Statistica Sinica **12**(2002), 1133-1144

A MINIMAX TWO-STAGE PROCEDURE FOR COMPARING TREATMENTS: LOOKING AT A HYBRID TEST AND ESTIMATION PROBLEM AS A WHOLE

Wolfgang Bischoff and Frank Miller

University of Karlsruhe, Germany

Abstract: We discuss a problem occuring when a new manufacturing process is investigated. For such applications, it is typical that the relationship among the input variables and the output variables is unknown. After obtaining information on this relationship by experiments, the goal is not only to make statistical inference on this relationship, but to come to a decision for a problem that depends upon it. In our context, the problem is to first test a certain hypothesis against an alternative and then to estimate a certain parameter in case the hypothesis is rejected. The main objective is to develop a solution for the problem as a whole, i.e., a solution of the joint test and estimation problem. We determine the optimal minimax procedure in a certain class by numerical integration. Moreover, we show that the optimal two-stage minimax procedure is better than the optimal one-stage minimax procedure.

Key words and phrases: Comparing treatments, decision theory, hybrid test and estimation problem, minimax procedure, two-stage procedure.

1. Introduction

In many applications one is interested in the relationship among the input variables and the output variables of a process. After obtaining information on this relationship by experiments, the goal is to infer this relationship and to come to a decision for a problem that depends on it. In our context, we have to come to a decision for a hybrid problem, namely first to test a certain hypothesis against an alternative and then to estimate a certain parameter in case the hypothesis is rejected. To be more concrete, we discuss such a practical problem in the following technical, engineering context.

Example 1.1. Assume a company is to begin a manufacturing process. Typically, the production cost of one unit with given input variables (production conditions) can be calculated. On the other hand, the relationship among input variables and quality (characteristic features) of the product is not known. The company has some idea of how selling price depends on different characteristic features of the product. Hence, if the relationship of the production conditions and the characteristic features were known, the company would know how selling price depends on production conditions. Since this relationship is unknown, it is typical to obtain information on it through experiments in which production conditions are systematically varied.

Let T be the set of experimental/special production conditions. We assume the cost function $c : T \to [0, \infty)$ is known and the selling price function $p : T \to [0, \infty)$ is unknown. Thus the selling price $p(t)$ depends on t via the unknown characteristic features of the product under t. The crucial function $g := p - c$, usually unknown since p is. With samples produced under various conditions, the company wants information on the characteristic features as they depend on production condition, thus on p and hence on g. Two problems are of interest.

- Is there a $t_0 \in T$ such that $g(t_0) > 0$? If not, the company will dispense with new production. We test $H_0 : g \leq 0$ against $K : g(t_0) > 0$ for some $t_0 \in T$.
- If K holds, where is g maximized? The company is interested in estimating $\theta := \arg\max_{t \in T} g(t)$.

We develop a solution for the whole problem, for testing and estimation, by introducing a natural loss function, see Section 2.1. Then we can show the intuitive result that a two-stage minimax procedure is superior to a one-stage minimax procedure (see Section 6). Note that for related problems concerning clinical trials, certain tests and estimators have been investigated separately, see for example Siegmund (1993), Coad (1995), Bauer, Bauer, and Budde (1998), Friede, Miller, Bischoff, and Kieser (2000), but not together. Wetherill and Glazebrook (1986) present decision theoretic methods for sequential problems, but they do not focus on the hybrid test and estimation problem considered here.

In Section 2.2, we introduce the experimental model. A one-stage decision procedure is described in Section 3, and its risk function is derived, and probabilities of wrong decisions are presented. Section 4 deals with a two-stage procedure for the experiments and its more complicated risk function. Confidence intervals for the one- and two-stage procedures are developed in Sections 3 and 4, respectively. An example with artificial data illustrates our two-stage design. The results of Sections 3 and 4 are specialized to the normal distribution in Section 5 and the connection to large sample results is discussed there. In our opinion a minimax procedure is to be preferred and, in Section 6, we determine the optimal minimax procedure by numerical integration. If prior information is available, a Bayes procedure can be numerically developed as well. Moreover, we show that the two-stage minimax procedure is better than the one-stage minimax procedure. Technical proofs of the assertions of Sections 3 and 4 are postponed to the Appendix.

2. The Loss Function and the Model

2.1. The loss function for the problem

Let $a = 0$ be the decision against production, $a = 1$ be the decision for production. If H_0 is true and $a = 0$, there is no loss: $L = 0$; if $a = 1$ and the production condition s is chosen, the loss (per unit) is $L = -g(s) \geq 0$. Next assume that K is true. Then if $a = 0$ the company makes no profit instead of the possible profit $g(\theta)$ and we take the loss to be $L = g(\theta)$; if $a = 1$ and the production condition s is chosen, the loss is the difference between $g(\theta)$ and $g(s)$: $L = g(\theta) - g(s)$. Putting the four cases together, $L(g,(a,s)) = \mathbf{1}\{g(\theta) > 0\} \cdot g(\theta) - a \cdot g(s)$.

2.2. The model for the experiments, assumptions and notation

For the results of the experiments, we assume

$$Y(t_i) = g(t_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where n is the total number of experiments, t_i, $i = 1, \ldots, n$, are production conditions and the ϵ_i are independent and identically distributed according to a distribution function F. We first assume that F is an arbitrary, continuous distribution function which is known. In Section 5, we specialize to independent, normally distributed random variables with a common known variance and possibly different mean values. There is asymptotic justification for the assumption of normality there. Usually, for the practical problem described in this paper, it can be assumed that the regression function g is bounded.

Let $t_1, \ldots, t_k$ be the distinct experimental conditions among which the $n \geq k$ experimental conditions can be chosen for the experiments, and let $T = \{t_1, \ldots, t_k\}$. For the sake of simplicity we assume n is a multiple of k.

3. A One-Stage Decision Procedure

If one has no information about g, run $m := n/k$ experiments at every $t \in T$ and let $Y_{t,j}$, $t \in T$, $1 \leq j \leq m$, be the result of the j-th experiment with experimental condition t. In a first step, we estimate $g(t)$ by $\bar{Y}_t = \frac{1}{m} \sum_{j=1}^{m} Y_{t,j}$. We need a decision $(a,s) \in \{0,1\} \times T$ based on the data $Y_{t,j}, t \in T, 1 \leq j \leq m$. For that consider $\xi := \max_{t \in T} g(t) = g(\theta)$, where $\theta = \arg\max_{t \in T} g(t)$. If we know ξ, we choose $a = \mathbf{1}\{\xi > 0\}$. Since ξ is unknown, we consider $\hat{\xi} = \max\{\bar{Y}_t \mid t \in T\}$ and choose $a = \mathbf{1}\{\hat{\xi} > u\}$, where $u \in I\!R$ is chosen suitably. The choice of u is discussed in Section 5. Further, it is natural to choose the experimental condition which corresponds to the largest value of $\{\bar{Y}_t \mid t \in T\}$ as decision for $s \in T$. We denote this estimator for θ by $\hat{\theta}$.

Thus, we obtain a class of decision procedures $(a,s) = (\mathbf{1}\{\hat{\xi} > u\}, \hat{\theta}),\ u \in I\!R$. Note that on one hand, the risk function depends on k, n, and u. These parameters are fixed or can be chosen by the experimenter. On the other hand, we have the unknown parameters g and F. We suppress the dependence of the risk function on k and n and take them as fixed in the sequel. Then the risk function for the above class of decision procedures is given by

$$\begin{aligned} R_F(g,(\mathbf{1}\{\hat{\xi}>u\},\hat{\theta})) &= E_{g,F}(L(g,(\mathbf{1}\{\hat{\xi} > u\},\hat{\theta}))) \\ &= \mathbf{1}\{g(\theta) > 0\} \cdot g(\theta) - E_{g,F}(\mathbf{1}\{\hat{\xi} > u\} \cdot g(\hat{\theta})) \\ &= \mathbf{1}\{g(\theta) > 0\} \cdot g(\theta) - \sum_{s \in T} g(s) \cdot P(\hat{\theta} = s, \hat{\xi} > u). \end{aligned}$$

If there is no risk of confusion, we write P instead of $P_{g,F}$. Denote the distribution of $\bar{Y}_t = \frac{1}{m}\sum_{j=1}^m Y_{t,j}$ by $F_{g(t)}^{(m)}$.

Theorem 3.1. *For the one-stage procedure, the risk function is given by*

$$R_F(g,(\mathbf{1}\{\hat{\xi}>u\},\hat{\theta}))=\mathbf{1}\{g(\theta)>0\} \cdot g(\theta)-\sum_{s\in T} g(s) \cdot \int_u^\infty \prod_{t\in T\setminus\{s\}} F_{g(t)}^{(m)}(z) F_{g(s)}^{(m)}(dz).$$

The proof is in the Appendix.

In the next theorem, we give a confidence interval for the optimal yield $\xi = g(\theta)$. Of even greater practical appeal is an interval I with

$$P_{g,F}(g(\hat{\theta}) \in I) \geq 1 - \alpha \quad \text{for all } g : T \to I\!R. \tag{3.1}$$

Although the latter is formally not a confidence interval we use this language. We derive only one-sided confidence intervals $[l,\infty)$. Confidence intervals of the form $(-\infty,u]$ or $[l,u]$ can be computed along the same lines.

Theorem 3.2. *For the one-stage procedure, a confidence interval for $g(\theta)$ and $g(\hat{\theta})$ of size $1-\alpha$ is given by $[\hat{\xi} - d, \infty)$, where $d = (F_0^{(m)})^{-1}(\sqrt[k]{1-\alpha})$. Here, $(F_0^{(m)})^{-1}$ is the quantile function of the distribution of $\bar{Y}_t$ assuming $g(t) = 0$ for all $t \in T$.*

The theorem can be proved by showing $\inf_{g \in I\!R^T} P_{g,F}(g(\theta) \in [\hat{\xi} - d, \infty)) = P_{0,F}(0 \in [\hat{\xi} - d, \infty))$ and by showing the same equality when θ is replaced by $\hat{\theta}$.

4. A Two-Stage Decision Procedure

A two-stage approach with an adaptive choice of the design for the second stage seems to be a promising alternative to the one-stage procedure. In the first stage, experiments are done to get a first impression about g. In a second stage, only those input variables are used which have given good results in the first stage. In Section 6, we show by numerical computations that minimax risk can be reduced in this way.

Here, let n_1 be the number of experiments at the first stage and $n_2 = n - n_1$ at the second. We assume $m_1 := n_1/k_1$ experiments are carried out at each $t \in T$ at the first stage. With the information from the first stage, we select a subset T_2 of T and carry out $m_2 := n_2/k_2$ experiments at each $t \in T_2$.

It is natural to include those design points (input variables) in T_2 which gave the higher outcomes in the first stage. In particular we take those points into the second stage where the outcome is not less than the maximum outcome minus δ, say, where $\delta \geq 0$ is a fixed constant. For other selection rules, the same methodology is possible with a (perhaps) more complicated analysis.

We modify the notation for the two-stage case. In the first stage, let $Y_{1,t,j}$, $t \in T$, $1 \leq j \leq m_1$, be the result of the j-th experiment with experimental condition t and estimate $g(t)$ by $\bar{Y}_{1,t} := \frac{1}{m_1}\sum_{j=1}^{m_1} Y_{1,t,j}$. Then $T_2 = \{t \in T \mid \bar{Y}_{1,t} \geq \max_{s\in T}\{\bar{Y}_{1,s}\} - \delta\}$. In the second stage, the results of our experiments are $Y_{2,t,j},\ t\in T_2,\ 1 \leq j \leq m_2$. We define the estimator $\hat{\xi}$ by $\hat{\xi} = \max\left\{\frac{m_1}{m_1+m_2}\cdot \bar{Y}_{1,t} + \frac{m_2}{m_1+m_2}\cdot \bar{Y}_{2,t} \mid t \in T_2\right\}$. Note that $\frac{m_1}{m_1+m_2}\cdot \bar{Y}_{1,t} + \frac{m_2}{m_1+m_2}\cdot \bar{Y}_{2,t}$ is equal to the empirical mean $\frac{1}{m_1+m_2}(\sum_{j=1}^{m_1} Y_{1,t,j} + \sum_{j=1}^{m_2} Y_{2,t,j})$ of all observations at t. We estimate θ by $\hat{\theta}$,the $t \in T_2$ yielding the largest value.

We stress the dependence of the estimators $\hat{\xi}$ and $\hat{\theta}$ on δ by denoting $\hat{\xi}_\delta := \hat{\xi}$, $\hat{\theta}_\delta := \hat{\theta}$, and we denote the distribution of $\bar{Y}_{r,t} = \frac{1}{m}\sum_{j=1}^{m} Y_{r,t,j}, r = 1, 2$, by $F^{(m)}_{g(t)}$. The proof of the following theorem can be found in the Appendix.

Theorem 4.1. *The risk function is* $R_F(g, (\mathbf{1}\{\hat{\xi}_\delta > u\}, \hat{\theta}_\delta)) = \mathbf{1}\{g(\theta) > 0\}\cdot g(\theta) - \sum_{s\in T} g(s)\cdot P(\hat{\theta}_\delta = s, \hat{\xi}_\delta > u)$. *Let* $\gamma_{1r} := m_1/(m_1 + n_2/r), \gamma_{2r} := (n_2/r)/(m_1 + n_2/r), r = 1, \ldots, k_1$. *Then, for* $s \in T, u \in \mathbb{R}$,

$$
\begin{aligned}
&P(\hat{\theta}_\delta = s, \hat{\xi}_\delta > u) \\
&= \sum_{r=1}^{k_1} \sum_{a_1\in T} \sum_{\{a_2,\ldots,a_r\}\subseteq T\setminus\{a_1\}} \mathbf{1}\{s = a_j \textit{ for some } j \in \{1, \ldots, r\}\} \\
&\quad \cdot \int_{-\infty}^{\infty}\int_{z_1-\delta}^{z_1}\cdots\int_{z_1-\delta}^{z_1}\int_{(u-\gamma_{1r}z_j)/\gamma_{2r}}^{\infty} \left[\prod_{\substack{i=1\\ i\neq j}}^{r} F^{(n_2/r)}_{g(a_i)}\left(z^* + \frac{m_1}{n_2/r}(z_j - z_i)\right)\right] F^{(n_2/r)}_{g(a_j)}(dz^*) \\
&\quad \cdot \left[\prod_{a\in T\setminus\{a_1,\ldots,a_r\}} F^{(m_1)}_{g(a)}(z_1 - \delta)\right] F^{(m_1)}_{g(a_r)}(dz_r)\cdots F^{(m_1)}_{g(a_2)}(dz_2)\ F^{(m_1)}_{g(a_1)}(dz_1) \qquad (4.1)
\end{aligned}
$$

where the product $\left[\prod_{\substack{i=1\\i\neq 1}}^{1}\dots\right] := \left[\prod_{\emptyset}\dots\right] := 1.$

To clarify the right side of (4.1), note that an arbitrary element a_1 of T is selected in the second of the three sums. Then if $r>1$, every subset of $T\backslash\{a_1\}$ with $r-1$ elements is chosen in the third sum. In the special case $r=1$, the set $\{a_2,\dots,a_r\}$ is empty, the last sum-sign and the integrals corresponding to $a_2,\dots,a_r$, that is $\int_{z_1-\delta}^{z_1}\cdots\int_{z_1-\delta}^{z_1}$ and $F_{g(a_r)}^{(m_1)}(dz_r)\dots F_{g(a_2)}^{(m_1)}(dz_2)$, are left out. For example in the special case $k_1=2$ (two design points in the first stage), the probability $P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u)$ of Theorem 4.1 is (with $\bar{s}$ such that $\{s,\bar{s}\}=T$)

$$\int_{-\infty}^{\infty}\int_{(u-\gamma_{11}z_1)/\gamma_{21}}^{\infty}1F_{g(s)}^{(n_2/1)}(dz^*)F_{g(\bar{s})}^{(m_1)}(z_1-\delta)F_{g(s)}^{(m_1)}(dz_1)$$
$$+\sum_{a_1\in T}\sum_{a_2\in T\backslash\{a_1\}}\mathbf{1}\{s=a_j \text{ for some } j\in\{1,2\}\}$$
$$\cdot\int_{-\infty}^{\infty}\int_{z_1-\delta}^{z_1}\int_{(u-\gamma_{12}z_j)/\gamma_{22}}^{\infty}\left[\prod_{\substack{i=1\\i\neq j}}^{2}F_{g(a_i)}^{(n_2/2)}\left(z^*+\frac{m_1}{n_2/2}(z_j-z_i)\right)\right]F_{g(a_j)}^{(n_2/2)}(dz^*)F_{g(a_2)}^{(m_1)}(dz_2)F_{g(a_1)}^{(m_1)}(dz_1).$$

Note, that the first summand corresponds to $r=1$ and the rest of the expression to $r=2$.

Note further, that the limit of the probability $P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u)$ in Theorem 4.1 for $\delta\to\infty$ is the probability $P(\hat{\theta}=s,\hat{\xi}>u)$ of the one-stage case. This can be seen by expression (A.1) in the proof of Theorem 4.1.

It is worth mentioning that probabilities of wrong decisions ($P(\hat{\theta}_\delta\neq\theta)$, $P(\hat{\xi}_\delta>u)$ if H_0 is true and $P(\hat{\xi}_\delta\leq u)$ if K is true) can be computed by using formula (4.1). Theorem 3.2 can be modified for the two-stage case as follows. The proof is contained in the Appendix.

Theorem 4.2. *For the two-stage procedure, a confidence interval for* $g(\theta)$ *and* $g(\hat{\theta}_\delta)$ *of size* $1-\alpha$ *is given by* $[\hat{\xi}_\delta-d,\infty)$, *where* d *is the solution of* $P_{0,F}(\hat{\xi}_\delta\leq d)=1-\alpha$.

Example 4.3. We illustrate our two-stage procedure with an example. We assume errors are normally distributed with variance 1, that two different production conditions t_1,t_2 are considered ($k_1=2, T=\{t_1,t_2\}$) in the first stage. The total number of observations is $n=40$. $n_1=n_2=20$ observations are made in each stage. The parameters of the procedure u and δ are chosen according to the minimax criterion: $u=-0.01,\delta=0.5$, see Section 6. For the simulation, we assume the true values of the model parameters are $g(t_1)=-0.2, g(t_2)=0.2$. In the first stage, the mean values of our observations were 0.07 and 0.91 for t_1

and t_2, respectively. Since the difference between the two mean values was larger than $\delta = 0.5$, we had to carry out $n_2 = 20$ experiments with production condition t_2 in the second stage. The mean of these twenty observations was 0.29. Thus, we have $\hat{\theta}_\delta = t_2, \hat{\xi}_\delta = 0.50$. The procedure rejects the hypothesis and decides correctly that the production condition t_2 is better than t_1.

Next, we derive the probabilities of wrong decisions. The estimation error is given by $P(\hat{\theta}_\delta \neq \theta) = 1 - P(\hat{\theta}_\delta = \theta, \hat{\xi}_\delta > -\infty) = 0.106$. Generally, the corresponding one-stage procedure has a lower estimation error, since here all observations are used for estimation. The estimation error of the one-stage procedure is only slightly better: 0.103. The type II error of not rejecting the hypothesis (while K is true in this example) is $P(\hat{\xi}_\delta \leq u) = 1 - \sum_{s \in T} P(\hat{\theta}_\delta = s, \hat{\xi}_\delta > u) = 0.142$. For the one-stage minimax procedure (see Section 6), the type II error is 0.163.

The solution d of the equation $P_{0,F}(\hat{\xi}_\delta \leq d) = 1 - \sum_{s \in T} P_{0,F}(\hat{\theta} = s, \hat{\xi}_\delta > d) = 1 - \alpha$ is $d = 0.40$ (with $1 - \alpha = 0.95$). Hence, the confidence interval of Theorem 4.2 for $g(\theta)$ and for $g(\hat{\theta}_\delta)$ is $[\hat{\xi}_\delta - d, \infty) = [0.10, \infty)$.

5. Normally Distributed Errors and Large Sample Results

We specialize the above considerations and formulas by the following assumptions.

Assumption 5.1. (a) The error distribution is given by the $N(0, \sigma^2)$-distribution ($\sigma^2 > 0$). (b) There are two design points in the first stage.

Note that we only need the distribution of the mean of the experiments at each design point and each stage, thus the assumption of normality is approximatively true if the variance of the error distribution exists and m_1 and m_2 are sufficiently large.

The corollary below follows from Theorem 3.1 and Theorem 4.1. The proof is omitted but is available from the authors. The distribution function and the density of $N(0,1)$ are denoted by Φ and ϕ, respectively.

Corollary 5.2. *Suppose Assumption* 5.1 *holds,* $T = \{s, \bar{s}\}$ *and* $u \in \mathbb{R}$ *is arbitrary. Then in the one-stage case* ($n_2 = 0, n := n_1$)*, the joint distribution of* $\hat{\theta}$ *and* $\hat{\xi}$ *is given by*

$$P(\hat{\theta} = s, \hat{\xi} > u) = \int_u^\infty \Phi\Big((z - g(\bar{s}))\frac{\sqrt{n/2}}{\sigma}\Big) \cdot \phi\Big((z - g(s))\frac{\sqrt{n/2}}{\sigma}\Big) \cdot \frac{\sqrt{n/2}}{\sigma} dz. \quad (5.1)$$

In the two-stage case, we have

$$P(\hat{\theta}_\delta = s, \hat{\xi}_\delta > u)$$
$$= \int_{-\infty}^\infty \Phi\Big(-u\frac{m_1 + n_2}{\sigma\sqrt{n_2}} + z_1 \frac{m_1}{\sigma\sqrt{n_2}} + g(s)\frac{\sqrt{n_2}}{\sigma}\Big) \cdot \Phi\Big((z_1 - \delta - g(\bar{s}))\frac{\sqrt{m_1}}{\sigma}\Big)$$

$$\cdot\phi\Big((z_1 - g(s))\frac{\sqrt{m_1}}{\sigma}\Big)\cdot\frac{\sqrt{m_1}}{\sigma}\ dz_1 + \int_{-\infty}^{\infty}\int_{-\delta}^{\delta}\int_{u\frac{m_1+n_2/2}{n_2/2}}^{\infty}\frac{m_1\sqrt{n_2/2}}{\sigma^3}$$
$$\cdot\Phi\Big((z_0 - (z_1+\bar{z})\frac{m_1}{n_2/2} - g(\bar{s}))\frac{\sqrt{n_2/2}}{\sigma}\Big)\cdot\phi\Big((z_0 - z_1\frac{m_1}{n_2/2} - g(s))\frac{\sqrt{n_2/2}}{\sigma}\Big)$$
$$\cdot\phi\Big((\bar{z} + z_1 - g(\bar{s}))\frac{\sqrt{m_1}}{\sigma}\Big)\cdot\phi\Big((z_1 - g(s))\frac{\sqrt{m_1}}{\sigma}\Big)\ dz_0\ d\bar{z}\ dz_1. \tag{5.2}$$

If there is no prior information it seems suitable to choose the minimax criterion. Therefore, we minimize the maximal risk by choosing the parameter $\delta > 0$ and the critical value u such that $\sup_{g\in\mathcal{J}^T} R_F(g,(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta))$ is minimal, where $\mathcal{J}$ is a set that contains the image of g. For practical reasons, we can assume that the image of g is a subset of $\mathcal{J} = [b_l, b_u]$ where $b_l < 0$ is the minimal and $b_u > 0$ is the maximal possible value of g, see Section 2.2. The minimization of the maximal risk is done numerically by an appropriate discretization in Section 6.

6. Computations

For the numerical computations, we have chosen $k_1 = 2$ and $F(x) = \Phi(x)$, see Assumption 5.1 with $\sigma^2 = 1$. Further, we consider the setting $n = 40$ in the one-stage case and compare it with the setting $n_1 = 20$, $n_2 = 20$ (20 experiments in each stage) for the two-stage case. Recall that the parameters $c_1 := g(s)$ and $c_2 := g(t)$ ($\{s,t\} = T$) are unknown but bounded (see Section 5). We assume $c_1, c_2 \in \mathcal{J} := [-1,1]$. Here, we denote the risk function by $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta)) = R_F(g,(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta))$.

For fixed values $u \in \mathcal{U} := \{-0.10, -0.09, \ldots, 0.10\}$ in the one-stage case and fixed values $u \in \mathcal{U}, \delta \in \Delta := \{0, 0.1, \ldots, 1\}$ in the two-stage case, we compute the risk $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi} > u\},\hat{\theta}))$ and $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta))$, respectively, for every $c_1, c_2 \in \mathcal{C} := \{-1, -0.9, \ldots, 1\}$ using Theorem 4.1 and Corollary 5.2. Selected values of the maximal risks $\max_{c_1,c_2\in\mathcal{C}} R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi} > u\},\hat{\theta}))$ and $\max_{c_1,c_2\in\mathcal{C}} R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta))$ are reported for $\delta \in \Delta$ (in the two-stage case) and $u \in \mathcal{U}$ in Table 1.

In the one-stage case, the computations yield $u = 0.01$ as the minimax choice. In the two-stage case, it turned out that the minimax decision has the values $u = -0.01, \delta = 0.5$. A contour plot of $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_\delta > u\},\hat{\theta}_\delta))$ for the minimax two-stage procedure is shown in Figure 1 as a function of the true mean values c_1 and c_2.

In Figure 2, we compare the risks of the optimal one-stage and the optimal two-stage minimax procedure depending on the true mean values: contour plot of the risk difference $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi} > 0.01\},\hat{\theta})) - R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_{0.5} > -0.01\},\hat{\theta}_{0.5}))$ is given. This picture shows that for a large set of true mean values

c_1, c_2, the two-stage minimax procedure has a lower risk than the one-stage minimax procedure. In some cases, however, the one-stage procedure has a slightly lower risk (region in the upper right corner and the small region in the center of the picture).

Table 1. Let $F = N(0,1)$, $k_1 = 2$, $n = 40$ and additionally $n_1 = 20$, $n_2 = 20$ in the two-stage case. Then the maximal risk $\max_{c_1,c_2 \in \mathcal{C}} R_F((c_1,c_2), (\mathbf{1}\{\hat{\xi}_\delta > u\}, \hat{\theta}_\delta))$ is shown with respect of the true mean values $c_1, c_2 \in \mathcal{C} = \{-1, -0.9, \ldots, 1\}$ as a function of u and δ.

	Two-stage							One-stage
u	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.5$	$\delta = 0.6$	$\delta = 0.8$	$\delta = 1$	$\delta = \infty$
-0.02	0.0803	0.0665	0.0649	0.0676	0.0698	0.0728	0.0744	0.0753
-0.01	0.0810	0.0676	0.0644	0.0639	0.0656	0.0687	0.0703	0.0713
0.00	0.0818	0.0687	0.0656	0.0651	0.0651	0.0651	0.0664	0.0673
0.01	0.0827	0.0700	0.0668	0.0665	0.0664	0.0665	0.0665	0.0664
0.02	0.0837	0.0714	0.0682	0.0679	0.0679	0.0679	0.0679	0.0679

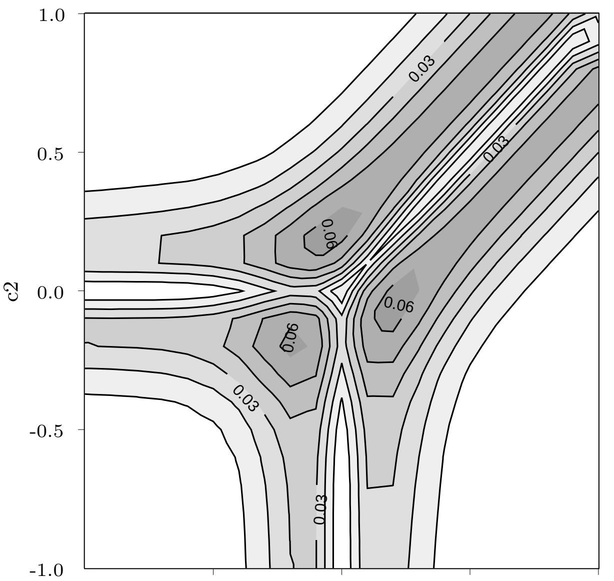

Figure 1. Contour plot of the risk $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_\delta > u\}, \hat{\theta}_\delta))$ of the minimax two-stage procedure ($\delta = 0.5, u = -0.01$) as a function of the true mean values $c_1, c_2 \in [-1,1]$. The contour lines are drawn for $0.01, \ldots, 0.06$.

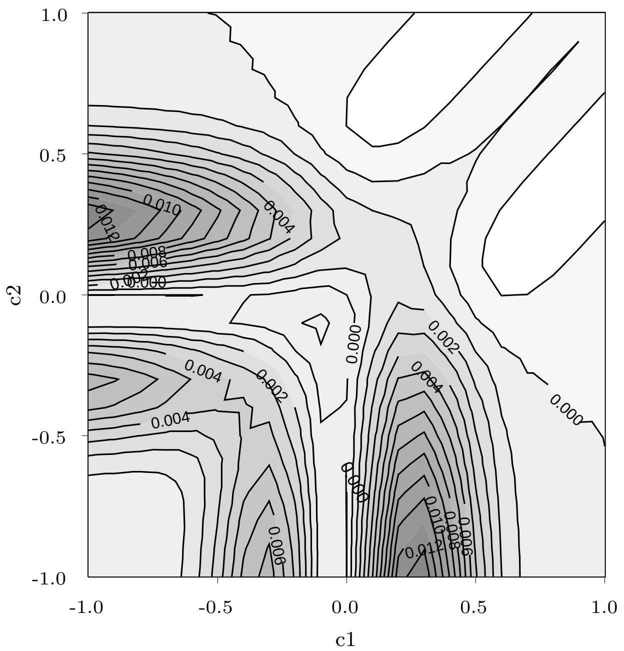

Figure 2. Contour plot of the risk difference $R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}>0.01\},\hat{\theta}))-R_F((c_1,c_2),(\mathbf{1}\{\hat{\xi}_{0.5}>-0.01\},\hat{\theta}_{0.5}))$ between the minimax one-stage and the minimax two-stage procedure as a function of the true mean values $c_1,c_2\in[-1,1]$. The contour lines are drawn for $-0.001,0,0.001,\dots,0.013$.

In practical problems, the variance σ^2 is usually unknown. It is also possible to apply this minimax approach to unknown σ^2 by computing the risks for values σ^2 in a certain set and then choosing the maximal risk. Further, it is also possible to choose $n_1\in\{k_1,\dots,n\}$ for fixed $n\in\mathbb{N}$ according to the minimax criterion.

The numerical calculation of the risk was done on an IBM RS/6000 SP parallel computer with a C program. This C program used an integration routine of the IMSL library for the multidimensional integration. This integration routine is based on iterated applications of Gauss' formulas. The C code for the computation is available from the authors upon request.

Remark. It is worth mentioning that the minimax procedures do not control the type I error. Further, the sample size n is not adjusted in order to get required probabilities of making correct decisions. Such modifications are possible but go beyond the scope of this paper.

Acknowledgements

We thank two unknown referees for suggestions which improved this paper. This work was partly supported by the Deutsche Forschungsgemeinschaft Grant Bi655.

Appendix

Proof of Theorem 3.1.

$$
\begin{aligned}
&P(\hat{\theta}=s,\hat{\xi}>u)=P(\forall t\in T\backslash\{s\}:\bar{Y}_s\geq\bar{Y}_t,\bar{Y}_s>u)\\
&=\int_u^\infty P(\forall t\in T\backslash\{s\}:z\geq\bar{Y}_t)\ F^{(m)}_{g(s)}(dz)=\int_u^\infty\prod_{t\in T\backslash\{s\}}F^{(m)}_{g(t)}(z)\ F^{(m)}_{g(s)}(dz).
\end{aligned}
$$

Proof of Theorem 4.1. The first equation is obvious. Next, we compute the probability $P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u)$ under the condition of the results of the first stage. Let $z_t\in I\!R, t\in T$, be fixed. Then define $T_2=T_2((z_t)_{t\in T})=\{t\in T\mid z_t\geq\max\{z_u\mid u\in T\}-\delta\}$, and let $r:=\#T_2$. Note that $\hat{\theta}_\delta\in T_2$ by the definition of the decision rule. For $s\in T_2, u\in I\!R$, we have (where $\gamma_{1r}=m_1/(m_1+n_2/r), \gamma_{2r}=(n_2/r)/(m_1+n_2/r)$)

$$
\begin{aligned}
&P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u\mid\bar{Y}_{1,t}=z_t,\ t\in T)\\
&=P(\forall t\in T_2\backslash\{s\}:\ \gamma_{1r}z_t+\gamma_{2r}\bar{Y}_{2,t}<\gamma_{1r}z_s+\gamma_{2r}\bar{Y}_{2,s},\gamma_{1r}z_s+\gamma_{2r}\bar{Y}_{2,s}>u)\\
&=\int_{(u-\gamma_{1r}z_s)/\gamma_{2r}}^\infty P(\forall t\in T_2\backslash\{s\}:\ \gamma_{2r}\bar{Y}_{2,t}<\gamma_{2r}z^*+\gamma_{1r}(z_s-z_t))\ F^{(n_2/r)}_{g(s)}(dz^*)\\
&=\int_{(u-\gamma_{1r}z_s)/\gamma_{2r}}^\infty\prod_{t\in T_2\backslash\{s\}}F^{(n_2/r)}_{g(t)}\left(z^*+\frac{m_1}{n_2/r}(z_s-z_t)\right)\ F^{(n_2/r)}_{g(s)}(dz^*).
\end{aligned}
$$

In the following, a_1 is the index of the largest sample mean from the first stage. Hence the probability $P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u)$ is

$$
\begin{aligned}
&\sum_{r=1}^{k_1}\sum_{a_1\in T}\sum_{\{a_2,\ldots,a_r\}\subseteq T\backslash\{a_1\}}\mathbf{1}\{s=a_j\text{ for some }j\in\{1,\ldots,r\}\}\\
&\cdot P(\forall b\in\{a_2,\ldots,a_r\},a\in T\backslash\{a_1,\ldots,a_r\}:\hat{\theta}_\delta=s,\hat{\xi}_\delta>u,\bar{Y}_{1,a_1}>\bar{Y}_{1,b}\geq\bar{Y}_{1,a_1}-\delta>\bar{Y}_{1,a})\\
&=\sum_{r=1}^{k_1}\sum_{a_1\in T}\sum_{\{a_2,\ldots,a_r\}\subseteq T\backslash\{a_1\}}\mathbf{1}\{s=a_j\text{ for some }j\in\{1,\ldots,r\}\}\\
&\quad\cdot\int_{-\infty}^\infty\int_{z_1-\delta}^{z_1}\cdots\int_{z_1-\delta}^{z_1}\int_{-\infty}^{z_1-\delta}\cdots\int_{-\infty}^{z_1-\delta}P(\hat{\theta}_\delta=s,\hat{\xi}_\delta>u\mid\bar{Y}_{1,a_i}=z_i,i=1,\ldots,k_1)\\
&\quad F^{(m_1)}_{g(a_{k_1})}(dz_{k_1})\ldots F^{(m_1)}_{g(a_{r+1})}(dz_{r+1})F^{(m_1)}_{g(a_r)}(dz_r)\ldots F^{(m_1)}_{g(a_2)}(dz_2)F^{(m_1)}_{g(a_1)}(dz_1)\\
&=\sum_{r=1}^{k_1}\sum_{a_1\in T}\sum_{\{a_2,\ldots,a_r\}\subseteq T\backslash\{a_1\}}\mathbf{1}\{s=a_j\text{ for some }j\in\{1,\ldots,r\}\}\\
&\quad\cdot\int_{-\infty}^\infty\int_{z_1-\delta}^{z_1}\cdots\int_{z_1-\delta}^{z_1}\int_{(u-\gamma_{1r}z_j)/\gamma_{2r}}^\infty\left[\prod_{\substack{i=1\\i\neq j}}^r F^{(n_2/r)}_{g(a_i)}\left(z^*+\frac{m_1}{n_2/r}(z_j-z_i)\right)\right]F^{(n_2/r)}_{g(a_j)}(dz^*)
\end{aligned}
$$

$$\Big[\prod_{a\in T\setminus\{a_1,\ldots,a_r\}} F^{(m_1)}_{g(a)}(z_1-\delta)\Big] F^{(m_1)}_{g(a_r)}(dz_r)\ldots F^{(m_1)}_{g(a_2)}(dz_2)F^{(m_1)}_{g(a_1)}(dz_1). \qquad \text{(A.1)}$$

Proof of Theorem 4.2. We have to determine $d \in \mathbb{R}$ with

$$\inf_{g\in\mathbb{R}^T} P_{g,F}(g(\theta)\in[\hat{\xi}_\delta - d,\infty)) = 1-\alpha$$

(and the same with θ replaced by $\hat{\theta}_\delta$). We claim that in both cases, the infimum of the left side is attained for $g(s)=0, s\in T$. Then the assertion of the theorem follows immediately. The claim is only proved for $g(\hat{\theta}_\delta)$.

Let $\bar{\epsilon}_{1t} = (\sum_{j=1}^{m_1}\epsilon_{1tj})/m_1$ so $\bar{Y}_{1t} = g(t)+\bar{\epsilon}_{1t}$, $t\in T$, and let $z_t, t\in T$, be arbitrary real numbers. Then, under the condition $\bar{\epsilon}_{1t} = z_t, t\in T$,

$$\begin{aligned}
&P_{g,F}(g(\hat{\theta}_\delta)\in[\hat{\xi}_\delta - d,\infty) \mid \bar{\epsilon}_{1t}=z_t, t\in T)\\
&= P_{g,F}(\exists s\in T_2 : \hat{\xi}_\delta - g(s)\le d, \hat{\theta}_\delta = s \mid \bar{\epsilon}_{1t}=z_t, t\in T)\\
&\ge P_{g,F}(\forall s\in T_2 : (m_1+m_2)^{-1}(m_1 z_s + m_2(\bar{Y}_{2s}-g(s)))\le d \mid \bar{\epsilon}_{1t}=z_t, t\in T)\\
&= P_{0,F}(\hat{\xi}_\delta \le d \mid \bar{\epsilon}_{1t}=z_t, t\in T).
\end{aligned}$$

Note that, as in the proof of Theorem 4.1, T_2 and k_2 are fixed under the condition $\bar{\epsilon}_{1t}=z_t, t\in T$. Since g and the numbers z_t were arbitrary, the claim follows.

References

Bauer, M., Bauer, P. and Budde, M. (1998). A simulation program for adaptive two stage designs. *Comput. Statist. Data Anal.* **26**, 351-371.

Coad, D. S. (1995). Sequential allocation involving several treatments. In *IMS Lecture Notes Monograph. Ser.* **25** (Edited by N. Flournoy, W. F. Rosenberger), 95-109. Adaptive designs. Selected proceedings of a 1992 joint AMS-IMS-SIAM summer conference.

Friede, T., Miller, F., Bischoff, W. and Kieser, M. (2001). A note on change point estimation in dose-response trials. *Comput. Statist. Data Anal.* **37**, 219-232.

Siegmund, D. (1993). A sequential clinical trial for comparing three treatments. *Ann. Statist.* **21**, 464-483.

Wetherill, G. B. and Glazebrook, K. D. (1986). *Sequential Methods in Statistics*, 3rd edition. Chapman and Hall, New York.

Institut für Mathematische Stochastik, Universität Karlsruhe, D-76128 Karlsruhe, Germany.

E-mail: Wolfgang.Bischoff@math.uni-karlsruhe.de

Institut für Mathematische Stochastik, Universität Karlsruhe, D-76128 Karlsruhe, Germany.

E-mail: Frank.Miller@math.uni-karlsruhe.de

(Received March 2001; accepted April 2002)