Biometrika (1977), 64, 1, pp. 115–21
Printed in Great Britain

Distribution of the chi-squared test in nonstandard situations

BY LUCIANO MOLINARI

Department of Statistics, University of California, Berkeley, and Eidgenössische Technische Hochschule, Zürich

SUMMARY

The distribution of the χ^2 test under the null hypothesis is studied, when the parameters are estimated by the method of moments. A general formula, applicable also to other situations, is given. Three examples are studied in more detail and numerical results are given, indicating how unsafe it can be to use a χ^2 distribution with a number of degrees of freedom smaller than or equal to the number of cells.

Some key words: Chi-squared test; Distribution of quadratic forms; Goodness of fit.

1. INTRODUCTION

The standard theorem on the asymptotic distribution of the χ^2 test,

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i},$$

is given, e.g. by Cramér (1946, p. 427) or Rao (1973, p. 391), and rests on several assumptions about the true distribution of the observations and the estimators of any unknown parameters. These assumptions are not satisfied in practice mainly for the following reasons:

(*a*) estimators are obtained from the ungrouped data rather than from the grouped observations;

(*b*) the cells are often random, i.e. are determined by a first look at the observations;

(*c*) the method of estimation is not efficient.

Points (*a*) and (*b*) were discussed by Chernoff & Lehmann (1954) and by Watson (1958). In both cases it is safe to consider X^2 as having a χ^2 distribution with $k-1$ degrees of freedom, the level of significance being then underestimated.

A discussion of (*c*) for the method of moments is given. It seems that the formulae derived could be easily applied to other estimates. It turns out that in some cases, not at all 'pathological' ones, it would be very unsafe to use a χ^2 distribution with $k-1$ degrees of freedom.

The following notation will be used in this paper. Let $f(x, \theta_1, \ldots, \theta_q) = f(x, \theta)$ be the probability density of each observation under the null hypothesis. Let $\Delta_1, \ldots, \Delta_k$ be the χ^2 cells and

$$p_i = p_i(\theta) = \int_{\Delta_i} f(x, \theta)\, dx \quad (i = 1, \ldots, k)$$

their probabilities.

Denote by $n_1, \ldots, n_k$ the observed frequencies and let $V = (v_1, \ldots, v_k)^{\mathrm{T}} = (v_i)^{\mathrm{T}}$ be the vector with components

$$v_i = (n_i - np_i)/\sqrt{(np_i)} \quad (i = 1, \ldots, k), \quad \Lambda = (\sqrt{p_1}, \ldots, \sqrt{p_k})^{\mathrm{T}},$$

$$M = ((m_{ij})) = \left(\left(\frac{1}{\sqrt{p_i}} \frac{\partial p_i}{\partial \theta_j}\right)\right) \quad (i = 1, \ldots, k; j = 1, \ldots, q).$$

Notice that $J = M^{\mathrm{T}} M$ is the information matrix of the grouped data and that $M^{\mathrm{T}} \Lambda = 0$.

Finally, let $X \sim F$ mean that F is the distribution function of the random variable X.

2. Main theorem and some examples

THEOREM. *Let X have a p-dimensional normal distribution with expectation* 0 *and covariance matrix* Σ; *let A be a $p \times p$ symmetric matrix. The quadratic form X^TAX is then distributed as* $\lambda_1 y_1^2 + \ldots + \lambda_p y_p^2$ *with* $y_1, \ldots, y_p$ *independent* $N(0,1)$ *and* $\lambda_1, \ldots, \lambda_p$ *the not necessarily different eigenvalues of* $A\Sigma$.

Remark. This theorem is well known and is quoted, e.g. in a slightly different form, by Searle (1971, pp. 57, 69), who also gives several references. Therefore, I will not give the proof; the theorem is, however, sometimes misquoted in the form that X^TAX has a χ^2 distribution if and only if $A\Sigma$ is idempotent. This is correct if Σ is not singular, but not in general. Also, the theorem is mostly given and proved under the assumption that Σ is not singular. The following corollary is an immediate consequence of the theorem.

COROLLARY. *Consider the χ^2 test*

$$X^2 = \sum_{i=1}^{k} (n_i - n\hat{p}_i)^2/(n\hat{p}_i),$$

where the $\hat{p}_i$'s are either given or estimated. If, asymptotically, $V \sim N_k(0, \Sigma)$, *then* $X^2 = V^TV$ *is asymptotically distributed as* $\Sigma\lambda_i y_i^2$ *with* $\lambda_1, \ldots, \lambda_k$ *the eigenvalues of* Σ *and* $y_1, \ldots, y_k$ *independent* $N(0,1)$.

The following three examples are taken from the classical theory of χ^2 tests:

Example 1. If $p_1, \ldots, p_k$ are given, V is asymptotically normal with $\Sigma = I - \Lambda\Lambda^T$, Σ being idempotent with rank $k-1$, X^2 is χ^2 with $k-1$ degrees of freedom.

Example 2. For the estimation of $p_1, \ldots, p_k$ by maximum likelihood from the grouped data (Rao, 1973, p. 392), $\Sigma = (I - MJ^{-1}M^T)(I - \Lambda\Lambda^T)(I - MJ^{-1}M^T)$, and Σ is idempotent with rank $k-q-1$, where q is the number of estimated parameters. Therefore, X^2 has a χ^2 distribution with $k-q-1$ degrees of freedom.

Example 3. For the estimation of $p_1, \ldots, p_k$ by maximum likelihood from the ungrouped data, it was proved by Chernoff & Lehmann (1954) that

$$X^2 \sim \chi^2_{k-q-1} + \lambda_1 y_1^2 + \ldots + \lambda_q y_q^2,$$

with $y_1, \ldots, y_q$ independent $N(0,1)$ and $0 \leqslant \lambda_1, \ldots, \lambda_q \leqslant 1$ the eigenvalues of a certain matrix. In order to reject the null hypothesis it is safe to take X^2 distributed as χ^2 with $k-1$ degrees of freedom.

3. A General Formula

Assume now that the parameter θ is estimated by some method giving an estimate $\bar{\theta}$ and corresponding estimates of p_i as

$$\bar{p}_i = \int_{A_i} f(x, \bar{\theta})\, dx = \bar{p}_i(\bar{\theta}).$$

To find the asymptotic distribution of X^2 the asymptotic distribution of

$$\bar{V} = (\bar{v}_i) = ((n_i - n\bar{p}_i)/\surd(n\bar{p}_i))$$

is required. Let us make the following assumption.

Assumption A. Assume that, asymptotically,

$$\left\{\frac{n_1 - np_1}{\surd(np_1)}, \ldots, \frac{n_k - np_k}{\surd(np_k)}, \surd n(\bar{\theta}_1 - \theta_1), \ldots, \surd n(\bar{\theta}_q - \theta_q)\right\}^T \sim N_{k+q}\left\{0, \begin{pmatrix} I - \Lambda\Lambda^T & C \\ C^T & T \end{pmatrix}\right\}$$

for some C and T. Notice that $C^T\Lambda = 0$.

The assumption will usually be satisfied in practice, because many statistics like central and raw moments and quantiles, and therefore functions of them, are asymptotically normal. If the assumption is satisfied and all functions involved are regular enough, e.g. with continuous derivatives, the following result can be easily obtained, in almost exactly the same way as by Rao (1973, p. 394):

$$\Sigma = (I - I)\begin{bmatrix} I - \Lambda\Lambda^{\mathrm{T}} & CM^{\mathrm{T}} \\ MC^{\mathrm{T}} & MTM^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} I \\ -I \end{bmatrix} = I - \Lambda\Lambda^{\mathrm{T}} - (MC^{\mathrm{T}} + CM^{\mathrm{T}}) + MTM^{\mathrm{T}}. \quad (1)$$

Because $C^{\mathrm{T}}\Lambda = M^{\mathrm{T}}\Lambda = 0$ and $\Lambda\Lambda^{\mathrm{T}} = 1$ the matrix $I - \Sigma = \Lambda\Lambda^{\mathrm{T}} + (MC^{\mathrm{T}} + CM^{\mathrm{T}}) - MTM^{\mathrm{T}}$ has always an eigenvalue one, while for its rank we have

$$r(I-\Sigma) \leqslant r(\Lambda\Lambda^{\mathrm{T}}) + r\{(C - MT)M^{\mathrm{T}}\} + r(MC^{\mathrm{T}}) \leqslant 2q+1.$$

Therefore assuming $k > 2q$, which will generally be true in practice, Σ has always at least $k-2q-1$ eigenvalues equal to 1 and one eigenvalue 0; that is, asymptotically,

$$X^2 = \overline{V}^{\mathrm{T}}\overline{V} \sim \sum_{i=1}^{2q} \lambda_i y_i^2 + \sum_{i=2q+1}^{k-1} y_i^2,$$

with $y_1, \dots, y_{k-1}$ independent $N(0, 1)$ and $\lambda_1, \dots, \lambda_{2q}$ the eigenvalues of Σ possibly different from 0 and 1.

4. One-parameter families of densities

In a family $f(x, \theta)$ of densities let θ be estimated by the method of moments.

With $\mu = E(X) = \int xf(x,\theta)\,dx = g(\theta)$, $\sigma^2 = \mathrm{var}(X) = \sigma^2(\theta)$, and assuming that μ is a function of θ, we define $\theta = f_1(\mu)$ and its estimate $\overline{\theta} = f_1(m)$, where $m = \overline{X} = m_1 = n^{-1}\Sigma X_i$.

Assumption *A* is certainly satisfied under weak assumptions on f and f_1.

From, for example,

$$\mathrm{cov}\{(n_i - np_i)/\surd(np_i)\} = \frac{1}{\surd p_i}\int_{\Delta_i}(u-\mu)f(u,\theta)\,du = A_i(\theta),$$

and by setting $G^{\mathrm{T}} = (A_1, \dots, A_k)$ and omitting for simplicity the argument θ, we obtain

$$I - \Sigma = \Lambda\Lambda^{\mathrm{T}} + \frac{df_1}{d\mu}(MG^{\mathrm{T}} + GM^{\mathrm{T}}) - \left(\frac{df_1}{d\mu}\sigma\right)^2 MM^{\mathrm{T}} \quad (2)$$

with rank $\leqslant 3$ and an eigenvalue 1. With

$$a = \frac{df_1}{d\mu}G^{\mathrm{T}}M, \quad b = \frac{df_1}{d\mu}G^{\mathrm{T}}G, \quad c = \frac{df_1}{d\mu}M^{\mathrm{T}}M, \quad v = \frac{df_1}{d\mu}\sigma^2,$$

the nontrivial eigenvalues of $I - \Sigma$ can be found from the equation

$$\lambda^2 - (2a - vc)\lambda + a^2 - bc = 0.$$

Notice that

$$a^2 - bc = \left(\frac{df_1}{d\mu}\right)^2\{(G^{\mathrm{T}}M)^2 - (G^{\mathrm{T}}G)(M^{\mathrm{T}}M)\} \leqslant 0,$$

with equality holding, for $df_1/d\mu \neq 0$, if and only if G and M are linearly dependent. If this is not the case the roots of the above equation are of different sign and one of the eigenvalues of Σ is > 1.

Example 4. For a normal distribution with σ known,

$$f(x,\theta) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\} \quad (\mu = \theta).$$

It is easy to verify that $G = \sigma^2 M$, $v = \sigma^2$ and $1-\lambda = 1-M^{\mathrm{T}}M\sigma^2 < 1$.

Example 5. For a gamma distribution with p known,

$$f(x,\theta) = \frac{\theta^p}{\Gamma(\theta)} e^{-\theta x} x^{p-1} \quad (p > 0,\, \theta > 0).$$

Here $\mu = p/\theta$ and $\sigma^2 = p/\theta^2$. Also $G = -M$, $v = -1$ and $1-\lambda = 1-M^{\mathrm{T}}Mp/\mu^2 < 1$. In both examples the moment estimate coincides with the maximum likelihood estimate based on the ungrouped data. Therefore, the result $0 < 1-\lambda < 1$ could have been predicted following Chernoff & Lehmann (1954).

Example 6. For a gamma distribution with α known,

$$f(x,\theta) = \frac{\alpha^\theta}{\Gamma(\theta)} e^{-\alpha x} x^{\theta-1} \quad (\alpha > 0,\, \theta > 0).$$

Here M and G are linearly independent, giving for Σ one eigenvalue > 1.

5. Two-parameter families of densities

Let $f(x, \theta_1, \theta_2)$ be a family of densities. Similarly to § 4, define

$$\mu = E(X) = \mu_1(\theta_1, \theta_2), \quad \sigma^2 = \operatorname{var}(X) = \mu_2(\theta_1,\theta_2), \quad \theta_1 = f_1(\mu_1,\mu_2), \quad \theta_2 = f_2(\mu_1,\mu_2)$$

with moment estimates as

$$\bar\theta_1 = f_1(m_1, m_2), \quad \bar\theta_2 = f_2(m_1, m_2),$$

where

$$m_1 = m = \bar X, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar X)^2 = s^2.$$

Again, Assumption A appears to be satisfied in many practical situations. The asymptotic joint distribution of m_1 and m_2 is given by Rao (1973, p. 437). Following the method of § 4, it can be shown that equation (1) becomes

$$\Sigma = I - \Lambda\Lambda^{\mathrm{T}} - (MFG^{\mathrm{T}} + GF^{\mathrm{T}}M^{\mathrm{T}}) + MF\Omega F^{\mathrm{T}}M^{\mathrm{T}}, \tag{3}$$

where

$$F = \left(\left(\frac{\partial f_i}{\partial \mu_j}\right)\right), \quad G = \begin{bmatrix} A_1 & B_1 - 2\mu A_1 \\ \vdots & \vdots \\ A_k & B_k - 2\mu A_k \end{bmatrix},$$

$$B_i = \frac{1}{\sqrt{p_i}} \int_{\Delta_i} (u - \mu^2 - \mu_2) f(u, \theta_1, \theta_2)\, du = B_i(\theta_1, \theta_2), \quad \Omega = \begin{bmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{bmatrix}.$$

In equation (3), without loss of generality we can set $\theta_1 = \mu_1$ and $\theta_2 = \mu_2$, so that $F = I$. Then we have following cases.

(*a*) If the columns of G are linearly dependent on those of M, that is $G = MR$, then the rank of $I-\Sigma$ is $\leqslant 3$ and Σ has two nontrivial eigenvalues λ_1, λ_2 which are $\leqslant 1$ if and only if $R^{\mathrm{T}}+R-\Omega$ is positive-semidefinite.

(*b*) If only one of the columns of G is a linear combination of those of M, say $G_1 = a_1 M_1 + a_2 M_2$, then $I-\Sigma$ has rank $\leqslant 4$ and its nontrivial eigenvalues are the roots of a cubic polynomial whose last coefficient is $(2a_1 - \Omega_{11})\Delta$, with $\Delta > 0$. If this is positive either all roots are negative or two are positive and one is negative. See § 6 for an example.

(*c*) If the columns of G and M are independent, then $I-\Sigma$ has rank 5 and its four nontrivial eigenvalues are the roots of a polynomial of degree 4 whose last coefficient is

$$\det\{(MG)^{\mathrm{T}}(MG)\} > 0.$$

Therefore, all the roots are positive, or all are negative, or two are positive and two are negative. See § 6 for an example.

6. Three numerical examples

For a normal distribution, with

$$\theta_1 = \mu, \quad \theta_2 = \sigma^2, \quad f(x,\theta_1,\theta_2) = \frac{1}{\sqrt{(2\pi\theta_2)}}\exp\{-\tfrac{1}{2}(x-\theta_1)^2/\theta_2\},$$

we obtain $\Omega_{11} = \sigma^2$, $\Omega_{12} = 0$ and $\Omega_{22} = 2\sigma^4$. It is easy to verify that $G = M\Omega$ and

$$\Sigma = I - \Lambda\Lambda^{\mathrm{T}} - MM^{\mathrm{T}}.$$

Therefore, Σ has $k-3$ eigenvalues 1, one eigenvalue 0, and two eigenvalues $0 < \lambda_1 \leqslant \lambda_2 < 1$.

Table 1. *Eigenvalues λ_1 and λ_2 for the normal distribution as functions of the mean, M, and the standard deviation, S*

M	$S = 0{\cdot}5$		$S = 1{\cdot}0$		$S = 2{\cdot}0$		$S = 3{\cdot}0$		$S = 4{\cdot}0$	
1·0	0·269	0·764	0·100	0·693	0·040	0·673	0·030	0·672	0·041	0·686
3·0	0·245	0·458	0·078	0·165	0·024	0·267	0·026	0·405	0·052	0·515
5·0	0·245	0·458	0·077	0·145	0·025	0·103	0·037	0·272	0·067	0·434
7·0	0·245	0·458	0·078	0·165	0·024	0·267	0·026	0·405	0·052	0·515
9·0	0·269	0·764	0·100	0·693	0·040	0·673	0·030	0·672	0·041	0·686

Table 1 gives λ_1 and λ_2 as a function of μ and σ^2 for $\mu = 1, 3, 5, 7, 9$, $\sigma = 0{\cdot}5, 1, 2, 3, 4$, for a χ^2 test based on cells $(-\infty, 1], (1, 2], \ldots, (8, 9], (9, \infty)$.

For a gamma distribution with

$$\theta_1 = \alpha, \quad \theta_2 = \gamma, \quad f(x,\theta_1,\theta_2)f(x,\alpha,\gamma) = \frac{\alpha^\gamma}{\Gamma(\gamma)}e^{-\alpha x}x^{\gamma-1},$$

we have

$$E(X) = \mu_1 = \mu = \gamma/\alpha, \quad \operatorname{var}(X) = \mu_2 = \sigma^2 = \gamma/\alpha^2, \quad \mu_3 = 2\gamma/\alpha^3, \quad \mu_4 - \mu_2^2 = (2\gamma^2+6\gamma)/\alpha^4.$$

Define

$$p_i = p_{i\gamma} = \int_{\Delta_i} f(x,\alpha,\gamma)\,dx.$$

It is easy to show that $\partial p_i/\partial\alpha = -\sqrt{p_{i\gamma}}A_i$ and case (*b*), § 5 applies.

Therefore, Σ has $k-4$ eigenvalues 1, one eigenvalue 0, and eigenvalues λ_1, λ_2 and λ_3. With the notation of (*b*), § 5, we obtain

$$a_1 = -\partial\mu/\partial\alpha = \gamma/\alpha^2, \quad \Omega_{11} = \gamma/\alpha^2, \quad 2a_1 - \Omega_{11} = \gamma/\alpha^2 > 0,$$

and all nontrivial eigenvalues of $I-\Sigma$ are negative, or one is negative and two are positive. For reasons of continuity the same alternative must hold for all α and λ. It turns out that the second alternative is true. Therefore, the nontrivial eigenvalues of Σ are $0 < \lambda_1 \leqslant \lambda_2 < 1 < \lambda_3$. Table 2 gives λ_1, λ_2 and λ_3 for $\mu = 1, 2, 3, 4, 6, 8$, σ^2 and χ^2 cells as for Table 1.

Table 2. *Eigenvalues λ_1, λ_2 and λ_3 for the gamma distribution as functions of M and S*

M	$S = 0·5$	$S = 1·0$	$S = 2·0$	$S = 3·0$	$S = 4·0$
1·0	0·184 0·482 1·02	0·034 0·282 1·14	0·053 0·568 1·43	0·044 0·916 2·03	0·089 0·977 2·66
2·0	0·224 0·485 1·05	0·055 0·145 1·19	0·034 0·278 1·36	0·030 0·721 1·58	0·041 0·912 1·99
3·0	0·235 0·472 1·02	0·065 0·151 1·11	0·026 0·204 1·35	0·021 0·600 1·49	0·029 0·845 1·75
4·0	0·239 0·467 1·01	0·069 0·149 1·06	0·025 0·205 1·24	0·018 0·563 1·41	0·024 0·806 1·62
6·0	0·242 0·462 1·01	0·075 0·155 1·03	0·027 0·337 1·09	0·021 0·636 1·20	0·027 0·818 1·38
8·0	0·244 0·467 1·00	0·089 0·397 1·01	0·038 0·649 1·03	0·037 0·805 1·07	0·047 0·894 1·18

Table 3. *Eigenvalues λ_1, λ_2, λ_3 and λ_4 for the log normal distribution as functions of M and S*

M	$S = 0·5$	$S = 1·0$	$S = 2·0$	$S = 3·0$	$S = 4·0$
1·0	0·133 0·393 1·00 1·10	0·154 0·495 1·00 1·93	0·193 0·999 1·03 22·4	0·453 1·000 1·08 201	0·712 1·000 1·15 1190
2·0	0·201 0·471 1·00 1·08	0·050 0·154 1·00 1·64	0·043 0·831 1·00 4·23	0·131 0·996 1·01 15·7	0·300 1·000 1·04 544
3·0	0·223 0·471 1·00 1·05	0·055 0·146 1·00 1·24	0·031 0·593 1·00 2·35	0·058 0·936 1·00 5·79	0·154 0·997 1·01 14·3
4·0	0·232 0·467 1·00 1·03	0·062 0·148 1·00 1·14	0·029 0·475 1·00 1·60	0·039 0·876 1·00 3·15	0·098 0·974 1·00 6·55
6·0	0·239 0·464 1·00 1·01	0·074 0·166 1·00 1·06	0·031 0·514 1·00 1·20	0·036 0·855 1·00 1·61	0·073 0·952 1·00 2·55
8·0	0·243 0·472 1·00 1·01	0·093 0·445 1·00 1·02	0·046 0·767 1·00 1·06	0·059 0·926 1·00 1·25	0·095 0·968 1·00 1·63

For a log normal distribution, we define

$$\theta_1 = \alpha, \quad \theta_2 = \beta^2, \quad f(x, \alpha, \beta^2) = \frac{1}{x\sqrt{(2\pi\beta^2)}} \exp\{-\tfrac{1}{2}(\log x - \alpha)^2/\beta^2\}.$$

By some tedious computations it can be shown that case (c) in § 5 applies. The columns of G and M are independent and can be computed by the exponential and cumulative normal functions only. Of the four nontrivial eigenvalues of $I-\Sigma$ two are positive and two are negative. Therefore, Σ has $k-5$ eigenvalues 1, one eigenvalue 0, and four eigenvalues

$$0 < \lambda_1 \leqslant \lambda_2 < 1 < \lambda_3 \leqslant \lambda_4.$$

Table 3 gives λ_1, λ_2, λ_3 and λ_4, respectively.

In conclusion, we note that the χ^2 test is still a standard technique during a preliminary analysis, whose users might sometimes be unaware of the underlying assumptions or unable to get efficient estimates. The results should be seen as a warning against using χ^2 tables with $k-1$ or $k-q-1$ degrees of freedom to reject the null hypothesis. For instance the 5 % rejection point based on 9 degrees of freedom is 16·9, while the approximate correct level corresponding to 16·9 in the log normal situation with $M = 4$ and $S = 3$ is about 11 %, and the correct rejection point for 5 % is 21.

This work was prepared with the support of the Swiss National Science Foundation and the U.S. Office of Naval Research.

References

CHERNOFF, H. & LEHMANN, E. L. (1954). The use of the maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25**, 579–86.

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

RAO, C. R. (1973). *Linear Statistical Inference*, 2nd edition. New York: Wiley.

SEARLE, S. R. (1971). *Linear Models*. New York: Wiley.

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

[*Received November* 1975. *Revised June* 1976]