CONTRIBUTION TO THE THEORY OF THE χ^2 TEST

J. NEYMAN

UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

In the present paper several alternative definitions of the familiar symbol χ^2 are discussed. The body of the paper is divided into two parts. In the first part (sec. 3) a class of estimates is defined, termed best asymptotically normal estimates (BAN estimates, for short), all having the same asymptotic properties as the maximum likelihood estimates but varying in the ease with which they can be computed. In the second part (sec. 4) a class of tests is developed which are all equivalent in the limit to λ-tests. Both the computation of BAN estimates and the application of the statistical tests considered involve the minimization of the alternatively defined χ^2's.

Some of the results given below were announced in 1940 [**8**].*

2. General conditions

The problems considered refer to the following situation. Consider s sequences of independent trials and let n_i denote the number of trials in the ith sequence. Each trial of the ith sequence is capable of producing one of the ν_i mutually exclusive results, say

$$R_{i,1}, R_{i,2}, \cdots, R_{i,\nu_i}, \tag{1}$$

with probabilities

$$p_{i,1}, p_{i,2}, \cdots, p_{i,\nu_i}, \tag{2}$$

where

$$\sum_{j=1}^{\nu_i} p_{i,j} = 1. \tag{3}$$

Denote by $n_{i,j}$ the number of occurrences of $R_{i,j}$ in the course of the n_i trials forming the ith sequence. Also let $q_{i,j} = n_{i,j}/n_i$. Finally let $N = n_1 + n_2 + \cdots + n_s$ and $Q_i = n_i/N$. The symbols $n_{i,j}$ and $q_{i,j}$ will be treated as random variables. The Q_i's will be considered as constants. N, the total number of observations, will be assumed to increase without limit.

The problems treated below arise when the values of the probabilities $p_{i,j}$ are unknown but it is given that each $p_{i,j}$ ($i = 1, 2, \cdots, s; j = 1, 2, \cdots, \nu_i$) is a specified function of several parameters $\theta_1, \theta_2, \cdots, \theta_k$. The reasoning which follows does not depend very much on the value of k, provided $k \geqq 2$. In order

* **Boldface numbers in brackets refer to references at the end of the paper (p. 273).**

to simplify the writing we shall assume $k = 2$. Thus we shall consider the situation where it is known that

$$p_{i,j} = f_{i,j}(\theta_1, \theta_2) > 0, \qquad i = 1, 2, \cdots, s; \quad j = 1, 2, \cdots, \nu_i, \tag{4}$$

with the functions $f_{i,j}$ satisfying s identities

$$\sum_{j=1}^{\nu_i} f_{i,j}(\theta_1, \theta_2) \equiv 1, \qquad i = 1, 2, \cdots, s. \tag{5}$$

It will be further assumed that the inequalities (4) and the identities (5) hold for the whole range of variation of θ_1 and θ_2. There will be no need to specify this range. The functions $f_{i,j}$ will be assumed to be continuous with respect to θ_1 and θ_2 and to possess continuous partial derivatives up to the second order. We shall use the notation

$$\frac{\partial f_{i,j}}{\partial \theta_k} = f_{i,j,k}, \qquad \frac{\partial^2 f_{i,j}}{\partial \theta_k \partial \theta_l} = f_{i,j,k,l}. \tag{6}$$

Then the identities (5) imply, for each $i = 1, 2, \cdots, s$,

$$\sum_{j=1}^{\nu_i} f_{i,j,k} \equiv \sum_{j=1}^{\nu_i} f_{i,j,k,l} \equiv 0. \tag{7}$$

The parameters θ_1 and θ_2 will be assumed independent, meaning that there exists at least one determinant

$$\begin{vmatrix} f_{i,j,1} & f_{i,j,2} \\ f_{\alpha,\beta,1} & f_{\alpha,\beta,2} \end{vmatrix} \neq 0. \tag{8}$$

In the situation considered, where the values of the $p_{i,j}$ are unknown, the actual values of θ_1 and θ_2 are also unknown. Let these actual values be $p_{i,j}^0$, θ_1^0, and θ_2^0.

3. Best asymptotically normal estimates

The BAN estimates are a generalization of the maximum likelihood estimates (ML estimates, for short). As proved by Hotelling [**5**] and, in a more general way, by Doob [**2**], the ML estimates, say $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ_1^0 and θ_2^0, are functions of the $q_{i,j}$, do not depend directly on N, and possess the following properties:

(i) $\hat{\theta}_i$ is a consistent estimate of θ_i^0. That is to say, as $N \to \infty$, the estimate $\hat{\theta}_i$ tends in probability to θ_i^0 or, in symbols,

$$\lim_{N \to \infty} p\, \hat{\theta}_i = \theta_i^0, \qquad i = 1, 2. \tag{9}$$

The equality (9) means that, whatever be $\epsilon, \eta > 0$, there exists a number $N_{\epsilon,\eta}$, such that $N > N_{\epsilon,\eta}$ implies

$$P\{|\hat{\theta}_i - \theta_i^0| > \epsilon\} < \eta. \tag{10}$$

(ii) As $N \to \infty$, the distribution of $\hat{\theta}_i$ tends to be normal, $N\left(\theta_i^0, \frac{\sigma_i}{\sqrt{N}}\right)$.

More specifically, whatever the real number t,

$$\lim_{N\to\infty} P\left\{\frac{(\hat{\theta}_i - \theta_i^0)\sqrt{N}}{\sigma_i} < t\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx = \Phi(t) \text{ (say)}, \tag{11}$$

where σ_i is a sure number, independent of N.

(iii) If ϑ is any other function satisfying (i) and (ii), but with σ taking the place of σ_i, then

$$\sigma \geqq \sigma_i. \tag{12}$$

To the writer's knowledge, these three important properties of the ML estimates form the only rational basis for preferring these estimates to others. It is obvious, however, that the ML estimates are not the only statistics which possess properties (i), (ii), and (iii). For example, if φ is any bounded function of the q's and $\hat{\theta}_i$ is the ML estimate of θ_i^0, then the function

$$\hat{\theta}_i + \frac{\varphi}{N} \tag{13}$$

will also possess properties (i) through (iii) and thus, as far as its asymptotic properties are concerned, be just as good an estimate of θ_i^0 as $\hat{\theta}_i$.

At this place it is convenient to notice that the ML estimates $\hat{\theta}_i$ also possess the following fourth property:

(iv) The ML estimates, $\hat{\theta}_i$ considered as functions of the $q_{i,j}$, possess continuous partial derivatives with respect to each $q_{i,j}$.

The foregoing circumstances, combined with the fact that the effective determination of the ML estimates is often very tedious, suggests the following two problems:

a) To determine the class of all estimates, to be denoted as BAN estimates, which satisfy conditions (i) through (iv), with the hope that some of them can be more easily computed than the ML estimates.

b) To investigate the class of BAN estimates and to see whether their distributions, corresponding to a fixed value of N, would make the use of some of them preferable to others.

The present paper deals with the first problem only.

DEFINITION. *A function ϑ of the random variables $q_{i,j}$ which does not depend directly on N is called a BAN estimate of the parameter θ_i if it satisfies the four conditions (i) through (iv) listed above.*

The search for BAN estimates will be preceded by several lemmas.

LEMMA 1. *If* $a_{i,j}$ $(i = 1, 2, \cdots, s; j = 1, 2, \cdots, \nu_i)$ *are any fixed numbers, then the variance* σ_X^2 *of the variable*

$$X = \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} a_{i,j} (q_{i,j} - p_{i,j}^0) \tag{14}$$

is given by

$$\sigma_X^2 = \sum_{i=1}^{s} \frac{\sigma_i^2}{n_i} = \frac{1}{N} \sum_{i=1}^{s} \sigma_i^2 Q_i^{-1}, \tag{15}$$

where

$$\sigma_i^2 = \sum_{j=1}^{\nu_i} (a_{i,j} - a_{i\cdot})^2 p_{i,j}^0 \tag{16}$$

and

$$a_{i\cdot} = \sum_{j=1}^{\nu_i} a_{i,j} p_{i,j}^0. \tag{17}$$

Lemma 1 is easily verified by direct computation.

LEMMA 2. *If* ϑ *is a BAN estimate of* θ_1, *then it can be presented in the form*

$$\vartheta = \theta_1^0 + X + Y_N \sigma_X, \tag{18}$$

where X *and* σ_X *have the structure of (14) and (15) with the* $a_{i,j}$ *depending on the* $p_{k,l}^0$*'s but being independent of* N *and the* $q_{k,l}$*'s, and where* Y_N *stands for a random variable which tends in probability to zero as* $N \to \infty$. *Moreover, it may always be assumed that, for* $i = 1, 2, \cdots, s$,

$$a_{i\cdot} = \sum_{j=1}^{\nu_i} a_{i,j} p_{i,j}^0 = 0, \tag{19}$$

so that

$$\sigma_X^2 = \frac{1}{N} \sum_{i=1}^{s} Q_i^{-1} \sum_{j=1}^{\nu_i} a_{i,j}^2 p_{i,j}^0 = \frac{\sigma_0^2}{N} \text{ (say)}. \tag{20}$$

PROOF. Let

$$\rho^2 = \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} (q_{i,j} - p_{i,j}^0)^2. \tag{21}$$

Since every BAN estimate has continuous partial derivatives with respect to every $q_{i,j}$, Taylor's formula gives

$$\vartheta = \vartheta_0 + \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} a_{i,j}(q_{i,j} - p_{i,j}^0) + R, \tag{22}$$

where

$$a_{i,j} = \left.\frac{\partial \vartheta}{\partial q_{i,j}}\right|_{q_{k,l} = p_{k,l}^0} \tag{23}$$

and where

$$\lim_{\rho \to 0} \frac{R}{\rho} = 0. \tag{24}$$

It is now necessary to show that the three terms on the right-hand side of (22) have the properties stated in lemma 2. We notice first that, should the derivatives (23) fail to satisfy equations (19), then it would be possible to subtract from (22) the expression

$$\sum_{i=1}^{s} a_{i\cdot} \sum_{j=1}^{\nu_i} (q_{i,j} - p_{i,j}^0) , \tag{25}$$

which is identically equal to zero. Then the BAN estimate would be put into a form analogous to (22),

$$\vartheta = \vartheta_0 + \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} (a_{i,j} - a_{i\cdot}) (q_{i,j} - p_{i,j}^0) + R , \tag{26}$$

with the coefficients $a_{i,j} - a_{i\cdot}$ necessarily satisfying condition (19).

In order to show that $\vartheta_0 = \theta_1^0$, it will be sufficient to show that the last two terms in the right-hand side of (22) both tend in probability to zero as $N \to \infty$. Denote by X_N the first of these terms. It is obvious that the expectation of X_N is equal to zero. According to lemma 1 and relation (19), the variance of X_N is given by (20) and tends to zero as N is indefinitely increased. Hence X_N tends in probability to zero as $N \to \infty$.

Instead of proving that the remainder R in (22) tends to zero in probability, we shall prove a stronger statement, namely, that

$$Y_N = \frac{R}{\sigma_X} = \sqrt{N}\frac{R}{\sigma_0} \tag{27}$$

tends in probability to zero. For this purpose we shall use the generalized theorem of Laplace (see David [**1**]), which implies that, as N tends to infinity, the distribution of products

$$(q_{i,j} - p_{i,j}^0)\sqrt{N} \tag{28}$$

tends to the multivariate normal law with expectations zero and with variances equal to $Q_i p_{i,j}^0(1 - p_{i,j}^0)$.

In particular, whatever be $\gamma > 0$, the probability that the variables $q_{i,j}$ will satisfy the relation

$$N\rho^2 = N\sum_{i=1}^{s} \sum_{j=1}^{\nu_i} (q_{i,j} - p_{i,j}^0)^2 < \gamma^2 \tag{29}$$

tends, uniformly for $\gamma > 0$, to a limit $F(\gamma)$ obtainable from the multivariate normal law. Thus, whatever be $\eta > 0$, there exists a number M_η, such that $N > M_\eta$ implies

$$P\{\rho\sqrt{N} < \gamma\} \geqq F(\gamma) - \frac{1}{2}\eta \tag{30}$$

for all values of γ.

In order to prove that Y_N tends to zero in probability, select two arbitrarily small positive numbers ϵ and η and show that there always exists a number $N_{\epsilon,\eta}$, such that $N > N_{\epsilon,\eta}$ implies

$$P\{|Y_N| < \epsilon\} = P\left\{\frac{|R|\sqrt{N}}{\sigma_0} < \epsilon\right\} > 1 - \eta . \tag{31}$$

Take the selected η and determine M_η such that $N > M_\eta$ implies (30) for all values of γ. Consider only values of N exceeding M_η. Next select γ_η so large that $F(\gamma_\eta)$ exceeds $1 - \frac{1}{2}\eta$. Then

$$P\{\rho\sqrt{N} < \gamma_\eta\} \geqq 1 - \eta . \tag{32}$$

The inequality in braces in the middle term of (31) may be written as follows:

$$\rho\sqrt{N}\frac{|R|}{\rho\sigma_0} < \epsilon . \tag{33}$$

It may be satisfied when the two factors on the left-hand side have various combinations of values, but in particular when simultaneously

$$\rho\sqrt{N} < \gamma_\eta \qquad \text{and} \qquad \frac{|R|}{\rho} < \frac{\sigma_0\epsilon}{\gamma_\eta}. \tag{34}$$

Thus the probability in the left-hand side of (31) cannot be less than the probability of the simultaneous fulfillment of the two inequalities (34),

$$P\{|Y_N| < \epsilon\} \geqq P\{\rho\sqrt{N} < \gamma_\eta\}\, P\left\{\frac{|R|}{\rho} < \frac{\sigma_0\epsilon}{\gamma_\eta}\,\middle|\, \rho\sqrt{N} < \gamma_\eta\right\}, \tag{35}$$

or, owing to (32),

$$P\{|Y_N| < \epsilon\} \geqq (1 - \eta)P\left\{\frac{|R|}{\rho} < \frac{\sigma_0\epsilon}{\gamma_\eta}\,\middle|\, \rho\sqrt{N} < \gamma_\eta\right\}. \tag{36}$$

Here the probability on the right-hand side of (36) is the conditional probability of the inequality on the left of the vertical bar, computed on the assumption that ρ fulfills the inequality on the right of the vertical bar. It will be seen that there exists a number $K_{\epsilon,\eta}$, such that $N > K_{\epsilon,\eta}$ implies

$$P\left\{\frac{|R|}{\rho} < \frac{\sigma_0\epsilon}{\gamma_\eta}\,\middle|\, \rho\sqrt{N} < \gamma_\eta\right\} = 1 . \tag{37}$$

Fix, for a moment, a value of $\rho > 0$ and denote by $\varphi(\rho)$ the corresponding maximum value of $|R|/\rho$. Condition (24) implies that

$$\lim_{\rho\to 0}\varphi(\rho) = 0 . \tag{38}$$

Denote by ρ_0 a number such that $\rho < \rho_0$ implies

$$0 \leqq \varphi(\rho) < \frac{\sigma_0 \epsilon}{\gamma_\eta}. \tag{39}$$

Then define

$$K_{\epsilon,\eta} = (\gamma_\eta/\rho_0)^2 \tag{40}$$

and consider the probability in (37) for values of $N > K_{\epsilon,\eta}$. The assumption underlying this probability implies that

$$\rho < \frac{\gamma_\eta}{\sqrt{N}} < \rho_0 . \tag{41}$$

This inequality implies (39) and, owing to the definition of $\varphi(\rho)$, it also implies that

$$\frac{|R|}{\rho} < \frac{\sigma_0 \epsilon}{\gamma_\eta}. \tag{42}$$

Thus equality (37) is proved, and it follows that if $N > \max(M_\eta, K_{\epsilon,\eta})$ then inequality (31) must be satisfied.

Thus the two last terms in (22) both tend in probability to zero, and therefore ϑ must tend in probability to ϑ_0. Since ϑ is assumed to be a BAN estimate of θ_1^0, it follows that $\vartheta_0 = \theta_1^0$, which completes the proof of lemma 2.

Consider two sequences of random variables $\{X_n\}$ and $\{Y_n\}$. Let $F_n(t)$ represent the probability law of X_n, that is to say,

$$F_n(t) = P\{X_n \leqq t\} . \tag{43}$$

LEMMA 3. *If as* $n \to \infty$ *the probability law* $F_n(t)$ *tends to a continuous probability law* $F(t)$ *and if the sequence* $\{Y_n\}$ *tends in probability to zero, then the probability law of the sum* $X_n + Y_n$ *tends uniformly to* $F(t)$,

$$\lim_{n\to\infty} P\{X_n + Y_n \leqq t\} = F(t) = \lim_{n\to\infty} P\{X_n \leqq t\} . \tag{44}$$

PROOF. In proving lemma 3 it will be convenient to use the following notation: If A and B stand for some inequalities which might be satisfied by some random variables, then the probability of the simultaneous fulfillment of both A and B will be denoted by $P\{(A)(B)\}$.

In order to prove lemma 3 it is necessary to show that whatever be $\epsilon > 0$ there exists a number N_ϵ, such that $n > N_\epsilon$ implies

$$|P\{X_n + Y_n \leqq t\} - F(t)| < \epsilon \tag{45}$$

for all values of t. We have

$$\begin{aligned}&|P\{X_n + Y_n \leqq t\} - F(t)| \\ &\quad \leqq |P\{X_n + Y_n \leqq t\} - P\{X_n \leqq t\}| + |P\{X_n \leqq t\} - F(t)| .\end{aligned} \tag{46}$$

Since it is assumed that $F(t)$ is continuous, a theorem of G. Pólya [**10**] implies that the convergence of $F_n(t)$ to $F(t)$ is uniform. Therefore, given any number $\eta > 0$, there exists a number N_1, such that $n > N_1$ implies

$$|P\{X_n \leqq t\} - F(t)| < \eta . \tag{47}$$

It follows that we need consider only the first term in the right-hand side of (46). Let $\Delta > 0$ be so small that

$$F(t + \Delta) - F(t - \Delta) < \eta . \tag{48}$$

We have

$$\begin{aligned} &|P\{X_n + Y_n \leqq t\} - P\{X_n \leqq t\}| \\ &\quad = |P\{(X_n > t)(Y_n \leqq t - X_n)\} - P\{(X_n \leqq t)(Y_n > t - X_n)\}| . \end{aligned} \tag{49}$$

Since the probabilities are non-negative, it follows that the left-hand side of (49) does not exceed the greater of the two probabilities in the right-hand side. It will be sufficient to show that one of them tends to zero as n is indefinitely increased. The proof relating to the other probability runs on exactly similar lines.

Consider then the probability

$$\begin{aligned} &P\{(X_n > t)(Y_n \leqq t - X_n)\} \\ &\quad = P\{(t < X_n \leqq t + \Delta)(Y_n \leqq t - X_n)\} + P\{(t + \Delta < X_n)(Y_n \leqq t - X_n)\} . \end{aligned} \tag{50}$$

Obviously

$$P\{(t + \Delta < X_n)(Y_n \leqq t - X_n)\} \leqq P\{Y_n \leqq -\Delta\} . \tag{51}$$

Since, by hypothesis, Y_n tends in probability to zero as $n \to \infty$, there exists a number N_2, such that $n > N_2$ implies

$$P\{Y_n \leqq -\Delta\} < \eta . \tag{52}$$

Thus, for $n > N_2$, the second term on the right-hand side of (50) will be less than η. Further

$$\begin{aligned} P\{(t < X_n \leqq t + \Delta)(Y_n \leqq t - X_n)\} &\leqq P\{t < X_n \leqq t + \Delta\} \\ &= F_n(t + \Delta) - F_n(t) , \end{aligned} \tag{53}$$

and it is seen that for $n > N_1$ the right-hand side of (53) must be smaller than 3η. It follows that if n exceeds both N_1 and N_2 then the probability in (50) is smaller than 4η. Since η is an arbitrarily small positive number, this completes the proof of lemma 3.

REMARK. It is believed that the results presented in the foregoing lemmas either are known or are easily obtainable from those published by other authors. However, the writer failed to find publications in which the necessary results are given in exactly the form needed here. For example, lemma 3 is obtainable from certain results published by Maurice Fréchet [**4**]. Also, it is similar to lemma 1 in the paper of J. Wolfowitz [**13**].*

Now let ϑ denote a function of $q_{1,1}, q_{1,2}, \cdots, q_{s,\nu_s}$. It will be assumed that ϑ has continuous partial derivatives with respect to each independent variable $q_{i,j}$. Then the proof of lemma 2 implies that

$$\vartheta = \vartheta_0 + X + Y_N \sigma_X , \tag{54}$$

where X and σ_X are defined by formulas (14) and (20) and where Y_N is a random variable which tends to zero in probability as $N \to \infty$.

LEMMA 4. *If* $N \to \infty$, *then the distribution of* $(\vartheta - \vartheta_0)/\sigma_X$ *tends to the normal with zero mean and unit variance,*

$$\lim_{N \to \infty} P\left\{\frac{\vartheta - \vartheta_0}{\sigma_X} < t\right\} = \Phi(t) . \tag{55}$$

PROOF. We have

$$\frac{\vartheta - \vartheta_0}{\sigma_X} = \frac{X}{\sigma_X} + Y_N . \tag{56}$$

The generalized theorem of Laplace implies that as $N \to \infty$ the probability law of X/σ_X tends to $\Phi(t)$. Since $\Phi(t)$ is continuous and since Y_N tends in probability to zero, lemma 3 implies lemma 4.

THEOREM 1. *For a statistic* $\vartheta(q)$, *function of* $q_{1,1}, q_{1,2}, \cdots, q_{s,\nu_s}$, *but independent of* N, *to be a BAN estimate of the parameter* θ_1, *it is necessary and sufficient:*

a) that ϑ *have continuous partial derivatives with respect to all the independent variables* $q_{i,j}$;

b) that the result of substituting

$$q_{i,j} = f_{i,j}(\theta_1, \theta_2) , \qquad i = 1, 2, \cdots, s; \ j = 1, 2, \cdots, \nu_i , \tag{57}$$

in $\vartheta(q)$ *leads to the identity*

$$\vartheta(f) \equiv \theta_1 ; \tag{58}$$

c) that if $\vartheta^*(q)$ *is any function satisfying* (*a*) *and* (*b*) *and if*

$$\left.\frac{\partial \vartheta}{\partial q_{i,j}}\right|_{q_{\alpha,\beta} = f_{\alpha,\beta}} = a_{i,j} \quad \text{and} \quad \left.\frac{\partial \vartheta^*}{\partial q_{i,j}}\right|_{q_{\alpha,\beta} = f_{\alpha,\beta}} = b_{i,j} , \tag{59}$$

then

$$\sum_{i=1}^{s} \frac{1}{Q_i} \sum_{j=1}^{\nu_i} a_{i,j}^2 f_{i,j} \leqq \sum_{i=1}^{s} \frac{1}{Q_i} \sum_{j=1}^{\nu_i} b_{i,j}^2 f_{i,j} \tag{60}$$

for all combinations of values of θ_1 *and* θ_2.

* After this paper was submitted for publication, an elegant proof of Lemma 3 was published by Harald Cramér in his remarkable book, "Mathematical Methods of Statistics," Princeton University Press, 1946.

Theorem 1 is a direct consequence of the preceding lemmas.

Up to the present, the number of parameters whose values determine the probabilities $p_{i,j}$ played no role. On the other hand, the statement of the following theorem does depend on that number and it will be expedient to assume that each function $f_{i,j}$ depends on m independent parameters $\theta_1, \theta_2, \cdots, \theta_m$,

$$p_{i,j} = f_{i,j}(\theta_1, \theta_2 \cdots, \theta_m) = f_{i,j}(\theta) \text{ (say)} . \tag{61}$$

As formerly, the independence of parameters will be understood to mean that for every system of their values there exists at least one determinant of mth order,

$$\begin{vmatrix} f_{\alpha_1,\beta_1,1} & f_{\alpha_1,\beta_1,2} & \cdots & f_{\alpha_1,\beta_1,m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_{\alpha_m,\beta_m,1} & f_{\alpha_m,\beta_m,2} & \cdots & f_{\alpha_m,\beta_m,m} \end{vmatrix} \neq 0 . \tag{62}$$

Let

$$G_{u,v} = \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{f_{i,j,u} f_{i,j,v}}{f_{i,j}} = G_{v,u} \tag{63}$$

for $u, v = 1, 2, \cdots, m$. The inequality (62) implies, then, that

$$\Delta = \begin{vmatrix} G_{1,1} & G_{1,2} & \cdots & G_{1,m} \\ G_{2,1} & G_{2,2} & \cdots & G_{2,m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ G_{m,1} & G_{m,2} & \cdots & G_{m,m} \end{vmatrix} > 0 . \tag{64}$$

Denote by $\Delta_{i,j}$ the minor of Δ corresponding to $G_{i,j}$.

THEOREM 2. *For a statistic* $\vartheta(\mathrm{q})$, *function of* $\mathrm{q}_{1,1}, \mathrm{q}_{1,2}, \cdots, \mathrm{q}_{s,\nu_s}$, *to be a BAN estimate of* θ_1, *it is sufficient that it satisfy the conditions* (*a*) *and* (*b*) *of theorem 1 and* (*d*) *that*

$$\left.\frac{\partial \vartheta}{\partial q_{i,j}}\right|_{q_{\alpha,\beta} = f_{\alpha,\beta}} = a_{i,j} = \frac{Q_i}{f_{i,j}\Delta} \sum_{k=1}^{m} f_{i,j,k} \Delta_{1,k} . \tag{65}$$

It will be seen that values (65) satisfy the condition

$$\sum_{j=1}^{\nu_i} a_{i,j} p_{i,j}^0 = 0 , \tag{66}$$

and that the asymptotic variance of ϑ is given by, say,

$$\sigma_e^2 = \frac{1}{N} \sum_{i=1}^{s} Q_i^{-1} \sum_{j=1}^{\nu_i} a_{i,j}^2 f_{i,j} = \frac{\Delta_{1,1}}{N\Delta} . \tag{67}$$

PROOF. If $\vartheta(q)$ is a function satisfying the conditions (a) and (b) of theorem 1, then by differentiating equation (58) with respect to the parameters θ_1, $\theta_2, \cdots, \theta_m$ a system of m linear equations is obtained:

$$\sum_{i=1}^{s} \sum_{j=1}^{\nu_i} a_{i,j} f_{i,j,1} = 1$$
$$\sum_{i=1}^{s} \sum_{j=1}^{\nu_i} a_{i,j} f_{i,j,k} = 0, \qquad k = 2, 3, \cdots, m, \tag{68}$$

which must be satisfied by the derivatives $a_{i,j}$. It is then easily found that the system of values of these derivatives minimizing the expression in the left-hand side of (60) under the sole restriction that equations (68) are satisfied is given by formula (65). This proves theorem 2.

REMARK. There is just one system of values of $a_{i,j}$ which minimize the left-hand side of (60) subject to conditions (68). However, the question of the necessity of condition (65) is still left in doubt, because it is not certain that the expressions on the right-hand side of (65) necessarily possess the property of being the partial derivatives of the same function taken with respect to the independent variables on which it depends. In other words, thus far it is uncertain whether there always must exist a function $\vartheta(q)$ whose derivatives satisfy equations (65). This question may be settled either by studying the properties of expressions (65) or by exhibiting functions $\vartheta(q)$ whose derivatives equal (65). Since the purpose of this study is to indicate several new forms of BAN estimates, the second of the two methods of proving the necessity of condition (65) will be followed.

The next three theorems indicate methods of obtaining three alternative BAN estimates. The proofs of these theorems are completely analogous and consist in verifying that a particular method determines a function $\vartheta(q)$ satisfying the conditions of theorem 2. For this reason the proof will be given for only one theorem, namely theorem 5.

The likelihood function of the parameters $\theta_1, \theta_2, \cdots, \theta_m$, given the observable random variables $q_{i,j}$, is given by

$$P = C \prod_{i=1}^{s} \left[\prod_{j=1}^{\nu_i} f_{i,j}^{q_{i,j}} \right]^{n_i}, \tag{69}$$

where C does not depend on $\theta_1, \theta_2, \cdots, \theta_m$. The ML estimates of the parameters θ are obtained by maximizing P or by maximizing its logarithm

$$L = \log C + N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} q_{i,j} \log f_{i,j}. \tag{70}$$

The equations determining the ML estimates are obtained by equating the derivatives of L to zero:

$$\psi_k = \frac{\partial L}{\partial \theta_k} = N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{q_{i,j}}{f_{i,j}} f_{i,j,k} = 0, \qquad k = 1, 2, \cdots, m. \tag{71}$$

THEOREM 3. *Under the general conditions described in section 2, the system of* k *equations (71) possesses solutions,* $\hat{\theta}_t(q)$, $(t = 1, 2, \cdots, m)$ *independent of* N, *which have the following properties:*

a) The substitution $q_{i,j} = f_{i,j}(\theta_1, \cdots, \theta_m)$ *gives, for* $t = 1, 2, \cdots, m$,

$$\hat{\theta}_t(f) \equiv \theta_t . \tag{72}$$

b) The functions $\theta_t(q)$ *possess continuous partial derivatives with respect to each variable* $q_{i,j}$.

c) The substitution of $q_{i,j} = f_{i,j}\ (\theta_1, \cdots, \theta_m)$ *into each partial derivative of* $\hat{\theta}_t$ *gives*

$$\left.\frac{\partial \theta_t}{\partial q_{i,j}}\right|_{q_{\alpha,\beta} = f_{\alpha,\beta}} = a_{i,j}^{(t)} , \tag{73}$$

where $a_{i,j}^{(t)}$ *is determined by formula (65) in which* $\Delta_{t,k}$ *should be substituted for* $\Delta_{1,k}$.

It follows that the solutions of (72) so obtained are BAN estimates of the θ'*s*.

Consider the familiar expression of the symbol χ^2 as defined by K. Pearson,

$$\chi^2 = \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} \frac{(n_{i,j} - n_i p_{i,j})^2}{n_i p_{i,j}} = N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{(q_{i,j} - p_{i,j})^2}{p_{i,j}} , \tag{74}$$

and substitute $p_{i,j} = f_{i,j}(\theta_1, \theta_2 \cdots, \theta_m)$. We have, say,

$$\chi^2(q, f) = N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{(q_{i,j} - f_{i,j})^2}{f_{i,j}} . \tag{75}$$

Keep the $q_{i,j}$ fixed and minimize this expression with respect to the unrestricted variation of the parameters θ. This leads to the solution of the system of equations, say

$$V_k = \frac{\partial \chi^2}{\partial \theta_k} = -N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \left(\frac{q_{i,j}}{f_{i,j}}\right)^2 f_{i,j,k} = 0 , \qquad k = 1, 2, \cdots, m . \tag{76}$$

THEOREM 4. *The system of equations (76) possesses a system of solutions, say* $\bar{\theta}_t(q)$, *which have the properties (a), (b), and (d) enumerated in theorem 2. It follows that the functions* $\bar{\theta}_t(q)$ *are BAN estimates of the parameters* θ_t.

Consider now an expression for χ^2 which is similar to but not identical with (74). The difference between the two expressions occurs in the denominators of the particular terms. The new expression is

$$\chi_1^2(q, f) = N \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{(q_{i,j} - f_{i,j})^2}{q_{i,j}} . \tag{77}$$

In writing this formula it is assumed that none of the $q_{i,j}$ is equal to zero.

The problem of minimizing χ_1^2 with respect to the unrestricted variation of $\theta_1, \cdots, \theta_m$, while the values of the $q_{i,j}$ are kept constant, leads to the solution of the following system of equations:

$$W_k = \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{f_{i,j}}{q_{i,j}} f_{i,j,k} = 0\,, \qquad k = 1, 2, \cdots, m\,. \tag{78}$$

THEOREM 5. *The system of equations (78) possesses a system of solutions, say $\theta^*_t(q)$, which have the properties (a), (b), and (d) enumerated in theorem 2. It follows that the functions $\theta^*_t(q)$ are BAN estimates of the parameters θ_t ($t = 1, 2, \cdots, m$).*

PROOF. In order to prove the existence of the solutions θ^*_t it is first necessary to show that, by substituting into (78) any system of values of the θ's and $q_{i,j} = f_{i,j}$ the equations (78) will be satisfied. This, however, is a direct consequence of the identities (7),

$$\sum_{j=1}^{\nu_i} f_{i,j,k} \equiv 0\,. \tag{79}$$

Next it is necessary to check whether the derivatives of the W_k with respect to the parameters θ and with respect to the $q_{i,j}$ are continuous in the vicinity of the point $q_{i,j} = f_{i,j}(\theta_1, \cdots, \theta_m)$. We have

$$\frac{\partial W_k}{\partial \theta_t} = \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \left. \frac{f_{i,j,t}f_{i,j,k} + f_{i,j}f_{i,j,k,t}}{q_{i,j}} \right|_{q_{i,j} = f_{i,j}} = G_{k,t}\,, \tag{80}$$

and, remembering that for each $i = 1, 2, \cdots, s$,

$$q_{i,\nu_i} = 1 - \sum_{j=1}^{\nu_i - 1} q_{i,j}\,, \tag{81}$$

$$\frac{\partial W_k}{\partial q_{i,j}} = -Q_i \left. \left(\frac{f_{i,j}}{q_{i,j}^2} f_{i,j,k} - \frac{f_{i,\nu_i}}{q_{i,\nu_i}^2} f_{i,\nu_i,k} \right) \right|_{q_{\alpha,\beta} = f_{\alpha,\beta}} = -Q_i \left(\frac{f_{i,j,k}}{f_{i,j}} - \frac{f_{i,\nu_i,k}}{f_{i,\nu_i}} \right) \tag{82}$$

for $j = 1, 2, \cdots, \nu_i - 1$. Owing to the assumption that within the region of variation of the parameters $\theta_1, \cdots, \theta_m$ all the functions $f_{i,j}$ have positive values, it is seen that the derivatives (80) and (82) are continuous in the vicinity of the point $q_{i,j} = f_{i,j}$.

Further, it is seen that the partial derivatives, say $\theta^*_{t,i,j}$, of the θ^*_t with respect to $q_{i,j}$ ($i = 1, 2, \cdots, s$; $j = 1, 2, \cdots, \nu_i - 1$), taken at the point $q_{i,j} = f_{i,j}$, are obtained from the system of equations

$$\sum_{t=1}^{m} G_{k,t} \theta^*_{t,i,j} = Q_i \left(\frac{f_{i,j,k}}{f_{i,j}} - \frac{f_{i,\nu_i,k}}{f_{i,\nu_i}} \right) \tag{83}$$

for $k = 1, 2, \cdots, m$. Comparing these equations with (65), it is found that

$$\theta^*_{t,i,j} = a_{i,j}^{(t)} - a_{i,\nu_i}^{(t)} . \tag{84}$$

Thus the first-order term of the Taylor expansion of θ^*_t in terms of the differences $q_{i,j} - f_{i,j}$, for $i = 1, 2, \cdots, s$ and $j = 1, 2, \cdots, \nu_i - 1$, must be, say,

$$X_t = \sum_{i=1}^{s} \sum_{j=1}^{\nu_i - 1} (a_{i,j}^{(t)} - a_{i,\nu_i}^{(t)}) (q_{i,j} - f_{i,j}) = \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} a_{i,j}^{(t)} (q_{i,j} - f_{i,j}) \tag{85}$$

because of the obvious equality

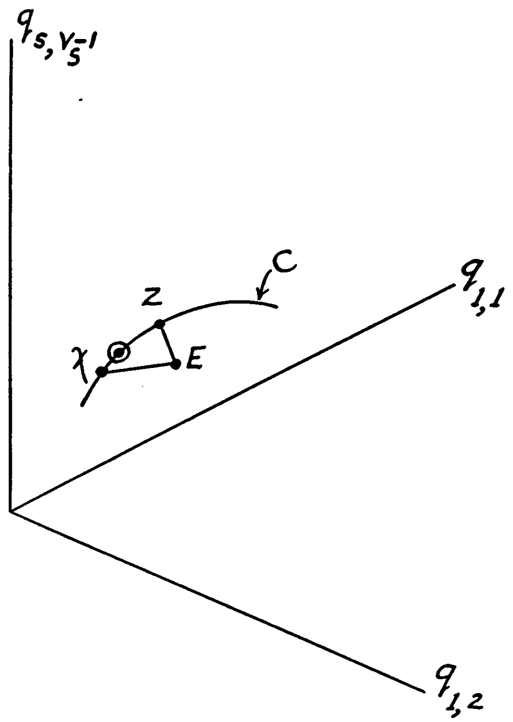
$$q_{i,\nu_i} - f_{i,\nu_i} = - \sum_{j=1}^{\nu_i - 1} (q_{i,j} - f_{i,j}) . \tag{86}$$

Thus equations (78) must possess solutions θ^*_t satisfying all the conditions of theorem 5.

REMARK 1. Since both expressions (75) and (77) are weighted sums of squares of deviations $(q_{i,j} - f_{i,j})$, the two methods of determining BAN estimates described in theorems 4 and 5 are, in effect, modifications of the least-square procedure, with weights selected in a particular way. In these circumstances it is interesting that the plain least-square method, or the weighted least-square method with arbitrarily selected weights, will always give consistent estimates of the θ's which, however, do not necessarily satisfy condition (d) of theorem 2 and, therefore, have a greater asymptotic variance than the BAN estimates.

REMARK 2. Theorems 3, 4, and 5, together with the preceding remark, have an interesting geometric interpretation. Consider the sample space of the $q_{i,j}$ $(i = 1, 2, \cdots, s; j = 1, 2, \cdots, \nu_i - 1)$ as illustrated in the accompanying figure. Equating $q_{i,j} = f_{i,j}(\theta_1, \theta_2, \cdots, \theta_m)$ and letting the θ's vary within appropriate limits, a locus of points (parameter points) is obtained represented by the curve C. The coördinates of points on this curve are the possible values of the probabilities $p_{i,j}$ and each such point determines a system of values of the θ's. The heavy point within a circle on C represents the "true" values of the θ's. The observations determine the frequencies $q_{i,j}$ represented in the figure by the point E (the "event point"). The procedure of ordinary unweighted least squares, applied to the determination of estimates of the true values of the θ's, reduces itself to finding, on the curve C, the particular point Z whose distance from E is the least. Here the word "distance" is understood in the ordinary way, that is to say, as the square root of the sum of the squares of differences between the coördinates of these two points.

The procedures indicated in theorems 4 and 5 can be interpreted similarly except that the conception of "distance" between two points is modified by the presence of weights depending either on the point, say χ, on the curve C (theorem 4) or on the event point E (theorem 5). As a result, the point χ on C minimizing the specially defined distance need not coincide with the point Z



on C, whose ordinary distance from E is a minimum. From this point of view, the maximum likelihood procedure is essentially similar to those just described, differing only in the definition of "distance." In order to see the analogy, it is sufficient to notice that the absolute maximum of the likelihood P, corresponding to the totally unrestricted variation of the $p_{i,j} \geqq 0$, subject only to the condition that

$$\sum_{j=1}^{\nu_i} p_{i,j} = 1 , \tag{87}$$

is attained when $p_{i,j} = q_{i,j}$. Let $A(q)$ be that maximum value of P. Now it will be seen that to maximize the likelihood function P with respect to some variation of the $p_{i,j}$'s means exactly the same as to minimize the logarithm of the ratio $A(q)/P$. The absolute minimum value of this logarithm is equal to zero and is attained when $p_{i,j} = q_{i,j}$. Thus, if the "distance" between the points with coördinates $p_{i,j}$ and another point with coördinates $q_{i,j}$ is defined to be the logarithm of the ratio $A(q)/P$, then the maximum likelihood procedure will fit the general description of the methods of determining the estimates of the parameters θ as follows: the search for estimates of the parameters θ reduces itself to the search for that point on the locus C whose generalized distance from the event point E is the least.

Theorems 3, 4, and 5 indicate that, with large values of N, the repeated observation of the event point E and the subsequent determination of the least "distant" point on C will determine the estimates of the θ's whose distribution about the true values depends on the definition "distance." In fact, the definitions of "distance" implied by theorems 3, 4, and 5 would (eventually, when N is increased) give tighter clustering of the estimates of the θ's than some other definitions, in particular tighter than the ordinary definition of distance between two points.

REMARK 3. The procedures for obtaining BAN estimates implied by theorems 3, 4, and 5 will vary in their difficulty according to the nature of the particular problem. The method of theorem 5 has particular advantages in all those problems where the functions $f_{i,j}$ are linearly connected with the parameters θ. In this special case the BAN estimates θ^* are determined by a system of linear equations. Theorem 6 extends this possibility to some other more complicated cases. The presentation of this theorem must be preceded by some introductory remarks.

It is obvious that the problem of estimating the parameters $\theta_1, \theta_2, \cdots, \theta_m$ is equivalent to that of estimating all the probabilities $p_{i,j}$. In the following, it will be convenient to use this form of the general problem. Accordingly, it will be considered, for each $i = 1, 2, \cdots, s$, that

$$p_{i,\nu_i} = 1 - \sum_{j=1}^{\nu_i - 1} p_{i,j}, \tag{88}$$

and, similarly, that

$$q_{i,\nu_i} = 1 - \sum_{j=1}^{\nu_i - 1} q_{i,j}. \tag{89}$$

The information that each $p_{i,j}$ is a known function $f_{i,j}$ of m independent parameters $\theta_1, \theta_2, \cdots, \theta_m$ is equivalent to the restriction on

$$\nu = \sum_{i=1}^{s} (\nu_i - 1) \tag{90}$$

independent variables $p_{i,j}$ imposed by means of $\mu = \nu - m$ equations of the form

$$F_t(p) = F_t(p_{1,1}, \cdots, p_{s,\nu_s - 1}) = 0, \qquad t = 1, 2, \cdots, \mu, \tag{91}$$

which are obtainable by eliminating the parameters θ from the equations $p_{i,j} = f_{i,j}$.

The formulation of theorem 6 does not require a distinction between the probabilities $p_{i,j}$ or the frequencies $q_{i,j}$ relating to particular sequences of trials. Therefore, to simplify the typographical work, the previous notation will be altered. Thus, instead of writing $p_{i,j}$ and $q_{i,j}$, the symbols p_k and q_k will be used for $k = 1, 2, \cdots, \nu$.

The assumptions on the functions $f_{i,j}$ imply that the functions F_t possess continuous partial derivatives up to the second order. Also the true values, say p_i^0, of the probabilities p_i satisfy equations (91). Finally, the independence of the parameters θ implies that, for each system of values of the p's satisfying (91), there exists at least one system of the p's such that the Jacobian

$$\frac{\partial(F_1, F_2, \cdots, F_m)}{\partial(p_{i_1}, p_{i_2}, \cdots, p_{i_m})} \neq 0. \tag{92}$$

In these circumstances, Taylor's formula may be applied to each function $F_t(p)$ to give its expansion about any point satisfying the conditions $p_{i,j} > 0$ and

$$\sum_{j=1}^{\nu_i - 1} p_{i,j} < 1, \qquad i = 1, 2, \cdots, s. \tag{93}$$

Taylor's formula will be applied to obtain the expansion of $F_t(p)$ about the point $p_k = q_k$ $(k = 1, 2, \cdots, \nu)$, which we shall denote by E. Thus

$$F_t(p) = F^*_t(q, p) + \frac{1}{2} \sum_{i=1}^{\nu} \sum_{j=1}^{\nu} C_{t,i,j} (p_i - q_i)(p_j - q_j), \tag{94}$$

where

$$F^*_t(q, p) = F_t(q) + \sum_{i=1}^{\nu} b_{t,i}(p_i - q_i). \tag{95}$$

Here $b_{t,i}$ represents the partial derivative of $F_t(p)$ with respect to p_i taken at the point E. Thus $b_{t,i}$ does not depend upon the p's, so that $F^*_t(q, p)$ is a linear function of the p's. On the other hand, the coefficients $C_{t,i,j}$ are functions of both the p's and the q's.

Denote generally by $\Delta(p, q)$ the generalized distance between the point with coördinates $p_1, p_2, \cdots, p_\nu$ and the point E. It will be assumed that the distance $\Delta(p, q)$ possesses the following two properties:

(i) $\Delta(p, q)$ possesses continuous partial derivatives of second order with respect to all the independent variables p_i, q_j $(i, j = 1, 2, \cdots, \nu)$.

(ii) $\Delta(p, q)$ possesses an absolute minimum of zero at the point E.

It will be noticed that the "distances" underlying the three methods of obtaining BAN estimates considered above possess the properties (i) and (ii).

Consider now two problems of minimizing the distance $\Delta(p, q)$. In the first, the minimization will be effected with respect to such variation of the p's as is consistent with the μ restrictions

$$F_t(p) = 0, \qquad t = 1, 2, \cdots, \mu. \tag{96}$$

In the second problem, the minimization will be effected with respect to such variation of the p's as is consistent with the μ restrictions of the form

$$F^*_t(p) = 0, \qquad t = 1, 2, \cdots, \mu \tag{97}$$

In either problem the method of Lagrange will be used, and it will be necessary to give some detail. Owing to the perfect analogy between the two, the details will be given only once in relation to restrictions (96). Corresponding equations referring to restrictions (97) will be obtained from those relative to (96) by adding asterisks to the appropriate symbols.

Let $a_1, a_2, \cdots, a_m$ be some constants. The method of minimizing $\Delta(p, q)$ with respect to the variation of the p's consistent with (96), while the q's are kept constant, consists of differentiating the function, say

$$\psi = \Delta(p, q) + \sum_{t=1}^{\mu} a_t F_t, \tag{98}$$

with respect to all the p's, as if they were entirely independent, and in setting the derivatives to zero. Thus ν equations are obtained,

$$\psi_k = \frac{\partial \Delta}{\partial p_k} + \sum_{t=1}^{\mu} a_t F_{t,k} = 0\,, \qquad k = 1, 2, \cdots, \nu\,, \tag{99}$$

where $F_{t,k}$ stands for the derivative of F_t with respect to p_k. Solving (99) for the p_i's, the solutions $\pi_i(q, a)$ are obtained, which obviously are functions of the q's and of the a's. We shall have to deal with the derivatives of the $\pi_i(q, a)$. These are obtained by substituting $\pi_i(q, a)$ into each equation (99) and by differentiating. Thus

$$\frac{\partial \psi_k}{\partial q_u} = \sum_{i=1}^{\nu} \left(\frac{\partial^2 \Delta}{\partial p_k \partial p_i} + \sum_{t=1}^{\mu} a_t \frac{\partial F_{t,k}}{\partial p_i} \right) \frac{\partial \pi_i}{\partial q_u} + \frac{\partial^2 \Delta}{\partial p_k \partial q_u} = 0 \tag{100}$$

for $k, u = 1, 2, \cdots, \nu$ and

$$\frac{\partial \psi_k}{\partial a_\tau} = \sum_{i=1}^{\nu} \left(\frac{\partial^2 \Delta}{\partial p_k \partial p_i} + \sum_{t=1}^{\mu} a_t \frac{\partial F_{t,k}}{\partial p_i} \right) \frac{\partial \pi_i}{\partial a_\tau} + F_{\tau,k} = 0 \tag{101}$$

for $k = 1, 2, \cdots, \nu$; $\tau = 1, 2, \cdots, \mu$. As long as the values of the a's are arbitrary, the functions $\pi_i\ (q, a)$ need not satisfy restrictions (96). In order to satisfy these restrictions, the functions $\pi_i(q, a)$ are substituted into each of the equations (96) instead of the p's and the equations obtained, say

$$F_t[\pi(q, a)] = 0\,, \qquad t = 1, 2, \cdots, \mu\,, \tag{102}$$

are considered as determining implicit functions $a_\tau(q)$ for $\tau = 1, 2, \cdots, \mu$. The derivatives of the $a_\tau(q)$ are obtained from the following linear equations:

$$\sum_{i=1}^{\nu} F_{t,i} \left(\frac{\partial \pi_i}{\partial q_u} + \sum_{\tau=1}^{\mu} \frac{\partial \pi_i}{\partial a_\tau} \frac{\partial a_\tau}{\partial q_u} \right) = 0 \tag{103}$$

for $t = 1, 2, \cdots, \mu$ and $u = 1, 2, \cdots, \nu$. The same equations (103) could be written as

$$\sum_{\tau=1}^{\mu} \frac{\partial a_\tau}{\partial q_u} \sum_{i=1}^{\nu} F_{t,i} \frac{\partial \pi_i}{\partial a_\tau} + \sum_{i=1}^{\nu} F_{t,i} \frac{\partial \pi_i}{\partial q_u} = 0\,. \tag{104}$$

Once the functions $a(q)$ are obtained, they are substituted into the functions $\pi_k(q, a)$ instead of the a's. In this way the functions

$$P_k(q) = \pi_k[q, a(q)]\,, \qquad k = 1, 2, \cdots, \nu\,, \tag{105}$$

are obtained. Among systems of solutions (105) there must be one system which for $q_k = p_k^0$ $(k = 1, 2, \cdots, \nu)$ reduces to $P_k = p_k^0$ $(k = 1, 2, \cdots, \nu)$. In fact, the numbers p_k^0 satisfy restrictions (96) and (97) and also ascribe the minimum value to $\Delta(p^0, p^0) = 0$. Just those solutions which reduce to p_k^0 at $q_k = p_k^0$ will be considered below and will be denoted by $P_k(q)$ and $P^*_k(q)$, respectively.

Notice that under the conditions considered the functions $P_i(q)$ and $P^*_i(q)$ possess continuous partial derivatives of the first order taken with respect to each independent variable q_i $(i = 1, 2, \cdots, \nu)$.

THEOREM 6. *The first-order partial derivatives of* $P_k(q)$ *and* $P^*_k(q)$, *taken at the point* $q_i = p_i^0$ $(i = 1, 2, \cdots, \nu)$, *with respect to the same variable* q_u $(u = 1, 2, \cdots, \nu)$ *have equal values,*

$$\left.\frac{\partial P_k}{\partial q_u}\right|_{q_i = p_i^0} = \left.\frac{\partial P^*_k}{\partial q_u}\right|_{q_i = p_i^0} = A_{k,u} \text{ (say)} . \tag{106}$$

Theorem 6 implies that the linear terms of the Taylor expansions of $P_k(q)$ and $P^*_k(q)$ made about the point $q_i = p_i^0$ $(i = 1, 2, \cdots, \nu)$ coincide,

$$\left.\begin{aligned} P_k(q) &= p_k^0 + \sum_{i=1}^{\nu} A_{k,i}(q_i - p_i) + R , \\ P^*_k(q) &= p_k^0 + \sum_{i=1}^{\nu} A_{k,i}(q_i - p_i) + R^* . \end{aligned}\right\} \tag{107}$$

It follows that if the generalized distance $\Delta(p, q)$ is such that its minimization under the restrictions (96) leads to BAN estimates of the p_k^0, then the minimization of the same distance $\Delta(p, q)$ under the simpler conditions (97) will also lead to the BAN estimates of the p_k^0.

PROOF. The derivatives of $P_k(q)$ and $P^*_k(q)$ with respect to any variable q_u are obtained as follows: Substitute in each equation (99) the function $P_k(q)$ instead of p_k and the function $a_t(q)$ instead of a_t. This substitution will result in the identity, say,

$$\Psi_k(q) \equiv 0 , \qquad k = 1, 2, \cdots, \nu . \tag{108}$$

The derivative of the function $\Psi_k(q)$ with respect to q_u is expressed in terms of derivatives of all the functions considered,

$$\frac{\partial \Psi_k(q)}{\partial q_u} = \sum_{i=1}^{\nu}\left(\frac{\partial^2 \Delta}{\partial p_k \partial p_i} + \sum_{t=1}^{\mu} a_t \frac{\partial F_{t,k}}{\partial p_i}\right)\frac{\partial P_i}{\partial q_u} + \frac{\partial^2 \Delta}{\partial p_k \partial q_u} + \sum_{t=1}^{\mu} \frac{\partial a_t}{\partial q_u} F_{t,k} \equiv 0 \tag{109}$$

for $k = 1, 2, \cdots, \nu$ and $u = 1, 2, \cdots, \nu$. Theorem 6 will be proved if it is shown that, at $q_i = p_i^0$ $(i = 1, 2, \cdots, \nu)$, the system of equations (109) coincides with the analogous one corresponding to $P^*_i(q)$. For this purpose, notice first that, at the point considered,

$$F_{t,k} = F^*_{t,k} = b_{t,k} . \tag{110}$$

Next it is easy to see that, at the same point,

$$a_t(p^0) = a^*_t(p^0) = 0 . \tag{111}$$

This follows from the assumption (ii) (p. 255) that $\Delta(p, q)$ possesses an absolute minimum at $p_k = q_k$. In fact, this assumption implies that at $q_k = p_k = p_k^0$ all the derivatives of $\Delta(p, q)$ with respect to the p_i's must be equal to zero. As a result, the system of equations (99) becomes homogeneous with respect to the a's and, owing to (92), the only solution is provided by (111). Returning to equations (109), it will be seen that to complete the proof of theorem 6 it is sufficient to show that, at the point $q_i = p_i^0$ $(i = 1, 2, \cdots, \nu)$,

$$\frac{\partial a_t}{\partial q_u} = \frac{\partial a^*_t}{\partial q_u} \tag{112}$$

for $t = 1, 2, \cdots, \mu$ and $u = 1, 2, \cdots, \nu$. The derivatives in the left-hand side of (112) are determined from equations (104); those on the right-hand side from a system similar to (104) corresponding to minimization of Δ under restrictions (97). In order to prove the identity of these two systems of equations at the point $q_u = p_u^0$ it is sufficient to show that at this point

$$\frac{\partial \pi_i}{\partial a_r} = \frac{\partial \pi^*_i}{\partial a_r} \tag{113}$$

and

$$\frac{\partial \pi_i}{\partial q_u} = \frac{\partial \pi^*_i}{\partial q_u} \tag{114}$$

for all combinations of values of the subscripts. But the derivatives in (113) and (114) are determined by equations of the type (100) and (101), and it is seen that at the point $q_u = p_u^0$ $(u = 1, 2, \cdots, \nu)$ the equations (113) and (114) must be satisfied. Thus the values of the derivatives $\partial P_k/\partial q_u$ and $\partial P^*_k/\partial q_u$ at the point $q_u = p_u^0$ $(u = 1, 2, \cdots, \nu)$ are obtained from identical systems of linear equations and the proof of theorem 6 is completed.

It will be seen that by combining theorem 5 with theorem 6 the search of the BAN estimates of the probabilities $p_{i,j}^0$ is reduced to the solution of a system of linear equations. If

$$\theta_i = \varphi_i(p) \tag{115}$$

represents the expression of the parameter θ_j in terms of some probabilities $p_{u,v}$, then a BAN estimate θ_j of θ_j^0 is obtained by substituting in (115), instead of $p_{u,v}$, the corresponding BAN estimate.

Whereas the BAN estimates of the θ's so obtained possess the asymptotic properties stated in the definition of the BAN estimates, the question remains open concerning how good these estimates are when the number of observations is only moderate.

Before concluding, it may be interesting to notice that the results given in this section seem to contradict the assertion of R. A. Fisher [**3**], not a very clear one, to the effect that "the maximum likelihood equation may indeed be derived from the conditions that it shall be linear in frequencies, and efficient for all values of θ."

4. Class of χ^2-tests equivalent in the limit to the λ-test

Referring to the situation described in section 2, it will be considered as known for certain that

$$p_{i,j} = f_{i,j}(\theta_1, \theta_2, \cdots, \theta_f, \theta_{f+1}, \cdots, \theta_m), \tag{116}$$

with the values of the θ's being unknown. The hypotheses ascribing to the θ's particular values within certain limits form the set Ω of admissible hypotheses. The problem considered in this section is that of tests of a hypothesis H, ascribing specific values to some $f \leqq m$ parameters

$$\theta_k = \dot{\theta}_k, \qquad k = 1, 2, \cdots, f. \tag{117}$$

The number f, equal to the number of parameters specified by the hypothesis tested out of all the parameters particularizing the admissible hypotheses, plays an important role in the theory given below. It will be useful to illustrate its significance.

Example 1. A random variable X is known to be able to assume non-negative integer values 0, 1, 2, $\cdots$. However, nothing is known about the probabilities

$$p_i = P\{X = i\}. \tag{118}$$

A total of n independent observations of the variable X is to be made, and n_i represents the number of those in which $X = i$. The hypothesis H_o to test is that X follows a Poisson law, so that

$$p_i = e^{-\lambda}\lambda^i/i! \tag{119}$$

Limiting the sequence of possible values of X to ν categories $X = 0$, $X = 1$, $\cdots$, $X = \nu - 2$ and $X \geqq \nu - 1$, it is easy to establish the relation of the situation described with the above theoretical one, as follows:

Since there is no a priori information concerning the probabilities $P\{X=m\}$, it will have to be assumed that the set, say Ω_0, of admissible hypotheses contains every hypothesis ascribing some specified values to the probabilities

$$p_m = P\{X = m\} > 0, \qquad m = 0, 1, \cdots, \nu - 2, \tag{120}$$

and to

$$p_{\nu-1} = 1 - \sum_{m=0}^{\nu-2} p_m > 0. \tag{121}$$

Referring to formula (116), it will be seen that in the present case the probabilities (120) play the role of independent parameters θ, whose number is $\nu - 1$.

The hypothesis tested H_0, expressed by means of (119), contains just one parameter λ whose value is left unspecified. In other words, out of the $\nu - 1$

parameters involved in Ω_0, the hypothesis H_0 specifies $\nu - 2$. Thus, in this particular case, $f = \nu - 2$. In order to bring the situation in this example into complete correspondence with (116) and (117), it will be sufficient to rewrite (120) as follows:

$$p_0 = e^{-\lambda}, \tag{122}$$

$$p_m = e^{-\lambda} \frac{\lambda^m}{m!} \theta_1 \theta_2 \cdots \theta_m, \qquad m = 1, 2, \cdots, \nu - 2, \tag{123}$$

where λ and the θ's are arbitrary positive parameters subject to the restriction that $p_{\nu-1}$ of formula (121) is a positive number.

The hypothesis tested H_0 specifies, then, that

$$\theta_k = \dot{\theta}_k = 1, \qquad k = 1, 2, \cdots, \nu - 2. \tag{124}$$

The test considered in this example may be described as the test of the Poisson law against the unrestricted set of alternatives.

Example 2. Consider the random variable X as in example 1 but assume as given that X follows a contagious distribution of type A with two parameters (see [7]) whose probability generating function is

$$G(u) = e^{-\theta_1 \frac{1 - e^{-\theta_2(1-u)}}{\theta_2}}. \tag{125}$$

It follows that in this case the set of admissible simple hypotheses, say Ω_1, is more restricted than the set Ω_0 of example 1. In fact, each of the simple hypotheses belonging to Ω_1 specifies the probabilities corresponding to (125) with some particular values of the parameters $\theta_1 > 1$ and $\theta_2 \geqq 0$. If $\theta_2 = 0$, then (125) reduces to the probability generating function of the Poisson law.

Consider now the same hypothesis tested as in example 1, namely, that X follows a Poisson law. Out of the two parameters θ_1 and θ_2 characterizing the admissible simple hypotheses, the hypothesis tested specifies one $\dot{\theta}_2 = 0$. Hence in this case $f = 1$.

In relation to the χ^2-tests discussed below, the number f is called the number of degrees of freedom. Thus, in each case, the number of degrees of freedom is the difference between the total number of independent parameters which specify the particular simple hypotheses considered as admissible and the number of parameters which are left unspecified by the hypothesis tested.

Referring again to the situation described in section 2, it will be necessary to explain the conception of a test of a single hypothesis H relating to the case where the number of observations N indefinitely increases.

Imagine a rule R_1 which associates with every value of N a region, say w_N, in the $\nu = \Sigma\nu_i - s$ dimensioned space W of the q's. This region w_N will be called the Nth critical region. Imagine further that in a particular case some N' observations are to be made and that there is another rule, R_2, to reject H

whenever the event point, $E_{N'}$, falls within $w_{N'}$ and not to reject H in other cases. The combination of the two rules R_1 and R_2 will be regarded as a test of the hypothesis H.

Consider now two different tests, T_1 and T_2, associated with sequences of critical regions $\{w'_N\}$ and $\{w''_N\}$ respectively.

DEFINITION. *If, whatever be the admissible simple hypothesis* h, *the probability of the two tests* T₁ *and* T₂ *contradicting each other tends to zero as* N *is indefinitely increased, then the tests* T₁ *and* T₂ *are called equivalent in the limit or asymptotically equivalent.*

The tests T_1 and T_2 will contradict each other either when E_N falls within w'_N but outside of w''_N or when E_N falls within w''_N but outside of w'_N. Denoting by ab the common part of any two regions a and b, the probability of contradiction of the two tests can be written as follows:

$$P\{E_N \epsilon (w'_N - w'_N w''_N) | h\} + P\{E_N \epsilon (w''_N - w'_N w''_N) | h\} . \qquad (126)$$

It follows that, for T_1 and T_2 to be equivalent in the limit, the expression (126) must tend to zero as $N \to \infty$, for every $h \epsilon \Omega$.

It is important to notice that, if T_1 and T_2 are both consistent so that, whatever be the simple hypothesis h' inconsistent with the hypothesis tested H,

$$\lim_{N \to \infty} P\{E_N \epsilon w_N | h'\} = \lim_{N \to \infty} P\{E_N \epsilon w'_N | h'\} = 1 , \qquad (127)$$

the equivalence of T_1 and T_2 in the limit will depend only on the properties of the critical regions w_N and w'_N with respect to the hypotheses tested. In other words, to prove the asymptotic equivalence of T_1 and T_2, it would be sufficient to show that

$$\lim_{N \to \infty} P\{E_N \epsilon (w_N - w_N w'_N) | H\} + P\{E_N \epsilon (w'_N - w_N w'_N | H\} = 0 . \quad (128)$$

Let $\hat{p}_{i,j}$ denote the ML estimate of $p_{i,j}$ computed without any reference to the hypothesis tested H, and let $\hat{\hat{p}}_{i,j}$ stand for the ML estimate of the same probability $p_{i,j}$ computed on the assumption that H is true. In other words, $\hat{\hat{p}}_{i,j}$ is a function of the q's obtained by maximizing the likelihood function with respect to the unrestricted variation of $\theta_{f+1}, \theta_{f+2}, \cdots, \theta_m$, while the q's are kept constant and the other parameters are ascribed the values $\dot{\theta}_1, \dot{\theta}_2, \cdots, \dot{\theta}_f$ specified by H. If it happens that H is true, then $\hat{\hat{p}}_{i,j}$ will have the properties of a BAN estimate of $p_{i,j}^0$, but not necessarily so if H is wrong.

Further, let χ_ϵ be the root of the equation

$$\int_{\chi_\epsilon}^{\infty} \chi^{f-1} e^{-\frac{1}{2}\chi^2} d\chi = \epsilon 2^{\frac{1}{2}f-1} \Gamma(\tfrac{1}{2}f) , \qquad (129)$$

where ϵ is the chosen level of significance. In other words, χ_ϵ^2 is the tabled value of the χ^2 with f degrees of freedom, corresponding to the level of significance ϵ.

Then the λ-test (see [**9**]), of the hypothesis H consists in the rule of rejecting H when the expression, say,

$$\lambda_N = \left\{ \prod_{i=1}^{s} \left[\prod_{j=1}^{\nu_i} \left(\frac{\hat{\hat{p}}_{i,j}}{\hat{p}_{i,j}} \right)^{q_{i,j}} \right]^{Q_i} \right\}^{N} \leqq e^{-\frac{1}{2}\chi_\epsilon^2} \tag{130}$$

and in not rejecting it otherwise. It will be noticed that inequality (130) defines the sequence of critical regions $\{w_N\}$ associated with the λ-test.

In the following, it will be shown that certain χ^2-tests are equivalent in the limit to the λ-test of the hypothesis H just described. The definition of the χ^2 considered involves BAN estimates of the parameters θ or of the probabilities $p_{i,j}^0$. A certain point in the proof requires the discussion of the method by which a BAN estimate is determined. Therefore the following theorem 7 relates only to BAN estimates discussed in this paper, i.e., to ML estimates and to those mentioned in theorems 4, 5, and 6. As was the case with respect to the λ-test, we shall need both the BAN estimates of the probabilities $p_{i,j}^0$ computed without reference to the hypothesis tested H and those computed on the assumption that H is true. The former will be denoted by $p_{i,j}(\Omega)$, and the latter by $p_{i,j}(H)$. This notation, then, applies to the four categories of BAN estimates mentioned above.

THEOREM 7. *The λ-test of the hypothesis* H *is equivalent in the limit to any one of the following tests*:

Test T$_1$: *Reject* H *whenever, say,*

$$\chi_b^2 = \chi_r^2 - \chi_a^2$$

$$= \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{[q_{i,j} - p_{i,j}(H)]^2}{p_{i,j}(H)} - \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{[q_{i,j} - p_{i,j}(\Omega)]^2}{p_{i,j}(\Omega)} \geqq \chi_\epsilon^2 . \tag{131}$$

Test T$_2$: *Reject* H *whenever, say,*

$$\chi_b^2(*) = \chi_r^2(*) - \chi_a^2(*)$$

$$= \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{[q_{i,j} - p_{i,j}(H)]^2}{q_{i,j}} - \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{[q_{i,j} - p_{i,j}(\Omega)]^2}{q_{i,j}} \geqq \chi_\epsilon^2 . \tag{132}$$

Denote by u_N and v_N respectively the critical regions defined by (131) and (132). For any region w, let $\bar{w} = W - w$ denote its complementary region.

The proof of theorem 7 is based on the generalized theorem of Laplace.

In order to formulate this theorem in full detail, consider a ν dimensioned space W_x of continuous variables $x_{i,j}$ $(i = 1, 2, \cdots, s; j = 1, 2, \cdots, \nu_i)$ subject to the restrictions

$$\sum_{j=1}^{\nu_i} x_{i,j} \sqrt{p_{i,j}^0} = 0 , \qquad i = 1, 2, \cdots, s . \tag{133}$$

For every fixed N we shall consider a correspondence between points in W_x and those in the original space W of the q's as determined by the formulas

$$q_{i,j} = p_{i,j}^0 + x_{i,j}\sqrt{\frac{p_{i,j}^0}{NQ_i}}. \tag{134}$$

Here $q_{i,j}$ is considered merely as a function of $x_{i,j}$ and need not have only rational values with the denominators equal to n_i. Whatever be the open set τ in W_x, formula (134) will determine the corresponding set, say τ_N in W.

The generalized theorem of Laplace states, then, that for every $\eta > 0$ and for every open set τ in W_x, there exists a number $N_{\eta,\tau}$ such that the inequality $N > N_{\eta,\tau}$ implies

$$|P\{E_N \epsilon \tau_N\} - I(\tau)| < \frac{1}{2}\eta, \tag{135}$$

with

$$I(\tau) = c\int \cdots \int_\tau e^{-\frac{1}{2}\sum_i\sum_j x_{i,j}^2} dW_x, \tag{136}$$

where c is a constant so selected that $I(W_x) = 1$.

Denote by $\tau(\eta)$ the region in W_x defined by the inequality

$$\sum_{i=1}^{s}\sum_{j=1}^{\nu_i} x_{i,j}^2 < R^2, \tag{137}$$

where R is a constant satisfying the equation

$$I[\tau(\eta)] = 1 - \frac{1}{2}\eta. \tag{138}$$

Then, for $N > N_\eta$,

$$P\{E_N \epsilon \tau_N(\eta)\} > 1 - \eta. \tag{139}$$

In order to prove the first part of theorem 7 it will be sufficient to show that, for N sufficiently large,

$$P\{E_N \epsilon \tau_N(\eta)[w_N - w_N u_N] | h\} < \eta \tag{140}$$

and

$$P\{E_N \epsilon \tau_N(\eta)[u_N - w_N u_N] | h\} < \eta \tag{141}$$

for every $h \epsilon \Omega$. In fact,

$$w_N - w_N u_N = \tau_N(\eta)[w_N - w_N u_N] + \bar{\tau}_N(\eta)[w_N - w_N u_N], \tag{142}$$

and the probability of E_N falling within the region represented by the second term in the right-hand side of (142), being at most equal to that of E_N falling within $\bar{\tau}_N(\eta)$, must be less than η. Similar argument applies to (141).

The convenience of dealing with products like $\tau_N(\eta) \cdot w_N$, etc., consists in that the $x_{i,j}$'s are bounded within $\tau(\eta)$,

$$|x_{i,j}| \leqq R. \tag{143}$$

Thus, using (134),

$$|q_{i,j} - p_{i,j}^0| < R\sqrt{\frac{p_{i,j}^0}{NQ_i}} \tag{144}$$

within all the region $\tau_N(\eta)$.

Proof of theorem 7 requires further the following known facts from the theory of normal variables. Let us use again the letters $x_1, x_2, \cdots, x_n$ to denote some normal random variables [having no relation to variables $x_{i,j}$ of (134)] independent or correlated, none being a function of the others, and having variances equal to unity. Let the probability law of the x's be

$$p(x_1, \cdots, x_n) = ce^{-\frac{1}{2}Q(x)}, \tag{145}$$

where $Q(x)$ is a positive definite form

$$Q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{i,j}(x_i - \xi_i)(x_j - \xi_j), \tag{146}$$

with the ξ_i denoting some real numbers.

(i) There exists a system of n random variables $y_1, y_2, \cdots, y_n$ linearly connected with the x's so that

$$x_i = \sum_{j=1}^{n} a_{i,j}y_i, \qquad i = 1, 2, \cdots, n, \tag{147}$$

and a system of constants $\eta_1, \eta_2, \cdots, \eta_n$ such that the substitution of (147) in $Q(x)$ gives, say,

$$Q[x(y)] = \sum_{i=1}^{n} (y_i - \eta_i)^2, \tag{148}$$

which implies that the probability law of the y's is represented by the function

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\eta_i)^2}. \tag{149}$$

(ii) Under the general conditions of (i), consider n linear combinations of some s independent parameters $\theta_1, \theta_2, \cdots, \theta_s$,

$$\xi'_i = \sum_{j=1}^{s} b_{i,j}\theta_j, \qquad i = 1, 2, \cdots, n, \tag{150}$$

such that for a certain system of values $\theta_1^0, \theta_2^0, \cdots, \theta_s^0$ of the parameters each ξ'_i coincides with ξ_i. Consider also the form, say,

$$Q'(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{i,j}(x_i - \xi'_i)(x_j - \xi'_j) . \tag{151}$$

The transformation (147) applied to (151) will give, say,

$$Q''(y) = \sum_{i=1}^{n} (y_i - \eta'_i)^2 , \tag{152}$$

where η'_i stands for a linear combination of the same parameters $\theta_1, \theta_2, \cdots, \theta_s$, say

$$\eta'_i = \sum_{j=1}^{s} c_{i,j}\theta_j , \tag{153}$$

such that for $\theta_j = \theta_i^0$ $(i = 1, 2, \cdots s)$ each η'_i has the value η_i.

Then there exists a system of n random variables $z_1, z_2, \cdots, z_n$ linearly connected with the y's,

$$y_i = \sum_{j=1}^{n} \beta_{i,j}z_j , \qquad i = 1, 2, \cdots, n , \tag{154}$$

(and therefore with the x's) such that the substitution of (154) into (152) gives, say,

$$Q''[y(z)] = \sum_{i=1}^{n-s} z_i^2 + \sum_{j=1}^{s} \left(z_{n-s+j} - \sum_{k=1}^{j} d_{j,k}\theta_k \right)^2 , \tag{155}$$

with the $d_{j,k}$ denoting constant numbers and

$$d_{j,j} \neq 0 , \qquad j = 1, 2, \cdots, s . \tag{156}$$

A direct proof of the existence of the transformations described is given by P. C. Tang [11]. The distribution of the variables z_i is given by the function

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left\{\sum_{i=1}^{n-s} z_i^2 + \sum_{j=1}^{s}\left(z_{n-s+j} - \sum_{k=1}^{s} d_{j,k}\theta_k^0\right)^2\right\}} . \tag{157}$$

(iii) Denote by Q_a the minimum value of $Q'(x)$ of (151) computed for an unrestricted variation of the parameters $\theta_1, \cdots, \theta_s$ while keeping the random variables $x_1, \cdots, x_n$ fixed. Q_a will be termed the absolute minimum of $Q'(x)$. Obviously it is a function of the x's. Applying to $Q'(x)$ in turn the two transformations (147) and (154), it will be seen from (155) that, for any fixed system of values of the x's, say $x'_1, x'_2, \cdots, x'_n$,

$$Q_a = \sum_{i=1}^{n-s} z'^2_i , \tag{158}$$

where the z' stand for the corresponding values of the z's.

It follows from (157) that the z's are mutually independent normal variables, all having unit variances and, for $i = 1, 2, \cdots, n - s$, having their means equal to zero. Thus the random variable Q_a is distributed as a central χ^2 with $n - s$ degrees of freedom.

(iv) Let f be a positive integer less than s and let $\dot{\theta}_1, \dot{\theta}_2, \cdots, \dot{\theta}_f$ be any real numbers. Assign some fixed values to $x_1, x_2, \cdots, x_n$, let $\theta_i = \dot{\theta}_i$ for $i = 1, 2, \cdots, f$, and vary $\theta_{f+1}, \cdots, \theta_s$. Under these conditions, let Q_r stand for the minimum value of $Q'(x)$. This minimum will be termed the relative minimum. Obviously $Q_r \geqq Q_a$. Let $Q_b = Q_r - Q_a$. Referring again to (155), it will be seen that, if $x'_1, \cdots, x'_n$ and $z'_1, \cdots, z'_n$ are the values of the x's and the z's which correspond to each other through (147) and (154), then

$$Q_b = \sum_{j=1}^{f} (z'_{n-s+j} - \sum_{k=1} d_{j,k}\dot{\theta}_k)^2 . \tag{159}$$

Thus Q_b considered as a random variable may be presented as a sum of squares of mutually independent normal variables all having variances equal to unity. It follows that the distribution of Q_b is that of a χ^2 with f degrees of freedom. If it happens that $\dot{\theta}_k = \theta_k^0$ for $k = 1, 2, \cdots, f$, then Q_b is distributed as a central χ^2, otherwise as a non-central.

We shall now prove several easy lemmas.

Consider the Taylor expansion of the functions $f_{i,j}$ about the "true" parameter point $\theta_k = \theta_k^0$ $(k = 1, 2, \cdots, m)$,

$$p_{i,j} = f_{i,j} = \varphi_{i,j} + R_{i,j} , \tag{160}$$

with

$$\varphi_{i,j} = p_{i,j}^0 + \sum_{k=1}^{m} f_{i,j,k}^0 (\theta_k - \theta_k^0) , \tag{161}$$

where $f_{i,j,k}^0$ stands for the value of $f_{i,j,k}$ at the true parameter point. Write

$$\chi_0^2 = \sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{(q_{i,j} - \varphi_{i,j})^2}{p_{i,j}^0} , \tag{162}$$

and denote by $\bar{p}_{i,j}$ the value of $\varphi_{i,j}$ minimizing χ_0^2 with respect to the unrestricted variation of the θ's. Also let Q_a stand for the minimum value of χ_0^2 in (162).

LEMMA 5. *Whatever be the BAN estimate* $p_{i,j}(\Omega)$ *of the four categories considered, at each point of the region* $\tau_N(\eta)$

$$p_{i,j}(\Omega) = \bar{p}_{i,j} + o\left(\frac{1}{\sqrt{N}}\right). \tag{163}$$

Lemma 5 is proved by following up the process of determining $p_{i,j}(\Omega)$ and by noticing that the values of the derivatives

$$\frac{\partial p_{i,j}(\Omega)}{\partial q_{\alpha\beta}} , \tag{164}$$

taken at the point $q_{u,v} = p_{u,v}^0$ ($u = 1, 2, \cdots, s; v = 1, 2, \cdots, \nu_u$) are determined by the numbers $f_{i,j,k}^0$. Thus the Taylor expansion of $p_{i,j}(\Omega)$ about the particular point will differ from $\bar{p}_{i,j}$ only by terms in differences $(q_{u,v} - p_{u,v}^0)$ of order of magnitude higher than the first. However, within $\tau_N(\eta)$ the differences $q_{u,v} - p_{u,v}^0$ are all of order $O\left(\frac{1}{\sqrt{N}}\right)$, which proves lemma 5.

LEMMA 6. *Within the region* $\tau_N(\eta)$ *the values of* χ_a^2 *and of* $\chi_a^2(*)$ *differ from* Q_a *no more than by terms of order* $O\left(\frac{1}{\sqrt{N}}\right)$. *Thus there exists a number* C_1 *such that within* $\tau_N(\eta)$

$$|\chi_a^2 - Q_a| < \frac{C_1}{\sqrt{N}} \qquad \text{and} \qquad |\chi_a^2(*) - Q_a| < \frac{C_1}{\sqrt{N}}. \tag{165}$$

Simple algebra shows that lemma 6 is a consequence of lemma 5.

LEMMA 7. *As* N *is indefinitely increased, the differences*

$$\Delta = \chi_a^2 - Q_a \qquad \text{and} \qquad \Delta^* = \chi_a^2(*) - Q_a \tag{166}$$

tend in probability to zero.

Lemma 7 is a direct consequence of lemma 6 and of the fact that for sufficiently large values of N the probability of the event point falling within $\tau_N(\eta)$ exceeds $1 - \eta$.

LEMMA 8. *As* N *is indefinitely increased, the distribution of* Q_a *tends to that of a central* χ^2 *with* $\nu - m$ *degrees of freedom.*

Lemma 8 is a consequence of the generalized theorem of Laplace and of the property (iii) of normal variables quoted above (see p. 265).

LEMMA 9. *As* N *is indefinitely increased, the distribution of either* χ_a^2 *or* $\chi_a^2(*)$ *tends to that of a central* χ^2 *with* $\nu - m$ *degrees of freedom.*

Since the χ^2 distribution is continuous, lemma 9 is implied by lemmas 3, 7, and 8.

Let

$$\psi_{i,j} = p_{i,j}^0 + \sum_{k=f+1}^{m} f_{i,j,k}^0 (\theta_k - \theta_k^0) \tag{167}$$

and Q_r denote the minimum of

$$\sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i} \frac{(q_{i,j} - \psi_{i,j})^2}{p_{i,j}^0} \tag{168}$$

computed with respect to the variation of $\theta_{f+1}, \theta_{f+2}, \cdots, \theta_m$. Also let $Q_b = Q_r - Q_a \geqq 0$.

LEMMA 10. *If the hypothesis tested* H *is true, then within* $\tau_N(\eta)$ *the values of* χ_b^2 *and* $\chi_b^2(*)$ *differ from* Q_b *no more than by terms of order* $O\left(\frac{1}{\sqrt{N}}\right)$. *Thus there exists a number* C_2 *such that within* $\tau_N(\eta)$

$$|\chi_b^2 - Q_b| < \frac{C_2}{\sqrt{N}} \qquad \text{and} \qquad |\chi_b(*) - Q_b| < \frac{C_2}{\sqrt{N}}. \tag{169}$$

LEMMA 11. *As* N $\to \infty$ *the distribution of* Q_b *tends to that of a central* χ^2 *with* f *degrees of freedom.*

LEMMA 12. *If the hypothesis tested* H *is true, then, as* N $\to \infty$, *the distribution of* χ_b^2 *and that of* $\chi_b^2(*)$ *tend to that of a central* χ^2 *with* f *degrees of freedom.*

Proofs of lemmas 10, 11, and 12 follow exactly the lines of proofs of lemmas 6, 8, and 9.

LEMMA 13. *The* λ*-test of the hypothesis* H *is consistent.*

Lemma 13 is a direct consequence of a general theorem of Wald [**12**] concerning the consistency of the λ-tests.

LEMMA 14. *Both tests* T_1 *and* T_2 *mentioned in theorem 7 are consistent.*

PROOF. Notice first that, since as $N \to \infty$ the distributions of χ_a^2 and $\chi_a^2(*)$ tend to that of a central χ^2 with $\nu - m$ degrees of freedom, whatever be $\epsilon > 0$, there exist two numbers A and N_ϵ, such that the inequality $N > N_\epsilon$ implies

$$\left.\begin{aligned} &P\{\chi_a^2 > A\} < \epsilon\,, \\ &P\{\chi_a^2(*) > A\} < \epsilon\,. \end{aligned}\right\} \tag{170}$$

The next step is to show that, if the hypothesis tested H is wrong and therefore some contradictory simple hypothesis h is true, then, whatever be $\eta > 0$ and M, there exists a number $N_{\eta,M}$, such that the inequalities $N > N_{\eta,M}$ imply

$$\left.\begin{aligned} &P\{\chi_r^2 > M|h\} > 1 - \eta\,, \\ &P\{\chi_r^2(*) > M|h\} > 1 - \eta\,. \end{aligned}\right\} \tag{171}$$

The proof consists in showing that for sufficiently large values of N the inequalities in the braces in (171) will be satisfied within the whole region $\tau_N(\eta)$.

It will be sufficient to carry out the proof only in relation to the first of the inequalities (171). The value of χ_r^2 can be written as

$$\chi_r^2 = NV(q)\,, \tag{172}$$

with

$$V(q) = \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{[q_{i,j} - p_{i,j}(H)]^2}{p_{i,j}(H)}\,. \tag{173}$$

Whatever be the method of determining $p_{i,j}(H)$, the value $V(p^0)$ of $V(q)$ taken at the point $q_{\alpha,\beta} = p_{\alpha,\beta}^0$ ($\alpha = 1, 2, \cdots, s$; $\beta = 1, 2, \cdots, \nu_\alpha$) must be positive,

$$V(p^0) = \sum_{i=1}^{s} Q_i \sum_{j=1}^{\nu_i} \frac{[p_{i,j}^0 - p_{i,j}(H)]^2}{p_{i,j}(H)} = 2\Delta > 0\,. \tag{174}$$

In fact, $V(p^0)$ could be equal to zero only if each $p_{i,j}(H) = p_{i,j}^0$. This, however, is impossible because, if the true hypothesis h contradicts H, then at least one of the estimates $p_{i,j}(H)$ must be different from the true value of $p_{i,j}^0$. This is immediately evident with such BAN estimates as are based on minimizing the generalized distance $\Delta(p, q)$ under the exact restrictions on the p's by which the hypothesis tested is expressed and which, therefore, must be satisfied by the $p_{i,j}(H)$ but not all by the $p_{i,j}^0$. However, the same conclusion is also true for the BAN estimate based on theorem 6, obtained by minimizing a generalized distance under the linear restrictions (97),

$$F^*_t(p) \equiv F_t(q) + \sum_{i=1}^{s} \sum_{j=1}^{\nu_i} b_{t,i,j}(p_{i,j} - q_{i,j}) = 0, \qquad t = 1, 2, \cdots, \nu - m + f. \tag{175}$$

In fact, if H is wrong, then the substitution of the $p_{i,j}^0$ instead of $p_{i,j}$ will fail to turn into zero at least one of the functions $F_t(p)$,

$$F_t(p^0) \neq 0 . \tag{176}$$

Thus if each $q_{i,j}$ is put equal to $p_{i,j}^0$ and if then each $p_{i,j}$ is put equal to $p_{i,j}^0$, then the corresponding value of F^*_t will be different from zero. It follows that, whichever of the four categories of BAN estimated is used, if H is wrong, then (174) must hold good. Since $V(q)$ is continuous, there must exist a vicinity S of the point $q_{\alpha,\beta} = p_{\alpha,\beta}^0$, say,

$$\sum_{i=1}^{s} \sum_{j=1}^{\nu_i} (q_{i,j} - p_{i,j}^0)^2 < \delta , \tag{177}$$

such that, within S,

$$V(q) > \Delta . \tag{178}$$

In the same vicinity

$$\chi_r^2 > N\Delta , \tag{179}$$

and may be made as large as desired provided N is taken sufficiently large. However, the region $\tau_N(\eta)$ is defined by the inequality (137) and it will be seen that, for sufficiently large values of N, $\tau_N(\eta)$ will be entirely contained in S.

This proves the assertion (171).

It is now easy to see that, in all conditions,

$$P\{\chi_r^2 > \chi_\epsilon^2\} \geqq P\{\chi_r^2 > A + \chi_\epsilon^2\} - P\{\chi_a^2 > A\} . \tag{180}$$

The application of the results just obtained leads to the conclusion that, for sufficiently large values of N,

$$P\{\chi_b^2 > \chi_\epsilon^2 | h\} > 1 - 2\eta , \tag{181}$$

which means that test T_1 is consistent. The same argument applies to test T_2.

Lemma 15. *If the hypothesis tested* H *is true, then within the region* $\tau_N(\eta)$ *the value of* $-2\log\lambda_N$ *differs from* Q_b *no more than by terms of order* $O\left(\frac{1}{\sqrt{N}}\right)$. *Thus there exists a number* C_3 *such that, within* $\tau_N(\eta)$,

$$|-2\log\lambda_N - Q_b| < \frac{C_3}{\sqrt{N}}. \tag{182}$$

Proof. The expression of the $\log\lambda_N$ can be written as follows:

$$\log\lambda_N = \sum_{i=1}^{s}\sum_{j=1}^{\nu_i} n_{i,j}\left\{\log\frac{\hat{\hat{p}}_{i,j}}{q_{i,j}} - \log\frac{\hat{p}_{i,j}}{q_{i,j}}\right\}. \tag{183}$$

Let now $p_{i,j}$ mean any positive numbers, bounded from zero and such that

$$\sum_{j=1}^{\nu_i} p_{i,j} = 1. \tag{184}$$

Writing

$$q_{i,j} = p_{i,j} + u_{i,j}\sqrt{\frac{p_{i,j}}{n_i}} \tag{185}$$

and applying the familiar expansions, it is easy to find that so long as the $p_{i,j}$ remain bounded from zero and $q_{i,j} = p_{i,j} + O\left(\frac{1}{\sqrt{N}}\right)$,

$$\sum_{i=1}^{s}\sum_{j=1}^{\nu_i} n_{i,j}\log\left(\frac{q_{i,j}}{p_{i,j}}\right) = \frac{1}{2}\sum_{i=1}^{s} n_i \sum_{j=1}^{\nu_i}\frac{(q_{i,j}-p_{i,j})^2}{p_{i,j}} + O\left(\frac{1}{\sqrt{N}}\right). \tag{186}$$

Since the true probabilities $p_{i,j}^0$ are different from zero and, with the hypothesis H being true, the estimates $\hat{p}_{i,j}$ and $\hat{\hat{p}}_{i,j}$ tend to $p_{i,j}^0$ as $N\to\infty$, uniformly in $\tau_N(\eta)$, formula (186) may be applied to $\log\lambda_N$ of (183) giving

$$-2\log\lambda_N = \sum_{i=1}^{s} n_i\sum_{j=1}^{\nu_i}\frac{(q_{i,j}-\hat{\hat{p}}_{i,j})^2}{\hat{\hat{p}}_{i,j}} - \sum_{i=1}^{s} n_i\sum_{j=1}^{\nu_i}\frac{(q_{i,j}-\hat{p}_{i,j})^2}{\hat{p}_{i,j}} + O\left(\frac{1}{\sqrt{N}}\right), \tag{187}$$

and the proof of lemma 15 is concluded by a reference to lemma 10.

Lemma 16. *If the hypothesis tested* H *is true, then the probability that the* λ*-test and either of the tests* T_1 *and* T_2 *described in theorem 7 will contradict each other tends to zero as* N *is indefinitely increased.*

It will be sufficient to prove lemma 16 in relation to test T_1 only. The Nth critical regions w_N and u_N respectively of the λ-test and of test T_1 are defined by the inequalities (130) and (131), which we shall rewrite as

$$\left.\begin{aligned} -2\log\lambda_N &\geqq \chi_\epsilon^2, \\ \chi_b^2 &\geqq \chi_\epsilon^2. \end{aligned}\right\} \tag{188}$$

According to the remark made previously, it will be sufficient to show that as $N \to \infty$ the probability of E_N falling within each of the regions

$$\tau_N(\eta)[w_N - u_N w_N] \qquad \text{and} \qquad \tau_N(\eta)[u_N - u_N w_N] \tag{189}$$

tends to zero. Since in the two cases the reasoning is the same, only the case of the first of the two regions (189) will be discussed in detail.

The point E_N falls within $(w_N - u_N w_N)$ when simultaneously

$$-2 \log \lambda_N \geqq \chi_\epsilon^2 \tag{190}$$

and

$$\chi_b^2 < \chi_\epsilon^2. \tag{191}$$

Owing to lemmas 10 and 15, there exists a number C such that within $\tau_N(\eta)$

$$Q_b - \frac{C}{\sqrt{N}} < \chi_b^2 \tag{192}$$

and

$$-2 \log \lambda_N < Q_b + \frac{C}{\sqrt{N}}. \tag{193}$$

It follows that if E_N falls within that part of τ_N which belongs to $w_N - w_N u_N$, then Q_b must satisfy the inequalities

$$Q_b - \frac{C}{\sqrt{N}} < \chi_\epsilon^2 < Q_b + \frac{C}{\sqrt{N}} \tag{194}$$

or

$$\chi_\epsilon^2 - \frac{C}{\sqrt{N}} < Q_b < \chi_\epsilon^2 + \frac{C}{\sqrt{N}}, \tag{195}$$

which means that

$$P\{E_N \epsilon\, \tau_N(\eta)[w_N - w_N u_N] \,|H\} \leqq P\{\chi_\epsilon^2 - \frac{C}{\sqrt{N}} < Q_b < \chi_\epsilon^2 + \frac{C}{\sqrt{N}} \,|H\}. \tag{196}$$

But as $N \to \infty$ the probability in the right-hand side of (196) tends to zero. Therefore the probability in the left-hand side of (196) must tend to zero as N is indefinitely increased. This proves lemma 16.

Lemmas 13, 14, and 16 imply theorem 7. It will be noticed also that whichever of the three tests mentioned in theorem 7 is used lemma 12 implies that, as $N \to \infty$, the probability of rejecting the hypothesis tested H when it is true tends to the selected level of significance ϵ.

5. Summary

1. In many important problems of application the use of the maximum likelihood estimates and the application of the λ-test of statistical hypotheses

are prohibitive because of the difficulty in solving systems of equations which these methods involve.

In relation to the situation described in section 2, theorem 6 gives a method of determining estimates, termed BAN estimates, which have the same asymptotic properties as the maximum likelihood estimates. When parameters to be estimated are expressed as functions of probabilities $p_{i,j}$, this method reduces to the solution of a system of linear equations.

2. Theorem 7 determines two types of the χ^2-test both of which are consistent and equivalent in the limit to the λ-test. The application of these tests depends on the possibility of computing the BAN estimates of the unknown parameters. Owing to theorem 6, this procedure is frequently reduced to the solution of a system of linear equations.

3. The machinery of the χ^2-tests mentioned in paragraph 2 above was first presented in 1929 by the writer [**6**]. However, in that early paper the justification of the tests was based on consideration of the probabilities a posteriori, of which the present paper is entirely free.

REFERENCES

1. DAVID, F. N. "Limiting distributions connected with certain methods of sampling human populations," *Stat. Res. Mem.*, vol. 2 (1938), pp. 69–90.
2. DOOB, J. L. "Probability and statistics," *Trans. Amer. Math. Soc.*, vol. 36 (1934), pp. 759–775.
3. FISHER, R. A. "On a property connecting the measure of discrepancy with the method of maximum likelihood," *Atti del Congresso Internazionale dei Matematici*, vol. 6 (1932), pp. 95–100; particularly p. 97.
4. FRÉCHET, MAURICE. *Recherches Théoriques Modernes sur la Théorie des Probabilités.* Paris, 1937.
5. HOTELLING, HAROLD. "The consistency and ultimate distribution of optimum statistics," *Trans. Amer. Math. Soc.*, vol. 32 (1930), pp. 847–859.
6. NEYMAN, J. "Contribution to theory of certain test criteria," *XVIII Session de l'Institut International de Statistique* (Varsovie, 1929), pp. 1–48.
7. ———. "On a new class of 'contagious' distributions, applicable in entomology and bacteriology," *Annals of Math. Stat.*, vol. 10 (1939), pp. 35–57.
8. ———. "Conception of equivalence in the limit of tests and its application to certain λ- and χ^2-tests," *ibid.*, vol. 11 (1940), pp. 477–478.
9. NEYMAN, J., and E. S. PEARSON. "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, vol. 20-A (1928), pp. 175–240 and 263–294.
10. PÓLYA, G. "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momenten problem," *Math. Zeitschrift*, vol. 8 (1920), pp. 171–181.
11. TANG, P. C. "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, vol. 2 (1938), pp. 126–149.
12. WALD, ABRAHAM. "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, vol. 54 (1943), pp. 426–482.
13. WOLFOWITZ, J. "Additive partition functions and a class of statistical hypotheses," *Annals of Math. Stat.*, vol. 13 (1942), pp. 247–279.