Nonparametric Estimation of Quadratic Functionals

In Gaussian White Noise

by

Jianqing Fan
Department of Statistics
University of California
Berkeley, CA 94720

Technical Report No. 166
July 1988

Research supported by NSF Grant No. DMS84-51753

Department of Statistics
University of California
Berkeley, California

Nonparametric Estimation of Quadratic Functionals In Gaussian White Noise

Jianqing Fan

Department of Statistics
University of California, Berkeley
Berkeley, CA 94720

ABSTRACT

Suppose that we observe the process Y(t) from

$$Y(t) = \int_0^t f(u)du + \sigma W(t), \quad t \in [0,1]$$

the optimal rates of convergence (when $\sigma \to 0$) of quadratic functionals under hyperrectangle type of constraint on unknown function $f(t)$ is addressed, where $W(t)$ is a standard Wiener Process on [0, 1]. Specially, the optimal rate of estimating $\int_0^1 [f^{(k)}(x)]^2 dx$ under the hyperrectangle constraint $\Sigma = \{f : x_j(f) \le n^{-p}\}$ is σ when $p \ge 0.75 + 2k$, and $\sigma^{2-(4k+1)/(4p-1)}$, when $k + 0.25 < p < 0.75 + 2k$, where $x_j(f)$ is the jth Fourier-Bessel coefficient of unknown function f. The lower bounds are found by applying the hardest 1-dimensional approach as well as testing method. We use the lower bounds to compare how far the lower bounds apart from the upper bounds, not only in rates but also in constant factors, which gives us an idea how efficient the best "unbiased" truncation estimator is in comparison to the possible best estimator. Indeed, with such a comparison, we find the asymptotic minimax estimator for estimating the functional $\int_0^1 f^2(x)dx$ when $p \ge 0.75$. The more general testing bound are also discussed in the current setting.

Key Words and Phrases: Quadratic functionals, Lower bounds, Rates of convergence, Minimax risk, Gaussian white noise, Estimation of bounded squared-mean, testing of hypotheses, Hardest 1-dimensional subproblem, Bayesian approach.

1. Introduction

Suppose the information available for unknown function f(t) is the observation from

$$Y(t) = \int_0^t f(u)\, du + \sigma W(t), \quad t \in [0, 1] \tag{1.1}$$

with constraint $f(t) \in \Sigma$, a subset of $L^2[0, 1]$, where W(t) is a standard Wiener process on [0, 1]. We want to estimate a functional $T(f)$ nonparametrically. The functionals of particular interest in this paper are quadratic functionals, with Σ a hyperrectangle. For example, we want to estimate

$$T(f) = \int_0^1 [f^{(k)}(t)]^2\, dt \tag{1.2}$$

where $f^{(k)}(t)$ is the kth derivative of the unknown function f(t). A natural question is how well can we estimate the functional? How well does the best quadratic estimator behave? Ibragimov *et al* (1986) made a start on questions of this kind.

Such a model is apparently the most useful model for studying the properties of nonparametric estimates, as pointed out by Ibragimov *et al* (1986), since the statistical nature of the problem is not complicated. Donoho and Liu (1988) connect such model with nonparametric density estimation by heuristic argument that the information available from an empirical process is almost the same as model (1.1) because the empirical process tends weakly to a Browian bridge. Indeed, the result of Efroimovich and Pinsker (1982) -- under ellipsoid type of constraint, the best linear estimator of density estimator (found explicitly) is an efficient estimator among all possible estimators -- is motivated by Pinsker's(1980) study of model (1.1). Our results together with recent findings of Bickel and Ritov (1988) also support such a heuristic.

The discretized version of model (1.1) is as follows: fix an orthonormal basis $\{\phi_j(t)\}_1^\infty$ in $L^2[0, 1]$, expand the functional $dY(t)$, $f(t)$, $dW(t)$ on this fixed basis. Then the observation from (1.1) is equivalent to observe their ith coordinate from

$$y_i = x_i + \sigma z_i \qquad (i = 1, 2, \cdots) \tag{1.3}$$

where y_i, x_i, z_i are the ith coordinate of $dY(t)$, $f(t)$, $dW(t)$ defined by

$$y_i = \int_0^1 \phi_i(t)\, dY(t), \quad x_i = \int_0^1 \phi_i(t)\, f(t)\, dt, \quad z_i = \int_0^1 \phi_i(t)\, dW(t) \tag{1.4}$$

and the functional $T(f)$ of interest is equivalent to $T(x)$, and constraint Σ is the same as a subset in R^∞.

The quadratic functionals under consideration are

$$Q(x) = \sum_1^\infty \lambda_i\, x_i^2 \tag{1.5}$$

with hyperrectangle type of constraint

$$\Sigma = \{x \in R^\infty : |x_i| \le A_i\} \tag{1.6}$$

To make problem meaningful ($Q(x) < \infty$, for $x \in \Sigma$), assume additionally that

$$\sum_1^\infty \lambda_i\, A_i^2 < \infty \tag{1.7}$$

Specially, when $\lambda_j = j^{2k}$, and $A_j = j^{-p}$ ($p \ge k + 0.25$), then $Q(x)$ defined by (1.5) is the same as that defined in (1.2) with the smoothness constraint on f(t).

For finding the optimal rates of convergence, mathematically it involves on one hand, to find a lower bound, which says that no estimator can do better than the lower bound, and on the other hand, to find an reasonable estimator to achieve the rate of the lower bound. For the current setting, it is not hard to find an intuitive estimator, which is good enough to allow us to obtain an optimal rate of the convergence.

Finding a lower bound usually involves more mathematical techniques. An usual way to find a minimax lower bound is by assigning an appropriate prior. However, we will show that no *intuitive* Bayes method works in the current setting. For the setting of nonparametric density estimation, the mathematical background behind the construction of Farrell (1972) and Stone (1980) is that no perfect test, no good estimate even though they didn't state clearly.

The generalization of such an idea is illustrated in Donoho and Liu (1987b,c, 1988).

When T is a linear functional, Donoho and Liu (1987a) shows that when the constraint is convex, balance and symmetric, under quadratic loss, the maximum risk of best linear estimator is no worse than that of the best estimator times 1.34 by using the hardest 1-dimensional trick back to Stein (1956). The optimal rate of the convergence is equivalent to the modulus function defined by

$$b_T(\sigma) = sup \ \{ \ |T(x_1) - T(x_2)|: \ ||x_1 - x_2||^2 = \sigma^2, \ x_1 \in \Sigma, \ x_2 \in \Sigma \} \tag{1.8}$$

We will see that no matter what kind of functional we want to estimate and what kind of constraint we have, the lower bound $b_T(\sigma)$ is always available to be a lower bound. However, the attainability of the lower bound $b_T(\sigma)$ fails already even for quadratic functionals.

Looking inside of the modulus function defined above, it says by no mean that once we cannot test between x_1, and x_2 from the observations, then the change of the functional of $T(x)$ is the least mistake we will make for estimation of $T(x)$. For the case that modulus bound doesn't work, one may argue that the pair of $x_1 \in \Sigma$, and $x_2 \in \Sigma$ may not be one of the most difficult pairs of subsets to be tested among all subsets having the same change of the functional based on our observation (1.3). Thus, a more economical and automatic way to formulate a pair of subsets (usually composite) to be tested is that let the change of functional of two sets be at least Δ, and then see how large Δ will be in order to have a good but no perfect test to separate two subsets. The idea of this kind is due to Donoho and Liu (1987c).

Contents of Paper: We begin with finding the best "unbiased" truncation quadratic estimator, and compute its MSE error. Then, we apply the hardest 1-dimensional method to give a lower bound, which shows that when $p \geq q + 0.75$, the best "unbiased" truncation quadratic estimator is quite efficient in asymptotic minimax sense. When $(q + 1)/2 < p < q + 0.75$, the hardest 1-dimensional approach does not give sharp lower bounds. To handle this case, we develop a new lower bound, based on the Bayes risk in testing between two highly composite priors, which shows that the best "unbiased" truncation quadratic estimator gives the optimal rate in this case. However, the lower bound and upper bound is quite different in the constant factor. Motivated by the testing method, we will suggest how to assign an appropriate prior to give a bigger (possible) lower bound (in constant factor). Finally, we generalize the idea of testing method to more general setting.

2. Quadratic estimators

Let start with the model (1.3). Suppose we observe

$$y = x + \sigma z \tag{2.1}$$

with hyperrectangle type of constraint (1.6). A intuitive class of quadratic estimators to estimate

$$Q(x) = \sum_1^\infty \lambda_i \, x_i^2 \tag{2.2}$$

(where $\lambda_j \geq 0$) is the the class of estimators defined by

$$q_B(y) = y'By + c \tag{2.3}$$

where B is a symmetric matrix, and c is a constant. Simple algebra shows the risk of $q_B(y)$ under quadratic loss is

$$R(B, x) \triangleq E \, (q_B(y) - Q(x))^2 \tag{2.4}$$

$$= (x'Bx + \sigma^2 \, trB + c - Q(x))^2 + 2\sigma^4 \, trB^2 + 4\sigma^2 \, x'B^2x \tag{2.5}$$

Following Proposition tells us the class of quadratic estimators with diagonal matrix B is a complete class among all estimators defined by (2.3).

Proposition 2.1: Let D_B be a diagonal matrix, whose diagonal elements are those of B. Then for each symmetric B ,

$$\max_{x \in \Sigma} R(B, x) \geq \max_{x \in \Sigma} R(D_B, x)$$

where Σ is defined by (1.6).

Thus only diagonal matrix B is needed to be considered. For diagonal matrix

$$B = diag(b_1, b_2, \ldots.) \tag{2.6}$$

the estimator (2.3) has risk

$$R(B, x) = (\sum_1^\infty b_j x_j^2 + \sigma^2 \sum_1^\infty b_j + c - \sum_1^\infty \lambda_j x_j^2)^2 + \sum_1^\infty b_j^2 (2\sigma^4 + 4\sigma^2 x_j^2) \tag{2.7}$$

A natural question is when the estimator (2.3) with B defined by (2.6) converges almost surely. Following Proposition may state in some standard probability book.

Proposition 2.2: Under model (2.1), $q_B(y)$ converges almost surely for each $x \in \Sigma$ iff

$$\sum_1^\infty b_i (A_i^2 + \sigma^2) < \infty \tag{2.8}$$

Even though for diagonal matrix B, it is hard to find the best quadratic estimator. For the infinite dimensional estimation problem, usually the bias is a one of major contribution to the risk. Thus, we would prefer to use a unique unbiased quadratic estimator

$$\sum_1^\infty \lambda_j (y_j^2 - \sigma^2)$$

but it may not convergence in L^2, and even it does converge, it may contribute too much on variance. Because of convergence condition (1.7), we know that after certain dimension m (which will go to infinite, as $\sigma \to 0$), we are not worth to estimate $\sum_m^\infty \lambda_j \, x_j^2$. Thus, "unbiased" truncation quadratic estimator

$$q_{UT}(y) = \sum_1^m \lambda_j (y_j^2 - \sigma^2) \tag{2.9}$$

is a good candidate for estimating Q(x), and m is chosen to minimize its MSE. For estimator $q_{UT}(y)$ just defined above, the maximum MSE is

$$\max_{x \in \Sigma} R(q_{UT}, x) = (\sum_m^\infty \lambda_j A_j^2)^2 + \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 A_j^2) \tag{2.10}$$

When $\lambda_j = j^q$ and $A_j = j^{-p}$, $(p > (q+1)/2)$, (2.10) can be simplified a lot. Let's study its asymptotic property as $\sigma \to 0$. By (2.10), the estimator

$$\sum_1^m j^q (y_j^2 - \sigma^2) \tag{2.11}$$

has its maximum risk

$$R(m) \triangleq \frac{2m^{2q+1}}{2q+1}\sigma^4(1+o(1)) + 4\sigma^2 \sum_1^m j^{2q-2p} + (\frac{m^{-(2p-q-1)}}{2p-q-1})^2(1+o(1))$$

Case I: $(q+1)/2 < p \leq q + 0.75$.

Choose $m = [c_0 \sigma^{-\frac{4}{4p-1}}]$, where c_0 is specified below. Let

$$g(c) = \frac{2c^{2q+1}}{2q+1} + c^{-2(2p-q-1)}/(2p-q-1)^2 \tag{2.12}$$

Then the minimizer of g(c) is $c_0 = (2p-q-1)^{-\frac{1}{4p-1}}$. Thus, (2.11) is simplified as

$$g(c)\,\sigma^{4-\frac{4(2q+1)}{4p-1}}\,(1+o(1)) \tag{2.13}$$

and the optimal m is

$$m_0 = [(2p-q-1)^{-\frac{1}{4p-1}}\,\sigma^{-\frac{4}{4p-1}}] \tag{2.14}$$

Thus risk of the optimal estimatior "unbiased" is

$$g(c_0)\sigma^{4-\frac{4(2q+1)}{4p-1}}\,(1+o(1)) \tag{2.15}$$

Case II: $p > q + 0.75$

The optimal $m_0 = [\sigma^{-\frac{4}{4p-1}}]$, with risk

$$4\sum_1^\infty j^{2q-2p}\sigma^2(1+o(1)) \tag{2.16}$$

Let summarize the result we obtained above

Theorem 1: For $\lambda_j = j^q$ and $A_j = j^{-p}$, $(p > (q+1)/2)$, the best "unbiased" truncation estimated is given by (2.11) with optimal m_0 defined by (2.14) when $(q+1)/2 < p \le q + 0.75$, and the maximum risk of the optimal estimator is given by (2.15), and the optimal $m_0 = [\sigma^{-\frac{4}{4p-1}}]$ when $p > q + 0.75$ with the maximum risk given by (2.16).

Note that when $(q+1)/2 < p \le q + 0.75$ the rate given above is bigger than σ^2. In fact, in the next two section, we will show that the "unbiased" truncation estimator is quite efficient in the sense that its risk is quite close to lower bounds, and the rates given above are optimal. Note the when q = 0, we will actually prove that the best "unbiased" truncation estimator is asymptotic efficient in minimax sense.

In the following examples, we always assume that the constraint is $\Sigma = \{x : |x_j| \le j^{-p}\}$.

Example 1: Suppose we want to estimate $T(f) = \int_0^1 f^2(t)\,dt$ from model (1.1). Let $\{\phi_j(t)\}$ be a fixed orthonormal basis. Then $T(f) = \sum_{j=1}^{\infty} x_j^2$. Thus, the optimal "unbiased" estimator is $\sum_1^{m_0} \lambda_j (y_j^2 - \sigma^2)$, where

$$m_0 = \begin{cases} [\sigma^{-\frac{4}{4p-1}}], & \text{if } p > 0.75 \\ [(2p-1)^{-\frac{1}{4p-1}} \sigma^{-\frac{4}{4p-1}}], & \text{if } 0.5 \le p \le 0.75 \end{cases}$$

Moreover, when $p > 0.75$, the estimator is an asymptotic minimax estimator. For $0.5 \le p \le 0.75$, the optimal rates are achieved.

Example 2: Let orthonormal basis be $\{\phi_j(t)\} = \{1, \frac{1}{\sqrt{2}}\cos 2\pi jt, \frac{1}{\sqrt{2}}\sin 2\pi jt\}$. We want to estimate

$$T(f) = \int_0^1 [f^{(k)}(t)]^2\,dt = (2\pi)^k/2 \sum_{j=2}^{\infty} j^{2k}\, x_j^2$$

The optimal "unbiased" estimator is $(2\pi)^k/2 \sum_{j=2}^{m_0} j^{2k}\,(y_j^2 - \sigma^2)$, where

$$m_0 = \begin{cases} [\sigma^{-\frac{4}{4p-1}}], & \text{if } p > 2k + 0.75 \\ [(2p - 2k - 1)^{-\frac{1}{4p-1}} \sigma^{-\frac{4}{4p-1}}], & \text{if } k + 0.5 \le p \le 2k + 0.75 \end{cases}$$

Moreover, the estimators achieve the optimal rates.

Example 3: (Inverse Problem) Suppose we are interesting in recovering the indirectly observed function f(t) from data of the form (Donoho & MacGibbon (1987))

$$y(u) = \int_0^u \int_0^s K(t, s) f(t) dt \, ds + \sigma \int_0^u dW, \quad u \in [0, 1]$$

where again W is a Wiener Process and K is known. Let $K: L_2[0, 1] \to L_2[0, 1]$ have a singular system decomposition (Bertero *et al* (1982)), i.e. a representation

$$Kf = \sum_{i=1}^{\infty} \lambda_i (f, \xi_i) \eta_j$$

where the λ_i are singular values, the $\{\xi_i\}$ and $\{\eta_i\}$ are orthogonal sets in $L_2[0, 1]$. Then the observation is equivalent to

$$y_i = \lambda_i \theta_i + \sigma \varepsilon_i$$

where ε_i are i.i.d. $N(0, 1)$, θ_i is the Fourier-Bessel coefficient of (f, ξ_i), and y_i is a Fourier-Bessel coefficient of the observed data. Now suppose want to estimate

$$\int_0^1 f^2(t) \, dt = \sum_1^{\infty} \theta_i^2 = \sum_1^{\infty} \lambda_i^{-2} x_i^2$$

where $x_i = \lambda_i \theta_i$. If the non-parametric constraint on θ is a hyperrectangle, then the constraint of x is also a hyperrectangle in R^{∞}. Applying Theorem 1, we can get an "optimal" estimator, which achieve the optimal rate. Moreover, we will know roughly how efficient the estimator is if we apply the Table 3.1 and 4.1.

3. Hardest 1-dimensional lower bound

The hardest 1-dimensional trick is suggested by Stein (1956). The idea is to use the difficulty of the hardest 1-dimensional problem as a lower bound of that of nonparametric problem. Thus, for estimation of quadratic functionals (infinite dimensional), we will end up with the difficulty of estimation of θ^2 from observation $Y \sim N(\theta, \sigma^2)$ (1-dimensional).

Let Y be a random variable distributed as N(θ, σ^2), and suppose it is known that $|\theta| \le \tau$. Let the minimax risk of estimating θ^2 be

$$\rho(\tau, \sigma^2) = \inf_{\delta} \sup_{|\theta| \le \tau} E_\theta(\delta(Y) - \theta^2)^2 \tag{3.1}$$

By definition, it is easy to show that

$$\rho(\tau, \sigma^2) = \sigma^4 \rho(\tau/\sigma, 1) \tag{3.2}$$

For estimation of bounded mean, the quantity

$$\inf_{\delta} \sup_{|\theta| \le \tau} E_\theta(\delta(Y) - \theta)^2 \tag{3.3}$$

has been studied by Casella and Strawderman (1981), Bickel (1982). However $\rho(\tau, \sigma^2)$ behaves much different from the minimax risk of estimation of the bounded mean (3.3). For the purpose of later use, we give the following asymptotic result of $\rho(a, 1)$.

Theorem 2: As $a \to \infty$,

$$\rho(a, 1) = 4a^2(1 + o(1)) \tag{3.4}$$

Moreover the least favorable prior is the limit (as $n \to \infty$) of the prior density of

$$g_n(\theta, a) = \frac{(2n+1)\theta^{2n}}{2a^{2n+1}} 1_{(|\theta| \le a)} \tag{3.5}$$

Now we are ready to find a lower bound by using the hardest 1-dimensional trick. Fix a point $x \in \Sigma$. Let $x_t = tx$, $t \in [-1, 1]$, be a line segment passing the origin. Then the problem of estimating of $Q(x)$ in $\{ x_t \}$ is just that of estimating

$$\theta^2 = Q(tx) = t^2 Q(x) \tag{3.6}$$

from observation

$$y = tx + \sigma z$$

where z is normal with mean 0 and variance identity, and t is unknown to be estimated. As $(y, x) \sim N(t\|x\|^2, \sigma^2\|x\|^2)$ is a sufficient statistics for t, the observation available for estimation t or θ is equivalent to that from $\frac{\sqrt{Q(x)}}{\|x\|^2}(y, x) \sim N(\theta, \frac{Q(x)}{\|x\|^2}\sigma^2)$. Thus the minimax risk of estimation θ^2 in the problem is

$$\rho(\sqrt{Q(x)}, \frac{Q(x)}{\|x\|^2}\sigma^2) \tag{3.7}$$

according to our notation. By (3.2),

$$\rho(\sqrt{Q(x)}, \frac{Q(x)}{\|x\|^2}\sigma^2) = \frac{Q(x)^2\sigma^4}{\|x\|^4}\rho(\|x\| / \sigma, 1) \tag{3.8}$$

But, the minimax risk of estimation $Q(x)$ is at least as difficulty as that of estimation θ^2. Consequently,

$$\inf_{\delta} \sup_{x \in \Sigma} E_x(\delta(y) - Q(x))^2$$

$$= \inf_{\delta} \sup_{x \in \Sigma} \sup_{|t| \leq 1} E_{tx}(\delta(y) - t^2 Q(x))^2$$

$$\geq \sup_{x \in \Sigma} \inf_{\delta} \sup_{|t| \leq 1} E_{tx}(\delta(y) - \theta^2)^2$$

$$\geq \sup_{x \in \Sigma} \frac{Q^2(x)\sigma^4}{\|x\|^4}\rho(\|x\| / \sigma, 1)$$

In conclusion,

Theorem 3: The minimax risk of estimation $Q(x)$ from observation (2.1) is at least

$$\sup_{x \in \Sigma} \frac{Q(x)^2\sigma^4}{\|x\|^4}\rho(\|x\| / \sigma, 1) \tag{3.9}$$

and as $\sigma \to 0$, the minimax lower bound is at least

$$\sup_{x \in \Sigma, \|x\| > 0} \frac{4Q^2(x)}{\|x\|^2}\sigma^2(1 + o(1)) \tag{3.10}$$

Comparing the lower bound and upper bound given by (2.16), when $p \geq q + 0.75$ the rate σ^2 is the optimal one. When p = 0 , the best truncation is the asymptotic minimax estimator. How close is the upper bound to lower bound? By (2.16) and (3.10), as $\sigma \to 0$

$$\frac{\text{Lower Bound}}{\text{Upper Bound}} \geq \frac{C_{2p-q}^2}{C_{2p}\, C_{2p-2q}} \tag{3.11}$$

where $C_r = \sum_1^{\infty} j^{-r}$ can be calculated numerically. Following table shows the result:

Table 3.1: Comparison of the lower bound and upper bound
p = 1 + 1 + 0.5 i

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8
q=0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
=0.5	0.976	0.991	0.996	0.998	0.999	1.000	1.000	1.000
=1.0	0.940	0.977	0.990	0.995	0.998	0.999	0.999	1.000
=1.5	0.910	0.963	0.984	0.992	0.996	0.998	0.999	0.999
=2.0	0.887	0.952	0.979	0.990	0.995	0.998	0.999	0.999
=2.5	0.871	0.944	0.975	0.988	0.995	0.997	0.998	0.999
=3.0	0.859	0.938	0.972	0.987	0.994	0.997	0.998	0.999
=3.5	0.851	0.934	0.970	0.986	0.993	0.996	0.998	0.999
=4.0	0.845	0.931	0.968	0.985	0.993	0.996	0.998	0.999
=4.5	0.842	0.929	0.967	0.984	0.992	0.996	0.998	0.999
=5.0	0.839	0.928	0.966	0.984	0.992	0.996	0.998	0.999

4. Testing of the vertices of hypercube

Note that when $(q+1)/2 \le p < q + 0.75$, the upper bound (2.15) is not at the rate of σ^2, and consequently, the hardest 1-dimensional lower bound (3.10) is much smaller. We may wonder whether 1-dimensional bounds cannot capture the difficulty of estimating quadratic functionals or else the best "unbiased" truncation estimator is not a good estimator in this case. Indeed, we will construct a bigger lower bound, which says the best "unbiased" truncation estimator gives the optimal rate.

The idea of the construction is as follows: take a largest hypercube of dimensional n (which depends on σ) in the hyperrectangle, and assign probability $\frac{1}{2^n}$ to each vertices, and then test them against the origin. When no perfect test exists (by choosing some critical value n, depending on σ), the difference in functional values at two hypercubes supplies a lower bound. The approach we use is related to one of Ibragimov *et al* (1986), who, however use hypersphere rather than hyperrectangle.

To carry out the idea, let's formulate testing problem

$$H_0\colon x_i = 0, \leftrightarrow H_1\colon x_i = \pm t_n \ (i=1, \cdots, n), 0\ (i > n), \tag{4.1}$$

i.e. we want to test whether the observation y_i from H_0 or H_1, which is specified as follows:

$$H_0\colon y_i \sim N(0, \sigma^2)\ (i = 1, \cdots, n) \leftrightarrow \tag{4.2}$$

$$H_1\colon y_i \sim \frac{1}{2}\,[\phi(y, t_n, \sigma) + \phi(y, -t_n, \sigma)],\ \ (i = 1, \cdots, n).$$

where $\phi(y, t, \sigma)$ is the density function of $N(t, \sigma^2)$. The result of the testing can be summarized as follows:

Lemma 4.1: The sum of type I and type II error of the best testing procedure is

$$2\,\Phi\Big(-\frac{\sqrt{n}\,(t_n / \sigma)^2}{\sqrt{8}}\Big)(1 + o(1)) \tag{4.3}$$

if $n^{1/2}\,(t_n / \sigma)^2 \to c$, where $\Phi(\cdot)$ is the probability function of a standard normal distribution.

Now we are ready to derive a lower bound. Let

$$r_n = |Q(H_1) - Q(H_0)|/2 = t_n^2 \sum_1^n \lambda_j/2 \tag{4.4}$$

be the half of change of the functional. For any estimator $T(y)$, the minimax risk under quadratic loss is

$$\begin{aligned}
&\sup_{x \in \Sigma} E_x\,(T(y) - Q(x))^2 \\
&\geq \frac{1}{2}\left[E_0\,(T(y) - Q(x))^2 + E_1\,(T(y) - Q(x))^2\right] \\
&\geq \frac{1}{2} r_n^2 \left[P_0\,(|T(y)| \geq r_n) + P_1(|T(y) - Q(x)| \geq r_n)\right] \\
&\geq \frac{1}{2} r_n^2 \left[P_0\,(|T(y)| \geq r_n) + P_1(|T(y)| \leq r_n)\right] \qquad (4.5) \\
&\geq r_n^2\,\Phi(-\frac{\sqrt{n}\,(t_n / \sigma)^2}{\sqrt{8}})(1 + o(1))
\end{aligned}$$

where E_0 and E_1 mean that take the expectation under H_0 and H_1, respectively.

The last inequality holds because we can view the second term of (4.5) as sum of type I error and type II error, which is no smaller than that of best test procedure given by Lemma 4.1. Thus, take n such that the last term is bounded away from 0, then r_n^2 is the order of the lower bound.

Theorem 4: Suppose that A_j is increasing in j, when j is large. Let n_σ be the largest number of the equation

$$\sqrt{n}\,(A_n/\sigma)^2 \leq c$$

Then, for any estimator $T(y)$, the maximum risk of estimating $Q(x)$ is no smaller than

$$\sup_c \{ \frac{\Phi(-c/\sqrt{8})}{4} (\sum_1^{n_\sigma} \lambda_j)^2 A_{n_\sigma}{}^4 \}\ (1 + o(1)) \quad (as\ \sigma \to 0). \tag{4.6}$$

Moreover,

$$\sup_{x \in \Sigma} P_x\{|T(y) - Q(x)| \geq \frac{(\sum_1^{n_\sigma} \lambda_j) A_{n_\sigma}^2}{2}\} \geq \Phi(-\frac{c}{\sqrt{8}})\,(1 + o(1))$$

and for any symmetric increasing loss function $l(\cdot)$,

$$\sup_{x \in \Sigma} E_x\,[l(\frac{2|T(y) - Q(x)|}{\sum_1^{n_\sigma} \lambda_j\, A_{n_\sigma}^2})] \geq l(1)\,\Phi(-\frac{c}{\sqrt{8}}) + o(1)$$

When $A_j = j^{-p}$ and $\lambda_j = j^q$, we can calculate the rate in Theorem 4 explicitly.

Corollary 4.1: When $A_j = j^{-p}$ and $\lambda_j = j^q$, for any estimator the minimax risk under quadratic loss is no better than

$$\xi_{p,q}\,\sigma^{4 - \frac{4(2q+1)}{4p-1}}\,(1 + o(1))$$

Moreover, for any estimator T(y),

$$\sup_{x \in \Sigma} P_x\{|T(y) - Q(x)| \geq c^{1 - \frac{2q+1}{4p-1}}/(2(q+1))\,\sigma^{2 - \frac{2(2q+1)}{4p-1}}\} \geq \Phi(-\frac{c}{\sqrt{8}})(1 + o(1))$$

where

$$\xi_{p,q} = \sup_{c > 0} c^{2 - \frac{4q+2}{4p-1}}\,\Phi(-\frac{c}{\sqrt{8}})/(4(q+1)^2)$$

When $(q+1)/2 \leq p < q + 0.75$, by comparing the lower bound and upper bound given in section 2, again we show that the best "unbiased" truncation estimator gives the optimal rate. Following table shows how close the lower bound and the upper bound are.

Table 4.1: Comparison of the lower bound and the upper bound

q = 0		q = 1		q = 2		q = 3		q = 4	
p =	ratio	p =	ratio	p =	ratio	p=	ratio	p=	ratio
0.55	0.0333	1.15	0.0456	1.75	0.0485	2.35	0.0495	2.95	0.0499
0.60	0.0652	1.30	0.0836	2.00	0.0864	2.70	0.0864	3.40	0.0858
0.65	0.0949	1.45	0.1159	2.25	0.1170	3.05	0.1153	3.85	0.1132
0.70	0.1218	1.60	0.1431	2.50	0.1419	3.40	0.1383	4.30	0.1345

where $ratio = \sqrt{\dfrac{\text{lower bound by testing method}}{\text{upper bound by } q_{UT}(y)}}$

The above table tells us that there is a large discrepancy at the level of constants between the upper and lower bounds. It appears to us that the hypersphere bound of Ibragimov *et al* (1986) would not give a lower bound of the same order as we are able to get via hypercubes.

5. Bayes Method

Often we use Bayes method to find a minimax lower bound. However, as mentioned in the introduction, the *intuitive* Bayes method gives too small lower bound in the case of $(q + 1)/2 \le p < q + 0.75$. By *intuitive* Bayes method, we mean that assign prior uniformly on the hyperrectangle, which is equivalent to say that independently assign prior uniformly on each coordinate, or more generally assign prior $x_j \sim \pi_j(\theta)$ independently. Mathematically, assigning the prior in this way does not include the interaction term (which has order n^2 terms) and hence give a smaller lower bound. To see this, let $\rho_{\pi_j}(j^{-p}, \sigma)$ be the Bayes risk of estimation θ^2 from $Y \sim N(\theta, \sigma^2)$ knowing $|\theta| \le j^{-p}$ with prior π_j. Let $\delta(y) = E(\sum_1^\infty j^q x_j^2 \mid y)$ be the Bayes solution of the problem. By independent assumption of prior,

$$\delta(y) = \sum_1^\infty j^q E(x_j^2 | y_j)$$

Then the Bayes risk is

$$E_\pi E_x (\hat{\delta}(y) - \sum_1^\infty j^q x_j)^2 = \sum_1^\infty j^{2q} \rho_{\pi_j}(j^{-p}, \sigma)$$

$$\le \sum_1^n j^{2q} \rho_{\pi_j}(j^{-p}, \sigma^2) + \sum_{n+1}^\infty j^{2q-4p} \tag{5.1}$$

because the Bayes risk is no larger than the maximum risk of estimator 0. As Bayes risk is no bigger than the minimax risk, by Theorem 2, the last quantity is no bigger than

$$\sum_1^n j^{2q} \rho_{\pi_j}(\frac{j^{-p}}{\sigma}, 1)\sigma^4 + O(n^{-4p+2q+1})$$

$$\le O(n^{2q-2p+1}\sigma^2) + O(n^{-4p+2q+1})$$

$$= \max(O(\sigma^2), O(\sigma^{4-\frac{2q+1}{p}}))$$

$$= o(\sigma^{4-\frac{4(2q+1)}{4p-1}}) \quad (when\ (q+1)/2 < p < q+0.75)$$

by choosing $n = \sigma^{-\frac{1}{p}}$.

To find a bigger lower bound, motivated by the testing method, take a largest n dimensional hypercube in Σ. Assign prior $\pi(x)$ on the diagonal line segments starting from the origin uniformly, with probability $\frac{1}{2^n}$ to each line segment of the hypercube.

Mathematically, assigning the prior in this way can be reduced to find the Bayes risk of the estimating θ^2 based on the n i.i.d. observations y_i having density

$$y_i \sim \frac{1}{2}(\phi(\theta, 1) + \phi(-\theta, 1))$$

with prior $\pi(\theta)$ is uniformly distributed on $[0, t_n]$, and denote the Bayes risk by $B(t_n, n)$, where $\phi(\theta, 1)$ is the density of normal distribution with mean θ and variance 1. Then, a minimax lower bound of estimating $Q(x)$ is at least as big as the Bayes risk, i.e. no smaller than the quantity

$$\min_{\delta(y)} \max_{|\theta| \le A_n} E(\delta(y) - \sum_1^n \lambda_j \theta^2)^2$$

$$\ge (\sum_1^n \lambda_j)^2 \sigma^4 B(A_n/\sigma, n)$$

We believe that for the case $A_n = n^{-p}$, and $\lambda_j = j^q$ by choosing suitable dimension $n = [(c\sigma)^{-\frac{4}{4p-1}}]$, the lower bound might be improved. We are not going to pursue that here.

6. General testing lower bound

In this section, we will generalize the testing bound to more general setting, which applies to all functionals to be estimated and any kind of constraint. Suppose we have observation from (1.3), and want to estimate $T(x)$ knowning $x \in \Sigma$. Then no estimator can estimate T(x) faster than the modulus bound defined by (1.8).

Theorem 5: For any estimator $\delta(y)$ from observation (1.3),

$$\sup_{x \in \Sigma} E_x (\delta(y) - T(x))^2 \ge \sup_c \Phi(-c/2) b_T^2(c\sigma)/4$$

where b_T is defined by (1.8).

Remark: When the hardest 1-dimensional lower bound and modulus bound both give the correct rates for estimating quadratic functional ($p \ge q + 0.75$), the modulus bound gives a smaller constant factor. In fact, by Cauchy-Schwartz inequality,

$$b_T(\varepsilon) \le 2\sqrt{\sum j^{2q} j^{-2p}}\,\varepsilon = 2\sqrt{C_{2(p-q)}}\,\varepsilon \quad (p - q > 0.5)$$

and

$$\frac{\text{Modulus bound}}{\text{1-dim bound}} \le \frac{\sup_c [\Phi(-c/2)c^2]\, C_{2(p-q)} C_{2p}}{4\, C_{2p-q}^2} = 0.167 \times \frac{C_{2(p-q)}\, C_{2p}}{C_{2p-q}^2},$$

where $C_r = \sum_1^\infty j^{-r}$. The second is evaluated numerically in Table 3.1. For p=0, the last term is no bigger than 0.167 and sometimes even much worse, if we evaluate both bounds more carefully.

Applying the testing trick in section 4, we can give a more general result which allow us to judge whether the modulus lower bound is too small or not. To state the result, without loss of generality, assume that $T(0) = 0$. Let l_n be the length of the largest n-dimensional hypercube in Σ, with $l_n > 0$ and $l_n \to 0$, and N_σ be the largest integer such that

$$n^{0.25}\, l_n \geq \sigma$$

Theorem 6: Under the above notations, assume that in the neighborhood of origin, $|T(x)| \geq Q(x)$, which is symmetric in each argument. Then the rate of $Q(x_\sigma)$ is a lower rate, in the sense that no estimator can estimate T(x) faster than that of $Q(x_\sigma)$ (as $\sigma \to 0$), where $x_\sigma = (\sigma, \cdots, \sigma, 0, 0, \cdots)/N_\sigma^{0.25}$ with N_σ non-zero elements.

If moreover $b_T(c\sigma)/Q(x_\sigma) \to 0$ for each c. Then modulus bound $b_T(c\sigma)$ does not give an achievable lower rate under quadratic loss, where b_T is defined by (1.8).

As mentioned in the introduction, 2-points testing bound may not be the best pair to be tested. To give a more powerful lower bound, let $\Sigma_{T \leq t} = \{x \in \Sigma: T(x) \leq t\}$ and $\Sigma_{T \geq t + \Delta}$ defined similarly be the two subsets to be tested. Let $conv(\Sigma_{T \geq t + \Delta})$ be the convex hull of $\Sigma_{T \geq t + \Delta}$ in the space of distributions, and define $conv(\Sigma_{T \leq t})$ similarly. Le Cam (1985) shows minimax risk of the sum of type I and type II error of testing

$$H_0: x \in \Sigma_{T \geq t + \Delta} \leftrightarrow H_1: x \in \Sigma_{T \leq t} \tag{6.1}$$

is just the maximum testing affinity of the hardest pair in $conv(\Sigma_{T \geq t + \Delta})$ and $conv(\Sigma_{T \leq t})$, and denote it by $\pi(conv(\Sigma_{T \geq t + \Delta}), conv(\Sigma_{T \leq t})$,i.e.

$$\pi(conv(\Sigma_{T \geq t + \Delta}), conv(\Sigma_{T \leq t}) =$$

$$\min_{0 \leq \phi \leq 1} \max_{x \in \Sigma_{T \geq t + \Delta} \bigcup \Sigma_{T \leq t}} (E_x \phi(y) + E_x(1 - \phi(y))). \tag{6.2}$$

The sum of type I and type II of the most difficult subsets to is tested be

$$\alpha_T(\Delta, \sigma) = \sup_t \pi(conv(\Sigma_{T \geq t + \Delta}), conv(\Sigma_{T \leq t}) \tag{6.3}$$

Hence the largest change of the functional Δ of the pair that we cannot test

$$\Delta_T(\alpha, \sigma) = sup\{\Delta\colon \alpha_T(\Delta,\sigma) \ge \alpha\} \tag{6.4}$$

Then, we have following theorem, which generalizes all kinds of testing lower bound given above.

Theorem 7: For any estimator $\delta(y)$,

$$\inf_{\delta} \sup_{x \in \Sigma} P_x\{|\delta(y) - T(x)| \ge \Delta_T(\alpha,\sigma)/2\} \ge \alpha/2$$

and for any symmetric increasing loss $l(t)$, the minimax risk lower bound is given

$$\inf_{\delta} \sup_{x \in \Sigma} E_x\, l\left(\frac{2(\delta(y) - Q(x))}{\Delta_T(\alpha,\sigma)}\right) \ge l(1)\, \frac{\alpha}{2}$$

7. Proofs

Proof of Proposition 2.1:

For any subset $S \subset \{1, 2, \ldots\}$, let prior μ^S be the probability measure of independently assigning $x_j = \pm A_j$ with probability $\frac{1}{2}$ each, for $j \in S$ and assigning probability 1 to the point 0 for $j \notin S$. Then by the Jensen's inequality,

$$\max_{x \in \Sigma} R(B, x)$$

$$\ge \max_{S} E_{\mu^S} R(B, x)$$

$$\ge \max_{S}\{(E_{\mu^S}(x'Bx) + \sigma^2\, trB + c - E_{\mu^S}[Q(x)])^2 + 2\sigma^4\, trB^2 + 4\sigma^2 E_{\mu^S}(x'B^2x)\}. \tag{7.1}$$

Let $D_A = E_{\mu^S}(xx')$, which is a diagonal matrix. Simple calculation shows that

$$E_{\mu^S}(x'Bx) = tr(BD_A) = tr(D_B D_A)$$

$$E_{\mu^S}(x'B^2x) = tr(B^2D_A) \ge tr(D_B)^2 D_A$$

$$trB^2 = trB'B = trD_B^2$$

Thus by (7.1) and the last 3 display,

$$\max_{x \in \Sigma} R(B, x) \geq \max_{S} E_{\mu^S} R(D_B, x) = \max_{x \in \Sigma} R(D_B, x)$$

The last equality holds because (2.7) is convex in x_j^2, and consequently attains its maximum at either $x_j^2 = 0$ or $x_j^2 = A_j^2$.

Proof of Proposition 2.2

Sufficiency follows from the monotone convergence theorem:

$$Eq_B(x) = \sum_1^{\infty} b_i (x_i^2 + \sigma^2) + c < \infty$$

Suppose that (2.3) convergence a.s. for each $x \in \Sigma$, then according to Kolmogrov 3-series theorem,

$$\sum_1^{\infty} P\{b_i y_i^2 \geq 1\} < \infty \tag{7.2}$$

$$\sum_1^{\infty} E b_i y_i^2 \, 1_{(b_i y_i^2 \leq 1)} < \infty \tag{7.3}$$

As the distribution of y_i is normal, it is easy to check that

$$E\ y_i^4 \leq 3(E\ y_i^2)^2$$

Thus by Cauchy-Schwartz inequality,

$$E\ b_i y_i^2 \, 1_{(b_i y_i^2 \geq 1)}$$

$$< \sqrt{3}\, E b_i y_i^2\, P\{b_i y_i^2 \geq 1\}$$

$$= o(\, E\ b_i y_i^2\,) \tag{7.4}$$

Hence the assertion follows from (7.3) and (7.4).

Proof of Theorem 2

Let the prior distribution of θ is $g_n(\theta, a)$ defined by (3.5), for fixed n. Then the posterior density of θ given Y is

$$\frac{\theta^{2n} \exp(-(y-\theta)^2/2) 1_{(|\theta| \le a)}}{\int_{-a}^{a} \theta^{2n} \exp(-(y-\theta)^2/2)\, d\theta} \tag{7.5}$$

Thus, the Bayes solution under quadratic loss is

$$E(\theta^2 \mid y) = \frac{\int_{-a}^{a} \theta^{2n+2} \exp(-(y-\theta)^2/2)\, d\theta}{\int_{-a}^{a} \theta^{2n} \exp(-(y-\theta)^2/2)\, d\theta}$$

$$\triangleq y^2 + \delta_a(y) \tag{7.6}$$

Now, the risk of the Bayes estimate is

$$E(y^2 + \delta_a(y) - \theta^2)^2$$

$$= 4\theta^2 + 3 + E\delta_a^2(\varepsilon + \theta) + 4\theta\, E\varepsilon\, \delta_a(\varepsilon + \theta) + 2E\varepsilon^2\, \delta_a(\varepsilon + \theta)$$

where $\varepsilon \sim N(0, 1)$. Thus, the Bayes risk is

$$4\,\frac{2n+1}{2n+3}\, a^2 + E\delta_a^2(\varepsilon + \theta) + 4\, E\varepsilon\, \theta\, \delta_a(\varepsilon + \theta) + 2E\varepsilon^2\, \delta_a(\varepsilon + \theta) + 3 \tag{7.7}$$

where E represents taking the expectation over $\varepsilon \sim N(0, 1)$, and $\theta \sim g_n(\theta, a)$.

We will prove that

$$E\delta_a^2(\varepsilon + \theta) = o(a^2) \tag{7.8}$$

Suppose (7.8) is true, then By Cauchy-Schwartz inequality,

$$E\varepsilon\,\theta\,\delta_a(\varepsilon+\theta) \le (E\theta^2\,E\varepsilon^2)^{1/2}\,(E\,\delta_a^2(\varepsilon+\theta))^{1/2} = o\,(a^2)$$

Thus, by (7.7), the Bayes risk is

$$4\,\frac{2n+1}{2n+3}\,a^2(1+o(1)) \to 4a^2(1+o(1)),\ (\,as\ n \to \infty)$$

Consequently,

$$\rho(a\,,\,1) \ge 4a^2(1+o(1))$$

On the other hand,

$$\sup_{|\theta| \le a} E_\theta(Y^2-\theta^2)^2 = 4\,a^2+3$$

We conclude that

$$\rho(a\,,\,1) = 4a^2(1+o(1))$$

Now, we have to check (7.8). By definition of (7.6) and integration by parts,

$$\delta_a(y) = \frac{\int_{-a}^{a} \theta^{2n}\,(\theta^2-y^2)\,\exp(-\,(y-\theta)^2/2)\,d\theta}{\int_{-a}^{a} \theta^{2n}\,\exp(-\,(y-\theta)^2/2)\,d\theta} \tag{7.9}$$

$$= -\,\delta_1(y)+\delta_2(y)+(2n+1)+2n\,\delta_3(y) \tag{7.10}$$

where

$$\delta_1(y) = a^{2n}(y+a)\,\exp(-\,(y-a)^2/2)\,/\,I \tag{7.11}$$

$$\delta_2(y) = a^{2n}(y-a)\,\exp(-\,(y+a)^2/2)\,/\,I \tag{7.12}$$

$$\delta_3(y) = y\int_{-a}^{a} \theta^{2n-1}\,\exp(-\,(y-\theta)^2/2)\,d\theta\,/\,I \tag{7.13}$$

$$I = \int_{-a}^{a} \theta^{2n}\,\exp(-\,(y-\theta)^2/2)\,d\theta \tag{7.14}$$

Thus, to prove (7.8), we need only to prove

$$E\,\delta_j^2(y) = o(a^2), \quad (j = 1,2,3)$$

which will be proved by following three Lemmas, where $y \overset{d}{=} \theta + \varepsilon$

Lemma 2.1: Under the above notations, $E\ \delta_1^2(y) = o(a^2)$

Proof: When $y \le a - 1$, and a is large,

$$I \ge (a-1)^{2n} \int_{a-1}^{a} \exp(-(y-\theta)^2/2)\, d\theta$$

$$\ge (a-1)^{2n} \exp(-(y-a)^2/2)$$

$$\ge \frac{1}{2} a^{2n} \exp(-(y-a)^2/2)$$

Thus, on the set $\{ y: y \le a-1 \}$,

$$\delta_1^2(y) \le 4(a+y)^2 \le 16a^2$$

Consequently, by Lebesque's dominated convergence theorem,

$$E\ \delta_1^2(y) \mathbf{1}_{(y \le a-1)} = o(a^2) \tag{7.15}$$

On the set $\{ a - 1 \le y \le a \}$, we have

$$I \ge \int_{a-1}^{a} \theta^{2n} \exp(-(y-\theta)^2/2)\, d\theta$$

$$> e^{-0.5}(a-1)^{2n}$$

$$\ge 0.5e^{-0.5}\, a^{2n} \exp(-(y-a)^2/2) \tag{7.16}$$

Thus, by (7.16)

$$E\ \delta_1^2(y)\mathbf{1}_{(a-1\le y\le a)}$$

$$\le 16e\ a^2 P\{a-1\le y\le a\} = o(a^2) \tag{7.17}$$

When $y > a$,

$$I \ge (a-1)^{2n}\int_{a-1}^{a}\exp(-(y-\theta)^2/2)\,d\theta$$

$$\ge \frac{(a-1)^{2n}}{y-a+1}[\exp(-(y-a)^2/2)-\exp(-(y-a+1)^2/2)]$$

$$\ge \frac{a^{2n}}{2(y-a+1)}(1-e^{-0.5})\exp(-(y-a)^2/2)$$

Hence,

$$E\ \delta_1^2(y)\mathbf{1}_{(y>a)} \tag{7.18}$$

$$\le 4(1-e^{-0.5})^{-2}\,E(y-a+1)^2(y+a)^2\mathbf{1}_{(y>a)}$$

Note that

$$P\{y>a\}\le P\{a-a^{0.25}<\theta\}+P\{a^{0.25}<\varepsilon\} = O(a^{-0.75}) \tag{7.19}$$

Note also that

$$P\{y>a+a^{0.25}\}\le P\{\varepsilon > a^{0.25}\}$$

$$= O(a^{-0.25})\exp(-\sqrt{a}/2) \tag{7.20}$$

By the easily checking fact that

$$E\ |y|^m = O(a^m)$$

and Cauchy-Schwartz inequality,

$$E\ \delta_1^2(y)\mathbf{1}_{(y>a)}$$

$$\le 4(1-e^{-0.5})^2\,[E\ (y-a+1)^2(y+a)^2\mathbf{1}_{(a^{0.25}+a\ge y>a)}$$

$$+ E\,(y - a + 1)^2 (y + a)^2 \mathbf{1}_{(y > a^{0.25} + a)}]$$

$$\leq O(a^{2.5})\, P\{a^{0.25} + a \geq y > a\} + 4O(a^4)\, [P(y > a + a^{0.25})]^{1/2}$$

$$= O(a^{1.75}) \tag{7.21}$$

Consequently, by (7.15), (7.17), and (7.21), the desired assertion follows.

Lemma 2.2: Under the above notations, $E\ \delta_2^2(y) = o(a^2)$

Proof: The proof is similar to that of Lemma 2.1.

Lemma 2.3: Under the above notations, $E\ \delta_3^2(y) = o(a^2)$

Proof: When $a \geq 2$,

$$\delta_3^2(y) \leq 2\Big[y^2 \Big(\int_{|\theta| \geq 1} \theta^{2n-1} \exp(-(y - \theta)^2 / 2)\, d\theta \,/\, I\Big)^2$$

$$+ y^2 \Big(\int_{|\theta| < 1} \theta^{2n-1} \exp(-(y - \theta)^2 / 2)\, d\theta \,/\, I\Big)^2\Big]$$

$$\leq 2\Big[y^2 + y^2 \Big(\frac{\int_{|\theta| < 1} \theta^{2n-1} \exp(-(y - \theta)^2 / 2)\, d\theta}{\int_{-2}^{2} \theta^{2n} \exp(-(y - \theta)^2 / 2)\, d\theta} \Big)^2\Big] \tag{7.22}$$

Let

$$g(y) = \frac{\int_{|\theta| < 1} \theta^{2n-1} \exp(-(y - \theta)^2 / 2)\, d\theta}{\int_{-2}^{2} \theta^{2n} \exp(-(y - \theta)^2 / 2)\, d\theta}$$

Obviously, by the continuity of g, when $|y| \leq 2$, $g(x) \leq c$, a finite constant. And when $y > 2$,

$$\int_{|\theta| < 1} \theta^{2n-1} \exp(-(y - \theta)^2 / 2)\, d\theta \leq 2\exp(-(y - 1)^2 / 2)$$

while

$$\int_{-2}^{2} \theta^{2n} \exp(-(y-\theta)^2/2)\, d\theta \geq \exp(-(y-1)^2/2)$$

Hence, $|g(x)| \leq 2$, when $x > 2$. Similarly, when $x < -2$, $|g(x)| \leq 2$. Thus, we concluded that

$$\delta_3^2(y) \leq 2\, max\,(c^2+1,\, 5)\, y^2$$

As $\frac{y^2}{a^2}$ is uniformly integrable, and so is $\frac{\delta_3^2(y)}{a^2}$. Consequently,

$$E\ \frac{\delta_3^2(y)}{a^2} \rightarrow 0\ (as\ a \rightarrow \infty)$$

The assertion follows.

Proof of Lemma 4.1

Without loss of generality, assume that $\sigma = 1$. Then the likelihood ratio of the density under H_0 and H_1 is

$$L_n = \prod_1^n L_{n,i} \tag{7.23}$$

where

$$L_{n,i} = \exp(-t_n^2/2)\, [\exp(t_n\, y_i) + \exp(-t_n\, y_i)]/2$$

Denote $\phi_{n,i,t_n} = \log L_{n,i}$, then

$$\phi_{n,i,t_n} = -t_n^2/2 + \log[1 + \frac{t_n^2 y_i^2}{2} + \frac{t_n^4 y_i^4}{24} + O_p(t_n^6)]$$

$$= -t_n^2/2 + \frac{t_n^2 y_i^2}{2} + \frac{t_n^4 y_i^4}{24} - \frac{t_n^4 y_i^4}{8} + O_p(t_n^6)$$

Consequently,

$$\frac{\log L_n + n t_n^4/4}{\sqrt{n/2}\, t_n^2} = \frac{\sum_1^n (y_i^2 - 1) t_n^2/2 - t_n^4 (y_i^4 - 3)/12}{\sqrt{n/2}\, t_n^2} + O_p(\sqrt{n}\, t_n^4)$$

By invoking the central limit theorem for i.i.d. case, we conclude that

$$\frac{\log L_n + nt_n^4/4}{\sqrt{n/2}\, t_n^2} \xrightarrow{L} N(0, 1)$$

under H_0. Note that

$$\frac{\log L_n - nt_n^4/4}{\sqrt{n/2}\, t_n^2} = \frac{\sum_1^n (y_i^2 - 1 - t_n^2)/2 - t_n^2(y_i^4 - Ey_i^4)/12 + O(nt_n^4)}{\sqrt{n/2}} + O_P(\sqrt{n}\, t_n^4)$$

Now, under H_1,

$$nE(|y_i^2 - 1 - t_n^2|/\sqrt{n}\,)^4 = O(n^{-1})$$

and

$$nE(|y_i^4 - Ey_i^4|/\sqrt{n}\,)^4 = O(n^{-1})$$

Hence, the Lyapounov's condition holds for the triangular arrays. By triangular array central limit theorem, under H_1,

$$\frac{\log L_n - nt_n^4/4}{\sqrt{n/2}\, t_n^2} \xrightarrow{L} N(0, 1)$$

Consequently, the sum of type I and type II error is

$$P_{H_0}[L_n > 1] + P_{H_1}[L_n \le 1] = 2\,\Phi(-\frac{\sqrt{n}\ t_n^2}{\sqrt{8}})(1 + o(1))$$

Proof of Theorem 4:

The first two results is actually proved by the argument before stating Theorem 4. The third result follows from the inequality

$$l(|t|) \ge l(1)\mathbf{1}_{(|t| \ge 1)} \tag{7.24}$$

Proof of Corollary 4.1:

Take $n_\sigma = [(\sqrt{c}\ \sigma)^{-\frac{4}{4p-1}}]$. Then,

$$(\sum_{1}^{n_\sigma} j^q) n_\sigma^{-2p}$$

$$= n_\sigma^{q+1-2p}/(q+1)\,(1+o(1))$$

$$= c^{1 - \frac{2q+1}{4p-1}} \sigma^{2 - \frac{2(2q+1)}{4p-1}} / (q+1)$$

Hence, the assertion follows from Theorem 4.

Proof of Theorem 5:

Consider the testing problem

$$H_0: y \sim N(x_1, \sigma^2) \leftrightarrow H_1: y \sim N(x_2, \sigma^2) \tag{7.25}$$

Then, the likelihood ratio is

$$L = \exp\left(\frac{y(x_2 - x_1)}{\sigma^2} - \frac{\|x_2\|^2 - \|x_1\|^2}{2\sigma^2}\right)$$

and the minimum sum of the type I and type II error is

$$\min_{0 \le \phi \le 1} (E_{x_1}\phi(y) + E_{x_2}(1 - \phi(y))) = P_{x_1}(L > 1) + P_{x_2}(L \le 1)$$

$$= 2\Phi\left(-\frac{\|x_1 - x_2\|}{2\sigma}\right)$$

where $\Phi(\cdot)$ is the distribution of the standard normal distribution. Let $r = \frac{1}{2}|T(x_1) - T(x_2)|$ be the half of the change of the functional. Then, by the same argument as those in (4.5), we have

$$\sup_{x \in \Sigma} E_x\,(\delta(y) - T(x))^2 \ge \frac{1}{2} r^2 [P_{x_1}(|\delta(y) - T(x_1)| > r] + P_{x_2}(|\delta(y) - T(x_1)| < r)]$$

$$\ge \frac{|T(x_1) - T(x_2)|^2}{4} \Phi\left(-\frac{\|x_1 - x_2\|}{2\sigma}\right)$$

Taking "sup" over $x_1, x_2 \in \Sigma$ subject to $\|x_1 - x_2\|^2 = c\sigma$, we get the desire result.

Proof of Theorem 6

Consider the testing problem (4.2) with $t_n = l_n$, and $n = N_\sigma$. By Lemma 4.1, there is no perfect test between H_0, and H_1, i.e. the minimum sum of type I and type II error is bounded away from 0. Hence, by the similar algebra as those in (4.5), the half of change of the functional from H_0 to H_1 is a lower bound of estimating $T(x)$. Now when σ is small, by assumption, the change of functional from H_0 to H_1 is at least $Q(x_\sigma)$. Hence, $Q(x_\sigma)$ is a lower rate. As $B_T(c\sigma) = o(Q(x_\sigma))$, we conclude the second result.

Proof of Theorem 7

Without loss of generality, let the supremum over t in the definition of α_T be attained, at t_0, and the supremum over Δ in the definition of Δ_T be attained, at Δ. Then by definition, it follows the minimax risk for testing between H_0 and H_1 is at least α, i.e.

$$\min_{0 \le \phi \le 1} \sup_{x_0 \in \Sigma_{T \le t_0},\, x_1 \in \Sigma_{T \ge t_0 + \Delta}} [E_{x_0}\phi + E_{x_1}(1-\phi)] \ge \alpha \tag{7.26}$$

Now for any estimator δ, and $x_1 \in \Sigma_{T \ge t_0 + \Delta}$

$$P_{x_1}\{|\delta(y) - T(x_1)| \ge \Delta/2\} \ge P_{x_1}\{|\delta(y) - T(x_0)| < \Delta/2\}$$

Hence, by (7.26)

$$\sup_{x_0 \in \Sigma_{T \le t_0},\, x_1 \in \Sigma_{T \ge t_0 + \Delta}} \max_{x_0,\, x_1} P_{x_i}\{|\delta(y) - T(x)| \ge \Delta/2\}$$

$$\ge \sup_{x_0 \in \Sigma_{T \le t_0},\, x_1 \in \Sigma_{T \ge t_0 + \Delta}} \frac{1}{2}[\,P_{x_0}\{|\delta(y) - T(x_0)| \ge \Delta/2\} + P_{x_1}\{|\delta(y) - T(x_0)| < \Delta/2\}]$$

$$\ge \alpha/2$$

Thus, the conclusion first result follows. The second result follows from the inequality (7.24)

ACKNOWLEDGEMENTS

The author gratefully acknowledges Prof. D.L. Donoho for leading me to this research area, and for his valuable suggestions and fruitful discussions. The author would like to thank Prof. L. Le Cam for his valuable remarks and discussions.

REFERENCES

1. Bickel, P.J. (1981) Minimax estimation of the mean of a normal distribution when the parameter space is restricted, *Ann. Stat.*, **9**, 1307-1309.

2. Bickel, P.J. and Ritov(1988) Estimating integrated squared density derivatives, *Technical Report 146*, Department of Statistics, University of California, Berkeley.

3. Bertero, M., Baccoaci, P., and Pike, E.R. (1982) On the recovery and resolution of exponential relaxation rates from experimental data I, *Proc. Roy. Sco. Lond.*, A **383**, 15-.

4. Casella, G. and Strawderman, W.E. (1981) Estimating a bounded normal mean, *Ann. Stat.* **9**, 870-878.

5. Donoho, D.L. and Liu, R.C. (1987a) On Minimax Estimation of Linear Functionals, *Technical Report 105*, Department of Statistics, University of California, Berkeley.

6. Donoho, D.L. and Liu, R.C. (1987b) Geometrizing Rate of Convergence, I, *Tech. Report 137a*, Department of Statistics, University of California, Berkeley.

7. Donoho, D.L. and Liu, R.C. (1987c) Geometrizing Rate of Convergence, II, *Technical Report 120*, Department of Statistics, University of California, Berkeley.

8. Donoho, D.L. and Liu, R.C. (1988) Geometrizing Rate of Convergence, III *Technical Report 138*, Department of Statistics, University of California, Berkeley.

9. Donoho, D.L. and MacGibbon, B. (1987) Minimax risk for hyperrectangles, *Technical Report 123*, Department of Statistics, University of California, Berkeley.

10. Efroimovich, S.Y. and Pinsker, M.S. (1982) Estimation of square-integrable probability density of a random variable, *Problemy Peredachi Informatsii*, **18**, **3**, 19-38 (in Russian); *Problems of Information Transmission* (1983), 175-189.

11. Farrell, R.H. (1972) On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, *Ann. Math. Stat.*, **43, #1**, 170-180.

12. Ibragimov, I.A. and Khas'minskii (1981), *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York-Berlin-Heidelberg.

13. Ibragimov, I.A., Nemirovskii, A.S. and Khas'minskii, R.Z. (1986) Some problems on nonparametric estimation in Gaussian white noise, *Theory Prob. Appl.*, **31, 3**, 391-406.

14. Khas'minskii, R.Z. (1978) A lower bound on the risks of nonparametric estimates densities in the uniform metric, *Theory Prob. Appl.*, **23**, 794-798.

15. Le Cam, L. (1973) Convergence of estimates under dimensionality restrictions, *Ann. Stat.* **1**, 38-53.

16 Le Cam, L. (1985) *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York-Berlin-Heidelberg.

17. Pinsker, M.S. (1980) Optimal filtering of square integrable signals in Gaussian white noise, *Problems of Information Transmission*, **16, 2**, 52-68.

18. Sacks, J. and Ylvisaker, D. (1981) Variance estimation for approximately linear models, *Math. Operationsforch. Statist., Ser. Statistics*, **12**, 147-162.

19. Stein, C. (1956) Efficient nonparametric estimation and testing, *Proc. 3rd Berkeley Symp.* **1**, 187-195.

20. Stone, C. (1980) Optimal rates of convergence for nonparametric estimators, *Ann. Stat.* **8**, 1348-1360.

21. Stone, C. (1982) Optimal global rates of convergence for nonparametric regression, *Ann. Stat.*, **10, 4**, 1040-1053.

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review. Statistical Science, 1988, Vol.3 No. 2 239-271.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN; D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984, 579-611.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data. Scandinavian J. Statist., 1988, 15, 1-23.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data. Biometrika, 1987, 74, 799-808.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (Nov 1986, revised Aug 1988). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model. J. Amer. Statist. Assoc., 1988, 83, 744-749.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. CANCELLED

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. <u>Statistics a Guide to the Unknown</u>, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (Jan. 1988). Two reports on trend analysis: a) An elementary trend analysis of Rio negro levels at Manaus, 1903-1985. b) Consistent detection of a monotonic trend superposed on a stationary time series.

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (August 1988). Maximum likelihood an introduction.

169. BREIMAN, L. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed case and little bootstrap.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.

178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.

179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.

180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).

181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989). A model for step-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.

182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.

183. BERAN, R. (November 1988). Calibrating prediction regions.

184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.

185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.

186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes

187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.

188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.

189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.

190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.

191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.

192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.

193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.

194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.

195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.

196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.

197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.

198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.

199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2° congreso lantinoamericano de probabilidad y estadistica matematica*, Caracas, 1985.

200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.

201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.

202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.

203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.

204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p-Balls.

205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.