DISTANCE FUNCTIONS AND REGULAR BEST ASYMPTOTICALLY NORMAL ESTIMATES

BY WILLIAM F. TAYLOR[1]

School of Aviation Medicine, Randolph Field, Texas

Summary. Among the methods of obtaining satisfactory parameter estimates are maximum likelihood, minimum chi-square, minimum "reduced" chi-square, etc. This paper presents a generalization of the minimum chi-square method which yields regular best asymptotically normal (RBAN) estimates and which is often very simple to apply. It is shown that the least squares expressions associated with the logit and probit transformations are a type which lead to RBAN estimates.

1. Introduction. In 1945, J. Neyman [1] presented at the Berkeley Symposium on Mathematical Statistics and Probability his work on "best asymptotically normal" estimates (now called "regular best asymptotically normal" or RBAN.) He gave for multinomial situations several methods of estimation which yield estimates having desirable asymptotic properties. The estimation techniques developed by Neyman were all based on the minimization of a special kind of distance function, namely, the χ^2 goodness-of-fit expression or a similar one called the "reduced" χ^2.

Certain work by J. Berkson [2] brought the author's attention to functions which were a generalization of the χ^2 distance function and which yielded estimates upon minimization. In this paper there is presented a class of distance functions which lead to RBAN estimates and which includes minimum χ^2, logit, and probit estimates. The theorem of Section 3 is proved via a lemma from results given in [1]. It has been pointed out to the author that this theorem may also be obtained readily from the work of Barankin and Gurland [4].

The author wishes to thank Professor Neyman and Dr. Berkson for their assistance in this work.

2. Distance functions leading to RBAN estimates. Suppose the situation is the one described in [1], page 239. There are s sequences of independent trials, each sequence consisting of n_i trials. A trial of the ith sequence can produce ν_i exclusive results with probabilities

$$p_{i1}, p_{i2}, \cdots, p_{i\nu_i}; \sum_{j=1}^{\nu_i} p_{ij} = 1.$$

Received 9/17/52.

[1] This work was done while the author was at the University of California, Berkeley, California.

Let $N = \sum_{i=1}^{s} n_i$ and $Q_i = n_i/N$. By θ is meant the set of m unknown parameters $(\theta_1, \theta_2, \cdots, \theta_m)$. The p_{ij} are assumed equal to $f_{ij}(\theta)$, where the f_{ij} are continuous functions having continuous partial derivatives up to the second order. It is assumed that $f_{ij}(\theta) > 0$ and

$$\sum_{j=1}^{\nu_i} f_{ij}(\theta) = 1, \qquad i = 1, \cdots, s. \tag{1}$$

$\theta_1, \cdots, \theta_m$ are assumed functionally independent in that for some m functions, f_{ij}, the determinant

$$\begin{vmatrix} f_{\alpha_1\beta_1 1} & \cdots & f_{\alpha_1\beta_1 m} \\ \vdots & & \vdots \\ f_{\alpha_m\beta_m 1} & \cdots & f_{\alpha_m\beta_m m} \end{vmatrix} \neq 0; \qquad f_{\alpha_k\beta_k j} = \frac{\partial f_{\alpha_k\beta_k}}{\partial\theta_j}. \tag{2}$$

Definition 1. *$\delta(p, q)$ will be called a distance function of $p = (p_{11}, \cdots, p_{s\nu_s})$ and $q = (q_{11}, \cdots, q_{s\nu_s})$ if it is such that*

(i) $\delta(p, p) = 0$,

(ii) $\delta(p, q) > 0$ for $p \neq q$,

(iii) *$\delta(p, q)$ is continuous with continuous partial derivatives up to the second order.*

Letting $p_{ij} = f_{ij}(\theta)$, the problem is to estimate the θ's. Let $\vartheta_i(q)$, $i = 1, \cdots m$, be functions of q which are estimates of $\theta_1, \cdots, \theta_m$, respectively. Let,

$$\partial\delta(p, q)/\partial\theta_k = \psi_k(\theta, q).$$

The following lemma is due to Werner Leimbacher, formerly of the Statistical Laboratory at the University of California.

Lemma 1. *Those values, $\vartheta_t(q)$, functions of q_{ij}, $i = 1, \cdots, s$, $j = 1, \cdots, \nu_i$, which minimize $\delta(p, q)$ are RBAN estimates of θ_t, $t = 1, \cdots, m$, if $\delta(p, q)$ is such that*

$$\left.\frac{\partial^2\delta(p, q)}{\partial q_{ij}\partial\theta_k}\right|_{q=f} = \left.\frac{\partial\psi_k}{\partial q_{ij}}\right|_{q=f} = CQ_i\frac{f_{ijk}}{f_{ij}}, \tag{3}$$

$$\left.\frac{\partial^2\delta(p, q)}{\partial\theta_t\partial\theta_k}\right|_{q=f} = \left.\frac{\partial\psi_k}{\partial\theta_t}\right|_{q=f} = -C\sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i}\frac{f_{ijt}f_{ijk}}{f_{ij}}, \tag{4}$$

where C is a constant.

Proof. It is shown in [1], (Theorem 2, page 248), that for a statistic, $\vartheta_1(q)$, function of $q_{11}, \cdots, q_{s\nu_s}$, to be a RBAN estimate of θ_1, it is sufficient that it satisfy the conditions

(a) that $\vartheta_1(q)$ have continuous partial derivatives with respect to all the independent variables, q_{ij},

(b) that the result of substituting $q_{ij} = f_{ij}(\theta_1, \cdots, \theta_m)$, $i = 1, \cdots, s$; $j = 1, \cdots, \nu_i$, in $\vartheta_1(q)$ leads to the identity

$$\vartheta_1(f) = \theta_1, \tag{5}$$

(c) that

$$\left.\frac{\partial \vartheta_1}{\partial q_{ij}}\right|_{q_{\alpha\beta}=f_{\alpha\beta}} = \frac{Q_i}{f_{ij}\Delta} \sum_{k=1}^{m} f_{ijk} \Delta_{1k}, \tag{6}$$

where

$$\Delta = \begin{vmatrix} G_{11} & \cdots & G_{1m} \\ \vdots & & \vdots \\ G_{m1} & \cdots & G_{mm} \end{vmatrix}, \tag{7}$$

$$G_{uv} = \sum_{i=1}^{s} Q_i \sum_{j=1}^{r_i} \frac{f_{iju} f_{ijv}}{f_{ij}} = G_{vu}, \tag{8}$$

and Δ_{1k} is the cofactor of G_{1k}.

That $\Delta \neq 0$ follows from the initial assumptions on the f_{ij}'s.

The proof of the lemma consists mainly in showing that condition (c) is satisfied when δ satisfies equations (3) and (4). Suppose the equations $\psi_k = 0$, $k = 1, \cdots, m$, have been solved for the θ's. (These solutions are the $\vartheta_t(q)$, $t = 1, \cdots, m$, which minimize $\delta(pq)$.) In order to find $\partial\vartheta_t/\partial q_{ij}$ one substitutes the ϑ_t's into the ψ_k's and differentiates with respect to q_{ij}. This results in equations of the form

$$\frac{\partial \psi_k}{\partial q_{ij}} + \sum_{t=1}^{m} \frac{\partial \psi_k}{\partial \vartheta_t} \frac{\partial \vartheta_t}{\partial q_{ij}} = 0, \qquad k = 1, \cdots, m \tag{9}$$

Solving for $(\partial\vartheta_t/\partial q_{ij})$, one gets

$$\left.\frac{\partial \vartheta_t}{\partial q_{ij}}\right|_{q=f} = \left.\frac{-\sum_{k=1}^{m} \frac{\partial \vartheta_k}{\partial q_{ij}} \Delta_{tk}}{\begin{vmatrix} \frac{\partial \psi_1}{\partial \theta_1} & \cdots & \frac{\partial \psi_1}{\partial \theta_m} \\ \vdots & & \vdots \\ \frac{\partial \psi_m}{\partial \theta_1} & \cdots & \frac{\partial \psi_m}{\partial \theta_m} \end{vmatrix}}\right|_{q=f} \tag{10}$$

If

$$\left.\frac{\partial \psi_k}{\partial q_{ij}}\right|_{q=f} = CQ_i \frac{f_{ijk}}{f_{ij}}, \qquad \left.\frac{\partial \psi_k}{\partial \theta_t}\right|_{q=f} = -CG_{tk},$$

then

$$\left.\frac{\partial \vartheta_t}{\partial q_{ij}}\right|_{q=f} = \frac{Q_i}{f_{ij}\Delta} \sum_{k=1}^{m} f_{ijk} \Delta_{tk}. \tag{11}$$

Conditions (a) and (b) are satisfied since $\delta(p, q)$ is assumed to be a distance function with continuous partial derivatives up to the second order.

3. A particular class of distance functions.

DEFINITION 2. *The symbol* $\mathfrak{F}$ *denotes a class of distance functions satisfying the conditions of the previous lemma; i.e.* $\delta(p, q)$ *is said to belong to* $\mathfrak{F}$ *if it is a distance function and if it satisfies the conditions* (*i*) *and* (*ii*) *as given by equations* (3) *and* (4) *above.*

THEOREM. *If* $h(x)$ *is a strictly monotonic function of* x *for* $0 < x < 1$ *possessing continuous derivatives up to the third order and if the function* $g(u, v)$ *is positive for* $0 < u < 1$, $0 < v < 1$, *has continuous partial derivatives up to the second order, and satisfies the condition*

$$g(f_{ij}, f_{ij}) = \frac{1}{f_{ij}}\left[\frac{dh(x)}{dx}\bigg|_{x=f_{ij}}\right]^{-2} \tag{12}$$

for all i, j, *then the function*

$$\delta_1(p, q) = \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} g(f_{ij}, q_{ij})[h(q_{ij}) - h(f_{ij})]^2 \tag{13}$$

belongs to class $\mathfrak{F}$.

In other words, this theorem asserts that the functions $\vartheta_t(q)$, $t = 1, \cdots, m$, which minimize $\delta_1(p, q)$ are RBAN estimates.

PROOF. $\delta_1(pq)$ is a distance function. This is apparent by inspection since $h(x)$ is strictly monotonic and continuous and possesses continuous derivatives. Also

$$\frac{\partial \delta_1}{\partial \theta_k} = \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i}\left[g(f_{ij}, q_{ij})(-2)(h(q_{ij}) - h(f_{ij}))\frac{dh}{dx}\bigg|_{x=f_{ij}} f_{ijk} + \frac{\partial g}{\partial \theta_k}(h(q_{ij}) - h(f_{ij}))^2\right], \tag{14}$$

$$\frac{\partial^2 \delta_1}{\partial q_{ij}\,\partial \theta_k}\bigg|_{q=f} = -2n_i\left[f_{ijk}\, g(f_{ij}, q_{ij})\frac{dh}{dx}\bigg|_{x=f_{ij}}\frac{dh}{dx}\bigg|_{x=q_{ij}}\right]\bigg|_{q=f} = \frac{-2}{N} Q_i \frac{f_{ijk}}{f_{ij}}, \tag{15}$$

$$\frac{\partial^2 \delta_1}{\partial \theta_t\,\partial \theta_k}\bigg|_{q=f} = 2\sum_{i=1}^{s}\sum_{j=1}^{\nu_i}\left[g(f_{ij}, q_{ij})\left(\frac{dh}{dx}\right)^2\bigg|_{x=f_{ij}} f_{ijk} f_{ijt}\right]\bigg|_{q=f} = \frac{2}{N}\sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i}\frac{f_{ijk} f_{ijt}}{f_{ij}}. \tag{16}$$

Thus $\delta_1(pq)$ is a member of class $\mathfrak{F}$.

COROLLARY 1. *If the number of exclusive results which can be produced by the* ith *trial is* 2, $i = 1, \cdots, s$, *then one can put*

$$f_{i1} = f_i, \qquad f_{i2} = 1 - f_i, \qquad q_{i1} = q_i, \qquad q_{i2} = 1 - q_i .$$

If in addition $h(x) = -h(1 - x)$, $0 < x < 1$, *then* $\delta_1(p, q)$ *reduces to the form* $\delta_1^*(p, q)$ *given by the equation*

$$\delta_1^*(p, q) = \sum_{i=1}^{s} n_i[g(f_i, q_i) - g(1 - f_i, 1 - q_i)][h(q_i) - h(f_i)]^2. \tag{17}$$

The implications of the theorem are that another class of estimates has been shown to consist of RBAN estimates. If computation is clumsy using other methods, perhaps the one which involves minimizing functions like $\delta_1(p, q)$ is easy. Berkson's short cut logit technique is one case of this. It, together with an example using probits, is given in the following paragraphs.

4. Logit estimates. Assume that a sequence of s independent experiments is performed. The jth experiment consists in giving dose x_j of some drug to n_j individuals. Each individual responds or fails to respond to the drug and the proportion responding, q_j, is observed. Let the doses be $x_1, x_2, \cdots, x_s$, the number of individuals tested $n_1, n_2, \cdots, n_s$ with $\sum_{j=1}^{s} n_j = N$, the proportions responding (independent random variables) $q_1, q_2, \cdots, q_s$, and the probabilities of responding $p_1, p_2, \cdots, p_s$.

It is assumed that for $j = 1, \cdots, s$

$$1 > p_j = f_j(\alpha, \beta) > 0. \tag{18}$$

The $f_j(\alpha, \beta)$ are assumed to be continuous and to have continuous partial derivatives up to the second order. Also the parameters α and β are assumed independent in the sense that for at least two values of j, say, i and k,

$$\begin{vmatrix} f_{i\alpha} & f_{i\beta} \\ f_{k\alpha} & f_{k\beta} \end{vmatrix} \neq 0, \tag{19}$$

where $f_{i\alpha} = \partial f_i/\partial\alpha$, $f_{i\beta} = \partial f_i/\partial\beta$.

The short cut logit technique of estimation as described by Berkson, [2], be gins with the assumption that the $f_j(\alpha, \beta)$ are given by the logistic function,

$$f_j(\alpha, \beta) = \frac{1}{1 + e^{-(\alpha+\beta x_j)}}, \qquad j = 1, \cdots, s. \tag{20}$$

Noting that log $[f_j/(1 - f_j)]$ equals the linear expression, $\alpha + \beta x_j$, Berkson suggests for estimates of α and β the functions $a(q)$ and $b(q)$, respectively, which minimize

$$\chi_A^2 = \sum_{j=1}^{s} n_j q_j(1 - q_j)\left(\log\frac{q_j}{1 - q_j} - \log\frac{f_j}{1 - f_j}\right)^2 \tag{21}$$

These estimates are very simple to find and techniques have been developed to facilitate computation, (see [2]). The question has been raised, however, as to whether $a(q)$ and $b(q)$ are RBAN estimates of α and β. Corollary 2 answers this.

COROLLARY 2. *Let p_i be assumed equal to $f_i(\alpha, \beta)$, $i = 1, \cdots, s$, where the f_i are any functions of α and β which are continuous with continuous partial derivatives up to the second order and which satisfy the conditions* (18) *and* (19). *Then the function*

$$\chi_A^2 = \sum_{i=1}^{s} n_i q_i(1 - q_i)\left(\log\frac{q_i}{1 - q_1} - \log\frac{f_i}{1 - f_i}\right)^2$$

is a member of class $\mathfrak{F}$.

It follows immediately from this corollary that if $f_i(\alpha, \beta) = 1/[1 + e^{-(\alpha+\beta x_i)}]$ the values, $a(q)$ and $b(q)$, which minimize χ_A^2 are RBAN estimates of α and β.

PROOF. Corollary 1 will be applied. Let $h(x) = \log[x/(1 - x)]$. This has all the properties of $h(x)$ in the theorem and in addition $h(x) = -h(1 - x)$. Also

$$\left.\frac{dh}{dx}\right|_{x=f_i} = \frac{1}{f_i(1 - f_i)}.$$

Obviously

$$q_i(1 - q_i) = \frac{q_i^2(1 - q_i)^2}{q_i} + \frac{(1 - q_i)^2 q_i^2}{1 - q_i}.$$

Also, $g(f_i, q_i) = q_i^2(1 - q_i)^2/q_i$ satisfies the conditions imposed by the theorem; that is, $g(u, v)$ is positive for $0 < u, v < 1$, it has continuous partial derivatives up to the second order and it satisfies the condition

$$g(f_i, f_i) = \frac{1}{f_i}\left(\left.\frac{dh}{dx}\right|_{x=f_i}\right)^{-2} = \frac{f_i^2(1 - f_i)^2}{f_i}. \tag{22}$$

Substituting in χ_A^2 one gets

$$\chi_A^2 = \sum_{i=1}^{s} n_i(g(f_i, q_i) - g(1 - f_i, 1 - q_i))(h(q_i) - h(f_i))^2,$$

which is the form of $\delta_1^*(p, q)$. Hence χ_A^2 is a member of class $\mathfrak{F}$. Note that in the above it is not necessary to write $g(f_i, q_i)$ as a function of two variables when one argument does not appear. It is done merely to be consistent with the notation of the theorem.

5. Probit estimates. Next a probit method will be taken up as another example of this distance function notion. Suppose there exists a situation which is the same as that in the preceding example, only instead of the logistic function for p_i it is assumed that p_i is given by the cumulative normal distribution function

$$\begin{aligned} p_i = f_i(\mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{k_i} e^{-\frac{1}{2}((r-\mu)/\sigma)^2}\,dr = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x_i-\mu}{\sigma}} e^{-\frac{1}{2}r^2}\,dr \\ &= \Phi\left(\frac{x_i - \mu}{\sigma}\right) \end{aligned} \tag{23}$$

say. A probit method of estimating μ and σ is outlined below. Define χ_B^2 as in (24), where Φ^{-1} is the inverse function of Φ, that is, Φ^{-1} is a function such that $\Phi^{-1}(\Phi(x)) = x$. Let

$$\chi_B^2 = \sum_{i=1}^{s} n_i G(f_i, q_i)(\Phi^{-1}(q_i) - \Phi^{-1}(p_i))^2. \tag{24}$$

$G(f_i, q_i)$ will be defined below by (29). Also

$$\frac{\partial \chi_B^2}{\partial \mu} = 2 \sum_{i=1}^{s} n_i G(f_i, q_i) \left(\Phi^{-1}(q_i) - \frac{x_i - \mu}{\sigma} \right) = 0, \tag{25}$$

$$\frac{\partial \chi_B^2}{\partial \sigma} = 2 \sum_{i=1}^{s} n_i G(f_i, q_i) \left(\Phi^{-1}(q_i) - \frac{x_i - \mu}{\sigma} \right)\left(\frac{x_i - \mu}{\sigma^2} \right) = 0. \tag{26}$$

Let these last two equations be simplified and written as

$$\sigma\left(\sum_{i=1}^{s} Q_i G(f_i, q_i)\Phi^{-1}(q_i)\right) + \mu\left(\sum_{i=1}^{s} Q_i G(f_i, q_i)\right) - \sum_{i=1}^{s} Q_i G(f_i, q_i)x_i = 0, \tag{27}$$

$$\begin{aligned} &\sigma\left(\sum_{i=1}^{s} Q_i G(f_i, q_i)\Phi^{-1}(q_i)x_i\right) \\ &\qquad + \mu\left(\sum Q_i G(f_i, q_i)x_i\right) - \sum_{i=1}^{s} Q_i G(f_i, q_i)x_i^2 = 0, \end{aligned} \tag{28}$$

and let $G(f_i, q_i)$ be defined by the equation

$$\begin{aligned} G(f_i, q_i) &= \frac{1}{q_i(1-q_i)}\left(\frac{d\Phi^{-1}(p_i)}{dp_i}\bigg|_{p_i=q_i}\right)^{-2} = \frac{1}{q_i(1-q_i)}\left[\frac{1}{\frac{d\Phi(k)}{dx}}\bigg|_{x=\Phi^{-1}(q_i)}\right]^{-2} \\ &= \frac{1}{q_i(1-q_i)}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\{\Phi^{-1}(q_i)\}^2}\right]^2. \end{aligned} \tag{29}$$

Since the coefficients of μ and σ in (27) and (28) are easily found, the solutions $\hat{\mu}(q)$ and $\hat{\sigma}(q)$ of (27) and (28) can be obtained. $\hat{\mu}$ and $\hat{\sigma}$ minimize χ_B^2.

Corollary 3. *Let χ_B^2 be given by*

$$\chi_B^2 = \sum_{i=1}^{s} \frac{n_i}{q_i(1-q_i)}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\{\Phi^{-1}(q_i)\}^2}\right]^2 (\Phi^{-1}(q_i) - \Phi^{-1}(f_i))^2, \tag{24}$$

and let it be assumed that

$$f_i(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x_i} e^{-\frac{1}{2}((r-\mu)/\sigma)^2}\, dr = \Phi\left(\frac{x_i - \mu}{\sigma}\right). \tag{23}$$

Then χ_B^2 is a member of class $\mathfrak{F}$ and it follows that $\hat{\mu}(q)$ and $\hat{\sigma}(q)$ which minimize χ_B^2 are RBAN estimates of μ and σ.

Proof. Again apply Corollary 1. $\Phi^{-1}(x)$ is seen to possess the properties $h(x)$ of the theorem and is such that $\Phi^{-1}(x) = -\Phi^{-1}(1 - x)$. As is seen immediately $G(f_i, q_i)$ can be written as

$$\begin{aligned} &\frac{1}{q_i}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\{\Phi^{-1}(q_i)\}^2}\right)^2 + \frac{1}{1-q_i}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\{\Phi^{-1}(1-q_i)\}^2}\right]^2 \\ &\qquad = g(f_i, q_i) + g(1 - f_i, 1 - q_i), \end{aligned} \tag{30}$$

where $g(u, v)$ satisfies the conditions of the theorem. It follows that χ_B^2 can be put in the same form as $\delta_1^*(p, q)$ and hence χ_B^2 is a member of class $\mathfrak{F}$.

A practice in bio-assay has been to get maximum likelihood estimates of μ and σ and by a somewhat lengthy iterative process. Corollary 3 shows, however, that if the limiting situation in which $N \to \infty$ with $Q_i = n_i/N$ held constant for all i is considered and if the asymptotic properties of the estimates are the criteria for the goodness of an estimate, then there is nothing to favor the maximum likelihood estimates over the simpler ones derived above.

REFERENCES

[1] 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.

[2] J. BERKSON, "Application of the logistic function to bioassay," *J. Amer. Stat. Assn.*, Vol. 39 (1944), p. 357.

[3] J. BERKSON, "Minimum χ^2 and maximum likelihood solution in terms of a linear transform, with particular reference to bio-assay." *J. Amer. Stat. Assn.*, Vol. 44 (1949), pp. 273–278.

[4] E. W. BARANKIN AND J. GURLAND, "On asymptotically normal, efficient estimators: I." *Univ. California Publ. Stat.*, Vol. 1, pp. 89–130, University of California Press, 1951.