*The Annals of Statistics*
1987, Vol. 15, No. 1, 432–442

# MINIMAX ESTIMATION OF THE MEAN OF A GENERAL DISTRIBUTION WHEN THE PARAMETER SPACE IS RESTRICTED

BY AVRAHAM A. MELKMAN AND YA'ACOV RITOV

*Ben Gurion University of the Negev and Hebrew University of Jerusalem*

**1. Introduction.** We consider the one-dimensional additive model $Y = \vartheta + X$. If $X$ is a (standard) normal random variable and $\vartheta$ is completely unknown then of course $\delta(y) = y$ is the minimax estimator. This same estimator is no longer minimax, however, given the added prior information $|\vartheta| \le s$. In fact, the minimax estimate is then Bayes with respect to a least favorable prior distribution that is supported on $[-s, s]$. This distribution was investigated by Casella and Strawderman (1981) for small values of $s$, and by Bickel (1981) and Levit (1980a–c) and Levit and Berhin (1980) for large values of $s$. Our interest was captured particularly by Bickel's somewhat surprising result that if the least favorable distributions are rescaled to $[-1, 1]$ then they converge weakly, as $s \to \infty$, to a distribution with density $\cos^2(\pi x/2)$ [the distribution with minimum Fisher information among all those supported on $[-1, 1]$, see Huber (1974)], and the corresponding minimax risks behave like $1 - \pi^2/s^2 + o(1/s^2)$. Moreover, Bickel produced a family of estimates that have this risk asymptotically, and proved that they have the property that $s(y - \delta(y))$ is approximately $\pi \tan(\pi y/(2s))$.

The main point of this paper is that all the above mentioned results of Bickel (1981) remain valid without the normality assumption. Namely, all that is needed is that the specified distribution of $X$ be such that $EX = 0$, $EX^2 = 1$ and $EX^4 < \infty$. We prove actually a slightly stronger result, Theorem 2, wherein $X$ may be any member of a family of distributions that satisfies a weakened set of requirements. We do not know, however, whether these requirements are necessary, except for the fact that some moment higher than the second has to be bounded.

Suppose $Y_i = \vartheta + X_i$, $i = 1, \dots, n$. The results of this paper can then be used if we replace the vector of observations by a one-dimensional statistic that preserves the translation structure, e.g., the sample mean. If the distribution of $X_1$ is specific enough we may use the best invariant estimator for $\vartheta$, i.e., the Pitman estimator. When this estimator is also a sufficient statistic then the above reduction does not lose any information. If, however, $Y_i = \vartheta_i + X_i$, $i = 1, \dots, n$, then no easy reduction is possible and only partial results are known. Speckman (1982) restricted attention to linear estimates for the vector of means and then found the form of the minimax estimate. Melkman and Micchelli (1979)

Received February 1985; revised May 1986.
*AMS* 1980 *subject classifications*. 62F10, 62C20.
*Key words and phrases*. Minimax, Fisher information.

did not require linearity, a priori, but assumed that the distributions of $X_i$ have compact support.

**2. Main results.** Let $\mathbf{K}$ be some family of distributions on the line such that for all $K \in \mathbf{K}$

(a) $\int xK(dx) = 0$,
(b) $V_K = \int x^2K(dx) \le 1$,
(c) for some $\gamma > 0$ and $\mu_{2+\gamma} > 0$, $\int |x|^{2+\gamma}K(dx) \le \mu_{2+\gamma} < \infty$

and, finally, for some $\mu_4 < \infty$ there is a $K \in \mathbf{K}$ such that $\int x^4K(dx) \le \mu_4$ while

(d) $$\sup\left\{\int x^2K(dx)\colon K \in \mathbf{K}, \int x^4K(dx) \le \mu_4\right\} = 1.$$

Note that $\mathbf{K}$ may consist of a single distribution, i.e., the error distribution is known to have zero expectation, variance 1 and finite fourth moment.

Suppose we observe $Y = \vartheta + X$, where it is known that $\vartheta \in [-s, s]$ and $X \sim K$ for some $K \in \mathbf{K}$ (but unknown otherwise). The risk of an estimate $\delta$ for a specific $\vartheta$ is

$$R(\delta, \vartheta, K) = \int [\vartheta - \delta(\vartheta + x)]^2K(dx).$$

Thus the maximum risk of an estimate $\delta$ is

$$R_s(\delta) = \sup\{R(\delta, \vartheta, K)\colon |\vartheta| \le s,\ K \in \mathbf{K}\}.$$

Our interest is in finding that family of estimates which minimizes $R_s$ asymptotically in $s$. This is an easy task when attention is restricted to linear estimates [see Speckman (1982) for a multidimensional generalization].

PROPOSITION 1. *The estimate* $\delta_s^l(x) = [s^2/(s^2 + 1)]x$ *minimizes* $R_s$ *over all estimators that are linear functions of the observations. Moreover* $R_s(\delta_s^l) = s^2/(s^2 + 1)$.

PROOF. A linear estimator is of the form $\delta(y) = a + by$, hence

$$R(\delta, \vartheta, K) = [a + (b - 1)\vartheta]^2 + b^2V_K,$$
$$R_s(\delta) = [|a| + |b - 1|s]^2 + b^2.$$

Clearly $R_s$ is minimized by setting $a = 0,\ b = s^2/(s^2 + 1)$. □

For these linear estimators $\lim_{s\to\infty} s^2[1 - R_s(\delta_s^l)] = 1$. As we will see this is the appropriate normalization also for the general, nonlinear estimators. Consequently we call a family of estimates, $\{\delta_s\colon s > 0\}$, asymptotically minimax if

$$\lim_{s\to\infty} s^2[1 - R_s(\delta_s)] = \max!.$$

THEOREM 2. *Define the family of estimates* $\{\delta_s\colon s > 0\}$ *as follows: Let* $a_s$, $s > 0$, *be such that* $0 < a_s < 1$, $\lim_{s\to\infty} a_s = 0$, $\lim_{s\to\infty} a_s s^{\gamma/(3\gamma+2)} = \infty$, *where* $\gamma$

*is taken from condition* (c) *of the definition of* **K**. *For* $s > 0$ *define*

$$\psi_s(y) = \begin{cases} (1 - a_s)\pi \tan\left[(1 - a_s)\frac{\pi}{2}y\right], & |y| < 1 + \frac{a_s}{2}, \\ 0, & \textit{otherwise}, \end{cases}$$

*and define*

$$\delta_s(y) = y - \frac{1}{s}\psi_s\left(\frac{y}{s}\right).$$

*This family is asymptotically minimax and moreover*

$$\lim_{s \to \infty} s^2[1 - R_s(\delta_s)] = \pi^2.$$

REMARK. Note that, for simplicity, $\delta_s$ was permitted to exceed $s$, and is therefore inadmissible. A more reasonable estimate would crop $\delta_s$ to $[-s, s]$.

PROOF. We present here the skeleton of the proof, referring the reader interested in the more technical details to the Appendix.

First we show that $\sup s^2[1 - R_s(\delta_s)] = \pi^2$. To that end consider

$$\begin{aligned} R(\delta_s, s\vartheta, K) &= \int\left[-x + \frac{1}{s}\psi_s\left(\vartheta + \frac{x}{s}\right)\right]^2 K(dx) \\ &= V_K - \frac{2}{s}\int x\psi_s\left(\vartheta + \frac{x}{s}\right)K(dx) + \frac{1}{s^2}\int\psi_s^2\left(\vartheta + \frac{x}{s}\right)K(dx) \\ &= V_K - \frac{2}{s^2}\int\psi_s'\left(\vartheta + \frac{x}{s}\right)\bar{k}(x)\,dx \\ &\quad + \frac{1}{s^2}\int\psi_s^2\left(\vartheta + \frac{x}{s}\right)K(dx) + o\left(\frac{1}{s^2}\right), \end{aligned} \tag{1}$$

where the function $\bar{k}(x) = \int_x^\infty tK(dt)$ is positive, unimodal with the mode at zero, and $\int\bar{k}(x)\,dx = V_K$. In the Appendix (1) is justified and the following lemma is proven.

LEMMA 3. *If* $\mu_{2+\gamma} < \infty$, *but not necessarily* $\mu_4 < \infty$, *then*

$$\int\psi_s'\left(\vartheta + \frac{x}{s}\right)\bar{k}(x)\,dx = V_K\psi_s'(\vartheta) + o(1),$$

$$\int\psi_s^2\left(\vartheta + \frac{x}{s}\right)K(dx) = \psi_s^2(\vartheta) + o(1).$$

Therefore we conclude that

$$R(\delta_s, s\vartheta, K) = V_K - \frac{1}{s^2}\left[2V_K\psi_s'(\vartheta) - \psi_s^2(\vartheta)\right] + o\left(\frac{1}{s^2}\right),$$

$$\sup_{|\vartheta| \le 1} R(\delta_s, s\vartheta, K) = V_K - \left[\frac{(1 - a_s)\pi}{s}\right]^2 \frac{V_K - \sin^2(1 - a_s)\frac{\pi}{2}}{\cos^2(1 - a_s)\frac{\pi}{2}} + o\left(\frac{1}{s^2}\right)$$

and finally that

$$R_s(\delta_s) = 1 - \left[\frac{(1-a_s)\pi}{s}\right]^2 + o\left(\frac{1}{s^2}\right).$$

It remains to establish that no other estimator can do better. To that end we consider the Bayesian problem where $\vartheta$ is a random variable with a known distribution,

$$\rho(G, K) = E_{G,K}(\vartheta - \delta_B(Y))^2 = \int\int[\vartheta - \delta_B(\vartheta + x)]^2 G(d\vartheta)K(dx),$$

with $\delta_B$ the Bayes estimate of $\vartheta$ given $y$. We want, of course, a distribution $G$ which is unfavorable as possible. An excellent candidate for the least favorable distribution is one found by Bickel (1981). Thus we will use $G_s(x) = G(x/s)$, where the distribution $G$ has density

(2)
$$g(t) = \begin{cases} \cos^2 \frac{\pi}{2} t, & |t| \leq 1, \\ 0, & |t| > 1. \end{cases}$$

This is the distribution that minimizes the Fisher information among all distributions supported on $[-1, 1]$, Huber (1974). Note that for any estimate $\delta$

$$\begin{aligned} R_s(\delta) &= \sup_{|\vartheta| \leq s,\, K \in \mathbf{K}} \int[\vartheta - \delta(\vartheta + x)]^2 K(dx) \\ &\geq \int[\vartheta - \delta(\vartheta + x)]^2 G_s(d\vartheta)K(dx) \geq \rho(G_s, K), \end{aligned}$$

and hence we need only show that

$$\lim_{s \to \infty} s^2\left(1 - \sup_{K \in \mathbf{K}} \rho(G_s, K)\right) \leq \pi^2.$$

Now from $Y = E(\vartheta|Y) + E(X|Y) = \delta_B(Y) + E(X|Y)$ we get

$$V_K = EE(\Theta - Y)^2|Y = E\left\{E(\Theta - \delta_B(Y))^2|Y + (E(\Theta - Y)|Y)^2\right\}.$$

Hence

$$\rho(G, K) = V_K - E_{G,K}(EX|Y)^2$$

and in particular

$$\rho(G_s, K) = V_K - \int \frac{[\int x g_s(y - x)K(dx)]^2}{\int g_s(y - x)K(dx)}\, dy.$$

REMARK. Here and wherever appropriate the integrand is considered to be zero whenever its numerator is zero.

Integrating the numerator by parts we get

(3)
$$\rho(G_s, K) = V_K - \frac{V_K^2}{s^2} I_s(K),$$

where

$$I_s(K) = \int \frac{\left[\int g'\left(y - \frac{x}{s}\right)\bar{k}(x)\,dx\right]^2}{V_K^2 \int g\left(y - \frac{x}{s}\right)K(dx)}\,dy.$$

The following lemma is proven in the Appendix.

LEMMA 4. *If* $\int x^4 K(dx) \le \mu_4 < \infty$, *then*

$$\lim_{s\to\infty} I_s(K) = \int_{-1}^{1} \frac{[g'(y)]^2}{g(y)}\,dy = \pi^2.$$

Thus

$$\lim_{s\to\infty} s^2\Big(1 - \sup_{K\in\mathbf{K}} \rho(G_s, K)\Big) = \lim_{s\to\infty} \inf_{K\in\mathbf{K}} I_s(K) \le \pi^2.$$

This concludes the proof of the theorem. □

Clearly the unfavorable distribution $G$, defined in (2) is the least favorable in some sense. Certainly if $\{F_s\colon s > 0\}$ is a family of distributions such that $F_s$ is supported on $[-s, s]$ and at least as unfavorable as $G_s$, i.e.,

$$\sup_{K\in\mathbf{K}} \rho(F_s, K) \ge \sup_{K\in\mathbf{K}} \rho(G_s, K),$$

then

$$\lim_{s\to\infty} s^2\Big(1 - \sup_{K\in\mathbf{K}} \rho(F_s, K)\Big) = \pi^2.$$

This is so because for any $\delta$ $R_s(\delta) \ge \rho(F_s, K)$, and in particular for $\delta_s$,

$$\begin{aligned}\pi^2 = \lim_{s\to\infty} s^2(1 - R_s(\delta_s)) &\le \lim_{s\to\infty} s^2\Big(1 - \sup_{K\in\mathbf{K}} \rho(F_s, K)\Big) \\ &\le \lim_{s\to\infty} s^2\Big(1 - \sup_{K\in\mathbf{K}} \rho(G_s, K)\Big) \le \pi^2.\end{aligned} \tag{4}$$

A stronger uniqueness result for $G$ is the following.

PROPOSITION 5. *Let* $\{F_s\colon s > 0\}$ *be a family of distributions such that* $F_s$ *is supported on* $[-s, s]$, *and define* $F_s^*(y) = F_s(sy)$, $|y| \le 1$. *If*

$$\sup_{K\in\mathbf{K}} \rho(F_s, K) \ge \sup_{K\in\mathbf{K}} \rho(G_s, K),$$

*then any weak limit of a subsequence* $\{F_{s_1}^*, F_{s_2}^*, \dots\}$, $s_i \to \infty$, *is equal to* $G$.

The proof of this proposition, given in the Appendix, is a modification (to incorporate smoothing of $F_s$) of the proof of Theorem 2.1 of Bickel (1981).

We close this section with some observations concerning the sharpness of the conditions of Theorem 2.

Comparing the assumptions for Lemmas 3 and 4 we see that condition (d) imposed on the set $\mathbf{K}$ was used only to ensure the existence of a distribution so bad that no estimator can handle it better than $\delta_s$. To show the sharpness of this condition we need therefore to present a distribution with unbounded fourth moment for which there is a better estimator than $\delta_s$.

COUNTEREXAMPLE. Take $K$ to be a distribution with density $k(x) = c/|x|^5$ for $|x| > x_0$. Certainly $\int |x|^\alpha k(x)\,dx < \infty$, $0 \le \alpha < 4$ and $\int |x|^4 k(x)\,dx = \infty$. To prove that for this $K$ the minimax risk in estimating $\vartheta$ is less than $1 - \pi^2/s^2$ we will look at the behavior of $E_{F,K}(EX|Y)^2$ in the expression $\rho(F, K) = V_K - E_{F,K}(EX|Y)^2$, $F \in \mathbf{F}_s$ where $\mathbf{F}_s$ is any family of distributions supported on $[-s, s]$. Now, the proof of Proposition 5 demonstrated that

$$\liminf_{s\to\infty} \inf_{F\in\mathbf{F}_s} s^2 E_{F,K}\{\chi(|Y| \le s)(EX|Y)^2\} \ge \pi^2,$$

where $\chi(\cdot)$ is the indicator function. It remains to prove that

$$\liminf_{s\to\infty} \inf_{F\in\mathbf{F}_s} s^2 E_{F,K}\{\chi(|Y| > s)(EX|Y)^2\} > 0.$$

Given any distribution $F$ the marginal distribution $h$ of $Y$ is such that $h(y) \ge c/|3y|^5$ for $|y| > 2s$. Certainly $|E(X|Y = y)| > s$ for $|y| > 2s$, since $X = Y - \Theta$ and $|\Theta| < s$. Hence

$$s^2 E_{F,K}\{\chi(|Y| > 2s)(EX|Y)^2\} \ge c\frac{s^4}{3^5}\int_{2s}^{\infty}\frac{dy}{y^5} > \frac{1}{4}\frac{c}{3^5 2^4}.$$

Nevertheless, some restriction on moments higher than the second is necessary, as the following proposition shows.

PROPOSITION 6. *Let $\mathbf{M}$ be the family of distributions on the line such that $K \in \mathbf{M}$ if and only if $\int xK(dx) = 0$ and $\int x^2K(dx) = 1$. For this family the linear law $\delta_s^l(x) = [s^2/(1 + s^2)]x$ is minimax and its risk is $s^2/(1 + s^2)$.*

PROOF. Without loss of generality we can restrict ourselves to estimates $\delta$ which are antisymmetric. For fixed $s$ consider the distribution $K_s$ that has a mass of $1/(1 + s^2)$ at $-s$ and a mass of $s^2/(1 + s^2)$ at $1/s$. Clearly $K_s \in \mathbf{M}$, and hence for any estimate $\delta$,

$$R(\delta, s, K_s) = \frac{1}{1+s^2}(\delta(0) - s)^2 + \frac{s^2}{1+s^2}\left[\delta\left(\frac{1+s^2}{s}\right) - s\right]^2 \ge \frac{s^2}{1+s^2}.$$

Proposition 1 implies therefore that here too the linear law is minimax. □

## APPENDIX

Before turning to the proofs of the various lemmas we consider briefly the derivation of formula (1). The difficulty with the integration by parts is that $\psi_s$

is not absolutely continuous. Put differently, the integration is not over the whole real line but rather from $t_1 = -(1+\vartheta)s - \frac{1}{2}sa_s$ to $t_2 = (1-\vartheta)s + \frac{1}{2}sa_s$. Hence to prove (1) we should show that

$$\psi_s\left(1 + \frac{a_s}{2}\right)\left[\int_{t_2}^{\infty} tK(dt) + \int_{-\infty}^{t_1} tK(dt)\right] = o\left(\frac{1}{s^2}\right).$$

But

$$\psi_s\left(1 + \frac{a_s}{2}\right) \le \pi \tan\left[\left(1 - \frac{a_s}{2}\right)\pi\right] \le \frac{\pi}{\sin\frac{\pi}{2}a_s}$$

and

$$\int_{t_2}^{\infty} tK(dt) - \int_{-\infty}^{t_1} tK(dt) \le \int_{|t|>\frac{1}{2}sa_s} |t|K(dt) \le \frac{\mu_{2+\gamma}}{(sa_s)^{1+\gamma}}.$$

Hence

$$\lim_{s\to\infty} \psi_s\left(1 + \frac{a_s}{2}\right)\left[\int_{t_2}^{\infty} tK(dt) + \int_{-\infty}^{t_1} tK(dt)\right] \le \lim_{s\to\infty} \frac{\pi}{\sin\frac{\pi}{2}a_s}\,\frac{\mu_{2+\gamma}}{s^{\gamma}a_s^{1+\gamma}} = 0.$$

Proof of Lemma 3. The rationale of the proof is that for $x$ small compared to $s$, $\psi_s'(\vartheta + x/s)$ is essentially $\psi_s'(\vartheta)$, while for $x$ large $\bar{k}(x)$ is small. Accordingly we break the interval of integration up into the region inside $[-s^{\beta}, s^{\beta}]$, $\beta = 2/(2+3\gamma)$, and the region outside this interval. When $|x| \le s^{\beta}$

$$\left|\psi_s'\left(\vartheta + \frac{x}{s}\right) - \psi_s'(\vartheta)\right| \le \frac{|x|}{s}\operatorname{ess\,sup}\left\{\psi_s''(t)\colon |t-\vartheta| \le \frac{|x|}{s}\right\} \le \frac{c_1}{a_s^3 s^{1-\beta}},$$

and so

$$\sup_{|\vartheta|\le 1}\left|\int_{|x|\le s^{\beta}}\left[\psi_s'\left(\vartheta + \frac{x}{s}\right) - \psi_s'(\vartheta)\right]\bar{k}(x)\,dx\right| \le \frac{c_1}{a_s^3 s^{1-\beta}} V_K.$$

On the other hand for $|x| \ge s^{\beta}$ and $s$ large enough

$$\left|\psi_s'\left(\vartheta + \frac{x}{s}\right)\right| \le \psi_s'\left(1 + \frac{a_s}{2}\right) \le \frac{c_2}{a_s^2},$$

$$\int_{|x|\ge s^{\beta}} \bar{k}(x)\,dx \le \int_{|x|\ge s^{\beta}} x^2K(dx) + s^{\beta}\int_{|x|\ge s^{\beta}}|x|K(dx) \le 2\frac{\mu_{2+\gamma}}{s^{\beta\gamma}}.$$

Hence

$$\left|\int_{|x|\ge s^{\beta}} \psi_s'\left(\vartheta + \frac{x}{s}\right)\bar{k}(x)\,dx\right| \le 2\frac{c_2}{a_s^2}\,\frac{\mu_{2+\gamma}}{s^{\beta\gamma}}$$

and finally

$$\left|\int \psi_s'\left(\vartheta + \frac{x}{s}\right)\bar{k}(x)\,dx - V_K\psi_s'(\vartheta)\right| \le \frac{c_1 V_K}{a_s^3 s^{1-\beta}} + \frac{2c_2\mu_{2+\gamma}}{a_s^2 s^{\beta\gamma}} \xrightarrow[s\to\infty]{} 0.$$

The proof of the second equality is entirely similar. □

Proof of Lemma 4. Denote the integrand in the definition (3) of $I_s$ by $w_s(y, K)$. Since we expect the integrand to approach $[g'(y)]^2/g(y)$, which for $|y| \ge 1$ is $0/0$, we break the interval of $y$-integration into the region inside $(-1,1)$ and the region outside it.

For $y > 1$ revert back to

$$\int g'\left(y - \frac{x}{s}\right)\bar{k}(x)\,dx = s\int xg\left(y - \frac{x}{s}\right)K(dx)$$

whence, using Cauchy–Schwarz,

$$\int_1^\infty \frac{\left[\int g'\left(y - \frac{x}{s}\right)\bar{k}(x)\,dx\right]^2}{\int g\left(y - \frac{x}{s}\right)K(dx)}\,dy \le s^2\int_1^\infty \int x^2 g\left(y - \frac{x}{s}\right)K(dx)\,dy$$

$$= s\int_0^\infty \int x^2 g\left(1 + \frac{z-x}{s}\right)K(dx)\,dz.$$

Now for $z \le x$, $0 \le g(1 + (z-x)/s) \le |\sin(\pi/2(z-x)/s)| \le (\pi/2)(x-z)/s$, whereas for $x \le z$, $g(1 + (z-x)/s) = 0$. Hence

$$s\int_0^\infty \int x^2 g\left(1 + \frac{z-x}{s}\right)K(dx)\,dz \le \frac{\pi}{2}\int_0^\infty \int_z x^2(x-z)K(dx)\,dz$$

$$= \frac{\pi}{2}\int_0^\infty \int_0^x x^2(x-z)\,dz\,K(dx) \le \frac{\pi}{4}\mu_4.$$

Since also

$$sg\left(1 + \frac{z-x}{s}\right) \le \frac{\pi^2}{4}\frac{(z-x)^2}{s} \xrightarrow[s\to\infty]{} 0$$

pointwise, it follows from the dominated convergence theorem that

$$\lim_{s\to\infty}\int_1^\infty w_s(y, K)\,dy = 0.$$

Similarly one proves that

$$\lim_{s\to\infty}\int_{-\infty}^{-1} w_s(y, K)\,dy = 0.$$

Turning our attention now to the interval $[-1,1]$, observe that the problem is to show that $w_s(y, K)$ is bounded there. Indeed, once this is accomplished the dominated convergence theorem assumes that

$$\int_{-1}^{1} w_s(y, K)\, dy \underset{s\to\infty}{\longrightarrow} \int_{-1}^{1} \frac{[g'(y)]^2}{g(y)}\, dy,$$

because both $g$ and $g'$ are continuous and bounded.

Now for $|y| \le \frac{1}{2}$, $w_s(y, K)$ is uniformly bounded since its denominator satisfies

$$\int g\left(y - \frac{x}{s}\right) K(dx) \ge g\left(\frac{3}{4}\right) \int_{s(y-3/4)}^{s(y+3/4)} K(dx) \ge c_1 g\left(\frac{3}{4}\right),$$

for $s \ge s_0$ and $s_0$ such that $\int_{-s_0/4}^{s_0/4} K(dx) > 0$.

When $\frac{1}{2} < |y| \le 1$ a slightly more delicate argument is needed. Choose $s_0$ also such that $\int_0^{s_0/2} xK(dx) > 0$. Since $g(x) \ge (1-x)^2$ for $0 \le x \le 1$,

$$\begin{aligned}\int g\left(y - \frac{x}{s}\right) K(dx) &\ge \int_0^{sy} \left(1 - y + \frac{x}{s}\right)^2 K(dx)\\ &\ge \int_0^{s_0/2} \left(1 - y + \frac{x}{s}\right)^2 K(dx)\\ &\ge \frac{a_0}{s^2} + \frac{a_1}{s}(1-y) + a_2(1-y)^2,\end{aligned}$$

for some $a_0$, $a_1$, $a_2 > 0$ (and independent of $s$).

To bound the numerator use $|g'(x)| \le \pi^2/2(1-x)$, $x \le 1$, so that

$$\begin{aligned}\left|\int g'\left(y - \frac{x}{s}\right) \bar{k}(x)\, dx\right| &\le \frac{\pi^2}{2} \int_{s(y-1)}^{\infty} \left(1 - y + \frac{x}{s}\right) \bar{k}(x)\, dx\\ &\le \frac{\pi^2}{2}\left[(1-y)V_K + \frac{\mu_3}{s}\right],\end{aligned}$$

since $\int |x| \bar{k}(x)\, dx = \frac{1}{2}\mu_3$. Combining these estimates and denoting $t = s(1-y)$

$$w_s(y, K) \le \frac{b_0 + b_1 t + b_2 t^2}{a_0 + a_1 t + a_2 t^2},$$

which clearly is bounded for $t \ge 0$ as $a_0$, $a_1$, $a_2$ are all strictly positive. □

Proof of Proposition 5. Let $H$ be any distribution supported on $[-1, 1]$ with $H' = h$ such that $h'$ and $h''$ exist and are bounded. Let $F_{s_i} \to F^*$. Fix any $\varepsilon > 0$ and denote $\bar{F} = F^* * H(\cdot/\varepsilon)$ and $\bar{F}_i = F_{s_i}^* * H(\cdot/\varepsilon)$. It is easy to see that for all $i = 1, 2, \dots$ and $K \in \mathbf{K}$

$$\rho(F_{s_i}, K) \le \rho\left(\bar{F}_i\left(\frac{\cdot}{s_i}\right), K\right) = V_K + \frac{V_K}{s^2} \bar{I}_i(K),$$

where with $\bar{f}_i = \bar{F}_i'$

$$\bar{I}_i(K) = \frac{1}{V_K^2} \int \frac{\left[\bar{f}_i'\left(y - \frac{x}{s}\right)\bar{k}(x)\, dx\right]^2}{\int \bar{f}_i\left(y - \frac{x}{s}\right) K(dx)}\, dy.$$

Hence the assumption of the proposition implies that

$$\liminf_{i\to\infty} \inf_{K\in\mathbf{K}} \bar{I}_i(K) \le \pi^2. \tag{A.1}$$

We are going to show that (A.1) implies that $F_{s_i} \to_w G$. Now, the families of distributions $K$ and $\{\bar{k};\ K \in \mathbf{K}\}$ are both tight as $\mu_{2+\gamma} < \infty$. Moreover $\sup_t \bar{f}_i^{(j)}(t) \le \varepsilon^{-(j+1)} \sup_t h^{(j)}(t)$, $j = 0,1,2$. Therefore with $\bar{f} = \bar{F}'$ we have for all $y$

$$\begin{aligned}
&\limsup_{i\to\infty} \sup_{K\in\mathbf{K}} \left| \bar{f}(y) - \int \bar{f}_i\left(y - \frac{x}{s_i}\right) K(dx) \right| \\
&\quad \le \limsup_{i\to\infty} |\bar{f}(y) - \bar{f}_i(y)| \\
&\qquad + \limsup_{i\to\infty} \sup_{K\in\mathbf{K}} \int_{|x|<s^{1/2}} \left| \bar{f}_i(y) - \bar{f}\left(y - \frac{x}{s_i}\right) \right| K(dx) \\
&\qquad + \limsup_{i\to\infty} \sup_{K\in\mathbf{K}} \int_{|x|\ge s^{1/2}} \left| \bar{f}_i(y) - \bar{f}\left(y - \frac{x}{s_i}\right) \right| K(dx) = 0.
\end{aligned}$$

In exactly the same way one proves that

$$\sup_{K\in\mathbf{K}} \left| \frac{1}{V_K} \int \bar{f}_i'\left(y - \frac{x}{s_i}\right) \bar{k}(x)\, dx - \bar{f}'(y) \right| \to 0.$$

Using now Fatou's lemma

$$\begin{aligned}
\liminf_{i\to\infty} \inf_{K\in\mathbf{K}} \bar{I}_i(K) &\ge \inf_{K\in\mathbf{K}} \frac{1}{V_K^2} \int_{-(1+\varepsilon)}^{1+\varepsilon} \liminf_{i\to\infty} \frac{\left[ \int \bar{f}_i'\left(y - \frac{x}{s_i}\right) \bar{k}(x)\, dx \right]^2}{\int \bar{f}_i\left(y - \frac{x}{s_i}\right) K(dx)}\, dy \\
&\ge \int_{-(1+\varepsilon)}^{1+\varepsilon} \frac{[\bar{f}'(y)]^2}{\bar{f}(y)}\, dy \ge \frac{\pi^2}{1+\varepsilon}.
\end{aligned} \tag{A.2}$$

Reasoning analogously to the proof of Bickel (1981, Lemma 2.1), we conclude from (A.1) and (A.2) that $F_s^* = G$. □

## REFERENCES


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

CASELLA, G. and W. STRAWDERMAN (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–876.

HUBER, P. J. (1974). Fisher information and spline interpolation. *Ann. Statist.* **2** 1029–1034.

LEVIT, B. YA. (1980a). On the second order asymptotically minimax estimates. *Theory Probab. Appl.* **25** 561–576.

LEVIT, B. YA. (1980b). On the second order optimality in estimation theory. *Theory Probab. Appl.* **25** 640–642.

LEVIT, B. YA. (1980c). On some new results in the theory of second order optimality. *Theory Probab. Appl.* **25** 655–657.

LEVIT, B. YA. and P. E. BERHIN (1980). Second order asymptotically minimax estimates of the mean of a normal distribution. *Problems Inform. Transmission* **16** 60–79.

MELKMAN, A. A. and C. A. MICCHELLI (1979). Optimal estimation of linear operators in Hilbert spaces from inaccurate data. *SIAM J. Numer. Anal.* **16** 87–105.

SPECKMAN, P. (1982). Minimax estimates of linear functionals in a Hilbert space. Unpublished manuscript.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
BEN GURION UNIVERSITY OF THE NEGEV
P.O.B. 653
BEER SHEVA 84120
ISRAEL

DEPARTMENT OF STATISTICS
FACULTY OF SOCIAL SCIENCES
HEBREW UNIVERSITY
MOUNT SCOPUS
JERUSALEM 91905
ISRAEL