A METHOD OF GENERATING BEST ASYMPTOTICALLY NORMAL ESTIMATES WITH APPLICATION TO THE ESTIMATION OF BACTERIAL DENSITIES[1]

By Thomas S. Ferguson[2]

University of California, Berkeley

0. Summary. Various minimum χ^2 methods used for generating B.A.N. estimates are summarized, and a new method which generates B.A.N. estimates as roots of certain linear forms is introduced and investigated. As a particular application of the method, the estimation of the bacterial density in an experiment using dilution series is considered.

1. Introduction. The purpose of the present paper is to describe a simple method by which estimates having the usual asymptotic properties of Best Asymptotically Normal (B.A.N.) estimates can be obtained.

Originally B.A.N. estimates were introduced by J. Neyman [1] for a situation in which the underlying probability distributions have a multinomial-like structure. This was followed by a paper by E. W. Barankin and J. Gurland [2] who extended the class of estimation problems for which B.A.N. estimates could be used and also described quite general methods of generation of such estimates. Other results in this direction have been obtained by C. L. Chiang [3] and L. Le Cam [4] and W. Taylor [5].

A best asymptotically normal estimate θ^* of a parameter θ is, loosely speaking, one which is asymptotically normally distributed about the true parameter value, and which is best in the sense that out of all such asymptotically normal estimates it has the least possible asymptotic variance. Thus a B.A.N. estimate will be asymptotically the "most accurate" estimate of a parameter; but the value to a statistician of obtaining such estimates is even greater than is indicated by this. In the aforementioned paper of Neyman, a simple method of testing hypotheses is described which is asymptotically equivalent to the likelihood ratio test and involves the use of the χ^2 statistic and a B.A.N. estimate. It usually turns out that the hardest work in applying this technique is in computing the estimate. Thus it is important to have a number of different methods for computing B.A.N. estimates available to the applied statistician. The usual methods of obtaining B.A.N. estimates will be summarized briefly in section 2.

The objective of all these methods is at least in part a practical one and is essentially two-fold. First, it is hoped that some of these estimates will be easily computable. Second, even though all these estimates have the same

Received July 11, 1956; revised May 7, 1958.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

[2] Now at the University of California at Los Angeles.

asymptotic properties, they may differ widely in their small sample properties, and it seems reasonable that the choice of the proper estimate to use should depend in part on the behavior of the estimate for small samples. As a consequence, a large class of estimates with best asymptotic properties is proposed with the hope that some of the easily computable estimates will have small sample properties which are reasonably good. Blind adherence to the principle of maximum likelihood, for example, may lead to more difficult computations and still yield less accurate estimates than other methods of estimation.

A new approach to generating B.A.N. estimates as roots of linear forms of certain variables is suggested in section 3. In cases where minimum distance methods are applicable, the procedure proposed here leads to estimates which are solutions of equations obtained by simplifying in a suitable manner the equations obtained by the original methods. By way of an example, section 4 contains an application of this approach to the problem of estimation of bacterial density.

2. A review of the minimum χ^2 methods of generating B.A.N. estimates. Since the following methods are to be found in the literature at various levels of generalities, a complete mathematical description of the hypotheses necessary for their validity will be omitted.

Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of independent identically distributed s-dimensional random vectors whose distribution depends upon a parameter θ belonging to an open subset Θ of k-dimensional Euclidean space R_k with $k \leqq s$. Let $P(\theta) = E(X \mid \theta)$ be the s-dimensional vector of the expectations of the vector X_n, and let $\Sigma(\theta) = var(X \mid \theta) = E\{[X - P(\theta)]\,[X - P(\theta)]'\}$ be the $s \times s$ covariance matrix which is assumed to be finite and non-singular for each $\theta \varepsilon \Theta$. Furthermore, it is assumed that $P(\theta)$ is a one-to-one bicontinuous map from Θ to a subset of s-dimensional Euclidean space with continuous partial derivatives of the second order. Let Z_n be the s-dimensional random vector defined by $nZ_n = \sum_{j=1}^n X_j$.

The quadratic form

$$n[Z_n - P(\theta)]'\,\Sigma(\theta)^{-1}\,[Z_n - P(\theta)] \tag{2.1}$$

will be designated by the name of χ^2. The value $\hat{\theta}(Z_n)$ of θ which minimizes this quadratic form will be called the minimum χ^2 estimate of θ. An as example take the multinomial case where there are n independent trials each capable of producing any of $s + 1$ possible outcomes. Let the probability on each successive trial be $p_i(\theta)$ of producing the ith outcome. Let z_i denote the proportion of the trials which result in the ith outcome. Then

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{(z_i - p_i(\theta))^2}{p_i(\theta)} \tag{2.2}$$

is the familiar Pearson χ^2. It may be shown that (2.2) is algebraically equal to the χ^2 of the form (2.1) where the vector Z_n is the vector of the first s z_i's. The advantage of (2.1) lies in the fact that it describes a method for estimating parameters of continuous distributions.

Barankin and Gurland [2] have shown that the minimum χ^2 estimate, as defined above, is B.A.N. where the X_n have distributions belonging to a Koopman's family, and Z_n is a vector of sufficient statistics. When the distributions under consideration do not form a Koopman's family with sufficient statistics Z_n, the term B.A.N. estimate is perhaps not entirely justifiable but will be retained for convenience. The precise definition of B.A.N. estimate to be adopted is somewhat irrelevant, because the methods reviewed in this section and the method developed in section 3, give estimates which have the same asymptotic behavior as the minimum χ^2 estimates. In section 3.3, the sense in which the estimates are best is stated more precisely.

Starting with this basic minimum χ^2 estimate, several methods may be used to generate large classes of estimates. These methods will be described below. Method I is due essentially to Karl Pearson. Method II as a general method may be found in Barankin and Gurland [2] and Taylor [5], but special cases were used earlier (see Berkson [6]). Method III evolved from practical work and is of unknown authorship. Method IV is due to Neyman [1].

Method I. Modification. Let $M_n(Z_n, \theta)$ be an $s \times s$ symmetric positive definite matrix. The quadratic form

$$Q_n(\theta) = n[Z_n - P(\theta)]'M_n(Z_n, \theta)[Z_n - P(\theta)] \tag{2.3}$$

will be called the modified or reduced χ^2. The estimate $\hat{\theta}_M(Z_n)$ which minimizes the modified χ^2 with the function $M_n(Z_n, \theta)$ depending only on Z_n and not on θ or n, will be called the minimum modified χ^2 estimate of θ. For example, the estimate which minimizes the Pearson modified χ^2,

$$\chi_M^2 = n \sum_{i=1}^{s+1} \frac{(z_i - P_i(\theta))^2}{z_i} \tag{2.4}$$

is such an estimate.

Under the condition that $M_n(Z_n, \theta) \to \Sigma^{-1}(\theta)$ in probability as $n \to \infty$ when θ is the true value of the parameter, and under certain regularity conditions, the minimum modified χ^2 estimate of θ will have the same asymptotic properties as the minimum χ^2 estimate of θ (or simply $\theta_M(Z_n)$ will be B.A.N., according to the conventions made.)

Method II. Transformation. Let $g(x)$ be any function from R_s to R_s with continuous first partial derivatives

$$g(x) = \begin{pmatrix} g_1(x_1, \cdots, x_s) \\ \vdots \\ g_s(x_1, \cdots, x_s) \end{pmatrix} \tag{2.5}$$

Let the $s \times s$ matrix of first partial derivatives be denoted by

$$\dot{g}(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} g_1 & \cdots & \frac{\partial}{\partial x_1} g_s \\ \vdots & & \\ \frac{\partial}{\partial x_s} g_1 & \cdots & \frac{\partial}{\partial x_s} g_s \end{pmatrix} \tag{2.6}$$

We shall call the quadratic form

$$n[g(Z_n) - g(P(\theta))]'[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}[g(Z_n) - g(P(\theta))] \tag{2.7}$$

the transformed χ^2. More generally, we may consider the combinations of Methods I and II, and replace the matrix of the quadratic form (2.7) by an estimate

$$Q_n(\theta) = n[g(Z_n) - g(P(\theta))]'M_n(Z_n, \theta)[g(Z_n) - g(P(\theta))]. \tag{2.8}$$

We assume that $M_n(Z_n, \theta) \to [\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}$ in probability and the regularity conditions needed for Method I. In addition, one needs regularity conditions on g, namely that g is a one-to-one bicontinuous map from a neighborhood of $P(\Theta)$ into R_s, with continuous partial derivatives of the second order and that the matrix $\dot{g}(P(\theta))$ is nonsingular for each $\theta \in \Theta$. Then the minimum transformed χ^2 estimates, that is the value $\hat{\theta}_T(Z_n)$ of θ minimizing (2.7), will be a B.A.N. estimate of θ.

This method of generating B.A.N. estimates also applies to the χ^2 of (2.2); for example, letting $g_i(x)$ be the real-valued transformation applied to the ith cell

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{p_i(\theta)g_i'(p_i(\theta))} \tag{2.9}$$

or modified,

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{z_i\, g_i'(z_i)^2}. \tag{2.10}$$

The well-known example of Berkson [6] is of the type (2.10).

Many times the functions g_i may be chosen so that $g_i(p_i(\theta))$ is a linear function of the parameters $\theta_1, \cdots, \theta_k$. In such cases finding the value of θ which minimizes the χ^2 of equation (2.10) results in solving k linear equations in k unknowns. The reader may consult the paper of W. Taylor [5] for examples.

Method III. Expansion in a Taylor series about a $O(\sqrt{n})$-consistent estimate. An estimate θ_n^* of θ will be called $O(\sqrt{n})$-consistent if $\sqrt{n}(\theta_n^* - \theta)$ is bounded in probability uniformly in n when θ is the true value of the parameter; that is, for every $\epsilon > 0$ and $\theta \in \Theta$, there exists a number B so large that for every $n = 1, 2, \cdots$

$$P[\sqrt{n} \mid \theta_n^* - \theta \mid > B \mid \theta] < \epsilon. \tag{2.11}$$

Many types of estimates satisfy this requirement. For example, under certain regularity conditions, estimation by the method of moments yields estimates θ_n^* for which $\sqrt{n}(\theta_n^* - \theta)$ is asymptotically normal when θ is the true value of the parameter. This follows from a theorem of Cramér [7], p. 366, which states that certain functions of the moments are asymptotically normal. Such asymptotically normal estimates as this are obviously $O(\sqrt{n})$-consistent.

One may try to apply a correction to θ_n^* by an application of the method of expansion in a Taylor series to get an estimate closer to the minimum χ^2 estimate. It is known, however, that one such application to a $O(\sqrt{n})$-consistent estimate will give a B.A.N. estimate. More specifically, consider the expansion of some one of the previously mentioned χ^2's (modified and/or transformed) in a Taylor series to the second degree terms about a $O(\sqrt{n}$-consistent estimate θ_n^* of θ.

$$\chi^2(\theta) = \chi^2(\theta_n^*) + \dot{\chi}^2(\theta_n^*)'(\theta - \theta_n) + \tfrac{1}{2}(\theta - \theta_n^*)'\ddot{\chi}^2(\theta_n^*)(\theta - \theta_n^*) + \text{Rem.} \tag{2.12}$$

where $\dot{\chi}^2(\theta)$ is the $k \times 1$ vector of first derivatives of $\chi^2(\theta)$ and $\ddot{\chi}^2(\theta)$ is the $k \times k$ matrix of second derivatives of $\chi^2(\theta)$.

$$\dot{\chi}^2(\theta) = \begin{pmatrix} \frac{\partial}{\partial\theta_1}\chi^2(\theta) \\ \vdots \\ \frac{\partial}{\partial\theta_k}\chi^2(\theta) \end{pmatrix} \tag{2.13}$$

$$\ddot{\chi}^2(\theta) = \begin{pmatrix} \frac{\partial^2}{\partial\theta_1^2}\chi^2(\theta) & \cdots & \frac{\partial^2}{\partial\theta_1\,\partial\theta_k}\chi^2(\theta) \\ \vdots & & \\ \frac{\partial^2}{\partial\theta_k\,\partial\theta_1}\chi^2(\theta) & \cdots & \frac{\partial^2}{\partial\theta_k^2}\chi^2(\theta) \end{pmatrix}. \tag{2.14}$$

Instead of finding that value of θ which minimizes $\chi^2(\theta)$, one may discard the remainder term and find that value $\hat{\theta}_n$ of θ which minimizes the first three terms of the expansion. This estimate $\hat{\theta}_n$ will then be a B.A.N. estimate of θ. This method of generating B.A.N. estimates is important because it leads to k linear equations in k unknowns and is thus comparatively easy to apply.

Method IV. Linearization of the side conditions. This method, due to Neyman [1], was proposed with the specific intention of finding a B.A.N. estimate which could be computed by solving linear equations. In minimizing some χ^2 like (2.1), one may consider the vector P as the vector of parameters which are subject to certain restrictions, called side conditions, due to the dependence of P on θ. If there are s independent components of the vector P and k parameters, there will be $s - k$ side conditions on the p's.

$$f_j(p_1, \cdots, p_s) = 0 \qquad \text{for } j = 1, \cdots, k - s \tag{2.16}$$

One may then minimize χ^2 subject to these side conditions by the method of Lagrange multipliers. However, a simpler procedure would be to minimize χ^2 subject to the linearized counterpart of (2.15), that is, the first two terms of the Taylor series expansion about the point z_n. The solution for the estimate then only requires solution of linear equations. For a fuller account of the subject, the reader should consult the papers of Neyman and of Barankin and Gurland. The outline of the method given here is added only for the sake of completeness and no mention of the method will be made in the later sections of the paper.

3. B.A.N. estimates as roots of linear forms. The method customarily used to find a minimum χ^2 estimate is to differentiate χ^2 with respect to each of the parameters separately, set the results equal to zero and solve the resulting system of simultaneous equations. For example, one may differentiate the χ^2 of the equation (2.4) and obtain the equations

$$-2n \sum_{i=1}^{s} \frac{z_i - p_i(\theta))}{z_i} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0 \qquad \text{for } j = 1, 2, \cdots, k, \tag{3.1}$$

or one may differentiate the χ^2 of equation (2.3) with $M_n(Z_n, \theta)$ a function of Z_n only, such that $M_n(Z_n) \to \Gamma(\theta)$ in probability and the regularity conditions hold, and obtain

$$-n\dot{P}(\theta)M(Z_n)(Z_n - P(\theta)) = 0 \tag{3.2}$$

where $\dot{P}(\theta)$ is the $k \times s$ matrix of first partial derivatives of the vector $P(\theta)$,

$$\dot{P}(\theta) = \begin{pmatrix} \frac{\partial P_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial P_s(\theta)}{\partial \theta_1} \\ \vdots & & \\ \frac{\partial P_1(\theta)}{\partial \theta_k} & \cdots & \frac{\partial P_s(\theta)}{\partial \theta_k} \end{pmatrix} \tag{3.3}$$

and the 0 is the $k \times 1$ vector with a zero term in each component so that (3.2) represents k equations in k unknowns.

Well-chosen roots to equations such as (3.1) and (3.2) are B.A.N. estimates of the unknown parameters. This suggests that instead of starting with a quadratic form in $(Z_n - P(\theta))$ and finding values of θ which make the form a minimum, it may be simpler to start with an arbitrary linear form in $(Z_n - P(\theta))$ and find the roots. Roots of certain such linear forms, namely, (3.1) and (3.2), will be B.A.N. estimates. Furthermore, such a method of generating B.A.N. estimates will probably satisfy the requirement that they be easy to compute. It is the purpose of this section to investigate the asymptotic distribution of roots of linear forms in $(Z_n - P(\theta))$, and the conditions for such roots to be B.A.N. estimates of the parameters.

3.1. Preliminary lemma. This section contains an implicit function theorem needed for the proof of the main theorem. First an implicit function theorem which can be found in Pierpont [8], p. 293, for example is stated, from which the lemma of this section will follow. The unicity of the implicit function is stated in a somewhat stronger form than found in Pierpont. This strengthening can be obtained by modifying his proof slightly and the details of the proof need not be given here.

Let $F(x, u)$ *be a function of variables* $x \varepsilon R_s$ *and* $u \varepsilon R_k$ *with values in* R_k. *Let* $a \varepsilon R_s$ *and* $b \varepsilon R_k$, *and assume that*

(i) $F(x, u)$ *is continuous and* $F_u(x, u)$ *exists and is continuous in a neighborhood of the point* (a, b).

(ii) $F(a, b) = 0$ *and* $F_u(a, b)$ *is nonsingular. Then, there exists a neighborhood* N *of* a, *and a function* $\phi(x)$ *from* R_s *to* R_k, *such that*

(1) $\phi(x)$ *is continuous in* N,

(2) $\phi(a) = b$,

(3) $F(x, \phi(x)) = 0$ *for* $x \varepsilon N$, *and*

(4) (*uniqueness*) *there exists a neighborhood* N' *of the point* b *such that for* $u \varepsilon N'$ *and* $x \varepsilon N$, $F(x, a) \neq 0$ *unless* $u = \phi(x)$.

In the above theorem $F_u(x, u)$ represents the $k \times k$ matrix of partial derivatives of $F(x, u)$ with respect to u, as in equation (3.3). The assumption of continuity of $F_u(x, u)$ means that each component of the matrix is assumed to be continuous.

The following lemma is an extension of this theorem, similar to that found in Graves [9], p. 144, to the situation in which $F(x, u)$ is known to vanish along some curve in R_{s+k}, rather than just at one point.

LEMMA. *Let* $F(x, u)$ *be a function of variables* $x \varepsilon R_s$ *and* $u \varepsilon R_k$ *with values in* R_k, $k \leqq s$. *Let* $p(u)$ *be a function from some set* $D \subset R_k$ *to* R_s, *and assume that*

(i) D *is an open set*,

(ii) $p(u)$ *is one-to-one and inversely continuous from* D *into* R_s,

(iii) *there is a neighborhood of the curve* $\{(p(u), u) : u \varepsilon D\}$ *in which* $F(x, u)$ *is continuous and* $F_u(x, u)$ *exists and is continuous.*

(iv) $F(p(u), u) = 0$ *and* $F_u(p(u), u)$ *is nonsingular for every* $u \varepsilon D$.

Then, there exists a neighborhood N *of the set* $S = \{p(u) : u \varepsilon D\}$ *and a function* $\phi(x)$ *from* R_s *to* R_k *such that*

(a) $\phi(x)$ *is continuous in* N,

(b) $\phi(p(u)) = u$ *for* $u \varepsilon D$,

(c) $F(x, \phi(x)) = 0$ *for* $x \varepsilon N$, and

(d) *there exists a neighborhood of the curve* $\{(p(u), u) : u \varepsilon D\}$ *in which the only zeros of the function* $F(x, u)$ *are the points* $(x, \phi(x))$.

PROOF. From the previous implicit function theorem, for every $u \varepsilon D$, there is a neighborhood $N_{p(u)}$ of the point $p(u)$ and a function $\phi_u(x)$ from R_s to R_k such that

(1) $\phi_u(x)$ is continuous in $N_{p(u)}$,

(2) $\phi_u(p(u)) = u$,

(3) $F(x, \phi_u(x)) = 0$ for $x \varepsilon N_{p(u)}$, and

(4) for y in some neighborhood N_u of the point u, and $x \varepsilon N_{p(u)}$

$$F(x, y) \neq 0 \quad \text{unless} \quad y = \phi_u(x).$$

Using the inverse continuity of the function $p(u)$, and the continuity of the function $\phi_u(x)$, we may replace the neighborhoods $N_p(u)$ by spherical neighborhoods $N'_{p(u)}$ with the two additional properties that

(5) if $p(u_1) \varepsilon N'_{p(u_2)}$ for some u_1 and $u_2 \varepsilon D$, then $u_1 \varepsilon N_{u_2}$

and

(6) if $x \varepsilon N'_{p(u)}$ for some $u \varepsilon D$, then $\phi_u(x) \varepsilon N_u$.

Now consider spherical neighborhoods $N''_{p(u)}$ with radii equal to $\frac{1}{3}$ that of $N'_{p(u)}$, and let N denote $\bigcup_{u \varepsilon D} N''_{p(u)}$ The set N is then obviously a neighborhood of the set S.

We will show that if $x_0 \varepsilon N''_{p(u_1)} \cap N''_{p(u_2)}$, then $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$. For since $N''_{p(u_1)} \cap N''_{p(u_2)}$ is not empty, either $p(u_1) \varepsilon N'_{p(u_2)}$ or $p(u_2) \varepsilon N'_{p(u_1)}$. Suppose without loss of generality that the former is true; then since $u_1 \varepsilon N_{u_2}$ and

$$F(p(u_1), \phi_{u_2}(p(u_1))) = 0,$$

we have $\phi_{u_2}(p(u_1)) = u_1$. Furthermore, for $x \varepsilon N''_{p(u_1)} \cap N'_{p(u_2)}$, $\phi_{u_2}(x)$ is continuous and satisfies $F(x, \phi_{u_2}(x)) = 0$; but $\phi_{u_1}(x) \varepsilon N_{u_1}$ for $x \varepsilon N'_{p(u_1)}$ and thus $\phi_{u_1}(x)$ is the unique function, continuous in $N''_{p(u_1)}$ and such that

$$\phi_{u_1}(p(u_1)) = u_1 \qquad \text{and} \qquad F(x, \phi_{u_1}(x)) = 0.$$

Hence, $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$.

Thus for $x \varepsilon N$ we may define $\phi(x) = \phi_u(x)$ for any u for which $x \varepsilon N''_{p(u)}$, since such a definition is unique. Now parts (a), (b), and (c) of the conclusion of the lemma are obvious. As for (d), the neighborhood can be chosen to be $\bigcup_{u \varepsilon D}[N''_{p(u)} \times N_u]$.

3.2. The main theorem. Let Z_n, $n = 1, 2, \cdots$ be a sequence of s-dimensional random vectors whose distribution depends upon a parameter θ in some set $\Theta \subset R_k$, $k \leqq s$. Let $P(\theta)$ be a function from Θ to R_s.

Assumption 1. *Θ is an open set.*

Assumption 2. *$\mathfrak{L}\{\sqrt{n}(Z_n - P(\theta)) \mid \theta\} \rightarrow \mathfrak{L}(Z)$ where Z is a normal random vector with mean zero and variance-covariance matrix $\Sigma(\theta)$. (That is,*

$$EZ = O, \quad EZZ' = \Sigma(\theta).)$$

The convergence used above is convergence in law or in distribution. Assumption 2 states that when θ is the true value of the parameter, the distribution of $\sqrt{n}(Z_n - P(\theta))$ converges to a normal distribution with mean zero and variance-covariance matrix $\Sigma(\theta)$. The law degenerate at some point a will be denoted by $\mathfrak{L}(a)$. Thus $\mathfrak{L}(X_n) \rightarrow \mathfrak{L}(a)$ means that X_n converges in probability to a.

Assumption 3. *The mapping $P(\theta)$ from Θ into R_s is homeomorphic (that is, one-to-one and bicontinuous) and continuously differentiable.*

Let $f(x, \theta)$ be a $k \times s$ matrix for each $x \varepsilon R_s$ and $\theta \varepsilon \Theta$.

Assumption 4. *There is a neighborhood $N_0 \subset R_s \times \Theta$ of the set*

$$\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$$

within which $f(x, \theta)$ *and* $\partial/\partial\theta_j f(x, \theta)$ *for* $j = 1, 2, \cdots, k$ *are continuous jointly in* (x, θ).

Let $b(\theta) = f(P(\theta), \theta)$ and let $\dot{P}(\theta)$ be the $k \times s$ matrix of partial derivatives of $P(\theta)$, given by equation (3.3).

ASSUMPTION 5. *The matrix* $\dot{P}(\theta)b(\theta)'$ *is nonsingular for each* $\theta \varepsilon \Theta$.

Let

$$F(x, \theta) = f(x, \theta)(x - P(\theta)). \tag{3.4}$$

This is the linear form which will be used in the sequel to generate B.A.N. estimates of the parameter θ. The following theorem shows immediately that the root to the equation $F(Z_n, \theta) = 0$ will be a $O(\sqrt{n})$-consistent estimate of θ.

THEOREM 1. *Under assumptions* 1 *through* 5, *there exists a neighborhood* N *of the set* $S = \{P(\theta): \theta \varepsilon \Theta\}$ *and a unique function* $\hat{\theta}(x)$ *from* R_s *to* R_k *continuous in* N, *such that* $\hat{\theta}(P(\theta)) = \theta$ *for* $\theta \varepsilon \Theta$, *and* $F(x, \hat{\theta}(x)) = 0$ *for* $x \varepsilon N$. *Moreover,* $\mathcal{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) | \theta\} \rightarrow \mathcal{L}(Y)$ *where* Y *is a normal random vector with mean zero and variance-covariance matrix given by*

$$[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}$$

PROOF. $F(P(\theta), \theta) = 0$ and

$$F_\theta(x, \theta) = f_\theta(x, \theta)(x - P(\theta)) - \dot{P}(\theta)f(x, \theta)' \tag{3.5}$$

where $f_\theta(x, \theta)$ represents the $k \times k \times s$ cubic matrix of partial derivatives of the $k \times s$ matrix $f(x, \theta)$ with respect to θ. To avoid confusion we will write out the first term of this difference completely. Denote the function in the ith row, jth column of $f(x, \theta)$ by $f_{ij}(x, \theta)$, and let $P_j(\theta)$ and x_j represent the jth component of the vectors $P(\theta)$ and x. Then,

$$f_\theta(x, \theta)(x - P(\theta)) = \sum_{j=1}^{s} \begin{pmatrix} \frac{\partial}{\partial\theta_1} f_{1j} & \cdots & \frac{\partial}{\partial\theta_1} f_{sj} \\ \vdots & & \\ \frac{\partial}{\partial\theta_k} f_{1j} & \cdots & \frac{\partial}{\partial\theta_k} f_{sj} \end{pmatrix} (x_j - P_j(\theta)). \tag{3.6}$$

It is now easily checked that formula (3.5) holds. Hence,

$$F_\theta(P(\theta), \theta) = -\dot{P}(\theta)b(\theta)' \tag{3.7}$$

which, by assumption, is nonsingular for every $\theta \varepsilon \Theta$. Thus the hypotheses of the lemma of the previous section are satisfied and the first part of the theorem is proved.

To prove the second part, expand $F(x, \theta)$ about the point $\hat{\theta}(x)$ to one term using the formula

$$F(x, \theta) = F(x, \hat{\theta}(x)) + \left[\int_0^1 F_\theta\{x, \hat{\theta}(x) + \lambda(\theta - \hat{\theta}(x)\} \, d\lambda\right]' (\theta - \hat{\theta}(x)) \tag{3.8}$$

which may easily be checked. By the integral of a matrix we mean the matrix of the integrals of each term separately. For each $\theta \varepsilon \Theta$, formula (3.8) is valid whenever x is sufficiently close to $p(\theta)$, so that $(x, \hat{\theta}(x))$ is in a spherical neighborhood of $(p(\theta), \theta)$ contained entirely in N_0. We may replace x by Z_n in (3.8) and multiply both sides by $\sqrt{n}$.

$$\sqrt{n}\left[-\int_0^1 F_\theta\{Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))\}\, d\lambda\right]'(\hat{\theta}(Z_n) - \theta) = f(Z_n, \theta)\sqrt{n}\,(Z_n - P(\theta)). \tag{3.9}$$

We now invoke the theorems of Slutsky (see [10], section 2, theorem 2, or [4]). From assumption 1, $\mathcal{L}(Z_n \mid \theta) \to \mathcal{L}(P(\theta))$. Hence by Slutsky's theorem, since $f(x, \theta)$ is continuous in a neighborhood of $(p(\theta), \theta)$,

$$\mathcal{L}(f(Z_n, \theta) \mid \theta) \to \mathcal{L}(f(P(\theta), \theta)) = \mathcal{L}(b(\theta)). \tag{3.10}$$

Slutsky's theorem also gives

$$\mathcal{L}(f(Z_n, \theta)\sqrt{n}(Z_n - P(\theta)) \mid \theta) \to \mathcal{L}(b(\theta)Z) \tag{3.11}$$

where Z is a normal vector with zero mean and variance-covariance matrix $\Sigma(\theta)$. Since $\mathcal{L}(Z_n \mid \theta) \to \mathcal{L}(P(\theta))$ and $\mathcal{L}(\hat{\theta}(Z_n) \mid \theta) \to \mathcal{L}(\hat{\theta}(P(\theta))) = \mathcal{L}(\theta)$, we may apply the Lebesgue bounded convergence theorem to the integral in (3.9).

$$\mathcal{L}\left\{\int_0^1 F_\theta[Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))]\, d\lambda \mid \theta\right\} \to \mathcal{L}\left\{\int_0^1 F_\theta[P(\theta), \theta]\, d\lambda\right\} = \mathcal{L}\{F_\theta(P(\theta), \theta)\} = \mathcal{L}\{-\dot{P}(\theta)b(\theta)'\} \tag{3.12}$$

by equation (3.7). Another application of Slutsky's theorem allows us to conclude

$$\mathcal{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) \mid \theta\} \to \mathcal{L}\{[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z\}. \tag{3.13}$$

Denoting $[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z$ by Y, we see that Y is a normal random vector, with mean zero and covariance matrix

$$\begin{aligned} EYY' &= E[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)ZZ'b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1} \\ &= [b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}. \end{aligned} \tag{3.14}$$

3.3. Applications. The theorem just proved allows some immediate inferences. The important point to notice in this theorem is that the asymptotic distribution of $\sqrt{n}(\hat{\theta}(Z_n) - \theta)$ depends on the function $f(x, \theta)$ only through its values along the curve $\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$. Thus if the linear form

$$F(Z_n, \theta) = f(Z_n, \theta)(Z_n - P(\theta))$$

has a root which is already a B.A.N. estimate of θ, any linear form

$$g(Z_n, \theta)(Z_n - P(\theta)),$$

in which the function $f(x, \theta)$ is replaced by any function $g(x, \theta)$ satisfying assumption 4 and for which $g(P(\theta), \theta) = f(P(\theta), \theta)$, will have a root which is also a B.A.N. estimate of θ, since the asymptotic distribution of the two roots will be the same.

For example, equation (3.2) (neglecting the factor n which is immaterial as far as roots are concerned) is a linear form of the type $f(Z_n, \theta)(Z_n - P(\theta))$ for which

$$f(Z_n, \theta) = \dot{P}(\theta)M(Z_n). \tag{3.15}$$

Since $M(Z_n)$ converges in probability to $\Sigma(\theta)^{-1}$ when θ is the true value of the parameter, $M(P(\theta)) = \Sigma(\theta)^{-1}$ so that

$$b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1} \tag{3.16}$$

Now consider functions

$$f_1(Z_n, \theta) = b(\theta) \quad \text{and} \quad f_2(Z_n, \theta) = L(Z_n)M(Z_n) \tag{3.17}$$

where L is a matrix continuous in a neighborhood of $\{P(\theta):\theta \varepsilon \Theta\}$, such that $L(P(\theta)) = \dot{P}(\theta)$. If $f_1(Z_n, \theta)$ is used, we must also assume that $b(\theta)$ has a continuous derivative. In these circumstances, whenever the root to equation (3.2) is a B.A.N. estimate, roots to the linear forms involving $f_1(Z_n, \theta)$ and $f_2(Z_n, \theta)$ will be B.A.N. also.

Now we will show directly the exact conditions under which there will be a root of a linear form which will be "best" out of the class of all roots of linear forms; that is, the exact conditions under which there is a value of $b(\theta)$ which minimizes the variance (3.14).

Of two n by n matrices, A and B, A will be said to be smaller than B, in symbols $A < B$, if and only if $B - A$ is positive semi-definite; that is, if

$$x'[B - A]x \geqq 0$$

for every n-dimensional vector x. Thus of two unbiased estimates of a vector parameter θ, T_1 and T_2, with covariance matrices respectively A and B, T_1 would be preferred to T_2 if $A < B$, since the unbiased estimate $x'T_1$ of the parameter $x'\theta$ will have a smaller variance than the unbiased estimate $x'T_2$ of the same parameter.

THEOREM 2. *If in addition to assumptions* 1 *through* 5 *there exists an s by s nonsingular matrix* $\Sigma_0(\theta)$ *such that*

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = \dot{P}(\theta)' \tag{3.18}$$

then the asymptotic covariance matrix of $\hat{\theta}(Z_n)$ *taken on its minimum value when* $b(\theta) = \dot{P}(\theta)\Sigma_0(\theta)$. *The minimum value is then* $[\dot{P}(\theta)\Sigma_0(\theta)\dot{P}(\theta)']^{-1}$.

PROOF. For simplicity of notation the θ will be omitted. From assumption 5, $\dot{P}$ is of full rank so that $[\dot{P}\Sigma_0\dot{P}]$ is nonsingular. The inequality

$$(b'[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}']^{-1})\Sigma(b[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}]^{-1}) \geqq 0 \tag{3.19}$$

which holds since Σ is positive semi-definite, yields

$$[b\dot{P}']^{-1}b\Sigma b'[\dot{P}b']^{-1} - [\dot{P}\Sigma_0\dot{P}']^{-1} \geqq 0. \tag{3.20}$$

Yet it is easily checked that equality is attained if $b = \dot{P}\Sigma_0'$. qed.

The assumption of the existence of a matrix $\Sigma_0(\theta)$ satisfying (3.18) holds for example when $\Sigma(\theta)$ is nonsingular. Then $b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1}$ as was found in equation (3.16). However, in other important cases, for example in the multinomial case with the χ^2 of equation (2.2), the matrix $\Sigma(\theta)$ is singular. The following lemma which may be proved without difficulty, will perhaps be of aid in checking whether a Σ_0 satisfying (3.18) exists at all.

LEMMA. *In order that there exist a nonsingular matrix $\Sigma_0(\theta)$ satisfying* (3.18), *it is necessary and sufficient that the range space of $\dot{P}(\theta)'$ be contained in the range space of $\Sigma(\theta)$: that is, for every vector x there exists a vector $y(\theta)$ such that*

$$\Sigma(\theta)y(\theta) = \dot{P}(\theta)'x.$$

In certain cases one can find the matrix Σ_0 which satisfies (3.18). We shall do it now for the multinomial case. In this case the vector $P(\theta)$ is simply the vector of cell probabilities, and is $s + 1$ dimensional. The matrix $\Sigma(\theta)$ is found to be

$$\Sigma(\theta) = \begin{bmatrix} p_1(\theta) - p_1^2(\theta) & -p_1(\theta)p_2(\theta) & \cdots & -p_1(\theta)p_{s+1}(\theta) \\ -p_1(\theta)p_2(\theta) & p_2(\theta) - p_2^2(\theta) & & \\ \vdots & & & \\ -p(\theta)p_{s+1}(\theta) & & \cdots & p_{s+1}(\theta) - p_{s+1}^2(\theta) \end{bmatrix} \tag{3.21}$$

which may be expressed simply as

$$\Sigma(\theta) = B(\theta) - P(\theta)P(\theta)' \tag{3.22}$$

where $B(\theta)$ is the diagonal matrix

$$B(\theta) = \begin{bmatrix} p_1(\theta) & 0 & \cdots & 0 \\ 0 & p_2(\theta) & & \\ \vdots & & \ddots & \\ 0 & & & p_{s+1}(\theta) \end{bmatrix} \tag{3.23}$$

Then, as suggested by the χ^2 of (2.2), put $\Sigma_0(\theta) = B(\theta)^{-1}$.

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = B(\theta)B(\theta)^{-1}\dot{P}(\theta)' - P(\theta)P(\theta)'B(\theta)^{-1}\dot{P}(\theta)'. \tag{3.24}$$

It is easily seen that

$$P(\theta)'B(\theta)^{-1}\dot{P}(\theta)' = \left(\sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_1} p_i(\theta), \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_2} p_i(\theta), \cdots, \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_k} p_i(\theta)\right). \tag{3.25}$$

This vector must be zero since $\sum_{i=1}^{s+1} p_i(\theta) = 1$. Hence, the equality (3.18) is satisfied. Thus applying Theorem 2, roots of the linear form

$$\sum_{i=1}^{s+1} (z_i - p_i(\theta))f_{ij}(s_1, \cdots, s_{s+1}, \theta) = 0 \qquad j = 1, 2, \cdots k, \tag{3.26}$$

will be "best" when $f_{ij}(p_1(\theta), \cdots, p_{s+1}(\theta), \theta) = \partial/\partial\theta_j \log p_i(\theta)$.

It may further be shown in the multinomial case, that if the $f_{ij}(z, \theta)$ are chosen to be independent of z, and equal to $\partial/\partial\theta_j \log p_i(\theta)$, equation (3.26) will b the derivative of the log of the likelihood function set equal to zero, so that one has immediately that the maximum likelihood estimate, in addition to the minimum modified χ^2 estimate, is an estimate which is given by the root of a certain linear form. One would expect that the linear form (3.26) in which the functions f_{ij} do not depend on θ at all would be somewhat easier to solve for θ. It is this type of linear form which is suggested in section 4 as a method for estimating the bacterial density in a liquid.

We will now apply the preceding theorem to the various minimum χ^2 methods discussed previously.

Application to the transformed χ^2. The method of generating B.A.N. estimates described in Theorems 1 and 2 also applies easily to the transformed χ^2 of equations (2.8) and (2.10). For example, the derivative of the χ^2 of equation (2.8) with $T(Z_n)$ depending on Z_n only, and not on θ, is found to be

$$\frac{\partial}{\partial\theta}\chi_T^2 = n\dot{P}(\theta)\dot{g}(P(\theta))T(Z_n)(g(Z_n) - g(P(\theta))). \tag{3.27}$$

Assumption 1 of Theorem 1 becomes in this case

$$\mathcal{L}\{\sqrt{n}[g(Z_n) - g(P(\theta))] \mid \theta\} \to \mathcal{L}(Z) \tag{3.28}$$

where Z is a normal random vector with zero mean and variance-covariance matrix $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']$. This may easily be checked by expanding $g(Z_n)$ in a Taylor series about the point $P(\theta)$, and invoking asymptotic normality of $\sqrt{n}(Z_n - P(\theta))$. The only requirement on the function $g(x)$ is that it have a continuous derivative in a neighborhood of the curve $\{P(\theta):\theta \varepsilon \Theta\}$. If in addition $g(P(\theta))$ is nonsingular for each $\theta \varepsilon \Theta$, $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))]^{-1}$ will exist and $b(\theta)$ is found to be

$$b(\theta) = \dot{P}(\theta)\dot{g}(P(\theta))[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}. \tag{3.29}$$

Thus, if the root to equation (3.27) is a B.A.N. estimate, the root to the linear form

$$f(Z_n, \theta)(g(Z_n) - g(P(\theta)) = 0 \tag{3.30}$$

will also be a B.A.N. estimate, provided that f satisfies Assumption 4, and that $f(P(\theta), \theta) = b(\theta)$.

The linear form corresponding to the transformed multinomial χ^2 of (2.10) may be computed as before. It becomes

$$\sum_{i=1}^{s+1} [g_i(z_i) - g_i(p_i(\theta))]f_{ij}(z_1, \cdots, z_{s+1}, \theta) = 0 \qquad j = 1, 2, \cdots, k \tag{3.31}$$

where

$$f_{ij}(p_1(\theta), \cdots, p_{s+1}(\theta), \theta) = \left[\frac{\partial}{\partial\theta_j} p_i(\theta)\right]\frac{1}{p_i(\theta)g_i'(p_i(\theta))}. \tag{3.32}$$

Under assumptions 1 through 5, and the assumptions that each $g_i(x)$ is continuous in a neighborhood of the curve $\{x:x = p_i(\theta),\ \theta \varepsilon \Theta\}$ and that

$$g_i'(p_i(\theta)) \neq 0,$$

the roots to equation (3.31) will be B.A.N. estimates of the parameters.

Application to the expansion of χ^2 *in a Taylor series.* Let θ_n^* be a $O(\sqrt{n})$-consistent estimate of the parameter θ. To find the minimum value of the right hand side of equation (2.12) without the remainder term, we take a derivative and solve for the root $\hat{\theta}$.

$$\hat{\theta}_n = \theta_n^* - \ddot{\chi}^2(\theta_n)^{-1}\dot{\chi}^2(\theta_n) \tag{3.33}$$

If we use the modified χ^2 of equation (2.3) for this procedure with M a function of Z_n only, for example $M(Z_n) = \Sigma(\theta_n^*)^{-1}$, the first two derivatives are

$$\begin{aligned} \dot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)) \\ \ddot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta)' - 2n\ddot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)). \end{aligned} \tag{3.34}$$

where $\ddot{P}(\theta)$ is the $k \times k \times s$ cubic matrix of second partial derivatives of the vector $P(\theta)$.

If, on the other hand, we take the linear form with the function $f(Z_n, \theta)$ not depending on θ, say to be $\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}$, and expand it about θ_n^* to the first power and solve for θ, we have

$$\hat{\theta}_n = \theta_n^* + [\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta_n^*)). \tag{3.35}$$

If one compares the estimates (3.35) with the estimates (3.33) with equations (3.34) substituted, one sees that the former require less computation, and that by the amount in the second term of the expression for $\ddot{\chi}^2(\theta)$, involving all the second partial derivatives of the vector $P(\theta)$. Furthermore, computation of $[\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}$ would give an estimate of the limiting variance-covariance matrix of the B.A.N. estimate $\hat{\theta}_n$.

This method would be good for example in estimating the parameters of a Neyman type A distribution, where the vector $P(\theta)$ is a rather complicated function of the parameters, and other methods of getting B.A.N. estimates are rather difficult. This method has been applied by Robert Read of the Statistical Laboratory of the University of California, to estimating the parameters in a probabilistic model describing ionization in a cloud chamber, using as the preliminary estimates, those given by the method of moments. It has also been applied by Dr. Irene Rosenthal of the Psychology Department at the University of California, to estimate the parameters of a latent structure, using as first estimates those of Lazarsfeld [11].

4. Application to the problem of estimating bacterial density by the dilution method. The method of estimating the bacterial density of a liquid by taking samples in fermentation tubes at several levels of dilution of the liquid is well known. As far back as 1915 [12] the maximum likelihood estimate, called the most probable number (M.P.N.) by Biometricians, was suggested for the problem, and is still being used today in Public Health for water, milk, and sewage analysis. This and other estimates have been studied by Fisher [13], Halvorson and Ziegler [14], and Matuszewski, Neyman, and Supinska [15].

The situation is the following. We are given a large volume V of a liquid containing a large number N of bacteria, and we are interested in estimating the bacterial density $\lambda = N/V$, the number of bacteria per unit volume. A sample of size α unit volume is withdrawn and tested by some device such as placing the sample in a fermentation tube to see if any bacteria are present. It is assumed that each bacterium acts independently and that each has the same probability α/V of being in the sample. Thus the number of bacteria in the sample will be binomially distributed with probability α/V and size N; however, if α/V is small and N is large the distribution may conveniently be replaced by a Poisson with parameter $N\alpha/V = \alpha\lambda$. The probability that no bacteria appear in the sample is then $p = e^{-\alpha\lambda}$. If n independent samples of size α are withdrawn and tested, the number K of sterile samples will be binomially distributed with probability p and size n, and may be used to estimate the parameter λ. However, the value of the experiment depends to a great extent on choosing α so that $p = e^{-\alpha\lambda}$ will be in a good estimating range, for if p is too small or too close to one, one will obtain too many fertile or too many sterile samples to be able to estimate λ with much accuracy. And since λ is unknown it will usually be impossible to choose α so that $e^{-\alpha\lambda}$ will be moderately between zero and one. So one usually takes several sizes of sample volumes α_1, $\alpha_2, \cdots, \alpha_s$, called dilution levels, and numbers of samples $n_1, n_2, \cdots, n_s$ at each of the levels, with the hope that at least one of the $e^{-\alpha_i\lambda}$ will be in a good estimating range. Then the numbers $k_1, k_2, \cdots, k_s$, of sterile samples at each of the levels will be used to estimate λ.

The most frequently used B.A.N. estimate of the bacterial density seems to be the maximum likelihood estimate, since the minimum χ^2 estimates appear to be much more difficult to compute. The maximum likelihood estimate of λ is that value of λ which is a root of the equation

$$\sum_{i=1}^{s} \frac{(n_i - k_i)\alpha_i}{(1 - e^{-\alpha_i\lambda})} = \sum_{i=1}^{s} n_i\alpha_i. \tag{4.1}$$

Methods of solving this equation have been discussed by Halvorson and Zeigler [14], Barkworth and Irwin [16], and Finney [17]. Tables of the estimate for certain situations may be found in Halvorson and Zeigler and in Hoskins [18].

An application of the methods of the previous section will yield a B.A.N. estimate which is slightly easier to compute. Linear forms which lead to B.A.N. estimates are of the type

$$\sum_{i=1}^{s} n_i f_i(z, \lambda)(z_i - e^{-\alpha_i\lambda}) \tag{4.2}$$

where z_i represents the frequency of sterile tubes at the ith level of dilution, $z_i = k_i/n_i$, and $f_i(z, \lambda)$ converges in probability to $\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$, z representing the vector $(z_1, \cdots, z_s)$. Equation (4.2) with $f_i(z, \lambda)$ always equal to

$$\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

is equivalent to the maximum likelihood equation (4.1).

We would like to replace $f_i(z, \lambda)$ in equation (4.2) completely by an estimate, that is, $f_i(z, \lambda) = \alpha_i/(1 - z_i)$, but we must take care of the cases in which z_i is equal to one. So we may choose $f_i(z, \lambda) = \alpha_i/(1 - z_i)$ if $z_i \neq 1$ and

$$f_i(z, \lambda) = \alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

if $z_i = 1$. This will lead to a B.A.N. estimate since eventually as the n_i get large without bound, all the z_i will be different from one. We have the equation

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i}(z_i - e^{-\alpha_i\lambda}) + \sum_{z_i = 1} n_i\alpha_i = 0. \tag{4.3}$$

Written in simpler form, this equation becomes

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} e^{-\alpha_i\lambda} = \sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} z_i + \sum_{z_i = 1} n_i\alpha_i. \tag{4.4}$$

This equation is simpler to solve than equation (4.1) in that it only requires tables of e^{-x} which are readily available, while equation (4.1) requires for its solution the computation of $(1 - e^{-\alpha_i\lambda})^{-1}$ separately for each i or tables of $(e^x - 1)^{-1}$ or $(1 - e^{-x})^{-1}$. The method by which it is suggested that (4.4) be solved is the same as that suggested by other authors in connection with the solution of (4.1), and that is Newton's method. For a function $f(x)$ with a continuous first derivative, if x_0 is taken to be the initial guess at the solution of $f(x) = 0$, x_n is defined inductively by

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}. \tag{4.5}$$

Applying this procedure to equation (4.4), we obtain the inductive formula

$$\lambda_n = \lambda_{n-1} + \frac{\sum_{z_i \neq 1} \frac{n_i\alpha_i}{1 - z_i} e^{-\alpha_i\lambda_{n-1}} - \sum_{z_i \neq 1} \frac{n_i\alpha_i}{1 - z_i} z_i - \sum_{z_i = 1} n_i\alpha_i}{\sum_{z_i \neq 1} \frac{n_i\alpha_i^2}{1 - z_i} e^{-\alpha_i\lambda_{n-1}}}. \tag{4.6}$$

The author has made a numerical study of the small sample properties of this estimate, the minimum χ^2 estimate and the maximum likelihood estimate, which he hopes to publish at a later date. An indication is given in this study that in general the estimate given by equation (4.4) has slightly better small sample properties in the sense of bias and root mean square error, than either the maximum likelihood or the minimum χ^2 estimate.

In conclusion, I would like to express my thanks to Professor L. Le Cam for his generous advice and helpful discussions concerning this paper.

REFERENCES

[1] J. Neyman, "Contribution to the theory of the χ^2 test," *Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press (1949).

[2] E. Barankin and J. Gurland, "On asymptotically normal efficient estimators: I," *University of California Publications in Statistics*, Vol. 1, No. 6 (1951), pp. 86–130.

[3] C. L. Chiang, "On regular best asymptotically normal estimates with an application to a stochastic process," Doctoral Dissertation, University of California (1952).

[4] L. Le Cam, Lecture notes on large sample theory (mimeographed), University of California, 1955.

[5] W. Taylor, "Distance functions and regular best asymptotically normal estimates," *Ann. Math. Stat.*, Vol. 24, No. 1 (1953), pp. 85–92.

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

[7] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press (1946).

[8] J. Pierpont, *The Theory of Functions of Real Variables*, Vol. I, Ginn and Co., (1905).

[9] L. M. Graves, *The Theory of Functions of Real Variables*, McGraw-Hill Book Co., Inc. (1946).

[10] H. Chernoff, "Large sample theory: parametric case," *Ann. Math. Stat.*, Vol. 27, No. 1 (1956), pp. 1–22.

[11] S. A. Stouffer, *Measurement and Prediction*, Princeton University Press (1950).

[12] M. H. McCrady, "The numerical interpolation of fermentation-tube results," *Jour. Infec. Dis.*, Vol. 17 (1915), pp. 183–212.

[13] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc., London, Ser. A.*, Vol. 222 (1922), pp. 309–368.

[14] H. O. Halvorsen and N. R. Ziegler, "Application of statistics to problems in bacteriology," *Jour. Bact.*, Vol. 25 (1933), pp. 101–121.

[15] T. Matuszewski, J. Neyman, and J. Supinska, "Statistical studies in questions of bacteriology," *J. Roy. Stat. Soc., Ser. B.*, Vol. 2 (1932), pp. 63–82.

[16] H. Barksworth and J. O. Irwin, "Distribution of coliform organisms in milk and the accuracy of the presumptive coliform test," *Jour. Hyg.*, Cambridge, Vol. 38 (1938), pp. 446–457.

[17] D. J. Finney, "The principle of biological assay," *J. Roy. Stat. Soc., Ser. B.*, Vol. 9 (1947), pp. 46–91.

[18] J. K. Hoskins, "Most probable numbers for evaluation of coli aerogenes tests by fermentation tube method," Public Health Reports, Vol. 49 (1934), p. 393, Reprint No. 1621 (revised 1940).