The Annals of Statistics
1984, Vol. 12, No. 3, 1035–1049

BAYES DOUBLE SAMPLE ESTIMATION PROCEDURES[1,2]

BY ARTHUR COHEN AND HAROLD B. SACKROWITZ

Rutgers University

Bayes double sample procedures are given for estimating the mean of exponential family distributions. The distribution can be multiparameter and so the case of normal mean with unknown variance is included. Linear combination loss functions are usually assumed. Stein's double sample procedure, and other intuitive double sample procedures for estimating a binomial parameter, are studied. Recommendations are given, including guidelines for the size of the initial samples.

1. Introduction and summary. Double sample inference procedures are intuitively appealing and are widely used by statisticians. It appears natural in a study to take a "pilot" sample that would help determine how many additional sample points should be drawn. Miller and Freund (1977), page 245, discuss estimating a binomial parameter p with a bounded precision. One can interpret their suggestion to be to take a first sample, estimate p by $\hat{p}_1$ say, and use this estimate to determine the size of the second sample in a standard way. Stein (1945) offered a double sample method of confidence estimation of the mean of a normal distribution. Stein's procedure can also be viewed as a double sample point estimate of the mean with specified precision. In both the Miller and Freund and the Stein references, the choice of the first sample size is arbitrary.

In this study we develop double sample Bayes estimation procedures for the mean of exponential family distributions. The procedures consist of stating n_1, the size of the first sample; $n_2(\mathbf{X}_1)$, the size of the second sample which depends on $\mathbf{X}_1$, the data from the first sample; and finally the point estimate. The loss functions usually are linear combinations of loss due to terminal decision and loss due to sampling. We also study vector losses where the components of the vector consist of a loss due to terminal decision and a loss due to sampling.

When the terminal decision loss is squared error, a general and applicable procedure is obtained for the general exponential family using conjugate priors. The exponential family can be multidimensional, although in this paper we estimate only one mean. Thus our treatment includes estimating a normal mean when the variance is unknown. Miller and Freund's procedure and Stein's procedure are seen to be appropriate for some loss functions but not others. The

---

Received March 1983; revised March 1984.

[1] Research supported by NSF-MCS-81-18564-01.

[2] This paper was delivered at a Statistical Research Conference dedicated to the memory of Jack Kiefer and Jacob Wolfowitz. The conference was held at Cornell University, July 6–9, 1983, and was a special meeting of the Institute of Mathematical Statistics, cosponsored by the American Statistical Association. Financial support was provided by the National Science Foundation, the office of Naval Research, the Army Research Office-Durham, and units of Cornell University.

AMS 1980 *subject classifications*. Primary 62L12; secondary 62C10.

Key words and phrases. Double sample, Bayes procedures, expected sample size, exponential family, Stein's 2 stage procedure, linear combination loss function, vector loss function.

Bayes procedures are also studied numerically and compared with optimal one-stage sampling procedures, and compared with Miller-Freund procedures. The double sample Bayes procedures compare very favorably. For the normal case, considerable attention and computation is devoted to Stein type procedures and some guidelines are given for how to choose the first sample in a Stein type procedure. Recommendations are given that also pertain to possible variations of the procedure.

In Section 2 we discuss Bayes double sample estimation procedures in the exponential family model. Section 3 is concerned with these procedures for the binomial case, whereas the normal case with unknown mean and unknown variance is studied in Section 4.

2. Double sample Bayes estimators. Let X be a $d \times 1$ random vector with distribution

$$dP_\theta(x) = e^{x \cdot \theta - M(\theta)} d\mu(x),$$

where $M(\theta) = \log \int e^{\theta x} \, d\mu(x)$, μ is a σ-finite measure on the Borel sets of R^d, $\theta \in \Theta = \{\theta \mid M(\theta) < \infty\}$. Assume that $\mathscr{X}$, the interior of the convex hull of the support of μ is a nonempty open set in R^d and Θ is a nonempty open set in R^d. It is known, (See for example Diaconis and Ylvisaker, 1979) that

$$\begin{aligned} E(X \mid \theta) = E_\theta(X) &= \nabla M(\theta) = (\partial M(\theta)/\partial\theta_1, \cdots, \partial M(\theta)/\partial\theta_d)^t \\ &= (M_1(\theta), M_2(\theta), \cdots, M_d(\theta))^t \end{aligned} \tag{2.1}$$

and

$$E_\theta(X - \nabla M(\theta))(X - \nabla M(\theta))^t = M''(\theta) = \left(\frac{\partial^2 M(\theta)}{\partial\theta_i \partial\theta_j}\right)_{i,j=1}^d = (M_{ij})_{i,j=1}^d. \tag{2.2}$$

Let the conjugate prior probability distribution on Θ be

$$\pi_{n_0, x_0}(\theta) = \beta(n_0, x_0)\exp(n_0 x_0 \cdot \theta - n_0 M(\theta)) \, d\theta, \quad n_0 \in R^1, \quad x_0 \in R^d, \tag{2.3}$$

where n_0 and x_0 satisfy the conditions of Theorem 1 of Diaconis and Ylvisaker (1979) and $\beta(n_0, x_0)$ is to enable (2.3) to integrate to 1. From this same reference we conclude that if $X_1, X_2, \cdots, X_n$ is a random sample of size n from P_θ, then the posterior distribution is $\pi_{n_0+n, (n_0 x_0 + n\bar{X})/(n_0+n)}$ and

$$E(\nabla M(\theta) \mid X_1, X_2, \cdots, X_n) = (n_0 x_0 + n\bar{X})/(n_0 + n), \tag{2.4}$$

where $\bar{X}$ is the sample mean.

The problem is to estimate $\omega = M_1(\theta)$, namely the expected value of $X^{(1)}$, the first coordinate of X, with a double sample. Hence the actions consist of triples $a = (n_1, n_2, \tau)$, where n_1 represents the size of the first sample, n_2 represents the size of the second sample, and τ represents the estimate. Note $n = n_1 + n_2$. Let c be a constant representing the cost of an observation. For the most part the loss functions are either

$$L_1(\omega, a) = (\tau - \omega)^2 + c(n_1 + n_2) \tag{2.5}$$

or

$$L_2(\omega, a) = (\tau - \omega)^2 + c \log(n_1 + n_2). \tag{2.6}$$

For (2.6) we assume at least one observation will be taken or that when $(n_1 + n_2) = 0$, $\log(n_1 + n_2)$ is replaced by 0. We also study the vector loss functions

$$L_{v_1}(\omega, a) = ((\tau - \omega)^2, c(n_1 + n_2)) \tag{2.7}$$

or

$$L_{v_2}(\omega, a) = ((\tau - \omega)^2, c \log(n_1 + n_2)). \tag{2.8}$$

In some cases we replace $(\tau - \omega)^2$ by $(\tau - \omega)^2/M_{11}(\theta)$. Now let $Y_1 = \sum_{i=1}^{n_1} X_i$, $Y_2 = \sum_{i=n_1+1}^{n_2} X_i$, and let $dP(y_1)$ denote the marginal distribution of Y_1 (not dependent on θ). Let $Z_1 = E[M_{11}(\theta) \mid Y_1]$ and $Z_2 = E[M_{11}(\theta) \mid Y_1, Y_2]$. Consider

$$n_2 = [[(Z_1/c)^{1/2} - n_1 - n_0]]^+, \tag{2.9}$$

where $[[b]]^+$ means the largest integer $\le b$ if $b > 0$ and 0 otherwise;

$$B = \{y_1: Z_1 \le c(n_1 + n_0)^2\}, \tag{2.10}$$

and

$$\int_B \left[\left(\frac{Z_1}{n_1 + n_0}\right) + n_1 c\right] dP(y_1) + \int_{B'} [2(Z_1 c)^{1/2} - n_0 c]\, dP(y_1). \tag{2.11}$$

THEOREM 2.1. *For the loss function in* (2.5) *the Bayes procedure with respect to the prior in* (2.1) *is to choose* τ *to be the first coordinate of* (2.4), n_2 *as in* (2.9) *and* n_1 *to minimize* (2.11).

PROOF. It is well known that the Bayes procedure is obtained by working backwards. (See for example, Ferguson, 1967, Section 7.2.) That is, we decide on the terminal decision assuming we have observed (Y_1, Y_2). We then decide on n_2 assuming n_1, y_1 are fixed and finally we decide on n_1. The choice of τ for the terminal decision is clear. Now write the Bayes risk as

$$\begin{aligned} &EE_\theta(E(M_1(\theta) \mid Y_1, Y_2) - M_1(\theta))^2 + cEE_\theta(n_1 + n_2) \\ &= E(\mathrm{Var}(M_1(\theta) \mid Y_1, Y_2)) + cEE_\theta(n_1 + n_2) \\ &= \int E\{[\mathrm{Var}(M_1(\theta) \mid Y_1, Y_2)] \mid Y_1\}\, dP(y_1) + c \int (n_1 + n_2)\, dP(y_1). \end{aligned} \tag{2.12}$$

An integration by parts in the expression for the posterior variance of $M_1(\theta)$ given (Y_1, Y_2) yields

$$\mathrm{Var}(M_1(\theta) \mid Y_1, Y_2) = E(M_{11}(\theta) \mid Y_1, Y_2)/(n_0 + n_1 + n_2), \tag{2.13}$$

when the left hand side is finite. (When the left hand side is infinite, (2.13) still holds. See Woodroofe, 1981 and 1983.)

Substitute (2.13) into (2.12) to find that the Bayes risk is

$$\int \left[\frac{Z_1}{n_0 + n_1 + n_2} + c(n_1 + n_2)\right] dP(y_1). \tag{2.14}$$

To minimize (2.14) we first minimize the integrand for fixed n_1 and y_1. Treating n_2 as a continuous variable, we find by differentiating with respect to n_2 that the integrand is minimized by choosing n_2 according to (2.9). Substituting this value of n_2 into (2.14) then yields the fact that n_1 need be chosen to minimize (2.11). □

REMARK. In this section and some of the other sections, some small degree of approximation is implicitly assumed. The need for approximation is because n_2 and n_1 must be integers and yet in the minimization problems we treat them as continuous variables. Hence by using $[[b]]^+$ we are introducing an approximation. The approximation can be avoided when the computation of a procedure can be executed and one can distinguish the posterior risks given $[[b]]^+$ and $[[b]]^+ + 1$ for every value of (n_1, y_1).

Later we will develop the Bayes procedures for the binomial case and the normal mean case when the variance is unknown. It will be noted that neither Miller and Freund's suggestion nor Stein's method will ensue. However, both of these methods will essentially be shown to be limits of Bayes procedures when the loss function is (2.6). Toward this end, consider

$$n_2 = [[(Z_1/c) - n_1]]^+, \tag{2.15}$$

$$B = \{y_1 \colon Z_1 \le cn_1\} \tag{2.16}$$

$$\int_B \left[\frac{Z_1}{n_1} + n_1 c\right] dP(y_1) + \int_{B'} [2(Z_1 c)^{1/2}]\, dP(y_1). \tag{2.17}$$

THEOREM 2.2. *Consider the loss function in* (2.6). *Consider the sequence of priors in* (2.1) *with* $n_0 \to 0$. *Then the* a.e. *limit of the sequence of Bayes procedures is to choose* $\tau = \bar{X}^{(1)}$, n_2 *as in* (2.15) *and* n_1 *to minimize* (2.17).

PROOF. The proof is essentially the same as in Theorem 2.1. That is, one derives the Bayes procedure for fixed n_0. The determination of τ is the same as before. The determination of n_2 is the solution of a quadratic equation whose solution in the limit is (2.15). Continuity implies that n_1 is chosen to minimize (2.17).

REMARK 2.3. In the normal mean case with unknown variance, we also consider generalized Bayes procedures using a uniform generalized prior on the mean and a proper prior on the variance. For both loss functions (2.5) and (2.6), this amounts to regarding $n_0 = 0$ and the resulting generalized Bayes procedures are as in Theorems 2.1 and 2.2.

REMARK 2.4. If in (2.5) and (2.6) squared error loss is replaced by squared error divided by $M_{11}(\theta)$, then it is known that the Bayes double sample procedure is a one-sample procedure. (See Whittle and Lane, 1967.) This, however, will not necessarily be the case if in (2.7) and (2.8) we replace squared error by squared error divided by $M_{11}(\theta)$ and utilize the flexibility this formulation permits in enlarging on the conjugate priors. See Cohen and Sackrowitz (1984b), Section 3.

REMARK 2.5. The determination of n_1 involves the minimization of (2.11) in the case of loss function (2.5) or (2.17) in the case of loss function (2.6). In all cases the determination of B, Z_1 and $P(y_1)$ is straightforward. In some cases the minimization step requires a computer. However, the computer program required is straightforward.

3. Binomial case. Let X_i, $i = 1, 2, \cdots$ be i.i.d. Bernoulli variables with parameter p. Then the natural parameter is $\theta = \log p/(1 - p)$, and $M_{11}(\theta) = e^{\theta}/(1 + e^{\theta})^2 = p(1 - p)$. Thus

$$Z_1 = E(M_{11}(\theta) \mid Y_1) = (y_1 + n_0 x_0) \frac{n_1 - y_1 + n_0(1 - x_0)}{(n_1 + n_0 + 1)(n_1 + n_0)}. \tag{3.1}$$

The Bayes estimator is $\tilde{\tau} = (n_0 x_0 + y_1 + y_2)/(n_0 + n_1 + n_2)$. Note that if $n_0 = 0$, (a limiting Bayes case) the estimator is the usual sample proportion $\hat{p}$. The quantity Z_1 may be expressed as $\hat{p}_1(1 - \hat{p}_1)(n_1/(n_1 + 1))$ where $\hat{p}_1 = y_1/n_1$. For the loss function in (2.5), we see from (2.9) that when (2.9) is positive, n_2 is determined by setting $(n_1 + n_2) = k(\hat{p}_1(1 - \hat{p}))^{1/2}$ where k is a constant. Thus $(n_1 + n_2)$ is set equal to a multiple of the estimated standard deviation of X_i. If the loss function is (2.6) however, from (2.15) we determine n_2 by setting $(n_1 + n_2)$ equal to a multiple of the sample variance. This then corresponds somewhat to one aspect of the intuitive procedure recommended by Miller and Freund (1977). This latter procedure determines n_2 by setting $\hat{p}_1(1 - \hat{p}_1)/(n_1 + n_2) = \Delta$, where Δ is a preselected constant. The procedure then estimates p by the sample proportion based on $n_1 + n_2$ observations.

Note that the best fixed sample size procedure for the prior (2.3) is to choose $N = [[\sqrt{n_0 x_0(1 - x_0)/c(n_0 + 1)} - n_0]]^+$ and estimate p by (2.4). The risk for this procedure when the loss function is (2.5) is

$$p^2(n_0^2 - N) + p(N - 2n_0^2 x_0) + (n_0 x_0)^2/(N + n_0)^2. \tag{3.2}$$

Some computations were done to assess the value of the double sample procedures. For the loss functions in (2.5) and (2.6) we studied the risks of the double sample Bayes procedure for the uniform prior, i.e. $n_0 x_0 = 1$, $x_0 = ½$, the single sample Bayes procedure, and some Miller and Freund (MF) procedures with a first sample size and Δ that would yield comparable risks. We varied the values of c. For the loss function in (2.5) we also found some optimal first sample sizes for varying values of c for the uniform prior. These are given in Table 3.1.

Table 3.2 gives risk functions for the single sample Bayes procedures, double sample Bayes procedures, and Miller-Freund procedures. The last column gives the risk for the fully sequential procedure which can serve as kind of a baseline. The loss is (2.5) and three values of c were chosen so that the single sample procedures had sample sizes of $N = 50, 100, 406$. MF-n_1, Δ means the Miller-

TABLE 3.1

Optimal values of n_1			
c	n_1	c	n_1
.05	0	.005	3
.03	0	.001	8
.02	1	.0005	11
.01	2	.0001	21

TABLE 3.2

Risk functions for loss (2.5)

$c = .000062 \quad N = 50$

p	2-sample	1-sample	MF-26, .064	MF-30, .064	MF-35, .064	sequential
0.000	0.00289	0.00346	0.00161	0.00186	0.00217	0.00297
0.050	0.00341	0.00427	0.00325	0.00338	0.00352	0.00332
0.100	0.00446	0.00500	0.00432	0.00438	0.00453	0.00439
0.150	0.00543	0.00564	0.00562	0.00538	0.00527	0.00540
0.200	0.00618	0.00619	0.00695	0.00657	0.00621	0.00618
0.250	0.00673	0.00666	0.00784	0.00752	0.00717	0.00674
0.300	0.00712	0.00704	0.00822	0.00801	0.00779	0.00714
0.350	0.00740	0.00734	0.00830	0.00819	0.00808	0.00742
0.400	0.00759	0.00756	0.00830	0.00824	0.00818	0.00760
0.450	0.00771	0.00768	0.00828	0.00825	0.00822	0.00771
0.500	0.00776	0.00773	0.00828	0.00826	0.00823	0.00775

$c = .000016 \quad N = 100$

P	2-sample	1-sample	MF-47, .0455	sequential
0.000	0.00117	0.00170	0.00075	0.00122
0.050	0.00166	0.00214	0.00170	0.00163
0.100	0.00235	0.00253	0.00229	0.00235
0.150	0.00284	0.00287	0.00306	0.00284
0.200	0.00308	0.00317	0.00357	0.00318
0.250	0.00344	0.00343	0.00387	0.00345
0.300	0.00364	0.00363	0.00395	0.00364
0.350	0.00379	0.00379	0.00400	0.00379
0.400	0.00389	0.00391	0.00404	0.00389
0.450	0.00395	0.00398	0.00408	0.00395
0.500	0.00397	0.00400	0.00409	0.00396

$c = .000001 \quad N = 406$

P	2-sample	1-sample	MF-136, .023	sequential
0.000	0.00019	0.000406	0.00014	0.000192
0.050	0.00043	0.000521	0.00046	0.000410
0.100	0.00060	0.000623	0.00069	0.000600
0.150	0.00071	0.000714	0.00084	0.000715
0.200	0.00080	0.000792	0.00087	0.000800
0.250	0.00086	0.000859	0.00091	0.000860
0.300	0.00092	0.000913	0.00094	0.000915
0.350	0.00095	0.000955	0.00097	0.000951
0.400	0.00096	0.000986	0.00099	0.000977
0.450	0.00099	0.001004	0.00100	0.000994
0.500	0.00100	0.001010	0.00101	0.000997

Freund procedure with initial sample n_1 and Δ. Symmetry permits values of p ranging between 0 and ½. Table 3.3 offers risks for the loss (2.6). In Table 3.2 the pattern is the same for all values of N. The double sample Bayes procedure is superior to the one-sample procedure for values of p near 0 and 1. MF is best at the very low and very high values, but for other values of p it is clearly the worst. Overall, the double sample procedures do very well since even when their

TABLE 3.3

Risk functions for loss (2.6)

$c = .00308$ $N = 50$

P	**2-sample**	**1-sample**	**MF-5, .069**	**MF-15, .069**	**MF-22, .069**	**sequential**
0.000	0.01126	0.01242	0.00496	0.00834	0.00952	0.01163
0.050	0.01152	0.01323	0.00865	0.01054	0.01138	0.01105
0.100	0.01263	0.01395	0.01442	0.01294	0.01279	0.01212
0.150	0.01412	0.01459	0.02028	0.01547	0.01443	0.01382
0.200	0.01521	0.01514	0.02516	0.01729	0.01600	0.01525
0.250	0.01585	0.01561	0.02865	0.01819	0.01706	0.01609
0.300	0.01621	0.01599	0.03078	0.01842	0.01755	0.01643
0.350	0.01643	0.01629	0.03185	0.01831	0.01768	0.01655
0.400	0.01656	0.01650	0.03224	0.01810	0.01767	0.01661
0.450	0.01663	0.01663	0.03230	0.01795	0.01763	0.01665
0.500	0.01666	0.01667	0.03229	0.01790	0.01762	0.01666

$c = .001601$ $N = 100$

P	**2-sample**	**1-sample**	**MF-25, .045**	**sequential**
0.000	0.00641	0.00747	0.00515	0.00662
0.050	0.00678	0.00791	0.00660	0.00650
0.100	0.00781	0.00830	0.00828	0.00764
0.150	0.00863	0.00865	0.00942	0.00867
0.200	0.00906	0.00895	0.00985	0.00918
0.250	0.00930	0.00920	0.00991	0.00936
0.300	0.00945	0.00941	0.00988	0.00947
0.350	0.00956	0.00957	0.00986	0.00956
0.400	0.00964	0.00968	0.00986	0.00963
0.450	0.00957	0.00975	0.00987	0.00968
0.500	0.00970	0.00978	0.00987	0.00969

$c = .000408$ $N = 404$

P	**2-sample**	**1-sample**	**MF-73, .025**	**sequential**
0.000	0.00193	0.00246	0.00175	0.00196
0.050	0.00229	0.00257	0.00227	0.00226
0.100	0.00264	0.00267	0.00279	0.00266
0.150	0.00277	0.00276	0.00292	0.00277
0.200	0.00285	0.00284	0.00296	0.00285
0.250	0.00291	0.00291	0.00299	0.00291
0.300	0.00296	0.00296	0.00302	0.00296
0.350	0.00299	0.00301	0.00305	0.00299
0.400	0.00300	0.00304	0.00397	0.00301
0.450	0.00302	0.00306	0.00308	0.00302
0.500	0.00303	0.00306	0.00308	0.00303

risks are not the lowest they are not exceeding the other risks by very much. Furthermore the risk for the double sample is extremely close to the risk of the fully sequential procedure. The pattern in Table 3.3 is similar although the MF-procedures do somewhat better here and look quite reasonable.

4. Estimating a normal mean when variance is unknown. Let X_i, $i = 1, 2, \cdots$ be i.i.d. normal variables with mean μ and unknown variance σ^2. It is easily seen that for this case $\theta_1 = \mu/\sigma^2$, $\theta_2 = -\frac{1}{2}\sigma^2$,

$$M(\theta) = (\mu^2/2\sigma^2) + \tfrac{1}{2}\log\sigma^2 = -\theta_1^2/4\theta_2 - \tfrac{1}{2}\log(-2\theta_2)$$

so that $M_1(\theta) = \mu$ and $M_{11}(\theta) = \sigma^2$. Let

$$y_1 = \textstyle\sum_{i=1}^{n_1} X_i, \quad y_2 = \sum_{i=n_1+1}^{n_2} X_i,$$

$$\bar{y}_1 = y_1/n_1, \quad \bar{y}_2 = y_2/n_2, \quad y = y_1 + y_2, \quad \bar{y} = y/(n_1 + n_2),$$

$$n = n_1 + n_2, \quad u = \textstyle\sum (x_i - \bar{y}_1)^2, \quad v = u/(n-1), \quad h = (1/\sigma^2).$$

In deriving the Bayes procedures it is convenient to put the priors directly on (μ, σ^2). Consider the normal-gamma prior (See Raiffa and Schlaifer, 1961, page 300) with 4 hyperparameters

$$\begin{aligned} f_{N_\gamma}(\mu, h \mid m', n_0, \rho, \nu) &= f_N(\mu \mid m', h, n_0) f_{\gamma^2}(h \mid \rho, \nu) \\ &\propto \exp\{-\tfrac{1}{2}hn_0(\mu - m')^2\}h^{\delta(n_0)/2}\exp\{-\tfrac{1}{2}\rho\nu h\}h^{(\nu/2)-1}, \end{aligned} \tag{4.1}$$

where $\rho, \nu, n_0 \geq 0$ and $\delta(n_0) = 1$ if $n_0 > 0$, $\delta(n_0) = 0$ if $n_0 = 0$. Note if $n_0 = 0$ the marginal prior on μ is a generalized prior. The posterior distribution for the above conjugate prior when n observations are taken is normal-gamma with parameters $m'' = (n_0 m' + n\bar{y})/(n + n_0)$, $n'' = n_0 + n$,

$$\rho'' = \frac{[\rho\nu + n_0 m'^2] + [u + n\bar{y}^2] - n''m''^2}{[\rho + \delta(n_0)] + (n-1)},$$

$\nu'' = [\nu + \delta(n_0)] + (n - 1)$. Thus the estimator is m'' and either from the posterior variance of μ or from the posterior expectation of $(1/h)$ (both derivable from Raiffa and Schlaifer) n_2 is determined from (2.9) with

$$Z_1 = \frac{[\rho\nu + n_0 m'^2] + u + n_1\bar{y}_1^2 - [(m'n_0)^2 + 2m'n_0 + n_1^2\bar{y}_1^2]/(n_1 + n_0)}{[(n_1 - 3) + \nu + \delta(n_0)]}. \tag{4.2}$$

The set B defined in (2.10) is readily determined as is the marginal joint distribution of $(\bar{y}_1, u)$. However, completing the description of the Bayes procedure for the loss functions (2.5) and (2.6) would require choosing n_1 to minimize (2.11), which in this case would require a two dimensional numerical intergration. It is of greater interest to treat the special case $n_0 = 0$. This is easily justified by intuition, invariance, and other considerations. When $n_0 = 0$, (4.2) becomes

(4.3) $$Z_1 = (\rho\nu + u)/[(\nu - 2) + (n_1 - 1)] = (\lambda\alpha + u)/[(n_1 - 1) + 2(\alpha - 1)],$$

if we set $\rho = \lambda/2$ and $\nu = 2\alpha$. (The parameterization in terms of λ and α is preferred since now the generalized prior is the product of a uniform prior on μ and a gamma prior with parameters (λ, α) on $h = (1/\sigma^2)$). For this prior, it follows that the marginal distribution of u is such that $w = \lambda/(u + \lambda)$ has a beta distribution with parameters $(\alpha, (n_1 - 1)/2)$. Thus the Bayes procedures for the loss functions in (2.5) and (2.6) can be determined. We summarize for the loss function in (2.6). (In the remainder of this section, we will assume the loss function is (2.6) and remark that (2.5) could be used with the obvious changes.)

THEOREM 4.1. *The generalized Bayes procedure for the generalized prior which is a uniform prior on μ and a gamma prior with parameters (λ, α) on h is to estimate by $\bar{y}$, take*

$$n_2 = [[(\lambda\alpha + u)/c[(n_1 - 1) + 2(\alpha - 1)] - n_1]]^+,$$

and choose n_1 to minimize

(4.4) $$\int_0^d \left[1 + \log\left\{\frac{wa(\alpha - 1) + a}{w(2\alpha + n_1 - 3)}\right\}\right] f(w)\,dw + \int_d^1 \left[\log n_1 + \frac{wa(\alpha - 1) + a}{w(2\alpha + n_1 - 3)}\right] f(w)\,dw$$

where $a = \lambda/c$, $d = a/(n_1[(n_1 - 3) + 2\alpha] - (\alpha - 1)a)$ and $f(w)$ has a beta distribution with parameters $(\alpha, (n_1 - 1)/2)$.

We remark that for given values of n_1 and α, (4.4) can be expressed as a finite sum and so with the computer, determination of n_1 is not difficult.

It will prove helpful to treat the case of Theorem 4.1 when $\alpha = 1$. Let n_1 be odd. Consider the following function of n_1:

(4.5) $$\begin{cases} 1 + \log(a/(n_1 - 1)) + \sum_{j=1}^{(n_1-1)/2} 1/j & \text{if } a \geq n_1(n_1 - 1) \\ 1 + \log(a/(n_1 - 1)) + \sum_{j=1}^{(n_1-1)/2} 1/j - (1 - (a/n_1(n_1 - 1)))^{(n_1+1)/2} \\ \quad - (1 + (a/2n_1))[\log(a/n_1(n_1 - 1)) \\ \quad + \sum_{j=1}^{(n_1-1)/2} (1 - (a/n_1(n_1 - 1)))^j/j] & \text{if } a < n_1(n_1 - 1). \end{cases}$$

COROLLARY 4.2. *The generalized Bayes procedure for the generalized prior which is a uniform prior on μ and an exponential prior with parameter $\lambda = ac$ on h is to estimate by $\bar{y}$, take $n_1 = [[((\lambda + u)/c(n_1 - 1)) - n_1]]^+$ and choose n_1 to minimize* (4.5).

PROOF. When $\alpha = 1$ (4.4) becomes

$$\int_0^{a/(n_1(n_1-1))} \left[1 + \log \frac{a}{w(n_1-1)}\right] \frac{n_1-1}{2} (1-w)^{(n_1-1)/2}\, dw$$

$$+ \int_{a/(n_1(n_1-1))}^1 \left[\log n_1 + \frac{a}{w(n_1-1)}\right] \frac{n_1-1}{2} (1-w)^{((n_1-1)/2)-1}\, dw$$

$$(4.6)\quad \begin{aligned} &= \left(1 + \log \frac{a}{(n_1-1)}\right)\left(1 - \left(1 - \frac{a}{n_1(n_1-1)}\right)^{(n_1-1)/2}\right) \\ &+ \left(1 - \frac{a}{n_1(n_1-1)}\right)^{(n_1-1)/2} \ln n_1 \\ &- \frac{(n_1-1)}{2} \sum_{j=0}^{((n_1-1)/2)-1} (-1)^j \binom{((n_1-1)/2)-1}{j} \int_0^{a/(n_1(n_1-1))} (\log w) w^j\, dw \\ &+ \frac{a}{2} \sum_{j=0}^{((n_1-1)/2)-1} (-1)^j \binom{((n_1-1)/2)-1}{j} \int_{a/(n_1(n_1-1))}^1 w^{j-1}\, dw. \end{aligned}$$

Performing the integrations in (4.6) and using the equalities

$$(4.7)\quad \sum_{j=1}^m (-1)^j \frac{1}{j} \binom{m}{j} (1 - r^j) = -\sum_{j=1}^m \frac{1}{j} (1-r)^j$$

$$(4.8)\quad \begin{aligned} &\sum_{j=0}^m (-1)^j \binom{m}{j} (j+1)^{-2} r^{j+1} \\ &= -\left(\frac{\sum_{j=1}^{m+1} (1-r)^j}{j} + \frac{\sum_{j=1}^{m+1} (-1)^j \binom{m+1}{j}}{j}\right) \Big/ (m+1) \end{aligned}$$

yields the result. □

Before proceeding to develop specific procedures which can be recommended, we note an interesting property of the risk of these generalized Bayes double sample procedures when compared to the risk of any one sample procedure with estimate $\bar{y}$. The risk for the one sample procedure is $(\sigma^2/n) + c \ln n = O(\sigma^2)$ as $\sigma^2 \to \infty$.

THEOREM 4.3. *The risk for the generalized Bayes procedures of Theorem* 4.1 *is* $O(\ln \sigma)$.

PROOF. Let $K = n_1 c[(n_1 - 1) + 2(\alpha - 1)] - \lambda\alpha$. Then the risk is

$$\int_0^K \left[\frac{\sigma^2}{n_1} + c \ln n_1\right] f(u; \sigma^2)\, du$$

$$(4.9) \qquad + c \int_K^\infty \sigma^2 \frac{(n_1 - 1) + 2(\alpha - 1)}{\lambda\alpha + u} + \ln\left[\frac{\lambda\alpha + u}{c((n_1 - 1) + 2(\alpha - 1))}\right] f(u; \sigma^2)\, du$$

where $f(u; \sigma^2)$ is the chi-squared density with scale parameter σ^2 and $(n_1 - 1)$ degrees of freedom. Substituting the chi-squared density in (4.9), making the change of variable $t = u/2\sigma^2$, integrating and using crude upper bounds yields that (4.9) is $O(\ln \sigma)$ as $\sigma \to \infty$. □

REMARK 4.4. For the loss function (2.5), the risk of the generalized Bayes double sample procedures can be shown to be $O(\sigma)$.

The significance of Theorem 4.3 and Remark 4.4 is that the double sample procedures will lead to dramatic improvements in risk for large σ when compared to any one sample procedure. Furthermore, this result does not depend on n_1.

Now we set out to study Stein's procedure and procedures which can be recommended. What we have been referring to as Stein's procedure is to choose $n_2 = [[(u/(n - 1)k) - n_1]]^+ + 1$ for k a positive constant. (The constant k is related to the coverage probability and width of a confidence interval.) Lehmann (1959), page 203 suggests $n_2 = [[(u/(n - 1)k) + 1 - n_1]]^+$. No mention is made of how to choose n_1.

The Bayes procedures of Theorem 4.1 and Corollary 4.2 choose $n_2 = [[c_1 + c_2 s^2 - n_1]]^+$ while the Stein-Lehmann procedures essentially choose $n_2 = [[1 + c_3 s^2 - n_1]]^+$ with c_1, c_2, c_3 constants. Hence we can address the question: Can Stein's procedure be regarded (approximately) as some generalized Bayes procedure? Furthermore, if so, what value of n_1 would correspond to such a procedure? In an effort to answer these questions, we ask whether the constant a of Corollary 4.2 can be chosen so that $c_1 = a/(n_1^* - 1) < 1$ where n_1^* is the optimal n_1. If we let $a \downarrow 0$, it can be verified by taking the limit in (4.4) that the Bayes risk becomes a multiple of $\ln n_1$ and so the optimal value for n_1 is $n_1 = 2$. (The same result is true for loss function (2.5).) Such a procedure approximates the Stein procedure but other choices of the constant a will also approximate the Stein procedure. Furthermore, it is worth considering the slight variations of the Stein procedure, namely those given by Corollary 4.2. We offer one more result that enhances the important computer work to be discussed. If σ^2 were known, then the optimal fixed sample size procedure in which the estimator is $\bar{y}$ is to choose $n = [[\sigma^2/c]]$. The risk for such a procedure is $c + c \ln \sigma^2/c$.

THEOREM 4.4. *The risk for any double sample procedure with estimator $\bar{y}$ is greater than or equal to* $(c + \ln \sigma^2/c)$.

PROOF. The risk for the double sample procedure is

$$E\{[\sigma^2/(n_1 + n_2(y_1, u))] + c \ln[n_1 + n_2(y_1, u))\}. \tag{4.10}$$

For fixed (y_1, u) the integrand is minimized at $n_1 + n_2(y_1, u) = \sigma^2/c$ and the minimum value is $c + c \ln \sigma^2/c$. □

REMARK 4.5. The lower bound in Theorem 4.4 is appropriate not only for double sample procedures but for k stage sample procedures $k \geq 2$ and also sequential procedures. Thus if a double sample procedure has a risk close to the lower bound, then that double sample procedure is very good since not much improvement can be achieved, not even among procedures that sample at more than 2 stages.

We now discuss computer work. Table 4.1 yields values of n_1^*, the optimal value of n_1 for various values of the parameter a of Corollary 4.2. For $a = 4$ $(a/(n_1 - 1)) = 1$ implying that this procedure is essentially Stein's procedure. Table 4.2 offers risks as a function of σ^2 of the double sample Bayes procedures when the cost is $c = .01$ and the sample sizes for the first sample are $n_1 = 2, 5, 15$. They are contrasted with the optimal risk given in Theorem 4.4. Table 4.3 is the same except that $c = .1$.

In both cases the procedure with $n_1 = 5$ is doing well, particularly when $c = .1$. When $c = .1$, the risk of the procedure with $n_1 = 5$ compares very favorably with the optimal risk except for very small values of σ^2. When $c = .01$, the procedure with $n_1 = 15$ does even better than $n_1 = 5$ and its risk compares extremely well with the optimal risk. These tables can serve as guidelines for recommendations. For costs lower than .01, n_1 can be increased. For costs near .1 or higher, $n_1 = 5$ is very reasonable. These guidelines can perhaps be used as suggestions for how to choose the first sample size in Stein's double sample fixed width confidence procedure. If $2d$ is the fixed width then c relates to d by $d^2/t^2(\alpha)$ where $t^2(\alpha)$ is a two tailed critical t-value. For example if $c = .1$, $\alpha = .05$, $t^2 \cong 8$ so $d \cong .9$ which means that if $2d$ is near 2 then $n_1 = 5$ could be recommended.

TABLE 4.1

Optimal values of n_1

a	n_1	a	n_1
4	5	34	13
11	7	42	15
17	9	50	17
25	11	60	19

TABLE 4.2

Risks for double sample procedures $c = .01$

Var	$n_1 = 2$ 2-Stage Risk	$n_1 = 5$ 2-Stage Risk	$n_1 = 15$ 2-Stage Risk	Optimal Risk
0.01	0.0125	0.0181	0.0277	0.0100
0.11	0.0466	0.0356	0.0347	0.0340
0.21	0.0652	0.0436	0.0409	0.0404
0.31	0.0792	0.0483	0.0449	0.0443
0.41	0.0908	0.0517	0.0478	0.0471
0.51	0.1009	0.0542	0.0500	0.0493
0.61	0.1099	0.0568	0.0519	0.0511
0.71	0.1181	0.0580	0.0534	0.0526
0.81	0.1257	0.0595	0.0547	0.0539
0.91	0.1328	0.0608	0.0559	0.0551
1.01	0.1395	0.0620	0.0570	0.0562
1.11	0.1458	0.0630	0.0579	0.0571
1.21	0.1518	0.0640	0.0588	0.0580
1.31	0.1575	0.0648	0.0596	0.0588
1.41	0.1630	0.0656	0.0603	0.0595
1.51	0.1683	0.0664	0.0610	0.0602
1.61	0.1734	0.0671	0.0617	0.0608
1.71	0.1783	0.0677	0.0623	0.0614
1.81	0.1830	0.0683	0.0628	0.0620
1.91	0.1877	0.0689	0.0634	0.0625

Var	$n_1 = 2$ 2-Stage Risk	$n_1 = 5$ 2-Stage Risk	$n_1 = 15$ 2-Stage Risk	Optimal Risk
1.00	0.1388	0.0619	0.0569	0.0561
6.00	0.3187	0.0809	0.0749	0.0740
11.00	0.4224	0.0871	0.0810	0.0800
16.00	0.5031	0.0909	0.0847	0.0838
21.00	0.5715	0.0937	0.0874	0.0865
26.00	0.6318	0.0958	0.0896	0.0886
31.00	0.6864	0.0976	0.0913	0.0904
36.00	0.7367	0.0991	0.0928	0.0919
41.00	0.7834	0.1004	0.0941	0.0932
46.00	0.8273	0.1016	0.0953	0.0943
51.00	0.8689	0.1026	0.0963	0.0954
56.00	0.9084	0.1036	0.0972	0.0963
61.00	0.9461	0.1044	0.0981	0.0972
66.00	0.9823	0.1052	0.0989	0.0979
71.00	1.0171	0.1059	0.0996	0.0987
76.00	1.0507	0.1066	0.1003	0.0994
81.00	1.0831	0.1073	0.1009	0.1000
86.00	1.1146	0.1079	0.1015	0.1006
91.00	1.1451	0.1084	0.1021	0.1012
96.00	1.1748	0.1090	0.1026	0.1017

TABLE 4.3

Risks for double sample procedures $c = .1$

Var	$n_1 = 2$ 2-Stage Risk	$n_1 = 5$ 2-Stage Risk	$n_1 = 15$ 2-Stage Risk	Optimal Risk
0.01	0.0743	0.1629	0.2715	−0.1303*
0.11	0.1303	0.1830	0.2781	0.1095
0.21	0.1836	0.2044	0.2848	0.1742
0.31	0.2293	0.2271	0.2915	0.2131
0.41	0.2692	0.2491	0.2981	0.2411
0.51	0.3049	0.2695	0.3048	0.2629
0.61	0.3372	0.2880	0.3116	0.2808
0.71	0.3670	0.3047	0.3184	0.2960
0.81	0.3946	0.3197	0.3255	0.3092
0.91	0.4204	0.3334	0.3327	0.3208
1.01	0.4448	0.3459	0.3401	0.3313
1.11	0.4679	0.3574	0.3474	0.3407
1.21	0.4898	0.3679	0.3547	0.3493
1.31	0.5108	0.3777	0.3618	0.3573
1.41	0.5309	0.3867	0.3687	0.3646
1.51	0.5502	0.3952	0.3753	0.3715
1.61	0.5688	0.4032	0.3817	0.3779
1.71	0.5867	0.4106	0.3877	0.3839
1.81	0.6041	0.4176	0.3935	0.3896
1.91	0.6210	0.4243	0.3991	0.3950

Var	$n_1 = 2$ 2-Stage Risk	$n_1 = 5$ 2-Stage Risk	$n_1 = 15$ 2-Stage Risk	Optimal Risk
1.00	0.4424	0.3447	0.3393	0.3303
6.00	1.0904	0.5612	0.5169	0.5094
11.00	1.4515	0.6293	0.5783	0.5700
16.00	1.7286	0.6701	0.6161	0.6075
21.00	1.9610	0.6993	0.6434	0.6347
26.00	2.1648	0.7219	0.6649	0.6561
31.00	2.3483	0.7404	0.6826	0.6737
36.00	2.5164	0.7562	0.6976	0.6886
41.00	2.6724	0.7697	0.7106	0.7016
46.00	2.8185	0.7816	0.7222	0.7131
51.00	2.9563	0.7923	0.7325	0.7234
56.00	3.0871	0.8019	0.7419	0.7328
61.00	3.2119	0.8107	0.7505	0.7413
66.00	3.3313	0.8188	0.7584	0.7429
71.00	3.4460	0.8263	0.7657	0.7565
76.00	3.5566	0.8333	0.7725	0.7633
81.00	3.6633	0.8398	0.7789	0.7697
86.00	3.7667	0.8460	0.7849	0.7757
91.00	3.8669	0.8517	0.7905	0.7813
96.00	3.9643	0.8572	0.7959	0.7867

* The optimal value of n_1 is less than 1 and risk has ln n_1 terms.

REFERENCES

COHEN, A. and SACKROWITZ, H. B. (1984a). Decision theory results for vector risks with applications. *Statistics and Decisions*, to appear.

COHEN, A. and SACKROWITZ, H. B. (1984b). Results in double sample estimation for the binomial distribution. *Ann. Statist.* **12** 1109–1116.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281.

FERGUSON, T. S. (1967). *Mathematical Statistics*. Academic, New York.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

MILLER, I. and FREUND, J. E. (1977). *Probability and Statistics for Engineers*, Second Ed. Prentice-Hall, Englewood Cliffs, New Jersey.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Graduate School of Business Administration, Harvard University, Boston.

STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

WHITTLE, P. and LANE, R. O. D. (1967). A class of situations in which a sequential procedure is non-sequential. *Biometrika* **54** 220–234.

WOODROOFE, M. (1981). A.P.O. rules are asymptotically non-deficient for estimation with squared error loss. *Z. Wahrsch. verw. Gebiete* **58** 331–341.

WOODROOFE, M. (1983). Personal communication.

DEPARTMENT OF STATISTICS
HILL CENTER–BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903.