Minimax Risk for Hyperrectangles

David L. Donoho

Department of Statistics
University of California, Berkeley

Richard C. Liu

Department of Mathematics
University of California, Los Angeles

Brenda MacGibbon

Département de mathématiques et d'informatique
Université du Québec à Montréal

Technical Report No. 123

November 1987

First Revision July, 1988
Second Revision October, 1988
Third Revision December, 1988

The work of the first-named author was supported by NSF grant DMS-84-51753, by grants from SUN Microsystems and from Schlumberger. The work of the second-named author was supported by NSF grant DMS-86-02018. The work of the third-named author was supported by grants from NSERC of Canada and FCAR of Québec.

Department of Statistics

University of California

Berkeley, California

Minimax Risk for Hyperrectangles

David L. Donoho

Department of Statistics
University of California, Berkeley

Richard C. Liu

Department of Mathematics
University of California, Los Angeles

Brenda MacGibbon

Department of Decision Sciences and MIS
Concordia University

Consider estimating the mean of a standard Gaussian shift when that mean is known to lie in a quadratically convex set in l_2. Such sets include ellipsoids, hyperrectangles, and l_p-bodies with $p > 2$. The minimax risk among linear estimates is within 25% of the minimax risk among all estimates. The minimax risk among truncated series estimates is within a factor 4.44 of the minimax risk. This implies that the difficulty of estimation -- a statistical quantity -- is measured fairly precisely by the n-widths -- a geometric quantity.

If the set is not quadratically convex, as in the case of l_p-bodies with $p<2$, things change appreciably. Minimax linear estimators may be outperformed arbitrarily by nonlinear estimates. The (ordinary, Kolmogorov) n-widths still determine the difficulty of *linear* estimation, but the difficulty of *nonlinear* estimation is tied to the (inner, Bernstein) n-widths, which can be far smaller.

Essential use is made of a new heuristic: that the difficulty of the hardest rectangular subproblem is equal to the difficulty of the full problem.

Key Words and Phrases: Estimating a bounded normal mean, Estimating a function observed with white noise, hardest rectangular subproblems, Ibragimov-Hasminskii constant, quadratically convex sets, Bernstein and Kolmogorov n-widths.

Acknowledgements: The authors would like to thank Lucien Birgé for a very revealing seminar on the work of M.S. Pinsker for ellipsoids. Iain Johnstone stimulated our interest in such topics by a number of helpful conversations. Jianqing Fan made valuable comments on an earlier draft. The work of the first-named author was supported by NSF grant DMS-84-51753, by grants from SUN Microsystems and from Schlumberger. The work of the second-named author was supported by NSF grant DMS-86-02018. The work of the third-named author was supported by grants from NSERC of Canada and FCAR of Québec.

1. Introduction

Pinsker (1980) considered the problem of estimating the mean of a certain Gaussian process when the mean is known to lie in an infinite-dimensional "ellipsoid". He found an exact value for the minimax risk of linear estimates and an asymptotic value for the minimax risk among nonlinear estimates. These evaluations allow one to obtain precise constants on the asymptotic minimax risk for certain "real" function estimation problems: density estimation — Efroimovich and Pinsker (1981, 1982) — and regression estimation — Nussbaum (1985). This is an improvement over usual treatments of nonparametric estimation problems, where only rates, and not constants, are available. A remarkable feature of the Pinsker solution is that it shows the minimax linear estimator to be asymptotically minimax among all estimates. Thus, in minimax theory at least, there is little to be gained by nonlinear procedures.

Because Pinsker's result is specifically for the case where the unknown mean lies in an ellipsoid, the question arises whether similar results hold when the unknown mean lies in a set with a different "shape". In this paper they show that if the mean is known to lie in a quadratically convex set, the minimax linear risk is within a factor 1.25 of the minimax risk nonasymptotically. Thus, for ellipsoids, hyperrectangles, and l_p bodies the minimax linear risk is not very different from the minimax risk. Almost certainly, the constant 1.25 can be replaced by 1.247.

More generally we might ask: in the problem of estimating a mean θ known to lie in a convex compact subset Θ of l_2, does there exist a constant, *independent of* Θ, bounding the ratio of minimax risk to minimax linear risk. If such a constant exists independently of Θ (provided Θ is convex), many of the usual lower bound arguments in rates of convergence theory might be dispensed with altogether. One would simply determine the behavior of the minimax linear estimator; then no nonlinear estimator could improve on this except by a constant factor. On the other hand, if there exists a class of convex sets Θ for which the minimax linear and minimax risks behave essentially differently, this seems also intrinsically interesting. We show here that for a large class of cases, the constant 1.25 applies.

Our approach also gives results on the minimax risk of truncated series estimates. Let the set Θ have the sequence of Kolmogorov n-widths (d_n). Then by using an optimal truncated series estimate, the worst-case risk $\inf_n d_n^2 + n\,\sigma^2$ is attainable. We show that if Θ is quadratically convex, this upper bound based on n-widths is within a factor 4.44 of the minimax risk, and a factor 4 of the minimax linear risk. Moreover, even if Θ is not quadratically convex, the minimax truncation risk and the minimax linear risk are within a factor 4. Thus, from a minimax point of view, general linear estimates do not improve dramatically on truncation schemes.

Our results have other implications. Consider the problem of estimating the linear functional $L(\theta)$ of the unknown mean θ, when θ is known to lie in a convex set Θ. Results of Ibragimov and Hasminskii (1984), combined with our Theorem 1, show that when Θ is symmetric, the ratio of minimax linear to minimax risks is less than 1.25. Results of Donoho and Liu (1988b), combined with our Theorem 1, show that for *any* convex set Θ, the ratio of minimax *inhomogeneous* linear risk to minimax risk is bounded by 1.25. Thus, for estimating a single linear functional, an absolute bound on improvement by nonlinearity holds quite generally, independent of the shape of the convex set in which the mean is known to lie.

These results provide a partial answer to the question raised by Sacks and Strawderman (1981) -- namely, is it possible to improve significantly on minimax linear estimators by nonlinear schemes. They also provide a concrete working out of the Birgé-Le Cam program to express minimax risks in terms of geometric quantities; we show that for quadratically convex sets, the geometric quantity $\inf_n d_n^2 + n\,\sigma^2$ is within a factor 4.44 of the minimax risk.

However, we also get negative results which are perhaps more interesting. We show that for l_p-bodies with $p < 2$, the minimax linear risk need not tend to zero at the same rate as the minimax risk. This shows that the l_p bodies with $p<2$ represent in a certain sense an answer to the question posed by Sacks and Stradwerman (1982).

An interesting feature of our approach is the use of geometric ideas, including that of hardest rectangular subproblem and quadratic hull, to explain these phenomena.

Section 11 shows that analogs of these results hold for other loss functions; it discusses the case of l_1-loss.

2. The Problem

The basic model is as follows. We are given

$$y_i = \theta_i + \varepsilon_i \qquad i = 0,1,2,... \tag{2.1}$$

where ε_i are iid $N(0,\sigma^2)$ and θ_i are unknown, but it is known that

$$|\theta_i| \le \tau_i, \qquad i=0,1,2,.... \tag{2.2}$$

Thus $\theta = (\theta_i)$ lies in the hyperrectangle $\Theta = \Theta(\tau) = \{\theta: |\theta_i| \le \tau_i\}$. We wish to estimate θ with small squared error loss, i.e. to make the squared l_2-norm $||\hat{\theta}-\theta||^2 = \sum(\hat{\theta}_i - \theta_i)^2$ small. We will use the minimax principle to evaluate estimates; an estimator $\hat{\theta}^*$ is minimax if

$$\sup_{\theta\in\Theta} E||\hat{\theta}^* - \theta||^2 = \inf_{\hat{\theta}} \sup_{\theta\in\Theta} ||\hat{\theta}-\theta||^2. \tag{2.3}$$

We also speak of restricted minimax estimates. Thus, if $\hat{\theta}^*$ is linear and satisfies (2.3) with the infimum over $\hat{\theta}$ referring only to linear procedures, we say that $\hat{\theta}^*$ is linear minimax.

Let us indicate briefly how this estimation problem is related to estimating an unknown function. See also Pinsker (1980), Ibragimov and Hasminskii (1984), Nussbaum (1985). Suppose we are interested in estimating the function $f(t)$, $t \in [a,b]$, but f is observed in a white noise:

$$y(t) = \int_a^t f(t)\,dt + \sigma\int_a^t dW(t) \quad t \in [a,b] \tag{2.4}$$

where $W(t)$ is a Wiener process. We wish to find an estimate $\hat{f}$ of f which makes $\int(\hat{f}-f)^2$ small, and we have *a priori* information that f is smooth.

If the smoothness information is of a particular kind, the problem reduces to the hyperrectangle model (2.1)-(2.2). Let $dm = dt/(b-a)$ and suppose we have a set $\{\phi_i\}$ of functions orthonormal for $L_2(m,[a,b])$. Let $\theta_i = \int f\phi_i\,dm$ be the i-th Fourier-Bessel coefficient of f with respect to this set, and suppose we know *a priori* that the Fourier-Bessel coefficients of f decay rapidly:

$$|\theta_i| \le \tau_i, \quad \tau_i \to 0 \text{ as } i \to \infty. \tag{2.5}$$

Let us see how these assumptions reduce to (2.1)-(2.2). To start with, (2.5) is precisely of the form (2.2); on the other hand, if we take the Fourier-Bessel coefficients of (2.4) we get

$$y_i = \frac{1}{b-a}\int \phi_i\,dy(t) = \theta_i + \varepsilon_i$$

where ε_i are i.i.d. $N(0,\sigma^2)$. Thus (2.4) and (2.1) are equivalent, if f lies in the span of $\{\phi_i\}$. Finally, a good estimate of f leads to a good estimate of θ, and vice versa. If $\hat{f}$ is an estimate of f, it induces an estimate $\hat{\theta}$ of θ via $\hat{\theta}_i = \int \hat{f} \phi_i dm$ and for this estimate we have $\int (\hat{f} - f)^2 dm = \sum (\hat{\theta}_i - \theta_i)^2$ Similarly, given an estimate $\hat{\theta}$ of θ, we obtain a 'series' estimate $\hat{f}$ of f via $\hat{f}(t) = \sum \hat{\theta}_i \phi_i(t)$ and again (2.6) holds.

A concrete example of the isomorphism between (2.1)-(2.2) and (2.4)-(2.5) is provided by Fourier series. Let $[a,b] = [-\pi,\pi]$, and let the orthonormal set $\{\phi_i\}$ be the usual sinusoids: $\phi_0 = 1$, and for $i > 0$, $\phi_{2i-1}(t) = \sqrt{2}\sin(it)$, and $\phi_{2i}(t) = \sqrt{2}\cos(it)$. Then the coefficients θ_i are just the Fourier Coefficients of f, and (2.6) is Parseval's relation. In this setup, the prior "smoothness" condition (2.5) does really correspond to smoothness. For example, suppose that f and $(q-1)$ derivatives of f are of bounded variation, and that f and these derivatives satisfy periodic boundary conditions at π and $-\pi$. Then $|\theta_{2i}|, |\theta_{2i-1}| \le ci^{-q}$ for an appropriate c. Thus the condition (2.5) with $\tau_{2i} = \tau_{2i-1} = ci^{-q}$ is a weakening of the condition that f have $(q-1)$ derivatives of bounded variation.

The white-noise model (2.4) is closely related to problems of density estimation and spectral density estimation. Indeed, it can appear as the limiting Gaussian shift experiment in such problems. Thus it should be no surprise that results on hyperrectangles allow one to attack certain asymptotic minimax problems. Bentkus and his school have used this connection to get expressions for the asymptotic minimax risk among linear estimates in density estimation problems with smoothness constraints (2.5) (Bentkus and Kazbaras, 1981), for the asymptotic minimax risk among kernel estimates of a spectral density also using (2.5) (Bentkus and Sushinskas, 1982), (Bentkus, 1985a,b), and for the minimax risk among kernel estimates in estimating a periodic function from sampled data (Jakimauskas, 1984). Similarly, if the hyperrectangle constraint (2.5) is replaced by a quadratic constraint, Pinsker's results on Ellipsoids become relevant, and may be used to study asymptotic minimaxity with L_2 smoothness constraints in density estimation (Efroimovich and Pinkser, 1982), in regression estimation (Nussbaum, 1985), and in spectral density estimation (Efroimovich and Pinsker, 1981).

In this paper, we consider only the problem for observations (2.1); we take it for granted that the results have a variety of applications, such as those just mentioned. We also take it for granted that behavior as $\sigma \to 0$ is important, which may not be seem like a natural question in the model (2.1), but which is natural when the connection with e.g. density estimation is considered.

3. The 1-dimensional Problem

Consider estimating a *single* bounded normal mean, i.e. estimating $\theta \in \mathbf{R}$ from the single observation, $y \sim N(\theta, \sigma^2)$ with the prior information that $|\theta| \le \tau$. This problem has been studied by Casella and Strawderman (1981), Levit (1980), Bickel (1981), and Ibragimov and Hasminskii (1984). It is known that the minimax estimator for this problem is Bayes with respect to a prior concentrated at a finite number of points in $[-\tau, \tau]$. Let $\delta^N_{\tau,\sigma}(y)$ denote this minimax estimator. $\delta^N_{\tau,\sigma}$ is nonlinear in y (i.e. it derives from a nonGaussian prior). Let $\rho_N(\tau,\sigma)$ denote the minimax risk. More information will be given below.

Consider estimating θ in this setup by a (possibly biased) linear estimator. The minimax linear estimator can be worked out using calculus; it is

$$\delta^L_{\tau,\sigma}(y) = \frac{\tau^2}{\tau^2+\sigma^2}\, y$$

and the minimax linear risk is

$$\rho_L(\tau,\sigma) = \inf_{\delta \text{ linear}} \sup_{|\theta| \le \tau} E(\delta(y)-\theta)^2 = \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}. \tag{3.1}$$

As it turns out, the minimax linear risk in this problem is not very different from the nonlinear minimax risk. Consider the ratio of the two: $\rho_L(\tau,\sigma)/\rho_N(\tau,\sigma)$. By the invariance $\rho(\tau,\sigma) = \sigma^2\rho(\tau/\sigma,1)$ which holds for both ρ_L and ρ_N, this ratio depends on τ and σ only through the "signal-to-noise" ratio $v = \tau/\sigma$. Let $\mu(v)$ denote the ratio of the two risks for a given value of v. Ibragimov and Hasminskii (1984) pointed out 3 basic facts about $\mu(v)$: (1) it is continuous on $(0,\infty)$; (2) it is near 1 for v large:

$$\lim_{\tau/\sigma \to \infty} \frac{\rho_L(\tau,\sigma)}{\rho_N(\tau,\sigma)} = 1 \tag{3.2}$$

and (3) also near 1 for v small:

$$\lim_{\tau/\sigma \to 0} \frac{\rho_L(\tau,\sigma)}{\rho_N(\tau,\sigma)} = 1. \tag{3.3}$$

Let μ^* denote the maximum value of $\mu(v)$, i.e. the worst-case ratio of ρ_L to ρ_N

$$\mu^* = \sup_{\tau,\sigma} \frac{\rho_L(\tau,\sigma)}{\rho_N(\tau,\sigma)}. \tag{3.4}$$

Ibragimov and Hasminskii (1984) argued that (3.2), (3.3), and continuity of $\mu(\nu)$ imply that $\mu^* < \infty$.

We can interpret (3.2) and (3.3) as follows. In the extremes where the prior information $|\theta| \le \tau$ is weak compared to the noise level (i.e. τ/σ large) and also where it is strong compared to the noise level (i.e. τ/σ small) the minimax linear estimate is nearly minimax.

Actually, much more is true. $\mu(\nu)$ never gets very far from 1 even at moderate ν. Lucien Birgé, in a talk on the work of M.S. Pinsker at the Mathematical Sciences Research Institute in Berkeley in April, 1983 mentioned that he had convinced himself that $\mu^* < 1.7$. In fact, as we shall explain in a moment, *the Ibragimov-Hasminskii constant* μ^* *is less than 1.25.*

In studying the ratio $\mu(\nu) = \rho_L(\nu,1) / \rho_N(\nu,1)$, we have information on ρ_L from (3.1). However, information on $\rho_N(\nu,1)$ is harder to come by. For small ν we can use the fact that, for $\nu < 1.05$,

$$\rho_N(\nu,1) = \nu^2 e^{-\nu^2/2} \int_{-\infty}^{+\infty} \frac{\phi(t)}{\cosh(\nu t)}\, dt. \tag{3.5}$$

where ϕ denotes the N(0,1) density. This is proved in the appendix. For large ν we can use the inequality

$$\rho_N(\nu,1) \ge (1 - \frac{\sinh \nu}{\nu \cosh \nu}) \tag{3.6}$$

which follows from Donoho and Liu (1988a, section 6.1). Actually, (3.5) implies that $\mu(\nu) \le 1.25$ for $\nu \le .5$, and (3.6) implies that $\mu(\nu) \le 1.25$ for $\nu \ge 3.1$. (We remark that the important relations (3.2) and (3.3) follow immediately from (3.1), (3.5), and (3.6)).

To get information about $\mu(\nu)$ for moderate ν, one has to resort to the implicit characterisation of ρ_N as the maximum of Bayes risks:

$$\rho_N(\nu,1) = \sup_{\pi \in \Pi_\nu} \rho(\pi) \tag{3.7}$$

where $\rho(\pi)$ denotes the Bayes risk

$$\rho(\pi) = \inf_\delta E_\theta E_{Y|\theta} (\delta(Y) - \theta)^2, \qquad \theta \sim \pi .$$

By L.D. Brown's identity $\rho(\pi) = 1 - I(\Phi * \pi)$, where $I(F)$ denotes the Fisher information $\int (f')^2/f$, and $\Phi * \pi$ denotes the convolution of π with the standard Normal distribution function Φ (see, for example, Bickel (1981)). Thus, putting $I^*(\nu) = \inf\{I(\Phi*\pi): \pi \in \Pi_\nu\}$, we have $\rho_N(\nu,1) = 1 - I^*(\nu)$. As I is convex, evaluation of $I^*(\nu)$ presents a problem of minimizing a convex functional subject to the convex constraint $\pi \in \Pi_\nu$. The appendix explains how a numerical approach was used to get numbers $\hat{I}(\nu)$ approximating upper bounds to $I^*(\nu)$. Assuming no programming error was committed, and that machine arithmetic is performed with advertised accuracy, the numbers $\hat{\rho}_N(\nu) = 1 - \hat{I}(\nu)$ may be shown to *rigorously* obey

$$\rho_N(\nu,1) \geq \hat{\rho}_N(\nu) - .0001 \qquad \nu \in [.42, 4.2] \tag{3.8}$$

Thus, they are "lower bounds to four digits accuracy".

Table 1 presents a small selection of the numerical results we have obtained; it shows the numbers $\hat{\rho}_N$, together with the corresponding ρ_L and the ratio $\hat{\mu} = \rho_l/(\hat{\rho}_N - .0005) \geq \mu$.

Table 1 Risks in the 1-dimensional problem

ν	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
ρ_L	0.038	0.138	0.265	0.390	0.500	0.590	0.662	0.719	0.764	0.800
$\rho_N \geq$	0.037	0.137	0.261	0.373	0.449	0.491	0.534	0.576	0.614	0.644
ratio $\frac{\rho_L}{\rho_N} \leq$	1.032	1.009	1.016	1.046	1.114	1.201	1.239	1.248	1.244	1.242
ν	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
ρ_L	0.829	0.852	0.871	0.887	0.900	0.911	0.920	0.928	0.935	0.941
$\rho_N \geq$	0.669	0.692	0.714	0.733	0.750	0.765	0.779	0.792	0.804	0.814
ratio $\frac{\rho_L}{\rho_N} \leq$	1.239	1.231	1.220	1.209	1.200	1.191	1.181	1.172	1.163	1.156

Professor Hasminskii has informed us that a set of calculations he performed in Moscow convinced him that μ^* is about 5/4. Professor Brown has informed us that a recent thesis at the Hebrew University by I. Feldman makes it practically certain that the precise value of μ^* is between 1.246 and 1.247. Taking into account all the limitations of numerical approaches the best we can say with certainty is

Theorem 1. *Suppose (3.8) holds. Then* $\mu^* \leq 1.25$.

The proof is given in the appendix, where considerably more information about our procedure and the claim (3.8) are available. An unconditional result is possible. Let $\rho_T(\tau,\sigma) = \min(\tau^2,\sigma^2)$. This is the minimax risk of the truncation rule which estimates θ by zero if $\tau < \sigma$ and by y if $\tau \geq \sigma$ (see section 6). We have

Theorem 2.

$$\max_{\nu} \frac{\rho_T(\nu,1)}{\rho_N(\nu,1)} = \frac{1}{\rho_N(1,1)} \approx 2.22 \tag{3.9}$$

The proof is in the appendix. As $\rho_T \geq \rho_L$ it follows that $\mu^* \leq 2.22$.

4. Hyperrectangles

Return now to the hyperrectangle problem. If we let θ_i be a random variable distributed according to the prior supporting the minimax rule $\delta^N_{\tau_i,\sigma}$ and independent of the other θ_i's, then the Bayes risk for estimation of θ is easy to calculate; due to the independence of y_i's it is just the coordinatewise sum $\sum_i \rho_N(\tau_i,\sigma)$ As the coordinatewise estimate $\hat{\theta}^N = (\delta^N_{\tau_i,\sigma}(y_i))$ is Bayes for the indicated prior, and as the indicated prior is least favorable for this estimator, this Bayes risk is the minimax risk and this estimator is minimax.

Proposition 3. *The minimax risk for Problem (2.1)-(2.2) is*

$$R_N^*(\sigma) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} E\, \|\hat{\theta} - \theta\|^2 = \sum \rho_N(\tau_i,\sigma). \tag{4.1}$$

By similar reasoning, the linear estimator $\hat{\theta}^L = (\delta^L_{\tau_i,\sigma}(y_i))$ is the minimax linear estimator, and

Proposition 4. *The minimax linear risk for Problem (2.1)-(2.2) is*

$$R_L^*(\sigma) = \sum \rho_L(\tau_i,\sigma).$$

The minimax linear risk has been studied intensively in several papers by R. Bentkus and members of his school; Proposition 4 appears implicitly in several of their papers. The minimax risk has apparently not been intensively studied, apparently because there is no tractable closed form expression for $\rho_N(\tau_i,\sigma)$. However, in view of Theorem 1, we know that each $\rho_L(\tau_i,\sigma) \leq \mu^* \rho_N(\tau_i,\sigma)$, giving

Corollary.

$$R_L^*(\sigma) \leq \mu^* R_N^*(\sigma) \leq 1.25 R_N^*(\sigma). \tag{4.4}$$

Thus the best nonlinear estimate of θ cannot improve on the best linear one by very much.

An asymptotic comparison, as $\sigma \to 0$, of the two different risks can be made as follows. Recalling the definition of $\mu(\nu)$,

$$R_N^*(\sigma) = \sum (\mu(\frac{\tau_i}{\sigma}))^{-1} \rho_L(\tau_i, \sigma)$$

and so

$$\frac{R_N^*(\sigma)}{R_L^*(\sigma)} = \frac{\sum \mu(\frac{\tau_i}{\sigma})^{-1} \rho_L(\tau_i, \sigma)}{\sum \rho_L(\tau_i, \sigma)}.$$

As $\rho_L \geq 0$ one may view this right hand side as defining an "average" of $\mu(\frac{\tau_i}{\sigma})^{-1}$ with respect to a "probability distribution" $\rho_L(\tau_i, \sigma) / \sum \rho_L(\tau_i, \sigma)$ on i. As many of the terms τ_i occur at τ_i / σ large, and an infinite number occur at τ_i / σ small, (3.2)-(3.3) might suggest that with high "probability" $\mu(\frac{\tau_i}{\sigma})$ is close to one. Consequently, the actual ratio of minimax risks will be closer than the bound 1.25.

Theorem 5. *Let $q > 1/2$. Suppose that $\tau_i = ci^{-q}$. Then*

$$\lim_{\sigma \to 0} \frac{R_N^*(\sigma)}{R_L^*(\sigma)} = \zeta_L(q) \equiv \int_0^\infty \mu(\nu)^{-1} g_q(\nu) d\nu$$

where the probability density g_q is supported on $[0,\infty]$ *and is defined by*

$$g_q(\nu) = \frac{\frac{\nu^2}{1+\nu^2} \nu^{-(1+1/q)}}{\int_0^\infty \frac{\nu^2}{1+\nu^2} \nu^{-(1+1/q)} d\nu}. \tag{4.5}$$

The proof is given in the Appendix. A table of lower bounds on $\zeta_L(q)$ is given below. The bounds were arrived at using techniques described in Gatsonis, MacGibbon, and Strawderman (1987), and in section 3 above.

Table 2.

Bounds on $\zeta_L(q)$ and $\zeta_T(q)$

q	$\zeta_L(q) \geq$	$\zeta_T(q) \leq$
.75		1.23
1.0		1.27
1.2	.897	1.27
1.4	.903	1.25
1.6	.904	1.24
1.8	.906	1.22
2.0	.912	1.21
2.2	.915	1.19
2.4	.918	1.18
2.6	.921	1.17
2.8	.926	1.16
3.0	.927	1.12
4.0	.940	1.12
5.0	.949	1.10
10.0	.971	1.05
25.0	.98	1.02
50.0	.99	1.01

Corollary. *For $q \in (1/2,\infty)$, $\zeta_L(q) < 1$. Consequently, $\hat{\theta}^L$ is not asymptotically minimax as $\sigma \to 0$. $\zeta_L(q) \to 1$ as $q \to 1/2$ or ∞. Consequently, $\hat{\theta}^L$ is* **nearly** *asymptotically minimax in the cases where the problem is very difficult (q near 1/2) or very easy (q near ∞).*

The proof of the first two sentences consists in the observation that $\mu(v) > 1$ for all $v \in (0,\infty)$, as the minimax estimator is not linear. (Indeed, a minimax estimator is Bayes for some prior supported on $[-v, v]$; it is therefore bounded in absolute value by v, whereas nontrivial linear estimators are not bounded). Thus, the expectation of $\mu(v)^{-1}$ with respect to g_q is strictly less than 1. Equivalently, $\zeta_L(q) < 1$, which prohibits minimaxity.

For sentences three and four, note that by (3.2)-(3.3), $\mu(v)$ is near 1 for v near 0 and ∞. Now the limit of g_q, as $q \to \infty$, is a measure concentrated at $+\infty$. Indeed, let $x > 1$ and $q > 1$. Then

$$\int_0^x \frac{v^2}{1+v^2} v^{-(1+1/q)} dv \;\leq\; \int_0^1 v^{1-1/q} dv + \int_1^x v^{-1} dv \;\leq\; \frac{1}{2} + \log(x).$$

Also if $q > 1$,

$$\int_0^\infty \frac{v^2}{1+v^2} v^{-(1+\frac{1}{q})} dv \;\geq\; \frac{1}{2} \int_1^\infty v^{-(1+\frac{1}{q})} dv \;=\; \frac{q}{2}.$$

Hence

$$\int_0^x g_q(v)dv \;\le\; \frac{1+log(x)/2}{q}$$

which tends to zero as $q\to\infty$. Then as $q\to\infty$ we must have

$$\int \mu(v)^{-1} g_q(v)dv \;\to\; \lim_{v\to\infty} \mu(v)^{-1} \;=\; 1.$$

On the other hand, the limit of g_q, as $q\to\frac{1}{2}$, is a measure concentrated at 0. To see this, note that if $x<1$ and $1/2<q<1$,

$$\int_x^\infty \frac{v^2}{1+v^2} v^{-(1+\frac{1}{q})} dv \;\le\; \int_x^1 v^{1-\frac{1}{q}} + \int_1^\infty v^{-2}dv \;\le\; \int_x^1 v^{-1} + 1 \;=\; 1-log(x),$$

while

$$\int_0^\infty \frac{v^2}{1+v^2} v^{-(1+\frac{1}{q})} dv \;\ge\; \frac{1}{2}\int_0^1 v^{1-\frac{1}{q}} dv \;=\; \frac{1}{2}\,\frac{1}{2-\frac{1}{q}}.$$

Consequently,

$$\int_x^\infty g_q(v)dv \;\le\; 2(2-\frac{1}{q})(1-log(x)),$$

which tends to zero as $q\to\frac{1}{2}$. It follows that as $q\to\frac{1}{2}$,

$$\int \mu(v)^{-1} g_q(v)dv \;\to\; \lim_{v\to 0} \mu(v)^{-1} = 1.$$

As $\zeta_L(q)$ is the expectation of $\mu(v)^{-1}$, this completes the proof. □

Thus, $\hat{\theta}^L$ is *not* asymptotically minimax for typical infinite dimensional hyperrectangles, although it is not far from minimax, as Table 2 shows. If Θ is a finite-dimensional hyperrectangle, of course, then $\hat{\theta}^L$ *is* asymptotically minimax as $\sigma\to 0$. This is just a consequence of

$$\frac{\sum_1^d \rho_L(\tau_i,\sigma)}{\sum_1^d \rho_N(\tau_i,\sigma)} \;\le\; \sup_{1\le i\le d} \frac{\rho_L(\tau_i,\sigma)}{\rho_N(\tau_i,\sigma)} \;=\; \sup_{1\le i\le d} \mu(\tau_i/\sigma) \;\to\; 1$$

as $\sigma\to 0$.

6. Quadratically Convex Sets

Suppose now that we observe data according to (2.1), but instead of (2.2) we know that $\theta\in\Theta$, where Θ is convex, but not a hyperrectangle. If Θ contains a hyperrectangle $\Theta(\tau)$, $\tau=(\tau_i)_{i=0}^{\infty}$, the problem of estimating θ under (2.1)-(2.2) is called a *rectangular subproblem.* The minimax linear risk of the full problem is as large as that of any subproblem, so

$$R_L^*(\sigma) \geq \sup\{ R_L^*(\sigma;\Theta(\tau)) \colon \Theta(\tau) \subset \Theta \} \qquad (5.1)$$

When equality holds here, we have

$$\begin{aligned} R_N^*(\sigma) &\geq \sup\{ R_N^*(\sigma;\Theta(\tau)) \colon \Theta(\tau) \subset \Theta \} \\ &\geq \sup\{ \frac{1}{\mu^*} R_L^*(\sigma;\Theta(\tau)) \colon \Theta(\tau) \subset \Theta \} \qquad \text{(by (4.4))} \\ &= \frac{1}{\mu^*} R_L^*(\sigma). \end{aligned} \qquad (5.2)$$

This proves

Lemma 6. *If the difficulty, for linear estimates, of the hardest rectangular subproblem, is* **equal** *to the difficulty, for linear estimates, of the full problem, then*

$$R_L^*(\sigma) \;\leq\; \mu^* \, R_N^*(\sigma) \;\leq\; 1.25 R_N^*(\sigma) \qquad (5.3)$$

We now show that equality often holds in (5.1). First, some definitions.

We say that Θ is *orthosymmetric* if, whenever $\theta = (\theta_i)_{i=0}^{\infty}$ belongs to Θ, $(\pm\theta_i)_{i=0}^{\infty}$ also belongs to Θ for all choices of signs $\pm$. Examples of orthosymmetric sets include: *Ellipsoids,* sets of the form $\{\theta \colon \sum a_i \theta_i^2 \leq 1\}$ where all $a_i \geq 0$; more generally, *weighted l_p-bodies*, of the form, $\{\theta \colon \sum a_i |\theta_i|^p \leq 1\}$, sets $\{\theta \colon \sum a_i \psi(|\theta_i|) \leq 1\}$, and of course hyperrectangles. We say Θ is *quadratically convex* if $\{(\theta_i^2)_{i=0}^{\infty}, \theta \in \Theta\}$ is convex. Ellipsoids and weighted l_p-bodies with $p \geq 2$ are quadratically convex, as are hyperrectangles, and sets $\{\theta \colon \sum a_i \psi(\theta_i^2) \leq 1\}$ where ψ is convex. (To make these examples more concrete, recall from the function smoothing interpretation in section 2 that constraints on the q-th derivative of a function can be expressed by weighted l_p bodies with weights $a_0 = 0$, $a_{2i-1} = a_{2i} = c i^{pq}$, $i \geq 1$.)

Theorem 7. *If Θ is compact, quadratically convex, and orthosymmetric, the difficulty, for linear estimates, of the hardest rectangular subproblem is* **equal** *to the difficulty, for linear estimates, of the full problem:*

$$R_L^*(\sigma) \;=\; \sup\{ R_L^*(\sigma;\Theta(\tau)) \colon \Theta(\tau) \subset \Theta \} \qquad (5.4)$$

Thus, the factor 1.25 which we have established applies not only to hyperrectangles, but also to compact ellipsoids and compact l_p bodies, $p > 2$. Note that the set $\{\theta \colon \sum a_i |\theta_i|^p \leq 1 \;\text{ and }\; \|\theta\|^2 \leq C\}$ is orthosymmetric and quadratically convex, and compact if all but a finite number of the a_i are nonzero and $a_i \to \infty$.

The result (5.4) is also true for some noncompact cases -- $\Theta = R^n$ being an obvious example. Also, if $\Theta = \Theta_0 \times \Theta_1$, and (5.4) is true for each factor Θ_i, then (5.4) is true for Θ. These two remarks may be combined. If a finite number of the a_i are zero, and if $a_i \to \infty$, then $\Theta = \{\theta : \sum a_i |\theta_i|^p \le 1\}$ is the product $\Theta = R^n \times \Theta'$, where Θ' satisfies the hypotheses of the theorem. Thus (5.4) is true for *all* ellipsoids and l_p-bodies with $p > 2$, not just compact ones. Probably (5.4) is true even if Θ is just closed.

Proof. The idea is as follows. First, we show there is a hardest rectangular subproblem $\Theta(\tau^*)$. Let $\hat{\theta}^*$ be the minimax linear estimator for that subproblem; we have automatically that for any linear estimator $\hat{\theta}$

$$\sup_{\Theta(\tau^*)} R(\hat{\theta}, \theta) \ \ge \ \sup_{\Theta(\tau^*)} R(\hat{\theta}^*, \theta).$$

The key step is to show that τ^* is as hard for $\hat{\theta}^*$ as the full problem:

$$R(\hat{\theta}^*, \tau^*) \ge R(\hat{\theta}^*, \theta) \quad \text{for all } \ \theta \in \Theta. \tag{5.5}$$

It follows that

$$R_L^*(\sigma) \ = \ R(\hat{\theta}^*, \tau^*) \ \equiv \ R_L^*(\sigma; \Theta(\tau^*)).$$

Hence, (5.4).

To start, we identify the hardest rectangular subproblem. Let Θ_+ denote the positive orthant of Θ. As Θ is orthosymmetric, if $\theta \in \Theta$, then so is $(\pm\theta_i)_{i=0}^{\infty}$ for all sequences of signs $\pm$. As Θ is convex, if $\tau \in \Theta_+$, all $(\pm\theta_i)_{i=0}^{\infty}$ with $|\theta_i| \le \tau_i$ must belong to Θ. Therefore, $\Theta(\tau) \subset \Theta$ iff $\tau \in \Theta_+$. Hence, if we define for $\tau \in \Theta_+$

$$J(\tau) \ = \ \sum \rho_L(\tau_i, \sigma) \ = \ R_L^*(\sigma, \Theta(\tau)),$$

then

$$\sup\{ R_L^*(\sigma; \Theta(\tau)) : \ \Theta(\tau) \subset \Theta \} \ = \ \sup_{\Theta_+} J(\tau)$$

We claim that J is an l_2-continuous functional on Θ_+. From

$$\frac{r^2\sigma^2}{r^2+\sigma^2} - \frac{s^2\sigma^2}{s^2+\sigma^2} \le |r^2-s^2|$$

we get $|J(\theta)-J(\tau)| \le \sum |\theta_i^2-\tau_i^2|$. Let (θ_n) be a sequence in Θ_+ converging l_2-strongly to τ. Putting $t_{n,i}=\theta_{n,i}^2$ and $t_i=\tau_i^2$, we have $t_{n,i}\ge 0$, and $t_i \ge 0$. From the convergence θ_n to τ, we have $t_{n,i}\to t_i$ for each i, and $\sum t_{n,i} \to \sum t_i$. Applying Sheffé's Lemma, t_n converges to t in l_1. Thus $\sum |\theta_{n,i}^2-\tau_i^2| \to 0$. By the inequality above $|J(\theta)-J(\tau)| \to 0$.

As J is continuous, it follows from compactness of Θ that J has a maximum in Θ_+; τ^*, say. $\Theta(\tau^*)$ is the hardest rectangular subproblem for linear estimates.

To avoid typographical excess, let τ_i denote the i-th component of τ^*. The minimax linear estimator for $\Theta(\tau^*)$ is of the form $(c_i y_i)_{i=0}^{\infty}$, where $c_i = \dfrac{\tau_i^2}{\tau_i^2+\sigma^2}$. For the mean-squared error of this estimator, we have

$$\begin{aligned} R(\hat\theta^*,\theta) &= Bias^2 + Variance \\ &= \sum (1-c_i)^2\theta_i^2 + \sigma^2 \sum c_i^2. \end{aligned}$$

As we saw earlier, the theorem follows from the inequality (5.5). As the variance of $\hat\theta^*$ does not depend on θ, the inequality is equivalent to saying that $Bias^2(\theta)$ is maximized at $\theta=\tau^*$. As $Bias^2(\theta)$ does not depend on the signs of the components of θ, it is enough to check that it is maximized in the positive orthant at τ^*, i.e.

$$\sum (1-c_i)^2(\tau_i^2-\theta_i^2) \ge 0 \quad \text{for all } \theta\in\Theta_+. \tag{5.6}$$

Consider once again the functional J. We are going to show that $J(\theta)\le J(\tau^*)$ implies (5.6); the theorem then follows by definition of τ^* as the maximizer of J in Θ_+. We first change variables. For a generic θ in Θ_+, put $t=(\theta_i^2)_{i=0}^{\infty}$; put Θ_+^2 for the set of all such t. As Θ is quadratically convex, Θ_+^2 is convex. Define $\tilde J(t) = \sum \dfrac{t_i\sigma^2}{t_i+\sigma^2}$, so that $\tilde J(t) \equiv J(\theta)$. With $t_0 = (\tau_i^2)_{i=0}^{\infty}$, we have

$$\tilde J(t) \le \tilde J(t_0), \quad t\in\Theta_+^2. \tag{5.7}$$

We claim $\tilde J$ is Gâteaux differentiable on l_2 at t_0, with derivative

$$\langle D_{t_0}\tilde J, h\rangle = \sum (1-c_i)^2 h_i. \tag{5.8}$$

Now the maximum condition (5.7) gives $\langle D_{t_0}\tilde J, h\rangle \le 0$ for all $h=(t-t_0)$. Using this and the definition of t and t_0 will establish (5.6).

We provide the needed details. Let r and s denote scalars; a bit of algebra yields

$$\frac{(r+\varepsilon s)\sigma^2}{r+\varepsilon s+\sigma^2} - \frac{r\sigma^2}{r+\sigma^2} = \varepsilon s(1-c)^2 + \varepsilon^2 s^2 \frac{(1-c)^2}{r+\varepsilon s+\sigma^2} \tag{5.9}$$

where $c = r/(r+\sigma^2)$. Now if both $r \ge 0$ and $r+\varepsilon s \ge 0$, then $\frac{(1-c)^2}{r+\varepsilon s+\sigma^2} \le \frac{1}{\sigma^2}$. Now let $h \in l_2$; if $t_0+\varepsilon h \ge 0$ coordinatewise, applying (5.9) coordinatewise to the components of $\tilde{J}$, with $r = t_i$ and $s = h_i$, gives

$$|(\tilde{J}(t_0+\varepsilon h) - \tilde{J}(t_0)) - \varepsilon \sum (1-c_i)^2 h_i| \le \frac{\varepsilon^2}{\sigma^2} \sum h_i^2. \tag{5.10}$$

Now let $\theta \in \Theta$ and let t be the corresponding element of Θ_+^2. Define $t_\varepsilon = (1-\varepsilon)t_0 + \varepsilon t$. By convexity of Θ_+^2, $t_\varepsilon \in \Theta_+^2$. By (5.7), $\tilde{J}(t_\varepsilon) - \tilde{J}(t_0) \le 0$. It follows that

$$\varepsilon^{-1}\{\tilde{J}(t_\varepsilon) - \tilde{J}(t_0)\} \le 0 \qquad for\ \varepsilon \in (0,1]. \tag{5.11}$$

Now $t_\varepsilon = t_0 + \varepsilon h$ for $h = t - t_0$. Also,

$$\sum h_i^2 = \sum (\theta_i^2 - \tau_i^2)^2 = \sum (\theta_i - \tau_i)^2 (\theta_i + \tau_i)^2 \le 4M^2 \sum (\theta_i - \tau_i)^2 \le 16M^4 \tag{5.12}$$

where $M = \sup\{\|\theta\| : \theta \in \Theta\} < \infty$, by compactness of Θ. Using (5.10) and (5.11) with (5.12) gives

$$\sum (1-c_i)^2 (t_i - t_{0,i}) \le \frac{\varepsilon}{\sigma^2} 16M^4$$

for all $\varepsilon \in (0,1]$. Taking into account the definitions of $t_i = \theta_i^2$ and $t_0 = \tau_i^2$, this implies that (5.6) holds for every $\theta \in \Theta_+$.□

Remarks.

1. The concept of hardest rectangular subproblems appears to be new. Pinsker (1980) established a maximin property for ellipsoids which can be shown to imply (5.4) for ellipsoids (see eqs. 17-18, page 122 of the English translation). Thus our result is an abstraction and generalization. However, even for ellipsoids, the implication (5.3) seems to be new.

2. Theorem 7 does not cover l_p-bodies with $p < 2$. In fact (5.4) is not true in those cases. However, see sections 8, 9, and 10.

3. Pinsker (1980) showed that for certain ellipsoids,

$$\frac{R_L^*(\sigma)}{R_N^*(\sigma)} \to 1 \tag{5.13}$$

as $\sigma \to 0$. Fundamental to his argument is the idea that the hardest rectangular subproblem be *finite dimensional*. This is not true for l_p-bodies with $p > 2$, as one could discover from straightforward calculations based on Theorem 7. Possibly, ellipsoids are the only sets where (5.13) holds. As we saw in Theorem 5 and its corollary, (5.13) cannot hold for most hyperrectangles. So the class of cases where (5.13) holds is strictly smaller than those where the 25% bound holds.

6. Truncation estimates

Suppose, once again, that $\Theta = \Theta(\tau)$ is a hyperrectangle, and recall that the minimax estimator and minimax linear estimator for this situation are $\hat\theta^N$ and $\hat\theta^L$. A simple alternative to these estimates is the *truncated series* estimate $\hat\theta^T$, obtained by letting y_i serve as the estimate of θ_i in those coordinates at which $\tau_i > \sigma$ and letting 0 serve as the estimate of θ_i at those coordinates where $\tau_i \le \sigma$. Thus

$$\hat\theta_i^T = y_i I_{\{\tau_i > \sigma\}}.$$

We remark that $\hat\theta^T$ uses the data to estimate θ at those coordinates where the "signal-to-noise" ratio τ_i / σ is bigger than one; at other coordinates it ignores the data and just uses zero.

The term "truncated series estimate" derives from the function-smoothing viewpoint. The estimate $\hat f^T(t) = \sum_i \hat\theta_i^T \phi_i(t)$ estimates f by a series which is truncated as soon as the estimated coefficient has signal/noise ≤ 1. The maximum risk of $\hat\theta_i^T$ as an estimate of θ_i,

$$\rho_T(\tau_i, \sigma) = \max_{|\theta_i| \le \tau_i} E(\hat\theta_i^T - \theta_i)^2$$

is just σ^2 or τ_i^2 depending on whether $\tau_i > \sigma$ or $\tau_i \le \sigma$. Thus we have the simple formula which was used already in section 3. From this, we have the worst-case risk of $\hat\theta^T$:

$$R_T^*(\sigma) = \sup_{\theta \in \Theta} E \, \| \hat\theta^T - \theta \|^2 = \sum \rho_T(\tau_i, \sigma).$$

In fact, $R_T^*(\sigma)$ is the minimax risk among *all* truncation estimates. Indeed, let $\hat\theta_i^P = y_i I_{\{i \in P(\sigma)\}}$ where $P(\sigma)$ is a set of indices. The worst-case risk of $\hat\theta^P$ is

$$\sum_i \sigma^2 I_{\{i \in P(\sigma)\}} + \tau_i^2 I_{\{i \notin P(\sigma)\}} \ge \sum_i \min(\sigma^2, \tau_i^2) = R_T^*(\sigma).$$

Thus $\hat\theta^T$ is minimax among truncation estimates.

A common objection to truncation estimates is that their transition from "using the data" to "ignoring the data" is too abrupt. Estimates such as $\hat{\theta}^N$ and $\hat{\theta}^L$ in some sense manage a smooth transition from using the data ($\tau_i \gg \sigma$) to ignoring the data ($\tau_i \ll \sigma$). Surprisingly, truncated estimates do not do too badly in terms of minimax risks. We have

$$\frac{\rho_T(\tau,\sigma)}{\rho_L(\tau,\sigma)} = \frac{\min(\tau^2,\sigma^2)}{[\tau^2\sigma^2/(\sigma^2+\tau^2)]} = \frac{\min(\tau^2,\sigma^2)(\tau^2+\sigma^2)}{\tau^2\sigma^2} = (\tau^2+\sigma^2)/\max(\tau^2,\sigma^2) \le 2.$$

so

$$R_T^*(\sigma) = \sum \rho_T(\tau_i,\sigma) \le \sum 2 \cdot \rho_L(\tau_i,\sigma) = 2R_L^*(\sigma).$$

From Theorem 2 we have, for similar reasons, $R_T^*(\sigma) \le 2.22 R_N^*(\sigma)$. This proves

Theorem 8. *To minimize, among truncation rules, the worst-case risk over the hyperrectangle $\Theta(\tau)$, the optimal rule is to truncate at signal-to-noise ratio 1. The resulting risk is never worse than twice the minimax linear risk, and never worse than 2.22 times larger than the minimax risk.*

For asymptotics as $\sigma \to 0$ we can use the same averaging argument that led to Theorem 5, but this time on the ratio ρ_T/ρ_L rather than on μ. This leads to

Theorem 9. *Let $q > \frac{1}{2}$. If $\tau_i = ci^{-q}$ then*

$$\lim_{\sigma\to 0} \frac{R_T^*(\sigma)}{R_L^*(\sigma)} = \zeta_T(q) \equiv \int_0^1 (1+v^2) g_q(v)\,dv + \int_1^\infty (1+v^2)/v^2 g_q(v)\,dv$$

where the density g_q is defined in (4.5).

We omit the proof. We find the relatively good performance of truncation in this minimax setting surprising. See table 2.

7. N-widths and Minimax Risk

Suppose now that Θ is convex but not a hyperrectangle, and we are interested in estimating θ from data (2.1). Consider truncation estimates defined using projections -- $\hat{\theta} = Py$, $P^2 = P$. Define

$$R_T^*(\sigma;\Theta) = \inf_P \sup_\Theta E \, \| Py - \theta \|^2$$

where the infimum is over all linear projections. For hyperrectangles, the optimal projections are of course parallel to the coordinates, so this definition agrees with the one in section 6, and $R_T^*(\sigma;\Theta(\tau)) = \sum \rho_T(\tau_i,\sigma)$. If Θ is not a hyperrectangle, there is an obvious lower bound -- the full problem is at least as bad as any rectangular subproblem. Under quadratic convexity, the bound is near sharp:

Theorem 10. *Let Θ be compact, quadratically convex, and orthosymmetric. Then the difficulty, for truncation estimates, of the hardest rectangular subproblem, is at least* **half** *the difficulty, for truncation estimates, of the full problem:*

$$R_T^*(\sigma) \le 2\cdot\sup\{R_T^*(\sigma;\Theta(\tau)) : \Theta(\tau) \subset \Theta\} \tag{7.1}$$

Proof. We use notation from the proof of Theorem 7. Put $J(\tau) = \sum \rho_T(\tau_i, \sigma)$ for $\tau \in \Theta_+$. We have $|J(\theta) - J(\tau)| \le \sum |\theta_i^2 - \tau_i^2|$, so arguing as in the proof of Theorem 7, J is l_2-continuous on Θ_+. A maximizer $\tau^* = (\tau_i)_{i=0}^\infty$ exists by compactness. $\Theta(\tau^*)$ is the hardest rectangular subproblem for truncation estimates.

For a generic $\theta \in \Theta_+$, define a corresponding $t \in \Theta_+^2$ by $t_i = \theta_i^2$; put $\tilde{J}(t) = \sum \min(t_i, \sigma^2)$ and $t_{0,i} = \tau_i^2$. Note that $J(\theta) = \tilde{J}(t)$. $\tilde{J}$ is a concave functional maximized over Θ_+^2 at t_0. The Gateaux differential of $\tilde{J}$ is not, in general, additive. Nevertheless, for the differential $D\tilde{J}$ of $\tilde{J}$ at t_0, in direction h, the maximum condition gives

$$D\tilde{J}_{t_0}(h) \le 0 \tag{7.2}$$

for every h of the form $t - t_0$, $t \in \Theta_+^2$. Let P denote the set of indices i such that $t_{0,i} \ge \sigma^2$, and let Q denote the set where $t_{0,i} = \sigma^2$. A calculation gives

$$D\tilde{J}_{t_0}(h) = \sum_{i \notin P} h_i - \sum_{i \in Q} (h_i)_- ; \tag{7.3}$$

where $(a)_- = |a| I_{a<0}$. We omit details here; they are similar to those given in the proof of Theorem 7. From (7.2) and (7.3) we get $\sum_{i \notin P} (t_i - t_{0,i}) \le \sum_{i \in Q} (t_i - t_{0,i})_-$, or, as $(t_i - t_{0,i})_- \le \sigma^2$,

$$\sum_{i \notin P} t_i \le \sum_{i \notin P} t_{0,i} + \sigma^2 \sum_{i \in Q} 1$$

Translating back to θ-coordinates, we get

$$\sum_{i \notin P} \theta_i^2 \le \sum_{i \notin P} \tau_i^2 + \sigma^2 \sum_{i \in Q} 1. \tag{7.4}$$

Consider the minimax truncation estimator $\hat{\theta}^*$ for $\Theta(\tau^*)$; given by $\hat{\theta}_i^* = y_i I_{\{i \in P\}}$. It has risk

$$R(\hat{\theta}^*,\theta) = \sum_{i \notin P} \theta_i^2 + \sigma^2 \sum_{i \in P} 1.$$

Since $Q \subset P$, (7.4) gives

$$R(\hat{\theta}^*,\theta) \leq \sum_{i \notin P} \tau_i^2 + 2 \cdot \sigma^2 \sum_{i \in P} 1 \leq 2 \cdot \sum \rho_T(\tau_i,\sigma).$$

The last step follows from the definition of P, via $\rho_T(\tau_i,\sigma) = \tau_i^2 I_{i \notin P} + \sigma^2 I_{i \in P}$.□

Corollary. *If Θ is orthosymmetric, compact, and quadratically convex, then*

$$R_T^*(\sigma) \leq 4.44 \cdot R_N^*(\sigma)$$

As in Theorem 9, one could show in specific cases a more precise result in the asymptotic case $\sigma \to 0$.

It follows that n-widths of the set Θ determine the difficulty of estimation quite precisely. The (Kolmogorov Linear) n-width of Θ is defined as (see Pinkus, 1984)

$$d_n = \inf_{P_n} \sup_{\Theta} \| P_n \theta - \theta \|$$

the infimum being over all n-dimensional projections. Then we have

$$R_T^*(\sigma) = \inf_n d_n^2 + n\sigma^2.$$

Thus, for Θ orthosymmetric and quadratically convex, the corollary shows that the purely geometric quantity $\inf_n d_n^2 + n\sigma^2$ is within a factor 4.44 of the minimax risk. In particular, if the n-widths go to zero at rate n^{-r}, then $R_N^*(\sigma) \to 0$ at rate $(\sigma^2)^{\frac{2r}{2r+1}}$.

8. Non Quadratically Convex Sets

Let Θ be a set. The *quadratically convex hull* of Θ is

$$QHull(\Theta) = \{\theta : (\theta_i^2) \in Hull(\Theta_+^2)\}. \tag{8.1}$$

For quadratically convex, closed orthosymmetric sets, of course, $QHull(\Theta) = \Theta$. On the other hand, for weighted l_p-bodies with $p < 2$, the hull is strictly larger than the set itself. Indeed, if $\Theta_p(a)$ denotes $\{\theta : \sum a_i |\theta_i|^p \leq 1\}$, one can easily compute

$$QHull(\Theta_p(a)) = \{\theta : \sum a_i^{2/p} |\theta_i|^p \leq 1\} \tag{8.2}$$

Thus for all the weighted l_p-bodies with $p \in (0,2)$, the quadratic hull is an *ellipsoid*. (More is true. Consider the function-smoothing interpretation, with $a_i = i^{pq}$ representing smoothness constraints on the q-th derivative. For every $p \in [0,2)$, the quadratic hull is the ellipsoid with weights $a_i = i^{2q}$!) The key fact about quadratic convexifications is that it preserves minimax risks of *linear* estimators.

Theorem 11. *Let* Θ *be orthosymmetric and compact.*

$$R_T^*(\sigma;\Theta) = R_T^*(\sigma;QHull(\Theta)) \tag{8.3}$$

$$R_L^*(\sigma;\Theta) = R_L^*(\sigma;QHull(\Theta)) \tag{8.4}$$

Before giving the proof, some remarks. First, for linear estimation, l_p-type constraints, with $p<2$, do not add anything new; by (8.2)-(8.4) the difficulty is the same as with the ellipsoidal constraints of the corresponding quadratic hull. Second, Theorems 7 and 11 together say that the minimax linear risk is still determined by the hardest rectangular subproblem -- of the *enlarged* set $QHull(\Theta)$. Finally, let $\Theta(\tau^*)$ be the hardest rectangular subproblem of $QHull(\Theta)$ for truncation estimates. Then

$$\begin{aligned} R_L^*(\sigma;\Theta) &\geq R_L^*(\sigma;\Theta(\tau^*)) \\ &\geq \frac{1}{2}R_T^*(\sigma;\Theta(\tau^*)) \geq \frac{1}{4}R_T^*(\sigma;QHull(\Theta)) = \frac{1}{4}R_T^*(\sigma;\Theta) \end{aligned}$$

which proves

Corollary. *Let* Θ *be orthosymmetric and compact. Then*

$$R_T^*(\sigma;\Theta) \leq 4R_L^*(\sigma;\Theta).$$

So for weighted l_p-bodies with $p \in (0,\infty)$, the minimax linear estimator never improves drastically on minimax truncated series estimators.

As a final remark, note that the formula $R_T^*(\sigma) = \inf_n d_n^2 + n\sigma^2$ always determines the difficulty of truncated series estimates. It follows from the Corollary that under orthosymmetry the n-widths determine the difficulty of linear estimation to within a factor 4.

Proof of Theorem 11.

Let C be a compact linear operator on l_2, and let $\hat{\theta} = Cy$ be the estimator it induces. Then

$$R(\hat{\theta},\theta) = \|(C-I)\theta\|^2 + \sigma^2\|C\|_{HS}$$

where I denotes the identity and $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm. The appendix proves the inequality

$$\sup_{(\theta_i)=(\pm\tau_i)} \|(C-I)\theta\|^2 \geq \sup_{(\theta_i)=(\pm\tau_i)} \|(Diag(C)-I)\theta\|^2 \tag{8.5}$$

and we also have

$$\|C\|_{HS} \geq \|Diag(C)\|_{HS} \quad . \tag{8.6}$$

Together, these imply that $Diag(C)$ has a smaller worst-case risk than C. Hence there is a minimax linear estimator of the diagonal form $\hat{\theta}_i = c_i y_i$, and in fact with each $c_i \in [0,1]$. Similarly, there is a minimax truncation estimator of the form $\hat{\theta}_i = c_i y_i$ with each $c_i \in \{0,1\}$. The risk of such estimators has the form

$$R(\hat{\theta},\theta) = \sum(1-c_i)^2\theta_i^2 + \sigma^2\sum c_i^2. \tag{8.7}$$

Now let $\tilde{\theta}$ be an element of $QHull(\Theta)$. Let $\tilde{t}$ be the corresponding point in $Hull(\Theta_+^2)$, defined by $\tilde{t}_i = \tilde{\theta}_i^2$. We have an integral representation $\tilde{t} = \int t \, d\mu(t)$ with μ a probability measure on Θ_+^2. Let π be the probability measure on Θ_+ induced by μ via the change of variables formula. Now obviously

$$\sup_{\Theta} R(\hat{\theta},\theta) \geq \int R(\hat{\theta},\theta) \, d\pi(\theta);$$

but, using (8.7)

$$\int R(\hat{\theta},\theta) d\pi(\theta) = \int \left[\sum(1-c_i)^2\theta_i^2 \right] d\pi(\theta) + \sigma^2\sum c_i^2.$$

Now by the construction of π and μ, and the change of variables formula,

$$\int \theta_i^2 \, d\pi(\theta) = \int t_i \, d\mu(t) = \tilde{t}_i = \tilde{\theta}_i^2$$

so

$$\int R(\hat{\theta},\theta) d\pi(\theta) = \sum(1-c_i)^2\tilde{\theta}_i^2 + \sigma^2\sum c_i^2 = R(\hat{\theta},\tilde{\theta}).$$

Hence $\sup_{\Theta} R(\hat{\theta},\theta) \geq R(\hat{\theta},\tilde{\theta})$ for every $\tilde{\theta} \in QHull(\Theta)$: $QHull(\Theta)$ is no harder for such an estimator than Θ itself. Results (8.3)-(8.4) follow. □

9. Difficulty of Non-Quadratically Convex Classes

If Θ is orthosymmetric but not quadratically convex, $QHull(\Theta)$ is larger than Θ itself. The two sets can, in fact, be quite different. Consider the l_1 body with weights $a_i = i^q$. A calculation based on the results of the last two sections reveals that the hardest rectangular subproblem of $QHull(\Theta)$ has risk which goes to zero as $(\sigma^2)^{\frac{2q}{2q+1}}$. However, as explained in section 10 below, the hardest rectangular subproblem in Θ has difficulty comparable to $(\sigma^2)^{\frac{2q+1}{2q+2}}$, which is much smaller.

A difference of this sort *guarantees* that linear estimators are *not* nearly minimax. This follows from

Theorem 12. *Let $p \in (0,\infty)$. Consider the l_p-body $\Theta_p(a)$ with weights $a_i \ge c i^{pq}$ for some $q>0$. Then*

$$R_N^*(\sigma;\Theta) \le M(\sigma) \sup\{R_N^*(\sigma;\Theta(\tau)) : \Theta(\tau) \subset \Theta\} \tag{9.1}$$

where

$$M(\sigma) = O(|\log\sigma|^2) \tag{9.2}$$

as $\sigma \to 0$.

In words, the hardest rectangular subproblem of $\Theta_p(a)$ is, to within logarithmic factors, as hard as the full problem. Hence if the difficulty of the hardest subproblem of $QHull(\Theta)$ tends to zero at a different rate from the difficulty of the hardest subproblem for Θ, the risk of linear estimators cannot tend to zero at the optimal rate. So, for example in the l_1-body case mentioned above, linear estimators are *not* nearly minimax.

Proof. By Theorem 8, the difficulty of the hardest subproblem is within a factor 2.22 of $\sup\{R_T^*(\sigma,\Theta(\tau)):\Theta(\tau)\subset\Theta\}$. The result (9.1) therefore follows if we can show that

$$R_N^*(\sigma;\Theta) \le M(\sigma) \sup_{\theta\in\Theta} \sum \min(\theta_i^2,\sigma^2) \tag{9.3}$$

with $M(\sigma)$ satisfying (9.2).

We now construct an estimator which proves that (9.2)-(9.3) hold. Pick $C = C(\sigma)$ so that $C \ge 1$ and $C^2(\sigma) \asymp |\log\sigma|^2$ as $\sigma \to 0$. Define

$$T = \{i : \sup_{\Theta} |\theta_i| > C\sigma\}.$$

Define the estimator $\hat\theta$ by the rule

$$\hat\theta_i = \begin{cases} sgn(y_i)(|y_i| - C\sigma)_+ & i \in T \\ 0 & i \notin T \end{cases} \tag{9.4}$$

In words, $\hat\theta$ is zero at those coordinates which cannot possibly be large, and translates towards zero in those coordinates which might possibly be large; compare Bickel (1983).

To analyze the worst-case behavior of $\hat\theta$, fix $\varepsilon \in (0,1)$. Given θ, define

$$B = \{i : |\theta_i| \ge \varepsilon\sigma\}$$
$$S = \{i : |\theta_i| < \varepsilon\sigma\}$$

the indices of the "big" and "small" coordinates of θ, respectively. Note that if $i \in T$, then $\hat\theta_i = y_i + \psi(y_i)$, where $|\psi(y_i)| \le C\sigma$. Therefore, if $i \in T$,

$$E(\hat{\theta}_i-\theta_i)^2 = E(y_i-\theta_i+\psi(y_i))^2$$
$$\le\left[\sqrt{E(y_i-\theta_i)^2}+\sqrt{E\psi^2(y_i)}\right]^2 \le (\sigma+C\sigma)^2 \qquad (9.5)$$

Also, if $i \notin T$

$$E(\hat{\theta}_i-\theta_i)^2 = \theta_i^2 \quad ; \qquad (9.6)$$

and, finally, if $i \in S\cap T$

$$E(\hat{\theta}_i-\theta_i)^2 \le 2\theta_i^2(1+4\phi(C-\varepsilon)) + 4\sigma^2[C+1]\phi(C-\varepsilon) \qquad (9.7)$$

where $\phi(t)$ is the N(0,1) density (this is proved in the appendix). For small σ, $C-\varepsilon>1$, and so $4\phi(C-\varepsilon)\le 1$. Combining (9.5)-(9.7),

$$\sum_i E(\hat{\theta}_i-\theta_i)^2 \le (C+1)^2\sum_{i\in B}\sigma^2 + 4\sum_{i\in S}\theta_i^2 + \sum_{i\in S\cap T}\sigma^2 4\,[C+1]\,\phi(C-\varepsilon)$$

Now as $C\ge 1$,

$$(C+1)^2\sigma^2 I_{\{|\theta_i|\ge\varepsilon\sigma\}} + 4\theta_i^2 I_{\{|\theta_i|<\varepsilon\sigma\}} \le \frac{(C+1)^2}{\varepsilon^2}\min(\theta_i^2,\sigma^2)$$

Recalling the definitions of B and S, we have

$$\sum_i E(\hat{\theta}_i-\theta_i)^2 \le \frac{(C+1)^2}{\varepsilon^2}\sum_i\min(\theta_i^2,\sigma^2) + Rem(C,\sigma)$$

where

$$Rem(C,\sigma) = 4\sigma^2[C+1]Card(T)\phi(C-\varepsilon)$$

Now, by the assumption that $a_i \ge c i^{pq}$, we have $Card(T) = O(\sigma^{-r})$ with $r=r(q)=1/q+1$. Also, $C+1 = O(|\log\sigma|)$ by definition of C. Therefore, as $\sigma\to 0$,

$$\frac{Rem(C,\sigma)}{\sigma^2} = O(|\log\sigma|\sigma^{-r}\exp(-|\log\sigma|^2/2)).$$

As $\sigma\to 0$, $\exp(-|\log\sigma|^2/2) = o(\exp(-R|\log\sigma|)) = o(\sigma^R)$ for every $R>0$. In particular, for $R>r$. We conclude that

$$\frac{Rem(C,\sigma)}{\sigma^2} \to 0 \qquad (9.8)$$

as $\sigma\to 0$. On the other hand, as Θ contains nonzero elements (otherwise the theorem is trivially true),

$$\sup_{\theta\in\Theta}\sum\min(\theta_i^2,\sigma^2) \ge \sigma^2(1+o(1)) \qquad (9.9)$$

as $\sigma\to 0$. Defining

$$M(\sigma) = \frac{(C+1)^2}{\varepsilon^2} + \frac{Rem(C,\sigma)}{\sigma^2(1+o(1))} \tag{9.10}$$

with the $o(1)$ term the same as in (9.9), we have

$$\begin{aligned} R_N^* &\le \sup_\theta \sum E(\hat{\theta}_i - \theta_i)^2 \\ &\le \sup_\theta \left[\frac{(C+1)^2}{\varepsilon^2} \sum \min(\theta_i^2, \sigma^2) + Rem(C,\sigma) \right] \\ &\le M(\sigma) \sup_\theta \sum \min(\theta_i^2, \sigma^2) \end{aligned}$$

This is of the same form as (9.3), where $M(\sigma)$ satisfies (9.2) because of (9.8). □

10. Hardest Cubical Subproblems of l_p bodies, $p \le 2$

Definition: a standard *n-cube of radius* τ is a set $\Theta_n(\tau,\mathbf{i})$ of elements θ such that $|\theta_i| \le \tau$ for indices $i \in \mathbf{i}$, $\theta_i = 0$ for indices $i \notin \mathbf{i}$, and $Card(\mathbf{i}) = n$.

Theorem 13. *Let $\Theta = \Theta_p(a)$ for $0 < p \le 2$. Let $n_0 = n_0(\sigma)$ be the largest n for which an n-cube of radius σ fits in Θ. Then the difficulty, for truncation estimates, of the hardest rectangular subproblem, is essentially the same as the difficulty of this n_0-cube:*

$$n_0\sigma^2 = \sup\{R_T^*(\sigma;\Theta_n(\sigma,\mathbf{i})) : \Theta_n(\sigma,\mathbf{i}) \subset \Theta\} \tag{10.1}$$

$$(n_0+1)\sigma^2 \ge \sup\{R_T^*(\sigma;\Theta(\tau)) : \Theta(\tau) \subset \Theta\}. \tag{10.2}$$

The proof is given in the appendix. Ignoring constants, the Theorem reduces the calculation of asymptotic behavior for the hardest subproblem to calculation of $n_0(\sigma)$. This is straightforward. Consider the l_p-body with weights $a_i = i^{pq}$ for $p < 2$. If an n-cube of size σ fits in Θ at all, it can be fit using the first n-coordinates for $\mathbf{i}$. Therefore, n_0 satisfies

$$\sigma^p \sum_0^{n_0-1} i^{pq} \le 1$$

$$\sigma^p \sum_0^{n_0} i^{pq} > 1.$$

One sees immediately that $\sigma^p n_0^{pq+1} \to pq+1$, and

$$n_0\sigma^2 = O\left((\sigma^2)^{\frac{2pq+2-p}{2pq+2}}\right). \tag{10.3}$$

As $p < 2$, this goes to zero faster than the risk for the linear minimax estimator in this case, which by section 7 is $(\sigma^2)^{\frac{2q}{2q+1}}$. Hence, the conclusion of the introduction: there exist settings in which nonlinear estimates improve on linear ones by an arbitrarily large factor in the worst case.

Remarks.

1. Formula (10.3) shows that p is, to some extent, a smoothness parameter. Think of the function-smoothing interpretation. With q, the "order of differentiability", fixed, the optimal rate of convergence improves as p gets smaller. As $p \to 0$, in fact, the rate tends (modulo logarithmic factors) to σ^2, which is the rate which would obtain if Θ were finite-dimensional.

2. The quantity n_0 is closely related to the so-called Bernstein (or inner) n-widths of Θ (Pinkus, 1985). Let $b_{n,\infty}$ denote the largest radius of an $n+1$-dimensional l_∞-ball which can be inscribed in Θ. Then $n_0 = 1 + \sup\{n : b_{n,\infty} \geq \sigma\}$. Theorems 12 and 13 attribute a central role for $b_{n,\infty}$ in determining the difficulty of estimation for l_p-bodies with $p \leq 2$. In particular, if the $b_{n,\infty}$ go to zero at rate n^{-s}, then, in the cases covered by Theorems 12 and 13, the minimax risk goes to zero as $(\sigma^2)^{\frac{-2s+1}{2s}}$ (ignoring logarithmic factors).

As seen above, the Kolmogorov n-widths of $\Theta_p(a)$ determine the performance of truncated series estimates, and more generally, of linear estimates. Thus, if the d_n go to zero at rate n^{-r}, the minimax linear risk goes to zero at rate $(\sigma^2)^{\frac{-2r}{2r+1}}$.

Comparing the last two paragraphs, we see that for the minimax linear risk and minimax risk to converge to zero at the same rate requires that $\frac{2s-1}{2s} = \frac{2r}{2r+1}$. Hence, $s = r + 1/2$. In other words, for n sufficiently large and some $c > 0$,

$$b_{n,\infty} \geq c \frac{d_n}{\sqrt{n}}. \tag{10.4}$$

A comparison of d_n and $b_{n,\infty}$ can be effected as follows. Let $b_{n,2}$ denote the largest radius of any $n+1$-dimensional l_2-ball which can be inscribed in Θ. (This is the classical Bernstein n-width; see Pinkus). As the sphere of radius 1 inscribes the cube of radius 1, and as the cube inscribes the sphere of radius $\sqrt{n+1}$,

$$b_{n,\infty} \leq b_{n,2} \leq \sqrt{n+1}\, b_{n,\infty}. \tag{10.5}$$

Also, we have (Pinkus, 1985, Page 13)

$$b_{n,2} \le d_n . \tag{10.6}$$

Combining (10.5) and (10.6), a sufficient condition for (10.4) is $b_{n,2} = d_n$. This equality of Bernstein and Kolmogorov n-widths occurs for ellipsoids (Pinkus, 1985, Chapter VI, Theorem 1.3, Page 199), but for very few other cases. The l_p-bodies, with $p < 2$ show that we can have

$$b_{n,2} \le \frac{d_n}{(n+1)^{1/p-1/2}}$$

If this sort of relation holds, and we put $p < 1$, (10.4) must fail, no matter how favorable the relation between $b_{n,2}$ and $b_{n,\infty}$ in (10.5).

To summarize, when Theorems 12 and 13 apply, the statement that the minimax linear and minimax nonlinear risks go to zero at different rates is basically equivalent to the statement that certain Bernstein n-widths are significantly smaller than the Kolmogorov n-widths. While this cannot happen for l_2-bodies, this is precisely what happens for l_p-bodies with $p<2$.

The linear n-widths of Kolmogorov have commonly been regarded as fundamental by approximation theorists, while Bernstein n-widths have been regarded as simply a tool for getting bounds on the n-widths of Kolmogorov (Pinkus, 1984, page 12). In this setting of statistical estimation, the reverse is true. Certain Bernstein n-widths determine (up to logarithmic factors) the difficulty of estimation, while the Kolmogorov n-widths measure the difficulty of linear estimation, which is in our view less fundamental.

11. Use of l_1-loss

We could have considered the problem (1.1)-(1.3) with the l_1-loss function: $\|\hat{\theta} - \theta\|_1 = \sum |\hat{\theta}_i - \theta_i|$. In order to do so, we would need to know the minimax risks in the bounded normal mean problem for l_1-loss. These apparently have not been studied previously. Let $\lambda_N(\tau,\sigma)$, $\lambda_L(\tau,\sigma)$ and $\lambda_T(\tau,\sigma)$ be the minimax nonlinear, linear, and truncation risks, respectively. We have

$$\lambda_T(\tau,\sigma) = \min\left(\tau, \sqrt{\frac{2}{\pi}}\,\sigma\right)$$

and from numerical work parallel to that described in this report,

$$\lambda_T(\tau,\sigma) \le 1.87\, \lambda_N(\tau,\sigma)$$

$$\lambda_L(\tau,\sigma) \le 1.23\ \lambda_N(\tau,\sigma)\ .$$

Also, the minimax risks over hyperrectangles are $\sum \lambda_N(\tau_i,\sigma)$, $\sum \lambda_L(\tau_i,\sigma)$, $\sum \lambda_T(\tau_i,\sigma)$ respectively. Finally, by an argument similar to the proof of Theorem 10 we have

Theorem 14. *Let Θ be orthosymmetric, convex, and compact for the l_1-norm. Then the l_1-difficulty for truncation estimates of the hardest rectangular subproblem in Θ is at least half the l_1-difficulty of the full problem.*

In short for the l_p-bodies $p \ge 1$, the minimax l_1 risk is within a factor 3.8 of the geometric quantity

$$\inf_n\ d_{n,1} + n\ \sigma\sqrt{\frac{2}{\pi}}$$

where $d_{n,1}$ denotes the Kolmogorov linear n-width of Θ in l_1-norm.

12. Appendix

Proof of (3.5)

Casella and Strawderman (1981) show that for $v < 1.05$, $\rho_N(v,1) = \rho(\pi_{2,v})$, where $\pi_{2,v} = \frac{1}{2}(\delta_v + \delta_{-v})$. Le Cam (1985, page 42) gives a formula for the exact Bayes Risk in estimation problems with squared error loss, which says that to estimate $\theta \in \{0,1\}$ from one observation of P_θ, the minimax risk is

$$\inf_{\hat\theta} \sup_{\{0,1\}} E_\theta(\hat\theta - \theta)^2 = \frac{1}{2}\int \frac{dP_0 dP_1}{dP_0 + dP_1} .$$

Now consider the problem of estimating $t \in \{-v, v\}$ from one observation from Φ_t, the distribution of $N(t,1)$. With $t = 2v(\theta - \frac{1}{2})$, $P_0 = \Phi_{-v}$, $P_1 = \Phi_v$, we have

$$\inf_{\hat t} \sup_{t \in \{-v,v\}} E_t(\hat t - t)^2 = 4v^2 \frac{1}{2}\int \frac{d\Phi_{-v}\, d\Phi_v}{d\Phi_{-v} + d\Phi_v} . \tag{12.1}$$

Now using

$$e^{-(y-v)^2/2} e^{-(y+v)^2/2} = e^{-(y^2+v^2)}$$
$$e^{-(y-v)^2/2} = e^{-(y^2+v^2)/2} e^{vy}$$
$$e^{-(y+v)^2/2} = e^{-(y^2+v^2)/2} e^{-vy}$$

we have, using ϕ_t for the density of Φ_t,

$$\int \frac{\phi_v \phi_{-v}}{\phi_v + \phi_{-v}} = \int \frac{e^{-(y^2+v^2)/2}/\sqrt{2\pi}}{e^{vy} + e^{-vy}} = e^{-v^2/2} \int \frac{\phi_0(y)}{2\cosh(vy)}\, dy$$

which, combined with (12.1), gives (3.5). □

Lemma 12.1 (Monotonicity) *For $v \ge 3$,*

$$m(v) \equiv \left[\frac{v^2}{1+v^2}\right] / \left[1 - \frac{sinh(v)}{v\ cosh(v)}\right]$$

is monotonically decreasing as v increases.

Proof. Symbolically differentiating $m(v)$ using the Macsyma symbolic manipulator, we have

$$\frac{dm}{dv} = \left[(v^2+1)\left[1-\frac{s}{v\,c}\right]\right]^{-1} \cdot v \cdot \left[2 - \frac{2\,v^2}{v^2+1} - \frac{v\left[\frac{s}{v^2\,c} + \frac{s^2}{v\,c^2} - \frac{1}{v}\right]}{\left[1-\frac{s}{v\,c}\right]}\right]$$

where $s = sinh(v)$, $c = cosh(v)$.

Note that, for $v \geq 1$,

$$\left[1 - \frac{s}{v\,c}\right] \geq 1 - \frac{s}{c} \geq 0\,.$$

Take the common denominator for the last term in $\frac{dm}{dv}$. The numerator will be

$$2\left[\left[1-\frac{s}{v\,c}\right] - (v^2+1)\left[\frac{s\,c + v^2\,s^2 - v\,c^2}{2\,v\,c^2}\right]\right].$$

Call the term in square brackets **I**. If $\mathbf{I} \geq 1$, then $m'(v) \leq 0$. Thus the lemma reduces to showing that $\mathbf{I} \geq 1$ for $v \geq 3$. Since $\frac{sinh}{cosh}$ is monotone increasing for $v \geq 0$, we have for $v \geq 3$ that

$$\frac{s}{v\,c} + \frac{v\,s^2}{c^2} \geq v\left[\frac{s}{c}\right]^2 \geq 3\left[\frac{s}{c}\right]^2 |_{v=3} = 2.970 \geq 2$$

which completes the proof. □

Description of Numerical Approach

Our approach to bounding $\rho_N(\tau,1)$ works in two stages.

Stage 1. With N, M, and Ω parameters, define $x_i = (i/N)\,(\tau+\Omega)$; $dx = x_i - x_{i-1}$. Put

$$I^o(F) = 2\sum_{i=0}^{N} (f'(x_i))^2/f(x_i)\cdot dx\qquad .$$

This is intended as a crude approximation to $\int_{-(\tau+\Omega)}^{\tau+\Omega} \frac{(f'(x))^2}{f(x)}\,dx$.

Let $t_j = \frac{j}{M}\cdot\tau$, $|j| \leq M$, and put $\Pi_\tau^M = \{\pi : supp(\pi) \subset \{t_j\}\}$. Now Π_τ^M is a convex, $2M+1$ dimensional set, and there are explicit formulas for f' and f when $F = \Phi * \pi$ with $\pi \in \Pi_\tau^M$. The problem

$$\min\,\{I^o(\Phi * \pi) : \pi \in \Pi_\tau^M\}$$

is therefore one of optimizing a smooth convex function over a finite dimensional convex set. We used the optimization system NPSOL developed in the Systems Optimization Laboratory at Stanford University --- see Gill, Murray, Saunders, and Wright (1986) --- to find a numerical "solution" to this problem; call it π^o. We claim *no* optimality of π^o.

Stage 2. with N_1, τ, Ω, and π^o parameters, we attempt to find an *upper bound* on $I(\Phi * \pi^o)$. Let $x_i = \frac{i}{N_1} \cdot (\tau + \Omega)$, $|i| \le N_1$, and put

$$I^1(F) = 2 \cdot dx \sum_{i=0}^{N_1-1} \sup_{x \in (x_i, x_{i+1})} \frac{(f'(x_i))^2}{f(x)} + 2 \cdot C .$$

Here $C = C(\tau,\Omega)$ is an absolute constant so that

$$C \ge \sup \{ \int_{\tau+\Omega}^{\infty} \frac{(f')^2}{f} : F = \Phi * \pi , \pi \in \Pi_\tau \} ;$$

for example, with $\Omega = 6$ and $\tau = 5$, $C \le 2 \cdot 10^{-7}$. Because of this, we have at once

$$I^1(\Phi * \pi) \ge I(\Phi * \pi^o) .$$

However, I^1 is not actually computable, because of the "sup" specified in its definition. Note, however, that for $f = \phi * \pi^o$, $g = (f')^2/f$ is an analytic function; an absolute upper bound on the number S of sign changes of g' follows immediately from just the fact $\pi^o \in \Pi_\tau^M$. In any interval $[x_i, x_{i+1}]$ where there is no zero of g', $\max \{g(x) : x \in [x_i, x_{i+1}]\} = \max \{g(x_i), g(x_{i+1})\}$. In any interval where there is a zero of g', a conservative bound on $\max \{g(x) : x \in [x_i, x_{i+1}]\}$ is

$$\max \{g(x_i), g(x_{i+1})\} + D \cdot dx / 2$$

with $D \ge \sup_x |g'(x)|$. Define now

$$I^2(F) = 2 \cdot dx \sum_{i=0}^{N_1-1} \max \left[\frac{(f'(x_i))^2}{f(x_i)}, \frac{(f'(x_{i+1}))^2}{f(x_{i+1})} \right] + (dx)^2 \cdot D \cdot S + 2C$$

we have

$$I^2(\Phi * \pi^o) \ge I^1(\Phi * \pi^o) \ge I(\Phi * \pi^o) .$$

Justification of (3.8)

The numbers printed in columns 2 and 6 of Tables 3.1 - 3.3 are *numerical* evaluations of $\hat{\rho}_N = 1 - I^2(\Phi * \pi^o)$ on a SUN-4 computer using IEEE-standard double precision arithmetic. From the above, (3.8) follows provided we can evaluate I^2 to 4 digits accuracy. This is the same as saying we can evaluate a sum of the form

$$\sum_{i=0}^{N_1 - 1} \frac{(f'(y_i))^2}{f(y_i)}$$

to 4 digits *relative* accuracy, where each $y_i = x_i$ or x_{i+1}. Now

$$f'(y) = - \sum_{j=-M}^{M} \pi^o(t_j)\,(x - t_j)\,\exp(-(x - t_j)^2 / 2) / \sqrt{2\pi}$$

$$f(y) = \sum_{j=-M}^{M} \pi^o(t_j)\,\exp(-(x - t_j)^2 / 2) / \sqrt{2\pi} \ .$$

By lengthy but standard arguments, it is possible to show that this is possible with double precision arithmetic, assuming that the exponential function can be evaluated on the computer to 14 digits accuracy, that $N_1 < 20{,}000$ and $M < 100$, and that addition, division, and multiplication work on the computer precisely according to IEEE standards. Details of the argument are available from the authors.

Proof of Theorem 1. We proceed in three steps, showing that $\frac{\rho_L(\nu,1)}{\rho_N(\nu,1)} \le 1.25$ on each of the three ranges [0, .42], [.42, 4.2], [4.2, ∞).

Range [0, .42]. As $\rho_L(\nu,1) \le \rho_T(\nu,1)$,

$$\sup_{\nu \le .42} \frac{\rho_L(\nu,1)}{\rho_N(\nu,1)} \le \sup_{\nu \le .42} \frac{\rho_T(\nu,1)}{\rho_N(\nu,1)} = \frac{\rho_T(.42,1)}{\rho_N(.42,1)} \le \frac{.1762}{(.145669 - 0.0005)} \le 1.25$$

by the monotonicity of $\rho_T(\nu,1) / \rho_N(\nu,1)$ for $\nu \in [0,1]$ (see the proof of Theorem 2).

Range [4.2, ∞). By (3.6),

$$\sup_{\nu \ge 4.2} \frac{\rho_L(\nu,1)}{\rho_N(\nu,1)} \le \sup_{\nu \ge 4.2} \frac{\nu^2 (1 + \nu^2)^{-1}}{\left[1 - \frac{sinh(\nu)}{\nu\ cosh(\nu)}\right]} = \frac{(4.2)^2 (1 + (4.2)^2)^{-1}}{\left[1 - \frac{sinh(4.2)}{(4.2)\ cosh(4.2)}\right]} \le 1.25$$

where we have used Lemma 12.1, which establishes the monotonicity of the ratio for $\nu \ge 3$.

Range [.42, 4.2]. Suppose we have numerical approximations $\hat\rho_N(\tau_i,1)$ accurate to within δ, at a sequence $\{\tau_i\}$. As $\rho_L(\tau,1)$ and $\rho_N(\tau,1)$ are both monotone in τ,

$$\frac{\rho_L(\tau,1)}{\rho_N(\tau,1)} \le \frac{\rho_L(\tau_{i+1},1)}{\hat\rho_N(\tau_i,1) - \delta}$$

where $\tau_i \le \tau \le \tau_{i+1}$. Therefore, picking $\{\tau_i\}$ appropriately

$$\sup_{.42 \le \tau \le 4.2} \frac{\rho_L(\tau,1)}{\rho_N(\tau,1)} \le \max_i \frac{\rho_L(\tau_{i+1},1)}{\hat{\rho}_N(\tau_i,1) - \delta} .$$

Our computations used the 656 points $\{\tau_i\} = \{.42,.44,.46, \cdots ,4.2\} \cup \{1.381,1.382, \cdots 1.859,1.860\}$. By (3.8) $\delta = .5\ (.1) \cdot 10^{-4}$ (4 digit accuracy), giving 1.2497... for the right hand side of the above display. See Tables 3.1 - 3.3 □

Table 3.1

τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(\nu) \le \mu_i$	τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(\nu) \le \mu_i$
0.40	0.136657	0.1379310	1.108271	1.86	0.624035	0.7757650	1.249858*
0.42	0.148601	0.1499490	1.101295	1.88	0.627203	0.7794640	1.249271*
0.44	0.160750	0.1621980	1.095188	1.90	0.630282	0.7830800	1.249524
0.46	0.173063	0.1746450	1.089829	1.92	0.633270	0.7866170	1.249030
0.48	0.185501	0.1872560	1.089829	1.94	0.636023	0.7900750	1.248597
0.50	0.198025	0.2000000	1.081075	1.96	0.638818	0.7934570	1.248510
0.52	0.210596	0.2128460	1.077567	1.98	0.641513	0.7967640	1.248225
0.54	0.223178	0.2257660	1.074587	2.00	0.644105	0.8000000	1.248025
0.56	0.235734	0.2387330	1.072101	2.02	0.646627	0.8031650	1.247916
0.58	0.248229	0.2517210	1.070088	2.04	0.649135	0.8062620	1.247838
0.60	0.260629	0.2647060	1.068530	2.06	0.651631	0.8092910	1.247684
0.62	0.272902	0.2776650	1.067414	2.08	0.654117	0.8122560	1.247454
0.64	0.285016	0.2905790	1.066728	2.10	0.656593	0.8151570	1.247148
0.66	0.296941	0.3034270	1.066468	2.12	0.659049	0.8179970	1.246769
0.68	0.308649	0.3161930	1.066629	2.14	0.661504	0.8207760	1.246340
0.70	0.320112	0.3288590	1.067208	2.16	0.663951	0.8234960	1.245826
0.72	0.331304	0.3414120	1.068208	2.18	0.666366	0.8261600	1.245246
0.74	0.342202	0.3538380	1.069631	2.20	0.668777	0.8287670	1.244646
0.76	0.352783	0.3661260	1.071477	2.22	0.671183	0.8313200	1.243976
0.78	0.363025	0.3782640	1.073751	2.24	0.673583	0.8338210	1.243241
0.80	0.372909	0.3902440	1.076461	2.26	0.675973	0.8362700	1.242446
0.82	0.382417	0.4020570	1.079611	2.28	0.678326	0.8386680	1.241601
0.84	0.391533	0.4136960	1.083209	2.30	0.680668	0.8410170	1.240757
0.86	0.400241	0.4251550	1.087262	2.32	0.683001	0.8433190	1.239869
0.88	0.408528	0.4364290	1.091780	2.34	0.685323	0.8455740	1.238935
0.90	0.416382	0.4475140	1.096772	2.36	0.687790	0.8477840	1.237960
0.92	0.423792	0.4584060	1.102249	2.38	0.690089	0.8499490	1.236667
0.94	0.430750	0.4691020	1.108223	2.40	0.692376	0.8520710	1.235622
0.96	0.437248	0.4796000	1.114702	2.42	0.694648	0.8541510	1.234543
0.98	0.443278	0.4899000	1.121700	2.44	0.696905	0.8561900	1.233440
1.00	0.448838	0.5000000	1.129234	2.46	0.699132	0.8581880	1.232312
1.02	0.453812	0.5099000	1.137312	2.48	0.701331	0.8601480	1.231188
1.04	0.458418	0.5196000	1.146231	2.50	0.703513	0.8620690	1.230067
1.06	0.462554	0.5291020	1.146231	2.52	0.705677	0.8639530	1.228929
1.08	0.466554	0.5384050	1.165242	2.54	0.707823	0.8658010	1.227778
1.10	0.470593	0.5475110	1.174781	2.56	0.709949	0.8676130	1.226615
1.12	0.474670	0.5564230	1.183645	2.58	0.712055	0.8693900	1.225444
1.14	0.478781	0.5651420	1.191855	2.60	0.714139	0.8711340	1.224268
1.16	0.482925	0.5736700	1.199441	2.62	0.716201	0.8728450	1.223090
1.18	0.487098	0.5820100	1.206425	2.64	0.718239	0.8745230	1.221911
1.20	0.491297	0.5901640	1.212837	2.66	0.720253	0.8761700	1.220736

1.22	0.495520	0.5981350	1.218702	2.68	0.722242	0.8777860	1.219566
1.24	0.499764	0.6059270	1.224045	2.70	0.724205	0.8793730	1.218403
1.26	0.504025	0.6135410	1.228892	2.72	0.726140	0.8809300	1.217250
1.28	0.508302	0.6209820	1.233270	2.74	0.727877	0.8824580	1.216110
1.30	0.512591	0.6282530	1.237200	2.76	0.729754	0.8839580	1.215269
1.32	0.516888	0.6353560	1.240709	2.78	0.731601	0.8854310	1.214161
1.34	0.521190	0.6422950	1.243823	2.80	0.733416	0.8868780	1.213072
1.36	0.525495	0.6490740	1.246564	2.82	0.735200	0.8882980	1.212005
1.38	0.529920	0.6556950	1.248954				

Table 3.2

τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(\nu) \le \mu_i$	τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(\nu) \le \mu_i$
2.84	0.736950	0.8896930	1.210961	3.82	0.805000	0.9358660	1.164876
2.86	0.738666	0.8910630	1.209943	3.84	0.806078	0.9364900	1.164065
2.88	0.740348	0.8924080	1.208953	3.86	0.807147	0.9371050	1.163271
2.90	0.741997	0.8937300	1.207991	3.88	0.808212	0.9377120	1.162481
2.92	0.743620	0.8950290	1.207056	3.90	0.809272	0.9383100	1.161688
2.94	0.745226	0.8963040	1.206137	3.92	0.810324	0.9388990	1.160895
2.96	0.746827	0.8975580	1.205219	3.94	0.811370	0.9394800	1.160105
2.98	0.748424	0.898790	1.204284	3.96	0.812410	0.9400540	1.159315
3.00	0.750012	0.900000	1.203331	3.98	0.813446	0.9406190	1.158526
3.02	0.751572	0.9011900	1.202369	4.00	0.814475	0.9411760	1.157736
3.04	0.753121	0.9023590	1.201428	4.02	0.815499	0.9417260	1.156947
3.06	0.754665	0.9035080	1.200483	4.04	0.816517	0.9422690	1.156159
3.08	0.756203	0.9046380	1.199523	4.06	0.817526	0.9428040	1.155372
3.10	0.757734	0.9057490	1.198552	4.08	0.818521	0.9433310	1.154591
3.12	0.759259	0.9068420	1.197571	4.10	0.819504	0.9438520	1.153823
3.14	0.760769	0.9079160	1.196580	4.12	0.820479	0.9443650	1.153066
3.16	0.762254	0.9089720	1.195592	4.14	0.821449	0.9448720	1.152312
3.18	0.763728	0.9100100	1.194625	4.16	0.822411	0.9453720	1.151560
3.20	0.765195	0.9110320	1.193656	4.18	0.823366	0.9458650	1.150812
3.22	0.766654	0.9120370	1.192681	4.20	0.824314	0.9463520	1.150068
3.24	0.768105	0.9130250	1.191699	4.22	0.825257	0.9468320	1.149328
3.26	0.769548	0.9139980	1.190714	4.24	0.826193	0.9473060	1.148588
3.28	0.770981	0.9149540	1.189723	4.26	0.827123	0.9477740	1.147853
3.30	0.772406	0.9158960	1.188732	4.28	0.828047	0.9482360	1.147120
3.32	0.773821	0.9168220	1.187738	4.30	0.828964	0.9486920	1.146390
3.34	0.775226	0.9177330	1.186743	4.32	0.829874	0.9491420	1.145664
3.36	0.776621	0.9186300	1.185749	4.34	0.830777	0.9495860	1.144943
3.38	0.778000	0.9195130	1.184755	4.36	0.831672	0.9500240	1.144225
3.40	0.779360	0.9203820	1.183771	4.38	0.832561	0.9504570	1.143514
3.42	0.780704	0.9212380	1.182802	4.40	0.833442	0.9508840	1.142806
3.44	0.782038	0.922080	1.181844	4.42	0.834316	0.9513060	1.142104
3.46	0.783360	0.9229080	1.180888	4.44	0.835182	0.9517230	1.141406
3.48	0.784672	0.9237250	1.179936	4.46	0.836040	0.9521340	1.140715
3.50	0.785973	0.9245280	1.178987	4.48	0.836890	0.9525400	1.140029
3.52	0.787261	0.9253200	1.178041	4.50	0.837732	0.9529410	1.139350
3.54	0.788538	0.9260990	1.177103	4.52	0.838566	0.9533370	1.138678
3.56	0.789801	0.9268660	1.176170	4.54	0.839391	0.9537290	1.138011
3.58	0.791053	0.9276220	1.175245	4.56	0.840208	0.9541150	1.137353
3.60	0.792290	0.9283670	1.174326	4.58	0.841018	0.9544970	1.136701

3.62	0.793515	0.9291000	1.173417	4.60	0.841817	0.9548740	1.136054
3.64	0.794725	0.9298230	1.172516	4.62	0.842606	0.9552460	1.135418
3.66	0.795921	0.9305340	1.171626	4.64	0.843388	0.9556140	1.134791
3.68	0.797103	0.9312360	1.170746	4.66	0.844164	0.9559770	1.134169
3.70	0.798270	0.9319260	1.169876	4.68	0.844936	0.9563360	1.133551
3.72	0.799421	0.9326070	1.169018	4.70	0.845705	0.9566910	1.132935
3.74	0.800557	0.9332780	1.168173	4.72	0.846470	0.9570420	1.132319
3.76	0.801679	0.9339390	1.167341	4.74	0.847231	0.9573880	1.131705
3.78	0.802795	0.9345910	1.166520	4.76	0.847986	0.9577300	1.131092
3.80	0.803904	0.9352330	1.165697	4.78	0.848735	0.9580680	1.130483

where $\mu_i = \dfrac{\rho_L(\tau_i,1)}{\rho_N(\tau_{i-1},1)-\delta}$, and $\delta = 0.0005$ for all nummbers except those numbers with a "*" are calculated by $\delta = 0.0001$.

Table 3.3

τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(v) \le \mu_i$	τ_i	ρ_N (≥)	ρ_L	$\max_{[\tau_{i-1},\tau_i]} \mu(v) \le \mu_i$
	continue				1.625	0.581700	0.7253220
1.385	0.531473	0.6573260	1.239541	1.630	0.582696	0.7265440	1.249217
1.390	0.532547	0.6589480	1.240085	1.635	0.583689	0.7277600	1.249167
1.395	0.533621	0.6605600	1.240611	1.640	0.584679	0.7289680	1.249112
1.400	0.534695	0.6621620	1.241117	1.645	0.585666	0.7301690	1.249051
1.405	0.535768	0.6637550	1.241604	1.650	0.586650	0.7313630	1.248985
1.410	0.536840	0.6653390	1.242074	1.655	0.587631	0.7325510	1.248914
1.415	0.537912	0.6669140	1.242527	1.660	0.588608	0.7337310	1.248838
1.420	0.538984	0.6684790	1.242960	1.665	0.589582	0.7349040	1.248759
1.425	0.540054	0.6700350	1.243375	1.670	0.590553	0.7360710	1.248675
1.430	0.541124	0.67158200	1.243776	1.675	0.591520	0.7372310	1.248586
1.435	0.542193	0.6731200	1.244159	1.680	0.592484	0.7383840	1.248494
1.440	0.543262	0.6746490	1.244526	1.685	0.593444	0.7395310	1.248398
1.445	0.544329	0.6761680	1.244874	1.690	0.594401	0.7406710	1.248299
1.450	0.545395	0.6776790	1.245210	1.695	0.595355	0.7418040	1.248196
1.455	0.546461	0.67918100	1.245530	1.700	0.596304	0.7429310	1.248088
1.460	0.547525	0.68067400	1.245833	1.705	0.597250	0.7440510	1.247980
1.465	0.548588	0.68215900	1.246123	1.710	0.598192	0.7451650	1.247868
1.470	0.549650	0.68363400	1.246398	1.715	0.599130	0.7462720	1.247754
1.475	0.550711	0.6851010	1.246659	1.720	0.600065	0.7473730	1.247638
1.480	0.551771	0.6865600	1.246905	1.725	0.600995	0.7484670	1.247518
1.485	0.552829	0.6880090	1.247137	1.730	0.601922	0.7495550	1.247398
1.490	0.553886	0.6894510	1.247357	1.735	0.602844	0.7506370	1.247275
1.495	0.554941	0.6908830	1.247564	1.740	0.603762	0.7517130	1.247152
1.500	0.555995	0.6923080	1.247759	1.745	0.604676	0.7527830	1.247027
1.505	0.557048	0.69372400	1.247940	1.750	0.605586	0.7538460	1.246901
1.510	0.558098	0.69513100	1.248108	1.755	0.606492	0.7549030	1.246773
1.515	0.559148	0.6965310	1.248267	1.760	0.607394	0.7559550	1.246644
1.520	0.560195	0.69792200	1.248411	1.765	0.608291	0.7570000	1.246513
1.525	0.561241	0.69930500	1.248546	1.770	0.609183	0.7580390	1.246383
1.530	0.562285	0.70067900	1.248669	1.775	0.610072	0.7590720	1.246254
1.535	0.563327	0.70204600	1.248781	1.780	0.610955	0.7601000	1.246122
1.540	0.564367	0.70340500	1.248884	1.785	0.611835	0.7611210	1.245993
1.545	0.565405	0.70475600	1.248975	1.790	0.612709	0.7621370	1.245861
1.550	0.566441	0.70609800	1.249058	1.795	0.613579	0.7631470	1.245732

1.555	0.567475	0.7074330	1.249130	1.800	0.614444	0.7641510	1.245602
1.560	0.568507	0.70876100	1.249192	1.805	0.615305	0.7651490	1.245474
1.565	0.569536	0.71008000	1.249245	1.810	0.616160	0.7661420	1.245344
1.570	0.570564	0.71139100	1.249291	1.815	0.617011	0.7671290	1.245218
1.575	0.571589	0.71269500	1.249326	1.820	0.617857	0.7681110	1.245091
1.580	0.572611	0.71399200	1.249353	1.825	0.618698	0.7690860	1.244966
1.585	0.573632	0.71528000	1.249374	1.830	0.619534	0.7700570	1.244842
1.590	0.574649	0.71656100	1.249383	1.835	0.620365	0.7710220	1.244720
1.595	0.575665	0.71783500	1.249389	1.840	0.621191	0.7719810	1.244599
1.600	0.576677	0.71910100	1.249383	1.845	0.622011	0.7729350	1.244479
1.605	0.577687	0.72036000	1.249373	1.850	0.622827	0.7738840	1.244364
1.610	0.578695	0.72161100	1.249355	1.855	0.623637	0.7748270	1.244248
1.615	0.579699	0.72285500	1.249329	1.860	0.624442	0.7757650	1.244136
1.620	0.580701	0.72409200	1.249299				

where $\mu_i = \dfrac{\rho_L(\tau_i,1)}{\rho_N(\tau_{i-1},1) - 0.0001}$.

Proof of Theorem 5.

Note that $\sum\rho_L(\tau_i,\sigma) < \infty$ iff $\sum\rho_T(\tau_i,\sigma) < \infty$ iff $\sum\tau_i^2 < \infty$ iff $q > 1/2$. Define the measure

$$M_\sigma[v_0,v_1] = \frac{\sum\rho_L(\tau_i,\sigma)I_{\{v_0\le\tau_i/\sigma\le v_1\}}}{\sum\rho_L(\tau_i,\sigma)}.$$

As $q > 1/2$, M_σ is a probability measure. Now if $\psi(v)$ is any function,

$$\frac{\sum\psi(\frac{\tau_i}{\sigma})\rho_L(\tau_i,\sigma)}{\sum\rho_L(\tau_i,\sigma)} = \int\psi(v)dM_\sigma(v).$$

Therefore, putting $\psi(v) = \mu(v)^{-1}$, the theorem is equivalent to $\int\psi(v)dM_\sigma(v) \to \int\psi(v)g_q(v)dv$. As $\mu(v)$ is bounded and continuous ((3.2)-(3.3)), this will follow if we can show that M_σ converges weakly to g_q, i.e.

$$M_\sigma[v_0,v_1] \to \int_{v_0}^{v_1} g_q(v)dv \tag{12.4}$$

for $0\le v_0\le v_1\le\infty$. Now define the measure

$$N_\sigma[v_0,v_1] = \sigma^{1/q}\,\#\{i: v_0\le\tau_i/\sigma\le v_1\}$$

From the definition of N_σ and τ_i, we have

$$N_\sigma[v_0,v_1] = \sigma^{1/q}\left[(\frac{\sigma v_0}{c})^{-1/q} - (\frac{\sigma v_1}{c})^{-1/q} + R_\sigma(v_0,v_1)\right]$$

where $|R_\sigma| \le 2$. Hence, if $0<v_0<v_1<\infty$, we have as $\sigma\to 0$ that

$$N_\sigma[v_0,v_1] \rightarrow H_q[v_0,v_1] \tag{12.5}$$

where $H_q[v_0,v_1] \equiv (\frac{v_0}{c})^{-1/q} - (\frac{v_1}{c})^{-1/q}$. Let $h_q(v) = c^{1/q}/q\, v^{-(1+1/q)}$ be the density of the measure H_q.

Now as $\rho_L(v,1)$ is continuous and bounded, (12.5) implies that for $\varepsilon > 0$,

$$\int_\varepsilon^\infty \rho_L(v,1)dN_\sigma(v) \rightarrow \int_\varepsilon^\infty \rho_L(v,1)h_q(v)dv. \tag{12.6}$$

Moreover, as lemma 7.2 (following) shows, for small σ and ε,

$$\int_0^\varepsilon \rho_L(v,1)dN_\sigma(v) \leq C\,\varepsilon^{2-\frac{1}{q}} \cdot \int_0^\infty \rho_L(v,1)N_\sigma(dv) \tag{12.7}$$

It follows from (12.6), (12.7), and Fatou's lemma that

$$\int_0^\infty \rho_L(v,1)dN_\sigma(v) \rightarrow \int_0^\infty \rho_L(v,1)h_q(v)dv. \tag{12.8}$$

Using (12.6) and (12.8) we then have

$$M_\sigma[v_0,v_1] = \frac{\int_{v_0}^{v_1} \rho_L(v,1)dN_\sigma(v)}{\int_0^\infty \rho_L(v,1)dN_\sigma(v)} \rightarrow \frac{\int_{v_0}^{v_1} \rho_L(v,1)h_q(v)dv}{\int_0^\infty \rho_L(v,1)h_q(v)dv} = \int_{v_0}^{v_1} g_q(v)dv$$

which establishes (12.4) and completes the proof. □

Lemma 11.2. *For all sufficiently small ε_0 and σ_0, there exists $C(\varepsilon_0,\sigma_0)$ so that (12.7) holds for all $\varepsilon < \varepsilon_0$ and all $\sigma < \sigma_0$.*

Proof.

$$\int_0^\varepsilon \rho_L(v,1)dN_\sigma(v) \leq \sum \tau_i^2 I_{\{\tau_i \leq \varepsilon\sigma\}} = \sum_i c^2 i^{-2q} I_{\{i \geq (\frac{\varepsilon\sigma}{c})^{-1/q}\}}$$

$$\leq \int_{(\frac{\varepsilon\sigma}{c})^{-1/q}}^\infty c^2(x-1)^{-2q}\,dx = \frac{c^2}{2q-1}\left[(\frac{\varepsilon\sigma}{c})^{-1/q} - 1\right]^{1-2q}$$

$$\int_0^\infty \rho_L(v,1)dN_\sigma(v) \geq \frac{1}{2}\int_0^1 \rho_L(v,1)dN_\sigma(v) = \frac{1}{2}\sum_i c^2 i^{-2q} I_{\{i \geq (\frac{\sigma}{c})^{-1/q}\}}$$

$$\geq \int_{(\frac{\sigma}{c})^{-1/q}+1}^\infty c^2(x+1)^{-2q}\,dx = \frac{1}{2}\frac{c^2}{2q-1}\left[(\frac{\sigma}{c})^{-1/q} + 2\right]^{1-2q}.$$

The ratio of the two terms is less than

$$2\,\varepsilon^{2-1/q}\;\frac{(1-(\frac{\varepsilon\sigma}{c})^{1/q})^{1-2q}}{(1+2(\frac{\sigma}{c})^{1/q})^{1-2q}}$$

so that (12.7) holds with

$$C \;=\; 2\frac{(1+2(\frac{\sigma_0}{c})^{1/q})^{2q-1}}{(1-(\frac{\varepsilon_0\sigma_0}{c})^{1/q})^{2q-1}}. \qquad \square$$

Proof of (8.5). Suppose first that τ has a finite number of nonzero coefficients. Then

$$\begin{aligned}\|(C-I)\theta\|^2 &= \sum_i\Big[(c_{ii}-1)\theta_i+\sum_{j\neq i}c_{ij}\theta_j\Big]^2\\ &= \sum_i(c_{ii}-1)^2\theta_i^2+2\sum_i\sum_{j\neq i}(c_{ii}-1)\,\theta_i\;c_{ij}\;\theta_j+\sum_i(\sum_{j!=i}c_{ij}\;\theta_j)^2\\ &\geq \|(Diag(C)-I)\theta\|^2+2\sum_i\sum_{j\neq i}(c_{ii}-1)\,\theta_i\;c_{ij}\;\theta_j \qquad (12.9)\end{aligned}$$

Let now s_i be an i.i.d. sequence of ± 1 gotten by tossing a fair coin. Let $\theta_i = s_i\tau_i$. Let E denote expectation with respect to coin-tossing measure. As E is linear, and all sums are finite,

$$E\left[\sum_i\sum_{j\neq i}(c_{ii}-1)\;\theta_i\;c_{ij}\;\theta_j\right]=\sum_i\sum_{j\neq i}(c_{ii}-1)\;c_{ij}\;\tau_i\;\tau_j\;E[s_i\;s_j]$$

As s_i and s_j are independent, zero mean random variables under coin tossing measure, $Es_is_j=0$. It follows that there exists θ of the form $(\theta_i)=(\pm\tau_i)$ which makes the last term in (12.9) nonnnegative. (8.5) follows.

The case of general τ follows by approximation. $\square$

Proof of (9.7)

$$E(\hat\theta_i-\theta_i)^2 \;=\; \theta_i^2P(\hat\theta_i=0)+E\{(\hat\theta_i-\theta_i)^2\mid\hat\theta_i\neq 0\}P(\hat\theta_i\neq 0).$$

Now from $(x-y)^2\leq 2(x^2+y^2)$

$$E\{(\hat\theta_i-\theta_i)^2\mid\hat\theta_i\neq 0\}\;\leq\;2\theta_i^2+2E\{\hat\theta_i^2\mid\hat\theta_i\neq 0\}$$

But, as $\hat\theta_i^2\leq y_i^2$ and $i\in S$

$$\begin{aligned}E\{\hat\theta_i^2\mid\hat\theta_i\neq 0\}P(\hat\theta_i\neq 0) &\leq E\{y_i^2\mid y_i^2>C^2\sigma^2\}P\{y_i^2>C^2\sigma^2\}\\ &\leq 2\int_{C-\varepsilon}^{\infty}(\theta_i+\sigma z)^2\phi(z)dz\end{aligned}$$

with ϕ the density of $N(0,1)$. Using $(x+y)^2\leq 2(x^2+y^2)$, we get

$$\leq 4\theta_i^2 \int_{C-\varepsilon}^{\infty} \phi(z)dz + 4\sigma^2 \int_{C-\varepsilon}^{\infty} z^2\phi(z)dz$$

$$= 4\theta_i^2(1-\Phi(C-\varepsilon)) + 4\sigma^2\{(C-\varepsilon)\phi(C-\varepsilon)+1-\Phi(C-\varepsilon)\}$$

Applying $1-\Phi(a) \leq \frac{1}{a}\phi(a)$ (Mills' Ratio) and putting the pieces together gives (9.7). □

Proof of Theorem 13.

We prove only the special case where all $a_i>0$. Define new variables w_i via $w_i = a_i\,\tau_i^p$. In terms of these variables, the problem of finding the hardest rectangle is to Maximize

$$J(w) = \sum_i \min(w_i^{2/p}/a_i^{2/p}, \sigma^2)$$

subject to the constraints (C1) each $w_i \geq 0$, and (C2) $\sum_i w_i \leq 1$. As J is monotone increasing in each w_i, a maximum exists satisfying (C3) $\sum_i w_i = 1$. Moreover, as J is constant in w_i as soon as $w_i^{2/p}$ is largerthan $\sigma^2 a_i^{2/p}$, it follows that a maximum exists satisfying (C4) each $w_i \leq \sigma^p a_i$. Let **W** denote the set of w satisfying the constraints (C1), (C3), and (C4). A maximum of J with respect to the original constraints (C1)-(C2) exists in the special set **W**, and **W** is convex.

The restriction of J to **W** is just $\sum_i w_i^{2/p}/a_i^{2/p}$ -- this functional is convex, as $p \leq 2$, and strictly convex if $p<2$. Any member of **W** may be expressed as a mixture of extreme points, and by convexity of J, the value of J at any member is less than the maximum value of J at some extreme point occurring in this representation. It follows that the desired maximum value of J is the maximum over extreme points.

An extreme point of **W** can be characterized as follows. First, the coordinates sum to 1. Second, in all but one coordinate, the coordinate value is either the minimum or the maximum value allowed for that coordinate. In the remaining coordinate, the value is determined by the condition that the coordinate sum be 1. Let now an extreme point w be given, and let **i** be the indices of the coordinates taking on their maximum possible values under (C4). The value of J at w is bounded by

$$\sum_{i:w_i \neq 0} (\textit{ maximum allowed value for coordinate } i\ )^{2/p}/a_i^{2/p} = (Card(\mathbf{i})+1)\,\sigma^2 \qquad (12.10)$$

We now interpret (C4) in terms of the original τ-variables. Given an extreme point w, define τ by $\tau_i = (w_i/a_i)^{1/p}$. The condition that w satisfy (C1) and (C2) implies that the corresponding point τ is in the positive orthant of Θ; as we have argued before, orthosymmetry implies that $\Theta(\tau) \subset \Theta$.The extreme point w has the property that $w_i = (\sigma^2 a_i)^{p/2}$ for $i \in \mathbf{i}$. This is completely equivalent to saying $\tau_i^2 = \sigma^2$ for $i \in \mathbf{i}$. The rectangle $\Theta(\tau)$ therefore contains the cube $\Theta_n(\sigma,\mathbf{i})$ $(n = Card(\mathbf{i}))$. Hence $\Theta_n(\sigma,\mathbf{i}) \subset \Theta$, and so $Card(\mathbf{i}) \le n_0(\sigma)$. Hence (12.10) implies inequality (9.2). (9.1) is immediate. □

References.

Bickel, P. J. (1981) Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Stat.* **9,** 1301-1309.

Bickel, P.J. (1983) Minimax estimation of the mean of a normal distribution subject to doing well at a point. *Recent Advances in Statistics.* 511-528. Academic Press.

Bentkus, R.J. and Kazbaras, A.R. (1981). On optimal statistical estimators of a distribution density. *Doklady AN SSSR* **258** 1300-1302 (in Russian); *Soviet Math. Doklady* **23,** 487-490 (in English).

Bentkus, R.J. and Sushinskas, J.V. (1982) On optimal statistical estimates of a spectral density (in Russian). *Doklady AN SSSR* **263**, 782-786 (in Russian); *Soviet Math. Doklady* **25**, 415-419 (in English).

Bentkus, R.J. (1985a) The asymptotics of the minimax mean square risk of statistical estimators of a spectral density in the space L_2. *Doklady AN SSSR* **281**, 11-15. (in Russian); *Soviet Math. Doklady* **31**, 259-263 (in English).

Bentkus, R.J. (1985b) Asymptotics of the minimax mean square risk of statistical estimators of a spectral density in the space L_2. *Litovskii Matematicheskii Sbornik* **25**, 23-42 (in Russian); *Lithuanian Mathematical Journal* **25**, 11-24 (in English).

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

Donoho, D.L. and Liu, R.C. (1988a) Geometrizing Rates of Convergence, III. Technical Report 138, Statistics Department, U.C. Berkeley.

Donoho, D.L. and Liu, R.C. (1988b) Hardest One-Dimensional Subfamilies. Technical Report 178, Statistics Department, U.C. Berkeley.

Efroimovich, S. Y., and Pinsker, M.S. (1981) Estimation of square-integrable [spectral] density based on a sequence of observations. *Problemy Peredachi Informatsii,* **17,** 3, 50-68 (in Russian); *Problems of Information Transmission* (1982) 182-196 (in English).

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 (in English).

Gatsonis, C., MacGibbon, B., and Strawderman, W. (1987) On the estimation of a restricted normal mean. *Statistics and Probability Letters,* **6,** 21-30.

Gill, P., Murray, W., Saunders, M. and Wright, M. (1986) User's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming. Technical Report SOL 86-2. Systems Optimization Laboratory, Stanford University.

Levit, B. Y. (1980) On asymptotic minimax estimates of the second order. *Theory of Prob. and its Applications.* **25,** 552-568.

Ibragimov, I. A. and Hasminskii, R. Z. (1984), Nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory of Probability and its Applications.* **29**, 1-32.

Jakimauskas, G. (1984) Optimal statistical estimates of a periodic function observed in random noise. *Litovskii Matematicheskii Sbornik* **24**, 201-210 (in Russian); *Lithuanian Mathematical Journal* **24,** 93-98 (in English).

Le Cam, L. (1985) *Asymptotic Methods in Statistical Decision Theory.* Springer, Berlin.

Nussbaum, Michael (1985) Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Stat.* **13,** 984-997.

Pinsker, M. S. (1980) Optimal filtering of square integrable signals in Gaussian white noise. *Problemy Peredachi Informatsii* **16, 2,** 52-68 (in Russian); *Problems of Information Transmission* (1980) 120-133 (in English).

Sacks, J. and Strawderman, W. (1982) Improvements on linear minimax estimates. in *Statistical Decision Theory and Related Topics III,* **2** (S. Gupta ed.) Academic, New York.