with $\hat{S}_i$ in* (3.6), *and τ_i^* be as in* (3.8). *Suppose* (3.7) *holds. If $\delta/(\sigma_n\sqrt{2}) \leq \tau_i^*$, then*

$$\begin{aligned}P\{d(\mathbf{X}) = i|H_i\} = 1 - \alpha_i &+ P\left\{\max_{j\neq i} \hat{S}_j < \hat{S}_i < \tau_i^*\sigma_n/\sqrt{2} + \delta/2|H_i\right\} \\ &- P\left\{\delta < \hat{S}_i < \tau_i^*\sigma_n/\sqrt{2} + \delta/2|H_i\right\}.\end{aligned}$$

If $\delta/(\sigma_n\sqrt{2}) \geq \tau_i^*$, then

$$\begin{aligned}P\{d(\mathbf{X}) = i|H_i\} = 1 - \alpha_i &+ P\left\{\max_{j\neq i} \hat{S}_j < \hat{S}_i < \tau_i^*\sigma_n/\sqrt{2} + \delta/2|H_i\right\} \\ &+ P\left\{\tau_i^*\sigma_n/\sqrt{2} + \delta/2 < \hat{S}_i < \delta|H_i\right\}.\end{aligned}$$

Consequently, $P\{\text{select } T_i|H_i\} \geq 1 - \alpha_i$ for all $1 \leq i \leq k$, provided that $\delta/(\sigma_n\sqrt{2}) \geq \max_{1\leq i\leq k} \tau_i^*$.

Remark. If (3.2) holds, then $\tau_i^* = z_{\alpha,k} + z_{\beta,k}$ for all $1 \leq i \leq k$.

The sample size will be chosen such that (3.1) is satisfied. For $k = 2$, (3.5) is the Neyman-Pearson test, which selects T_2 iff $(X_2 - X_1)/(\sigma_n\sqrt{2}) > z_{\alpha_1}$ with $\tau = \delta/(\sigma_n\sqrt{2}) = z_{\alpha_1} + z_{\alpha_2}$. For $k > 2$, our numerical experience indicates that

$$\delta/\left(\sigma_n\sqrt{2}\right) = \tau(\alpha_1, \ldots, \alpha_k) \approx \tau^{(a)} = (2k)^{-1}\sum_{i=1}^{k}\left\{\tau(\alpha_i) + \max_{j\neq i}\tau(\alpha_j)\right\} \tag{3.9}$$

provides reasonable approximation in (3.1), where $\tau(\alpha)$ is as in Table 1. If (3.2) holds, then $\tau^{(a)} = \{\tau(\alpha) + \tau(\beta)\}/2$. The computation of $\tau(\alpha)$ is essentially as difficult as that of Table A.1 of Gibbons, Olkin, and Sobel (1977) or the tables in Chen and Simon (1993). Some simulation results for (3.5) are summarized in Table 2.

Table 2. Simulation results for unequal α_i.

	P_1	P_2	P_3
$d(\mathbf{X})$, $\tau = 3.309$	0.9499	0.9185	0.9086
$d(\mathbf{X})$, $\tau = 3.272$	0.9456	0.9187	0.9007
$d^{(cs)}(\mathbf{X})$, $\tau = 3.272$	0.9479	0.9039	0.8995

Here $k = 3$ and the α_i satisfy (3.2) with $\alpha_1 = 0.05$ and $\alpha_2 = \alpha_3 = 0.10$. The values of $\delta/(\sigma_n\sqrt{2}) = \tau$ are the $\tau^{(a)} = 3.309$ of (3.9) in row 1 and the $\tau = 3.272$ of Chen and Simon (1993) in rows 2 and 3. The $d(\mathbf{X})$ is given by (3.5), while $d^{(cs)}(\mathbf{X})$ is the selection rule of Chen and Simon (1993). The P_i are the probabilities of correct selection under H_i, $1 \leq i \leq 3$. Each entry is based on 10,000 computer simulations. The first row of Table 2 shows that (3.1) holds approximately for (3.5) with the sample size approximation (3.9). The second and third rows show that (3.5) performs slightly better than $d^{(cs)}(\mathbf{X})$.

The error probability in (3.1) can be made exact by using scores $S_i + c_i$ for some suitable constants c_i. For this purpose, Proposition 4 suggests the recursion:

$$\begin{aligned}
\beta_i^{(m+1)} &= 1 - P\{d(\mathbf{X}, \alpha_1^{(m)}, \ldots, \alpha_k^{(m)}, \tau^{(m)}) = i | H_i\}, \\
\alpha_i^{(m+1)} &= \alpha_i^{(m)} + \alpha_i - \beta_i^{(m+1)}, \\
\tau^{(m+1)} &= \tau^{(m)} + \rho \left(\sum_{i=1}^{k} \beta_i^{(m+1)} - \sum_{i=1}^{k} \alpha_i \right)
\end{aligned}$$

with the initialization $\alpha_i^{(0)} = \alpha_i$ and a suitable constant ρ, where $d(\mathbf{X}, \alpha_1, \ldots, \alpha_k, \tau)$ is the selection rule given by (3.5) with the scores (3.6) and $\delta/\sigma_n = \tau\sqrt{2}$.

4. Connection to a Change Point Problem

Suppose we are only interested in the parameter space $\cup_{i=1}^{k} H_i$, where H_i are given in (2.2). Then, our selection problem becomes a change point problem in the sense that the mean changes from some unknown μ to $\mu + \delta$ at change point i under H_i. See for example Hinkley (1970). By the discussion in Section 2.2, this change point problem is invariant, and invariant selection rules are functions of statistics

$$\mathbf{Y} = \mathbf{Y}(\mathbf{X}) = (Y_2, \ldots, Y_k), \quad Y_i = Y_i(\mathbf{X}) = X_i - X_1,$$

which have a fixed joint distribution under H_1. Thus, an optimal invariant selection rule exists and maximizes $P\{\text{select}\, T_i \mid H_i\}$.

The probability density function of $\mathbf{Y}$ under H_1 is

$$f_1(\mathbf{y}) = f_1(y_2, \ldots, y_k) = (2\pi\sigma_n^2)^{-(k-1)/2} k^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} (y_i - \bar{y})^2 \right\}$$

with $y_1 = 0$ and $\bar{y} = \sum_{i=1}^k y_i/k$. If $d(\mathbf{X})$ is invariant, then $d(g_{c,j}(\mathbf{X})) = \tilde{g}_j(d(\mathbf{X}))$. In other words, $d(\mathbf{X}) = i$ implies $d(g_{c,k+1-i}(\mathbf{X})) = 1$ for all c. Since $d(\mathbf{X})$ can only be 1 at one of these k points $\{\mathbf{Y}(g_{c,j}(\mathbf{X})), 0 \le j < k\}$, the optimal invariant selection rule is

$$d^{(cp)}(\mathbf{X}) = i \quad \text{iff} \quad f_1(\mathbf{Y}(g_{c,k+1-i}(\mathbf{X}))) = \max_j f_1(\mathbf{Y}(g_{c,k+1-j}(\mathbf{X}))).$$

Since $f_1(\mathbf{Y})$ is decreasing in $\sum(Y_i - \bar{Y})^2 = \sum(X_i - \bar{X})^2$, this rule can be written as

$$d^{(cp)}(\mathbf{X}) = i \quad \text{iff} \quad SS_i = \min_{1 \le j \le k} SS_j$$

with $SS_1 = \sum_{j=1}^k (X_j - \bar{X})^2$ and $SS_i = SS_1(g_{\delta,k+1-i}(\mathbf{X}))$, or equivalently

$$\begin{aligned} SS_i &= \min \left\{ \sum_{j=1}^k (X_j - \mu_j)^2 : (\mu_1, \ldots, \mu_k) \in H_i \right\} \\ &= SS_1 - 2\delta(i-1)(k-i+1)k^{-1} \left(\bar{X}_{i,k} - \bar{X}_{1,i-1} - \delta/2 \right), \ 2 \le i \le k, \end{aligned}$$

where $\bar{X}_{j_1,j_2} = \sum_{j=j_1}^{j_2} X_j/(j_2 - j_1 + 1)$ for $j_1 \le j_2$. Since SS_i is the residual sum of squares when $\mathbf{X}$ is fitted by a vector in H_i via the least squares method, the selection rule $d^{(cp)}(\mathbf{X})$ is also the MLE for the change point.

For the selection rule $d^{(cp)}(\mathbf{X})$, T_1 is selected under H_1 with the probability

$$P \left\{ \max_{2 \le i \le k} \left(\bar{X}_{i,k} - \bar{X}_{1,i-1} - \delta/2 \right) < 0 \right\},$$

which is a decreasing function of $\tau_* = \delta/(\sigma_n \sqrt{2})$. Its inverse function $\tau_*(\alpha)$ gives a lower bound of the function $\tau(\alpha)$ in Table 1, as $d^{(cp)}(\mathbf{X})$ is the optimal invariant rule for $\boldsymbol{\theta} \in \cup_{i=1}^k H_i$. The selection rule $d^{(cp)}(\mathbf{X})$ may not perform well when $\boldsymbol{\theta} \notin \cup_{i=1}^k H_i$. It does not possess the asymptotic consistency property of the Corollary to Proposition 3. For example, if $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3)$ with $\mu_1 > 0$, $\mu_2 < -(\mu_1 + \delta)$, and $\mu_3 = 0$, then $\lim_{\sigma_n \to 0} P\{d^{(cp)}(\mathbf{X}) = 3\} = 1$, while the correct selection is T_1.

5. Applications

Recently, there has been a spate of multi-armed trials with prior preference among treatments. During the period 1991-1993, the National Cancer Institute supported and reviewed several Phase III multi-armed trials in major cancer sites, including three trials in non-small cell lung cancer, two in head and neck cancer, two in colorectal adjuvant, two in rectal adjuvant, one in non-Hodgkin's lymphoma, and two in breast cancer. In most of these cancer clinical trials, the variable of major interest was survival or tumor response, while the prior preference was based on toxicity, quality of life, and cost of administration. Two of these trials are given as examples in Chen and Simon (1993, 1994) where methods to obtain the sample size required for tumor response and survival endpoints are described in detail.

Acknowledgements

This research was partially funded by the National Science Foundation and the Army Research Office. Most part of this research was done while Cun-Hui Zhang visited the Biometric Research Branch, Cancer Therapy Evaluation Program, National Cancer Institute. He would like to thank the host for their hospitality and support.

References

Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B* **23**, 239-281.

Bechhofer, R. E. and Turnbull, B. W. (1978). Two (k+1)-decision selection procedures for comparing k normal means with a specified standard. *J. Amer. Statist. Assoc.* **73**, 385-392.

Chen, T. T. and Simon, R. U. (1993). A multiple-step selection procedure with sequential protection of preferred treatments. *Biometrics* **49**, 753-761.

Chen, T. T. and Simon, R. U. (1994). Extension of one-sided test to multiple treatment trials. *Controlled Clinical Trials* **15**, 124-134.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096-1121.

Dunnett, C. W. (1984). Selection of the best treatment in comparison to a control with an application to a medical trial. In *Design of Experiments, Essays in Honor of Robert E. Bechhofer* (Edited by T. J. Santner and A. C. Tamhane), 47-66. Marcel Dekker, New York.

Ferguson, T. S. (1967). *Mathematical Statistics.* Academic Press, New York.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comp. Graph. Statist.* **1**, 141-149.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* Wiley, New York.

Gupta, S. S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.* **29**, 235-244.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1-17.

Paulson, E. (1952). On the comparison of several experimental categories with a control. *Ann. Math. Statist.* **23**, 239-246.

Schervish, M. J. (1984). Algorithm AS 195. Multivariate normal probabilities with error bound. *Appl. Statist.* **33**, 81-94.

Thall, P. F., Simon, R. and Ellenberg, S. S. (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika* **75**, 303-310.

Biometric Research Branch, National Cancer Institute, Bethesda, Maryland 20892, U.S.A.

E-mail: tchen@helix.nih.gov

Department of Statistics, Hill Center, Busch Campus, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.

E-mail: czhang@stat.rutgers.edu

(Received August 1995; accepted August 1996)