SOME REMARKS ON ESTIMATING A NONCENTRALITY PARAMETER

by

Michael D. Perlman* and Uttara Rasmussen

Technical Report No. 203

March 1973

University of Minnesota
Minneapolis, Minnesota

*Research supported in part by National Science Foundation Grant No. GP-34482

ABSTRACT

Some Remarks on Estimating a Noncentrality Parameter

It is shown that for the problem of estimating the noncentrality parameter of a noncentral χ^2 - variate X with n degrees of freedom, the Bayesian and objectivist (frequentist) approaches lead to approximately the same estimator, X - n. Furthermore, an estimator uniformly better than X - n is obtained by an empirical Bayes method.

American Mathematical Society 1970 subject classification.

Primary 62A15, 62F10; Secondary 62C15.

Key words and phrases: noncentral chi-square variate, noncentrality parameter, estimator, Bayesian framework, objectivist framework, UMVUE, empirical Bayes method, inadmissibility, quadratic loss, two stage prior.

1. An example of apparently wide discrepancy between Bayesian and objectivist estimators.

Let X be a random variable distributed as $\chi_n^2(\theta)$, a noncentral chi-square distribution with n degrees of freedom and unknown noncentrality parameter θ. Consider the problem of estimating θ based on X. Efron (1970) presents this problem as an example where the Bayesian and objectivist approaches seemingly lead to widely differing estimators of θ. Stein (1959) presented essentially the same example in the context of interval estimation.

In the remainder of this section we follow Efron's discussion. For convenience we represent X as $\|Y\|^2$ and θ as $\|\mu\|^2$, where $Y : n \times 1$ has the multivariate normal distribution $N(\mu, I)$ with mean $\mu : n \times 1$ (only X is observed, not Y). To estimate θ based on X the objectivist approach, whereby θ is assumed fixed and it is desired to minimize the mean squared error, leads to the estimator

$$\theta^*(X) = X - n, \tag{1.1}$$

which is the uniformly minimum variance unbiased estimator (UMVUE). (This follows from the Rao-Blackwell Theorem and the completeness of X for θ, which in turn follows from the completeness of Y for μ).

For the Bayesian (subjectivist) approach, consider the natural conjugate prior distribution

$$\mu \sim N(0, \gamma I) \tag{1.2}$$

for μ, where $\gamma > 0$ is a given constant. This induces the prior distribution

$$\theta \sim \gamma \chi_n^2 \tag{1.3}$$

for θ. The joint distribution of Y and μ is

(1.4) $\begin{pmatrix} Y \\ \mu \end{pmatrix} \sim N\left[0, \begin{pmatrix} (1+\gamma)I & \gamma I \\ \gamma I & \gamma I \end{pmatrix}\right]$,

the posterior distribution of μ given Y is

(1.5) $\mu | Y \sim N[\gamma(1+\gamma)^{-1}Y, \gamma(1+\gamma)^{-1}I]$,

and hence the posterior distribution of θ given X is

(1.6) $\theta | X \sim \gamma(1+\gamma)^{-1}\chi_n^2(\gamma(1+\gamma)^{-1}X)$.

Now, in the absence of strong prior opinions it is customary to adopt a "noninformative," or "diffuse", prior distribution. This is achieved by letting $\gamma \to \infty$ in (1.2) or (1.3), and leads to the posterior distribution

(1.7) $\theta | X \sim \chi_n^2(X)$.

Then for quadratic loss, this Bayesian analysis yields the estimator

(1.8) $\tilde{\theta}(X) = E[\theta | X] = X + n$.

Since $\tilde{\theta} - \theta^* = 2n$, as $n \to \infty$ there is apparently a wide discrepancy between the objectivist and subjectivist estimators for this problem. Furthermore not only is the discrepancy large but $\tilde{\theta}$ has uniformly greater mean squared error than θ^*, so the subjectivist estimator is inadmissible in the objectivist framework.

At first glance there seems to be no way to avoid this unpleasant conflict. Efron (1970, p. 1051) writes that "the nub of the difficulty... is that both the objective and subjective decision theorists have very definite and very different 'frameworks of replication' in mind." Since the two approaches entail very different assumptions and therefore have very different goals it is not surprising, indeed it is inevitable, that the two approaches often lead to widely differing estimates.

In the next section we present an argument, essentially due to Leonard J. Savage, showing that the apparently large discrepancy vanishes on closer examination and that the Bayesian approach when properly interpreted leads approximately to the same estimator $\theta^* = X - n$ as the objectivist approach. We are indebted to Richard Olshen who communicated the essentials of Savage's argument to us.

2. Savage's reconciliation of the Bayesian and objectivist approaches.

Closer examination casts doubt on the "Bayes" estimator $X + n$. Since $E[X|\theta] = \theta + n$, for any proper prior distribution we have $E[X - n] = E[\theta]$ while $E[X + n] = E[\theta] + 2n$, where these expectations are computed with respect to the marginal distributions of X and θ. Therefore, we have for the Bayes estimator $E[\theta|X]$ that

$$E\{(X - n) - E[\theta|X]\} = 0,$$

$$E\{(X + n) - E[\theta|X]\} = 2n,$$

so

$$E\{(X + n) - E[\theta|X]\}^2 = E\{(X - n) - E[\theta|X]\}^2 + (2n)^2.$$

These equations suggest that for any proper prior distribution the Bayes estimator $E[\theta|X]$ is more likely to assume values near $X - n$ than $X + n$, where "likely" is interpreted with respect to the marginal distribution of X. We must consider the marginal, rather than the conditional, distribution of X since in the Bayesian framework of replication this marginal distribution governs the values assumed by X. In the remainder of this section we make these considerations more precise (see (2.2) and (2.3)) for the particular prior distributions introduced in section 1.

In the preceding section the "Bayes" estimator $\tilde{\theta} = X + n$ resulted from the improper prior distribution obtained by letting $\gamma \to \infty$ in (1.2) or (1.3). Although the posterior distribution (1.7) thus obtained is proper, the resulting estimator $X + n$ is inadmissible. The reader is undoubtedly aware of other inadmissible improper Bayes estimators (e.g., the estimator Y of μ when $n \geq 3$).

Rather than using an improper prior, let us remain proper Bayesians and adopt the proper prior distribution (1.2) or (1.3) with γ large but finite. From (1.6) we obtain the proper Bayes estimator

$$(2.1) \qquad \tilde{\theta}_\gamma(X) = \gamma(1+\gamma)^{-1}\,[\gamma(1+\gamma)^{-1} X + n].$$

Although $\tilde{\theta}_\gamma \to X + n$ as $\gamma \to \infty$ for fixed X, we show now that <u>for any finite value of</u> γ <u>no matter how large, under the marginal distribution of</u> X <u>the proper Bayes estimator</u> $\tilde{\theta}_\gamma$ <u>assume values much closer to</u> X - n <u>than to</u> X + n. Precisely, we show now that under the marginal distribution of X induced by the prior (1.3) we have, <u>uniformly</u> in γ, that

$$(2.2) \qquad |(X - n) - \tilde{\theta}_\gamma(X)| = 0_p(n^{\frac{1}{2}})$$

while

$$(2.3) \qquad |(X + n) - \tilde{\theta}_\gamma(X)| = 0_p(n^{\frac{1}{2}}) + 2n.$$

From (1.4) it is seen that marginally $Y \sim N[0,(1+\gamma)\,I]$, so X has the marginal distribution

$$(2.4) \qquad X \sim (1+\gamma)\,\chi_n^2 .$$

From (2.1) and some algebra we find that

$$(2.5) \qquad |(X - n) - \tilde{\theta}_\gamma(X)| = \frac{1+2\gamma}{1+\gamma}\left|\frac{X}{1+\gamma} - n\right|,$$

$$(2.6) \qquad |(X + n) - \tilde{\theta}_\gamma(X)| = \frac{1}{1+\gamma}\left[\left(\frac{1+2\gamma}{1+\gamma}\right) X + n\right].$$

However, under the marginal distribution (2.4),

$$(2.7) \qquad \frac{X}{1+\gamma} - n \sim \chi_n^2 - n = 0_p(n^{\frac{1}{2}})$$

as $n \to \infty$ since $\chi_n^2 - n \to N(0,2n)$, so that uniformly in γ

$$(2.8) \qquad |(X - n) - \tilde{\theta}_\gamma(X)| \le 2 0_p(n^{\frac{1}{2}}) = 0_p(n^{\frac{1}{2}}),$$

which proves (2.2). Also from (2.6) and (2.7) we have

$$(2.9) \qquad |(X + n) - \tilde{\theta}_\gamma(X)| = \frac{1}{1+\gamma}\,[(1+2\gamma)(n + 0_p(n^{\frac{1}{2}})) + n] = 2n + 0_p(n^{\frac{1}{2}})$$

uniformly in γ, proving (2.3).

Basically, the flaw in the "Bayesian" reasoning leading to the estimate X + n is that the marginal distribution of X was ignored. Earlier it was remarked that the Bayes estimator $\tilde{\theta}_\gamma(X) \to X + n$ as $\gamma \to \infty$ for fixed X. Under the marginal distribution (2.4), however, X is not fixed as $\gamma \to \infty$; rather, $X \cong (1 + \gamma)\, n \to \infty$ as $\gamma \to \infty$. If one realizes this and notices that $\tilde{\theta}_\gamma(X) = X - n$ when $X = (1 + \gamma)n$, one will not be misled.

One further remark is called for. In section 1 an improper prior distribution for θ, obtained by letting $\gamma \to \infty$, was used to represent an "absence of strong prior opinion". To the contrary, however, this improper prior strongly favors arbitrarily large θ values and lends to the overestimate $X + n$. Rather than resorting to improper priors in the absence of prior information it seems wiser to abandon the Bayes approach altogether and seek procedures which have desirable objectivist properties, such as admissibility, UMVUE, minimax, etc.

In this section we have essentially presented a Bayesian justification of the estimator $X - n$ for θ. The reader should consult Stein (1956, 1962) for an objectivist justification of $X - n$.

3. Obtaining an improved estimator by the empirical Bayes method.

Consider for a moment the problem of estimating the mean vector μ of the multivariate normal vector Y introduced in section 1. Stein's famous proof (James and Stein (1961)) of the inadmissibility under squared error loss of the UMVUE $\mu^*(Y) = Y$ can be motivated as follows (see Stein (1962, 1966)). If μ has the prior distribution $N(0,\gamma I)$ of (1.2), the Bayes estimator of μ based on Y is given by $\tilde{\mu}_\gamma(Y) = \gamma(1+\gamma)^{-1}Y$. If γ is unknown we can attempt to estimate $\gamma(1+\gamma)^{-1} = 1-(1+\gamma)^{-1}$ on the basis of the marginal distribution of Y, i.e., $Y \sim N[0,(1+\gamma)I]$. Since $\|Y\|^2 \sim (1+\gamma)\chi_n^2$ we have $E[(n-2)\|Y\|^{-2}] = (1+\gamma)^{-1}$ whenever $n \geq 3$, so we are led to consider the estimator $[1-(n-2)\|Y\|^{-2}]Y$ for μ. Calculations then show that this estimator has uniformly smaller mean squared error than the UMVUE Y.

The above method of motivating an improved estimator by estimating an unknown parameter of the prior distribution is called the empirical Bayes method by Efron and Morris (1972a). Additional examples of this technique are given by Efron and Morris (1972b) and Sutherland, Holland, and Fienberg (1974). It is of interest to apply this method to our original problem of estimating the noncentrality parameter θ on the basis of X, to see if the UMVUE X - n can be improved.

Under the prior distribution (1.3) for θ we know that the Bayes estimator is $\tilde{\theta}_\gamma(X)$, given by (2.1). Since the marginal distribution of X is $(1+\gamma)\chi_n^2$ we have $E(n^{-1}X - 1) = \gamma$, so we might estimate γ by $n^{-1}X - 1$ and $1+\gamma$ by $n^{-1}X$. Substituting these estimates into (2.1) we obtain the estimator X - n, the UMVUE.

Alternatively, let us write $\gamma(1+\gamma)^{-1} = 1-(1+\gamma)^{-1}$ and express the

Bayes estimator $\tilde{\theta}_\gamma(X)$ of (2.1) as

(3.1) $$\tilde{\theta}_\gamma(X) = X - 2(1+\gamma)^{-1}X + (1+\gamma)^{-2}X + n - (1+\gamma)^{-1}n.$$

Since $EX = (1+\gamma)n$ and $EX^2 = (1+\gamma)^2(2n+n^2)$ this suggests estimating $(1+\gamma)^{-1}$ and $(1+\gamma)^{-2}$ by nX^{-1} and $(2n+n^2)X^{-2}$, respectively. Substituting these estimates into (3.1) we obtain the estimator $X - n + 2nX^{-1}$. Thus the arguments in this and the preceding paragraph lead us to consider estimators of the form

(3.2) $$\hat{\theta}_b(X) = X - n + \frac{b}{X}$$

where b is a constant.

<u>Theorem</u>. If $n \geq 5$ and $0 < b < 4$, the estimator $\hat{\theta}_b$ has uniformly smaller mean square error than the UMVUE $X - n$.

<u>Proof</u>. Let E_θ represent expectation with respect to the conditional distribution of X given θ, i.e., $X \sim \chi^2_n(\theta) = \chi^2_{n+2K}$, where K is a Poisson variate with mean $\theta/2$. The mean squared error of $X - n$ is

$$E_\theta(X - n - \theta)^2 = 2n + 4\theta \equiv V_\theta,$$

while the mean squared error of $\hat{\theta}_b$ is

$$E_\theta(\hat{\theta}_b - \theta)^2 = V_\theta + 2bE_\theta\left[\frac{X-n-\theta}{X}\right] + b^2E_\theta\left[\frac{1}{X^2}\right]$$

$$= V_\theta + 2b - 2b(n+\theta)E_\theta\left[\frac{1}{n+2K-2}\right] + b^2E_\theta\left[\frac{1}{(n+2K-2)(n+2K-4)}\right].$$

Now $n + 2K - 4 \geq 1$ since $n \geq 5$, and $P[n + 2K - 4 > 1] > 0$, so

$$E_\theta(\hat{\theta}_b - \theta)^2 < V_\theta + 2b - [2b(n+\theta) - b^2]E_\theta\left[\frac{1}{n+2K-2}\right]$$

$$= V_\theta + b\left\{2 - [2(n+\theta) - b]E_\theta\left[\frac{1}{n+2K-2}\right]\right\}.$$

Now by Jensen's inequality,

$$E_\theta\left[\frac{1}{n+2K-2}\right] > \frac{1}{E_\theta[n+2K-2]} = \frac{1}{n+\theta-2}.$$

Also, $b[2(n + \theta) - b] > 0$ since $n \geq 5$ and $0 < b < 4$. Thus

$$E_\theta(\hat{\theta}_b - \theta)^2 < V_\theta + b\{2 - \frac{2(n + \theta) - b}{n + \theta - 2}\} = V_\theta + \frac{b(b - 4)}{n + \theta - 2}$$

Since $b(b - 4) < 0$,

$$E_\theta(\hat{\theta}_b - \theta)^2 < V_\theta = E(X - n - \theta)^2$$

for all θ, proving the theorem.

Without appealing to the empirical Bayes method, the form of the Bayes estimators $\tilde{\theta}_\gamma$ in (2.1) might suggest trying to find a linear estimator of the form $\alpha X + \beta$ to uniformly improve $X - n$. However, it can be shown that no linear estimator has uniformly smaller mean square error than $X - n$. The empirical Bayes argument in the parapraph preceding (3.2) led us to the estimator $\hat{\theta}_{2n} = X - n + 2nX^{-1}$, to which the theorem does not apply since $2n$ exceeds 4 if $n \geq 5$. Thus the empirical Bayes approach serves here primarily to suggest the form of an estimator which dominates $X - n$. We suggest taking $b = 2$ in $\hat{\theta}_b$, since this value minimizes $b(b - 4)$.

Of course, the inadmissibility of $X - n$ is clear without recourse to the Theorem, since it may assume negative values and so is obviously dominated by $(X - n)^+$. However, whereas $(X - n)^+$ differs from $X - n$ only for small values of X, $\hat{\theta}_b(X)$ differs from (exceeds) $X - n$ for all values of X and thus the Theorem provides the qualitative information that the UMVUE $X - n$ is somewhat too small an estimate of θ over the entire range of X.

Since $\min_x \hat{\theta}_b(x) = 2b^{\frac{1}{2}} - n$ is negative when $n \geq 5$ and $0 < b < 4$, the estimator $\hat{\theta}_b$ itself may assume negative values and so is dominated by $(\hat{\theta}_b)^+$. However, both estimators $\hat{\theta}_b$ and $(\hat{\theta}_b)^+$ have the highly undesirable property that they approach $+\infty$ as $X \to 0$. To avoid this we might consider estimators of the form

$$\hat{\theta}_{b,c} = X - n + \frac{b}{c + X}$$

where $b > 0$ and $c > 0$ are constants. Incomplete calculations suggest that the estimators $\hat{\theta}_{b,c}$, and a fortiori $(\hat{\theta}_{b,c})^+$, uniformly dominate $X - n$ for all $n \geq 1$ and $c > 0$, provided b is sufficiently small, but we do not have a rigorous proof.

Finally, turn to the question of finding an admissible estimator which dominates $X - n$. The dominating estimators which we have considered, namely $(X - n)^+$, $(\hat{\theta}_b)^+$, and possibly $(\hat{\theta}_{b,c})^+$, are not smooth functions of X and therefore are not likely to be admissible. To obtain a (nonlinear) proper Bayes estimator which is admissible and possibly dominates $X - n$ we might follow the approach of Strawderman (1971) and consider a two-stage prior distribution for θ. Additional discussion of this method is given by Lindley and Smith (1972) and Sutherland, Holland, and Fienberg (1974). Referring to the notation of section 1, we now assume that both θ and γ are random parameters. The conditional distribution of θ given γ remains as in (1.3), while we assume that $\lambda \equiv (1 + \gamma)^{-1}$ has unconditional density proportional to

(3.3) $$\lambda^{-a} e^{-\frac{1}{2}c\lambda}, \quad 0 < \lambda < 1,$$

where a, c are constants satisfying $-\infty < a < 1$, $-\infty < c < \infty$. Since $\gamma (1 + \gamma)^{-1} = 1 - \lambda$, we see from (2.1) that the Bayes estimator of θ is

(3.4) $$\begin{aligned} \tilde{\theta}_{a,c}(X) &= E[\theta|X] = E\{E[\theta|X,\lambda] \,|X\} \\ &= E\{(1 - \lambda)[(1 - \lambda) X + n] \,|\, X\} \\ &= X - 2E[\lambda|X] X + E[\lambda^2|X] X + n - n E[\lambda|X]. \end{aligned}$$

From Strawderman (1971, p. 387) we find that the conditional distribution of λ given X is

$$g(\lambda|x) = \frac{\lambda^{\frac{n}{2}-a} \exp[-\tfrac{1}{2}\lambda(c + x)]}{J(x;\ a,c)}$$

where

$$J(x;\ a,c) = \int_0^1 \lambda^{\frac{n}{2}-a} \exp[-\tfrac{1}{2}\lambda(c + x)]\ d\lambda$$

can be evaluated in terms of the incomplete gamma integral. Integrating by parts we obtain

$$E[\lambda|X] = \frac{n + 2 - 2a}{c + X} - \frac{2 \exp[-\frac{1}{2}(c + X)]}{(c + X)\ J(X;\ a,c)},$$

$$E[\lambda^2|X] = \frac{(n + 2 - 2a)\ (n + 4 - 2a)}{(c + X)^2} - \frac{(n + 4 - 2a)2\exp[-\frac{1}{2}(c + X)]}{(c + X)^2\ J(X;\ a,c)} - \frac{2\exp[-\frac{1}{2}(c + X)]}{(c + X)\ J(X;\ a,c)} .$$

Substituting these expressions into the last line of (3.4), we obtain an exact formula for the proper Bayes admissible estimator $\tilde{\theta}_{a,c}$. For the simplest case where a = c = 0, this formula becomes

$$\tilde{\theta}_{0,0}(X) = (X - n) + \frac{4}{X} \{ (n + 2) - \frac{2\exp[-\frac{1}{2}X]}{J(X;\ 0,0)} \} + \frac{2\exp[-\frac{1}{2}X]}{J(X;\ 0,0)} - 4.$$

Even in this simplest use, however, we have been unable to compute the mean square error and compare it to that of X - n. We conjecture that the family $\{ \tilde{\theta}_{a,c} | -\infty < a < 1, -\infty < c < \infty \}$ contains estimators which uniformly improve X - n.

Acknowledgment. We wish to thank Brad Efron and Richard Olshen for many helpful suggestions.

References

[1] Efron, B. (1970). Comments on Blyth's paper. Ann. Math. Statist. 41 1049-1054.

[2] Efron, B. and Morris, C. (1972 a). Limiting the risk of Bayes and Empirical Bayes estimators - Part II: the empirical Bayes case. J. Amer. Stat. Assoc. 67 130-139.

[3] Efron, B. and Morris, C. (1972 b). Empirical Bayes on vector observations: an extension of Stein's method. Biometrika 59 335-347.

[4] James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. 1 361-379.

[5] Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model (with discussion). J. Royal Statist. Soc. Ser. B. 34 1 - 42.

[6] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. 1 197-206.

[7] Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. Ann. Math. Statist. 30 877-880.

[8] Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution (with discussion). J. Royal Statist. Soc. Ser. B 24 265-296.

[9] Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In Research Papers in Statistics, Festschrift J. Neyman (ed. F. N. David). Wiley, London.

[10] Strawderman, W. (1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist. 42 385-388.

[11] Sutherland, M., Holland, P. and Fienberg, S. (1974). Combining Bayes and frequency approaches to estimating a multinomial parameter. In Studies in Bayesian Econometrics and Statistics (ed. S. E. Fienberg and A. Zellner). North Holland, Amsterdam.