ON THE THEORY OF BAN ESTIMATES[1]

By Robert A. Wijsman[2]

University of California, Berkeley

1. Introduction and summary. The notion of best asymptotically normal estimates—BAN estimates for short[3]—was introduced by Neyman [8] in the multinomial case. Applications have been made in biological problems, notably in bio-assay [2], [4], [5]. Generalizations of Neyman's work have been made by Barankin and Gurland [1], Chiang [3] and Ferguson [5]. The usual theory of BAN estimates requires differentiability of the estimates, and imposes rather strong conditions on certain functions given in advance (the functions ζ and Σ of Section 3). In this note a different definition of BAN estimates is made which does not require differentiability, at the same time relaxing the conditions on ζ and Σ, whereas in essence all important theorems in the theory of BAN estimates are retained.

2. Notation. Convergence in probability is denoted by $\xrightarrow{P}$. $X_n \sim Y_n$ means $X_n - Y_n \xrightarrow{P} 0$. $X_n \xrightarrow{\mathfrak{L}} \mathfrak{N}(0, \Sigma)$ means that the law of X_n tends to a multivariate normal law with mean 0 and covariance matrix Σ. A' is the transpose of a matrix A, I_m the $m \times m$ identity matrix. We shall write *regular* (1), *regular* (2), BAN (1), BAN (2), depending on whether Definition 1 or Definition 2 of Section 3 is used. For notation and terminology not explained here see Chiang's paper [3], with which the notation in this note is in fair agreement.

3. The usual definition and a new definition of BAN estimates. Let Z_n be a sequence of random vectors, taking values in a space Z which is a subspace of a k-dimensional Euclidean space R^k. The distribution of the Z_n depends on a parameter θ which takes values in an open subset Ω of an m-dimensional Euclidean space, where $m \leqq k$. The true value of θ will be denoted by θ_0. It is assumed that

$$(1) \qquad \sqrt{n}\,(Z_n - \zeta(\theta_0)) \xrightarrow{\mathfrak{L}} \mathfrak{N}(0, \Sigma(\theta_0))$$

in which ζ and Σ are functions on Ω, ζ into Z and Σ into the space of $k \times k$ positive semi-definite matrices. Let the set $\zeta(\Omega)$ be denoted by U. In the simplest theory of BAN estimates, an admissible estimate is a function $\hat{\theta}$ from Z to Ω, and if Z_n is observed then θ is estimated by $\hat{\theta}(Z_n)$. We shall occasionally write

Received November 11, 1957; revised November 7, 1958.

[1] This investigation was supported (in part) by a research grant (No. G-3666) from the National Institutes of Health, Public Health Service.

[2] Now at the University of Illinois.

[3] Some authors prefer "RBAN," where "R" stands for *regular*. In this note, regularity of a BAN estimate is part of the definition.

$\hat{\theta}_n$ instead of $\hat{\theta}(Z_n)$. In the usual theory of BAN estimates the following assumptions and definitions are made (our Definition 1 differs slightly from the one given in [3] but leads to the same definition of BAN (1)):

ASSUMPTION 1. (i) ζ is 1-1 and bicontinuous; (ii) $\Sigma(\theta)$ is nonsingular for every θ; (iii) ζ and Σ have continuous second derivatives; (iv) the matrix $\partial\zeta/\partial\theta$ is of rank m for every θ.

DEFINITION 1. $\hat{\theta}$ is called regular (1) if (i) $\hat{\theta}(Z_n) \xrightarrow{P} \theta_0$ whatever θ_0, that is, $\hat{\theta}$ is consistent; (ii) $\partial\hat{\theta}/\partial z$ exists and is continuous in a neighborhood of U.

We shall denote $\zeta(\theta_0) = \zeta_0$, $\Sigma(\theta_0) = \Sigma_0$. The $k \times m$ matrix derivative $\partial\zeta/\partial\theta$ will be denoted by $V(\theta)$, the $m \times k$ matrix derivative $\partial\hat{\theta}/\partial z$ by $A(z)$. $V(\theta_0)$ and $A(\zeta_0)$ will be written V_0, A_0 respectively. An immediate consequence of Assumption 1, Definition 1 and (1) is that

$$\hat{\theta}(\zeta(\theta)) \equiv \theta \text{ identically in } \theta. \tag{2}$$

This enables us to show that $\hat{\theta}_n$ is asymptotically normal, with asymptotic covariance matrix $A_0 \Sigma_0 A_0'$. From (2) one obtains by differentiation

$$A(\zeta(\theta))V(\theta) \equiv I_m \text{ identically in } \theta, \tag{3}$$

which is a very important relation since it has as a consequence that among all matrices of the form $A_0\Sigma_0A_0'$ there is a minimal one. A regular (1) estimate with minimal covariance matrix is called BAN (1).

A priori there is no reason why an estimate should have a derivative, apart from the convenience with which one can show the existence of a minimum covariance matrix among the covariance matrices of regular (1) estimates. The differentiability conditions in Assumption 1(iii) also seem stronger than they need be. The reason for assuming so much differentiability is to ensure the differentiability of $\hat{\theta}$ if the latter is generated by minimizing a quadratic form [3] or by the root of a linear form [5]. Chiang [3] in his theorem 5 makes the full Assumption 1(iii). In his theorem 6 he only assumes that ζ has continuous second derivatives, and makes no assumptions on Σ. However, the type of estimates allowed there goes out beyond the simplest theory of BAN estimates as treated in the present paper, since $\hat{\theta}$ is allowed to be a function of both Z_n and S_n, where S_n is a consistent estimate of Σ. If S_n is taken as a function of Z_n, which is usually the case in applications, it has to be a differentiable function in order that $\hat{\theta}$ is differentiable. Then, since $Z_n \xrightarrow{P} \zeta_0$, we must have $S_n(\zeta_0) = \Sigma_0$, whatever be θ_0; that is, $S_n(\zeta(\theta)) = \Sigma(\theta)$, so that Σ has continuous first derivatives. If a regular (1) estimate is to be obtained as a root of a linear form (equation (5) in the present paper) then the matrix B in this linear form has to have continuous first partial derivatives with respect to z and θ. If, in addition, this regular (1) estimate is to have minimum covariance matrix, then the matrix B has to satisfy $B(\zeta(\theta), \theta) = V'(\theta)\Sigma^{-1}(\theta)$ for all θ (the more general condition given in [5] reduces to the one given here if Σ is non-singular). It follows then that $V'(\theta)\Sigma^{-1}(\theta)$ has to have continuous first derivatives with respect to θ, and the most natural way to achieve this is to have both V and Σ continuously differentiable. Thus, we see that in order to generate BAN (1) estimates it is practically necessary to require that Σ is continuously differentiable and ζ continuously twice differentiable.

It should be said at once that Ferguson in [5] does not require the estimates to be regular (1), which allows at the same time for relaxation of the conditions on Σ and ζ. The class of estimates considered in [5] consists of all those estimates which can be obtained as a suitably chosen root of a linear form (our equation (5)), for the various choices of the matrix B. The restrictions on B—continuity in z and differentiability in θ—ensure the continuity of $\hat{\theta}$ in a neighborhood of U. The corresponding class of covariance matrices contains the minimum member $(V_0'\Sigma_0^{-1}V_0)^{-1}$, which can be attained if Σ and V are continuous in θ (our Assumption 2(iii)). This minimal covariance matrix has the same form as in the case of regular (1) estimates, so that Ferguson's BAN estimates have the same asymptotic properties as BAN (1) estimates whenever they exist.

It may be argued that from a theoretical point of view it is slightly unsatisfactory to define a class of estimates by the manner in which they are generated, rather than by their common properties. The approach in this paper is different from Ferguson's in that a class of estimates will be defined by the properties they are to have, in the same way as this is done in the case of regular (1) estimates, but with a weaker definition of regular. The new definition of regular (Definition 2, below) will henceforth be referred to as *regular* (2). In spite of this weakening, all important theorems remain valid, in particular the existence of the minimum covariance matrix $(V_0'\Sigma_0^{-1}V_0)^{-1}$. The Assumptions 1(iii) on Σ and ζ will also be relaxed. The only differentiability condition retained is that $V = \partial\zeta/\partial\theta$ exists and is continuous. This condition is certainly a very natural one, since the matrix V plays an important role in the theory of BAN estimates. Thus, compared to the theory of BAN (1) estimates, the field of applicability is enlarged, and at the same time the theory becomes somewhat neater since the conditions are better tuned to the essentials in the theory. The class of regular (2) estimates contains both the regular (1) and Ferguson's estimates. Furthermore, the asymptotic properties of BAN (2) estimates are the same as those of BAN (1) estimates whenever the latter exist. Thus, the class of estimates to be considered is enlarged, and if a BAN (1) estimate exists, it still belongs to the best estimates in this class.

Assumption 2. (i), (ii) and (iv) are the same as assumptions 1(i), 1(ii) and 1(iv); (iii) Σ is continuous and ζ has a continuous first derivative V.

Definition 2. $\hat{\theta}$ will be called regular (2) if (i) $\hat{\theta}$ is continuous in each point of U; (ii) for each θ there exists an $m \times k$ matrix $A(\theta)$, continuous in θ, such that

$$\sqrt{n}(\hat{\theta}(Z_n) - \theta_0) \sim A(\theta_0)\sqrt{n}(Z_n - \zeta_0) \text{ whatever } \theta_0 . \tag{4}$$

Definition 2 is weaker than Definition 1, in that the former is implied by the latter, but not vice versa. In fact, Definition 2 does not even guarantee the continuity of $\hat{\theta}$ in a neighborhood of U. This causes a little trouble, since $\hat{\theta}$ is then not necessarily measurable. For the reinterpretation of various probability statements the reader is referred to a remark by LeCam ([6], p. 132). With Definition 2 it is easy to show the validity of (2). The following theorem, however, is not immediate, and will be proved in section 4.

THEOREM 1. *$A(\theta)V(\theta) \equiv I_m$ identically in θ.*

Theorem 1 takes the place of (3), and $A(\theta_0)$ takes the place of what was previously called A_0. Thus, it is seen that the classes of covariance matrices of regular (1) and regular (2) estimates coincide. A regular (2) estimate with minimal covariance matrix will be called BAN (2), and, with a few modifications, the whole theory of BAN estimates is unchanged.

The next theorem is essentially due to Ferguson [5], except for the weaker conditions, and shows how to generate a BAN (2) estimate as a root of a linear form.

THEOREM 2. *Let $B(z, \theta)$ be an $m \times k$ matrix, continuous in θ for each z and continuous in (z, θ) at each point $(\zeta(\theta),\theta)$, such that B_0V_0 is nonsingular whatever θ_0, where $B_0 = B(\zeta_0, \theta_0)$. Then there exists a neighborhood N of U and a function $\hat{\theta}$ on N to Ω such that on N, $\hat{\theta}(z)$ satisfies the equation*

$$B(z, \theta)(z - \zeta(\theta)) = 0 \tag{5}$$

for θ and such that $\hat{\theta}$ is a regular (2) estimate. Furthermore, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim (B_0V_0)^{-1}B_0\sqrt{n}(Z_n - \zeta_0). \tag{6}$$

Lastly, if $B_0 = V_0'\Sigma_0^{-1}$ then $\hat{\theta}$ is BAN.

The proof of Theorem 2 proceeds along conventional lines and will not be reproduced here. It relies on Brouwer's fixed point theorem [7], in which the transformation is assumed to be continuous but not necessarily differentiable, and, as a consequence, uniqueness of the fixed point cannot be concluded. As a result, the estimate $\hat{\theta}$ is not necessarily unique. The matrix B may be chosen to be $V'(\theta)\Sigma^{-1}(\theta)$ and will then satisfy the conditions of Theorem 2, by assumption 2(iii). Moreover, this choice for B will generate a BAN (2) estimate, by the last part of Theorem 2. The same conclusions hold if $V(\theta)$ is replaced by a matrix $V^*(z)$, depending only on z, such that V^* is continuous in each point of U, and $V^*(\zeta(\theta)) = V(\theta)$ for all θ. $\Sigma(\theta)$ may be replaced similarly.

4. Proof of Theorem 1. We shall need the following lemma, whose proof is due to Dr. Lucien M. LeCam.

LEMMA 1. *Let (a, b) be a one-dimensional interval in R^k, and suppose that for each $x \in (a, b)$ there is a sphere $S(x)$ about x with radius $r(x)$. Then there are two distinct points, x_1 and x_2, in (a, b) such that $x_1 \in S(x_2)$ and $x_2 \in S(x_1)$.*

PROOF. Let $x_0 \in (a, b)$ be a point such that there is no $x \in (a, b) \cap S(x_0)$ for which $x_0 \in S(x)$. Then $r(x) \to 0$ as $x \to x_0$. On the other hand, $r(x_0) > 0$. It follows that r has a discontinuity of the first kind at x_0. The conclusion follows from the fact that the discontinuities of the first kind are denumerable.

We proceed to prove Theorem 1. Suppose that for some $\theta_1 \in \Omega$, $A(\theta_1)V(\theta_1) \neq I_m$.

Without loss of generality we may assume that $\zeta(\theta_1)$ is the origin 0 of the coordinate system in $\mathcal{Z}$. Let T be the tangent plane to U in 0, and consider in the following T fixed, i.e. not subject to transformations. There is a transformation $z \to z^* = g(z)$ which in a $\mathcal{Z}$-neighborhood of 0 is continuously differentiable in both directions, with $\partial g(0)/\partial z = I_k$, and which maps U into T. More precisely, there is a U-neighborhood N_u of 0, a T-neighborhood N_t of 0, and a $\mathcal{Z}$-neighborhood N_z of N_u, such that on N_z the transformation g has the differentiability properties mentioned in the preceding sentence, and such that $g(N_u) = N_t$. Let $N_\theta = \zeta^{-1}(N_u)$. In (4) we shall only consider values of θ_0 which are in N_θ. The estimate $\hat{\theta}$ is now to be considered as a function of z^*. If we write (4) down for the transformed variables, i.e. replacing Z_n by Z_n^*, ζ_0 by ζ_0^*, we have to replace $A(\theta_0)$ by $A^*(\theta_0) = (\partial g(\zeta_0)/\partial z)^{-1}A(\theta_0)$. $A^*(\theta)$, like $A(\theta)$, is continuous in θ. Since $\partial g(\zeta(\theta_1))/\partial z = I_k$, we have $A^*(\theta_1) = A(\theta_1)$. Furthermore, for the transformed variables we have $V^*(\theta) = \partial\zeta^*/\partial\theta$, where $\zeta^*(\theta) = g(\zeta(\theta))$, so that $V^* = (\partial g/\partial z)V$ (the arguments have been suppressed). At θ_1 we have $V^*(\theta_1) = V(\theta_1)$. Thus, if $A(\theta_1)V(\theta_1) \neq I_m$, then also $A^*(\theta_1)V^*(\theta_1) \neq I_m$. Dropping the asterisks, we consider a new problem, in which the new Ω is the old N_θ, the new U is the old N_t. Hence U is a subset of an m-dimensional subspace of R^k. For some $\theta_1 \in \Omega$, $A(\theta_1)V(\theta_1) \neq I_m$. We may further simplify the problem by making a suitable transformation $\theta \to \theta^*$, continuously differentiable in both directions. This transforms $V \to V^* = V(\partial\theta/\partial\theta^*)$ and $A \to A^* = (\partial\theta^*/\partial\theta)A$ (the arguments have been suppressed). Thus, if $AV \neq I_m$, then also $A^*V^* \neq I_m$. We choose the transformation $\theta \to \theta^* = \zeta(\theta)$. Dropping the asterisks, in the new problem Ω and U are identical, ζ is the identity function on U, and $\hat{\theta}$ is a function on $\mathcal{Z}$ to U which is the identity function on U, by (2). For each $u \in U$, $A(u)$ is a linear transformation of $\mathcal{Z}$ into the m-dimensional subspace in which U is embedded. If Theorem 1 is true, then, for each u, $A(u)$ is the identity transformation on U. We have assumed that $A(0)$ is not the identity transformation on U, and will show that this leads to a contradiction.

If $A(0)$, restricted to U, is not the identity transformation, then the same is true on at least one of the coordinate axes in U. We choose one of these coordinate axes, and call it the x-axis in the following. For simplicity, x, with or without subscripts, will stand both for a point on the x-axis and for its x-coordinate. In the following, $|z|$ will denote the norm of a vector z, $\|A\|$ the norm of a matrix A, I the identity transformation on U. Using the continuity of A, there is on the x-axis an interval $N_a = (-a, a)$ and there is an $\epsilon > 0$ such that for all $\xi, x_1, x_2 \in N_a$ we have

$$(7) \qquad |(A(\xi) - I)(x_2 - x_1)| \geqq 6\epsilon\, |x_2 - x_1| .$$

Furthermore, we can choose a so small that for all $x_1, x_2 \in N_a$ we have

$$(8) \qquad \|A(x_2) - A(x_1)\| < \epsilon.$$

We now put

$$(9) \qquad f(z, u) = \hat{\theta}(z) - u - A(u)(z - u),$$

then from (4) it follows that

$$P\{\sqrt{n}\,|f(Z_n, u_0)| \geqq \epsilon\} \to 0, \tag{10}$$

in which u_0 is the true value of the parameter. Denote by $S(u, n)$ the open sphere with radius $n^{-1/2}$ about u. If $u = 0$ we write simply $S(n)$. Using (10) we have

$$P\{Z_n \varepsilon S(u_0, n), \sqrt{n}\,|f(Z_n, u_0)| \geqq \epsilon\} \to 0. \tag{11}$$

Making the transformation $y = n^{1/2}(z - u_0)$, $Y_n = n^{1/2}(Z_n - u_0)$, $g_n(y) = n^{1/2}f(z, u_0)$, we can write (11) as

$$P\{Y_n \varepsilon S(1), |g_n(Y_n)| \geqq \epsilon\} \to 0. \tag{12}$$

Let μ be the Lebesgue measure in R^k. Since Y_n has a limiting density with respect to μ, we conclude from (12)

$$\mu\{y \varepsilon S(1), |g_n(y)| \geqq \epsilon\} \to 0. \tag{13}$$

If we divide the left-hand side of (13) by the constant $\mu S(1) = \mu\{y \varepsilon S(1)\}$, and make the transformation back from y to z, we obtain

$$[\mu S(n)]^{-1}\mu\{z \varepsilon S(u, n), |f(z, u)| \geqq \epsilon n^{-1/2}\} \to 0 \tag{14}$$

in which we have dropped the subscript 0 on u.

For any z, let $z = x_z + y_z$, where x_z is the x-component of z. We choose a number $\alpha > 0$ in such a way that for any x_1, x_2, n, with

$$\tfrac{1}{2}n^{-1/2} < x_2 - x_1 < n^{-1/2} \tag{15}$$

we have

$$\mu\{z \varepsilon S(x_1, n) \cap S(x_2, n), x_1 < x_z < x_2\} > \alpha\mu S(n). \tag{16}$$

The number α can obviously be chosen independently of the particular choice of x_1, x_2, n, so long as (15) holds. Restricting now u in (14) to points x in N_a, we have for each $x \varepsilon N_a$ an integer n_x such that

$$\mu\{z \varepsilon S(x, n), |f(z, x)| \geqq \epsilon n^{-1/2}\} < \tfrac{1}{2}\alpha\mu S(n) \tag{17}$$

provided $n \geqq n_x$. According to Lemma 1, there are two points x_1, $x_2 \varepsilon N_a$, with $x_1 < x_2$, such that $x_1 \varepsilon S(x_2, n_{x_2})$ and $x_2 \varepsilon S(x_1, n_{x_1})$. We can now choose an integer $n \geqq \max(n_{x_1}, n_{x_2})$ such that (15) holds, and therefore also (16). The two equations (16) and (17) together imply that

$$\begin{aligned}\mu\{z \varepsilon S(x_1, n) \cap S(x_2, n), x_1 < x_z < x_2, \\ |f(z, x_1)| < \epsilon n^{-1/2}, |f(z, x_2)| < \epsilon n^{-1/2}\} > 0,\end{aligned} \tag{18}$$

so that the set in braces in (18) is not empty. Choosing any point z in this set, and using (15), we have the following inequalities:

$$x_1 < x_z < x_2, \tag{19}$$

$$|y_z| < 2(x_2 - x_1), \tag{20}$$

$$|f(z, x_1)| < 2\epsilon(x_2 - x_1), |f(z, x_2)| < 2\epsilon(x_2 - x_1). \tag{21}$$

From (9) we compute

$$\begin{aligned} f(z, x_2) - f(z, x_1) = (A(x_1) - I)(x_z - x_1) \\ + (A(x_2) - I)(x_2 - x_z) + (A(x_1) - A(x_2))y_z. \end{aligned} \tag{22}$$

By continuity of A and by virtue of (19) there is a point $\xi \varepsilon (x_1, x_2)$ such that

$$|(A(\xi) - I)(x_2 - x_1)| = |(A(x_1) - I)(x_z - x_1) + (A(x_2) - I)(x_2 - x_z)|. \tag{23}$$

Using (8), (20), (21), (22) and (23), we have finally

$$|(A(\xi) - I)(x_2 - x_1)| < 6\epsilon(x_2 - x_1), \tag{24}$$

contradicting (7). Q.E.D.

Acknowledgment. The author is indebted to Dr. Lucien M. LeCam for suggesting the proof of Lemma 1, and to the referee for helpful comments.

REFERENCES

[1] E. W. BARANKIN AND J. GURLAND, "On asymptotically normal, efficient estimators I," *University of California Publications in Statistics*, Vol. 1 (1950), pp. 89–130.

[2] J. BERKSON, "Application of the logistic function to bio-assay," *J. Am. Stat. Assn.*, Vol. 39 (1944), pp. 357–365.

[3] C. L. CHIANG, "On regular best asymptotically normal estimates," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 336–351.

[4] C. L. CHIANG, "An application of stochastic processes to experimental studies on flour beetles," *Biometrics*, Vol. 13 (1957), pp. 79–97.

[5] T. S. FERGUSON, "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities," Dissertation, University of California, Berkeley, 1956, *Ann, Math. Stat.*, Vol. 29 (1958), pp. 1046–1062.

[6] L. M. LECAM, "On the asymptotic theory of estimation and testing hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956, pp. 129–156.

[7] S. LEFSCHETZ, *Introduction to Topology*, Princeton University Press, 1949.

[8] J. NEYMAN, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 239–273.