Using What We Know: Inference with Physical Constraints

Chad M. Schafer and Philip B. Stark
Department of Statistics, University of California, Berkeley, CA 94720, USA

Frequently physical scientists seek a confidence set for a parameter whose precise value is unknown, but constrained by theory or previous experiments. The confidence set should exclude parameter values that violate those constraints, but further improvements are possible: We construct *minimax expected size* and *minimax regret* confidence procedures. The resulting confidence sets include only values that satisfy the constraints; they have the correct coverage probability; and they minimize a measure of average size. We illustrate these approaches with three examples: estimating the mean of a normal distribution when this mean is known to be bounded, estimating a parameter of a bivariate normal distribution arising in a signal detection problem, and estimating cosmological parameters from MAXIMA-1 observations of the cosmic microwave background radiation. In the first two examples, the new methods are compared with two others: a standard approach adapted to force the estimate to conform to the bounds, and the likelihood-ratio testing approach proposed by Feldman and Cousins [1998]. Software that implements the new method efficiently is available online.

1. INTRODUCTION

In many statistical estimation problems parameters are just indices of stochastic models, but in the physical sciences parameters are often physical constants whose values have scientific interest. Previous experiments, theory and physical constraints often limit the possible or plausible values of unknown constants. In cosmology, for example, decades of observation and theoretical research have led to wide agreement on the range of possible values for key cosmological parameters, such as the Hubble constant and the age of the Universe. A good statistical method should use everything we know—data and physical constraints—to make inferences as sharp as possible. This paper looks at the problem of incorporating prior constraints into confidence sets from a frequentist perspective.

There is a duality between hypothesis tests and confidence sets. Suppose that Θ is the set of possible values of the parameter θ (either a scalar or a vector), and let η denote a generic element of Θ. Let $A(\eta)$ be an *acceptance region* for testing the hypothesis that $\theta = \eta$. If the data, a realization of the random variable X, fall within $A(\eta)$, we consider $\theta = \eta$ an adequate explanation of the data, while if the data fall outside $A(\eta)$, we reject the hypothesis $\theta = \eta$. The chance when $\theta = \eta$ that the data fall outside $A(\eta)$ is the probability of *type I error*—the significance level—of the test.

Suppose we have a family of acceptance regions $\{A(\eta) : \eta \in \Theta\}$, each with significance level at most α; that is,

$$P_\eta\{X \notin A(\eta)\} \leq \alpha, \ \ \forall \eta \in \Theta. \tag{1}$$

Then the set

$$C_A(x) \equiv \{\eta \in \Theta : x \in A(\eta)\} \tag{2}$$

is a confidence procedure for θ with *confidence level* at least $1 - \alpha$. That is,

$$P_\theta\{C_A(X) \ni \theta\} \geq 1 - \alpha, \ \ \forall \theta \in \Theta. \tag{3}$$

Tailoring the acceptance regions $\{A(\eta)\}$ lets us control properties of the resulting confidence set.

For example, we might want the confidence set to include the smallest possible range of parameter values. That would lead us to pick $A(\eta)$ to minimize the probability when $\theta \neq \eta$ that $X \in A(\eta)$, (the probability of *type II error*). It is generally not possible to minimize these false coverage probabilities simultaneously over all $\theta \in \Theta$. The constraint $\theta \in \Theta$ avoids tradeoffs in favor of impossible models.

Incorporating bounds is simple with Bayesian methods: Use a prior that assigns probability one to the set Θ. However, any prior does more than impose the constraint $\theta \in \Theta$: It also assigns probabilities to all measurable subsets of Θ. In problems with infinite-dimensional parameters, it can be impossible to find a prior that honors the physical constraints [Backus 1987, 1988].

1.1. Expected Size of Confidence Regions as Risk

We want a confidence procedure to produce sets that are as small (accurate) as possible, but still to have coverage probability $1-\alpha$, no matter what value θ has, provided it is in Θ. To quantify size, we use an arbitrary measure ν on Θ (typically ν is ordinary volume—Lebesgue measure). We study how the expected size of the region depends on the true value of the parameter θ. This embeds our problem in statistical decision theory: We compare estimators based on their *risk functions* over $\theta \in \Theta$, where risk is the expected measure of the confidence region.

It is rare that one procedure minimizes the expected size for every $\theta \in \Theta$. (Such procedures are *uniformly most accurate* (UMA) confidence procedures. See Schervish [1995], for example.) Making the expected size small for one value of θ tends to make it larger for other values, so minimizing the expected size for $\theta \notin \Theta$ tends to make the expected size unnec-

essarily large for some values of $\theta \in \Theta$. We seek the *minimax expected size* (MES) confidence procedure: the procedure that minimizes the maximum expected size for parameter values θ that are members of Θ, the set of possible theories. Thus, the parameter constraint $\theta \in \Theta$ enters in two ways: The confidence region includes only values in Θ, and the expected size is considered only for $\theta \in \Theta$.

MES is the inversion of a family of hypothesis tests that are most powerful against a *least favorable alternative* (LFA), a mixture of theories $\{P_\eta : \eta \in \Theta\}$; those tests are based on likelihood ratios. Evans et al. [2003] establish in some generality that MES is of this form. (Often in decision theory the minimax procedure is the Bayes procedure for the prior that yields the largest Bayes risk.) Typically, we can only approximate MES numerically.

Forming confidence regions by inverting hypothesis tests based on likelihood ratios is not new in the physical sciences. For example, Feldman and Cousins [1998] construct confidence intervals by inverting the likelihood ratio test (LRT). (See Bickel and Doksum [1977], Lehmann [1986], and Schervish [1995] for discussions of LRT.) MES has two advantages over LRT: First, it is optimal in a sense that clearly measures accuracy. Second, although approximating the LFA can be challenging, performing the likelihood ratio test in complex situations can be even more difficult because one must calculate the restricted MLE for all possible data.

On the other hand, LRT has an appealing invariance under reparametrization: The LRT confidence set for a transformation of a parameter is just the same transformation applied to the LRT confidence set for the original parameter. In contrast, a transformation of the MES confidence set is a confidence set for the transformation of the parameter, but typically it is not the same set as the MES confidence set designed for the transformed parameter—it has larger (maximum) expected measure. Bayesian credible regions based on "uninformative" priors also lack this kind of invariance, because a prior that is flat in one parametrization is not flat after a non-affine reparametrization. (None of these methods necessarily produces a confidence *interval* under reparametrizations. For example, a confidence interval for θ^2 that does not include zero would transform to a confidence set for θ that is the union of two disjoint intervals.)

Any procedure that has $1-\alpha$ coverage probability for all $\eta \in \Theta$ has strictly positive expected measure for all $\theta \in \Theta$. Let $r(\theta)$ be the infimum of the risks at θ of all $1-\alpha$ confidence procedures. The *regret* of a confidence procedure at the point θ is the difference between $r(\theta)$ and the risk at θ [DeGroot 1988]. MES is the $1-\alpha$ procedure whose supremal risk over $\eta \in \Theta$ is as small as possible. In contrast, the *minimax regret* procedure (MR) is the $1-\alpha$ confidence procedure for which the supremum of the regret is smallest. MR can be constructed in much the same was as MES, by finding a *least regrettable alternative* (LRA). MES and MR can be quite different, as illustrated in section 2.

The next section gives two simple examples demonstrating MES and MR, and contrasting them with a classical approach and LRT. Section 3 sketches the theory behind MES and MR in more detail. Section 4 applies the approaches to a more complicated problem: estimating cosmological parameters from observations of the cosmic microwave background radiation (CMB). Section 5 describes a computer algorithm for approximating MES and MR in complex problems such as the CMB problem.

2. SIMPLE EXAMPLES

2.1. The Bounded Normal Mean Problem

We observe a random variable X that is normally distributed with mean θ and variance one. We know *a priori* that $\theta \in [-\tau, \tau] = \Theta$. We seek a confidence interval for θ. Evans et al. [2003] discuss this problem in detail, and characterize the MES procedure. Compare the following three approaches:

1. **Truncating the standard confidence interval.** Let z_p be the p^{th} percentile of the standard normal distribution. A simple approach that honors the restriction $\theta \in [-\tau, \tau]$ is to intersect the usual confidence interval $[X - z_{1-\alpha/2}, X + z_{1-\alpha/2}]$ with $[-\tau, \tau]$. The resulting confidence interval corresponds to inverting hypothesis tests whose acceptance regions are

$$A_{\rm TS}(\eta) = \left[\eta - z_{1-\alpha/2},\ \eta + z_{1-\alpha/2}\right] \qquad (4)$$

for $\eta \in [-\tau, \tau]$. This is an intuitively attractive solution, and it is the only *unbiased* procedure: The parameter value that the interval is most likely to cover is the true value θ. However, some biased procedures have smaller maximum expected length.

2. **Inverting the likelihood ratio test.** Let $\hat{\theta}$ denote the restricted maximum likelihood estimate of θ: the parameter value in Θ for which the likelihood is greatest, given data $X = x$. Acceptance regions for the likelihood ratio test are formed by setting a threshold k_η for the ratio of the likelihood of the parameter η and the likelihood of $\hat{\theta}$; the hypothesis $\theta = \eta$ is rejected if the ratio is too small. The threshold is chosen so that when $\theta = \eta$, the probability that $X \in A(\eta)$ is at least $1-\alpha$. Thus,

$$A_{\rm LRT}(\eta) = \left\{ x \colon \frac{\phi(x-\eta)}{\phi(x-\hat{\theta})} \geq k_\eta \right\}, \qquad (5)$$

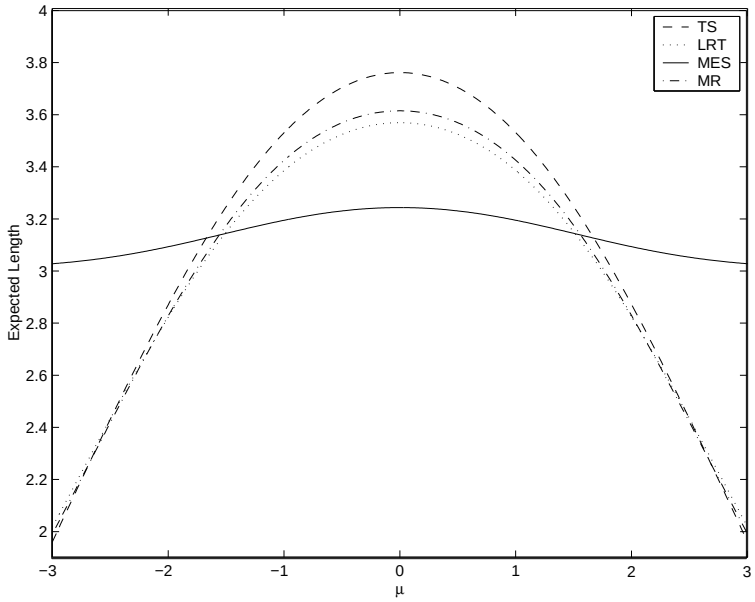

Figure 1: Expected lengths of the 95% confidence intervals for a bounded normal mean as a function of the true value θ for $\tau = 3$.

where $\phi(\cdot)$ is the standard normal density function:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-(z)^2/2\right) \tag{6}$$

and

$$\hat{\theta} = \begin{cases} -\tau, & x \le -\tau \\ x, & -\tau < x < \tau \\ \tau, & x \ge \tau \end{cases} . \tag{7}$$

3. **Minimax expected size procedure.** Both MES and LRT are based on inverting tests involving likelihood ratios, but the likelihoods in the denominator (the alternative hypotheses) are different. The MES acceptance regions are

$$A_{\mathrm{MES}}(\eta) = \left\{ x \colon \frac{\phi(x-\eta)}{\int_{-\tau}^{\tau} \phi(x-u)\lambda(du)} \ge c_\eta \right\}, \tag{8}$$

where λ is the least favorable alternative and c_η is chosen so that the coverage probability is $1-\alpha$.

Table I Maximum expected lengths of three 95% confidence procedures for estimating the mean of a normal distribution when the mean is known to be in the interval $[-\tau, \tau]$. TS is the truncated standard procedure, LRT is the inversion of the likelihood ratio test (the Feldman-Cousins approach), and MES is the minimax expected size procedure.

τ	**TS**	**LRT**	**MES**
1.75	2.9	2.7	2.6
2.00	3.2	2.9	2.8
2.25	3.4	3.1	3.0
2.50	3.6	3.3	3.1
2.75	3.7	3.5	3.2
3.00	3.8	3.6	3.2
3.25	3.8	3.7	3.3
3.50	3.9	3.7	3.3
3.75	3.9	3.8	3.4
4.00	3.9	3.8	3.4

Table I lists the maximum expected sizes for each of these three procedures for several values of the bound τ. The advantage of MES is larger when τ is larger. Figure 1 compares the expected lengths of the intervals as a function of the true value θ for $\tau = 3$. The MES procedure attains smaller expected length at $\theta = 0$ at the cost of larger expected length when $|\theta|$ is large. When $\tau \le 2z_{1-\alpha}$, as it is here, the MES procedure minimizes the expected size at $\theta = 0$ (equivalently, the regret of MES at $\theta = 0$ is zero: λ assigns probability one to $\theta = 0$). The MES interval in the bounded normal mean problem is a truncated version of the confidence interval proposed by Pratt [1961] for estimating an unrestricted normal mean. Figure 1 also shows the expected size of the MR interval. The expected size at zero is larger for MR than for MES, but that increase is offset by large decreases in expected size for large $|\theta|$. None of the methods dominates the rest for all $\theta \in \Theta$; other considerations are needed to choose among them.

Table II 95% confidence intervals for $\sin 2\beta$ using each of the four methods.

Method	Lower	Upper
TS	-0.07	1.00
LRT	-0.07	1.00
MES	0.00	1.00
MR	-0.08	1.00

The bounded normal mean problem arises in particle physics: Affolder et al. [2000] estimate the violation of charge-conjugation parity (CP) using observations of proton-antiproton collisions in the CDF detector at Fermilab. The parameter that measures CP violation is called $\sin 2\beta$, which must be in the interval $[-1, 1]$. In the model Affolder et al. [2000] use, the MLE of $\sin 2\beta$ has a Gaussian distribution with mean $\sin 2\beta$ and standard deviation 0.44. This standard deviation captures both systematic and random error in the estimate. This is equivalent to the situation described above, with $\tau = 1.0/0.44 \approx 2.27$. The observed measurement was 0.79, and the 95% confidence intervals for $\sin 2\beta$ are shown in table II. These results illustrate the strange behavior of MES in some cases: Since the LFA concentrates its mass on zero, the interval will always include a parameter value arbitrarily close to zero. Figure 2 compares acceptance regions and intervals for the four methods in this case. Note that for MES, the acceptance regions always extend to either $-\infty$ or $+\infty$.

2.2. Estimating a Function of Two Normal Means: An Example in Psychophysics

Suppose $\{X_{ij}\}_{i=1}^{2}{}_{j=1}^{n_j}$ are independent, normally distributed with variance one, that the expected values of $\{X_{1j}\}$ are all μ_1 and that the expected values of $\{X_{2j}\}$ are all μ_2. We know *a priori* that $-b \le \mu_2 \le \mu_1 \le b$. The goal is to estimate the two parameters $\theta_1 = 0.5(\mu_1 + \mu_2)$ and $\theta_2 = \mu_1 - \mu_2$ from observing the signs of $\{X_{ij}\}$. Thus,

$$\Theta \equiv \{(\eta_1, \eta_2) \in \Re^2 : -b \le \eta_1 - \eta_2/2 \le \eta_1 + \eta_2/2 \le b\}. \quad (9)$$

Let

$$Y_i \equiv \sum_j 1_{\{X_{ij} \ge 0\}} \quad i = 1, 2. \quad (10)$$

These observable variables are sufficient statistics for the signs of $\{X_{ij}\}$; they are independent; and Y_i has the binomial$(n_i, p_i \equiv \Phi(\mu_i))$ distribution, where Φ is the standard normal cumulative distribution function. We call (p_1, p_2) the *canonical parameters* because of their simple relationship to the distribution of the observations.

This is a stylized version of an estimation problem in signal detection theory [Miller 1996, Kadlec 1999]. A subject is presented with a randomized sequence of noisy auditory stimuli, and is asked to discern which stimuli contain "signal," and which are only "noise." In the standard model, the subject is assumed to have an internal scoring mechanism that assigns a number to each stimulus. If the number is positive, the subject reports that the stimulus contains signal; otherwise, the subject reports that the stimulus is just noise. Moreover, according to the model, scores for different stimuli are independent normal random variables with variance one.

For stimuli that consist of signal and noise, the expected scores are all equal to μ_1, while for stimuli that contain just noise, the expected scores are all equal to μ_2. The quantity of greatest interest is θ_2, the difference between these means, denoted d' in the psychophysics literature. It is a measure of the distance between the distributions of scores with and without noise, and (indirectly) provides an upper bound on the accuracy of signal detection. Of secondary interest is θ_1, the midpoint of the two means, which measures the "bias" in the decision rule: When $\theta_1 = 0$, so that $\mu_1 = -\mu_2$, the chance of claiming that the stimulus contains signal when it does not is equal to the chance of claiming that the signal is just noise when it contains signal. When $\theta_1 > 0$, the subject is biased in favor of claiming that the stimulus contains signal; when $\theta_1 < 0$, the subject is biased in favor of claiming that signal is not present. The restriction $-b \le \mu_2 \le \mu_1 \le b$ derives from the assumption that $\epsilon \le p_2 \le p_1 \le 1 - \epsilon$: The subject is more likely to report that signal is present when it is in fact present, and the subject has a strictly positive chance of misclassifying both types of stimuli. The constraints are related through $b = \Phi^{-1}(1 - \epsilon)$.

2.3. Confidence Regions for (θ_1, θ_2)

We compare methods for obtaining a $1 - \alpha$ confidence region for the parameter vector (θ_1, θ_2). Starting with a "good" confidence region for (p_1, p_2) and then finding its preimage in (θ_1, θ_2) space tends to produce unnecessarily large confidence regions for (θ_1, θ_2) because of the nonlinear relationship between these parametrizations. This distinction between the canonical parameters and the parameters of interest is crucial: We want the confidence region for models to constrain the values of the parameters of interest as well as possible. Whether that region corresponds to a small set of canonical parameters is unimportant.

The first approach we consider is based on the normal approximation to the distribution of the maximum likelihood estimator (MLE). For large enough samples, the MLE is approximately normally distributed with mean $\theta = (\theta_1, \theta_2)$ and covariance matrix

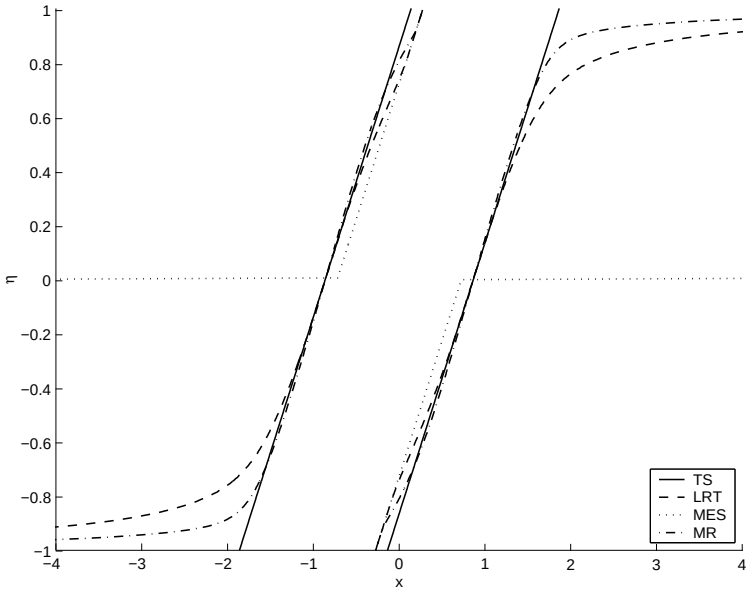

Figure 2: A depiction of 95% confidence intervals for an application of the bounded normal mean problem, the estimation CP violation parameter $\sin 2\beta$. Read across to see the acceptance region $A(\eta)$ for each of the four confidence procedures. Vertical sections are confidence intervals for different data values.

$\mathbf{I}^{-1}(\theta)$, where $\mathbf{I}(\theta)$ is the *Fisher information matrix* [Bickel and Doksum 1977]. In this case,

$$\mathbf{I}(\theta_1, \theta_2) = \begin{bmatrix} w_1 + w_2 & 0.5(w_1 - w_2) \\ 0.5(w_1 - w_2) & 0.25(w_1 + w_2) \end{bmatrix}, \quad (11)$$

where

$$w_i \equiv \frac{n_i \phi^2(\mu_i)}{p_i(1-p_i)}, \quad i = 1, 2, \quad (12)$$

and ϕ is the standard normal density. We can use this asymptotic distribution and the constraint to construct an approximate confidence region for (θ_1, θ_2) by intersecting Θ with an ellipse centered at the MLE. The light gray truncated ellipse in Figure 3 is an approximate 95% confidence region formed using this method. In this case, $n_1 = n_2 = 10$, the observed data are $y_1 = 8$ and $y_2 = 4$, and the bound b is $\Phi^{-1}(.99)$.

Figure 3 also illustrates the MES confidence region. The regular grid of points is the set of parameter values tested; those accepted are plotted as larger dots than those rejected. The MES region is the convex hull of the accepted parameter values. Table III compares the expected size of the confidence regions for these two procedures, along with LRT and MR, for various values of (θ_1, θ_2). MES has the smallest maximum expected size over this sample of parameter values, but small expected size for large θ_2 comes at the cost of increased expected size when θ_2 is small. TS is dominated by the others; there is no clear choice among the other three procedures.

Table III Expected sizes of four approximate 95% confidence regions for the parameter θ_1 in the psychophysics example in section 2.2: truncating the confidence ellipse based on the asymptotic distribution of the MLE (TS), inverting the likelihood ratio test (LRT), minimax expected size (MES), and minimax regret (MR).

$\lvert\theta_1\rvert$	θ_2	TS	LRT	MES	MR
0.00	0.00	1.87	1.55	2.42	1.57
0.00	1.50	3.78	3.20	2.68	2.97
0.00	3.00	5.49	3.13	2.84	3.08
0.00	4.50	6.32	2.61	2.73	2.63
1.00	0.00	2.40	1.69	2.58	1.94
1.00	1.00	3.05	2.45	2.68	2.52
1.00	2.00	3.77	2.68	2.68	2.55
2.00	0.00	2.96	1.37	2.48	1.72
2.00	0.50	2.96	1.49	2.50	1.80

3. SOME THEORY

This section presents some of the theory behind MES and MR informally; see also Evans et al. [2003] for a more rigorous and general treatment of MES.

Consider the following estimation problem. The compact set Θ, a subset of $\Re^p$, is the set of possible states of nature—the possible values of an unknown parameter θ. For each $\theta \in \Theta$, there is a distribution P_θ on the space of possible observations $\mathcal{X} = \Re^m$; X is a random variable with distribution P_θ; and x is a generic observed value of X. Each distribution P_θ has

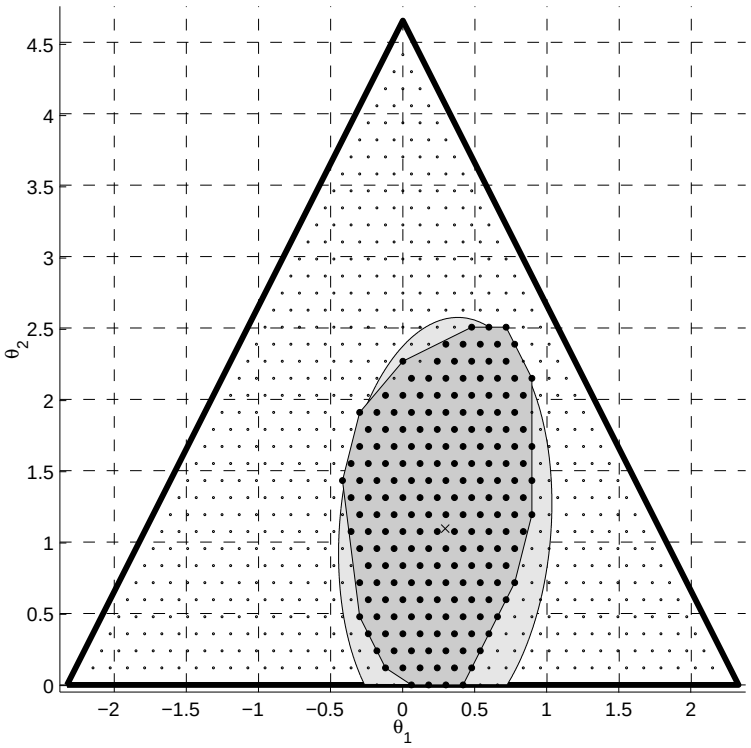

Figure 3: Approximate 95% confidence sets for an estimation problem in psychophysics. In this case $n_1 = n_2 = 10$, $y_1 = 8$, and $y_2 = 4$. The light gray truncated ellipse is a confidence region found using the asymptotic approximation to the distribution of the MLE. The "x" in the center of the ellipse is the MLE. The dots in the grid are the parameter values considered by MES. The larger dots are accepted values; the smaller are rejected. The darker, irregular region is the convex hull of these accepted parameter values, the MES confidence set.

a density $f(x|\theta)$ relative to Lebesgue measure; $f(x|\theta)$ is jointly continuous in x and θ.[1] We seek a confidence set for θ based on the observation $X = x$ and the *a priori* constraint $\theta \in \Theta$.

First consider testing the hypothesis $\theta = \eta$ at level α for an arbitrary fixed value $\eta \in \Theta$. Let $A(\eta)$ be the acceptance region of the test—the set of values $x \in \mathcal{X}$ for which we would not reject the hypothesis. Because the significance level of the test is α,

$$P_\eta\{X \in A(\eta)\} \geq 1 - \alpha. \tag{13}$$

The *power function* β of the test is the chance that the test rejects the hypothesis $\theta = \eta$ when in fact $\theta = \zeta$:

$$\beta(\zeta, \eta) \equiv 1 - P_\zeta\{X \in A(\eta)\}. \tag{14}$$

Because A_η has significance level α, $\beta(\eta, \eta) \leq \alpha$. Subject to that restriction, when testing a particular alternative hypothesis $\theta = \zeta$, it is natural to choose $A(\eta)$ to maximize $\beta(\zeta, \eta)$. Such a test is *most powerful (against the alternative $\theta = \zeta$)*. The following classical result characterizes the most powerful test in this situation.

The Neyman-Pearson Lemma: For fixed η, the acceptance region of the level α test that maximizes

$$\int_\Theta \beta(\zeta, \eta)\, \pi(d\zeta) \tag{15}$$

for an arbitrary measure π on Θ is

$$A_\pi(\eta) \equiv \{x \colon T_\pi(\eta, x) \geq c_\eta\}\,, \tag{16}$$

where

$$T_\pi(\eta, x) \equiv \frac{f(x|\eta)}{\int_\Theta f(x|\zeta)\pi(d\zeta)}, \tag{17}$$

with c_η chosen so that $\beta(\eta, \eta) = \alpha$.

The acceptance region $A_\pi(\eta)$ defined in equation 16 plays a crucial role in constructing optimal confidence sets. The set

$$C_A(x) = \{\eta \in \Theta : x \in A(\eta)\} \tag{18}$$

of all η that are accepted at significance level α is a $1 - \alpha$ confidence region for θ based on the observation

[1] This discussion assumes X is continuous. For X discrete, we could introduce an independent, uniformly distributed random variable U observed along with X. This is equivalent to considering randomized decision rules. See Evans et al. [2003] for more rigor.

$X = x$. We want to minimize the expected ν-measure of the confidence region $C_A(X)$ by choosing the acceptance regions $A(\eta)$ well. The measure ν on the parameter space Θ can be essentially arbitrary, but it needs to be defined on a broad enough class of subsets of Θ that $C_A(x)$ is ν-measurable for any value of x. In applications, ν is typically Euclidean volume.

The following theorem is due to Pratt [1961].

Pratt's Theorem:

$$\mathbf{E}_\zeta[\nu(C_A(x))] = \int_\Theta (1 - \beta(\zeta, \eta))\, \nu(d\eta), \qquad (19)$$

where $\mathbf{E}_\zeta[\cdot]$ is expectation when $\theta = \zeta$, and $\beta(\cdot, \cdot)$ is the power function of the family of tests $\{A_\eta\}$ corresponding to the confidence set C_A.

Pratt's theorem links maximizing the power function β to minimizing the expected size of the confidence region $C_A(X)$. The following result combines the Neyman-Pearson Lemma and Pratt's Theorem.

Corollary: The confidence set C_A that minimizes

$$\int_\Theta \mathbf{E}_\zeta[\nu(C_A(X))]\, \pi(d\zeta) \qquad (20)$$

is C_{A_π}.

What is the role of the measure π? The following is proved in great generality in Evans et al. [2003].

Theorem [Evans et al. 2003]: There exists a measure λ on Θ such that the acceptance regions A_λ give the confidence procedure that minimizes

$$\max_{\theta \in \Theta} \mathbf{E}_\theta[\nu(C_A(X))]\,. \qquad (21)$$

This is MES, and λ is referred to as the *least favorable alternative* because the alternative defined by λ maximizes the Bayes risk (see section 5).

This result can be adapted to show that there is another measure μ on Θ for which C_{A_μ} is the minimax regret procedure. Determining these priors exactly is not computationally feasible except in simple cases. Section 5 sketches an efficient method to approximate λ and μ numerically.

4. CMB DATA ANALYSIS

The cosmic microwave background radiation (CMB) consists of redshifted photons that have travelled since the *time of last scattering*, approximately 300,000 years after the Big Bang, when the Universe had cooled enough to allow atoms to form and photons to travel freely. The small fluctuations in the temperature of the CMB are the signature of the primordial variability that led to the structure visible in the Universe today, such as galaxies and clusters of galaxies. Theoretical research connects unknown physical constants that characterize the Universe—such as the fraction of ordinary matter in the Universe, the fraction of dark matter in the Universe, Einstein's cosmological constant, Hubble's constant, the optical depth of the Universe, and the spectral index—to the angular distribution of the fluctuations. See chapter two of Longair [1998] for an introduction.

Estimating these cosmological parameters from observed CMB fluctuations is conceptually similar to the example given in section 2.2. The physically interesting parameters are the cosmological parameters, while the canonical parameter is the angular power spectrum of the CMB. The data are assumed to be a realization of a normally distributed vector with mean zero and covariance matrix

$$\mathbf{N} + \sum_\ell \left(\frac{2\ell + 1}{4\pi}\right) C_\ell(\theta) B_\ell^2\, \mathbf{P}_\ell, \qquad (22)$$

where $\mathbf{N}$ is the measurement error covariance matrix (which is assumed to be known), $\{C_\ell(\theta)\}$ is the CMB power spectrum for the cosmological parameter vector θ, $\{B_\ell\}$ is the transfer function resulting from the beam pattern of the observing instrument, and $\mathbf{P}_\ell$ is a matrix whose (i, j) entry is the degree ℓ Legendre polynomial evaluated at the cosine of the angle between pixel i and pixel j. This representation is based on the spherical harmonic decomposition of a spherical, isotropic Gaussian process model for the CMB. The software package CMBFAST [Seljak and Zaldarriaga 1996] is the standard for calculating the spectrum from cosmological parameters; the nonlinearity of this mapping is a major complication in this problem.

Table IV lists the parameters we use and their *a priori* bounds, based on Abroe et al. [2002]. Figure 4 shows the data: the 5,972 observations in the MAXIMA-1 8 arcminute resolution data set [Hanany et al. 2000]. We compress the data to 2,000 linear combinations of the original observations, then form 95% MES and MR joint confidence regions for the parameters. Figure 5 shows the MES confidence set in the spectral domain. A total of 1,000 models were tested; 35 were accepted. (Generating spectra from the randomly selected parameter vectors is computationally expensive. These results are preliminary: We plan to test more models in the future.) Their spectra are the heavier curves in the figure. The lighter curves are spectra of 300 of the rejected models. The dark band is an approximate 95% confidence region for the angular power spectrum of CMB fluctuations.

The parameter values for each of the 1,000 tested spectra are known: Table V lists 15 the 35 accepted vectors along with the minimum and maximum accepted values for each parameter. For example, all the accepted values of the total energy density relative to the critical energy density, $\Omega = \Omega_m + \Omega_\Lambda$,

Table IV Cosmological parameters and their bounds, following Abroe et al. [2002]. The parameters also must satisfy $\Omega_b \leq \Omega_m$ and $0.6 \leq \Omega_m + \Omega_\Lambda \leq 1.4$.

Parameter (Symbol)	**Lower**	**Upper**
Total Matter (Ω_m) †	0.05	1.00
Baryonic Matter (Ω_b) †	0.005	0.15
Cosmological Constant (Ω_Λ) †	0.0	1.0
Hubble Constant (H_0) (km s^{-1} Mpc^{-1})	40.0	90.0
Scalar Spectral Index (n_s)	0.6	1.5
Optical Depth (τ)	0.0	0.5

† Relative to critical density.

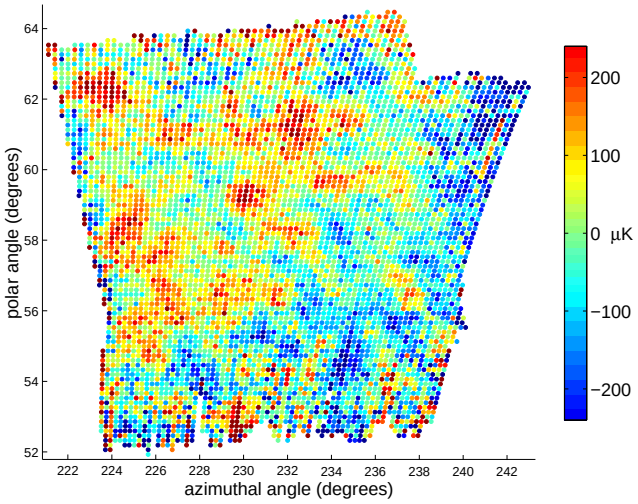

Figure 4: The MAXIMA-1 data set used in this analysis. There are 5,972 pixels at 8 arcminute resolution.

are between 0.915 and 1.334. The MAXIMA-1 experiment has much higher resolution than previous experiments, but it still does not constrain most of the parameters individually, owing partly to tradeoffs among the parameters. (Our data compression also might contribute to the uncertainty; we have not yet explored the sensitivity to the compression scheme.)

From a frequentist viewpoint, the fact that there is a parameter vector that accounts adequately for the data (that is accepted) and which has $\Omega = 1.334$ means that we cannot rule out the possibility that $\Omega = 1.334$ at significance level 0.05. Bayesian techniques make inferences starting with the marginal posterior distribution for each parameter by itself: Whether the posterior credible region includes $\Omega = 1.334$ depends on the posterior weight assigned to the set of *all* models with $\Omega = 1.334$. That weight, in turn, depends on the prior as well as the data.

Figure 5 also plots error bars given by Hanany et al. [2000], based on their analysis of the MAXIMA-1 data. The error bar at $\ell = 223$ extends far above all the accepted spectra. Close inspection shows that each accepted spectrum passes either through the bar at $\ell = 147$ or through the bar at $\ell = 300$. None of the 1,000 spectra (including the 665 spectra that are not plotted) passes through all three of these bars. This shows the fundamental problem with the "chi-by-eye" procedure for comparing spectra with error bars: It is not clear how well the spectra should fit the bars, especially when estimates at different frequencies are dependent, as they are here. MES allows more precise comparisons, and maximizes the power of the tests in the sense described in section 3. The MR results, shown in Figure 6, are similar but only 25 spectra are accepted. Figures 5 and 6 also show the best fitting model based on the recent WMAP experiment [Bennett et al. 2003], which has much higher resolution than MAXIMA-1. At low ℓ, the WMAP model is quite similar to the models accepted by MES and MR using the MAXIMA-1 data.

Table V Fifteen of the 35 cosmological parameter vectors accepted by MES. The final two rows list the minimum and maximum accepted values of each parameter.

Ω_b	Ω_m	Ω_Λ	τ	H_0	n_s	Ω
0.042	0.674	0.241	0.317	77.00	1.117	0.915
0.078	0.368	0.632	0.161	69.71	0.834	1.000
0.088	0.786	0.214	0.445	68.65	1.151	1.000
0.131	0.860	0.176	0.417	67.07	1.027	1.036
0.081	0.540	0.526	0.000	77.15	0.809	1.066
0.079	0.321	0.773	0.364	69.35	1.002	1.094
0.134	0.940	0.161	0.466	66.83	1.038	1.101
0.101	0.699	0.482	0.000	44.68	0.833	1.181
0.089	0.425	0.771	0.217	77.85	0.896	1.196
0.130	0.591	0.635	0.364	43.03	0.944	1.226
0.085	0.994	0.243	0.315	76.79	1.081	1.237
0.096	0.555	0.693	0.260	81.28	0.923	1.248
0.093	0.708	0.551	0.000	76.96	0.855	1.259
0.139	0.667	0.623	0.269	61.15	0.954	1.290
0.133	0.692	0.642	0.068	41.47	0.846	1.334
0.011	0.058	0.082	0.000	41.47	0.729	0.915
0.139	0.994	0.988	0.466	89.24	1.151	1.334

5. APPROXIMATING THE LFA

The LFA λ is the measure π on Θ that maximizes

$$\mathbf{B}(\pi) \equiv \int_\Theta \mathbf{E}_\zeta[\nu(C_{A_\pi}(X))]\,\pi(d\zeta)\,. \qquad (23)$$

This is an instance of Bayes/minimax duality: The

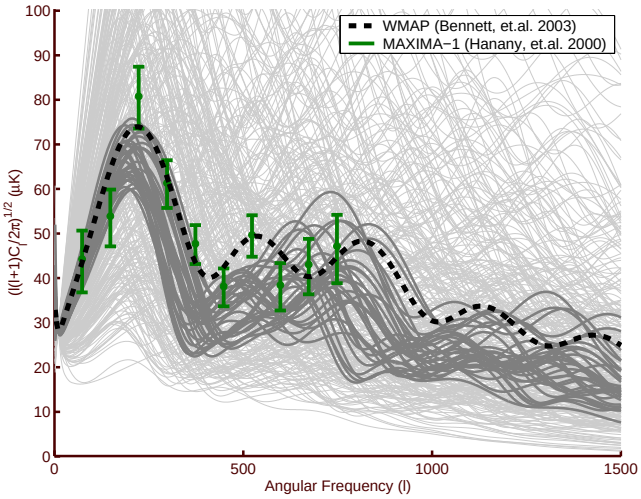

Figure 5: The 35 accepted spectra (dark curves) and 300 of the 965 rejected spectra (light curves) from the MES procedure applied to 8 arcminute MAXIMA-1 data. The vertical bars are Bayesian error bars based on the MAXIMA-1 data [Hanany et al. 2000]; the dashed curve is the best fitting model to the WMAP data [Bennett et al. 2003].

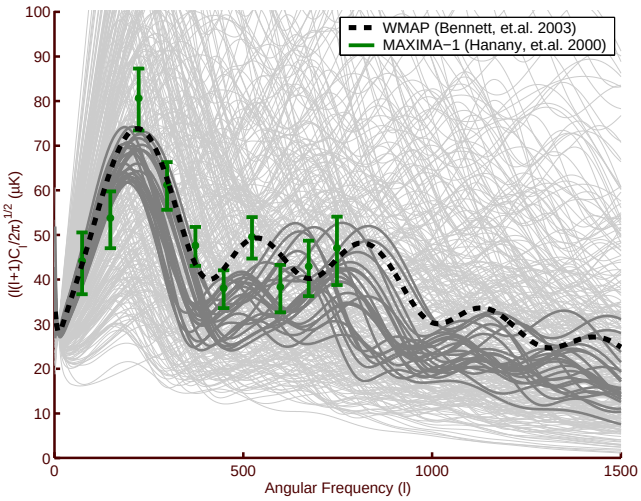

Figure 6: The 25 accepted spectra (dark curves) and 300 of the 975 rejected spectra (light curves) from the MR procedure applied to the 8 arcminute MAXIMA-1 data. The vertical bars are Bayesian credible intervals based on the MAXIMA-1 data [Hanany et al. 2000]; the dashed curve is the spectrum of the model that fits the WMAP data best [Bennett et al. 2003].

"worst" prior λ corresponds to the minimax procedure. It is computationally impractical to determine the LFA explicitly in all but the simplest situations. The main difficulty is the complicated relationship between π and A_π, which makes it hard to evaluate equation 23. Nelson [1966] and Kempthorne [1987] propose computational methods for determining least favorable priors in general situations, but they assume that calculating the Bayes risk (equation 23) is a solved problem: It is not part of their algorithms.

The approach we use here is described in greater detail in Schafer and Stark [2003]. It involves two levels of numerical approximation. First, the support of the prior is restricted to a finite set of points. Second, Monte Carlo methods are used to estimate equation 23 for any prior π supported on this discrete set. Schafer and Stark [2003] show that as the size of the Monte Carlo simulations increases, the estimate of $\mathbf{B}(\pi)$ converges uniformly in π to $\mathbf{B}(\pi)$.

Let $\{\theta_i\}_{i=1}^p$ be the support points of the prior. Let $\{\eta_j\}_{j=1}^q$ be parameter values selected at random from the compact parameter space Θ according to the measure ν. For each η_j, simulate a set of data $x_{j1}, x_{j2}, \ldots, x_{jn}$. Construct matrices $\mathbf{A}_j$, $j = 1, 2, \ldots, q$, with (i,k) entry

$$\mathbf{A}_j(i,k) \equiv \frac{f(x_{jk}|\theta_i)}{f(x_{jk}|\eta_j)}. \tag{24}$$

We need to estimate $\mathbf{B}(\pi)$, for a prior π supported on $\{\theta_i\}_{i=1}^p$. The empirical distribution of the vector $\mathbf{A}_j\pi$ is an approximation to the distribution of the test statistic under the null hypothesis $\theta = \eta_j$. It can be used to estimate the threshold to form the acceptance region $A_\pi(\eta_j)$.

Choosing a decision rule can be thought of as picking a strategy in a zero-sum two-person game: Statistician versus Nature. The Statistician chooses a set of decision vectors $\mathbf{d}_j$, $j = 1, 2, \ldots, q$. Nature chooses π, a distribution over possible true values of the parameter. The Statistician pays Nature the approximate risk

$$\tilde{\mathbf{R}}(\mathbf{d}, \pi) \equiv \sum_{j=1}^{q} \mathbf{d}_j^T \mathbf{A}_j \pi. \tag{25}$$

The Statistician can choose $\mathbf{d}$ to minimize $\tilde{\mathbf{R}}(\mathbf{d}, \pi)$ by setting the component d_{jk} to one if $x_{jk} \in A_\pi(\eta_j)$ and to zero otherwise. Let $\mathbf{d}^\pi$ be that optimal decision function. Define

$$\tilde{\mathbf{B}}(\pi) \equiv \tilde{\mathbf{R}}(\mathbf{d}^\pi, \pi). \tag{26}$$

Then $\tilde{\mathbf{B}}(\pi)$ is an approximation of the Bayes risk for prior π. Maximizing $\tilde{\mathbf{B}}$ over π amounts to finding the (approximate) optimal strategy for Nature, the prior that maximizes the payout by the Statistician. This is a matrix game, and finding an optimal strategy is a well-studied problem. A fictitious play algorithm proposed by Brown and Robinson [Robinson 1951] works well here because it can handle the constraint on the Statistician's strategies that ensures $1 - \alpha$ coverage. Solving this matrix game for large problems is computationally expensive; typically the most costly

steps are to simulate data from a randomly chosen parameter vector and to evaluate the likelihood function. An implementation in Fortran-90, runnable on parallel computers, is available at the URL `www.stat.berkeley.edu/~stark/Code/LFA_Search`. This subroutine also can find an approximate LRA for MR.

6. CONCLUSION

Expected size is a useful measure of the performance of a confidence estimator. It is directly related to the power of the procedure to reject false parameter values; this is a natural property to maximize. Generally there is no estimator that minimizes expected size for all parameter values simultaneously: Some tradeoff must be imposed. Minimax expected size (MES) and minimax regret expected size (MR) trade off expected sizes at the possible parameter values optimally—in different senses. MES and MR are alternatives to the likelihood ratio test approach to confidence sets proposed by Feldman and Cousins [1998]. MES and MR incorporate bounds on parameters by minimizing the maximum expected size only over the set of parameters that satisfy these bounds, and by including only parameters within the bounds. MR is less conservative than MES. These regions typically cannot be calculated analytically, but they can be approximated numerically, and we provide a Fortran-90 subroutine.

MES and MR can be used to estimate cosmological parameters from observations of the cosmic microwave background radiation, incorporating bounds on the parameters to produce confidence sets that are small in expectation. MES and MR use the subroutine CMBFAST to map cosmological parameters to power spectra. They do not involve the complicated relationship between the parameters of interest and the canonical parameters explicitly. MES and MR can test cosmological models formally, avoiding potentially misleading "chi-by-eye" comparisons between spectra and spectrum estimates.

Acknowledgments

MAXIMA-1 data are courtesy of the MAXIMA collaboration. Analysis of CMB data was performed at NERSC using resources made available by Julian Borrill. We had many helpful conversations with Andrew Jaffe regarding microwave cosmology, and with Poppy Crum regarding signal detection problems in psychophysics.

References

G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).
G. Backus, Proc. Natl. Acad. Sci. **84**, 8755 (1987).
G. Backus, Geophys. J. **94**, 249 (1988).
M. Schervish, *Theory of Statistics* (Springer-Verlag, New York, 1995).
S. Evans, B. Hansen, and P. Stark, Tech. Rep. 617, Univ. of California, Berkeley (2003).
P. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden Day, San Francisco, 1977).
E. Lehmann, *Testing Statistical Hypotheses* (John Wiley and Sons, New York, 1986), 2nd ed.
M. DeGroot, in *Encyclopedia of Statistical Science*, edited by S. Kotz, N. Johnson, and C. Read (John Wiley and Sons, New York, 1988), vol. 8, pp. 3–4.
J. Pratt, J. Am. Stat. Assoc. **56**, 549 (1961).
T. Affolder, H. Akimoto, A. Akopian, M. Albrow, P. Amaral, S. Amendolia, D. Amidei, J. Antos, G. Apollinari, T. Arisawa, et al., Phys. Rev. D **61**, 072005 (2000).
J. Miller, Perception and Psychophysics **58**, 65 (1996).
H. Kadlec, Psychological Methods **4**, 22 (1999).
M. Longair, *Galaxy Formation* (Springer-Verlag, New York, 1998).
U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
M. Abroe, A. Balbi, J. Borrill, E. Bunn, P. Ferreira, S. Hanany, A. Jaffe, A. Lee, K. Olive, B. Rabii, et al., Month. Not. Royal Astron. Soc. **334**, 1 (2002).
S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. Ferreira, V. Hristov, A. Jaffe, et al., Astrophys. J. Lett. **545**, L5 (2000).
C. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. Meyer, L. Page, D. Spergel, G. Tucker, et al., Astrophys. J. Suppl. **148**, 1 (2003).
W. Nelson, Ann. Math. Stat. **37**, 1643 (1966).
P. Kempthorne, SIAM J. Sci. Stat. Comput. **8**, 171 (1987).
C. Schafer and P. Stark (2003), In preparation.
J. Robinson, Ann. Math. **54**, 296 (1951).