A CHI-SQUARE STATISTIC FOR GOODNESS-OF-FIT TESTS WITHIN THE EXPONENTIAL FAMILY

BU-503-M

by

March, 1974

K. C. Rao* and D. S. Robson
Cornell University

Abstract

When the class boundaries used in constructing a chi-square goodness-of-fit statistic are predetermined and the unknown parameters are estimated by maximum likelihood from the ungrouped data, the resulting statistic does not have a limiting X^2-distribution but instead is asymptotically distributed as a linear function of chi-square variables. The same result applies in the more realistic and useful case where only the number of classes and their probability content are predetermined. It is shown here that in both of the above cases, in the case of exponential family, the quadratic form of the asymptotic multinormal conditional distribution of the class frequencies given the parameter estimates can be used to test the goodness-of-fit. The statistic does have a limiting X^2-distribution and the degrees of freedom are only one less than the number of classes after grouping, regardless of the number of parameters estimated.

---

* Part of the work was done at the National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

Paper No. BU-503-M in the Biometrics Unit Mimeo Series, Cornell University.

A CHI-SQUARE STATISTIC FOR GOODNESS-OF-FIT TESTS WITHIN THE EXPONENTIAL FAMILY

by

K. C. Rao* and D. S. Robson

Cornell University

SUMMARY

When the class boundaries used in constructing a chi-square goodness-of-fit statistic are predetermined and the unknown parameters are estimated by maximum likelihood from the ungrouped data, the resulting statistic does not have a limiting X^2-distribution but instead is asymptotically distributed as a linear function of chi-square variables. The same result applies in the more realistic and useful case where only the number of classes and their probability content are predetermined. It is shown here that in both of the above cases, in the case of exponential family, the quadratic form of the asymptotic multinormal conditional distribution of the class frequencies given the parameter estimates can be used to test the goodness-of-fit. The statistic does have a limiting X^2-distribution and the degrees of freedom are only one less than the number of classes after grouping, regardless of the number of parameters estimated.

Some key words: Conditional probability density function (c.p.d.f.), approximation to c.p.d.f., power comparisons.

1. INTRODUCTION

The classical procedure for testing whether a sample $x_1, \cdots, x_n$ is obtained from a specified univariate parametric family $f(x;\theta)$, such as Poisson or Normal,

---

* Part of the work was done at the National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

employs a statistic measuring goodness-of-fit between the observed (ν_i) and expected (np_i) numbers of observations falling into r predetermined classes. If $f(x;\theta)$ involves unknown parameters $\theta = (\theta_1, \cdots, \theta_s)$ these can be estimated as functions of ν_i using the maximum likelihood or minimum χ^2-procedure to obtain estimates $\tilde{p}_i = \tilde{p}_i(\nu_1, \cdots, \nu_r)$ of class probabilities $p_i (i = 1, \cdots, r)$. Under certain regularity conditions (cf. Cramer (1946) pp. 477-479) the goodness-of-fit statistic

$$\tilde{R} = \sum_{i=1}^{r} \frac{(\nu_i - n\tilde{p}_i)^2}{n\tilde{p}_i} \tag{1.1}$$

is then asymptotically distributed as χ^2 with r-s-1 degrees of freedom (χ^2_{r-s-1}, briefly). However, if the original observations $x_1, \cdots, x_n$ are available and if the class frequencies $\nu_1, \cdots, \nu_{r-1}$ are not a statistically sufficient reduction of $x_1, \cdots, x_n$, then more efficient estimators of p_i are available, such as maximum likelihood estimators $\hat{p}_i = p_i(\hat{\theta})$ obtained by maximizing the likelihood of $x_1, \cdots, x_n$ with respect to θ. Chernoff and Lehmann (1954) have shown that the statistic thus constructed

$$X^2 = \sum_{i=1}^{r} \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i} \tag{1.2}$$

is asymptotically distributed as a linear function of chi-square variables, $X^2 \overset{\mathcal{L}}{=} y_1^2 + \cdots + y_{r-s-1}^2 + \lambda_1 y_{r-s}^2 + \cdots + \lambda_s y_{r-1}^2$, where y_i are independent standard normal variables and the λ's, constrained by $0 \le \lambda_i < 1$, may depend on the s unknown parameters $\theta_1, \cdots, \theta_s$.

Chernoff and Lehmann (1954) considered only the case where the class boundaries are predetermined. Subsequently, A. R. Roy (1956) and Watson (1957, 1958) independently showed that this same result applies in the more realistic and useful case where only the number of classes, r, and the p_i are predetermined; the class boundaries are then functions of $\hat{\theta}$. Watson (1958) concludes that if the parameters involved are those of location and scale, the asymptotic distribution of (1.2) is independent of parameters. Moore (1971) and Dahiya and Gurland (1972) tabulated the percentile points of the asymptotic distribution of (1.2) when $f(x;\theta)$ is the normal probability density function with unknown mean and variance.

We show here that the asymptotic dependence on both the parameters and the functional form of $f(x;\theta)$ can be eliminated by adding a correction term Y^2 which converges in law to $(1 - \lambda_1)y^2_{r-s} + \cdots + (1 - \lambda_s)y^2_{r-1}$. The s degrees of freedom which are completely lost in (1.1) where the parameter estimates are based on grouped data, fractionally recovered in X^2 where θ is estimated before grouping, are thus totally recovered in the corrected statistic $X^2 + Y^2$.

The existence of such a statistic is evident in the special case where $f(x;\theta)$ is a member of the exponential family. Rao and Chakravarti (1956) following up the work of Fisher (1950) obtained a test statistic for testing the goodness-of-fit of a Poisson distribution. They developed the statistic from the distribution of class frequencies given the sufficient statistic. H. Levene (1949) used this technique to test if the frequency of any one cell is a violator of the binomial distribution $(q + p)^2$, and C. Vithayasai (1971) examined the small sample behavior of the statistic proposed by Levene. The method presented in this paper unifies the conditional approach and extends it to the continuous case when $f(x;\theta)$ belongs to the exponential family and hence admits a minimal sufficient statistic, $\hat{\theta}$. The following development holds if $\mathrm{Var}(\hat{\theta})$ is of the form c/n; and if $\mathrm{Var}(\hat{\theta})$ is of the form $c/n^{1+\delta}$, $\delta > 0$, it can be shown that (1.2) is asymptotically distributed as χ^2_{r-1}. Some comparisons are given between the power of the statistic proposed in this paper and the power of (1.1) and (1.2).

In what follows, i takes values from 1 to r and j,k assume values from 1 to s. The notation $y_n = o_p(r_n)$ means $|y_n|/r_n$ approaches zero in probability.

2. RESULTS

Let $x_1, \cdots, x_n$ be n independent random variables having a common probability density function (p.d.f.) $f(x;\theta)$, $\theta = (\theta_1, \cdots, \theta_s)$. We assume that $f(x;\theta)$ belongs to the s-parameter exponential family. Let $T = (T_1, \cdots, T_s)$ be a minimal sufficient statistic for θ. Let $\hat{\theta} = \hat{\theta}(T) = (\hat{\theta}_1(T), \cdots, \hat{\theta}_s(T))$ be the maximum likelihood estimator of θ so that $\hat{\theta}$ has asymptotically s-variate normal distribution. Let $\hat{\theta}$ have mean $\theta + b(\theta)/n$ and covariance matrix $V(\theta)/n + o(1/n)$.

Let $f(x_1|\hat{\theta})$ be the conditional p.d.f. of x_1 given $\hat{\theta}$ while $f(x_1;\hat{\theta})$ denotes $f(x_1;\theta)$ with θ replaced by $\hat{\theta}$. We assume the regularity conditions on $f(x;\theta)$ as described on p. 194 of Zacks (1971). In addition to these conditions, we assume that $\int f^2(x;\theta)dx$ is finite. Then one can show that

$$f(x_1|\hat{\theta}) = f_1 - \frac{b(\hat{\theta})}{n}\partial f_1 - \frac{1}{2n}\partial'\hat{V}\partial f_1 + o_p(1/n) \qquad (2.1)$$

and

$$f(x_1,x_2|\hat{\theta}) = f_1f_2 - \frac{b'(\hat{\theta})}{n}(f_1\partial f_2 + f_2\partial f_1) - \frac{1}{2n}(f_1\partial'\hat{V}\partial f_2 + 2\partial' f_1\hat{V}\partial f_2 + f_2\partial'\hat{V}\partial f_1) + o_p(1/n) \qquad (2.2)$$

where $f_i = f(x_i;\hat{\theta})$ $i = 1,2$, $\hat{V} = V(\hat{\theta})$, $\partial'(\cdot) = \left(\frac{\partial(\cdot)}{\partial\hat{\theta}_1}, \cdots, \frac{\partial(\cdot)}{\partial\hat{\theta}_s}\right)$ is a row vector, and the subscript 'p' denotes the probability with respect to the joint p.d.f. of x_1 and $\hat{\theta}$. We follow the convention that $\partial(\cdot)$ is an operator acting only on f_1 (or f_2) but not on $\hat{V}$.

Let (z_{i-1},z_i), for $i = 1, \cdots, r$, be the i^{th} class interval in the case of predetermined class intervals and $(z_{i-1}(\hat{\theta}), z_i(\hat{\theta}))$ be the i^{th} class interval when the class boundaries are selected as functions of the parameter estimates. In the first case the class boundaries are fixed and the class probabilities are unknown while in the latter case the class probabilities are predetermined with respect to $f(x_1;\hat{\theta})$. Let I_i denote the i^{th} class in either case. Define $g_i(x_\alpha) = 1$ or 0 according as the α^{th} observation falls in I_i or not. Let $p_i = F(z_i;\hat{\theta}) - F(z_{i-1},\hat{\theta})$ (or $F(z_i(\hat{\theta});\hat{\theta}) - F(z_{i-1}(\hat{\theta});\hat{\theta})$) where $F(z;\hat{\theta}) = \int_{-\infty}^{z} f(x_1;\hat{\theta})dx_1$. Using (2.1) and (2.2) we obtain omitting terms $o_p(1/n)$

$$E\{g_i(x_1)|\hat{\theta}\} = \int g_i(x_1)f(x_1|\hat{\theta})dx_1$$

$$= p_i - \frac{b'(\hat{\theta})}{n}\int_{I_i} \partial f_1 dx_1 - \frac{1}{2n}\int_{I_i} \partial'\hat{V}\partial f_1 dx_1$$

$$E\{g_i(x_1)g_\ell(x_2)|\hat{\theta}\} = p_i p_\ell - \frac{b'(\hat{\theta})}{n}\left(p_i\int_{I_\ell}\partial f_2 dx_2 + p_\ell\int_{I_i}\partial f_1 dx_1\right)$$

$$- \frac{1}{2n}\left(p_i\int_{I_\ell}\partial'\hat{V}\partial f_2 dx_2 + 2\int_{I_i}\partial' f_1 dx_1 \hat{V}\int_{I_\ell}\partial' f_2 dx_2\right.$$

$$\left. + p_\ell\int_{I_i}\partial'\hat{V}\partial f_1 dx_1\right).$$

From these expressions, after some simplification and again omitting terms $o_p(1/n)$, we obtain

$$E\left(\frac{\nu_i}{\sqrt{n}}|\hat{\theta}\right) = \sqrt{np_i}$$

$$\mathrm{Var}\left(\frac{\nu_i}{\sqrt{n}}|\hat{\theta}\right) = p_i(1 - p_i) - \int_{I_i}\partial' f_1 dx_1\hat{V}\int_{I_i}\partial f_2 dx_2 \tag{2.3}$$

$$\mathrm{Cov}\left(\frac{\nu_i}{\sqrt{n}},\frac{\nu_\ell}{\sqrt{n}}|\hat{\theta}\right) = -p_i p_\ell - \int_{I_i}\partial' f_1 dx_1\hat{V}\int_{I_\ell}\partial f_2 dx_2 .$$

Let

$$u_{ij} = \int_{I_i} \frac{\partial f_1}{\partial \hat{\theta}_j} dx_1 \qquad (2.4)$$

$V_{12} = (u_{ij}) = V'_{21}$ and $V_{11} = (\delta_{i\ell} p_i - p_i p_\ell)$ where $\delta_{i\ell} = 1$ or 0 according as $i = \ell$ or not. We assume that rank $V_{12} = s$.

The conditional p.d.f. of $(\nu_1, \cdots, \nu_{r-1})$ given $\hat{\theta}$ is not a multinomial distribution; and also the vectors $(g_1(x_\alpha), \cdots, g_{r-1}(x_\alpha))$, $(\alpha = 1, \cdots, n)$ are not independent in the conditional approach. Thus one cannot use either the approximation of multivariate normal distribution to multinomial distribution or the central limit theorem to prove the asymptotic normality of $(\nu_1, \cdots, \nu_{r-1})/\sqrt{n}$. Another method is to obtain the characteristic function of a linear function of ν's (cf. C. R. Rao (1965) p. 108) and show the convergence to that of a univariate normal distribution. To accomplish this, we first let $n \to \infty$ then take a subsequence of n_1 observations $x_1, \cdots, x_{n_1}$, and find the approximation for the conditional p.d.f. of $x_1, \cdots, x_{n_1}$ given $\hat{\theta}$ as in (2.2) up to terms $o_p(1/n^q)$ where q is the largest integer in $n_1/2$. Then put the n_1 observations into classes to obtain the class frequencies $(\nu_1, \cdots, \nu_{r-1})$ and derive the characteristic function (c.f.) of a linear function of these ν's, properly normalized. It can then be shown that as $n_1/n \to 1$, the c.f. of $(\nu_1, \cdots, \nu_{r-1})/\sqrt{n_1}$ approaches the c.f. of a normal distribution with mean zero and covariance $V_{11} - V_{12}\hat{V}V_{21}$. The quadratic form associated with this normal distribution can be used as the test statistic to test goodness-of-fit to the family $f(x;\theta)$. This statistic is asymptotically distributed as χ^2_{r-1},

$$Q_{r-1}(\nu;\hat{\theta}) = \frac{1}{n}(\nu - np)'(V_{11} - V_{12}\hat{V}V_{21})^{-1}(\nu - np) \overset{\mathcal{L}}{=} \chi^2_{r-1}$$

$$= \sum_i \frac{(\nu_i - np_i)^2}{np_i} + \frac{1}{n} \sum_{j,k} \left\{ \sum_i \left(\frac{\nu_i - np_i}{p_i}\right) u_{ij} \right\} \left\{ \sum_i \left(\frac{\nu_i - np_i}{p_i}\right) u_{ik} \right\} a^{jk}$$

$$= X^2 + Y^2$$

where $(a^{jk}) = (\hat{V}^{-1} - \tilde{J})^{-1}$, $\tilde{J} = \left(\sum_i \frac{1}{p_i} u_{ij}u_{ik}\right)$.

Since the decomposition of $Q_{r-1}(\nu;\hat{\theta})$ into $X^2 + Y^2$ corresponds to writing $B^{-1} = (V_{11} - V_{12}\hat{V}V_{21})^{-1} = V_{11}^{-1} + \{(V_{11} - V_{12}\hat{V}V_{21})^{-1} - V_{11}^{-1}\}$, then to study the limiting distribution of Y^2 it is sufficient to obtain the characteristic roots of $\{(V_{11} - V_{12}\hat{V}V_{21})^{-1}\} B = I - V_{11}^{-1}(V_{11} - V_{12}\hat{V}V_{21}) = V_{11}^{-1}V_{12}\hat{V}V_{21}$. The product $V_{12}\hat{V}V_{21}$ is the estimated covariance matrix of $\sqrt{n}V_{12}(\hat{\theta} - \theta)$ and hence is positive semi-definite of rank s (since rank $V_{12} = s$). Thus only s of the characteristic roots are non zero and are determined from $|V_{12}\hat{V}V_{21} - (1 - \lambda)V_{11}| = 0$. However, since $V_{11} - V_{12}\hat{V}V_{21}$ and V_{11} are positive definite it can be seen (cf. C. R. Rao (1965) p. 56) $|V_{11} - \mu(V_{11} - V_{12}\hat{V}V_{21})| = 0$ has all roots $\mu \geq 1$. From this we obtain $|V_{12}\hat{V}V_{21} - \frac{\mu - 1}{\mu} V_{11}| = 0$ with all $\mu \geq 1$ and hence deduce that $1 - \lambda = \frac{\mu - 1}{\mu}$ with $0 \leq \lambda \leq 1$. Thus $Y^2 \sim (1 - \lambda_1)y_{r-s}^2 + \cdots + (1 - \lambda_s)y_{r-1}^2$ where $y_i \sim N(0,1)$ random variables, and since the characteristic roots of $V_{11}^{-1}B$ are characteristic roots of $I - V_{11}^{-1}V_{12}\hat{V}V_{21}$ then $X^2 \sim y_1^2 + \cdots + y_{r-s-1}^2 + \lambda_1 y_{r-s} + \cdots + \lambda_s y_{r-1}^2$.

The following section contains results of simulation of the distribution of the statistic Q. These simulations support the results (2.1) - (2.3) and the asymptotic normality of $(\nu_1, \cdots, \nu_{r-1})/\sqrt{n_1}$. In the simulations we used all the n observations instead of n_1 described above.

3. SOME NUMERICAL EXAMPLES

In this section we calculate the statistic $Q_{r-1}(\nu;\hat{\theta})$ in the cases of testing the goodness-of-fit of binomial, exponential and normal distributions.

Binomial: $f(x;\theta) = \binom{r}{x}\theta^x(1 - \theta)^{r-x}$, $x = 0, 1, \cdots, r$. Let $x_1, \cdots, x_n$ be a sample of n independent observations. The m.ℓ. estimate $\hat{\theta}$ of θ is $\Sigma x_i/n$ and is a sufficient statistic for θ. Using (2.1) the conditional probability density of x_1 given $\hat{\theta}$ is

$$f(x_1|\hat{\theta}) = f(x_1;\hat{\theta}) - \frac{1}{2n}\frac{\hat{\theta}(1-\hat{\theta})}{r}\frac{d^2f(x_1;\hat{\theta})}{d\hat{\theta}^2} + o_p(\frac{1}{n}).$$

Suppose that the first m classes $(x = 0, 1, \cdots, m-1)$ are pooled while the last $k = r - m - 1$ classes are kept as they are. Then

$$Q_k(\nu;\hat{\theta}) = \sum\frac{(\nu_i - np_i)}{np_i} + \frac{1}{n}\left\{\sum\frac{(\nu_i - np_i)}{p_i}\frac{dp_i}{d\hat{\theta}}\right\}^2 \Big/ \left\{\frac{r}{\hat{\theta}(1-\hat{\theta})} - \sum\frac{1}{p_i}\left(\frac{dp_i}{d\hat{\theta}}\right)^2\right\} .$$

The exact size of the conditional test is evaluated in Table 1 for $r = 3$, $m = 2$, $\hat{\theta} = .6$ at the nominal values $\alpha = .05$ and $\alpha = .025$; critical regions are thus given by $C = \left\{(\nu_1,\nu_2,\nu_3) : Q_2(\nu;\hat{\theta}) \geq \chi^2_{2;\alpha}\right\}$ where $P(\chi^2_2 > \chi^2_{2;\alpha}) = \alpha$.

Table 1: Exact Distribution of Q for Samples from Binomial Distribution

Sample size	$P(\chi^2_2 > \chi^2_{2;\alpha}) = \alpha$	
	.05	.025
10	.04801	.00667
20	.03588	.02511
30	.04224	.02338
40	.04681	.02148
50	.05140	.02151
60	.04467	.02507

In the following two examples the r classes are selected such that the class boundaries are functions of $\hat{\theta}$ and each class has the same probability content under $f(x;\hat{\theta})$, that is,

$$\int_{z_{i-1}(\hat{\theta})}^{z_i(\hat{\theta})} f(x;\hat{\theta})dx = \frac{1}{r} . \qquad (3.1)$$

Exponential Distribution:

$$H_0\colon f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \ge 0, \ \theta \text{ unknown.}$$

The sample mean $\bar{x}$ is a sufficient statistic for θ. Let the r class intervals be $(\bar{x}z_{i-1}, \bar{x}z_i)$, $i = 1, \cdots, r - 1$, with $z_0 = 0$ and $z_r = \infty$, where z_i's are determined from (3.1), giving $z_i = -\log(1 - \frac{i}{r})$. Let

$$u_i = \int_{\bar{x}z_{i-1}}^{\bar{x}z_i} \frac{\partial f(x;\bar{x})}{\partial \bar{x}}\, dx = \frac{1}{\bar{x}} (z_{i-1}e^{-z_{i-1}} - z_i e^{-z_i})$$

and $v_i = \bar{x}u_i$, $(i = 1, \cdots, r)$. After some simplification, we obtain the test statistic as

$$Q_{r-1}(\nu;\bar{x}) = \frac{r}{n} \sum_i (\nu_i - \frac{n}{r})^2 + \frac{r^2}{n} \frac{\{\sum_i(\nu_i - n/r)v_i\}^2}{(1 - r\sum_i v_i^2)}$$

where ν_i = number of x's falling in $(\bar{x}z_{i-1}, \bar{x}z_i)$. Thirty-five hundred samples of size n (= 100) were generated and for each such sample $Q_{r-1}(\nu;\hat{\theta})$, (for r = 4, 6, 8, 10, 12) and X^2 were computed. In Table 2, (n = 100), the first line for each value of r gives the value $\hat{\alpha}$, the proportion of samples in which $Q_{r-1}(\nu;\hat{\theta})$ exceeds $\chi^2_{r-1;\alpha}$, and the second line gives the value $\hat{\alpha}_c$, the proportion of samples in which X^2 exceeds $\chi^2_{r-2;\alpha}$. The sampling distribution of Q is seen to agree with the nominal chi-square distribution for all values of r, while the distribution of X^2 becomes nominal only for sufficiently large r.

Normal Distribution:

$$H_0\colon f(x;\theta) = f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad \mu \text{ and } \sigma^2 \text{ unknown.}$$

Let $\bar{x}$ and s^2 be the sample mean and variance; i.e., $\hat{\theta} = (\bar{x}, s^2)$. Let

Table 2: Simulated sampling distribution of goodness-of-fit statistics Q and X^2 for samples of size 100 from an exponential distribution with unknown mean.

r \ α		.975	.95	.90	.80	.70	.50	.30	.20	.10	.05	.025	.01
4	Q	.982	.948	.900	.794	.702	.501	.302	.206	.102	.056	.025	.009
	X^2	.998	.978	.968	.882	.805	.574	.364	.241	.120	.057	.029	.010
6	Q	.973	.944	.900	.796	.692	.501	.291	.187	.090	.044	.022	.008
	X^2	.984	.960	.912	.823	.739	.528	.317	.204	.097	.045	.024	.008
8	Q	.972	.953	.900	.802	.697	.491	.306	.204	.096	.047	.024	.008
	X^2	.983	.956	.922	.818	.726	.511	.307	.214	.100	.052	.024	.008
10	Q	.975	.950	.897	.801	.694	.483	.276	.185	.090	.043	.020	.008
	X^2	.984	.955	.907	.826	.719	.503	.295	.190	.090	.047	.025	.009
12	Q	.973	.944	.896	.804	.708	.504	.296	.198	.099	.050	.022	.007
	X^2	.973	.948	.912	.813	.729	.516	.315	.200	.101	.058	.020	.008

Entries 1st line: Proportion of Q's > $X^2_{r-1;\alpha}$; 2nd line: Proportion of $X^2 > X^2_{r-2;\alpha}$ in 3500 samples.

$(\bar{x} + z_{i-1}s,\ \bar{x} + z_i s)$, $i = 1, \cdots, r$ be the r class intervals, where the z_i are determined from (3.1) with $z_0 = -\infty$, $z_r = \infty$. From (2.4) we obtain

$$u_{i1} = \frac{1}{s\sqrt{2\pi}} (e^{-z_{i-1}^2/2} - e^{-z_i^2/2}),\quad u_{i2} = \frac{1}{2s^2\sqrt{2\pi}} (z_{i-1}e^{-z_{i-1}^2/2} - z_i e^{-z_i^2/2}).$$

Let $v_{i1} = su_{i1}$, $v_{i2} = s^2 u_{i2}$, $(\hat{V}^{-1} - \tilde{J})^{-1} = (\hat{a}^{jk})$, $j,k = 1,2$; then $\hat{a}^{11} = -(r\Sigma v_{i2}^2 - 2)/D$, $\hat{a}^{12} = \hat{a}^{21} = r\Sigma v_{i1}v_{i2}/D$, $\hat{a}^{22} = -(r\Sigma v_{i1}^2 - 1)/D$ where $D = (r\Sigma v_{i1}^2 - 1)(r\Sigma v_{i2}^2 - 2) - r^2(\Sigma v_{i1}v_{i2})^2$. The test statistic is given by

$$\begin{aligned} Q_{r-1}(\nu;\hat{\theta}) = \frac{r}{n}\Sigma(\nu_i - n/r)^2 + \frac{r^2}{n}[(\Sigma(\nu_i - n/r)v_{i1})^2 \hat{a}^{11} \\ + 2(\Sigma(\nu_i - n/r)v_{i1})(\Sigma(\nu_i - n/r)v_{i2})\hat{a}^{12} \\ + (\Sigma(\nu_i - n/r)v_{i2})^2 \hat{a}^{22}]. \end{aligned}$$

The distribution of this statistic was simulated by Monte Carlo methods. If $y_1, \cdots, y_n$ are independent standard normal variates, then the variables $x_i = (y_i - \bar{y})/s_y$ will have $\bar{x} = 0$, $s_x^2 = 1$. Using these values of x's the statistics $Q_{r-1}(\nu;\bar{x};s^2)$ and X^2 were computed. Using y's the statistic $R = \frac{r}{n}\Sigma(m_i - n/r)^2$ where m_i is the number of y's falling in (z_{i-1}, z_i), $(i = 1, \cdots, r)$ was computed as the test statistic to test H_0: $f(y;\theta) = (1/\sqrt{2\pi})\exp(-y^2/2)$. Table 3 gives the comparison of simulated sampling distributions of the goodness-of-fit statistics Q, R, X^2. For fixed r, the first and second lines give the proportion of samples in which $Q_{r-1}(\nu;\bar{x};s^2)$ and R exceed $X^2_{r-1;\alpha}$, respectively; the third line gives the proportion of samples in which X^2 exceeds $X^2_{r-3;\alpha}$ for r = 4, 6, 8, 10, 12. As anticipated, the sampling distributions of Q and R agree well with their nominal distributions for all r.

Table 3: Simulated sampling distribution of goodness-of-fit statistics for samples of size 100 from a normal distribution with unknown mean and variance.

r \ α		.975	.95	.90	.80	.70	.50	.30	.20	.10	.05	.025	.01
	Q	.978	.949	.915	.809	.710	.510	.292	.184	.095	.051	.028	.010
4	R	.978	.955	.898	.830	.722	.509	.301	.207	.103	.050	.028	.012
	X^2	.999	.999	.999	.999	.964	.822	.529	.389	.209	.107	.055	.024
	Q	.977	.953	.900	.809	.715	.503	.296	.204	.094	.046	.023	.008
6	R	.972	.959	.913	.811	.707	.513	.312	.197	.105	.056	.032	.014
	X^2	.995	.984	.957	.900	.814	.609	.392	.251	.130	.061	.033	.014
	Q	.974	.946	.897	.792	.693	.490	.283	.185	.087	.041	.026	.010
8	R	.981	.956	.905	.813	.696	.505	.298	.199	.102	.055	.027	.011
	X^2	.988	.970	.928	.855	.765	.551	.340	.233	.111	.058	.027	.013
	Q	.970	.950	.899	.788	.685	.485	.288	.191	.095	.047	.023	.008
10	R	.978	.957	.910	.817	.716	.508	.306	.199	.105	.052	.027	.012
	X^2	.982	.968	.914	.817	.733	.541	.341	.221	.110	.054	.028	.011
	Q	.973	.950	.896	.789	.680	.494	.289	.196	.101	.053	.025	.009
12	R	.980	.957	.914	.817	.707	.500	.297	.210	.114	.058	.021	.011
	X^2	.982	.961	.907	.815	.725	.523	.327	.222	.119	.059	.031	.013

4. POWER COMPARISONS

The X^2 test of goodness-of-fit (cf. Cochran, 1952) is most commonly used when we do not have a clear-cut alternative in mind, and are not in a position to make computations of the power. In their discussion of the limiting distribution of (1.2) Chernoff and Lehmann (1954) expressed the hope that the power of the test would be increased if the maximum likelihood estimates based on the original observations are employed. In general, as pointed out by Chibisov (1971) power comparisons by analytic methods are very difficult and we have therefore resorted to simulation studies of the power functions of the three statistics under consideration. We have examined the alternatives including double exponential, mixtures of double exponential, N(0,4) and N(0,9) with N(0,1) variables. In each of these cases $Q_{r-1}(\nu;\hat{\theta})$ has larger power when compared to its competitors. The mixtures of normal variables with different variances were suggested by Tukey (1960).

In order to compare (1.1) with the others on a competitive basis we first remove the requirement of predetermined class boundaries, using class boundaries of the form $\bar{x} \pm z_i s$ for all three test statistics and with equal predicted class frequencies for both (1.2) and $Q_{r-1}(\nu;\hat{\theta})$. Predicted class frequencies for (1.1) are based on a sample mean and variance estimated from the class frequencies. Following Cramer (1946) these multinomial maximum likelihood estimates (m.m.l.e.) are approximated by moment estimates after replacing each observation in the i^{th} class by the midpoint a_i of this class. Midpoints of the extreme classes are here defined by

$$a_r = \sqrt{(r - \sum_{2}^{r-1} a_i^2)/2}$$

and $a_1 = -a_r$.

The results given in Table 4 reveal consistent and substantial differences in power of the three statistics, the power of $Q_{r-1}(\nu;\hat{\theta})$ being greatest and the power

Table 4: Comparison of power of the statistics Q, X^2, and $\tilde{R}$ for 1000 samples of size 80.

Sample Mixture	No. of Classes	.10 Q	.10 X^2	.10 $\tilde{R}$	.05 Q	.05 X^2	.05 $\tilde{R}$	.01 Q	.01 X^2	.01 $\tilde{R}$
DE-80 N(0,1)-0	4	.641	.608	.361	.525	.487	.216	.322	.257	.063
	6	.675	.570	.364	.585	.467	.251	.387	.244	.103
	8	.691	.517	.365	.580	.404	.245	.369	.200	.089
DE-40 N(0,1)-40	4	.400	.355	.222	.283	.235	.113	.124	.094	.021
	6	.417	.304	.159	.333	.215	.082	.182	.078	.024
	8	.436	.283	.157	.340	.181	.090	.164	.077	.024
DE-0 N(0,1)-80	4	.106	.108	.118	.046	.058	.051	.015	.017	.008
	6	.095	.094	.098	.036	.050	.042	.008	.009	.007
	8	.093	.090	.102	.050	.056	.046	.012	.011	.007
N(0,1)-60 N(0,4)-20	4	.298	.266	.164	.193	.171	.085	.077	.054	.017
	6	.321	.212	.121	.221	.130	.060	.083	.032	.010
	8	.332	.204	.117	.242	.106	.053	.103	.032	.010
N(0,1)-70 N(0,9)-10	4	.615	.573	.327	.507	.453	.187	.351	.261	.039
	6	.675	.507	.162	.591	.397	.088	.427	.217	.022
	8	.712	.463	.130	.633	.351	.071	.483	.164	.013

Critical values for the distribution of X^2 are obtained from Dahiya and Gurland (1972).

of (1.1) being least. Power comparisons were made for the mixtures of double exponential (DE) variable X when the density function is given by $g(x;\theta) = \exp(-|x|/\theta)/2$ with $\theta = 2$ and that of $N(0,1)$ variables; and also for the mixtures of $N(0,1)$ variables with $N(0,4)$ and $N(0,9)$ variables.

DISCUSSION

A practical disadvantage of the classical test statistic (1.1) is the awkwardness of the associated estimation problem, which has resulted in the customary practice of calculating (1.2) in place of (1.1). A practical disadvantage of (1.2) is that unless a large number of class intervals are used the asymptotic distribution of the test statistic depends on the null hypothesis. The proliferation of tables of critical values required to accommodate even the most commonly hypothesized parametric families can be avoided through the use of $Q_{r-1}(\nu;\hat{\theta})$. One further advantage of $Q_{r-1}(\nu;\hat{\theta})$ is the improvement in power that seems likely to occur in most applications as a result of the increase in degrees of freedom.

ACKNOWLEDGMENTS

We are thankful to the referee for his comments which resulted in improving the earlier versions of this paper and to Professor Jarslav Hájek for his comments on the need for the estimator of θ to be a sufficient statistic.

REFERENCES

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum likelihood estimates in X^2 tests for goodness-of-fit. Ann. Math. Statist. 25, 579-586.

CHIBISOV, D. M. (1971). Certain chi-square type tests for continuous distributions. Theor. Prob. and Appl. 16, 1-22.

COCHRAN, W. G. (1952). The X^2 test of goodness-of-fit. Jour. Amer. Statist. Assoc. 47, 315-345.

CRAMER, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

DAHIYA, R. C. and GURLAND, J. (1972). Pearson chi-squared test of fit with random intervals. Biometrika 59, 147-153.

FISHER, R. A. (1950). The significance deviations from expectations in a Poisson series. Biometrics 6, 17-24.

LEVENE, H. (1949). On a matching problem arising in genetics. Ann. Math. Statist. 20, 91-94.

MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. Ann. Math. Statist. 14, 217-226.

MOORE, D. S. (1971). A chi-square statistic with random cell boundaries. Ann. Math. Statist. 42, 147-156.

RAO, C. R. and CHAKRAVARTI, I. M. (1956). Some small sample tests of significance for a Poisson distribution. Biometrics 12, 264-282.

RAO, C. R. (1965). Linear Statistical Inference and Its Applications. John Wiley and Sons, Inc.

ROY, A. R. (1956). On chi-squared statistics with variable intervals. Department of Statistics, Stanford University, Technical Report.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions, Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling, Stanford University Press, 448-485.

VITHAYASAI, C. (1971). Exact critical values of the Hardy-Weinberg test statistic for two alleles. Biometrics and Computer Science Branch, Department of Fisheries and Forestry, Canada, Technical Report.

WATSON, G. S. (1957). The chi-squared goodness-of-fit test for normal distributions. Biometrika 44, 336-348.

WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. J. R. Statist. Soc., Ser. B 20, 44-61.

ZACKS, S. (1971). The Theory of Statistical Inference. John Wiley and Sons, Inc.