SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1063

August 1993

**A NOTE ON THE LOCAL
ASYMPTOTICALLY MINIMAX
RATE FOR ESTIMATING A
CROSSING POINT IN A
DIAGNOSTIC MARKER PROBLEM**

by

Fushing Hsieh¹
and Bruce Turnbull²

¹National Tsing Hua University, Institute of Statistics, Hsinchu, Taiwan 30043.

²Research supported in part by Grant GM 28364 from the U.S. National Institutes of Health.

**A NOTE ON THE LOCAL ASYMPTOTICALLY MINIMAX RATE
FOR ESTIMATING A CROSSING POINT IN A
DIAGNOSTIC MARKER PROBLEM**

Fushing Hsieh

National Tsing Hua University, Institute of Statistics, Hsinchu 30043 ,Taiwan

Bruce W. Turnbull

*School of Operations Research and Industrial Engineering,
227 E&TC Building, Cornell University, Ithaca, NY 14853-3801, U.S.A.*

Abstract. The optimal rate of convergence of estimating the crossing point of two densities in the local asymptotically minimax sense is discussed. This rate is shown to be the same as the optimal rate of density estimation under smoothness conditions on the two distribution functions.

Keywords. Classification problem, density estimation, Hölder continuity, Youden index.

1. Introduction

A diagnostic test giving a measurement on a continuous scale is used to classify patients into either the “healthy” or “diseased” categories. Typically, a cutoff point, c , is selected, and patients with test results greater than this are classified as “diseased”, otherwise as “healthy”. The test score of a healthy patient is represented as a real random variable X with distribution function F and density f . Similarly a diseased patient’s score will be denoted by Y with distribution function G , density g .

* This research was supported in part by NIH Grant GM 28364.

For a diagnostic test, when a cutoff point c is chosen, the probability of having a false negative is $G(c)$, while the probability of having a false positive is $1 - F(c)$. The desired cutoff point, c , can be chosen according to the relative cost of a false negative to that of a false positive. In this note we will consider the cutoff point, c , which minimizes the sum, $G(x) + (1 - F(x))$, which is essentially the performance index for a diagnostic marker proposed by Youden (1950) — see also Hsieh and Turnbull (1993). (Our results can be easily generalized to a weighted sum.) We will assume that F and G each have connected intervals as their supports which overlap, with an intersection which contains the interval (a_0, b_0) . The two densities f and g are assumed to have a unique intersection point θ , say, in (a_0, b_0) , *i.e.* $f(\theta) = g(\theta)$, with $f(x) > g(x)$, if $x < \theta$ and $f(x) < g(x)$, if $x > \theta$. This is satisfied, for example, if there is a monotone likelihood ratio. This θ is called the crossing point of f and g and it is the parameter of our prime interest because this implies that $\theta = \operatorname{argmax}[F(x) - G(x)]$ and is the value of the optimal cutoff point c for the Youden index.

We assume that θ is unique in the following sense:

(A1) for any small $\delta (> 0)$, there exist $\varepsilon (> 0)$, such that

$$\inf_{|x-\theta|>\delta} |g(x) - f(x)| > \varepsilon \text{ for } x \in [a_0, b_0].$$

Further, we impose a Hölder continuity smoothness condition upon F and G . First suppose $\gamma \geq 2$, let α be the largest integer less than γ and set $\beta = \gamma - \alpha$. Let $\mathcal{C}^{(k)}(C)$ denote the class of functions with a continuous k th derivative on interval C , $C \subset \mathfrak{R}$. Define $\mathcal{F}(\gamma, \gamma_1)$ to be the class of distribution functions $Q(x)$ of Hölder continuity of order γ . That is, $Q(x)$ satisfies the following conditions:

- (i) $Q(x) \in \mathcal{C}^{(\alpha)}(a_0, b_0)$.
- (ii) $\sup |x_1 - x_2|^{-\beta} |Q^{(\alpha)}(x_1) - Q^{(\alpha)}(x_2)| < \gamma_1$, over $x_1, x_2 \in (a_0, b_0)$

From here on, we will assume that

(A2) F and G are in $\mathcal{F}(\gamma, \gamma_1)$, for some γ_1 and $\gamma(\geq 2)$.

For the identifiability of θ , without loss of generality, we further assume

(A3) The difference $f(x) - g(x)$ has non-zero left and right derivatives at θ of a^+ and a^- , respectively

Our estimators of θ will be based on a training set containing independent observations from two samples:

$$X_1, \dots, X_n \sim F$$

from the healthy population and

$$Y_1, \dots, Y_n \sim G$$

from the diseased population. The sample sizes m and n are assumed to be of the same order, *i.e.* $\frac{m}{n} \rightarrow \lambda > 0$.

Hsieh and Turnbull (1993) show that when F and G are in $\mathcal{F}(\gamma, \gamma_1)$, with $\gamma \geq 2$, the empirical estimator, $\hat{\theta}_n = \operatorname{argmax}(F_m(x) - G_n(x))$, $\hat{\theta}_n$ converges to θ almost surely. Here $F_m(x)$ and $G_n(x)$ are the empirical distribution functions from the two samples. They show that, with an appropriate normalizing constant C ,

$$C \cdot n^{1/3}(\hat{\theta}_n - \theta)$$

converges in distribution to the random variable, $\operatorname{argmax}(Z(t) - t^2)$, $t \in \Re$ where $Z(\cdot)$ is a standard two-sided Brownian motion. Clearly the convergence rate of $\hat{\theta}_n$ is $n^{-1/3}$. (See also Chernoff(1964) and Kim and Pollard (1990)).

Hsieh and Turnbull (1993) also discuss a smoothed estimator for θ , based on kernel density estimates, f_m and g_n say, of the densities f and g . Their “smoothed” estimator $\tilde{\theta}$ of θ is defined as a solution to the equation $f_m(x) = g_n(x)$. They show that for $F, G \in$

$\mathcal{F}(\gamma, \gamma_1), \gamma > 2$, the smoothed estimator, $\tilde{\theta}$ is strongly consistent and, with C_0 and σ^2 constants depending on f and g :

$$(nh_n)^{1/2}(\tilde{\theta}_n - \theta) \rightarrow N(0, \sigma^2) + C_0 \text{ (in distribution)}.$$

when a bandwidth of order $O(n^{-\frac{1}{2\gamma-1}})$ is used to construct the kernel density estimates.

In the next section we state and prove the main result that $\hat{\theta}_n$, in the case $F, G \in \mathcal{F}(2, \gamma_1)$, and $\tilde{\theta}_n$, in the case $F, G \in \mathcal{F}(\gamma, \gamma_1), \gamma > 2$, achieve the optimal convergence rate in the sense of being local asymptotically minimax.

It is interesting to note that for the two cases of $\gamma = 2$ and $\gamma > 2$, we use different estimators to achieve the optimal rate of convergence. Essentially this is because when $\gamma = 2$, the optimal bandwidth in density estimation is of order $O(n^{-1/3})$, yet this choice of bandwidth is too fast to assure the strong consistency of the first derivatives of f and g (see Silverman (1978)) which is needed for a Taylor series expansion of $f_m(\tilde{\theta}) - g_n(\tilde{\theta})$ around θ . This explains the use of $\hat{\theta}$ instead of $\tilde{\theta}$ when $\gamma = 2$. Thus different characteristics of θ are exploited when the smoothness conditions are different.

2. The optimal convergence rate.

Let $\theta(f, g)$ be the crossing point of f and g and $\mathcal{D}(\gamma, \gamma_1)$ denote the class of distribution pairs (F, G) satisfying (A1), (A2), (A3). In this section, we show that the optimal rate of convergence for estimating θ is $n^{-\frac{\gamma-1}{2\gamma-1}}$ in the local asymptotically minimax sense.

We impose the topology on $\mathcal{D}(\gamma, \gamma_1)$ generated by the collection of neighborhood systems:

$$\{V_{\varepsilon, \delta}(F, G) \mid \delta, \varepsilon > 0, \text{ and } C(\theta, \varepsilon) = (\theta(f, g) - \varepsilon, \theta(f, g) + \varepsilon) \subset (a_0, b_0)\}.$$

Here $(F, G) \in \mathcal{D}(\gamma, \gamma_1)$, $F' = f$ and $G' = g$, and

$$V_{\varepsilon, \delta}(F, G) = \{(F^*, G^*) \mid \|(F, G) - (F^*, G^*)\|_{\gamma, C(\theta, \varepsilon)} < \delta\},$$

where $\| (F, G) - (F^*, G^*) \|_{\gamma, I} = \max(\| F - F^* \|_{\gamma, I}, \| G - G^* \|_{\gamma, I})$, I is an interval on $\mathfrak{R}$ and $\| \cdot \|_{\gamma, I}$ is defined as

$$\| L \|_{\gamma, I} = \sum_{j=0}^{\alpha} \sup_I | L^{(j)} | + \sup_{x_1, x_2 \in I} | x_1 - x_2 |^{-\beta} | L^{(\alpha)}(x_1) - L^{(\alpha)}(x_2) |$$

This topological structure on $\mathcal{D}(\gamma, \gamma_1)$ is generalized from Hsieh and Levit (1991).

The optimal local asymptotically minimax property of the estimators of the crossing point θ in a decision theoretic framework is stated in the following theorem.

Theorem. For $\gamma \geq 2$, $(F_0, G_0) \in \mathcal{D}(\gamma, \gamma_1)$, $\ell(\cdot)$ any bounded subconvex loss function and any sufficiently small neighborhood $V \subset \mathcal{D}(\gamma, \gamma_1)$ of (F_0, G_0) , we have

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{(F, G) \in V} E_{(F, G)}[\ell(s_n(T_n - \theta(f, g)))] > 0$$

where $s_n = n^{-\frac{\gamma-1}{2\gamma-1}}$ and T_n is any estimate of $\theta(f, g)$.

Proof. Let the underlying density pair $(f_0, g_0) \in \mathcal{D}(\gamma, \gamma_1)$ and $\theta(f_0, g_0) = 0$. without loss of generality. By our assumptions, $\min(f_0(0), g_0(0)) > 0$ and

$$\lim_{x \rightarrow 0^\pm} (x)^{-1} [f_0(x) - g_0(x)] = a^\pm.$$

Define $a = (| a^+ | \wedge | a^- |) > 0$. Now consider the auxiliary function $h(x)$ with the following properties:

- (i) $h(x) = 1$ if $| x | < a^{-1}$
- (ii) $h \in \mathcal{C}^{(\alpha)}$ and $| h'(x) | < \frac{a}{2}$, $\forall x$.
- (iii) $h(x) \equiv 0$ if $| x | > M_0$ for some constant M_0 .
- (iv) $\int h(x) dx = 0$.

We consider the one dimensional subfamily of $\mathcal{D}(\gamma, \gamma_1)$:

$$\begin{aligned} f_m(x) &= f_0(x) - \xi \cdot m^{-\frac{\gamma-1}{2\gamma-1}} \cdot h(x \cdot m^{\frac{1}{2\gamma-1}}), \text{ and} \\ g_n(x) &= g_0(x) + \xi \cdot n^{-\frac{\gamma-1}{2\gamma-1}} \cdot h(x \cdot n^{\frac{1}{2\gamma-1}}) \end{aligned}$$

for $\xi \in [0, 1]$. It is clear that, for sufficiently large m and n , and small ε, δ , then $(f_m(x, \xi), g_n(x, \xi)) \in V_{\varepsilon, \delta}(F_0, G_0)$ for any fixed $\xi \in [0, 1]$.

We also have that

$$\theta(f_m, g_n) = \xi \cdot n^{-\frac{\gamma-1}{2\gamma-1}} \cdot a^{-1}(1 + \lambda^{-\frac{2(\gamma-1)}{2\gamma-1}})(1 + o(1))$$

is the unique solution of the equation $f_m(x) = g_n(x)$ in $[a_0, b_0]$. Also the Fisher information for ξ in this experiment is;

$$\begin{aligned} \Psi^2(n, m, \xi) &= m \int \frac{(\frac{\partial}{\partial \xi} f_m(x, \xi))^2}{f_m(x, \xi)} + n \int \frac{(\frac{\partial}{\partial \xi} g_n(x, \xi))^2}{g_n(x, \xi)} \\ &\rightarrow \left(\frac{1}{f(0)} + \frac{1}{g(0)} \right) \int h^2(y) dy > 0. \end{aligned}$$

With this fact we can now check the LAN conditions in Lemma 3.2 of Hasminskii (1979), and then apply the Lemma 1.1 of the same paper. This completes the proof of the theorem.

References:

- Chernoff H. (1964), Estimation of the mode, *Ann. Inst. Statist. Math.*, **16**, 31-41.
- Hasminskii, R.Z. (1979), Lower bounds for the risk of nonparametric estimates of the mode, In *Contributions to Statistics, Jaroslav Hájek Memorial Volume* (J. Jurečková, ed.). 91-97. Academia, Prague.
- Hsieh F.S. and Levit B. (1991), On the optimal rates of improvement of the sample median, Technical Report No.684, Dept. of Mathematics. Univ. of Utrecht. The Netherlands. Submitted for publication.

Hsieh, F.S and Turnbull, B.W.(1993), Nonparametric methods for evaluating diagnostic tests. In revision for *Statistica Sinica*.

Kim J. and Pollard D. (1990), Cube root asymptotics, *Ann. Statist.* **18**, 191-219.

Silverman, B.W. (1978), Weak and strong consistency of the kernel estimate of a density and its derivatives, *Ann. Statist.*, **6**, 177-184.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer* **3**, 32-35.