A backward-induction algorithm for computing the best convex contrast of two bivariate samples

D. W. Müller

Institut für Angewandte Mathematik

Universität Heidelberg

Germany

Abstract. For real-valued $x_1, x_2, ..., x_n$ with real-valued "responses" $y_1, y_2, ..., y_n$ and "scores" $s_1, s_2, ..., s_n$ we solve the problem of computing the maximum of $C(k) = \sum_{h=1,...,n} s_h \mathbf{I}[y_h \geq k(x_h)]$ over all convex functions k on $\Re^1$. The article describes a recursive relation and an algorithm based on it to compute this value and an optimal k in $O(n^3)$ steps. For a special choice of scores, max C(k) can be interpreted as a generalized (one-sided) Kolmogorov-Smirnov statistic to test for treatment effect in nonparametric analysis of covariance.

Key words: analysis of covariance, contrast, backward induction, Kolmogorov-Smirnov test, minimal guaranteed effect, nonparametrics.

A backward-induction algorithm for computing the best convex contrast of two bivariate samples

D.W.Müller

Institut für Angewandte Mathematik

Universität Heidelberg

Germany

1. Introduction

Let $x_1, x_2, ..., x_n$ be real numbers with real-valued "responses" $y_1, y_2, ..., y_n$ and real-valued "scores" $s_1, s_2, ..., s_n$. For a function k on $\Re^1$ we consider the "contrast"

$C(k) = \sum_{h=1,...,n} s_h \mathbf{I}[y_h \geq k(x_h)]$. The following quantity is of interest:

$C^* = \max C(k)$,

where the maximum is taken over all *convex* functions k on $\Re^1$.

This article describes a recursive relation and an algorithm based on it to compute the optimal contrast C^* and a convex curve k^* with $C(k^*) = C^*$ (the optimal contrast curve) in $O(n^3)$ steps.

In the simplest special case the (x_i,y_i) are assumed to belong to one of two groups, the group identifier being s_i with values in $\{\alpha,\beta\}$ ($\alpha\beta < 0$). For example $\alpha = 1$, $\beta = -1$.

A statistical application to nonparametric analysis of covariance will be discussed in section 6. There one considers the difference of the empirical distributions formed by the two groups evaluated for the epigraphs of convex functions k. If the group sizes are n_1 and n_2 respectively ($n = n_1 + n_2$), the scores will be set $s_i = 1/n_1$ for the treatment group and $s_i = -1/n_2$ for the control group.

In general we have $C^* \geq 0$, because $C(\tilde{k}) = 0$ for the constant function $\tilde{k}(x) \equiv 1 + \max_{i=1,...,n} y_i$. Moreover $\max_{h=1,...,n} s_h \leq C^* \leq \sum_{h=1,...,n} s_h^+$; hence $C^* > 0$ if and only if $\max_{i=1,...,n} s_i > 0$.

Without restriction of generality we will assume the x-values being ordered: $x_1 \leq x_2 \leq ... \leq x_n$. Obviously, maximization can be restricted to the class $\mathcal{L}$ of those special

convex functions k which are broken linear functions with break points (x_{i_1},y_{i_1}), ... ,(x_{i_h},y_{i_h}) $(i_1 < i_2 < ... < i_h)$, i.e. functions k such that $k(x_{i_j}) = y_{i_j}$ (j = 1,2, ... ,h), $k(x_i) > y_i$ (i = 1,2, ... ,i_1, i_h + 1, ..., n), being linear in the intervals $(-\infty, x_1]$, $[x_{i_j}, x_{i_{j+1}}]$ (j = 1,2, ... ,h – 1), $[x_{i_h}, +\infty)$.

The algorithm is related to the algorithm designed by Hartigan (1987) for computing the minimal volume convex set in $\Re^2$ containing a given proportion of a sample. However, in the present situation, backward induction can be used, leading to a transparent structure of the algorithm. Both algorithms have complexity at most $O(n^3)$.

2. The recursion

First we assume the absence of ties, i.e. $x_1 < x_2 < ... < x_n$. In this case the structure of the recursion is most transparent. The necessary modifications in the presence of ties will be described in section 5.

We introduce the subfamilies of convex functions $\mathcal{K}_{i,\delta} = \{k \in \mathcal{K}: k(x_i) = y_i \text{ and } k'(x_i) \geq \delta\}$. Here $k'(x_i)$ denotes the right derivative of k at x_i. Moreover $\mathcal{K} = \mathcal{K}_{1,-\infty} \cup \mathcal{K}_{2,-\infty} \cup ... \cup \mathcal{K}_{n,-\infty}$.

The important quantity to be considered here is

$$C(i,\delta) \quad (i = 1,2, ... ,n;\ \delta \in \Re^1),$$

defined as

$$C(i,\delta) = \max_{k \in \mathcal{K}_{i,\delta}} \left\{ \sum_{h \geq i} s_h \mathbf{I}[y_h \geq k(x_h)] \right\}.$$

$C(i,+\infty)$ and $C(i,-\infty)$ will denote the limits of $C(i,\delta)$ for $\delta \to +\infty$ and $\delta \to -\infty$, respectively.

For each i, $C(i,\bullet)$ is a monotonically decreasing left-continuous function taking at most n – i +1 values. We have $C(n,\delta) = s_n$ (all δ) and $C(i,+\infty) = s_i$ (all i). We note that

$$C(i,\delta) = \max \left\{ C(k): k \in \mathcal{K}_{i,\delta}, k(x_1) > y_1, k(x_2) > y_2, ..., k(x_{i-1}) > y_{i-1} \right\}.$$

Our purpose is to compute C(i,•) from C(i+1,•), ... , C(n,•) ("backward induction"). The recursive relation will be given by proposition 1. For its formulation we need the quantities $\delta(i,j) = \frac{y_j - y_i}{x_j - x_i}$ $(i \neq j)$ and u(i,j) $(i \leq j)$:

$u(i,j) = \sum\{s_h: i \leq h < j,\ y_h \geq y_i + \delta(i,j)\ (x_h - x_i)\}$.

For given $i \leq j$ the value u(i,j) can be computed in $O(n)$ steps. Thus the function u can be tabulated in $O(n^3)$ steps. However, as the following recursion shows, there is no need to store the whole matrix u, since at each induction step i only the ith row will be required.

Proposition 1 (backward induction)**.** *For each* i = 1, ... , n – 1:

$C(i,\delta) = s_i \vee \max_{i<j\leq n}\{u(i,j) + C(j,\delta(i,j)): \delta(i,j) \geq \delta\}$.

Proof. The proof relies on the fact that $\omega_k([i,h)) = \sum_{i\leq j<h} s_j\ \mathbf{I}[y_j \geq k(x_j)]$ is an additive functional on the intervals [i,h): $\omega_k([i,h)) = \omega_k([i,j)) + \omega_k([j,h))$. It is related to C(i,δ) via

$C(i,\delta) = \max_{k\ \in\ \mathcal{L}_{i,\delta}}\{\omega_k([i,n])\}$

$= \max_{i<j\leq n} \max_{k\ \in\ \mathcal{L}_{i,\delta}}\{\omega_k([i,j)) + \omega_k([j,n])\}$ (by additivity) .

Then

$\mathcal{L}_{i,\delta} = \{k \in \mathcal{L}_{i,\delta}: k(x_j) > y_j \text{ for all } j = i+1, \dots, n\} \cup \{k: k \text{ linear in } [x_i, x_j], \text{ interpolating } y_i \text{ and } y_j \text{ (for some } i<j) \text{ with } \delta(i,j) \geq \delta, \textit{ and } k(x) = \ell(x)\ (x \geq x_j) \text{ for some } \ell \in \mathcal{L}_{j,\delta(i,j)}\}$. Thus

$C(i,\delta) = s_i \vee \max_{i<j\leq n}\{u(i,j) + \max_{\ell\in\mathcal{L}_{j,\delta(i,j)}} \omega_\ell([j,n]): \delta(i,j) \geq \delta\}$

$= s_i \vee \max_{i<j\leq n}\{u(i,j) + C(j,\delta(i,j)): \delta(i,j) \geq \delta\}$.■

By means of this recursion the functions C(1,•), ... , C(n,•) can now be computed in $O(n^3)$ steps. This is due to the fact that the value C(i,δ) will only be needed for arguments δ = δ(i,j) (j = 1,..., i – 1, i + 1, ... ,n) (see section 3). Then, by proposition 2, the quantity of interest C* can be obtained in another $O(n)$ steps.

Proposition 2.

$C^* = 0 \vee \max_{1\leq i\leq n} C(i,-\infty)$.

Proof. C(k) will be maximized either by $\tilde{k}$ (then C* = 0) or by some $k \in \mathcal{K}$; in the latter case the optimal $k^* \in \mathcal{K}$ will satisfy

$k^*(x_1) > y_1$, $k^*(x_2) > y_2$, ... , $k^*(x_{i-1}) > y_{i-1}$, $k^*(x_i) = y_i$ for some $i \geq 1$; this means that $k^* \in \mathcal{K}_{1,-\infty}$ and $C^* = C(i,-\infty)$.■

3. The algorithm

For computational purposes, the functions C(i,•) will be represented as vectors $v(i,\bullet) \in \Re^n$ giving the values of C(i,•) at all slopes δ(i,j) of line segments joining the point (x_i,y_i) with all the other n – 1 points *(component loop)*. Only these special values will be needed.

In detail, let

$$v(i,j) = C(i,\delta(i,j)) \quad (j = 1, \ldots , i-1, i+1, \ldots, n).$$

(v(i,i) will be left undefined). The last n – i cases correspond to possible jumps of C(i,•); the first i – 1 cases give values of the function C(i,•), which will be needed in the recursion. In the new notation the recursion of proposition 1 becomes

(a) $v(i,h) = s_i \vee \max_{i<j\leq n}\{u(i,j) + v(j,i): \delta(i,j) \geq \delta(i,h)\}$ $(i+1 \leq h \leq n)$.

On the other hand, for $h \leq i - 1$, v(i,h) is the value C(i,δ(i,h)) of the monotonically decreasing left-continuous function C(i,•) with jumps at δ(i,j), $i < j \leq n$; thus

(b) $v(i,h) = s_i \vee \max_{\,i<j\leq n}\{v(i,j): \delta(i,j) \geq \delta(i,h)\}$ $(1 \leq h \leq i-1)$.

In both cases (a) and (b), for given h, at most n – i + 1 operations for searching the max are required (*inner loop*). (The operations (b) can be replaced by a simultaneous calculation of v(i,h), $1 \leq h \leq i - 1$, using a sorting algorithm, thus reducing the computational complexity. This modification, however, will not affect the rate of the overall complexity of the algorithm, in view of the operations (a)).

In toto, v(i,•) will be calculated from v(i+1,•), ... ,v(n,•) in at most (n – 1)(n – i + 1) steps. Thus v(•,•) will be available after at most $\sum_i(n - 1)(n - i + 1) = O(n^3)$ steps (*outer loop*).

We now summarize the computational scheme for v(•,•) as follows:

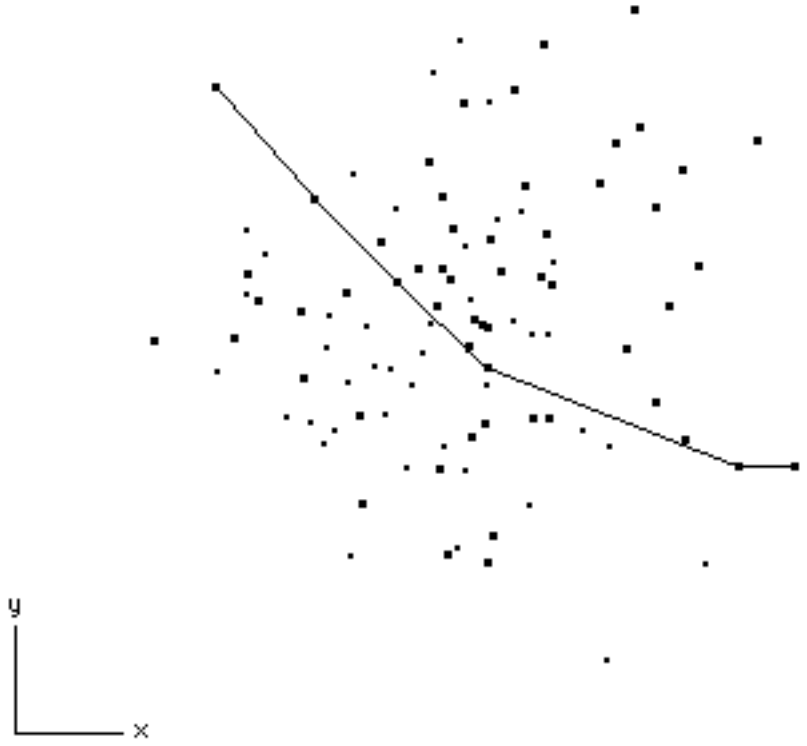
```
for h := 1 to n do v(n,h) := s_n ; {h, component loop}
for i := n – 1 downto 1 do
    begin
    for h := i + 1 to n do
        begin
        v := s_i;
        for j := i + 1 to n do
            if δ(i,j) ≥ δ(i,h) then v := max(v, u(i,j) + v(j,i));{j, inner loop (a)}
        v(i,h) := v
        end;{h, component loop}
    for h := 1 to i - 1 do
        begin
        v := s_i;
        for j := i+1 to n do
            if δ(i,h) ≤ δ(i,j) then v := max(v, v(i,j));{j, inner loop (b)}
        v(i,h) := v
        end;{h, component loop}
    end;{i, outer loop}.
```

In a computer experiment on an Apple Quadra 950 (68030 processor) x_i,y_i were sampled from a standard Gaussian distribution, with independently assigned scores ± 1; the running times were as follows

n	**$10^{-5}n^3$**	**seconds**
50	1.25	1.35
100	10	10.02
150	33.75	33.17
200	80	78.25
300	270	262.28
400	640	627.62

500 1250 1219.62

The following figure shows a computer output of the optimal contrast curve based on 100 samples (56 from N(0,1)⊗N(0.5,1) with scores = 1/56, plotted bold face, and 44 from N(0,1)⊗N(0,1) with scores -1/44. The contrast value is 0.401.



4. Optimal contrast curves

As a byproduct of the above algorithm, convex contrast curves k* maximizing C(k) can be computed. This can be done by *forward induction* using the list v(•,•), according to the following principle.

Let $k \in \mathcal{C}_{i,\delta(i,j)}$ (i < j) where δ(i,j) is a point of discontinuity (jump) of C(i,•). Then

$\sum_{h\geq i} s_h \mathbf{I}[y_h \geq k(x_h)] = C(i,\delta(i,j))$

if and only if

k *is linear in the interval* $[x_i,x_j]$ *interpolating* y_i *and* y_j, $k \in \mathcal{K}_{j,\delta(i,j)}$ *and* $\sum_{h\geq j} s_j \mathbf{I}[y_h \geq k(x_h)] = C(j,\delta(i,j))$.

According to this principle optimal contrast curves $k \in \mathcal{K}$ can be constructed by successively defining its linear segments.

Since a row index i with maximal v(i,j) for some j > i can be found from the table of v, it suffices to have an algorithm finding some $k_i \in \mathcal{K}_{i,-\infty}$ such that $C(k_i) = C(i,-\infty)$ for given i. Such an algorithm will now be desribed.

First step. Let $\delta_i = \max\{\delta: C(i,\delta) = C(i,-\infty)\}$. If $\delta_i = +\infty$, then we define k_i such that $k_i(x_j) > y_j$ for all j > i. In this case, the construction of k_i is complete. If $\delta_i < +\infty$, then δ_i is the position of a discontinuity (jump) of $C(i,\bullet)$. Therefore in this case there exists a $k \in \mathcal{K}_{i,\delta_i}$ such that $\sum_{h\geq i} s_h \mathbf{I}[y_h \geq k(x_h)] = C(i,\delta_i)$ and $k' = \delta_i = \delta(i,j)$ for some j > i. This means that k is linear in the interval $[x_i,x_j]$, interpolating y_i and y_j linearly. Hence we set $k_i(x) = y_i + \delta(i,j)(x - x_i)$ $(x \in [x_i,x_j])$.

Second step. Put $\delta_j = \max\{\delta: C(j,\delta) = C(j,\delta(i,j))\}$.If $\delta_j = +\infty$, then we define k_i such that $k_i(x_h) > y_h$ for all h > j. In this case, the construction of k_i is complete. If $a_j < +\infty$, then δ_j is the position of a discontinuity of $C(j,\bullet)$. Therefore there exists a $k \in \mathcal{K}_{j,a_j}$ such that $\sum_{h\geq j} s_h \mathbf{I}[y_h \geq k(x_h)] = C(j,\delta_j)$ and $k' = \delta_j = \delta(j,m)$ for some m > j. This means that k is linear in the interval $[x_j,x_m]$, interpolating y_j and y_m. Hence we set $k_i(x) = y_j + \delta(j,m)(x - x_j)$ $(x \in [x_j,x_m])$. This step is being repeated until the construction of k_i is complete.

5. The presence of ties

It is now only assumed that $x_1 \leq x_2 \leq \ldots \leq x_n$. This necessitates a few modifications. First of all, it is no restriction of generality to assume that $y_i \neq y_j$, if $x_i = x_j$. Otherwise the two points (x_i,y_i), (x_j,y_j) could be considered as only one point with score $s_i + s_j$.

By reordering the y's, it is possible to assume that for $i \leq j$, $x_i = x_j$, we always have $y_i \leq y_j$. We then introduce the new scores $\tilde{s_i} = \sum_{h\geq i} \{s_h: x_i = x_h\}$.

Then proposition 1 becomes

Proposition 1^t. (backward induction)**.** *For each* i = 1,..., n – 1:
$C(i,\delta) = \tilde{s}_i \vee \max_{i<j\le n}\{u(i,j) + C(j,\delta(i,j)): x_j \neq x_i, \delta(i,j) \ge \delta\}$.

(Here the maximum over an empty set is $-\infty$. The slope $\delta(i,j)$ is undefined for $x_i = x_j$). Proposition 2 remains valid without change when ties are present.

6. Statistical application (nonparametric analysis of covariance)

Let (x_1,y_1), ..., (x_n,y_n) be repeated observations of a covariable x and its response y consisting of two groups (treatment and control) of size n_1 and n_2 respectively ($n = n_1 + n_2$). For an observation (x_i,y_i) of the first group $s_i = 1/n_1$, otherwise $s_i = -1/n_2$. It is assumed in the simplest case that the observations follow a random process of the type $y_i = k_0(x_i) + \alpha_i + \varepsilon_i$ with a convex function k_0, an additive treatment effect α_i, depending only on the group, and the assumption that the error terms ε_i are independent identically distributed according to G, independent of x_i (i = 1,2, ... n) whose distribution is F. F and G are assumed to have strictly positive densities f,g w.r. to Lebesgue measure, respectivly. The deterministic quantities k_0 and α_i are unknown to the statistician. The problem is to test the hypothesis of no treatment effect ($\alpha_1 = \alpha_2$). In classical statistics it has been assumed that k_0 is a linear function. The performance of tests derived for this situation must be poor when the assumption of linearity is violated. Therefore nonparametric tests have been proposed which work under weak assumptions on k_0 (see Quade 1982). We consider a generalized Kolmogorov-Smirnov test, which rejects the hypothesis of no treatment effect if
$KS_n = \max_{k \text{ convex}}\left\{\frac{1}{n_1}\#\{i \le n: y_i \ge k(x_i): s_i > o\} - \frac{1}{n_2}\#\{i \le n: y_i \ge k(x_i): s_i < o\}\right\} \ge c$. It is easy to see that KS_n can be written as $\max\{C(k): k \text{ convex}\} = C(k_n^*)$. The critical constant c will be determined by rerandomization. We are going to argue that this test works reasonably if k_0 is a convex function.

If there is a treatment effect (i.e. $\alpha_1 > \alpha_2$), then under some conditions on the error distribution the optimal k_n^* is a consistent estimator of some translate of k_0 ($n \to +\infty$, such that $n_1 \wedge n_2 \to +\infty$). This is the content of proposition 3.

Proposition 3. *The density* g *of the error distribution* G *is assumed to be continuous and unimodal in the sense that, for some* z_0, g(z) *is strictly increasing for* $z < z_0$, *strictly decreasing for* $z > z_0$.

If $\alpha_1 > \alpha_2$, *not depending on* n, *then there exists a constant* γ_0 *such that the optimal* k_n^* *converges to* $k_0 + \gamma_0$ *uniformly on compacts, almost surely.*
$\gamma_0 = \alpha_1 + \lambda(\alpha_1 - \alpha_2)$, *where* λ *is the function on* $[0,+\infty)$ *defined by:* $g(z_1) = g(z_2)$,
$z_1 < z_2 \Rightarrow z_1 = \lambda(z_2 - z_1)$.

For the proof the following elementary lemma will be used.

Lemma. *Let* $\psi(\kappa,\pi)$ *be a function of abstract arguments* κ *and* π. *Assume that*

(i) κ *is the unique maximum of* $\psi(\bullet,\pi)$;

(ii) $\psi(\kappa,\pi_n) \to \psi(\kappa,\pi)$ *uniformly in* κ *as* $n \to +\infty$;

(iii) κ_n *maximizes* $\psi(\bullet, \pi_n)$.

Then $\psi(\kappa_n,\pi) \to \psi(\kappa,\pi)$ *as* $n \to +\infty$.

Proof (proposition 3).

Let $\mathcal{A}$ be the family of all closed convex sets A such that $(x,y) \in \mathcal{A}$, $y < z \Rightarrow (x,z) \in \mathcal{A}$. Hence the epigraph $U(k) = \{(x,y): y \geq k(x)\}$ of a convex function k is a member of $\mathcal{A}$. In this way, convex functions are imbedded in $\mathcal{A}$. The following defines a metric on the space $\mathcal{A}$: $d(A,B) = N(0,1)\otimes N(0,1)(A \Delta B)$ $(A,B \in \mathcal{A})$. In this metric, the space $\mathcal{A}$ is compact. If $d(U(k_n), U(k)) \to 0$ $(n \to +\infty)$ for convex functions k_n,k, then $k_n(x) \to k(x)$ uniformly on compact sets.Let

$P = \mathcal{L}(x, k_0(x) + \alpha_1 + \varepsilon \,|\, (x,\varepsilon) \sim F\otimes G)$, $Q = \mathcal{L}(x, k_0(x) + \alpha_2 + \varepsilon \,|\, (x,\varepsilon) \sim F\otimes G)$, and $\Phi(A,P,Q) = P(A) - Q(A)$ $(A \in \mathcal{A})$. After the assumptions on g, this functional on $\mathcal{A}$ has its unique maximum at $A = U(k_1)$ where $g(k_1(x) - k_0(x) - \alpha_1) = g(k_1(x) - k_0(x) - \alpha_2)$. Hence $k_1(x) - k_0(x) - \alpha_1 = \lambda(\alpha_1 - \alpha_2)$, that is, $k_1(x) - k_0(x) - \alpha_1$ is a constant. Thus $k_1(x) = k_0(x) + \gamma_0$, where $\gamma_0 = \alpha_1 + \lambda(\alpha_1 - \alpha_2)$. Let P_{n_1}, Q_{n_2} be the empirical distributions of two independent samples of P,Q of size n_1, n_2 respectively. The Glivenko-Cantelli theorem yields $P_{n_1}(A) \to P(A)$, $Q_{n_2}(A) \to Q(A)$ uniformly in $A \in \mathcal{A}$, almost surely (P,Q) as $n_1 \wedge n_2 \to +\infty$, hence $\Phi(A,P_{n_1},Q_{n_2}) \to \Phi(A,P,Q)$ uniformly in $A \in \mathcal{A}$, almost surely (P,Q) as $n_1 \wedge n_2 \to$

$+\infty$. Now $U(k_n^*)$ maximizes $\Phi(A,P_{n_1},Q_{n_2})$ over $A \in \mathcal{A}$. Therefore the lemma applies giving $\Phi(U(k_n^*), P, Q) \to \Phi(U(k_1), P, Q)$ almost surely (P, Q) as $n_1 \wedge n_2 \to +\infty$. Now $\Phi(U(k_n^*),P,Q) \to \Phi(U(k_1),P,Q)$ implies that $d(U(k_n^*), U(k_1)) \to 0$, by compactness. Thus $k_n^*(x) \to k_1(x)$ uniformly on compact sets.■

Also in a less restritive statistical model the above test statistic KS_n can be justified by being the empirical measure of the minimum guaranteed treatment effect (as in Müller (1980)). Let $Q = \mathcal{L}(x, k_0(x) + \varepsilon | (x,\varepsilon) \sim F \otimes H)$, the joint law of $x \sim F$ and $k_0(x) + \varepsilon$ ($\varepsilon \sim H$ independent of x) be the control distribution. Following Müller (1980) we model the treatment by means of a Markov kernel $K(z,\bullet)$, being the distribution of the error ε after treatment, given that $z = y - k_0(x)$ without treatment. This extends the approach of Doksum (1974) (see also Doksum and Sievers (1976)) who models the treatment effect by a deterministic function. The kernel K describes the stochastic change of the residual caused by the treatment. Thus $P = \mathcal{L}(x, k_0(x) + \varepsilon | (x,\varepsilon) \sim F \otimes G)$ where
$G(\bullet) = KH(\bullet) = \int K(z,\bullet)\, H(dz)$. The kernel K is unobservable, because only independent samples from the marginals P and Q will be taken. We assume that negative effects are excluded, that is, K is monotone in the sense that $K(z, [z,+\infty)) = 1$ for all z. The probability of a positive treatment effect given z will be $K(z, (z,+\infty))$. The minimum guaranteed treatment effect will then be the minimum of $\int K(z, (z,+\infty))\, H(dz)$, taken over all monotone kernels satisfying G = KH. This minimum can be seen to equal
$\pi = 1 - \|G \wedge H\| = \int g \wedge h(z)\, dz = \|(G - H)^+\|$ where h is the density of H. If g – h is increasing ($(g - h)(\zeta_0) = 0$) then

$$\begin{aligned}
\pi &= \|(G-H)^+\| \\
&= \iint (g - h)^+ [y - k_0(x)]\, dy\, F(dx) \\
&= \iint \mathbf{I}[y - k_0(x) \geq z_0]\, (g - h)\, [y - k(x)]\, F(dx)\, dy \\
&= \iint \mathbf{I}[y \geq \zeta_0 + k_0(x)]\, (P - Q)(dxy) \\
&= P(U(\zeta_0 + k_0)) - Q(U(\zeta_0 + k_0)).
\end{aligned}$$

On the other hand, for any k,

$P(U(k)) - Q(U(k)) = \iint \mathbf{I}[y - k(x) \geq 0]\, (g - h)[y - k(x)]\, F(dx)\, dy$ which is maximized in k if the inner integral is maximized for each x:

$$\int \mathbf{I}[y - k(x) \geq 0]\, (g - h)\, [y - k(x)]\, dy = (G - H)([k(x) - k_0(x), +\infty)).$$

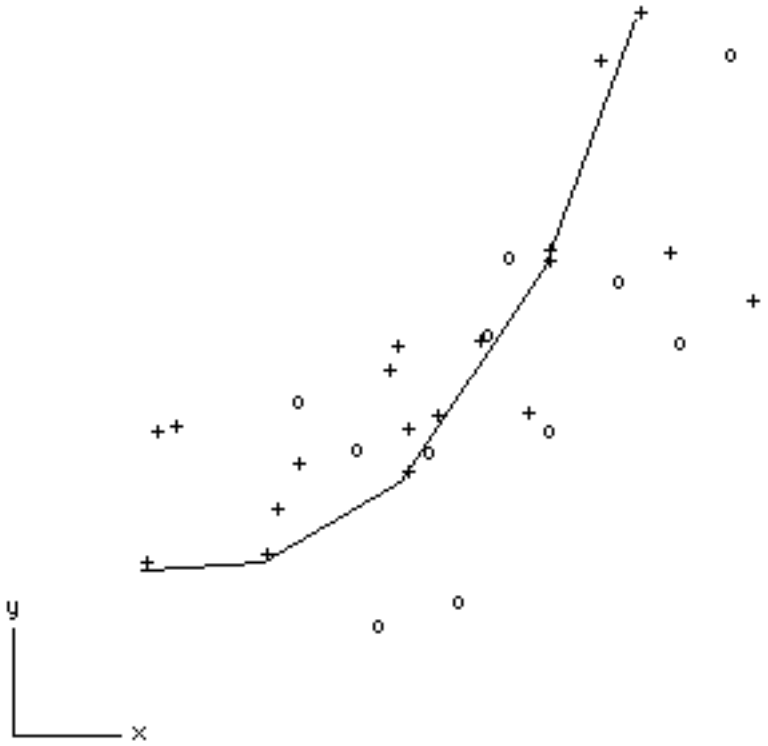
The maximum is attained for $k(x) = \zeta_0 + k_0(x)$. Thus $\pi = ||(G - H)^+|| \geq P(U(k)) - Q(U(k))$ with equality for $k = \zeta_0 + k_0$. Thus, using the Glivenko-Cantelli theorem, we have proved

Proposition 4. *Let the densities* g,h *of* G, H *respectively be such that* g – h *is increasing. With probability* $1, KS_n$ *converges to* $\pi = ||(G-H)^+||$, *the minimum guaranteed treatment effect.*

√n - consistency and a central limit theorem easily follow from the central limit theorem for the empirical process (see Bolthausen 1978). However, the asymptotic power for contiguous alternatives will depend on speculative assumptions about the error distributions. The large sample behaviour of the test using bootstrapped critical values can be derived following the lines of Romano (1989). The author does not expect that large sample results will be useful for sample sizes permitting computation.

7. Example

The following data from Snedecor and Cochran (1980) have been reanalyzed by Quade (1982). The variable y measures the cholesterol concentration for two groups of women (11 from Iowa serving as controls (o), 19 from Nebraska (+)), the covariable x is age. The figure shows an optimal contrast curve.



The resulting contrast value is C* = 0.4785. The significance probability was determined by a Monte Carlo simulation of 10,000 runs using Apple's SANE random number generator, an implementation of the portable pseudorandom number generator of Wichmann and Hill (1987). Under the hypothesis of no difference between the states the estimated probability that a contrast value of this size will be exceeded is 0.48. Thus, by our analysis (as well as by the analyses of the above-mentioned authors), there appears no significant difference between the states.

8. Further developments (S-shaped regression)

The analogous problem of an optimal S-shaped (i.e. convex-concave) contrast curve can be solved in the same manner; the corresponding algorithm has complexity at most $O(n^4)$. Also the problem of more than two samples can be treated as usual in the present context.

It is desirable to increase the power of the test by using the ranks of $y_i - k(x_i)$. The optimization of such contrast functionals seems to be an unsolved problem.

References

Bolthausen, E. (1978), "Weak Convergence of an Empirical Process Indexed by the Closed Convex Subsets of I^2", *Z.Wahrscheinlichkeitstheorie verw. Gebiete*, **43**, 173 - 181.

Doksum, K.A. (1974), "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two Sample Case," *The Annals of Statistics* **2**, 267 - 277.

Doksum, K.A. and Sievers, G.L. (1976), "Plotting with Confidence, Graphical Comparisons of Two Populations", *Biometrika* **63**, 421 - 434.

Hartigan, J.P. (1987), "Estimation of a Convex Density Contour in Two Dimensions", *Journal of the American Statistical Association*, **82**, 267 - 270.

Müller, D.W. (1980) "The analysis of 'minimum guaranteed effect' in the two sample case", Preprint Nr. **72** (7/1980), Sonderforschungsbereich 123 "*Stochastische Mathematische Modelle*", Universität Heidelberg.

Quade, D. (1982) "Nonparametric Analysis of Covariance by Matching", *Biometrics* **38**, 597 - 611.

Romano, J.P. (1989), "Bootstrap and Randomization Tests of Some Nonparametric Hypotheses", *The Annals of Statistics* **17**, 141 - 159.

Snedecor, G.W. and Cochran, W.G. (1980), *Statistical Methods*, 7th ed., Iowa State University Press.