Estimated Loss and Admissible Loss Estimators

by

Andrew L. Rukhin
Purdue University

Technical Report #85-26

Department of Statistics
Purdue University

September 1985
Revised April 1986

Research supported by the National Science Foundation Grant DMS-8401996.

Estimated Loss and Admissible Loss Estimators

by

Andrew L. Rukhin
Purdue University

Abstract

A loss function which combines the error in a decision problem and the accuracy of the statistical decision rule, is introduced. This loss provides a variable data-dependent measure of precision of the decision made, which admits frequentist interpretation. The Bayesian procedures with respect to this loss are the usual Bayes decision and the posterior loss, so that this loss function also may appeal to the Bayesian school. We give a necessary and sufficient condition for admissibility and demonstrate the inadmissibility of the sample standard deviation as an estimate of the accuracy of the normal sample mean.

1. Introduction

Let us start with general statistical decision problem as described by possible states of nature θ, decisions d and a loss function $W(\theta, d)$. Classical decision theory advocates making some decision $d = \delta(x)$ where x is the observation, with frequentist risk $R(\theta, \delta)$. This approach has been often criticized because it implicitly assumes that $R(\theta, \delta)$ is a good measure of accuracy of the procedure used (or of a measure of conclusiveness). In fact it is desirable "to have a methodology that gives a highly variable data-dependent measure of conclusiveness on the conclusion inferred from the experiment, with frequentist interpretation of that measure" (Brownie and Kiefer (1977)). There are many important examples where one would like to accompany the decision δ with an estimate, say, $\gamma = \gamma(x)$ of its accuracy or of the loss function $W(\theta, \delta)$. Exactly this necessity of measuring the accuracy or rather the inaccuracy is known to be behind the idea of confidence estimation (cf. Savage (1954), ch 17). Indeed while the center of a confidence interval may serve as a point estimator of the unknown parameter, its width represents the precision of this estimator.

The problem of estimating the risk function has been considered by Lehmann (1959) who mentioned estimated power of a test and by Sandved (1968) who found unbiased estimators of quadratic risk in several estimation situations. A lot of attention was brought to this problem by Jack Kiefer who in a series of papers (1975), (1976), (1977) developed conditional and estimated confidence theories which in particular provide estimates of confidence admitting frequentist interpretability. (See also Brown (1978) for some notions of admissibility in the problem of estimated confidence). Berger (1985a, 1985b, 1985c) compares the subjective Bayesian approach to this problem with the frequentist one. In particular he discusses the desirable properties of valid measures of performance of a statistical decision rule δ from the frequentist point of view. We also note that certain nonparametric and bootstrap methods for standard errors and other characteristics of statistical accuracy (Efron (1982), Efron and Tibshirani (1986)) may be of an estimated or conditional nature.

To develop a decision-theory approach to this problem one must specify an appropriate loss function. "Any serious attempt to take account of the consequences of unreliability in not capturing the true parametric value and of lack of usefulness in excessive width should, we feel, involve the specification of some reasonable loss function and the subsequent examination of the problem in terms of decision theory" (Aitchison and Dunsmore (1968)). In fact, a variety of loss functions in interval estimation has been considered (see Pratt (1961), Winkler (1972), Cohen and Strawderman (1973)).

In this paper we give a loss function which combines the general decision problem error with the inaccuracy estimate. This loss function is a very convenient tool in the problem of simultaneous "decision-precision" reporting. The corresponding risk, as any risk function, has frequentist interpretability in terms of long-run frequencies. The Bayes "decision-precision" pairs turn out to be the usual Bayes decision δ_B for θ and the posterior loss $\gamma_B = E\{W(\theta,\delta_B)/x\}$. Since admissible pairs in statistical decision theory are typically Bayes or generalized Bayes procedures a frequentist may accept posterior loss as an estimate of risk because of the admissibility argument (see Section 3 of this paper). Of course for him or her (and for this author too) there remains the often difficult problem of evaluating the risk function.

In Section 2 we introduce and discuss the mentioned loss function. Also a necessary and sufficient condition for admissibility is derived in Section 3. This condition is used in Section 4 to prove the inadmissibility of the sample standard deviation as an estimate of the accuracy of the normal sample mean.

2. "Decision-Precision" Loss Function

Denote by γ an estimator of the loss $W(\theta,\delta)$. A loss function which combines the non-negative decision loss $W(\theta,\delta)$ with the accuracy of δ, $L(\theta,\delta,\gamma)$, is given by

$$L(\theta,\delta,\gamma) = W(\theta,\delta)\ \gamma^{-1/2} + \gamma^{1/2}. \tag{2.1}$$

The important feature of this loss function is that for a fixed δ, the (unique) minimum in γ is attained at $\gamma_{\min} = W(\theta,\delta)$. A convenient convention is to put $\gamma_{\min} = 0$ if $W(\theta,\delta) = 0$. On the other hand for a fixed γ this loss function is just a linear transform of $W(\theta,\delta)$, so that the Bayes procedure δ_B for L is the Bayes procedure for W. The Bayes estimator γ_B of accuracy is

$$\gamma_B = E\{W(\theta,\delta_B)/x\};$$

i.e. γ_B coincides with the posterior loss of δ_B. This result allows one to interpret the posterior loss of the Bayes procedure δ_B as the Bayes estimator of the loss function of δ_B under (2.1). This estimator is related to Kiefer's conditional loss estimate for the "continuum" (finest possible) partition of the sample space.

To illustrate the properties of the loss function (2.1) let us consider two examples essentially treated by Kiefer (1977) and Berger (1985b).

Example 1. Let $x = (x_1, x_2)$ where independent variables x_1 and x_2 have the distribution

$$P_\theta(x_i = \theta - 1) = P_\theta(x_i = \theta + 1) = .5.$$

Let $W(\theta, d) = 1$ if $d \neq \theta; = 0$ if $d = \theta$. Consider the procedure $\delta(x) = (x_1 + x_2)/2$ if $|x_1 - x_2| = 2; = x_1 - 1$ if $x_1 = x_2$. Then

$$R(\theta, d) = P_\theta(\delta(x) \neq \theta) = .25 \qquad \text{for all } \theta.$$

Clearly the estimate $\gamma_1(x) \equiv .25$ of $W(\theta, \delta)$ is rather silly. Indeed if $|x_1 - x_2| = 2$ one is 100% certain that $\delta(x) = \theta$ while if $x_1 = x_2$ one is equally uncertain as to whether θ is $x_1 - 1$ or $x_1 + 1$.

Under the loss (2.1) the estimator γ_1 is inadmissible. Indeed

$$R(\theta, \delta, \gamma_1) = E_\theta\ L(\theta, \delta, \gamma_1) = .25(.25)^{-1/2} + (.25)^{1/2} = 1.$$

If one uses common sense and defines $\gamma_0(x) = .5$ if $x_1 = x_2; = 0$ if $|x_1 - x_2| = 2$, then for all θ

$$R(\theta, \delta, \gamma_0) = P_\theta\{x_1 = x_2\} \left[(.5)^{-1} + (.5)^{1/2}\right] = 2^{1/2}/2 < 1 = R(\theta, \delta, \gamma_1).$$

This fact seems to confirm the reasonable character of (2.1). It should be contrasted with the admissibility of γ_1 for the loss $(\gamma - W(\theta, \delta))^2$ (see Berger (1985a) p. 44).

Example 2. Suppose one observes a normal random variable x with mean θ and unit variance. The choice between two hypotheses $H_0\colon \theta \leq -\varepsilon$ and $H_1\colon \theta \geq \varepsilon$ has to be made. Here ε is a fixed positive number.

Consider the test δ which rejects H_0 if $x \geq 0$. Clearly, for $\theta < 0$

$$P_\theta(\text{Type 1 error}) \ = P_{|\theta|}(\text{Type II error}) \ \leq P_{-\varepsilon}(x > 0) = \Phi(-\varepsilon).$$

If, say, $\varepsilon = 2$, $\Phi(-2) = .0228$, but it seems to be rather unreasonable to state that H_0 is rejected with error probability not exceeding .0228 when observed value of x is 0.

Motivated by the fact that δ is a Bayes test against the prior distribution assigning equal mass to $\theta = -\varepsilon$ and $\theta = \varepsilon$ we put

$$\gamma(x) = E\{W(\theta, \delta)/x)\}$$

where W is zero-one loss. Then

$$\begin{aligned}\gamma(x) &= \left[W(-\varepsilon, \delta(x)) \exp\{-(x+\varepsilon)^2/2\} + W(\varepsilon, \delta(x)) \exp\{-(x-\varepsilon)^2/2\}\right] \\ &\quad / \left[\exp\{-(x+\varepsilon)^2/2\} + \exp\{-(x-\varepsilon)^2/2\}\right] \\ &= \begin{cases} \exp\{-(x+\varepsilon)^2/2\}/\left[\exp\{-(x+\varepsilon^2/2\} + \exp\{-(x-\varepsilon)^2/2\}\right] & x \geq 0 \\ \exp\{-(x-\varepsilon^2/2\}/\left[\exp\{-(x+\varepsilon)^2/2\} + \exp\{-(x-\varepsilon)^2/2\}\right] & x < 0 \end{cases} \\ &= 1/(1 + e^{2\varepsilon|x|}). \end{aligned} \tag{2.2}$$

When $x = 0$, γ takes its largest value, .5, but for large values of $|x|$ this function is small, which is intuitively appealing.

When ε tends to zero, i.e. one considers testing $\theta \leq \theta$ versus $\theta > 0$, the estimator (2.2) becomes useless. Since the test δ is Bayes against any symmetric prior distribution Λ, one may rather employ an estimator γ of the form

$$\gamma(x) = \int_0^{\infty} e^{-\theta|x|}\, d\tilde{\Lambda}(\theta) / \int_{-\infty}^{\infty} e^{x\theta}\, d\tilde{\Lambda}(\theta)$$

where $d\tilde{\Lambda}(\theta) = e^{-\theta^2/2}\, d\Lambda(\theta)$.

Notice that $\gamma(0) = .5$ and γ is decreasing as $|x|$ increases.

Kiefer (1977) (see also Rubin (1984)) suggested the use of conservatively biased estimators γ, i.e. γ for which

$$E_\theta\, \gamma(x) \geq R(\theta, \delta). \tag{2.3}$$

If Λ is proper prior distribution then

$$\int \int [\gamma_B(x) - W(\theta, \delta_B(x))]\ dP_\theta(x)\ d\Lambda(\theta) = \int [E_\theta\, \gamma_B(x) - R(\theta, \delta)]\ d\Lambda(\theta) = 0$$

where P_θ denotes the distribution of x. Thus (2.3) cannot hold unless for Λ-almost all θ

$$E_\theta\, \gamma_B(x) = R(\theta, \delta_B).$$

But the latter can be valid for a Bayes γ_B only if

$$\int \int [\gamma_B(x) - W(\theta, \delta_B(x))]^2\, dP_\theta(x)\ d\Lambda(\theta) = 0$$

(see Lehmann (1983) p. 244–245). It follows that typically there is no proper Bayes conservatively biased or unbiased estimator γ_B. Therefore it is hopeless to expect conservatively calibrated precision estimators in most of statistical problems with a compact parametric space.

Another remark is that the loss function (2.1) gives equal weights to the "decision" loss and to the "precision" loss. If these losses are of different importance, then this fact can be reflected by introducing a positive factor ω possibly dependent on θ in the following way:

$$L_\omega(\theta, \delta, \gamma) = W(\theta, \delta)\gamma^{-1/2} + \omega\gamma^{1/2}$$

Then the Bayes estimator of $R(\theta, \delta)$ corresponds to a new prior Λ_ω: $d\Lambda_\omega = \omega^{-1} d\Lambda$.

3. Admissibility Criterion

In this section it is assumed that the sample space $\mathcal{X}$ is Euclidean space of dimension n, $\mathcal{X} = \mathbb{R}^n$, the decision space $\mathcal{D}$ is an open convex subset of $\mathbb{R}^m$ and the parameter space

Θ is a separable locally compact metric space. We make the measurability and regularity assumptions of (i)-(v) of Theorem 1 of Farrell (1968). In particular we suppose that W is a continuous loss function over $\Theta \times \mathcal{D}$ which is strictly convex in d, and that there exist positive densities $p_\theta(x)$ with respect to some measure μ.

Theorem 1. Under assumptions (i)-(v) of Farrell (1968) the pair (δ_0, γ_0) is an admissible procedure under loss (2.1) if and only if there exists a sequence $G_k, k = 1, 2, \ldots,$ of finite measures over Θ such that for a compact subset E of Θ , $G_k(E) \geq 1$, $k = 1, 2, \ldots, \sup_k G_k(C) < \infty$ for compact C and

$$\int_{\mathcal{X}} \int_{\theta} [W(\theta, \delta_0) - W(\theta, \delta_k)] \gamma_o^{-1/2} p_\theta(x) \, d\mu(x) \, dG_k(\theta) \to 0, \tag{3.1}$$

$$\int_{\mathcal{X}} \int_{\Theta} \left[\gamma_0^{1/2} - \gamma_k^{1/2}\right]^2 \gamma_0^{1/2} p_\theta(x) \, d\mu(x) \, dG_k(\theta) \to 0 \tag{3.2}$$

where δ_k, γ_k are Bayes rules against G_k.

In particular if (δ_0, γ_0) is admissible under loss (2.1) then δ_0 is admissible under the loss $W(\theta, \delta)\gamma_0^{-1/2}$ and γ_0 is admissible under the loss $E_\theta W(\theta, \delta_0)\gamma^{-1/2} + \gamma^{1/2}$.

Proof. Let for any integrable function $f(x, \theta), \mathcal{E}_k f(x, \theta) = \int \int f(x, \theta) p_\theta(x) \, d\mu(x) \, dG_k(\theta)$. According to Theorem 1 of Farrell (1968) (δ_0, γ_0) is admissible if and only if

$$\rho_k = \mathcal{E}_k \{W(\theta, \delta_o(x)) \, [\gamma_0(x)]^{-1/2} + [\gamma_0(x)]^{1/2} - W(\theta, \delta_k(x)) \, [\gamma_k(x)]^{-1/2} + [\gamma_k(x)]^{1/2}\} \to 0. \tag{3.3}$$

Thus because of the property of iterated expected value

$$\begin{aligned} \rho_k &= \mathcal{E}_k \{ [W(\theta, \delta_0) - W(\theta, \delta_k)] \gamma_0^{-1/2} \\ &\quad + (\gamma_0^{1/2} - \gamma_k^{1/2}) + W(\theta, \delta_k)(\gamma_0^{-1/2} - \gamma_k^{-1/2}) \} \\ &= \mathcal{E}_k \{ [W(\theta, \delta_0) - W(\theta, \delta_k)] \gamma_0^{-1/2} \\ &\quad + \mathcal{E}_k \{ (\gamma_0^{1/2} - \gamma_k^{1/2}) + \gamma_k (\gamma_0^{-1/2} - \gamma_k^{-1/2}) \} \\ &= \mathcal{E}_k \{ [W(\theta, \delta_0) - W(\theta, \delta_k)] \gamma_0^{-1/2} \\ &\quad + \mathcal{E}_k \{ \left[\gamma_0^{1/2} - \gamma_k^{1/2}\right]^2 \gamma_0^{-1/2} \}. \end{aligned} \tag{3.4}$$

Thus (3.3) holds if and only if (3.1) and (3.2) are valid.

Clearly (3.1) means the admissibility of δ_0 as an estimator of θ under rescaled loss function $L_0(\theta, \delta) = W(\theta, \delta)\gamma_0^{-1/2}$ (which involves the observation x).

Formula (3.2) means that γ_0 is an admissible estimator of the parametric function $\varphi(\theta) = E_\theta W(\theta, \delta_0)$ under loss function $L_1(\theta, \gamma) = \varphi(\theta)\gamma^{-1/2} + \gamma^{1/2}$. Indeed an easy calculation shows that

$$\mathcal{E}_k \{L_1(\theta, \gamma) - L_1(\theta, \gamma_k)\} = \mathcal{E}_k \left\{ \left[\gamma_o^{1/2} - \gamma_k^{1/2}\right]^2 \gamma_0^{-1/2} \right\}$$

and the conclusion follows from Farrell's Theorem.

Notice that separate admissibility of δ_0 under $L_0(\theta, \delta)$ and of γ_0 under $L_1(\theta, \gamma)$ does not imply the admissibility of (δ_0, γ_0) under $L(\theta, \delta, \gamma)$. Indeed δ_0 and γ_0 may be Bayes rules with respect to two different prior distributions, and Theorem 1 can be used as a source of examples of this kind.

It is known (cf. Berger and Srinivasan (1978)) that if $p_\theta(x) = \beta(\theta) \exp\{\theta' x\}$ and $W(\theta, d) = ||\theta - d||^2$, then any admissible estimator has the form

$$\delta(x) = \nabla \log \hat{G}(x)$$

where G is a σ-finite measure with support in the closure of the natural parameter space and

$$\hat{G}(x) = \int \exp\{\theta x\} \, dG(\theta) \tag{3.5}$$

is the Laplace transform of G.

It is easy to see that the corresponding risk estimator has the form

$$\gamma(x) = \nabla^2 \log \hat{G}(x) = \Sigma \frac{\partial^2}{\partial x_i^2} \log \hat{G}(x). \tag{3.6}$$

A modification of the proof of Theorem 2.1 of Berger and Srinivasan (1978) shows that any admissible pair δ, γ under (2.1) has the form (3.5), (3.6) for some σ-finite measure supported by the closure of the natural parameter space. Formula (3.6) is convenient for calculation of risk estimators in an exponential family.

Notice that the admissibility notion associated with the loss (2.1) is more natural and convenient to work with than the admissibility definitions due to Kiefer (1975) and Brown (1978) in the problems of conditional confidence estimators.

From the author's point of view the main advantage of the loss function (2.1) consists in the possibility of weeding out inadmissible pairs (δ, γ) some of which might be generalized Bayes.

4. Inadmissibility of the Sample Standard Deviation

In this section let $x = (x_1, \ldots, x_n), n > 3$ be a random normal sample with unknown mean ξ and unknown variance σ^2. The sample mean $X = \sum_1^n x_j/n$ is a natural traditional estimator of the mean ξ, and this estimator possesses numerous optimality properties (is minimax, unbiased, best equivariant, admissible under a wide class of loss functions). A natural estimator γ of the quadratic loss of X, $\gamma(X,S) = cS^2$, $S^2 = \sum_1^n (x_j - X)^2$, for a suitable choice of the constant c, does not have all these optimality properties. In fact, all estimators cS^2 are inadmissible as estimators of σ^2 for a variety of invariant loss functions. This result was discovered first by Stein (1964), and the further study has been performed by Brown (1968), Brewster and Zidek (1974), Strawderman (1974). The practical importance of this inadmissibility remains unclear since only minor relative improvements over cS^2 are known (see Rukhin (1986) for numerical study).

We show here that all estimators of the form cS^2 are also inadmissible as estimators of the risk function of X under the loss function

$$L(\xi,\sigma;\gamma) = \sigma^{-1}\left[(X-\xi)^2\gamma^{-1/2} + \gamma^{1/2}\right]. \tag{4.1}$$

This loss is invariant under affine transformations, and the best equivariant estimator is

$$\gamma_0(X,S) = c_0 S^2, \qquad c_0 = [n(n-2)]^{-1}.$$

Notice that

$$\begin{aligned} E\gamma_0(X,S) &= (n-1)\sigma^2\,[n(n-1)]^{-1} \\ &> \sigma^2 n^{-1} = E(X-\xi)^2 \end{aligned}$$

so that γ_0 is conservatively calibrated.

One can also consider the loss function

$$L_e(\xi,\sigma;\delta,\gamma) = \sigma^{-1}\left[(\delta-\xi)^2\gamma^{-1/2} + \gamma^{1/2}\right] \tag{4.2}$$

which essentially coincides with (2.1) and which admits a general estimator δ of ξ. Notice the analogy of this loss function to the loss function considered sometimes in confidence estimation

$$L_i(\xi,\sigma;\delta,\gamma) = I_{(\delta-\gamma^{1/2},\delta+\gamma^{1/2})^c}(\xi) + \omega\gamma^{1/2}. \tag{4.3}$$

Here $\gamma^{1/2}$ represents half-width of the confidence interval $(\delta-\gamma^{1/2},\delta+\gamma^{1/2})$, and ω is a positive weight. In other terms (4.2) can be obtained by "smoothing" the indicator function of the complement of the confidence interval into a quadratic function. The analytical advantage of (4.2) over (4.3) is tremendous. Both loss functions yield the best equivariant estimator $\delta_0(X,S) = X, \gamma_0(X,S) = cS^2$. (For (4.3) this is essentially Student's confidence interval.) In Theorem 2 we prove the inadmissibility of (δ_0,γ_0) under (4.2) or of γ_0 under (4.1). The question of the admissibility of Student's confidence interval, however, remains unsettled.

Theorem 2. The estimator

$$\delta_0(X,S) = X, \qquad \gamma_0(X,S) = S^2/[n(n-2)]$$

is inadmissible under loss (4.2). The estimator

$$\gamma_0(X,S) = S^2/[n(n-2)]$$

is inadmissible under loss (4.1).

Proof. According to our Theorem 1, if (δ_0, γ_0) is admissible under (4.2) then γ_0 is admissible under loss

$$L_1(\xi, \sigma; \gamma) = \left[\sigma^2 n^{-1} \gamma^{-1/2} + \gamma^{1/2}\right] \sigma^{-1}. \tag{4.4}$$

Since L_1 is continuous and strictly bowl-shaped (in fact it is convex in $\gamma^{1/2}\sigma^{-1}$) the best equivariant estimator γ_0 is inadmissible (see Brewster and Zidek (1974)). The case of loss (4.1) is treated similarly.

Theorem 2 does not have a constructive character. In fact explicit improvement over (δ_0, γ_0) is unknown. Interestingly enough the generalized Bayes estimator γ_1 of Brewster-Zidek which improves upon γ_0 under (4.4) does not lead to an improvement upon γ_0 under (4.1).

Acknowledgement

The author benefited from an interesting discussion with Jim Berger to whom he is sincerely grateful.

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1968). Linear-loss interval estimation of location and scale parameters, Biometrika, 55, 141–148.

Berger, J. (1985a). In defense of the likelihood principle: axiomatics and coherency, in Bayesian Statistics II (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith eds.), North-Holland, Amsterdam, 33–65.

Berger, J. (1985b). The frequentist viewpoint and conditioning, in Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer. (L. LeCam and R. Olshen eds.). Wadsworth, Belmont, 15–44.

Berger, J. (1985c). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. Ann. Statist. 6, 783–801.

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators, Ann. Statist. 2, 21–38.

Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. Ann. Math. Statist. 39, 29–48.

Brown, L. (1978). A contribution to Kiefer's theory of conditional confidence procedures. Ann. Statist. 6, 59–71.

Brownie, C. and Kiefer, J. (1977). The ideas of conditional confidence in the simplest setting. Commun. Statist. Theor. Meth. A6(8), 691–751.

Cohen, A. and Strawderman, W. (1973). Admissible confidence interval and point estimation for translation or scale parameters, Ann. Statist. 1, 545–550.

Efron, B. (1982). The Jacknife, the Bootstrap and Other Resampling Plans. Philadelphia; Society for Industrial and Applied Mathematics.

Efron, B. and Tibshirani, R. (1968). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science, 1, 54–77.

Farrell, R. H. (1968). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used, Ann. Statist. 38, 23–28.

Kiefer, J. (1975). Conditional confidence approach in multi-decision problems. in Multivariate Analysis IV (P. R. Krishnaiah, ed.). New York, Academic Press.

Kiefer, J. (1976). Admissibility of conditional confidence procedures. Ann. Statist. 4, 836–865.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators. Journ. Amer. Stat. Assoc. 72, 789–827.

Lehmann, E. L. (1959). Testing Statistical Hypothesis. New York, J. Wiley & Sons.

Lehmann, E. L. (1983). Theory of Point Estimation, New York, J. Wiley & Sons.

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

Rubin, D. B. (1984). Bayesianly justifiably and relevant frequency calculations for the applied statistician, Ann. Statist. 12, 1151–1172.

Rukhin, A. L. (1986). How much better are better estimators of a normal variance, Technical Report #86-9, Department of Statistics, Purdue University.

Sandved, E. (1968). Ancillary statistics and estimation of the loss in estimation problems, Ann. Math. Statist. 39, 1755–1758.

Savage, L. J. (1954). The Foundations of Statistics, New York, J. Wiley & Sons.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. Ann. Inst. Statist. Math. 16, 155–160.

Strawderman, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. Ann. Statist. 2, 190–198.

Winkler, R. L. (1972). Decision-theoretic approach to interval estimation. Journ. Amer. Stat. Assoc. 67, 187–191.