Λ-MINIMAX ESTIMATION OF A SCALE PARAMETER

Daniel L. Solomon

BU-384-M August, 1971

Abstract

The intent of this note is to illustrate the application of the Λ-minimax principle to the estimation of a scale parameter. In particular, a Λ-minimax estimator of the variance of a Normal distribution is obtained when Λ is a subset of the class of natural conjugate prior distributions. Two cases considered are 1) the mean is known and 2) a fully specified prior distribution is available for the mean. Case 2) is reduced to case 1).

Biometrics Unit, Cornell University

Λ-MINIMAX ESTIMATION OF A SCALE PARAMETER

Daniel L. Solomon

The intent of this note is to illustrate the application of the Λ-minimax principle (see e.g., Zacks, [6, sec. 6.6]) to the estimation of a scale parameter. In particular, a Λ-minimax estimator of the variance of a Normal distribution is obtained when Λ is a subset of the class of natural conjugate prior distributions. Two cases considered are 1) the mean is known and 2) a fully specified prior distribution is available for the mean. Case 2) is reduced to case 1).

1. INTRODUCTION

One approach to the use of incomplete prior information is the application of the Λ-minimax principle. An earlier paper [4] lists some references to its use. It is assumed that although a prior distribution on the states of nature is not available, it is known to belong to some family, Λ of distributions. A Λ-minimax decision rule is then defined as one which minimizes the supremum over Λ of the Bayes risk. That is, if $r(\lambda,\delta)$ is the Bayes risk of δ with respect to the prior distrubution $\lambda\epsilon\Lambda$, then a Λ-minimax rule δ_o satisfies

$$\sup_{\lambda\epsilon\Lambda} r(\lambda,\delta_o) = \inf_{\delta} \sup_{\lambda\epsilon\Lambda} r(\lambda,\delta).$$

With a view towards determining a Λ-minimax estimate for the variance of a Normal distribution, suppose that $X_1, X_2,\ldots,X_n$ are independent, identically Normally distributed random variables with mean θ and variance $h^{-1} = \sigma^2$. Suppose that σ^2 is to be estimated with "percentage error" loss function with $[(a - \sigma^2)/\sigma^2]^2 = h^2(a - h^{-1})^2$ being the loss incurred when σ^2 is estimated by a.

Thus mis-estimation of small values of σ^2 is more costly than equal mis-estimation of large values.

2. MEAN KNOWN

Define $W = \frac{1}{n} \sum_{i=1}^{n} X_i^2$, with the convention that $W = 0$ for $n = 0$ and suppose that it is known that $\theta = 0$ (the results extend easily to known but arbitrary θ). Then W is a sufficient statistic and $(nh)W$ has the χ^2 distribution with n degrees of freedom. It is convenient to perform the calculation in terms of h, and if values of h are viewed as realizationsof a random variable $\tilde{h}$, then the natural conjugate prior distribution for $\tilde{h}$ is the gamma-2 with probability density

$$f_{\gamma_2}(h|v',n') = (\text{CONST.})e^{-\frac{1}{2}hn'v'}h^{\frac{1}{2}n'-1}, \quad n',v' > 0.$$

Suppose that the prior distribution is known except for v', and that the decision maker can specify

$$v'\epsilon U = \left\{v' \middle| \frac{\Delta}{A} \leq v' \leq \Delta A\right\}$$

where $\Delta \geq 0$, $A \geq 1$. Thus Δ "locates" U and A is a measure of its "size." Then Λ is the class of gamma-2 distributions, $\left\{f_{\gamma_2}(\cdot|v',n'),\ v'\epsilon U\right\}$.

Finally, as in Stone [5] and Griffin and Krutchkoff [1], we restrict the class of decision rules and consider only rules linear in the sufficient statistic. Thus, with w an observed value of W, we write $\delta(w) = bw + a$ as an estimator of $\sigma^2 = h^{-1}$ and seek a and b to minimize

$$\sup_{v'\epsilon U} E\tilde{h}^2(bW + a - \tilde{h}^{-1})^2$$

where the expectation is with respect to the joint distribution of W and $\tilde{h}$. We set $a = (1-b)c\frac{n'\Delta}{n'+2}$ (so that the minimization is now over b and c) and can show that

$$E\tilde{h}^2\left[bW + (1-b)c\frac{n'\Delta}{n'+2}\right] = \frac{2b^2}{n} + \left(\frac{2}{n'+2}\right)(1-b)^2 + \frac{n'}{n'+2}(1-b)^2\left[\frac{c\Delta}{v'} - 1\right]^2 .$$

This is a continuous convex function of $\frac{1}{v'}$ and the set

$$U = \left\{v' \middle| \frac{\Delta}{A} \le v' \le \Delta A\right\} = \left\{v' \middle| \frac{1}{\Delta A} \le \frac{1}{v'} \le \frac{A}{\Delta}\right\}$$

is convex and compact so the supremum over U occurs at an extreme point of U (See for example, Hadley [2, p.91]). Thus

$$B(b,c) \equiv \sup_{v'\in U} E\tilde{h}^2\left[bW + (1-b)c\frac{n'\Delta}{n'+2}\right]$$

$$= \frac{2b^2}{n} + \left(\frac{2}{n'+2}\right)(1-b)^2 + \frac{n'}{n'+2}(1-b)^2\max\left\{\left(\frac{c}{A} - 1\right)^2, (cA-1)^2\right\}.$$

We seek b and c to minimize B(b,c). With some work it can be shown that the minimizing values are (details are available in [3])

$$c_o = 2A/(A^2+1), \quad b_o = \frac{\frac{2+n'G^2}{n'+2}}{\frac{2}{n} + \frac{2+n'G^2}{n'+2}}, \quad G \equiv \frac{A^2-1}{A^2+1} .$$

The Λ-minimax rule is thus

$$\delta_o(w) = b_o w + (1-b_o)2\left(\frac{A}{A^2+1}\right)\left(\frac{n'}{n'+2}\right)\Delta$$

which is the Bayes rule

$$\delta(w) = \frac{\frac{w}{n' + 2} + \frac{v'}{n}\left(\frac{n'}{n' + 2}\right)}{\frac{1}{n' + 2} + \frac{1}{n}}$$

when A = 1. (The posterior distribution of $\tilde{h}$ is gamma-2 with parameters

$$n'' = n' + n \quad \text{and} \quad v'' = \frac{nw + n'v'}{n' + n} = \frac{\frac{w}{n'} + \frac{v'}{n}}{\frac{1}{n'} + \frac{1}{n}} .$$

The posterior mean is $\frac{1}{v''}$, and so the Bayes estimate of $\tilde{\sigma}^2$ for quadratic loss is v''. Recall that the loss function for this problem, $[(\delta(w) - \sigma^2)/\sigma^2]^2$ is not quadratic, which accounts for the apparent discrepancy.)

The Λ-minimax value of δ_o is

$$B(b_o, c_o) = \frac{2(2 + n'G^2)}{2(n' + 2) + n(2 + n'G^2)} \tag{1}$$

which does not depend on Δ. The Bayes risk (of the Bayes rule) is obtained by substituting A = 1 (and therefore G = 0) in $B(b_o, c_o)$ giving $2/(n + n' + 2)$. The efficiency of the Λ-minimax rule with respect to the Bayes rule is taken to be the ratio of these two risks. So

$$\text{Eff.} = \frac{\frac{2}{n + n' + 2}}{\frac{2(2 + n'G^2)}{2(n' + 2) + n(2 + n'G^2)}} = \frac{2(n' + 2) + n(2 + n'G^2)}{(n + n' + 2)(2 + n'G^2)} .$$

Figures 1 and 2 give contours of the efficiency for sample sizes n = 2 and 20. For given n, the label of the contour at the point (A,n') is the efficiency of the Λ-minimax rule with respect to the Bayes rule when the degrees of freedom parameter of the prior distribution is n', and the incompleteness specification U, is determined by A. Recall that as A decreases toward 1, the Λ-minimax rule becomes the Bayes rule. Also note that only for large values of n', do increases in A significantly decrease the efficiency. Under the prior distribution, the variance of $\tilde{h}$ is a decreasing function of n'. Thus large values of n' correspond to tight prior distributions and in this case the importance of their being well specified (small A) is apparent from the contours. Finally observe by comparing figures 1 and 2, that the importance of small values of A, decreases as the sample size, n, increases since the weight $(1-b_o)$ attached to the prior information decreases.

(Insert Figures 1 and 2)

3. MEAN UNKNOWN

Suppose now that the mean θ is not known, and let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $V = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where V = 0 if $n \leq 1$. The joint density of $\bar{X}$ and V is then

$$f(\bar{x},v|\theta,h) = (\text{CONST.})\left[e^{-\frac{1}{2}hn(\bar{x}-\theta)^2}h^{\frac{1}{2}}\right]\left[e^{-\frac{1}{2}hrv}h^{\frac{1}{2}r}v^{\frac{1}{2}r-1}\right]$$

where r = n - 1, and the natural conjugate prior for $\tilde{\theta}$ (now viewed as a random variable with generic values θ) and $\tilde{h}$ is the Normal-gamma with density

$$f_{N_\gamma}(\theta, h|\mu, v', n^*, n') = f_N(\theta|\mu, hn^*) f_{\gamma_2}(h|v', n') \tag{2}$$

$$= (\text{CONST.}) \left[e^{-\frac{1}{2}hn^*(\theta-\mu)^2} h^{\frac{1}{2}} \right] \left[e^{-\frac{1}{2}hn'v'} h^{\frac{1}{2}n'-1} \right]$$

for $-\infty \le \theta \le \infty$, $h \ge 0$, and v', n^*, $n' > 0$.

Suppose that h^{-1} is to be estimated by a function δ of the sufficient statistic, V, with loss $h^2(\delta(v) - h^{-1})^2$ as in the previous section and also as before that prior knowledge about v' is incomplete but that the decision maker can specify

$$v' \epsilon U = \left\{ v' | \frac{\Delta}{A} \le v' \le \Delta A \right\},$$

where $\Delta \ge 0$, $A \ge 1$. Here Λ is the set of Normal-gamma distributions (2), with $v' \epsilon U$.

Restricting, as before, the discussion to rules linear in V, we write $\delta(v) = bv + (1-b)c \dfrac{n'\Delta}{n' + 2}$ and seek b and c to minimize $\sup\limits_{v' \epsilon U} E\tilde{h}^2(\delta(V) - \tilde{h}^{-1})^2$. The required computations are virtually identical to those of section 2, and the Λ-minimax rule when θ is unknown is

$$\delta_1(v) = b_1 v + (1-b_1) 2 \frac{A}{A^2 + 1} \frac{n'}{n' + 2} \Delta,$$

where

$$b_1 = \frac{\dfrac{2 + n'G^2}{n' + 2}}{\dfrac{2}{r} + \dfrac{2 + n'G^2}{n' + 2}}, \qquad G = \frac{A^2 - 1}{A^2 + 1} .$$

Furthermore, the Λ-minimax value is, by comparison with (1)

$$\frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)} .$$

It follows that the contours given in figures 1 and 2, of the efficiency of the Λ-minimax rule with respect to the Bayes rule, can be used when θ is unknown if n is replaced by $r = n - 1$.

REFERENCES

[1] Griffin, B. S. and Krutchkoff, R. G. "Optimal linear estimators: an empirical Bayes version with application to the binomial distribution," Biometrika, 58 (1971), 195-201 .

[2] Hadley, G. Nonlinear and Dynamic Programming, Reading, Mass.: Addison-Wesley, (1964).

[3] Solomon, D. L. "Λ-minimax estimation I: two terminal action problems," Paper No. BU-278-M in the mimeo series, Biometrics Unit, Cornell University, (1970).

[4] Solomon, D. L. "Λ-minimax estimation of a multivariate location parameter.," Paper No. BU-375-M in the mimeo series, Biometrics Unit, Cornell University, (1971).

[5] Stone, M. "Robustness of non-ideal decision procedures," Journal of the American Statistical Association, 58, (1963), 480-486.

[6] Zacks, S. The Theory of Statistical Inference, New York: John Wiley and Sons, Inc., 1971.

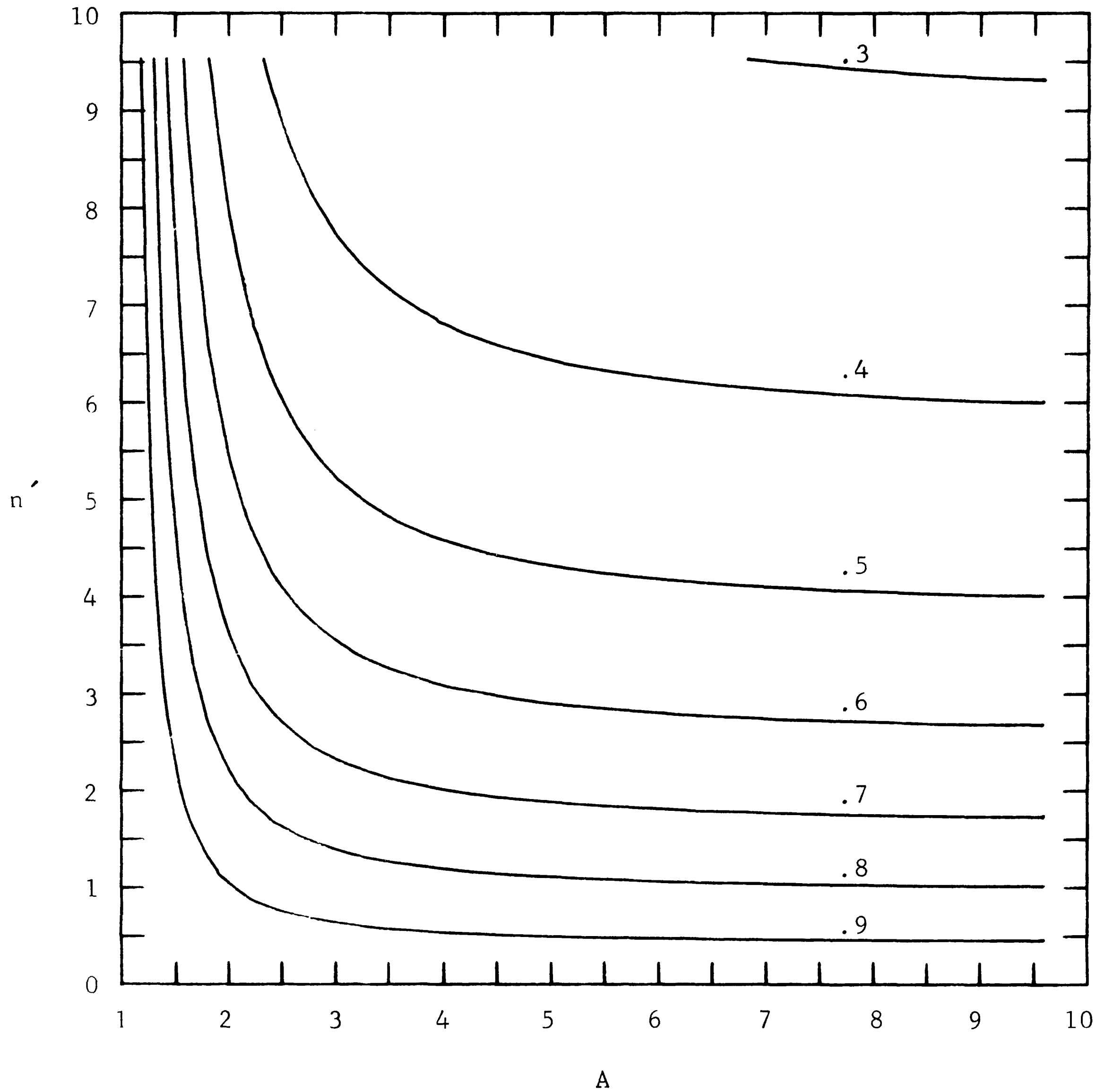

Fig. 1.--Contours of: (risk of Bayes estimate)/(risk of partially Bayes estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n', the incompleteness specification is determined by A, and the sample size is $n = 2$.

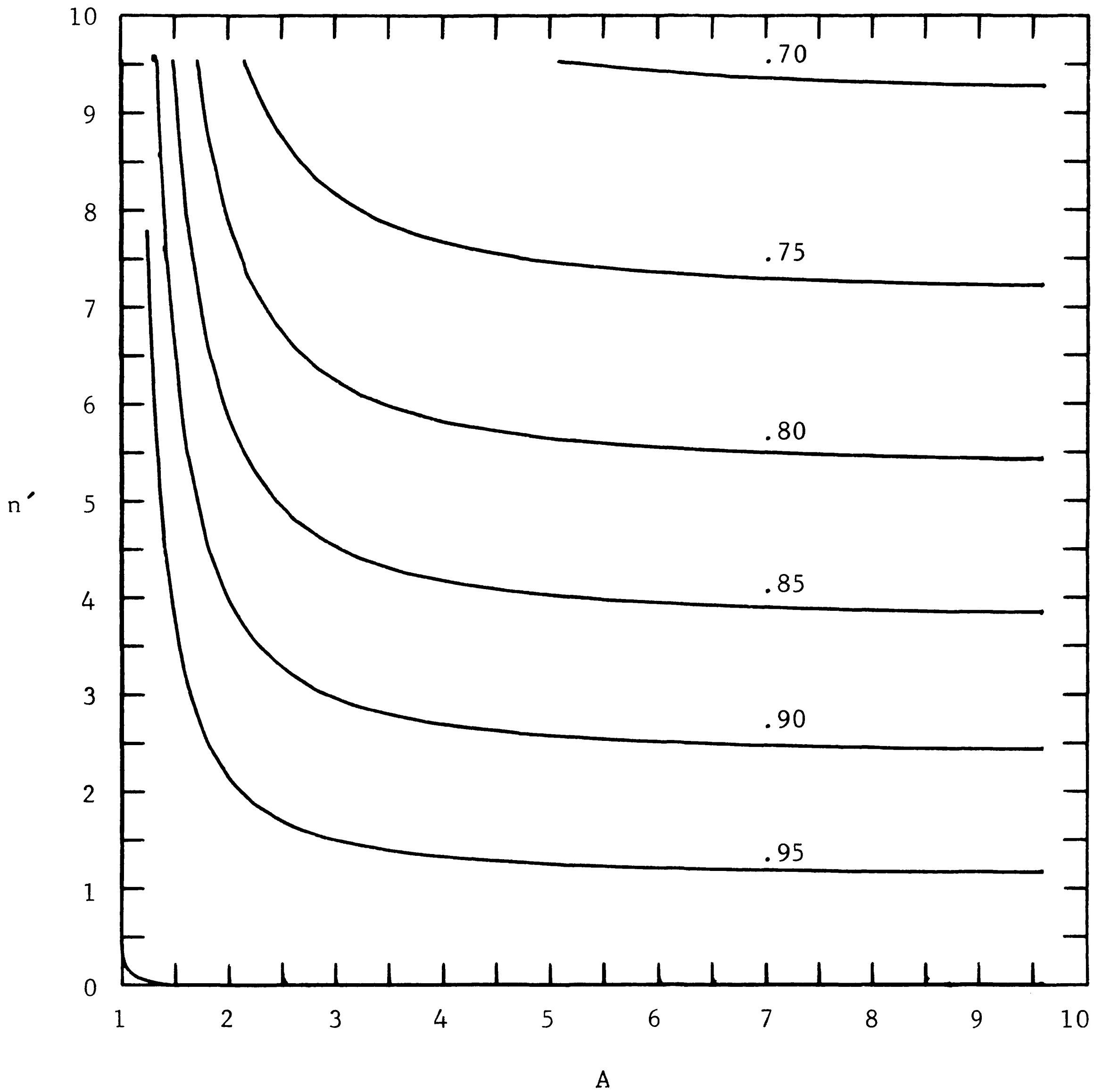

Fig. 2.--Contours of: (risk of Bayes estimate)/(risk of partially Bayes estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is n′, the incompleteness specification is determined by A, and the sample size is n = 20.