Bernoulli **11**(4), 2005, 571–590

Minimax expected measure confidence sets for restricted location parameters

STEVEN N. EVANS[1*], BEN B. HANSEN[2] and PHILIP B. STARK[1**]

[1]*Department of Statistics, University of California, Berkeley CA 94720-3860, USA. E-mail: *evans@stat.berkeley.edu; **stark@stat.berkeley.edu*
[2]*Department of Statistics, University of Michigan, Ann Arbor MI 48109-1092, USA. E-mail: bbh@umich.edu*

We study confidence sets for a parameter $\theta \in \Theta$ that have minimax expected measure among random sets with at least $1 - \alpha$ coverage probability. We characterize the minimax sets using duality, which helps to find confidence sets with small expected measure and to bound improvements in expected measure compared with standard confidence sets. We construct explicit minimax expected length confidence sets for a variety of one-dimensional statistical models, including the bounded normal mean with known and with unknown variance. For the bounded normal mean with unit variance, the minimax expected measure 95% confidence interval has a simple form for $\Theta = [-\tau, \tau]$ with $\tau \leqslant 3.25$. For $\Theta = [-3, 3]$, the maximum expected length of the minimax interval is about 14% less than that of the minimax fixed-length affine confidence interval and about 16% less than that of the truncated conventional interval $[X - 1.96, X + 1.96] \cap [-3, 3]$.

Keywords: Bayes-minimax duality; constrained parameters

1. Introduction

This paper studies how to construct confidence sets that are as small as they can be, in the sense of minimizing worst-case expected measure, while attaining at least their nominal confidence level.

1.1. The bounded normal mean

The *bounded normal mean* (BNM) problem, estimate $\theta \in [-\tau, \tau] \subseteq (-\infty, \infty)$ from the observation $X \sim \mathcal{N}(\theta, 1)$, is a special case. The difficulty of minimax estimation of linear functionals of infinite-dimensional parameters in Gaussian noise is related to the difficulty of estimating a BNM (Donoho and Liu 1991; Donoho 1994; Ibragivmov and Khas'minskii 1985). Estimating a BNM arises in robotics (Kamberova *et al*. 1996, Kamberova and Mintz 1999), and it is of theoretical interest in its own right (see, for example, Bickel 1981; Casella and Strawderman 1981; and references below). Bounded parameters often arise in physical problems, and finding sensible confidence intervals for bounded parameters is an interesting statistical challenge (Mandelkern 2002a, 2002b; Casella 2002; Gleser 2002; Wasserman 2002; van Dyk 2002; Woodroofe and Zhang 2002).

1350–7265

The constraint $\theta \in [-\tau, \tau]$ allows confidence sets for a normal mean to be smaller without sacrificing coverage probability: Consider the conventional 95% confidence set $\mathcal{I}(X) = [X - 1.96, X + 1.96]$ for a normal mean with unit variance. The conventional interval does not exploit the constraint $\theta \in [-\tau, \tau]$. In contrast, the variable-length 'truncated' interval

$$\mathcal{I}_T(X) = [X - 1.96, X + 1.96] \cap [-\tau, \tau] \tag{1}$$

also has 95% coverage probability provided $\theta \in [-\tau, \tau]$, and is shorter than $\mathcal{I}(X)$ for many values of X.

How much can the maximum expected length be reduced? One might optimize the trade-off between coverage and length as a decision problem using a measure of loss that combines the two. Casella *et al.* (1993) show that this can produce intervals with undesirable properties. Zeytinoglu and Mintz (1984; 1988) and Kamberova and Mintz (1999) fix the length of the interval, then find how to centre an interval of that length to maximize the minimum coverage probability for $\theta \in [-\tau, \tau]$. Their results can be used to find $1 - \alpha$ confidence intervals of minimal fixed length; see the Appendix.

We study how much smaller the maximum expected size of a $1 - \alpha$ confidence set can be when the size of the set can depend on the data. Lehmann (1986, p. 524) formulates this problem in general and relates it to accuracy. Minimax equivariant expected measure confidence sets have been constructed for some special cases (Hooper 1982; 1984; Lehmann 1986). Non-equivariant procedures – centred at shrinkage estimators and sometimes of variable size – can improve coverage probability uniformly without increasing expected volume (Brown 1966; Joshi 1967; 1969; Hwang and Casella 1982; Casella and Hwang 1983). We are not aware of previous work finding minimax expected measure, not necessarily equivariant, confidence sets in problems in which there is no uniformly most accurate set.

For inference about a normal mean $\theta \in [-\tau, \tau]$, $\tau \leqslant 2z_{1-\alpha}$, from $X \sim \mathcal{N}(\theta, 1)$, we show that the optimal procedure is the *truncated Pratt interval*

$$\mathcal{I}_{\mathrm{TP}}(X) \equiv \mathcal{I}_P(X) \cap [-\tau, \tau], \tag{2}$$

where $\mathcal{I}_P(X)$ is the Pratt interval (Pratt, 1961)

$$\mathcal{I}_{\mathrm{P}}(X) \equiv \begin{cases} [(x - c), 0 \vee (X + c)], & X \leqslant 0. \\ [0 \wedge (X - c), X + c], & X > 0, \end{cases} \tag{3}$$

with $c = z_{1-\alpha}$. The truncated Pratt interval is minimax for expected length when τ as large as $2z_{1-\alpha}$, nearly twice the value of τ for which the minimax mean square error point estimate has a simple form (Casella and Strawderman 1981). For $\tau \leqslant 2z_{1-\alpha}$, the truncated Pratt interval is minimax for expected Lebesgue measure among more general randomized $1 - \alpha$ confidence sets. For computational techniques to find minimax expected measure or minimax regret confidence sets numerically and for additional theory, see Hansen (2001), Schafer and Stark (2004) and Schafer (2004).

1.2. Improvements in expected length

Table 1 compares the maximum expected length of some 95% confidence intervals for a bounded normal mean. For $\tau = 2$, the maximum expected length of the truncated Pratt interval is 38% less than the length of the conventional interval $\mathcal{I}(X)$, 23% less than that of the affine minimax interval $\mathcal{I}_{\mathrm{A}}(X)$ (Stark 1992), 11% less than the maximum expected length of the truncated conventional interval $\mathcal{I}_T(X)$, and 16% less than the length of the minimax nonlinear fixed-length interval $\mathcal{I}_{\mathrm{N}}(X)$ (see the Appendix).

More generally, in estimating a bounded shift parameter of any shift family with monotone likelihood ratios, the truncated Pratt interval ((3), with c equal to the $1-\alpha$ quantile of the distribution of X when $\theta = 0$) has minimax expected measure provided τ is small.

1.3. Outline

Section 2.1 presents the notation and assumptions. Section 2.2 applies a minimax theorem due to Kneser (1952) and Fan (1953) to establish a Bayes-minimax duality for the expected measure of confidence sets. Section 2.3 uses the Bayes-minimax duality to study minimax expected measure confidence sets for restricted real-valued shift parameters of univariate distributions with monotone likelihood ratios. Section 2.4 extends the theory to situations

Table 1. Maximum expected lengths of several 95% confidence intervals for a bounded normal mean $\theta \in [-\tau, \tau]$. Previously proposed confidence sets for the BNM have maximum expected lengths up to 49% greater than that of the optimal measurable procedure, $\mathcal{I}_{\mathrm{OPT}}$

τ	Conventional $\mathcal{I} = [X \pm 1.96]$		Truncated conventional $\mathcal{I} \cap [-\tau, \tau]$		Best affine fixed-width[a] $\mathcal{I}_{\mathrm{A}}$		Best meas. fixed-width[b] $\mathcal{I}_{\mathrm{N}}$		Opt meas.[c] $\mathcal{I}_{\mathrm{OPT}}$
1.75	3.9	+49%	2.9	+10%	3.4	+28%	3.3	+25%	2.6
2.00	3.9	+38%	3.2	+11%	3.5	+23%	3.3	+16%	2.8
2.25	3.9	+31%	3.4	+13%	3.6	+19%	3.3	+10%	3.0
2.50	3.9	+26%	3.6	+14%	3.6	+17%	3.3	+6%	3.1
2.75	3.9	+22%	3.7	+15%	3.7	+15%	3.3	+3%	3.2
3.00	3.9	+21%	3.8	+16%	3.7	+14%	3.3	+1%	3.3
3.25	3.9	+19%	3.8	+16%	3.7	+14%	3.3	+0%	3.3
3.50	3.9	+18%	3.9	+16%	3.8	+13%	3.5	+5%	3.3[d]
3.75	3.9	+16%	3.9	+15%	3.8	+12%	3.6	+6%	3.4
4.00	3.9	+14%	3.9	+13%	3.8	+10%	3.6	+5%	3.4

[a]Affine fixed-width intervals have the form $[aX + b - e, aX + b + e]$, with a, b, and e constant.
[b]Measurable fixed-width intervals are of form $[\hat{\theta}(X) - e, \hat{\theta}(X) + e]$, with $\hat{\theta}(\cdot)$ measurable and e constant.
[c]General measurable confidence sets have form $\{\theta \in \Theta : (\theta, X) \in S\}$, where $S \subseteq \Theta \times X$ is product-measurable.
[d]The measurable 95% confidence set with smallest expected measure when $\tau \leqslant 3.29$ is the truncated Pratt interval I_{TP}. The entries in the rightmost column for $\tau = 3.50, 3.75,$ and 4.00 are the maximum expected lengths of optimal confidence sets I_{OPT}, approximated numerically.

with nuisance parameters, and studies confidence sets for the BNM where σ^2 is unknown, but for which τ/σ is small. Section 3 contains most of the proofs.

2. Principal results

2.1. Framework, notation and assumptions

Let Θ and $\mathcal{X}$ be measurable spaces. Let ν be a sigma-finite measure on Θ, and let μ be a sigma-finite measure on $\mathcal{X}$. Let $\{\mathbb{P}_\zeta\colon \zeta \in \Theta\}$ be a family of probability distributions on $\mathcal{X}$, absolutely continuous with respect to μ. For $\zeta \in \Theta$, let f_ζ denote the density of $\mathbb{P}_\zeta$ with respect to μ. Let $\mathbb{E}_\zeta$ denote the expectation with respect to $\mathbb{P}_\zeta$. Assume that the mapping $(\zeta, x) \mapsto f_\zeta(x)$ is product-measurable.

We observe an $\mathcal{X}$-valued random variable $X \sim \mathbb{P}_\theta$ (we sometimes write $X \sim f_\theta$) and an auxiliary independent uniform random variable $U \sim U[0, 1]$. The value of θ is unknown, except that $\theta \in \Theta$. We seek a confidence set $S(X, U)$ for θ based on (X, U) that has small ν-measure. (In Section 2.4 we allow θ to consist of two parts, the parameter of interest and a nuisance parameter.)

Let $\mathcal{M}$ be the set of product-measurable mappings of $\Theta \times \mathcal{X}$ to $\mathbb{R}$. Define

$$\mathcal{D} \equiv \{d \in \mathcal{M} : 0 \leqslant d(\zeta, x) \leqslant 1, (\nu \times \mu)\text{-almost surely}\}. \tag{4}$$

Note that $\mathcal{D}$ is a closed, norm-bounded subset of $L_\infty[\nu \times \mu]$, which is the dual of $L_1[\nu \times \mu]$, so, by the Banach–Alaoglu theorem, $\mathcal{D}$ is weak-star compact. Members of $\mathcal{D}$ can be thought of as families of acceptance functions for randomized tests of the hypotheses $\{H_\zeta : X \sim f_\zeta\}$ that are jointly measurable in the parameter value ζ and the datum X: if $U > d(\zeta, X)$, reject H_ζ; otherwise not. The significance level of the test $d(\zeta, \cdot)$ of H_ζ is $1 - \mathbb{E}_\zeta d(\zeta, X)$, the chance that $U > d(\zeta, X)$ when $X \sim f_\zeta$. If $\lambda\colon \zeta \mapsto \lambda_\zeta$ is a measurable function from Θ into $\mathbb{R}$, and if $\eta \in \Theta$,

$$1[f_\zeta(x) > \lambda_\zeta f_\eta(x)] \in \mathcal{D}, \tag{5}$$

so $\mathcal{D}$ includes likelihood ratio tests.

Each $d \in \mathcal{D}$ induces a randomized confidence set $S_d = S_d(X, U)$ for θ, where

$$S_d(x, u) \equiv \{\zeta \in \Theta : u \leqslant d(\zeta, x)\} \tag{6}$$

(Lehmann 1986). Because $d \in \mathcal{D}$, $S_d(x, u)$ is measurable for every $(x, u) \in \mathcal{X} \times [0, 1]$.

The probability that $S_d(X, U)$ correctly covers ζ is

$$\mathbb{C}_\zeta(d) \equiv \mathbb{E}_\zeta d(\zeta, X). \tag{7}$$

The nominal confidence level of S_d is $\inf_{\zeta\in\Theta} \mathbb{C}_\zeta(d)$. However, we shall regard

$$\mathbb{C}_\Theta(d) \equiv \nu\text{-ess}\inf_{\zeta\in\Theta} \mathbb{C}_\zeta(d) \tag{8}$$

as the confidence level of S_d: if $d \in \mathcal{D}$ and $\mathbb{C}_\Theta(d) = \beta$, then there exists $d' \in \mathcal{D}$, $(\nu \times \mu)$-almost everywhere equal to d, with $\inf_{\zeta\in\Theta} \mathbb{C}_\zeta(d') = \beta$, so that $\nu(S_d) = \nu(S_{d'})$ with probability one, whatever the value of θ. The functions $d(\cdot, \cdot)$ in

$$\mathcal{D}_\alpha \equiv \{d \in \mathcal{D} : \mathbb{C}_\Theta(d) \geqslant 1 - \alpha\} \tag{9}$$

are families of decision functions for randomized tests whose inversions are $1 - \alpha$ confidence sets for θ. We call members of $\mathcal{D}_\alpha$ decision functions, families of level-α tests, and $1 - \alpha$ randomized confidence sets (through the association (6)).

For $\theta = \zeta$, the expected ν-measure of the confidence set $S_d(X, U)$ is

$$\mathbb{L}_\zeta(d) = \mathbb{E}_\zeta \int_\Theta d(\eta, X)\nu(\mathrm{d}\eta). \tag{10}$$

The maximum expected ν-measure of S_d over Θ is

$$\mathbb{L}_\Theta(d) \equiv \sup_{\zeta \in \Theta} \mathbb{L}_\zeta(d). \tag{11}$$

Schafer (2004) extends the theory presented here to allow the measure ν to depend on the parameter θ.

We now characterize the decision functions $d \in \mathcal{D}_\alpha$ that minimize $\mathbb{L}_\Theta(d)$.

2.2. Bayes-minimax duality for confidence procedures

Let Π be the set of all probability measures on Θ. For $\pi \in \Pi$, the π-average expected ν-measure of the confidence set corresponding to d is

$$\mathbb{L}_\pi(d) \equiv \int_\Theta \mathbb{L}_\zeta(d)\pi(\mathrm{d}\zeta). \tag{12}$$

Theorem 1. *If $\tilde{\mathcal{D}} \subset \mathcal{D}$ is weak-star compact in $L_\infty[\nu \times \mu]$, then*

$$\inf_{d \in \tilde{\mathcal{D}}} \mathbb{L}_\Theta(d) = \sup_{\pi \in \Pi} \inf_{d \in \tilde{\mathcal{D}}} \mathbb{L}_\pi(d). \tag{13}$$

Theorem 1 is proved in Section 3.1. For $\pi \in \Pi$, define the average density

$$f_\pi(\cdot) \equiv \int_\Theta f_\zeta(\cdot)\pi(\mathrm{d}\zeta). \tag{14}$$

Fix $\alpha \in (0, 1)$ and let $\tilde{\mathcal{D}} \equiv \mathcal{D}_\alpha$. Given $\pi \in \Pi$, let $d^\pi = d^\pi(\zeta, x)$ be a family of decision functions for size-α randomized tests of the hypotheses $\{H_\zeta : X \sim f_\zeta, \zeta \in \Theta\}$ such that for each $\zeta \in \Theta$, the test $d^\pi(\zeta, \cdot)$ is most powerful against the alternative

$$H_\pi : X \sim f_\pi(\cdot). \tag{15}$$

Because each test is of a simple null hypothesis against a simple alternative, d^π is an amalgamation of likelihood ratio tests. For each $\zeta \in \Theta$, let

$$\lambda_\zeta \equiv \inf\left\{\lambda : \int_{f_\pi < \lambda f_\zeta} f_\zeta(x)\mu(\mathrm{d}x) \geqslant 1 - \alpha\right\}. \tag{16}$$

Define

$$d^\pi(\zeta, x) \equiv \begin{cases} 1, & f_\pi(x) < \lambda_\zeta f_\zeta(x), \\ c_\zeta, & f_\pi(x) = \lambda_\zeta f_\zeta(x), \\ 0, & f_\pi(x) > \lambda_\zeta f_\zeta(x), \end{cases} \tag{17}$$

with c_ζ chosen so that $\int d(\zeta, x) f_\zeta(x) \mu(\mathrm{d}x) = 1 - \alpha$. Then $d^\pi \in \mathcal{D}_\alpha$, and d^π minimizes $\mathbb{L}_\pi(\cdot)$ over $\mathcal{D}_\alpha$ – this follows from the optimality of $d^\pi(\zeta, \cdot)$ and the Ghosh–Pratt identity (Ghosh 1961; Pratt 1961, eq. 2); see Section 3.1.

Corollary 1. *For* $0 < \alpha < 1$,

$$\inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = \sup_{\pi \in \Pi} \mathbb{L}_\pi(d^\pi). \tag{18}$$

2.3. Bounded real shift parameters

In this section, we study confidence sets for bounded location parameters of one-dimensional shift families: $\mathcal{X} \subseteq \mathbb{R}$, Θ is a bounded subset of $\mathbb{R}$, and $f_\theta(x) \equiv f(x - \theta)$ for some density f with respect to Lebesgue measure.

Let $d^\eta \equiv d^{\delta_\eta}$ be the decision function most powerful against the alternative $\theta = \eta$. Pratt (1961) showed that the confidence set based on d^η minimizes expected Lebesgue measure when $\theta = \eta$:

$$\mathbb{L}_\eta(d^\eta) = \int_\Theta \mathbb{E}_\eta d^\eta(\zeta, X) \mathrm{d}\zeta. \tag{19}$$

Suppose $\{f_\theta : \theta \in \Theta\}$ has monotone likelihood ratios: $f_{\theta_2}/f_{\theta_1}$ is non-decreasing in x when $\theta_1 < \theta_2$. (The normal, uniform, logistic, and double exponential distributions have monotone likelihood ratios.) Then the acceptance region of the likelihood ratio test of a simple null hypothesis against a simple alternative hypothesis is a semi-infinite interval (Lehmann 1986):

$$d^\eta(\zeta, x) = \begin{cases} 1[x \leqslant \zeta + q_{1-\alpha}], & \zeta < \eta, \\ 1[x \geqslant \zeta + q_\alpha], & \zeta > \eta, \end{cases} \tag{20}$$

where q_β is the β quantile of $\mathbb{P}_0$, the distribution of X when $\theta = 0$.

Pratt (1963) studied the case $\Theta = \mathbb{R}$. When Θ is a bounded subset of $\mathbb{R}$, we call d^η the *truncated Pratt procedure*. We show here that for shift families with monotone likelihood ratios, when τ is sufficiently small there is a point $\eta \in \Theta = [-\tau, \tau]$ such that the truncated Pratt procedure d^η has minimax expected Lebesgue measure among randomized $1 - \alpha$ confidence sets.

Let $F(\cdot) \equiv \int_{-\infty}^{\cdot} f_0(x) \mathrm{d}x$ be the cdf of $\mathbb{P}_0$, and let η be any point in Θ such that

$$F(q_\alpha + \tau) - F(q_\alpha + \eta) = F(q_{1-\alpha} + \eta) - F(q_{1-\alpha} - \tau). \tag{21}$$

If $f_0(\cdot)$ is symmetric about any point, $\eta = 0$.

Theorem 2. *Let* $\Theta = [-\tau, \tau]$, *let* $\{f_\theta\}_{\theta \in \Theta} \equiv \{f_0(\cdot - \theta)\}_{\theta \in \Theta}$ *be a shift family of densities with respect to Lebesgue measure that has monotone likelihood ratios, and let* η *satisfy* (21). *Suppose* $\alpha < \frac{1}{2}$. *If*

$$\tau + |\eta| \leqslant q_{1-\alpha} - q_\alpha, \tag{22}$$

then

$$\inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = \int_{-\tau}^{\eta} F(\zeta + q_{1-\alpha})\mathrm{d}\zeta + \int_{\eta}^{\tau} (1 - F(\zeta + q_\alpha))\mathrm{d}\zeta, \tag{23}$$

and the truncated Pratt procedure d^η (20) *attains the infimum. When* f_0 *is symmetric (so that* $\eta = 0$ *suffices) the truncated Pratt procedure is not optimal if*

$$\tau = \tau + |\eta| > q_{1-\alpha} - q_\alpha = 2q_{1-\alpha}. \tag{24}$$

Corollary 2 (Bounded normal mean). *Let* $X \sim N(\theta, 1)$ *with* $\Theta = [-\tau, \tau]$, *and let* $\alpha < \frac{1}{2}$. *If* $\tau \leqslant 2z_{1-\alpha}$, *then*

$$\inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = 2\int_{0}^{\tau} \Phi(z_{1-\alpha} - \zeta)\mathrm{d}\zeta, \tag{25}$$

and the truncated Pratt procedure d^0 *attains the infimum. If* $\tau > 2z_{1-\alpha}$, *the truncated Pratt procedure is not minimax.*

Table 2 compares the performance of the truncated Pratt confidence interval for the BNM,

$$\mathcal{I}_{\mathrm{TP}}(X) = [(X - z_{1-\alpha}) \wedge 0, (X + z_{1-\alpha}) \vee 0] \cap [-\tau, \tau], \tag{26}$$

to those of the truncated conventional confidence interval $\mathcal{I}_{\mathrm{T}}(X)$ and the minimax affine confidence interval $\mathcal{I}_{\mathrm{A}}$.

2.4. Bounded normal mean with unknown variance: nuisance parameters

In this subsection, we change notation to allow the distribution of the data to depend on two parameters, the parameter $\theta \in \Theta$ of interest, and a nuisance parameter $\sigma \in \Sigma$. We denote this distribution $\mathbb{P}_{(\theta,\sigma)}$ and define the family of distributions

$$\mathbb{P}_{(\Theta,\Sigma)} \equiv \{\mathbb{P}_{(\theta,\sigma)} \colon \theta \in \Theta, \sigma \in \Sigma\}. \tag{27}$$

We assume as before that Θ is a measure space with measure ν, and we seek a confidence set for θ with small expected ν-measure. We assume that the family $\mathbb{P}_{(\Theta,\Sigma)}$ is dominated by a σ-finite measure μ. Let $f_{(\theta,\sigma)}$ be the density of $\mathbb{P}_{(\theta,\sigma)}$ with respect to μ. We also assume that for each fixed $\sigma \in \Sigma$, the mapping $(\theta, x) \mapsto f_{(\theta,\sigma)}(x)$ is product-measurable.

Let $\mathcal{D}$ contain the product-measurable mappings from $\Theta \times \mathcal{X} \to [0, 1]$ as before, but define

$$\mathbb{C}_{(\zeta,\sigma)}(d) \equiv \mathbb{E}_{(\zeta,\sigma)} d(\zeta, X), \tag{28}$$

$$\mathbb{C}_{(\Theta,\sigma)}(d) \equiv \nu\text{-ess}\inf_{\zeta \in \Theta} \mathbb{C}_{(\zeta,\sigma)}(d) \tag{29}$$

and

Table 2. Expected lengths of the truncated Pratt procedure and some others, for small to medium τ. The truncated Pratt dominates alternative procedures for small enough τ, but as τ increases above $2z_{1-\alpha}$, its worst-case behaviour deteriorates sharply

$1-\alpha$	τ	$\mathbb{E}_0$ $\mu(\mathcal{I}_{\mathrm{TP}}(X))$[a]	$\sup_\zeta \mathbb{E}_\zeta$ $\mu(\mathcal{I}_{\mathrm{TP}}(X))$	$\sup_\zeta \mathbb{E}_\zeta$ $\mu(\mathcal{I}_{\mathrm{T}}(X))$[b]	$\sup_\zeta \mathbb{E}_\zeta$ $\mu(\mathcal{I}_{\mathrm{A}}(X))$[c]
0.90	1.25	1.8*	1.8	2.0	2.5
	1.50	2.1*	2.1	2.3	2.7
	1.75	2.2*	2.2	2.6	2.9
	2.00	2.4*	2.4	2.8	2.9
	2.25	2.5*	2.5	3.0	3.0
	2.50	2.6*	2.6	3.1	3.1
	2.75	2.6	2.7	3.2	3.1
	3.00	2.6	3.0	3.2	3.1
	3.25	2.6	3.2	3.2	3.1
	3.50	2.6	3.5	3.3	3.2
0.95	1.75	2.6*	2.6	2.9	3.4
	2.00	2.8*	2.8	3.2	3.5
	2.25	3.0*	3.0	3.4	3.6
	2.50	3.1*	3.1	3.6	3.6
	2.75	3.2*	3.2	3.7	3.7
	3.00	3.3*	3.3	3.8	3.7
	3.25	3.3*	3.3	3.8	3.7
	3.50	3.3	3.5	3.9	3.8
	3.75	3.3	3.7	3.9	3.8
	4.00	3.3	4.0	3.9	3.8
	4.25	3.3	4.2	3.9	3.8
	4.50	3.3	4.5	3.9	3.8
0.99	2.50	4.0*	4.0	4.3	4.7
	2.75	4.2*	4.2	4.5	4.8
	3.00	4.4*	4.4	4.7	4.9
	3.25	4.5*	4.5	4.9	5.0
	3.50	4.5*	4.5	5.0	5.0
	3.75	4.6*	4.6	5.0	5.0
	4.00	4.6*	4.6	5.1	5.0
	4.25	4.6*	4.6	5.1	5.0
	4.50	4.7*	4.7	5.1	5.0
	4.75	4.7	4.8	5.1	5.0
	5.00	4.7	5.0	5.2	5.1
	5.25	4.7	5.3	5.2	5.1
	5.50	4.7	5.5	5.2	5.1

[a] μ denotes Lebesgue measure. $\mathbb{E}_0\mu(I_{\mathrm{TP}}(X))$ is the Bayes risk of d^0 (the Bayes decision rule for a prior that concentrates at zero), or I_{TP}. This is the worst-case risk of d^0 if and only if $\tau \leqslant 2z_{1-\alpha}$.
[b] I_{T}, the truncated conventional interval, is defined in (1).
[c] I_{A}, the minimax affine fixed-length interval, was determined and analysed numerically by the method of Stark (1992).
*I_{TP} is optimal.

$$\mathcal{D}_\alpha = \{d \in \mathcal{D} : \mathbb{C}_{(\Theta,\sigma)}(d) \geqslant 1 - \alpha, \forall \sigma \in \Sigma\}. \tag{30}$$

$\mathcal{D}_\alpha$ contains only decisions corresponding to confidence sets with probability at least $1 - \alpha$ of covering θ, whatever the value of $\theta \in \Theta$ and $\sigma \in \Sigma$. The decision rules in $\mathcal{D}$ do not depend on σ. Define

$$\mathbb{L}_{(\zeta,\sigma)}(d) \equiv \mathbb{E}_{(\zeta,\sigma)} \int_\Theta d(\eta, X)\nu(\mathrm{d}\eta) \tag{31}$$

and

$$\mathbb{L}_{(\Theta,\sigma)}(d) \equiv \sup_{\zeta \in \Theta} \mathbb{L}_{(\zeta,\sigma)}(d). \tag{32}$$

An optimal decision rule $d^* \in \mathcal{D}_\alpha$ would satisfy, for each fixed $\sigma \in \Sigma$,

$$\mathbb{L}_{(\Theta,\sigma)}(d^*) = \inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_{(\Theta,\sigma)}(d). \tag{33}$$

We specialize now to the BNM with unknown variance σ^2. We do not find a decision rule d^* that is optimal for all $\sigma \in \mathbb{R}^+$, but we do show that the truncated Pratt is optimal (among scale-invariant procedures) provided τ is not too large compared with σ.

We observe $X = (X_i)_{i=1}^n$, where $\{X_i\}_{i=1}^n$ are independently and identically distributed as $\mathcal{N}(\theta, \sigma^2)$ with $\theta \in \Theta = [-\tau, \tau]$ but otherwise unknown, and $\sigma \in \Sigma = \mathbb{R}^+$ but otherwise unknown. Let ν be Lebesgue measure on $[-\tau, \tau]$, and let μ be Lebesgue measure on $\mathbb{R}$. We seek a confidence set for θ that has $1 - \alpha$ coverage probability whatever the value of $\theta \in \Theta$ and $\sigma \in \Sigma$, and we want the expected measure of the set to be as small as possible at the worst θ, for each value of σ.

The Rao–Blackwell theorem and the principle of invariance lead us to focus on decision rules that depend on the data through $(\overline{X} - \zeta)/S$, where $\overline{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $S^2 \equiv (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Let $\mathcal{D}_i$ denote the set of such decision functions, and let $\mathcal{D}_{\alpha,i} \equiv \mathcal{D}_i \cap \mathcal{D}_\alpha$. We call $\mathcal{D}_{\alpha,i}$ the *scale-invariant* $1 - \alpha$ *confidence procedures*, even though the set does not contain all scale-invariant procedures. By sufficiency, for each σ it contains one that solves (33).

In general, which scale-invariant procedure is minimax for expected measure depends on σ, but the truncated Pratt procedure is minimax scale-invariant provided τ is not too big compared with σ:

Theorem 3. *Let $\overline{X}$ and S be independent random variables with $\overline{X} \sim \mathcal{N}(\theta, \sigma^2/n)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, for $\theta \in [-\tau, \tau]$ and $\sigma > 0$. Suppose $n \geqslant 3$ and $\alpha \in (\frac{1}{2}, 1)$. Let*

$$d_i^{\mathrm{TP}}(\zeta, (\bar{x}, s)) \equiv \begin{cases} 1[(\bar{x} - \zeta)/(sn^{-1/2}) \leqslant t_{1-\alpha}], & \zeta \leqslant 0, \\ 1[(\bar{x} - \zeta)/(sn^{-1/2}) \geqslant -t_{1-\alpha}], & \zeta > 0, \end{cases} \tag{34}$$

where $t_{1-\alpha}$ is the $1 - \alpha$ quantile of Student's t-distribution with $n - 1$ degrees of freedom. Then:

(i) $d_i^{\mathrm{TP}} \in \mathcal{D}_{\alpha,i}$.
(ii) *If*

$$\frac{\tau}{\sigma} \leqslant 2t_{1-\alpha}\sqrt{\frac{n-2}{n(n-1)}},$$

then $\mathbb{L}_{(\Theta,\sigma)}$ *attains its minimum on* $\mathcal{D}_{\alpha,i}$ *at* d_i^{TP}. *In addition,*

$$\inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\Theta,\sigma)}(d) = \mathbb{L}_{(\Theta,\sigma)}(d_i^{\mathrm{TP}}) = 2\int_0^\tau F_{\zeta\sqrt{n}/\sigma}(t_{1-\alpha})\mathrm{d}\zeta, \tag{35}$$

where F_x *is the cdf of the non-central t-distribution with* $n-1$ *degrees of freedom and non-centrality parameter* x.

Remark. The condition $\tau/\sigma \leqslant 2t_{1-\alpha}\sqrt{(n-2)/(n(n-1))}$ is sufficient, but not necessary, for d_i^{TP} to be minimax among scale-invariant procedures. Numerical experiments suggest that the largest τ/σ for which the result is true is between $2t_{1-\alpha}/\sqrt{n}$ and $2t_{1-\alpha}\sqrt{(n-2)/(n(n-1))}$. A statement similar to the theorem holds for $n=2$ provided $\tau/\sigma \leqslant \sqrt{\frac{1}{2}\ln 2}\,t_{1-\alpha}$.

3. Proofs

3.1. Theorem 1

Lemma 1. *For each* $\pi\in\Pi$, $d\mapsto\mathbb{L}_\pi(d)$ *is a weak-star lower semicontinuous mapping of* $L_\infty[\nu\times\mu]$ *into* $[0,\infty]$.

Proof. Fix $\pi\in\Pi$. Let $\{A_j\}_{j=1}^\infty$ be an increasing nested sequence of measurable subsets of Θ such that $\nu(A_j)<\infty$ and $\cup_j A_j=\Theta$. We have

$$\begin{aligned}\mathbb{L}_\pi(d) &= \int_\Theta\left(\int_{\Theta\times\mathcal{X}} d(\eta,x)\nu(\mathrm{d}\eta)f_\zeta(x)\mu(\mathrm{d}x)\right)\pi(\mathrm{d}\zeta)\\ &= \int_{\Theta\times\mathcal{X}} d(\eta,x)\left(\int_\Theta f_\zeta(x)\pi(\mathrm{d}\zeta)\right)\nu(\mathrm{d}\eta)\mu(\mathrm{d}x)\\ &= \sup_j\int_{\Theta\times\mathcal{X}} d(\eta,x)\left(1_{A_j}(\eta)\int_\Theta f_\zeta(x)\pi(\mathrm{d}\zeta)\right)\nu(\mathrm{d}\eta)\mu(\mathrm{d}x)\end{aligned} \tag{36}$$

by monotone convergence. The term in parentheses in (36) is in $L_1[\nu\times\mu]$, so for each j, the outer integral is a weak-star continuous functional of d. Because $\mathbb{L}_\pi(d)$ is the supremum of a collection of weak-star continuous functionals, it is weak-star lower semicontinuous. □

The next theorem is a special case of general minimax results of Kneser (1952), Fan (1953) and Sion (1958).

Theorem 4. *Let* M *be a convex set and let* $\mathbb{T}: M\times N\to[-\infty,\infty]$ *be linear in* M *and* *convex-like in N, in the sense that for each* $n_0, n_1 \in N$, $\gamma \in (0, 1)$, *there exists* $n_\gamma \in N$ *such that*

$$\gamma \mathbb{T}(m, n_0) + (1-\gamma)\mathbb{T}(m, n_1) \geqslant \mathbb{T}(m, n_\gamma)$$

for all $m \in M$. *If either*

(i) M *is a compact topological space and* $\mathbb{T}(m, n)$ *is upper semicontinuous in* m *for each* n, *or*
(ii) N *is a compact topological space and* $\mathbb{T}(m, n)$ *is lower semicontinuous in* n *for each* m,

then

$$\inf_{n \in N} \sup_{m \in M} \mathbb{T}(m, n) = \sup_{m \in M} \inf_{n \in N} \mathbb{T}(m, n). \tag{37}$$

Proof of Theorem 1. The set $\tilde{\mathcal{D}}$ is weak-star compact by assumption. The map $d \mapsto \mathbb{L}_\pi(d)$ is linear in d for fixed π, and the map $\pi \mapsto \mathbb{L}_\pi(d)$ is linear in π for fixed d. By Lemma 1, $d \mapsto \mathbb{L}_\pi(d)$ is weak-star lower semicontinuous, so Theorem 4 applies:

$$\inf_{d \in \tilde{\mathcal{D}}} \sup_{\pi \in \Pi} \mathbb{L}_\pi(d) = \sup_{\pi \in \Pi} \inf_{d \in \tilde{\mathcal{D}}} \mathbb{L}_\pi(d). \tag{38}$$

For any $d \in \mathcal{D}$ and $c \in \mathbb{R}$, the set $\{\theta \in \Theta : \mathbb{L}_\theta(d) \geqslant c\}$ is measurable, and some $\pi \in \Pi$ concentrates on it provided it is not empty. Therefore,

$$\sup_{\pi \in \Pi} \mathbb{L}_\pi(d) = \mathbb{L}_\Theta(d). \tag{39}$$

□

Lemma 2. *If* $\alpha \in [0, 1]$, *then* $\mathcal{D}_\alpha \subseteq L_\infty[\nu \times \mu]$ *is weak-star compact.*

Proof. $\mathcal{D}_\alpha \subseteq \mathcal{D}$, which is a weak-star compact subset of $L_\infty[\nu \times \mu]$, so it is enough to show that $\mathcal{D}_\alpha$ is closed. Now $d \in \mathcal{D}_\alpha$ if and only if, for ν-almost every $\zeta \in \Theta$,

$$1 - \alpha \leqslant \int_{\mathcal{X}} d(\zeta, x) f_\zeta(x) \mu(\mathrm{d}x). \tag{40}$$

For any measurable set $A \subset \Theta$ with $\nu(A) > 0$, define

$$C_A(d) \equiv \int_\Theta \int_{\mathcal{X}} d(\zeta, x) f_\zeta(x) \frac{1[\zeta \in A]}{\nu(A)} \mu(\mathrm{d}x)\nu(\mathrm{d}\zeta). \tag{41}$$

The function

$$(\zeta, x) \mapsto \frac{1}{\nu(A)} f_\zeta(x) 1[\zeta \in A] \tag{42}$$

is in $L_1[\nu \times \mu]$, so $d \mapsto C_A(d)$ is a weak-star continuous functional of d. Thus for each measurable A with $\nu(A) > 0$, $\{d \in \mathcal{D} : C_A(d) \geqslant 1 - \alpha\}$ is closed. But

$$\mathcal{D}_\alpha = \bigcap_{A:\nu(A)>0} \{d \in \mathcal{D} : C_A(d) \geqslant 1-\alpha\}. \tag{43}$$

$\square$

Proof of Corollary 1 from Theorem 1. By Lemma 2, $\mathcal{D}_\alpha$ is weak-star compact. Therefore,

$$\inf_{d\in\mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = \sup_{\pi\in\Pi} \inf_{d\in\mathcal{D}_\alpha} \mathbb{L}_\pi(d). \tag{44}$$

By construction, $\mathbb{C}_\zeta(d^\pi) = 1-\alpha$ for each $\zeta \in \Theta$, so $d^\pi \in \mathcal{D}_\alpha$. Fix $\pi \in \Pi$. For each $\zeta \in \Theta$, $x \mapsto d^\pi(\zeta, x)$ minimizes $\mathbb{E}_\pi d(\zeta, X)$ among $d(\zeta, \cdot) : \mathcal{X} \to [0, 1]$ satisfying $\mathbb{E}_\zeta d(\zeta, X) \geqslant 1-\alpha$. Therefore d^π minimizes $\int_\Theta \mathbb{E}_\pi d(\zeta, X)\nu(\mathrm{d}\zeta)$ among $d \in \mathcal{D}_\alpha : \mathbb{L}_\pi(d^\pi) = \inf_{d\in\mathcal{D}_\alpha}\mathbb{L}_\pi(d)$. $\square$

3.2. Theorem 2

Because $\{f_\theta : \theta \in \Theta\}$ has monotone likelihood ratios, d^η has the form (20). Without loss of generality, take $c_\zeta \equiv 1$. Let $F(\cdot)$ be the cdf of $\mathbb{P}_0$. The risk at $\theta \in \Theta$ of the decision procedure d^η is

$$\mathbb{L}_\theta(d^\eta) = \int_{-\tau}^{\eta} F(\zeta + q_{1-\alpha} - \theta)\mathrm{d}\zeta + \int_{\eta}^{\tau} (1 - F(\zeta + q_\alpha - \theta))\mathrm{d}\zeta. \tag{45}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{L}_\theta(d^\eta) = \int h(\zeta) f_\theta(\zeta)\mathrm{d}\zeta, \tag{46}$$

where

$$h(\zeta) \equiv 1[q_\alpha + \eta < \zeta \leqslant \tau + q_\alpha] - 1[-\tau + q_{1-\alpha} < \zeta \leqslant \eta + q_{1-\alpha}]. \tag{47}$$

Now $h(\cdot)$ has at most one strict sign change. The restriction $\tau + |\eta| \leqslant q_{1-\alpha} - q_\alpha$ implies that $q_\alpha + \eta \leqslant -\tau + q_{1-\alpha}$. Similarly, $\tau + |\eta| \leqslant q_{1-\alpha} - q_\alpha$ implies that $\tau + q_\alpha \leqslant \eta + q_{1-\alpha}$. Thus if h has a strict sign change, it is from positive to negative.

Shift families with monotone likelihood ratios are totally positive of order 2 (Lehmann 1986, p. 509), so f is totally positive of order 2. Integration against f is therefore variation-diminishing: the function

$$\theta \mapsto \int h(\zeta) f_\theta(\zeta)\mathrm{d}\zeta = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{L}_\theta(d^\eta) \tag{48}$$

has no more sign changes than h does, and its sign changes must be in the same directions as those of h (Karlin 1968, Theorem 1.3.1). Consequently, any local extremum of $\theta \mapsto \mathbb{L}_\theta(d^\eta)$ is a global maximum.

The definition of η (21) ensures that $\mathrm{d}\mathbb{L}_\theta(d^\eta)/\mathrm{d}\theta = 0$ at $\theta = \eta$. Therefore, $\theta \mapsto \mathbb{L}_\theta(d^\eta)$ attains a global maximum at $\theta = \eta$, and the maximum risk of the Bayes procedure for prior π_η (the point mass at $\{\eta\}$) is equal to the Bayes risk of π_η.

Suppose that f_0 is symmetric (so that $\eta = 0$ suffices) and that $\tau > 2q_{1-\alpha}$. Then h has a sign change from negative to positive: recall that h is a difference of indicators of two intervals, $[-z, \tau - z]$ and $[-\tau + z, z]$, where $z = q_{1-\alpha} = -q_\alpha > 0$. The sign pattern of h depends on the ordering of the endpoints. There are six cases:

1. $-z < \tau - z \leqslant -\tau + z < z$
2. $-z \leqslant -\tau + z \leqslant \tau - z \leqslant z$
3. $-z \leqslant -\tau + z \leqslant z \leqslant \tau - z$
4. $-\tau + z \leqslant -z < \tau - z \leqslant z$
5. $-\tau + z < z \leqslant -z < \tau - z$
6. $-\tau + z \leqslant -z \leqslant z \leqslant \tau - z$

Case 1 (case 2) occurs if and only if $\tau \leqslant z$ (if and only if $\tau \leqslant 2z$), but we have supposed that $\tau > 2z$. Cases 3 and 4 cannot occur because they require $\tau = 2z$. Case 5 is impossible because $z > -z$ (recall that $\alpha < \frac{1}{2}$). In case 6, h has a sign change from negative to positive, as asserted. A total positivity argument similar to the one above thus shows that when $\tau > 2q_{1-\alpha}$, $\mathbb{L}_\theta(d^\eta)$ attains a global minimum (rather than maximum) at $\theta = \eta = 0$ and hence the truncated Pratt procedure is not minimax for expected measure.

The following lemma is a more general version of a common result (see Lehmann and Casella 1998, Theorem 1.4, p. 310).

Lemma 3. *Suppose $\pi \in \Pi$, the set of probability measures on Θ. Let $\tilde{\mathcal{D}}$ be a closed set of decisions. Let the risk at ζ of a decision $d \in \tilde{\mathcal{D}}$ be $\mathbb{L}_\zeta(d)$, and let the Bayes risk of a decision $d \in \tilde{\mathcal{D}}$ with respect to prior $\pi \in \Pi$ be*

$$\mathbb{L}_\pi(d) = \int_\Theta \mathbb{L}_\zeta(d)\pi(\mathrm{d}\zeta). \tag{49}$$

Suppose $\tilde{\mathcal{D}}$ is compact in a topology in which $d \to \mathbb{L}_\pi(d)$ is lower semicontinuous, for all π. Then for each $\pi \in \Pi$, $\tilde{\mathcal{D}}$ contains at least one Bayes decision for prior π, $d^\pi \in \tilde{\mathcal{D}} : \mathbb{L}_\pi(d^\pi) = \inf_{d\in\tilde{\mathcal{D}}}\mathbb{L}_\pi(d)$. Suppose $\lambda \in \Pi$ satisfies $\mathbb{L}_\lambda(d^\lambda) = \sup_{\zeta\in\Theta}\mathbb{L}_\zeta(d^\lambda)$. Then d^λ is minimax, and λ is least favourable: $\mathbb{L}_\lambda(d^\lambda) \geqslant \mathbb{L}_\pi(d^\pi)$ for all $\pi \in \Pi$.

It follows from Lemma 3 that the prior π_η defined above is least favourable, and that d^η is minimax. Equation (23) follows from equation (45) and Theorem 1. □

3.3. Theorem 3

We first show that Theorem 1 essentially applies to the scale-invariant confidence procedures, so we can characterize the minimax procedures using duality.

Lemma 4. *$\mathcal{D}_i$ is a weak-star compact subset of $L_\infty[\nu \times \mu]$.*

Because $\mathcal{D}_\alpha$ is closed, it follows that $\mathcal{D}_{\alpha,i} = \mathcal{D}_\alpha \cap \mathcal{D}_i$ is weak-star compact in $L_\infty[\nu \times \mu]$.

Proof of Lemma 4. Suppose that, instead of observing $(\overline{X}, S)$, we observed $Z = (\overline{X} - \zeta)/S$.

The set Δ of measurable decision functions based on Z is weak-star compact in $L_\infty[\nu \times \lambda]$. Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(\zeta, \bar{x}, s) = (\zeta, (\bar{x} - \zeta)/s)$. Any $d \in \mathcal{D}_i$ can be written as the composition $d = \delta \circ T$ for some $\delta \in \Delta$. We want to show that the map $\delta \mapsto \delta \circ T$ from Δ onto $\mathcal{D}_i$ is weak-star continuous; that will establish that $\mathcal{D}_i$ is weak-star compact as the image of a weak-star compact set under a weak-star continuous map. For each $\zeta \in \Theta$, consider the bijective change of variables

$$(\bar{x}, s) \mapsto (z = (\bar{x} - \zeta)/s, s). \tag{50}$$

The Jacobian of this transformation is s, so

$$\begin{aligned}
&\int_{\Theta\times\mathbb{R}\times\mathbb{R}^+} \delta(\zeta, (\bar{x} - \zeta)/s) g(\zeta, \bar{x}, s) \nu(\mathrm{d}\zeta) \mu(\mathrm{d}\bar{x}, \mathrm{d}s) \\
&\quad = \int_{\Theta\times\mathbb{R}\times\mathbb{R}^+} \delta(\zeta, z) g(\zeta, sz + \zeta, s) s \nu(\mathrm{d}\zeta) \mu(\mathrm{d}z, \mathrm{d}s) \\
&\quad = \int_{\Theta\times\mathbb{R}} \delta(\zeta, z) \left(\int_{\mathbb{R}^+} s g(\zeta, sz + \zeta, s) \mathrm{d}s \right) \nu(\mathrm{d}\zeta) \lambda(\mathrm{d}z).
\end{aligned} \tag{51}$$

It follows as a special case (namely, $\delta \equiv 1$) that $(\zeta, z) \mapsto \int_{\mathbb{R}^+} s g(\zeta, sz + \zeta, s) \mathrm{d}s \in L_1[\nu \times \lambda]$, and thus that if $\delta_n \to \delta$ in the weak-star topology on $L_\infty[\nu \times \lambda]$ then $\delta_n \circ T \to \delta \circ T$ in the weak-star topology on $L_\infty[\nu \times \mu]$, as required. $\square$

The following lemma helps to characterize the risk function of d_i^{TP}.

Lemma 5. *If* $n \geqslant 3$ and $\tau/\sigma \leqslant 2u\sqrt{(n-2)/(n(n-1))}$, *then*

$$\theta \mapsto \frac{d}{d\theta} \mathbb{L}_{(\theta,\sigma)}(d_i^{\mathrm{TP}}) \tag{52}$$

$$= \frac{d}{d\theta} \mathbb{E}_{(\theta,\sigma)} \int_{-\tau}^{\tau} d_i^{\mathrm{TP}}(\zeta, (\overline{X}, S)) \mathrm{d}\zeta \tag{53}$$

is positive for $\theta < 0$, *negative for* $\theta > 0$, *and has a unique zero at* $\theta = 0$.

Lemma 5 is proved at the end of this subsection.

Proof of Theorem 3. Define d_i^{TP} as in Theorem 3. Let Π be a set of probability measures as specified in Section 2.2. For any $\pi \in \Pi$ and any fixed $\sigma \in \Sigma$, define

$$\mathbb{L}_{(\pi,\sigma)}(d) \equiv \int_\Theta \mathbb{L}_{(\zeta,\sigma)}(d) \pi(\mathrm{d}\zeta). \tag{54}$$

To prove Theorem 3, we apply Lemmas 2 and 4 to use Theorem 1 to obtain a result analogous to Corollary 1 for scale-invariant procedures:

$$\inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\Theta,\sigma)}(d) = \sup_{\pi\in\Pi} \inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\pi,\sigma)}(d). \tag{55}$$

For each ζ, the most powerful scale-invariant test of $H_\zeta : \theta = \zeta$ against the alternative $H_0 : \theta = 0$ is $d_i^{\mathrm{TP}}((X_1, \ldots, X_n), \zeta)$. This implies that for any fixed σ, d_i^{TP} minimizes, among scale-invariant level-α procedures, the expected confidence set Lebesgue measure when $\theta = 0$:

$$\inf_{d \in \mathcal{D}_{\alpha,i}} \mathbb{L}_0(d) = \mathbb{L}_0(d_i^{\mathrm{TP}}). \tag{56}$$

The procedure d_i^{TP} is thus a Bayes decision from $\mathcal{D}_{\alpha,i}$ for risk $\mathbb{L}_0$ and prior π_0, a point mass at 0. By Lemma 5, if $\tau/\sigma \leqslant 2t_{1-\alpha}\sqrt{(n-2)/(n(n-1))}$ then the risk of d_i^{TP}, $\mathbb{L}_{(\cdot,\sigma)}(d_i^{\mathrm{TP}})$, attains a global maximum at 0. The maximum risk of the Bayes procedure against π_0 is equal to the Bayes risk of π_0. It follows from Lemma 3 that d_i^{TP} is minimax. □

Proof of Lemma 5. Let $k \equiv n-1$ and $u = t_{1-\alpha}$. In terms of the value $(\bar{x}, s^2)$ of $X = (\overline{X}, S^2)$, the procedure d_i^{TP} is

$$d_i^{\mathrm{TP}}(\theta, (\bar{x}, s)) \equiv \begin{cases} 1[(\bar{x}-\theta)/(sn^{-1/2}) \leqslant u], & \theta \leqslant 0, \\ 1[(\bar{x}-\theta)/(sn^{-1/2}) \geqslant -u], & \theta > 0. \end{cases} \tag{57}$$

Fix σ, $\tau > 0$. Then

$$\mathbb{L}_{(\theta,\sigma)}(d) = \int_0^\tau \mathbb{E}_{(\theta,\sigma)}\left[1 - \Phi\left(\frac{\zeta - \theta - uSn^{-1/2}}{\sigma n^{-1/2}}\right)\right]\mathrm{d}\zeta + \int_{-\tau}^0 \mathbb{E}_{(\theta,\sigma)}\left[\Phi\left(\frac{\zeta - \theta + uSn^{-1/2}}{\sigma n^{-1/2}}\right)\right]\mathrm{d}\zeta. \tag{58}$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{L}_{(\theta,\sigma)}(d) \propto \mathbb{E}_{(\theta,\sigma)}g(\overline{X}, S), \tag{59}$$

where

$$g(\bar{x}, s) \equiv 1\left[\frac{-\bar{x}}{u} \leqslant \frac{s}{\sqrt{n}} \leqslant \frac{\tau - \bar{x}}{u}\right] - 1\left[\frac{\bar{x}}{u} \leqslant \frac{s}{\sqrt{n}} \leqslant \frac{\bar{x}+\tau}{u}\right]. \tag{60}$$

Note that, for all $\bar{x}$, $g(-\bar{x}, s) = -g(\bar{x}, s)$. Now

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{L}_\theta(d) \propto \int_{\mathbb{R}} \mathbb{E}_{(\theta,\sigma)}(g(\overline{X}, S)|\overline{X} = \bar{x})\phi\left(\frac{\bar{x}-\theta}{\sigma n^{-1/2}}\right)\mathrm{d}\bar{x}. \tag{61}$$

Because the normal density is totally positive, the number of sign changes of $\theta \mapsto \mathrm{d}\mathbb{L}_{(\theta,\sigma)}(d)/\mathrm{d}\theta$ is no larger than the number of sign changes of a version of $\bar{x} \mapsto \mathbb{E}_{(\theta,\sigma)}(g(\overline{X}, S)|\overline{X} = \bar{x})$. One version is $\bar{x} \mapsto Ch(\bar{x})$, where C is a constant that depends on σ, α, and k, but not on θ or $\bar{x}$:

$$h(\bar{x}) \equiv \int_0^\infty g(\bar{x},\, r) r^{k-1} \mathrm{e}^{-kr^2/(2\sigma^2/n)} \mathrm{d}r \tag{62}$$

$$= \begin{cases} -h(-\bar{x}), & \bar{x} \leqslant 0, \\ \displaystyle\int_0^{(\tau-\bar{x})/u} r^{k-1} \mathrm{e}^{-kr^2/(2\sigma^2/n)} \mathrm{d}r - \int_{\bar{x}/u}^{(\tau+\bar{x})/u} r^{k-1} \mathrm{e}^{-kr^2/(2\sigma^2/n)} \mathrm{d}r, & 0 \leqslant \bar{x} \leqslant \tau, \\ \displaystyle -\int_{\bar{x}/u}^{(\tau+\bar{x})/u} r^{k-1} \mathrm{e}^{-kr^2/(2\sigma^2/n)} \mathrm{d}r, & \bar{x} > \tau. \end{cases} \tag{63}$$

We make the following claims:

(i) h is antisymmetric about 0;
(ii) h is continuously differentiable in $\bar{x}$;
(iii) $h(0) = 0$;
(iv) $h(\bar{x}) < 0$ for sufficiently small positive $\bar{x}$, and $h'(0) < 0$;
(v) $h(\bar{x}) < 0$ for $\bar{x} \geqslant \tau$;
(vi) If $\tau/(\sigma n^{-1/2}) \leqslant 2u\sqrt{(k-1)/k}$, then h' takes the value 0 at most once on $[0, \tau]$.

Claims (i)–(v) are clear upon inspection of (63); (vi) is discussed below. Together (i)–(vi) imply that h changes sign once as $\bar{x}$ ranges from $-\infty$ to ∞, going from positive to negative as $\bar{x}$ increases through 0. Total positivity and (61) imply that $\theta \mapsto \mathrm{d}\mathbb{L}_{(\theta,\sigma)}(d)/\mathrm{d}\theta$ follows the same pattern, and by antisymmetry of h its zero must be at $\theta = 0$. That is, $\theta \mapsto \mathbb{L}_{(\theta,\sigma)}(d)$ attains its maximum at 0.

For (vi), observe that on $[0, \tau]$,

$$h'(\bar{x}) \propto (\bar{x}/\tau)^{k-1} \mathrm{e}^{-C(\bar{x}/\tau)^2/2} [-\mathrm{e}^{-C/2}(h_1(\bar{x}/\tau) + h_2(\bar{x}/\tau)) + 1], \tag{64}$$

where

$$h_1(\zeta) \equiv (1 + 1/\zeta)^{k-1} e^{-C\zeta},$$
$$h_2(\zeta) \equiv (1/\zeta - 1)^{k-1} e^{C\zeta}, \tag{65}$$
$$C \equiv \frac{k\tau^2}{u^2\sigma^2 n^{-1}}.$$

We now show that $h_1 + h_2$ is strictly decreasing on $(0, 1)$ provided $\tau/(\sigma n^{-1/2}) \leqslant 2u\sqrt{(k-1)/k}$, the bound in (vi). It follows that h' is zero at most once. Clearly, h_1 is strictly decreasing on $(0, \infty)$, regardless of $\tau/(\sigma n^{-1/2})$. Second, h_2 has derivative

$$h_2'(\zeta) = (k-1)(1/\zeta - 1)^{k-2} e^{-C\zeta} \left(-1/\zeta^2 + \frac{C}{k-1} 1/\zeta - \frac{C}{k-1} \right). \tag{66}$$

Because the term in large brackets is a quadratic function of $1/\zeta$, it has a zero on $(0, \infty)$ if and only if $C > 4(n-2)$. Otherwise, it does not change sign on the positive half-line. It must be negative as $\zeta \to \infty$, so if it does not change sign on $(0, \infty)$, it must be negative on that interval. Thus h_2' must be negative on $(0, 1)$, provided $\tau/(\sigma n^{-1/2}) \leqslant 2\sqrt{(k-1)/k}u$. But then h_2 is decreasing, so $h_1 + h_2$ is strictly decreasing, so, by (64), h has property (vi). □

Appendix: Minimax fixed-length confidence intervals

Zeytinoglu and Mintz (1984; 1988) study confidence intervals $[\hat{\theta} - l/2, \hat{\theta} + l/2]$ for a BNM that minimize $\sup_{\theta \in \Theta} \mathbb{P}_\theta\{\theta \notin [\hat{\theta} - l/2, \hat{\theta} + l/2]\}$, the maximum non-coverage probability, among random intervals of fixed length l. Their results can be used to find $1 - \alpha$ confidence intervals that are minimax for length among fixed-width $1 - \alpha$ confidence intervals.

Suppose $Z \sim \mathcal{N}(\theta, 1)$, $\theta \in [-\tau, \tau]$. Zeytinoglu and Mintz (1984, p. 949) show that if $l/2 < \tau \leqslant l$, the minimax non-coverage interval of fixed length l is centred at

$$\hat{\theta}(Z) = \begin{cases} Z, & |Z| \leqslant \tau - l/2, \\ \tau - l/2, & |Z| > \tau - l/2, \end{cases} \tag{67}$$

and has maximum non-coverage probability $\Phi(-l/2)$. If $l < \tau \leqslant 3l/2$, then the minimax non-coverage interval of fixed length l is centred at

$$\hat{\theta}(Z) = \begin{cases} 0, & |Z| < a, \\ Z - a, & a \leqslant |Z| < a + l, \\ l, & a + l \leqslant |Z|, \end{cases} \tag{68}$$

where a is the solution of $2\Phi(-a - l/2) = \Phi(a - l/2)$. In this case, the maximum non-coverage probability is $\Phi(a - l/2)$ (Zeytinoglu and Mintz, 1984, p. 948).

The left-hand half of Table 3 gives maximum non-coverage probabilities of the minimax non-coverage length-l procedure, assuming that $\tau \in (l/2, l]$. The right-hand half gives the a needed to specify the minimax non-coverage length-l procedure if $\tau \in (l, 3l/2]$, along with corresponding maximum noncoverage probabilities.

When $\tau \in [1.6, 3.25]$, the optimal fixed-width 95% interval is centred at a point $\hat{\theta}$ of form (67) and has width between 3.25 and 3.30. Since intervals of this form have maximum non-coverage chance $\Phi(-l/2)$, the minimax-width 95% interval has width precisely $2z_{0.95} \approx 3.28$.

If $\tau \in [3.6, 5.4]$, an interval of width 3.60 centred at a point of form (68) has 95% coverage. This minimax-width fixed-width 95% confidence interval is given by (68), with $a = 0.158$.

If $\tau \in [3.30, 3.60)$, then no interval with centring point given by (68) has sufficient uniform coverage probability. To obtain a 95% confidence interval one must centre it at a point of form (67). This means $\tau \in (l/2, l]$, implying that $l \geqslant \tau$. The maximum non-coverage at $l = 3.30$ falls under the 5% cut-off, so l need be no larger than τ. It thus turns out that for $\tau \in [3.30, 3.60)$ the maximum non-coverage probability of the minimax-width fixed-width 95% interval is strictly less than 5% ; for $\tau = 3.60$, this 95% interval has width 3.60 and is in fact a 96.4% interval.

Acknowledgements

This work was supported by the National Science Foundation through Presidential Young Investigator Award DMS-89-57573 and grants DMS-94-04276, AST-95-04410, DMS-97-

Table 3. Minimax non-coverage fixed-length intervals: maximum non-coverage probabilities and offset constants a

$l/2 < \tau \leqslant l$		$l < \tau \leqslant 3l/2$		
l[a]	p[b]	l	a[c]	p[d]
3.00	0.067			
3.05	0.064			
3.10	0.061			
3.15	0.058			
3.20	0.055	3.20	0.171	0.077
3.25	0.052	3.25	0.169	0.073
3.30	0.049	3.30	0.168	0.069
3.35	0.047	3.35	0.166	0.066
3.40	0.045	3.40	0.164	0.062
3.45	0.042	3.45	0.163	0.059
3.50	0.040	3.50	0.161	0.056
3.55	0.038	3.55	0.159	0.053
3.60	0.036	3.60	0.158	0.050
		3.65	0.156	0.048
		3.70	0.155	0.045
		3.75	0.153	0.043
		3.80	0.152	0.040

[a]Length of confidence interval.
[b]Maximum non-coverage probability of minimax non-coverage interval: $p = \sup_{\theta\in[-\tau,\tau]}\mathbb{P}_\theta(\theta \notin [\hat{\theta} - l/2, \hat{\theta} + l/2])$, for $\hat{\theta}$ as defined in (67).
[c]When $\tau \in (l, 3l/2]$, a combines with (68) to specify the minimax non-coverage interval of length l.
[d]Maximum non-coverage probability of minimax non-coverage interval: $p = \sup_{\theta\in[-\tau,\tau]}\mathbb{P}_\theta(\theta \notin [\hat{\theta} - l/2, \hat{\theta} + l/2])$, for $\hat{\theta}$ as defined in (68).

09320, DMS-98-72979, DMS-00-71468, and Postdoctoral Fellowship DMS-0102056, and by NASA through grants NAG5-3941 and NRA-96-09-OSS-034SOHO. Part of the work was performed while the first and third authors were on appointment as Miller Research Professors at the Miller Institute for Basic Research in Science.

We are grateful to D.A. Freedman for advice, direction, and comments on an earlier draft, and to R. Purves for valuable conversations. We thank the referees for helpful suggestions.

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.

Brown, L.D. (1966) On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.*, **37**, 1087–1136.

Casella, G. (2002) Comment on 'Setting confidence intervals for bounded parameters' by M. Mandelkern. *Statist. Sci.*, **17**(2), 159–160.

Casella G. and Hwang, J.T. (1983) Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.*, **78**, 688–698.

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

Casella, G. Hwang, J.T.G. and Robert, C. (1993) A paradox in decision-theoretic interval estimation. *Statist. Sinica*, **3**, 141–155.

Donoho, D.L. (1994) Statistical estimation and optimal recovery. *Ann. Statist.*, **22**, 238–270.

Donoho, D.L. and Liu, R.C. (1991) Geometrizing rates of convergence. III. *Ann. Statist.*, **19**, 668–701.

Fan, K. (1953) Minimax theorems. *Proc. Natl. Acad. Sci. USA*, **39**, 42–47.

Ghosh, J.K. (1961) On the relation among shortest confidence intervals of different types. *Calcutta Statist. Assoc. Bull.*, **10**, 147–152.

Gleser, L.J. (2002) Comment on 'Setting confidence intervals for bounded parameters' by M. Mandelkern. *Statist. Sci.*, **17**(2), 160–163.

Hansen, H.B. (2001) Minimax expected length confidence intervals. Master's thesis, University of California, Berkeley.

Hooper, P.M. (1982) Invariant confidence sets with smallest expected measure. *Ann. Statist.*, **10**, 1283–1294.

Hooper, P.M. (1984) Correction. *Ann. Statist.*, **12**, 784.

Hwang J.T. and Casella, G. (1982) Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.*, **10**, 868–881.

Ibragimov, I.A. and Khas'minskii, R.Z. (1985) On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.*, **29**, 18–32.

Joshi, V.M. (1967) Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.*, **38**, 1868–1875.

Joshi, V.M. (1969) Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.*, **40**, 1042–1067.

Kamberova, G. and Mintz, M. (1999) Minimax rules under zero-one loss for a restricted parameter. *J. Statist. Plann. Inference*, **79**, 205–221.

Kamberova, G., Mandelbaum, R. and Mintz, M. (1996) Statistical decision theory for mobile robotics: theory and application. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI '96*. New York: Institute of Electrical and Electronics Engineers.

Karlin, S. (1968) *Total Positivity*, Volume 1. Stanford, CA: Stanford University Press.

Kneser, H. (1952) Sur un théorème fondamental de la théorie des jeux. *C. R. Acad. Sci. Paris*, **234**, 2418–2420.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses* (2nd edn). New York: Wiley.

Lehmann, E.L. and Casella, G. (1998) *Theory of Point Estimation*. New York: Springer-Verlag. 2nd edition.

Mandelkern, M. (2002a) Setting confidence intervals for bounded parameters. *Statist. Sci.*, **17**(2), 149–159.

Mandelkern, M. (2002b) Setting confidence intervals for bounded parameters, rejoinder. *Statist. Sci.*, **17**(2), 171–172.

Pratt, J.W. (1961) Length of confidence intervals. *J. Amer. Statist. Assoc.*, **56**, 549–567.

Pratt, J.W. (1963) Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.*, **34**, 574–586.

Schafer, C.M. (2004) Constructing confidence regions of optimal expected size: theory and application to cosmic microwave inference. PhD thesis, University of California, Berkeley.

Schafer, C.M. and Stark, P.B. (2004) Using what we know: inference with physical constraints. In L. Lyons, R. Mount and R. Reitmeyer (eds), *Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology PHYSTAT2003*, pp. 25–34, Menlo Park, CA, 2004. Standford Linear Accelerator Center.

Sion, M. (1958) On general minimax theorems. *Pacific J. Math.*, **8**, 171–176.

Stark, P.B. (1992) Affine minimax confidence intervals for a bounded normal mean. *Statist. Probab. Lett.*, **13**, 39–44.

van Dyk, D.A. (2002) Comment on 'Setting confidence intervals for bounded parameters' by M. Mandelkern. *Statist. Sci.*, **17**(2), 164–168.

Wasserman, L. (2002) Comment on 'Setting confidence intervals for bounded parameters' by M. Mandelkern. *Statist. Sci.*, **17**(2), 163.

Woodroofe, M. and Zhang, T. (2002) Comment on 'Setting confidence intervals for bounded parameters' by M. Mandelkern. *Statist. Sci.*, **17**(2), 168–171.

Zeytinoglu, M. and Mintz, M. (1984) Optimal fixed size confidence procedures for a restricted parameter space. *Ann. Statist.*, **12**, 945–957.

Zeytinoglu, M. and Mintz, M. (1988) Robust fixed size confidence procedures for a restricted parameter space. *Ann. Statist.*, **16**, 1241–1253.

Received June 2003 and revised December 2004