The Annals of Statistics
1975, Vol. 3, No. 4, 959-964

UNBIASEDNESS OF THE CHI-SQUARE, LIKELIHOOD RATIO, AND OTHER GOODNESS OF FIT TESTS FOR THE EQUAL CELL CASE

BY ARTHUR COHEN[1] AND H. B. SACKROWITZ

Rutgers University

The chi-square test, likelihood ratio test, and other goodness of fit tests in a family are shown to have monotonic power functions, and hence are unbiased, for testing a simple hypothesis that all cell probabilities are equal. The chi-square test is also shown to be Type D for this situation.

1. Introduction and summary. Let $\mathbf{X} = (X_1, X_2, \cdots, X_N)$ be a random sample of size N from a population whose cumulative distribution function is $F(x)$. Partition the real line into k class intervals such that the probability that X_j falls into the ith class interval is p_i under $F(x)$. Let $\mathbf{n} = \mathbf{n}(\mathbf{X}) = (n_1(\mathbf{X}), \cdots, n_k(\mathbf{X}))$, where $n_i(\mathbf{X})$ is the number of observations in the sample falling into the ith class interval. Let $F_0(x)$ denote a specified distribution with cell probabilities p_{i0}, $i = 1, 2, \cdots, k$. Consider the case where all $p_{i0} = 1/k$. We wish to test the hypothesis $H_0: p_i = 1/k$, $i = 1, 2, \cdots, k$. Let h_i, $i = 1, 2, \cdots, k$ be convex functions of a single nonnegative variable. We study tests of the form, reject H_0 if

$$T_{\mathbf{h}} = \sum_{i=1}^{k} h_i(n_i) > c \tag{1.1}$$

where c is a positive constant and $\mathbf{h} = (h_1, h_2, \cdots, h_k)$.

We prove the following:

(1) For the tests in (1.1) the power functions are monotone on all lines passing through the null point $\mathbf{p}_0 = (1/k, 1/k, \cdots, 1/k)$, in the sense that the power functions increase away from the null point in either direction. This property implies unbiasedness of the tests (1.1). Mann and Wald (1942) claimed to have proven this for the chi-square test. However, as pointed out by Kendall and Stuart (1967), Mann and Wald proved only the local unbiased property of the chi-square test.

(2) The chi-square test ($h_i(n_i) = n_i^2/p_{i0}$) is Type D when all $p_{i0} = 1/k$. A type D test is locally strictly unbiased and among all tests (of the same size) with this property, it maximizes the determinant of the matrix of second partial derivatives of the power function evaluated at the null point. (See Lehmann (1959) page 342.)

This result is particularly interesting since this demonstrates an optimum local property of the chi-square test and also demonstrates an optimum property for

Received September 1973; revised October 1974.

[1] Research supported by N.S.F. Grant No. GP 36566X.

AMS 1970 *subject classifications.* Primary 62F05; Secondary 62G10.

Key words and phrases. Hypothesis testing, chi-square test, likelihood ratio tests, goodness of fit tests, unbiasedness, Type D.

fixed N and fixed k. Hoeffding (1965) and Bahadur (1967) show that the likelihood ratio test (LRT) ($h_i(n_i) = n_i \log (n_i/p_{i0})$) is preferable to the chi-square test as $N \to \infty$, k remains fixed, and one uses their criteria of efficiency. They demonstrate an asymptotic nonlocal optimality property of the LRT. Morris (1966) describes some situations where $k \to \infty$ in such a way that N/k is moderate, in which the chi-square test appears preferable to the LRT.

In Section 2 we prove the unbiasedness result. In Section 3 we prove the Type D result.

2. Power function property and unbiasedness. In this section we show that the power function of the test (1.1) has the following property: On any line passing through the null point $\mathbf{p}_0$, the power function increases away from the null point. Clearly such a property implies that the test (1.1) is unbiased.

We assume throughout this section that the critical constant c is chosen so that the test neither always accepts nor always rejects. To prove the result we need the following lemmas.

LEMMA 2.1. *Let h be a convex function. Let z have a binomial distribution with parameters ν and θ where, $\theta \geqq \frac{1}{2}$. Then for any real number d,*

$$P(h(z+1) + h(\nu - z) > d) - P(h(\nu - z + 1) + h(z) > d) \geqq 0 . \tag{2.1}$$

PROOF. Consider

$$\Delta(z) = [h(z+1) + h(\nu - z)] - [h(\nu - z + 1) + h(z)] . \tag{2.2}$$

First we note that $\Delta((\nu/2) - r) = -\Delta((\nu/2) + r)$ for all r. In addition, for $r > 0$, writing $((\nu/2) - r + 1)$ and $((\nu/2) + r)$ as convex combinations of $((\nu/2) + r + 1)$ and $((\nu/2) - r)$, the convexity of h implies

$$\begin{aligned} \Delta((\nu/2) - r) = {} & [h((\nu/2) - r + 1) + h((\nu/2) + r)] \\ & - [h((\nu/2) + r + 1) + h((\nu/2) - r)] \leqq 0 . \end{aligned} \tag{2.3}$$

Thus $\Delta(z) \leqq 0$ for $z < \nu/2$, $\Delta(z) \geqq 0$ for $z > \nu/2$, and $\Delta(\nu/2) = 0$.
Define the sets D_1, D_2 by

$$D_1 = \{z : h(z+1) + h(\nu - z) > d, h(\nu - z + 1) + h(z) \leqq d\} ,$$
$$D_2 = \{z : h(z+1) + h(\nu - z) \leqq d, h(\nu - z + 1) + h(z) > d\} .$$

Note that D_1, D_2 may be empty. We may now write

$$\begin{aligned} & P(h(z+1) + h(\nu - z) > d) - P(h(\nu - z + 1) + h(z) > d) \\ & \quad = P(z \in D_1) - P(z \in D_2) . \end{aligned} \tag{2.4}$$

By our established properties of $\Delta(z)$ it follows that $D_2 \subseteqq [0, \nu/2)$, $D_1 \subseteqq (\nu/2, \nu]$, and $D_2 = \{z : (\nu - z) \in D_1\}$. Since $\theta \geqq \frac{1}{2}$,

$$P(z \in D_1) - P(z \in D_2) \geqq 0 . \tag{2.5}$$

Thus from (2.5) and (2.4), (2.1) is proved. This completes the proof of the lemma.

Note that if $\frac{1}{2} < \theta < 1$ and if D_1 contains at least one integer then (2.5) > 0 which implies (2.1) > 0.

LEMMA 2.2. *Consider the test* (1.1) *with* $h_i = h$. *If the size of the test is* > 0, *then the power of the test is* 1 *for* $\{\mathbf{p} : p_j = 1, p_i = 0, i \neq j, j = 1, \cdots, k.\}$

PROOF. Since the size of the test is > 0 there must exist some point $(n_1, n_2, \cdots, n_k)$ for which (1.1) rejects. By the convexity of h we have $h(N) + (k-1)h(0) \geqq \sum_{i=1}^k h(n_i)$ for all $(n_1, \cdots, n_k)$ such that $\sum_{i=1}^k n_i = N$. Thus the point $\mathbf{n}^* = (n_1, n_2, \cdots, n_k)$, for which $n_j = N$, $n_i = 0$, $i \neq j$ must be a point for which (1.1) rejects. This completes the proof as $P(n_j = N, n_i = 0, i \neq j) = 1$, when $p_j = 1$.

Now we are ready to prove

THEOREM 2.1. *Consider the test of* H_0 *given in* (1.1) *with* $h_i = h$. *Then on any line passing through* $\mathbf{p}_0$, *the power function of the test increases away from the null point.*

PROOF. Let $R = \{(n_1, n_2, \cdots, n_k) : T_h = \sum_{i=1}^k h(n_i) > c\}$. Also let the line, in the parameter space of vectors $\mathbf{p}$, passing through $\mathbf{p}_0$ be represented by the points $(a_1 t + (1/k), a_2 t + (1/k), \cdots, a_k t + (1/k))$, where the a_i are fixed, $((-1/ka_i) < t < (k-1)/ka_i)$, $i = 1, 2, \cdots, k$ and $\sum_{i=1}^k (a_i t + (1/k)) = 1$. This latter condition implies $\sum_{i=1}^k a_i = 0$. Now the power of the given test at any point on the line passing through $\mathbf{p}_0$ is

$$\sum_R (N!/\textstyle\prod_{i=1}^k n_i!) \prod_{i=1}^k (a_i t + (1/k))^{n_i} . \tag{2.6}$$

Differentiate in (2.6) with respect to t and find that the derivative is

$$\begin{aligned} &\textstyle\sum_R (N!/\prod n_i!) \sum_{i=1}^k a_i n_i (a_i t + (1/k))^{n_i - 1} \prod_{j=1, j\neq i}^k (a_j t + (1/k))^{n_j} \\ &\quad = N \textstyle\sum_R \sum_{i=1}^k a_i \{[(N-1)!/\prod_{j=1, j\neq i}^k n_j! \,(n_i - 1)!](a_i t + (1/k))^{n_i - 1} \\ &\qquad \times \textstyle\prod_{j=1, j\neq i}^k (a_j t + (1/k)^{n_j}\} . \end{aligned} \tag{2.7}$$

(Note that in (2.7) and what follows, all terms involving $(n_i - 1)$ are zero when $n_i = 0$.) Since the bracketed term on the r.h.s. of (2.7) is a multinomial density for $m_i = n_i - 1$, $m_j = n_j$, for $j \neq i$, $j = 1, 2, \cdots, k$, and probabilities $p_i = (a_i t + (1/k))$, (2.7) may be rewritten as

$$N \textstyle\sum_{i=1}^k a_i \Pr\{\sum_{j=1, j\neq i}^k h(m_j) + h(m_i + 1) > c\} , \tag{2.8}$$

where Pr is with respect to the distribution of the m_i.

Let J_+ be those indices $i = 1, 2, \cdots, k$ for which $a_i > 0$, and J_- be those indices $i = 1, 2, \cdots, k$ for which $a_i < 0$. Since not all $a_i \neq 0$, and $\sum_{i=1}^k a_i = 0$, J_+ and J_- are not empty. Also let $A = \sum_{i \in J_+} a_i = -\sum_{i \in J_-} a_i$. Now fix a $t > 0$. Then (2.8) may be written as

$$\begin{aligned} &N[\textstyle\sum_{i \in J_+} a_i \Pr\{\sum_{j=1, j\neq i}^k h(m_j) + h(m_i + 1) > c\} \\ &\qquad + \textstyle\sum_{i \in J_-} a_i \Pr\{\sum_{j=1, j\neq i}^k h(m_j) + h(m_i + 1) > c\}] \\ &\qquad\quad \geqq NA[\min_{i \in J_+} \Pr\{\textstyle\sum_{j=1, j\neq i}^k h(m_j) + h(m_i + 1) > c\} \\ &\qquad\qquad - \max_{i \in J_-} \Pr\{\textstyle\sum_{j=1, j\neq i}^k h(m_j) + h(m_i + 1) > c\}] \\ &\qquad\quad = NA[\Pr\{\textstyle\sum_{j=1, j\neq i_0}^k h(m_j) + h(m_{i_0} + 1) > c\} \\ &\qquad\qquad - \Pr\{\textstyle\sum_{j=1, j\neq i_1}^k h(m_j) + h(m_{i_1} + 1) > c\}] , \end{aligned} \tag{2.9}$$

where i_0 and i_1 are indices in J_+, J_- respectively which give the min and max respectively for the relevant probability. Thus the theorem will be proved for $t > 0$, provided the term in brackets on the r.h.s. of (2.9) is greater than zero. Therefore write this term as

$$\begin{aligned}\int [\Pr\{h(m_{i_0}+1) + h(m_{i_1}) > c \\ - \textstyle\sum_{j\neq i_0,i_1} h(m_j) \mid m_j, j = 1, \cdots, k, j \neq i_0, i_1\} \\ - \Pr\{h(m_{i_1}+1) + h(m_{i_0}) > c \\ - \textstyle\sum_{j\neq i_0,i_1} h(m_j) \mid m_j, j = 1, \cdots, k, j \neq i_0, i_1\}] \\ \times dP(m_j, j = 1, \cdots, k, j \neq i_0, i_1) .\end{aligned} \tag{2.10}$$

On examining the integrand we realize that for fixed $m_j, j = 1, \cdots, k, j \neq i_0, i_1$, the pair (m_{i_0}, m_{i_1}) have a (conditional) two-dimensional multinomial distribution with parameters $N - 1 - \sum_{j\neq i_0,i_1} m_j$; $\theta = (a_{i_0}t + 1/k)/[(a_{i_0} + a_{i_1})t + 2/k]$, $1 - \theta$, (equivalently we can think of m_{i_0} as having a binomial distribution). In fact we may write the integrand as

$$\Pr\{h(z+1) + h(\nu - z) > d\} - \Pr\{h(\nu - z + 1) + h(z) > d\} , \tag{2.11}$$

where $d = c - \sum_{j\neq i_0,i_1} h(m_j)$ and z has a binomial distribution with parameters $\nu = N - 1 - \sum_{j\neq i_0,i_1} m_j, \theta$. Since $a_{i_0} > 0 > a_{i_1}$ we have $\theta > \frac{1}{2}$. Thus by Lemma 2.1 the integrand is always nonnegative. Hence, by virtue of (2.11) and (2.10) we conclude that (2.9) is greater than or equal to zero for $t > 0$. Thus the power function is nondecreasing as t ranges over positive values.

We now let $t > 0$, be such that $\mathbf{p} = (a_1 t + 1/k, a_2 t + 1/k, \cdots, a_k t + 1/k)$ is an interior point of the parameter space and argue that the integrand of (2.10) is greater than zero at some points $(m_j, j = 1, 2, \cdots, k, j \neq i_0, i_1)$ with positive probability. Suppose not. This means that

$$\begin{aligned}\Pr(\textstyle\sum_{j=1,j\neq i_0}^{k} h(m_j) + h(m_{i_0}+1) > c) \\ = \Pr(\textstyle\sum_{j=1,j\neq i_1}^{k} h(m_j) + h(m_{i_1}+1) > c) .\end{aligned} \tag{2.12}$$

By symmetry, since determination of D_1 with $\mathbf{p}$ an interior point does not depend on the pair of indices i_0 and i_1, we find $\Pr(\sum_{j=1,j\neq i}^{k} h(m_j) + h(m_i + 1) > c)$, $i = 1, 2, \cdots, k$, are all equal. This yields that the l.h.s. of (2.9) and (2.7) equals zero. This in turn implies that the power function is constant on interior points and by continuity must be constant everywhere. By the assumption that c is chosen so that we do not always reject, the constant power would have to be less than 1. But by Lemma 2.2 we know the power is 1 at some points in the parameter space. This is a contradiction and so (2.10) is greater than zero which implies (2.9) is greater than zero, for $t > 0$. This proves the theorem for $t > 0$. For $t < 0$ a similar argument yields that the derivative in (2.7) is negative. This completes the proof of the theorem.

REMARK 2.1. The chi-square test, with $h(n_i) = n_i^2$, and the likelihood ratio test, with $h(n_i) = n_i \log n_i$, clearly satisfy the conditions of the theorem.

REMARK 2.2. Kendall and Stuart (1967), page 436, say that the chi-square test, where p_{i0} are all equal, yields a locally unbiased test. They say that this is a recommendation of such a class formation since no such result is known to hold for the chi-square test in general.

One can slightly improve on the latter part of the above remark. That is, if the p_{i0} are not equal the chi-square test is not likely to be unbiased. In the following simple example it is seen that the chi-square test is not unbiased. Let $k = 2$, $N = 3$, $p_{10} = \frac{1}{4}$. Choose c so that chi-square rejects if $n_1 = 0, 2, 3$. The power of the test is easily computed to be $(1 - 3p + 6p^2 - 3p^3)$ which does not have its minimum at $\frac{1}{4}$.

3. Type D property of the chi-square test. Let $\varphi(\mathbf{n})$ denote a test function and let $\beta_\varphi(p_1, p_2, \cdots, p_{k-1})$ denote its power function. Let $\beta_\varphi^{ij} = (\partial^2/\partial p_i \partial p_j)\beta_\varphi(p_1, p_2, \cdots, p_{k-1})|_{\mathbf{p}=\mathbf{p}_0}$ and let Γ_φ be the $(k-1) \times (k-1)$ matrix with elements β_φ^{ij}. Let $\varphi^*(\mathbf{n})$ represent the chi-square test; i.e. $\varphi^*(\mathbf{n}) = 1$ if $\mathbf{n}'\mathbf{n} > c$, $\varphi^*(\mathbf{n}) = 0$ if $\mathbf{n}'\mathbf{n} < c$, and $\varphi^*(\mathbf{n})$ is arbitrary if $\mathbf{n}'\mathbf{n} = c$. (Assume c is such that the test does not always accept or does not always reject.). Let α be the size of the test. Finally, let $f_\mathbf{n} = (N!/\prod_{i=1}^k n_i!)(1/k^N)$. Now we prove

THEOREM 3.1. *The chi-square test $\varphi^*(\mathbf{n})$ is type* D.

PROOF. In Section 2 we proved that φ^* is strictly unbiased. Therefore we need only show

$$|\Gamma_{\varphi^*}| = \sup_\varphi |\Gamma_\varphi| \, . \tag{3.1}$$

First we show (3.1) where φ is permutation invariant. In this case all β_φ^{ii} are equal for $i = 1, 2, \cdots, k-1$ and β_φ^{ij} are all equal for all i, j, $i \neq j$. Letting $\sum_\mathbf{n}$ represent the sum over all possible vectors $\mathbf{n}$,

$$\beta_\varphi = \textstyle\sum_\mathbf{n} \varphi(\mathbf{n})(N!/\prod n_i!) \prod_{i=1}^k p_i^{n_i} \, . \tag{3.2}$$

From (3.2) it is easily found that

$$\begin{aligned} \beta_\varphi^{11} &= k^2 \textstyle\sum_\mathbf{n} \varphi(\mathbf{n})[n_1(n_1 - 1) - 2n_1 n_k + n_k(n_k - 1)]f_\mathbf{n} \\ \beta_\varphi^{12} &= k^2 \textstyle\sum_\mathbf{n} \varphi(\mathbf{n})[n_1 n_2 - n_1 n_k - n_2 n_k + n_k(n_k - 1)]f_\mathbf{n} \, . \end{aligned} \tag{3.3}$$

The permutation invariance of φ implies that Γ_φ is of the intraclass type and so Γ_φ has $(k-2)$ latent roots equal to $\beta^{11} - \beta^{12}$, and one root equal to $\beta^{11} + (k-2)\beta^{12}$. (See for example, Olkin and Pratt (1958), page 205.) Use permutation invariance and $\sum_{i=1}^k n_i = N$ to yield the identity

$$\textstyle\sum_\mathbf{n} \varphi(\mathbf{n}) n_1 n_2 f_\mathbf{n} = (N^2\alpha/k(k-1)) - (1/(k-1)) \sum_\mathbf{n} \varphi(\mathbf{n}) n_1^2 f_\mathbf{n} \, . \tag{3.4}$$

Permutation invariance (3.3), and (3.4) then yield that

$$\begin{aligned} \beta^{11} - \beta^{12} &= k^2\{(1/(k-1)) \textstyle\sum_\mathbf{n} \varphi(\mathbf{n})\mathbf{n}'\mathbf{n} f_\mathbf{n} - N\alpha/k - N^2\alpha/k(k-1)\} \\ \beta^{11} + (k-2)\beta^{12} &= k(\beta^{11} - \beta^{12}) \, . \end{aligned} \tag{3.5}$$

From (3.5) it is clear that $|\Gamma_\varphi|$ is maximized among permutation invariant tests by choosing φ to maximize

$$\sum_{\mathbf{n}} \varphi(\mathbf{n})\mathbf{n}'\mathbf{n}f_{\mathbf{n}} \tag{3.6}$$

subject to $\sum_{\mathbf{n}} \varphi(\mathbf{n})f_{\mathbf{n}} = \alpha$. Since the quantity

$$\sum_{\mathbf{n}} (\varphi^*(\mathbf{n}) - \varphi(\mathbf{n}))(\mathbf{n}'\mathbf{n} - c)f_{\mathbf{n}} \geqq 0 \,, \tag{3.7}$$

for all tests φ of size α, it follows that φ^* maximizes (3.6).

The proof of the theorem will be completed by showing that if φ' is a test which is not permutation invariant such that $|\Gamma_{\varphi'}| > |\Gamma_{\varphi^*}|$, then there is a permutation invariant test ψ such that $|\Gamma_{\psi}| > |\Gamma_{\varphi^*}|$. This will be a contradiction.

Let g_j, $j = 1, 2, \cdots, (k-1)!$ be the elements of the permutation group G. Let $\psi(\mathbf{n}) = (1/(k-1)!) \sum_{j=1}^{(k-1)!} \varphi'(g_j(\mathbf{n}))$. It is known that $\psi(\mathbf{n})$ is invariant. Furthermore, it is easy to see that $\Gamma_{\psi} = (1/(k-1)!) \sum_{j=1}^{(k-1)!} \Gamma_{\varphi'(g_j)}$ and that $|\Gamma_{\varphi'}| = |\Gamma_{\varphi'(g_j)}|$, $j = 1, 2, \cdots, (k-1)!$ Since the determinant is log concave (see for example, Marcus and Minc (1964), page 115, (8)), it follows that $|\Gamma_{\psi}| \geqq |\Gamma_{\varphi'}|$. This provides the contradiction and the proof of the theorem is complete.

REFERENCES

[1] Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

[2] Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

[3] Kendall, M. G. and Stuart, A. (1967). *The Advanced Theory of Statistics* **2**, 2nd ed. Hafner, New York.

[4] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[5] Mann, H. B. and Wald, A. (1942). On the choice of the number of intervals in the application of the chi-square test. *Ann. Math. Statist.* **13** 306–317.

[6] Marcus, M. and Minc, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities.* Allyn and Bacon, Boston.

[7] Morris, C. (1966). Admissible Bayes procedures and classes of epsilon Bayes procedures for testing hypotheses in a multinomial distribution. Technical Report No. 55, Department of Statistics, Stanford Univ.

[8] Olkin, I. and Pratt, J. (1958). Unbiased estimation of certain correlation coefficients. *Ann. Math. Statist.* **29** 201–211.

Statistics Center
Rutgers University
New Brunswick, N. J. 08903