# Minimax Expected Measure Confidence Sets for Restricted Location Parameters*

Steven N. Evans
Department of Statistics
University of California
Berkeley, CA 94720-3860

Bendek B. Hansen
Department of Statistics
University of Michigan
Ann Arbor, MI 48109-1092
and
Philip B. Stark*
Department of Statistics
University of California
Berkeley, CA 94720-3860

Technical Report 617
Department of Statistics, University of California
Berkeley, CA
Revised 23 May 2003

## 1 Summary

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

*Running Title: Minimax measure confidence sets. AMS Subject Classifications: 62C10, 62C20, 62F25, 62F30, 90C46. Keywords: constrained parameters, Bayes/Minimax duality. 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-97-09320, 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 in the Miller Institute for Basic Research in Science.

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 $\tau \leq 3.25$. 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 $[X - 1.96, X + 1.96] \cap [-3, 3]$.

## 2 Introduction

There are many procedures for constructing confidence sets. Classical considerations for choosing among them include accuracy, unbiasedness, equivariance, and combinations of these [Lehmann, 1986]. Accuracy seems quite natural: Given a pair of confidence procedures with the same probability of covering the correct value, the procedure with smaller chance of covering incorrect values is preferable.

Unbiasedness—the requirement that the probability of covering the true value of the parameter be at least as large as the probability of covering any other value—is related to accuracy and also seems desirable in many situations. Equivariance requires a bit more structure: The parameter space and the set of possible data both must be equipped with groups of transformations. There must be a correspondence between elements of the data group and elements of the parameter group. Then the confidence procedure is equivariant if the confidence set associated with the transformation of the data by an element of the data group is the transformation of the confidence set by the corresponding element of the parameter group. This limits the applicability of equivariance to situations with a high degree of symmetry. See Mandelkern [2002a] for a list of properties some view as desirable in a confidence interval for a bounded parameter; his list includes equivariance under one-to-one transformations of the parameter—which is rather restrictive.

Even when these three classical criteria can be applied, they can be at odds with the scientific goal of the estimation problem, can preclude intuitively reasonable optimality criteria [Woodroofe and Zhang, 2002], and can fail to specify a unique procedure.

The accuracy of a confidence procedure usually depends both on the procedure and on unknown parameters, making accuracy alone impractical as a criterion for choosing among confidence procedures. However, some problems do admit *uniformly most accurate* confidence sets. Uniformly most accurate confidence sets minimize expected measure for the worst-case values of the parameter [Lehmann, 1986, pp. 261, 524].

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. The structure required to study expected measure is both more and less restrictive than that used traditionally to study accuracy: The set of possible parameter values must be a measurable space, and the confidence sets must be measurable subsets of the set of parameters, but confidence sets with minimax expected measure can exist even when there is no uniformly most accurate confidence set. See § 3.

### 2.1 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, Ibragimov and Khas'minskii, 1984]. Estimating a BNM arises in robotics [Kamberova et al., 1996, Kamberova and Mintz, 1999], and it is of theoretical interest in its own right (*e.g.*, 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, Casella, 2002, Gleser, 2002, Wasserman, 2002, van Dyk, 2002, Woodroofe and Zhang, 2002, Mandelkern, 2002b].

The constraint $\theta \in [-\tau, \tau]$ allows point estimators to have smaller risk than otherwise would be possible. Bickel [1981], Casella and Strawderman [1981], Gourdin et al. [1994], Vidakovic and Dasgupta [1996], and Marchand and Perron [2001] studied minimax MSE estimates of the BNM. Point estimates of a BNM for loss functions other than squared-error also have been considered [Bischoff and Fieger, 1992, Eichenauer-Herrman and Ickstadt, 1992, Donoho, 1994], as have point estimates of a multi-dimensional BNM [Berry, 1990, Weiss, 1988, Marchand and Perron, 2001], point estimates of restricted parameters for distributions other than the normal [Johnstone and MacGibbon, 1992], and point estimates of the square of a bounded normal mean [Donoho and Nussbaum, 1990, Fan and Gijbels, 1992].

The constraint $\theta \in [-\tau, \tau]$ also allows confidence sets for a normal mean to be smaller without sacrificing coverage probability: Consider the conventional 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

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

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 tradeoff between coverage and length as a decision problem using a measure of loss that combines the two. However, Casella et al. [1993] show that this can produce interval estimates with undesirable properties. In contrast, Zeytinoglu and Mintz [1984], Zeytinoglu and Mintz [1988], and Kamberova and Mintz [1999] fix the length of the interval, then find how to center 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 appendix A.

By definition, the length of a minimax fixed-length interval is determined before the observation is made. Allowing the size of the confidence set to depend on the datum enlarges the collection of confidence procedures available, and variable-length intervals indeed can be shorter on the average than the minimax fixed-length interval without compromising uniform coverage probability.

Below, we determine how much the maximum expected size of a $1 - \alpha$ confidence set can be reduced by allowing the size to depend on the data. This minimax problem is not new. For example, Lehmann [1986, p. 524] states the general minimax problem for expected measure and relates it to accuracy. Minimax expected measure confidence sets have been constructed for some special cases in which the set of possible parameters has group structure and the procedure is restricted to be equivariant (*e.g.*, Hooper [1982, 1984] and Lehmann [1986]). Moreover, in many problems, equivariant confidence sets are not admissible, and using non-equivariant procedures—centered 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, when there is no uniformly most accurate procedure.

For inference about a normal mean $\theta \in [-\tau, \tau]$, $\tau \leq 2z_{1-\alpha}$, 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]

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

with $c = z_{1-\alpha}$. It is not surprising that the truncated Pratt interval has minimax expected length when $\tau$ is small: $\mathcal{I}_{\mathrm{P}}$ has minimal expected length at $\theta = 0$ among all confidence intervals with $1-\alpha$ coverage for all $\theta \in \Re$ [Pratt, 1961]. By continuity, it should nearly minimize expected length for a range of values of $\theta$ around zero, but if $\tau$ is sufficiently small, that range includes all permissible values of $\theta$. It is surprising to us how large $\tau$ can be: The truncated Pratt interval is minimax for expected length when $\tau$ as large as $2z_{1-\alpha}$, nearly twice as large as the value of $\tau$ for which the minimax MSE point estimate has a simple form [Casella and Strawderman, 1981]. Moreover, for $\tau \leq 2z_{1-\alpha}$, not only is the truncated Pratt interval minimax for expected length among non-randomized $1-\alpha$ confidence intervals, it is minimax for expected Lebesgue measure among more general randomized $1-\alpha$ confidence sets.

## 2.2 Improvements in expected length

Table 1 compares the maximum expected length of the optimal confidence interval (which is often the truncated Pratt) with the maximum expected lengths of some competing procedures, all at 95% confidence. With $\tau = 2.0$, 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 appendix A).

Insert Table 1

The truncated Pratt interval (*i.e.,* (3), with $c$ equal to the $1-\alpha$ quantile of the distribution of $X$ when $\theta = 0$) is minimax for any shift family of distributions with monotone likelihood ratios, provided that the shift parameter is restricted *a priori* to a sufficiently small set $\Theta = [-\tau, \tau]$.

### 2.3 Outline

This paper is organized as follows. § 3.1 presents the basic notation and assumptions. § 3.2 applies a minimax theorem due to Kneser [1952] and Fan [1953] to establish a Bayes-minimax duality for the expected measure of confidence sets, exploiting the representation of confidence sets in terms of families of randomized hypothesis tests. This leads to confidence sets of the form $S(x) \equiv \{\eta : f_\eta(x)/f_\pi(x) > \lambda_\eta\}$, where $f_\eta$ is the probability density of the observation if the parameter value is $\eta$, $f_\pi$ is a fixed mixture of densities corresponding to a Bayesian prior on parameters, and $\lambda_\eta$ are constants chosen so that the procedure has uniform $1-\alpha$ frequentist coverage probability. Such Bayesian/frequentist hybrid confidence sets have arisen in other contexts, *e.g.*, Brown et al. [1995]; similarly, see Casella [2002] for an argument in favor of frequentist-calibrated Bayesian credible regions.

§ 3.3 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. This is equivalent to the restriction that the density $f_0$ be strongly unimodal [Lehmann, 1986, p. 509]. Such distributions include the normal, uniform, logistic, and double exponential; a necessary and sufficient condition is that the cdf $F_0$ be continuous and that $\log F_0'$ be concave wherever neither one-sided derivative of $F_0$ vanishes [Ibragimov, 1956]. In particular, results for the BNM are corollary. § 3.4 extends the theory to situations with nuisance parameters, and studies confidence sets for the BNM where $\sigma^2$ is unknown, but for which the signal-to-noise ratio $\tau/\sigma$ is not too large. Proofs are postponed, for the most part, until § 4.

## 3 Principal Results

### 3.1 Framework, Notation and Assumptions

The framework that follows is similar to those of Joshi [1969], Hooper [1982, 1984] and Lehmann [1986].

Let $\Theta$ and $\mathcal{X}$ be measurable spaces. Let $\nu$ be a sigma-finite measure on $\Theta$, and let $\mu$ be a sigma-finite measure on $\mathcal{X}$. Let $\{\mathbb{P}_\zeta : \zeta \in \Theta\}$ be a family of probability distributions on $\mathcal{X}$, absolutely continuous with respect to $\mu$. For $\zeta \in \Theta$, let $f_\zeta$ denote the density of $\mathbb{P}_\theta$ with respect to $\mu$. 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

$X \sim f_\theta$ instead) and a uniform real-valued random variable $U \sim U[0,1]$ that is independent of $X$. The value of $\theta$ is unknown except that $\theta \in \Theta$. We seek a "small" confidence set $S(X,U)$ for $\theta$ based on the observation $X$ and the extra randomization $U$; the size of the set is measured by $\nu$. (In § 3.4, we allow $\theta$ to consist of two parts, the parameter of interest and a nuisance parameter. The nuisance parameter need not be subsumed into the measure space.) $\Theta$ captures possible *a priori* restrictions on $\theta$; for instance, in the BNM problem $\Theta = [-\tau, \tau]$, $\tau < \infty$.

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

$$\mathcal{D} \equiv \{d \in \mathcal{M} : 0 \le d(\zeta, x) \le 1, \text{ a.s. } (\nu \times \mu)\}. \tag{4}$$

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 $\mathcal{D}$ is weak-star compact according to the Banach-Alaoglu theorem. 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 hypothesized parameter value $\zeta$ and the datum $X$: If $U > d(\zeta, X)$, reject $H_\zeta$; otherwise not. The significance level of the test $d(\zeta, \cdot)$ of $H_\zeta$ is $1 - \mathbb{E}_\zeta d(\zeta, X)$, the chance that $U > d(\zeta, X)$ when $X \sim f_\zeta$. If $\lambda : \zeta \mapsto \lambda_\zeta$ is a measurable function of $\Theta$ into $\Re$, and if $\eta$ is any point in $\Theta$,

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

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

By virtue of the general duality between testing and confidence sets [Lehmann, 1986], each $d \in \mathcal{D}$ induces a randomized confidence set $S_d = S_d(X,U)$ for $\theta$, where

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

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 $\zeta$ is the chance that $U \le d(\zeta, X)$ when $X \sim \mathbb{P}_\zeta$:

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

The quantity $\mathbb{C}_\zeta(d)$ is well defined as a measurable function of $\zeta$ because, by assumption, $\mathbb{P}_\zeta$ has density $f_\zeta$ with respect to $\mu$ and $(\zeta, x) \mapsto f_\zeta(x)$ is product measurable. 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 $d' \in \mathcal{D}$, a.e. $(\nu \times \mu)$ 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 be $\theta$. The functions $d(\cdot, \cdot)$ in

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

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

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

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

which, like $\mathbb{C}_\zeta(d)$, is well defined as a measurable function of $\zeta$ because $\mathbb{P}_\zeta$ has density $f_\zeta$ with respect to $\mu$ and $(\zeta, x) \mapsto f_\zeta(x)$ is product measurable. The maximum expected $\nu$-measure of $S_d$ over $\Theta$ is

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

In this paper we characterize the decision functions $d \in \mathcal{D}_\alpha$ with minimal maximum risk $\mathbb{L}_\Theta(d)$.

### 3.2 Bayes-minimax duality for confidence procedures

Let $\Pi$ be the set of all probability measures on $\Theta$. For $\pi \in \Pi$, the $\pi$-average expected $\nu$-measure of the confidence set corresponding to the decision function $d$ is

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

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

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

Theorem 1 is proved in § 4.1. This theorem is useful because: (1) Procedures $d \in \widetilde{\mathcal{D}}$ that attain $\inf_{d \in \widetilde{\mathcal{D}}} \mathbb{L}_\pi(d)$ can be constructed using likelihood ratios, and (2) the set $\mathcal{D}_\alpha$ is weak-star compact in $L_\infty[\nu \times \mu]$ (see Lemma 2). This allows us to find, for any $\pi \in \Pi$, the decision function $d \in \mathcal{D}$ with uniform coverage probability $\mathbb{C}_\Theta(d) \geq 1 - \alpha$ that minimizes $\pi$-average expected $\nu$-measure.

For $\pi \in \Pi$, define the average density

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

Fix $\alpha \in (0,1)$ and let $\widetilde{\mathcal{D}} \equiv \mathcal{D}_\alpha$. Given $\pi \in \Pi$, let $d^\pi = d^\pi(\zeta, x)$ be a family of decision functions for size-$\alpha$ 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, $d^\pi$ is an amalgamation of likelihood ratio tests: For each $\zeta \in \Theta$ let

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

The function $\zeta \mapsto \lambda_\zeta$ is measurable because $(\zeta, x) \mapsto f_\zeta(x)$ is. 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 $c_\zeta$ chosen so that $\int d(\zeta, x) f_\zeta(x)\mu(dx) = 1 - \alpha$. Then $d^\pi \in \mathcal{D}_\alpha$, and $d^\pi$ minimizes $\mathbb{L}_\pi(\cdot)$ over $\mathcal{D}_\alpha$. (This follows from the optimality of each $d^\pi(\zeta, \cdot)$ as a level-$\alpha$ test and from the Ghosh-Pratt identity [Ghosh, 1961, Pratt, 1961, eq. 2]; see also § 4.1.)

**Corollary 1**

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

*for* $0 < \alpha < 1$.

### 3.3 Bounded real shift parameters

In this section, we study confidence sets for bounded location parameters of one-dimensional shift families, that is, the special case in which $\Theta, \mathcal{X} \subseteq \Re$, $\Theta$ is bounded, and $f_\theta(x) \equiv f(x - \theta)$ for some density $f$ with respect to Lebesgue measure.

Pratt [1961] constructed confidence sets for unrestricted parameters by inverting families of uniformly most powerful tests of the hypotheses $\theta = \zeta$

against a single alternative $\theta = \eta = 0 \in \Theta$. This corresponds to a decision function $d^\pi$ as in (16)–(17), with $\pi = \delta_\eta$ a point mass at $\eta$. Let $d^\eta \equiv d^{\delta_\eta}$ be the decision function that is most powerful against the alternative $\theta = \eta$. Pratt showed that $d^\eta$ yields the confidence set with smallest expected Lebesgue measure when $\theta = \eta$:

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

Suppose $\{f_\theta : \theta \in \Theta\}$ has monotone likelihood ratios ($f_{\theta_2}/f_{\theta_1}$ is nondecreasing in $x$, when $\theta_1 < \theta_2$). 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]:

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

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

Pratt [1963] was concerned primarily with the case $\Theta = \Re$. When $\Theta$ is a bounded subset of $\Re$, we call $d^\eta$ the *truncated Pratt procedure*. In this section, we show that for shift families with monotone likelihood ratios (including, for example, the normal, uniform, logistic, and double exponential), when $\tau$ is sufficiently small there is a point $\eta \in \Theta = [-\tau, \tau]$ such that the truncated Pratt procedure $d^\eta$ nas minimax expected Lebesgue measure among randomized $1 - \alpha$ confidence sets. Figure 1 shows the truncated Pratt procedure for $\eta = 0$, $\tau = 3$, and $\{\mathbb{P}_\theta : \theta \in \Theta\}$ the distributions with densities

$$\{f_\theta(\cdot) = \varphi(\cdot - \theta) : \theta \in [-\tau, \tau]\},$$

a normal shift family with bounded mean.

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

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

(The equation defining $\eta$ can be rewritten

$$\int_{q_\alpha}^{\tau + q_\alpha - \eta} f_0(x)dx = \int_{q_{1-\alpha} - \tau - \eta}^{q_{1-\alpha}} f_0(x)dx; \tag{22}$$

both integrals vary continuously as $\eta$ ranges from $-\tau$ to $\tau$, one decreases from a strictly positive quantity to 0, the other increases from 0, so there is a unique point at which they are equal.) If $f_0(\cdot)$ is symmetric about any point, $\eta = 0$.

Insert Figure 1

**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 $\eta$ satisfy (21). Suppose $\alpha < 1/2$. If*

$$\tau + |\eta| \le q_{1-\alpha} - q_\alpha, \tag{23}$$

*then*

$$\inf_{d\in\mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = \int_{-\tau}^{\eta} F(\zeta + q_{1-\alpha}) d\zeta + \int_{\eta}^{\tau} (1 - F(\zeta + q_\alpha)) d\zeta, \tag{24}$$

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

$$\tau = \tau + |\eta| > q_{1-\alpha} - q_\alpha = 2q_{1-\alpha}. \tag{25}$$

**Corollary 2 The bounded normal mean.** *Let $\Theta = [-\tau, \tau]$; let $z_\beta$ be the $\beta$-quantile of the $\mathcal{N}(0,1)$ distribution; and suppose that $\alpha < 1/2$. If $\tau \le 2z_{1-\alpha}$, then*

$$\inf_{d\in\mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = 2\int_0^\tau \Phi(z_{1-\alpha} - \zeta) d\zeta, \tag{26}$$

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

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

$$\mathcal{I}_{\mathrm{TP}}(X) = [(X - z_{1-\alpha}) \wedge 0, (X + z_{1-\alpha}) \vee 0] \cap [-\tau, \tau], \tag{27}$$

to that of the truncated conventional confidence interval,

$$\mathcal{I}_{\mathrm{T}}(X) = [X - z_{1-\alpha/2}, X + z_{1-\alpha/2}] \cap [-\tau, \tau], \tag{28}$$

and to that of the minimax affine confidence interval $\mathcal{I}_{\mathrm{A}}$.

Insert Table 2

### 3.4 Bounded normal mean with unknown variance: nuisance parameters

In this section, 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

$$\mathbb{P}_{(\Theta,\Sigma)} \equiv \{\mathbb{P}_{(\theta,\sigma)} : \theta \in \Theta,\ \sigma \in \Sigma\}. \tag{29}$$

We assume as before that $\Theta$ is a measure space with measure $\nu$, and we seek a confidence set for $\theta$ with small expected $\nu$-measure. We assume that the family $\mathbb{P}_{(\Theta,\Sigma)}$ is dominated by a $\sigma$-finite measure $\mu$, as we did in § 3.1. Let $f_{(\theta,\sigma)}$ be the density of $\mathbb{P}_{(\theta,\sigma)}$ with respect to $\mu$. We also assume that for each fixed $\sigma \in \Sigma$, the mapping $(\theta, x) \mapsto f_{(\theta,\sigma)}(x)$ is product measurable, as we did in § 3.1.

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

$$\mathbb{C}_{(\zeta,\sigma)}(d) \equiv \mathbb{E}_{(\zeta,\sigma)} d(\zeta, X), \tag{30}$$

$$\mathbb{C}_{(\Theta,\sigma)}(d) \equiv \nu\text{-ess}\inf_{\zeta\in\Theta} \mathbb{C}_{(\zeta,\sigma)}(d), \tag{31}$$

and

$$\mathcal{D}_\alpha = \{d \in \mathcal{D} : \mathbb{C}_{(\Theta,\sigma)}(d) \geq 1 - \alpha,\ \forall \sigma \in \Sigma\}. \tag{32}$$

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

$$\mathbb{L}_{(\zeta,\sigma)}(d) \equiv \mathbb{E}_{(\zeta,\sigma)} \int_\Theta d(\eta, X)\nu(d\eta) \tag{33}$$

and

$$\mathbb{L}_{(\Theta,\sigma)}(d) \equiv \sup_{\zeta\in\Theta} \mathbb{L}_{(\zeta,\sigma)}(d). \tag{34}$$

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

$$\mathbb{L}_{(\Theta,\sigma)}(d^*) = \inf_{d\in\mathcal{D}_\alpha} \mathbb{L}_{(\Theta,\sigma)}(d). \tag{35}$$

We specialize now to the bounded normal mean with unknown variance $\sigma^2$. We do not find a decision rule $d^*$ that is optimal for all $\sigma \in \Re^+$, but we do show that the truncated Pratt is optimal (among scale-invariant procedures) provided $\tau$ is not too large compared with $\sigma$. We observe $X = (X_i)_{i=1}^n$, where $\{X_i\}_{i=1}^n$ are i.i.d. $\mathcal{N}(\theta, \sigma^2)$ with $\theta \in \Theta = [-\tau, \tau]$ but otherwise unknown, and $\sigma \in \Sigma = \Re^+$ but otherwise unknown. Let $\nu$ be Lebesgue measure on $[-\tau, \tau]$, and let $\mu$ be Lebesgue measure on $\Re$. We seek a confidence set for $\theta$ that has $1-\alpha$ coverage probability whatever be $\theta \in \Theta$ and $\sigma \in \Sigma$, and we want the expected measure of the set to be as small as possible at the worst $\theta$, for each value of $\sigma$.

Let $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 \equiv \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ be the sample mean and sample variance. Because $(\bar{X}, S)$ is sufficient for $\mathbb{P}_{(\Theta,\Re^+)}$ and the loss $\mathbb{L}_{(\zeta,\sigma)}(d)$ is convex, by the Rao-Blackwell theorem [Lehmann and Casella, 1998] it suffices to consider decision functions that depend on the data only through $\bar{X}$ and $S$.

Because the scale parameter $\sigma$ is an unknown nuisance parameter, we restrict consideration just to decision rules $d(\zeta, (\bar{x}, s))$ that are invariant under changes of scale: The *principle of invariance* [Lehmann, 1986, §6.11] requires that for all $c > 0$,

$$d(\zeta, (\bar{x}, s)) = d(\zeta, (\zeta + c(\bar{x} - \zeta), cs)). \tag{36}$$

Combining these two restrictions leads us to focus on decision functions that depend on the data only through $(\bar{X} - \zeta)/S$. 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 $\sigma$ it contains one that solves (35).

In general, which scale-invariant procedure is minimax for expected measure depends on $\sigma$, but the following theorem asserts that the truncated Pratt procedure is minimax scale-invariant provided $\tau$ is not too big compared with $\sigma$.

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

$$d_i^{\mathrm{TP}}(\zeta, (\bar{x}, s)) \equiv \begin{cases} 1[(\bar{x} - \zeta)/s \le t_{1-\alpha}], & \zeta \le 0 \\ 1[(\bar{x} - \zeta)/s \ge -t_{1-\alpha}], & \zeta > 0, \end{cases} \tag{37}$$

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

1. $d_i^{\mathrm{TP}} \in \mathcal{D}_{\alpha,i}$,

2. *If*

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

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

$$\inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\Theta,\sigma)}(d) = \mathbb{L}_{(\Theta,\sigma)}(d_i^{\mathrm{TP}}) = 2\int_0^{\tau} F_{\zeta\sqrt{n}/\sigma}(t_{1-\alpha})d\zeta, \tag{38}$$

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

**Remark.** The condition $\tau/\sigma \leq 2t_{1-\alpha}\sqrt{\frac{n-2}{n(n-1)}}$ is sufficient, but not necessary, for $d_i^{\mathrm{TP}}$ to be minimax among scale-invariant procedures. Numerical experiments suggest that the largest $\tau/\sigma$ for which the result is true is between $\frac{2t_{1-\alpha}}{\sqrt{n}}$ and $2t_{1-\alpha}\sqrt{\frac{n-2}{n(n-1)}}$.

# 4 Proofs

## 4.1 Theorem 1

To prove theorem 1 we apply a general minimax theorem that requires that $\widetilde{\mathcal{D}}$ be compact in a topology in which $d \mapsto \mathbb{L}_\pi(d)$ is lower semicontinuous. The weak-star topology on $L_\infty[\nu\times\mu]$ suffices.

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

$$\begin{aligned}
\mathbb{L}_\pi(d) &= \int_\Theta \mathbb{L}_\zeta(d)\pi(d\zeta) \\
&= \int_\Theta \left(\mathbb{E}_\zeta \int_\Theta d(\eta,X)\nu(d\eta)\right)\pi(d\zeta) \\
&= \int_\Theta \left(\int_{\Theta\times\mathcal{X}} d(\eta,x)\nu(d\eta)f_\zeta(x)\mu(dx)\right)\pi(d\zeta) \\
&= \int_{\Theta\times\mathcal{X}} d(\eta,x)\left(\int_\Theta f_\zeta(x)\pi(d\zeta)\right)\nu(d\eta)\mu(dx) \\
&= \sup_j \int_{\Theta\times\mathcal{X}} d(\eta,x)\left(1_{A_j}(\eta)\int_\Theta f_\zeta(x)\pi(d\zeta)\right)\nu(d\eta)\mu(dx). \qquad (39)
\end{aligned}$$

by monotone convergence. The term in parentheses in (39) 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 is $n_\gamma \in N$ such that*

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

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

1. *$M$ is a compact topological space and $\mathbb{T}(m, n)$ is upper semi-continuous in $m$ for each $n$, or*

2. *$N$ is a compact topological space and $\mathbb{T}(m, n)$ is lower semi-continuous in $n$ for each $m$, then*

$$\inf_{n\in N} \sup_{m\in M} \mathbb{T}(m, n) = \sup_{m\in M} \inf_{n\in N} \mathbb{T}(m, n). \tag{40}$$

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

$$\inf_{d\in\widetilde{\mathcal{D}}} \sup_{\pi\in\Pi} \mathbb{L}_\pi(d) = \sup_{\pi\in\Pi} \inf_{d\in\widetilde{\mathcal{D}}} \mathbb{L}_\pi(d). \tag{41}$$

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

$$\sup_{\pi\in\Pi} \mathbb{L}_\pi(d) = \mathbb{L}_\Theta(d). \tag{42}$$

□

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

**Proof of Lemma 2.** $\mathcal{D}_\alpha \subseteq \mathcal{D}$, 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. Recall that $d \in \mathcal{D}_\alpha$ iff for $\nu$-almost-every $\zeta \in \Theta$,

$$\begin{aligned} 1 - \alpha &\leq \mathbb{E}_\zeta d(\zeta, X) \\ &= \int_{\mathcal{X}} d(\zeta, x) f_\zeta(x)\mu(dx). \end{aligned} \tag{43}$$

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

$$C_A(d) \equiv \int_\Theta \int_\mathcal{X} d(\zeta, x) f_\zeta(x) \frac{1[\zeta \in A]}{\nu(A)} \mu(dx)\nu(d\zeta). \tag{44}$$

The function

$$(\zeta, x) \mapsto \frac{1}{\nu(A)} f_\zeta(x) 1[\zeta \in A] \tag{45}$$

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 $\nu(A) > 0$, $\{d \in \mathcal{D} : C_A(d) \geq 1 - \alpha\}$ is closed. But

$$\mathcal{D}_\alpha = \bigcap_{A:\nu(A)>0} \{d \in \mathcal{D} : C_A(d) \geq 1 - \alpha\}. \tag{46}$$

□

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

$$\inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\Theta(d) = \sup_{\pi \in \Pi} \inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\pi(d). \tag{47}$$

The mapping $\zeta \mapsto \lambda_\zeta$ is measurable, so $(\zeta, x) \mapsto d^\pi(\zeta, x)$ is product measurable. 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

$$\mathbb{E}_\pi d(\zeta, X) = \int d(\zeta, x) f_\pi(x) \mu(dx) \tag{48}$$

among $d(\zeta, \cdot) : \mathcal{X} \to [0, 1]$ satisfying $\mathbb{E}_\zeta d(\zeta, X) \geq 1 - \alpha$. Therefore $d^\pi$ minimizes $\int_\Theta \mathbb{E}_\pi d(\zeta, X)\nu(d\zeta)$ among $d \in \mathcal{D}_\alpha$, *i.e.*,

$$\mathbb{L}_\pi(d^\pi) = \inf_{d \in \mathcal{D}_\alpha} \mathbb{L}_\pi(d).$$

Corollary 1 follows. □

## 4.2 Theorem 2

Because $\{f_\theta : \theta \in \Theta\}$ has monotone likelihood ratios, $d^\eta$ has the form (20). The value of $c_\zeta$ is inconsequential because Lebesgue measure is continuous; take $c_\zeta \equiv 1$. Let $F(\cdot)$ be the cdf of $\mathbb{P}_0$. We calculate the risk at $\theta \in \Theta$ of the decision procedure $d^\eta$:

$$\begin{aligned} \mathbb{L}_\theta(d^\eta) &= \int_\Theta \int d^\eta(\zeta, x) f_\theta(x) dx d\zeta & (49) \\ &= \int_{-\tau}^{\eta} F(\zeta + q_{1-\alpha} - \theta) d\zeta + \\ &\quad + \int_\eta^\tau (1 - F(\zeta + q_\alpha - \theta)) d\zeta. & (50) \end{aligned}$$

Therefore,

$$\frac{d}{d\theta}\mathbb{L}_\theta(d^\eta) = \int_\eta^\tau f(\zeta + q_\alpha - \theta)d\zeta - \int_{-\tau}^\eta f(\zeta + q_{1-\alpha} - \theta)d\zeta \tag{51}$$

$$= \int_{q_\alpha+\eta}^{\tau+q_\alpha} f_\theta(\zeta)d\zeta - \int_{-\tau+q_{1-\alpha}}^{\eta+q_{1-\alpha}} f_\theta(\zeta)d\zeta \tag{52}$$

$$= \int h(\zeta)f_\theta(\zeta)d\zeta, \tag{53}$$

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

Now $h(\cdot)$ is a difference of indicators of intervals, so it has at most one strict sign change. The restriction $\tau + |\eta| \leq q_{1-\alpha} - q_\alpha$ implies that $q_\alpha + \eta \leq -\tau + q_{1-\alpha}$. Similarly, $\tau + |\eta| \leq q_{1-\alpha} - q_\alpha$ implies that $\tau + q_\alpha \leq \eta + q_{1-\alpha}$. 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

$$\theta \mapsto \int h(\zeta)f_\theta(\zeta)d\zeta = \frac{d}{d\theta}\mathbb{L}_\theta(d^\eta) \tag{54}$$

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

The definition of $\eta$ ((21)) ensures that $\frac{d}{d\theta}\mathbb{L}_\theta(d^\eta) = 0$ 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 $\pi_\eta$ (the point mass at $\{\eta\}$) is equal to the Bayes risk of $\pi_\eta$.

Suppose that $f_0$ is symmetric (so that $\eta = 0$ suffices) and that $\tau > 2q_{1-\alpha}$. We claim that 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 to consider:

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

6. $-\tau + z \le -z \le z \le \tau - z$

Case 1 (case 2, resp.) occurs if and only if $\tau \le z$ (iff $\tau \le 2z$), 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 $\alpha < 1/2$). 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.

**Lemma 3** *(More general version of a common result; see Lehmann and Casella [1998, Th. 1.4, p. 310].) Suppose $\pi \in \Pi$, the set of probability measures on $\Theta$. Let $\widetilde{\mathcal{D}}$ be a closed set of decisions. Let the risk at $\zeta$ of a decision $d \in \widetilde{\mathcal{D}}$ be $\mathbb{L}_\zeta(d)$, and let the Bayes risk of a decision $d \in \widetilde{\mathcal{D}}$ with respect to prior $\pi \in \Pi$ be*

$$\mathbb{L}_\pi(d) = \int_\Theta \mathbb{L}_\zeta(d)\pi(d\zeta). \tag{55}$$

*Suppose that $\widetilde{\mathcal{D}}$ is compact in a topology in which $d \to \mathbb{L}_\pi(d)$ is lower semi-continuous, for all $\pi$. Then for each $\pi \in \Pi$, $\widetilde{\mathcal{D}}$ contains at least one Bayes decision for prior $\pi$, $d^\pi \in \widetilde{\mathcal{D}}$:*

$$\mathbb{L}_\pi(d^\pi) = \inf_{d \in \widetilde{\mathcal{D}}} \mathbb{L}_\pi(d). \tag{56}$$

*Suppose $\lambda \in \Pi$ satisfies*

$$\mathbb{L}_\lambda(d^\lambda) = \sup_{\zeta \in \Theta} \mathbb{L}_\zeta(d^\lambda). \tag{57}$$

*Then $d^\lambda$ is minimax, and $\lambda$ is least favorable:*

$$\mathbb{L}_\lambda(d^\lambda) \ge \mathbb{L}_\pi(d^\pi) \;\; \forall \pi \in \Pi. \tag{58}$$

The proof of Lemma 3 is essentially that of Theorem 1.4 on p. 310 of Lehmann and Casella [1998]. It follows from Lemma 3 that the prior $\pi_\eta$ defined above is least favorable, and that $d^\eta$ is minimax. Equation (24) now follows from equation (50) and Theorem 1. □

## 4.3 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. This requires showing that the set $\mathcal{D}_i$ of scale-invariant decision functions is weak-star compact in $L_\infty[\nu \times \mu]$, when $\nu$ is Lebesgue measure on $\Theta = [-\tau, \tau]$ and $\mu$ is Lebesgue measure on $\mathcal{X} = \Re \times \Re^+$.

**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 $(\bar{X}, S)$, we observed the scale-invariant quantity $Z = (\bar{X} - \zeta)/S$. For any $\theta \in \Theta$ and $\sigma > 0$, the distribution of $Z$ is absolutely continuous with respect to Lebesgue measure $\lambda$ on $\Re$. We know that the set $\Delta$ of measurable decision functions based on $Z$ is weak-star compact in $L_\infty[\nu \times \lambda]$. Define

$$T: \quad \mathbf{R}^3 \to \mathbf{R}^2 \tag{59}$$

$$(\zeta, \bar{x}, s) \mapsto (\zeta, (\bar{x} - \zeta)/s). \tag{60}$$

Any scale-invariant decision function $d \in \mathcal{D}_i$ 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 $\Delta$ 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. Suppose $\delta_n(\zeta, z)$ is a sequence of elements of $\Delta$ such that

$$\int_{\Theta \times \Re} \delta_n(\zeta, z) h(\zeta, z) \nu(d\zeta) \lambda(dz) \to \int_{\Theta \times \Re} \delta(\zeta, z) h(\zeta, z) \nu(d\zeta) \lambda(dz) \tag{61}$$

for some $\delta(\zeta, z) \in \Delta$ and all $h \in L_1[\nu \times \lambda]$. We need to show that (61) implies that

$$\begin{aligned} \int_{\Theta \times \Re \times \Re^+} \delta_n \circ T(\zeta, \bar{x}, s) g(\zeta, \bar{x}, s) \nu(d\zeta) \mu(d\bar{x}, ds) &= \\ \int_{\Theta \times \Re \times \Re^+} \delta_n(\zeta, (\bar{x} - \zeta)/s) g(\zeta, \bar{x}, s) \nu(d\zeta) \mu(d\bar{x}, ds) &\to \\ \int_{\Theta \times \Re \times \Re^+} \delta(\zeta, (\bar{x} - \zeta)/s) g(\zeta, \bar{x}, s) \nu(d\zeta) \mu(d\bar{x}, ds) & \end{aligned} \tag{62}$$

for all $g \in L_1[\nu \times \mu]$.

For each $\zeta \in \Theta$, consider the bijective change of variables

$$(\bar{x}, s) \mapsto (z = (\bar{x} - \zeta)/s, s). \tag{63}$$

The Jacobian of this transformation is $s$, so

$$\begin{aligned}
\int_{\Theta\times\Re\times\Re^+} \delta(\zeta,(\bar{x}-\zeta)/s)g(\zeta,\bar{x},s)\nu(d\zeta)\mu(d\bar{x},ds) &= \\
\int_{\Theta\times\Re\times\Re^+} \delta(\zeta,z)g(\zeta,sz+\zeta,s)s\nu(d\zeta)\mu(dz,ds) &= \\
\int_{\Theta\times\Re} \delta(\zeta,z)\left(\int_{\Re^+} sg(\zeta,sz+\zeta,s)ds\right)\nu(d\zeta)\lambda(dz) &= \\
\int_{\Theta\times\Re} \delta(\zeta,z)h_g(\zeta,z)\lambda(dz)\nu(d\zeta), & \qquad (64)
\end{aligned}$$

where $h_g$ is given by

$$\begin{aligned}
h_g : \Theta\times\Re &\rightarrow \Re \\
(\zeta,z) &\mapsto h_g(\zeta,z) \equiv \int_{\Re^+} sg(\zeta,sz+\zeta,s)ds.
\end{aligned}$$

It follows as a special case (namely, $\delta \equiv 1$) that $h_g \in L_1[\nu\times\lambda]$, and thus that if $\delta_n \rightarrow \delta$ in the weak-star topology on $L_\infty[\nu\times\lambda]$ then $\delta_n \circ T \rightarrow \delta\circ T$ in the weak-star topology on $L_\infty[\nu\times\mu]$, as required. □

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

**Lemma 5** *If $\tau/\sigma \le 2u\sqrt{\frac{n-2}{n(n-1)}}$, then*

$$\begin{aligned}
\theta \mapsto & \frac{d}{d\theta}\mathbb{L}_{(\theta,\sigma)}(d_i^{\mathrm{TP}}) & (65) \\
&= \frac{d}{d\theta}\mathbb{E}_{(\theta,\sigma)}\int_{-\tau}^{\tau} d_i^{\mathrm{TP}}(\zeta,(\bar{X},S))d\zeta & (66)
\end{aligned}$$

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

Lemma 5 is proved in § 4.5.

## 4.4 Proof of Theorem 3.

Define $d_i^{\mathrm{TP}}$ as in Theorem 3. Let $\Pi$ be a set of probability measures as specified in § 3.2. For any $\pi\in\Pi$ and any fixed $\sigma\in\Sigma$, define

$$\mathbb{L}_{(\pi,\sigma)}(d) \equiv \int_\Theta \mathbb{L}_{(\zeta,\sigma)}(d)\pi(d\zeta). \qquad (67)$$

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

$$\inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\Theta,\sigma)}(d) = \sup_{\pi\in\Pi}\inf_{d\in\mathcal{D}_{\alpha,i}} \mathbb{L}_{(\pi,\sigma)}(d). \qquad (68)$$

For each $\zeta$, 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 $\sigma$, $d_i^{\mathrm{TP}}$ minimizes, among scale-invariant level $\alpha$ procedures, the expected confidence set Lebesgue measure when $\theta = 0$:

$$\inf_{d \in \mathcal{D}_{\alpha,i}} \mathbb{L}_0(d) = \mathbb{L}_0(d_i^{\mathrm{TP}}). \tag{69}$$

The procedure $d_i^{\mathrm{TP}}$ is thus a Bayes decision from $\mathcal{D}_{\alpha,i}$ for risk $\mathbb{L}_0$ and prior $\pi_0$, a point mass at 0. By lemma 5, if $\tau/\sigma \le 2t_{1-\alpha}\sqrt{\frac{n-2}{n(n-1)}}$ then the risk of $d_i^{\mathrm{TP}}$, $\mathbb{L}_{(\cdot,\sigma)}(d_i^{\mathrm{TP}})$, attains a global maximum at 0. The maximum risk of the Bayes procedure against $\pi_0$ is equal to the Bayes risk of $\pi_0$. It follows from Lemma 3 that $d_i^{\mathrm{TP}}$ is minimax. $\Box$

## 4.5 Proof of Lemma 5

Let $\tilde{\sigma} \equiv \sigma/\sqrt{n}$, $k \equiv n-1$, and $u = t_{1-\alpha}$. In terms of the value $(\bar{x}, s^2)$ of $X = (\bar{X}, S^2)$, the procedure $d_i^{\mathrm{TP}}$ is

$$d_i^{\mathrm{TP}}(\theta, (\bar{x}, s)) \equiv \begin{cases} 1[(\bar{x}-\theta)/s \le u], & \theta \le 0 \\ 1[(\bar{x}-\theta)/s \ge -u], & \theta > 0. \end{cases} \tag{70}$$

Fix $\tilde{\sigma}, \tau > 0$.

$$\begin{aligned}
\mathbb{L}_{(\theta,\sigma)}(d) &= \mathbb{E}_{(\theta,\sigma)} \int_{-\tau}^{\tau} d(\zeta, (\bar{X}, S)) d\zeta && (71)\\
&= \int_{-\tau}^{\tau} \mathbb{E}_{(\theta,\sigma)}[\mathbb{E}_{(\theta,\sigma)}(d(\zeta, (\bar{X}, S))|S)] d\zeta && (72)\\
&= \int_0^{\tau} \mathbb{E}_{(\theta,\sigma)} \mathbb{P}\{(\bar{X}-\zeta)/S \ge -u|S\} d\zeta + \\
&\quad + \int_{-\tau}^{0} \mathbb{E}_{(\theta,\sigma)} \mathbb{P}\{(\bar{X}-\zeta)/S \le u|S\} d\zeta && (73)\\
&= \int_0^{\tau} \mathbb{E}_{(\theta,\sigma)} \left[1 - \Phi\left(\frac{\zeta-\theta-Su}{\tilde{\sigma}}\right)\right] d\zeta + \\
&\quad + \int_{-\tau}^{0} \mathbb{E}_{(\theta,\sigma)} \left[\Phi\left(\frac{\zeta-\theta+Su}{\tilde{\sigma}}\right)\right] d\zeta. && (74)
\end{aligned}$$

Thus,

$$\frac{d}{d\theta}\mathbb{L}_{(\theta,\sigma)}(d) \propto \mathbb{E}_{(\theta,\sigma)} \left\{ \int_0^{\tau} \phi\left(\frac{\zeta-\theta-Su}{\tilde{\sigma}}\right) d\zeta - \right.$$

$$
\begin{aligned}
& - \int_{-\tau}^{0} \phi\left(\frac{\zeta-\theta+Su}{\tilde{\sigma}}\right) d\zeta \Big\} && (75)\\
= \;& \mathbb{E}_{(\theta,\sigma)} \left\{ \int_{-Su}^{\tau-Su} \phi\left(\frac{\zeta-\theta}{\tilde{\sigma}}\right) d\zeta - \right. \\
& \left. - \int_{-\tau+Su}^{Su} \phi\left(\frac{\zeta-\theta}{\tilde{\sigma}}\right) d\zeta \right\} && (76)\\
\propto \;& \mathbb{E}_{(\theta,\sigma)} g(\bar{X}, S), && (77)
\end{aligned}
$$

where

$$
g(\bar{x}, s) \equiv 1\left[\frac{-\bar{x}}{u} \le s \le \frac{\tau-\bar{x}}{u}\right] - 1\left[\frac{\bar{x}}{u} \le s \le \frac{\bar{x}+\tau}{u}\right]. \tag{78}
$$

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

$$
\frac{d}{d\theta}\mathbb{L}_\theta(d) \propto \int_{\Re} \mathbb{E}_{(\theta,\sigma)}(g(\bar{X}, S)|\bar{X} = \bar{x}) \phi\left(\frac{\bar{x}-\theta}{\tilde{\sigma}}\right) d\bar{x}. \tag{79}
$$

Because the normal density is totally positive, the number of sign changes of $\theta \mapsto \frac{d}{d\theta}\mathbb{L}_{(\theta,\sigma)}(d)$ is no larger than the number of sign changes of a version of $\bar{x} \mapsto \mathbb{E}_{(\theta,\sigma)}(g(\bar{X}, S)|\bar{X} = \bar{x})$.

One version is $\bar{x} \mapsto Ch(\bar{x})$, where $C$ is a constant that depends on $\sigma$, $\alpha$, and $k$, but not on $\theta$ or $\bar{x}$:

$$
\begin{aligned}
h(\bar{x}) \;&\equiv\; \int_0^\infty g(\bar{x}, s) s^{k-1} e^{-ks^2/(2\tilde{\sigma}^2)} ds && (80)\\
&=\; \begin{cases} -h(-\bar{x}), & \bar{x} \le 0 \\ \int_0^{(\tau-\bar{x})/u} s^{k-1} e^{-ks^2/(2\tilde{\sigma}^2)} ds - \\ \int_{\bar{x}/u}^{(\tau+\bar{x})/u} s^{k-1} e^{-ks^2/(2\tilde{\sigma}^2)} ds, & 0 < \bar{x} \le \tau \\ -\int_{\bar{x}/u}^{(\tau+\bar{x})/u} s^{k-1} e^{-ks^2/(2\tilde{\sigma}^2)} ds, & \bar{x} > \tau \end{cases} && (81)
\end{aligned}
$$

(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 $\bar{x} \ge \tau$, and $h'(\tau) > 0$;

(vi) If $\tau/\tilde{\sigma} \le 2u\sqrt{\frac{k-1}{k}}$, then $h'$ takes the value 0 at most once on $[0, \tau]$.

Claims (i)–(v) are clear upon inspection of (81); (vi) is discussed below. Together (i)–(vi) imply that $h$ changes sign once as $\bar{x}$ ranges from $-\infty$ to $\infty$, going from positive to negative as $\bar{x}$ increases through 0. Total positivity and (79) imply that $\theta \mapsto \frac{d}{d\theta}\mathbb{L}_{(\theta,\sigma)}(d)$ 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 $[0, \tau]$,

$$\begin{aligned} h'(\bar{x}) \propto\ & -(\bar{x}+\tau)^{k-1}e^{-\frac{C(\bar{x}+\tau)^2}{2}} - (\tau-\bar{x})^{k-1}e^{-\frac{C(\tau-\bar{x})^2}{2}} + \\ & +\bar{x}^{k-1}e^{-\frac{C\bar{x}^2}{2}} && (82)\\ =\ & \bar{x}^{k-1}e^{-\frac{C\bar{x}^2}{2}}[-e^{-\frac{C\tau^2}{2}}(h_1(\bar{x}/\tau)+h_2(\bar{x}/\tau))+1], && (83)\end{aligned}$$

where

$$\begin{aligned} h_1(\zeta) &= (1+1/\zeta)^{k-1}e^{-C\tau^2\zeta}, \\ h_2(\zeta) &= (1/\zeta-1)^{k-1}e^{-C\tau^2\zeta}, \text{ and} \\ C &= \frac{k}{\tilde{\sigma}^2u^2}. && (84)\end{aligned}$$

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

$$h_2'(\zeta) = (k-1)(1/\zeta-1)e^{-C\tau^2\zeta}\underbrace{\left(-1/\zeta^2 + \frac{k}{k-1}\frac{\tau^2}{\tilde{\sigma}^2u^2}1/\zeta - \frac{k}{k-1}\frac{\tau^2}{\tilde{\sigma}^2u^2}\right)}_{*}. \tag{85}$$

By viewing (*) as a quadratic function of $1/\zeta$, one sees that it has a zero on $(0, \infty)$ iff $\frac{k}{k-1}\frac{\tau^2}{\tilde{\sigma}^2u^2} > 4$. Otherwise, it does not change sign on the positive half line. Evidently it must be negative as $\zeta \to \infty$, so if it does not change sign on $(0, \infty)$, it must be nonpositive on that interval. It follows that $h_2'$ must be nonpositive on $(0, 1)$, provided $\tau/\tilde{\sigma} \leq 2\sqrt{(k-1)/k}u$. But then $h_2$ is nonincreasing, forcing $h_1 + h_2$ to be strictly decreasing, and implying in turn by (83) that $h$ has property (vi) above. □

# A Minimax fixed-length confidence intervals

Zeytinoglu and Mintz [Zeytinoglu and Mintz, 1984, 1988] study the problem of determining 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 noncoverage 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 $Z \sim \mathcal{N}(\theta, 1)$, $\theta \in [-\tau, \tau]$. According to Zeytinoglu and Mintz [1984], if $l/2 < \tau \leq l$, then the minimax-noncoverage interval of fixed length $l$ is centered at

$$\hat{\theta}(Z) = \begin{cases} Z, & |Z| \leq \tau - l/2 \\ \tau - l/2, & |Z| > \tau - l/2 \end{cases} \tag{86}$$

and has maximum noncoverage probability $\Phi(-l/2)$ [p. 949]. If $l < \tau \leq 3l/2$, then the minimax-noncoverage interval of fixed length $l$ is centered at

$$\hat{\theta}(Z) = \begin{cases} 0, & |Z| < a \\ Z - a, & a \leq |Z| < a + l \\ l, & a + l \leq |Z|, \end{cases} \tag{87}$$

where $a$ is the solution of

$$2\Phi(-a - l/2) = \Phi(a - l/2). \tag{88}$$

In this case, the maximum noncoverage probability is $\Phi(a-l/2)$ [Zeytinoglu and Mintz, 1984, p. 948].

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

INSERT TABLE 3

Table 3 shows that if $\tau \in [1.6, 3.25]$, the optimal fixed-width 95% interval is centered at a point $\hat{\theta}$ of form (86) and has width between 3.25 and 3.30. Since intervals of this form have maximum noncoverage chance $\Phi(-l/2)$, the minimax-width 95% interval has width precisely $2z_{.95} \approx 3.28$.

If $\tau \in [3.6, 5.4]$, an interval of width 3.60 centered at a point of form (87) has 95% coverage. This minimax-width fixed-width 95% confidence interval is given by (87), with $a = 0.158$.

If $\tau \in [3.30, 3.60)$, then no interval with centering point given by (87) has sufficient uniform coverage probability. To get a 95% confidence interval one must center it at a point of form (86). This means $\tau \in (l/2, l]$, implying that $l \geq \tau$. The maximum noncoverage at $l = 3.30$ falls under the 5% cutoff,

so $l$ need be no larger than $\tau$. It thus turns out that for $\tau \in [3.30, 3.60)$ the maximum noncoverage 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.
**Acknowledgments.** We are grateful to D.A. Freedman for advice, direction, and comments on an earlier draft, and to R. Purves for helpful conversations.

## References

J.C. Berry. Minimax estimation of a bounded normal mean vector. *J. Multivariate Analysis*, 35:130–139, 1990.

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

W. Bischoff and W. Fieger. Minimax estimators and $\gamma$-minimax estimators for a bounded normal mean under the loss $l_p(\theta, d) = |\theta - d|^p$. *Metrika*, 39:185–197, 1992.

Lawrence David Brown. On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Stat.*, 37:1087–1136, 1966.

L.D Brown, G. Casella, and J.T.G. Hwang. Optimal confidence sets, bioequivalence, and the limacon of Pascal. *J. Am. Stat. Assoc.*, 90:880–889, 1995.

G. Casella. Comment on "Setting confidence intervals for bounded parameters" by M. Mandelkern. *Statistical Science*, 17(2):159–160, 2002.

G. Casella and J.T. Hwang. Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Am. Stat. Assoc.*, 78:688–698, 1983.

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

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

D.L. Donoho. Statistical estimation and optimal recovery. *Ann. Stat.*, 22: 238–270, 1994.

D.L. Donoho and R.C. Liu. Geometrizing rates of convergence. iii. *Ann. Stat.*, 19:668–701, 1991.

D.L. Donoho and M. Nussbaum. Minimax quadratic estimation of a quadratic functional. *J. Complexity*, 6:290–323, 1990.

J. Eichenauer-Herrman and K. Ickstadt. Minimax estimators for a bounded location parameter. *Metrika*, 39:227–237, 1992.

J. Fan and I. Gijbels. Minimax estimation of a bounded squared mean. *Statistics & Probability Letters*, 13:383–390, 1992.

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

J.K. Ghosh. On the relation among shortest confidence intervals of different types. *Calcutta Statistical Association Bulletin*, pages 147–152, 1961.

L.J. Gleser. Comment on "Setting confidence intervals for bounded parameters" by M. Mandelkern. *Statistical Science*, 17(2):160–163, 2002.

E. Gourdin, B. Jaumard, and B. MacGibbon. Global optimization decomposition methods for bounded parameter minimax risk evaluation. *SIAM J. Sci. Stat. Comput.*, 15:16–35, 1994.

Peter M. Hooper. Invariant confidence sets with smallest expected measure. *Ann. Stat.*, 10(4):1283–1294, 1982.

Peter M. Hooper. Correction. *Ann. Stat.*, 12(2):784, 1984.

Jiunn Tzon Hwang and George Casella. Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Stat.*, 10:868–881, 1982.

I. A. Ibragimov. On the composition of unimodal distributions. *Teor. Veroyatnost. i Primenen.*, 1:283–288, 1956.

I.A. Ibragimov and R.Z. Khas'minskii. On nonparametric estimation of the value of a linear functional in gaussian white noise. *Th. Prob. and its Appl. (transl. of Teorija Verojatnostei i ee Primenenija)*, 29:18–32, 1984. Durri-Hamdani (transl).

I.M. Johnstone and K.B. MacGibbon. Minimax estimation of a constrained Poisson vector. *Ann. Stat.*, 20:807–831, 1992.

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

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

G. Kamberova, R. Mandelbaum, and M. Mintz. Statistical decision theory for mobile robotics: theory and application. In *Proc. IEEE/SICE/RSJ Int. Conf. On Multisensor Fusion and Integration for Intelligent Systems, MFI '96*, 1996.

G. Kamberova and M. Mintz. Minimax rules under zero-one loss for a restricted parameter. *J. Statistical Planning and Inference*, 79:205–221, 1999.

S. Karlin. *Total Positivity*, volume 1. Stanford Univ. Press, Stanford, 1968.

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

E.L. Lehmann. *Testing Statistical Hypotheses*. John Wiley and Sons, New York, 2nd edition, 1986.

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

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

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

E. Marchand and F. Perron. Improving on the MLE of a bounded normal mean. *Ann. Stat.*, 29(4):1078–1093, 2001.

J.W. Pratt. Length of confidence intervals. *J. Am. Stat. Assoc.*, 56:549–567, 1961.

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

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

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

D.A. van Dyk. Comment on "Setting confidence intervals for bounded parameters" by M. Mandelkern. *Statistical Science*, 17(2):164–168, 2002.

B. Vidakovic and A. Dasgupta. Efficiency of linear rules for estimating a bounded normal mean. *Sankhyā*, 58:81–100, 1996.

L. Wasserman. Comment on "Setting confidence intervals for bounded parameters" by M. Mandelkern. *Statistical Science*, 17(2):163, 2002.

L. Weiss. Estimating multivariate normal means using a class of bounded loss functions. *Statistics & Decisions*, 6:203–207, 1988.

M. Woodroofe and T. Zhang. Comment on "Setting confidence intervals for bounded parameters" by M. Mandelkern. *Statistical Science*, 17(2): 168–171, 2002.

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

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

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

| $\tau$ | 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 |

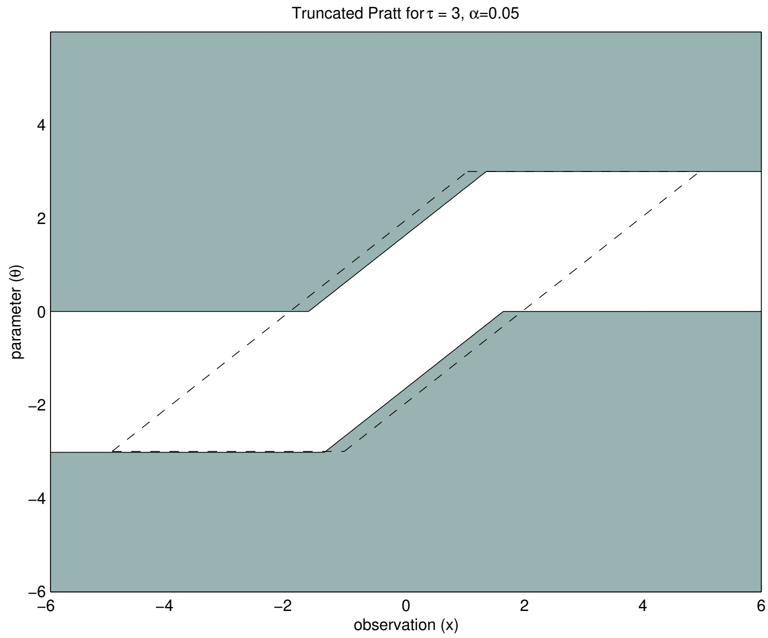
[a] Affine fixed-width intervals have the form $[aX + b - e, aX + b + e]$, with $a$, $b$, and $e$ constant.

[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 $S \subseteq \Theta \times \mathcal{X}$ is product-measurable.

[d] The measurable 95% confidence set with smallest expected measure when $\tau \leq 3.29$ is the truncated Pratt interval $\mathcal{I}_{\mathrm{TP}}$. The entries in the rightmost column for $\tau = 3.50, 3.75,$ and 4.00 are the maximum expected lengths of optimal confidence sets $\mathcal{I}_{\mathrm{OPT}}$, approximated numerically.

Figure 1: The truncated Pratt procedure for the BNM with $\sigma = 1$, $\tau = 3$, $\alpha = 0.05$. Viewed as a confidence interval or as a family of tests, this truncated Pratt (3) decision rule improves on the conventional procedure (1), by having smaller expected length or more power.
[a]




[a]Unshaded parts of vertical slices represent sample truncated Pratt confidence sets $\mathcal{I}_{\mathrm{TP}}(X)$; unshaded parts of horizontal slices represent truncated Pratt acceptance regions. The dashed lines show the endpoints of sample confidence sets and acceptance regions for the truncated conventional procedure.

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

| $1-\alpha$ | $\tau$ | $\mathbb{E}_0$ $\mu(\mathcal{I}_{\mathrm{TP}}(X))$[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 *[d] | 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] $\mu$ denotes Lebesgue measure. $\mathbb{E}_0\mu(\mathcal{I}_{\mathrm{TP}}(X))$ is the Bayes risk of $d^0$ (the Bayes decision rule for a prior that concentrates at zero), or $\mathcal{I}_{\mathrm{TP}}$. Iff $\tau \le 2z_{1-\alpha}$, this is the worst-case risk of $d^0$.

[b] $\mathcal{I}_{\mathrm{T}}$, the truncated conventional interval, is defined in (1).

[c] $\mathcal{I}_{\mathrm{A}}$, the minimax affine fixed-length interval, was determined and analyzed numerically by the method of Stark [1992].

[d] $*$ indicates that $\mathcal{I}_{\mathrm{TP}}$ is optimal.

Table 3: Minimax-noncoverage fixed-length intervals: maximum noncoverage probabilities and offset constants $a$.

| $l/2 < \tau \leq l$ | | $l < \tau \leq 3l/2$ | | |
|---|---|---|---|---|
| $l^a$ | $p^b$ | $l$ | $a^c$ | $p^d$ |
| 3.00 | .067 | | | |
| 3.05 | .064 | | | |
| 3.10 | .061 | | | |
| 3.15 | .058 | | | |
| 3.20 | .055 | 3.20 | .171 | .077 |
| 3.25 | .052 | 3.25 | .169 | .073 |
| 3.30 | .049 | 3.30 | .168 | .069 |
| 3.35 | .047 | 3.35 | .166 | .066 |
| 3.40 | .045 | 3.40 | .164 | .062 |
| 3.45 | .042 | 3.45 | .163 | .059 |
| 3.50 | .040 | 3.50 | .161 | .056 |
| 3.55 | .038 | 3.55 | .159 | .053 |
| 3.60 | .036 | 3.60 | .158 | .050 |
| | | 3.65 | .156 | .048 |
| | | 3.70 | .155 | .045 |
| | | 3.75 | .153 | .043 |
| | | 3.80 | .152 | .040 |

[a] Length of confidence interval.

[b] Maximum noncoverage probability of minimax-noncoverage 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 (86).

[c] When $\tau \in (l, 3l/2]$, $a$ combines with (87) to specify the minimax-noncoverage interval of length $l$.

[d] Maximum noncoverage probability of minimax-noncoverage 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 (87).