*The Annals of Statistics*
1977, Vol. 5, No. 2, 379–384

# SEQUENTIAL BAYES ESTIMATION OF THE DIFFERENCE BETWEEN MEANS

BY THOMAS A. KELLEY

*Los Alamos Scientific Laboratory*

It is desired to estimate the difference between the means of two independent normal distributions as accurately as possible and in a sequential manner when the total number of observations is fixed. The problem is posed in a Bayesian framework with conjugate prior distributions and squared error loss function. It is shown that the optimal sequential design depends on the ratio of the posterior variances of the two means. There exist constants (dependent on the prior parameters, the number of observations taken from each distribution, and the number of observations remaining) such that when the above-mentioned ratio exceeds this constant it is optimal to select the next observation from one distribution; otherwise it is optimal to select it from the other distribution.

**1. Introduction.** Suppose there are two experiments $\varepsilon_1$ and $\varepsilon_2$ which generate, independently of each other, sequences of i.i.d. random variables which are generically denoted by $X$ and $Y$ respectively. Let each $X$ have a normal distribution $N(\mu_1, P_1^{-1})$ where $P_1$ is the precision. Let each $Y$ have a normal distribution $N(\mu_2, P_2^{-1})$. Many different sequential schemes for estimating $\mu_1 - \mu_2$ have been proposed. If the observations are taken in pairs $(X_1, Y_1)$, $(X_2, Y_2)$, etc., then $Z_i = X_i - Y_i$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $P_1^{-1} + P_2^{-1}$. Thus all of the results on sequentially estimating the mean of a single normal distribution are applicable. For example, see Anscombe [1], Chow and Robbins [2], Geertsema [4], Ray [6], Robbins [7], Serfling and Wackerly [9], Simons [10], and Starr [12], [13]. Allowing for unequal sample sizes may increase the accuracy of the estimate or decrease the expected total sample size. Srivastava [11] and Robbins, Simons and Starr [8] have proposed a class of sequential rules incorporating both a sampling scheme and a stopping rule which are asymptotically optimal. The approach in this paper is to suppose that the total number of observations is fixed and to concentrate on the sampling scheme. From this point of view the problem closely resembles the two-armed bandit problem. Prior distributions and a loss function are assigned and the concomitant optimal sequential strategy is examined.

**2. The Bayesian model.** The vector of means and precisions $(\mu_1, P_1, \mu_2, P_2)$ is assigned a prior distribution so that $(\mu_1, P_1)$ and $(\mu_2, P_2)$ are priorly independent with normal-gamma distributions $NG(\eta_1, \tau_1, \alpha_1, \beta_1)$ and $NG(\eta_2, \tau_2, \alpha_2, \beta_2)$.

Received January 1975; revised May 1976.

*AMS* 1970 *subject classifications.* Primary 62L12; Secondary 93C40.

*Key words and phrases.* Sequential estimation, adaptive control process, dynamic programming algorithm, optimal strategy, myopic rule.

Properties of this conjugate family of distributions are well known (cf. DeGroot [3], Section 9.6).

The object is to estimate $\theta = \mu_1 - \mu_2$ with squared error loss, so that if $\hat{\theta}$ is an estimator of $\theta$, the loss is $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$. The estimator $\hat{\theta}$ which minimizes the expected loss (risk) with respect to a prior distribution is $E(\mu_1) - E(\mu_2)$ and the minimum risk itself is $\sigma^2(\mu_1) + \sigma^2(\mu_2)$ (cf. Zacks [15], pages 273–274). These are, respectively, the Bayes estimator and the Bayes risk.

Let $N$ be the fixed total number of trials and let $m$ and $n$ be the number of trials allocated to experiments $\varepsilon_1$ and $\varepsilon_2$ respectively. After all $N$ observations have been completed, the Bayes estimator of $\mu_1 - \mu_2$ is the difference in the means of the posterior marginal distributions of $\mu_1$ and $\mu_2$; the Bayes risk is the sum of the variances of these two distributions. Once this is established, the whole problem can be modeled as an adaptive control process.

**3. Adaptive control process model.** For a description of adaptive control processes (ACP) see Yakowitz [14]. In our situation, the state space may be thought of as the set of quadruples $(m, n, A_m, B_n)$ where $m \geqq 0$, $n \geqq 0$, $m + n \leqq N$, $A_m \geqq 0$, $B_n \geqq 0$. Here $A_m$ and $B_n$ are the variances of the posterior marginal distributions of $\mu_1$ and $\mu_2$ respectively and are functions of the sample means, sample variances, and prior parameters. From the properties of the normal-gamma distributions, it follows that

$$A_m = \frac{2\beta_1 + (m-1)S_X^2 + \tau_1 m(\tau_1 + m)^{-1}(\bar{X} - \eta_1)^2}{(\tau_1 + m)(m + 2\alpha_1 - 2)} \tag{3.1}$$

and

$$B_n = \frac{2\beta_2 + (n-1)S_Y^2 + \tau_2 n(\tau_2 + n)^{-1}(\bar{Y} - \eta_2)^2}{(\tau_2 + n)(n + 2\alpha_2 - 2)} . \tag{3.2}$$

In order that these quantities be finite and nonnegative for all $m$ and $n$, $\alpha_1$ and $\alpha_2$ must be larger than 1. The control set consists of two elements: choose $\varepsilon_1$ or choose $\varepsilon_2$. At time $t < N$ zero loss is incurred while at time $t = N$ a loss of $A_m + B_n$ is incurred. The loss function may then be written as the sum of these losses as $t$ ranges from 0 to $N$. The statistical law of motion for this ACP is embodied in the following lemma.

LEMMA 3.1. *Given $A_m$ and $B_n$, $U = a_m A_m / A_{m+1}$ and $V = b_n B_n / B_{n+1}$ have independent beta distributions $B(\alpha_1 + \frac{1}{2}m, \frac{1}{2})$ and $B(\alpha_2 + \frac{1}{2}n, \frac{1}{2})$ respectively where*

$$a_m = \frac{(\tau_1 + m)(m + 2\alpha_1 - 2)}{(\tau_1 + m + 1)(m + 2\alpha_1 - 1)} \tag{3.3}$$

*and*

$$b_n = \frac{(\tau_2 + n)(n + 2\alpha_2 - 2)}{(\tau_2 + n + 1)(n + 2\alpha_2 - 1)} . \tag{3.4}$$

PROOF. Tedious algebraic manipulation leads to the fact that

$$A_{m+1} = a_m(A_m + (m + 2\alpha_1 - 2)^{-1}(\tau_1 + m + 1)^{-1}(X_{m+1} - \eta_1')^2) \tag{3.5}$$

where

$$\eta_1' = (\tau_1\eta_1 + m\bar{X}_m)/(\tau_1 + m) . \tag{3.6}$$

Let $\mathscr{F}_m$ be the $\sigma$-field generated by $X_1, \cdots, X_m$. From the properties of the normal-gamma family of distributions, $X_{m+1}$ given $\mathscr{F}_m$ is distributed as $C^{\frac{1}{2}}W + \eta_1'$ where

$$C = (m + 2\alpha_1 - 2)(\tau_1 + m + 1)(m + 2\alpha_1)^{-1}A_m , \tag{3.7}$$

$\eta_1'$ is defined as in (3.6) and $W$ is a random variable having a Student's $t$-distribution with $2\alpha_1$ df. This means that, given $\mathscr{F}_m$, $A_{m+1}$ is distributed as $a_m A_m(1 + W^2(m + 2\alpha_1)^{-1})$. Thus $U$ has a beta distribution $B(\alpha_1 + \frac{1}{2}m, \frac{1}{2})$. An analogous argument can be applied to $B_{n+1}$ to get the distribution of $V$. Furthermore, $A_{m+1}$ and $B_{n+1}$ are independent, so $U$ and $V$ are independent. □

Because of the Markovian nature of the statistical law of motion and the separability of the loss function for this ACP, the dynamic programming algorithm yields the optimal strategy (cf. Yakowitz [14], Theorem 3.3).

**4. The optimal sequential design.** Suppose there are $k$ trials remaining where $m$ trials have already been allocated to $\varepsilon_1$ and $n$ trials to $\varepsilon_2$. Let $\mathscr{F}_{m,n}$ denote the sigma field generated by $(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$. Unless otherwise stated, all expectations in this section are conditional on $\mathscr{F}_{m,n}$. Also for notational convenience, the arguments $A_m$, $B_n$ will be suppressed from the functions in this section unless needed. Denote the anticipated risk over the remaining $k$ trials by $R_k(m, n)$. Then

$$R_1(m, n) = \min\{E(A_{m+1} + B_n), E(A_m + B_{n+1})\} . \tag{4.1}$$

From Lemma 3.1 it follows that

$$R_1(m, n) = A_m + B_n - \max\{A_m(\tau_1 + m + 1)^{-1}, B_n(\tau_2 + n + 1)^{-1}\} . \tag{4.2}$$

In general there exist functions $F_k(m, n)$ and $G_k(m, n)$ such that

$$R_k(m, n) = A_m + B_n - \max\{F_k(m, n), G_k(m, n)\} . \tag{4.3}$$

These functions may be defined recursively as follows.

$$F_0(m, n) = G_0(m, n) = 0 \tag{4.4}$$

and for $k \geqq 0$

$$F_{k+1}(m, n) = A_m(\tau_1 + m + 1)^{-1} + E(\max\{F_k(m + 1, n), G_k(m + 1, n)\}) \tag{4.5}$$

$$G_{k+1}(m, n) = B_n(\tau_2 + n + 1)^{-1} + E(\max\{F_k(m, n + 1), G_k(m, n + 1)\}) . \tag{4.6}$$

Once these functions have been determined, the optimal policy is to select experiment $\varepsilon_1$ provided $F_k(m, n) \geqq G_k(m, n)$ and experiment $\varepsilon_2$ otherwise. In order to gain more information about the optimal policy let $D_k(m, n) = F_k(m, n) - G_k(m, n)$, the relative advantage of $\varepsilon_1$ over $\varepsilon_2$. These functions may be defined recursively in the following manner.

$$D_1(m, n) = A_m(\tau_1 + m + 1)^{-1} + B_n(\tau_2 + n + 1)^{-1} . \tag{4.7}$$

Let $H_k(m, n) = E(\max\{F_k(m, n), G_k(m, n)\})$, then from (4.5) and (4.6) it follows that

$$D_{k+1}(m, n) = D_1(m, n) + H_k(m+1, n) - H_k(m, n+1)\,. \tag{4.8}$$

Since $\max\{F, G\} = G + (F-G)^+$ where $x^+$ denotes $\max\{x, 0\}$,

$$H_k(m+1, n) = E(G_k(m+1, n)) + E(D_k^+(m+1, n))\,. \tag{4.9}$$

Since $\max\{F, G\} = F + (G-F)^+ = F - (F-G)^-$ where $x^-$ denotes $\min\{x, 0\}$,

$$H_k(m, n+1) = E(F_k(m, n+1)) - E(D_k^-(m, n+1))\,. \tag{4.10}$$

Now using (4.5) and (4.6) again we find that

$$E(G_k(m+1, n)) = B_n(\tau_2 + n + 1)^{-1} + H_k(m+1, n+1)\,, \tag{4.11}$$

and also that

$$E(F_k(m, n+1)) = A_m(\tau_1 + m + 1)^{-1} + H_k(m+1, n+1)\,. \tag{4.12}$$

Thus $E(G_k(m+1, n) - F_k(m, n+1)) = -D_1(m, n)$ independent of $k$, so that for $k \geqq 1$

$$D_{k+1}(m, n) = E(D_k^+(m+1, n)) + E(D_k^-(m, n+1))\,. \tag{4.13}$$

Similar recursive relations have been derived by the author [5] for the Bernoulli two-armed bandit problem. These relations are now used to prove the first result about the optimal procedure.

THEOREM 4.1. *For any positive constant c,*

$$D_k(m, n, cA_m, cB_n) = cD_k(m, n, A_m, B_n)\,.$$

PROOF. By not suppressing the arguments $A_m$, $B_n$ the proof is by induction on $k$ and follows directly from (4.13). □

Since the optimal strategy depends on the sign of $D_k(m, n)$, the decision rule is a function of $(A_m, B_n)$ only through the ratio $A_m/B_n$. This means that a new function $\Delta_k(m, n, r)$ can be defined as follows.

$$\Delta_k(m, n, A_m/B_n) = B_n D_k(m, n, A_m, B_n)\,. \tag{4.14}$$

LEMMA 4.1. *The functions $\Delta_k(m, n, r)$ are defined recursively through the following equations.*

$$\Delta_1(m, n, r) = r(\tau_1 + m + 1)^{-1} - (\tau_2 + n + 1)^{-1} \tag{4.15}$$

*and for* $k = 1, \cdots, N-1$,

$$\Delta_{k+1}(m, n, r) = E(\Delta_k^+(m+1, n, a_m U^{-1} r)) + E(b_n V^{-1} \Delta_k^-(m, n+1, b_n^{-1} V r)) \tag{4.16}$$

*where U and V have independent beta distributions* $B(\alpha_1 + \frac{1}{2}m, \frac{1}{2})$ *and* $B(\alpha_2 + \frac{1}{2}n, \frac{1}{2})$ *respectively and* $a_m$ *and* $b_n$ *are given in* (3.3) *and* (3.4).

PROOF. This follows directly from (4.7), (4.13), and Lemma 3.1. □

LEMMA 4.2. *For the functions* $\Delta_k(m, n, r)$ *defined in Lemma* 4.1, *there exist two sequences* $Q_k(m)$ *and* $P_k(n)$ *such that* $Q_k(m) > 0$ *and* $P_k(n) < 0$ *and for all* $r \geqq 0$, $P_k(n) < \Delta_k(m, n, r) < Q_k(m)r$.

PROOF. The proof is by induction on $k$ and uses Lemma 4.1. □

The main use of this lemma is in the proof of the following theorem giving a characterization of the optimal sequential decision rule.

THEOREM 4.2. *For the functions* $\Delta_k(m, n, r)$ *defined in Lemma* 4.1, *there exists a unique sequence of constants* $\gamma_k(m, n)$ *such that* $\Delta_k(m, n, r) \geqq 0$ *if and only if* $r \geqq \gamma_k(m, n)$.

PROOF. Using Lemma 4.2, the following properties of $\Delta_k(m, n, r)$ may be established by induction:

(i) $\Delta_k(m, n, r)$ is a strictly increasing function of $r$,
(ii) $\Delta_k(m, n, r)$ is a continuous function of $r$,
(iii) $\Delta_k(m, n, r)$ is negative for sufficiently small values of $r$, and
(iv) $\Delta_k(m, n, r)$ is positive and arbitrarily large for sufficiently large values of $r$.

Then the existence of $\gamma_k(m, n)$ is guaranteed by the intermediate value theorem and the uniqueness of $\gamma_k(m, n)$ is guaranteed by property (i). □

The optimal sequential procedure is then of the following form. If $m$ trials have been allocated to $\varepsilon_1$, $n$ trials have been allocated to $\varepsilon_2$, and there are $k$ trials remaining, then on the basis of the prior parameters and the posterior variances $A_m$, $B_n$ it is optimal to allocate the next trial to $\varepsilon_1$ if and only if $A_m/B_n$ exceeds $\gamma_k(m, n)$.

**5. The myopic decision rule.** Unfortunately, little is known about the constants $\gamma_k(m, n)$. In approximating the optimal decision rule, one needs only to choose a set of constants $g_k(m, n)$ "close" to $\gamma_k(m, n)$. An appealing choice is

$$g_k(m, n) = \gamma_1(m, n) = (\tau_1 + m + 1)(\tau_2 + n + 1)^{-1}, \tag{5.1}$$

and the resulting decision rule, which is independent of $k$, is called the myopic rule. It "acts" as if there were always but one more trial remaining.

This decision rule also appears when one obtains the limiting form of the optimal strategy as more and more information is assumed to be available concerning the precisions $P_1$ and $P_2$. Consider prior distributions in the normal-gamma family where the marginals of $P_i$ ($i = 1, 2$) are gamma distributions with $\beta_i = \sigma_i^2\alpha_i$. Suppose $\sigma_i^2$ is a constant and $\alpha_i$ is allowed to increase to $+\infty$. These marginals approach singular distributions with the entire probability mass concentrated at $\sigma_i^{-2}$. The prior marginal distributions of $\mu_i$ approach normal distributions with mean $\eta_i$ and precision $\tau_i\sigma_i^{-2}$. Note also that as $\alpha_i \to +\infty$, $a_m \to (\tau_1 + m)(\tau_1 + m + 1)^{-1}$ and $b_n \to (\tau_2 + n)(\tau_2 + n + 1)^{-1}$. The random variables $A_m$, $B_n$, $U$ and $V$ all approach singular distributions with probability mass concentrated at $\sigma_1^2(\tau_1 + m)^{-1}$, $\sigma_2^2(\tau_2 + n)$, 1 and 1 respectively. The limiting form of $\Delta_k(m, n, r)$, from the Helly–Bray theorem, satisfies the recursive equations.

$$\Delta_1(m, n, r) = r(\tau_1 + m + 1)^{-1} - (\tau_2 + n + 1)^{-1}, \tag{5.2}$$

and for $k = 1, \cdots, N - 1$

$$\Delta_{k+1}(m, n, r) = \Delta_k^{+}(m + 1, n, a_m r) + b_n \Delta_k^{-}(m, n + 1, b_n^{-1} r). \tag{5.3}$$

THEOREM 5.1. *For the functions* $\Delta_k(m, n, r)$ *defined in* (5.2) *and* (5.3), $r > \gamma_1(m, n) \Rightarrow \Delta_k(m, n, r) \geqq 0$.

PROOF. The proof is by induction on $k$. □

So the limiting optimal decision rule is to allocate the next trial to $\varepsilon_1$ provided

$$\sigma_1^2 \sigma_2^{-2} \geqq (\tau_1 + m)(\tau_1 + m + 1)(\tau_2 + n)^{-1}(\tau_2 + n + 1)^{-1}. \tag{5.4}$$

Since this decision rule does not depend on the outcomes of the various trials, the sampling may be done in one stage, choosing as the total number of trials to be allocated to $\varepsilon_1$ the largest value of $m$ such that (5.9) is satisfied.

## REFERENCES


[1] ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

[2] CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

[3] DEGROOT, M. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

[4] GEERTSEMA, J. C. (1970). Sequential confidence intervals based on rank tests. *Ann. Math. Statist.* **41** 1016–1026.

[5] KELLEY, T. A. (1974). A note on the Bernoulli two-armed bandit problem. *Ann. Statist.* **2** 1056–1062.

[6] RAY, W. D. (1957). Sequential confidence intervals for the mean of a normal distribution with unknown variance. *J. Roy. Statist. Soc. Ser. B* **19** 133–143.

[7] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics* (Harald Cramér Volume) 235–245. Almquist and Wiksell, Uppsala.

[8] ROBBINS, H., SIMONS, G. and STARR, N. (1967). A sequential analogue of the Behrens–Fisher problem. *Ann. Math. Statist.* **38** 1384–1391.

[9] SERFLING, R. J. and WACKERLY, D. D. (1974). Asymptotic theory of sequential fixed width confidence intervals for location parameters. Technical report M316, Florida State Univ.

[10] SIMONS, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.* **39** 1946–1952.

[11] SRIVASTAVA, M. S. (1970). On a sequential analogue of the Behrens–Fisher problem. *J. Roy. Statist. Soc. Ser. B* **32** 144–148.

[12] STARR, N. (1966). The performance of a sequential procedure for the fixed-width estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

[13] STARR, N. (1966). On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* **37** 1173–1185.

[14] YAKOWITZ, S. J. (1969). *Mathematics of Adaptive Control Processes.* American Elsevier, New York.

[15] ZACKS, S. (1973). *The Theory of Statistical Inference.* Wiley, New York.



LOS ALAMOS SCIENTIFIC LABORATORY
MS 514 P.O. BOX 1663
LOS ALAMOS, NEW MEXICO 87545