On the Pearson-Fisher Chi-Squared Theorem

Aniello Buonocore and Enrica Pirozzi

University of Naples Federico II
Department of Mathematics and Applications “Renato Caccioppoli”
Via Cintia, Complesso Monte S. Angelo, 80126 - Napoli, Italy

Copyright © 2014 Aniello Buonocore and Enrica Pirozzi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

This paper outlines with modern notations the proof of Karl Pearson on the distribution of the statistic on which the chi-squared goodness of fit test with simple null hypothesis is based. It is then shown in details a proof due Ronald Aylmer Fisher, which in the opinion of the authors has not received the proper attention for its elegance and its immediate generalization to the case of composite null hypothesis. Finally, a new proof based on the Principle of mathematical induction is provided by making use of widely known probabilistic results.

Mathematics Subject Classification: 62-02; 62F03; 62G10

Keywords: Degree of freedom, goodness of fit test, multinomial distribution, multinormal distribution

1 Introduction

The chi-squared test for goodness of fit was designed and described by Karl Pearson in the paper [4] that, starting from the title, is rather cryptic and difficult to interpret. The proof provided by Pearson, in the opinion of many statistical readers, is affected by the employed terminology and by many allusions that appear archaic and not easily understandable. Despite this, no one

---

1 The paper was re-published in the collection: Karl Pearson’s Early Statistical Papers, Cambridge University Press, London (1948).
can doubt the final sentence of Robert Lewis Plackett paper [5] here entirely quoted: “The many positive aspects of originality and usefulness far outweigh the defects of presentation. Pearson’s 1900 paper on chi-squared is one of the great monuments of twentieth century statistics.”

The interested reader can refer to the above cited paper of Plackett to find informations on the first decade of the statistics career of Pearson and on the state of the art at that time; furthermore it also contains a precious work of interpretation of many details contained in the paper by Pearson.

For completeness, but also to agree on notations, we outline in the following the chi-squared test for a simple null hypothesis. With \( n \) and \( k \) we denote two integers: their meaning will be clarified in this section and it will be unchanged in the whole paper. Let \( Y \) be a random variable with unknown distribution function from which we take a simple random sample \( Y = (Y_1, Y_2, \ldots, Y_n) \). The cited test is used to verify the simple null hypothesis \( H_0 \) that a totally specified distribution function \( F \) assigned to \( Y \) is not significantly in contrast with the experimental evidence contained in \( Y \). Hence, we list the following essential points.

a) The range of \( Y \) is partitioned into \( k \) disjoint sets \( E_1, E_2, \ldots, E_k \).

b) For \( i = 1, 2, \ldots, k \), the random variable \( N_i \) represents the number of sample observations belonging to \( E_i \) with probability

\[
p_{0,i} := \mathbb{P}(Y \in E_i \mid H_0) = \int_{E_i} dF(y) > 0,
\]

with \( \sum_{i=1}^{k} p_{0,i} = 1 \).

c) The statistic \( Q_{k,n} := \sum_{i=1}^{k} \frac{(N_i - np_{0,i})^2}{np_{0,i}} \) is distributed as a chi-squared random variable with \( k - 1 \) degrees of freedom, asymptotically with respect to \( n \):

\[
Q_{k,n} \xrightarrow{n \to \infty} Q_k \sim \chi^2(k - 1).
\]

d) For \( n \gg k \), let \( q_{k,n} \) be the value of \( Q_{k,n} \) obtained by means of the random sample, \( H_0 \) is rejected if \( \mathbb{P}(Q_k > q_{k,n}) \) is less than the preassigned significance level.

The item c) of the previous list consists of the main theoretical result given in [4]; its proof and the extension to the case of the composite null hypothesis are the matter of the present paper.
2 The Pearson’s proof

In this section, we give the essentials of the Pearson proof in modern notations, pointing out original formulas, numbers and sentences, in squared brackets, to simplify comparisons with the mentioned Pearson’s paper. (For the sake of the reader we point out that Pearson used $n + 1$ in place of $k$ and $N$ in place of $n$.)

1) Let $r$ be an integer and $X = (X_1, X_2, \ldots, X_r)$ a random vector with non-singular multinormal distribution $N_r(\mathbf{0}, \mathbf{V})$ of order $r$. A rather cumbersome analytical proof allowed him to determine that

$$X\mathbf{V}^{-1}X^t \sim \chi^2(r).$$

See [4, formulas (i.), (ii.), (v.) and (vi.)].

2) For $i = 1, 2, \ldots, k$, the variable $N_i$ defined in b) of Section 1, under $H_0$ hypothesis, has binomial distribution with parameters $n$ and $p_{0,i}$:

$$N_i \sim \text{Bin}(n, p_{0,i}).$$

Furthermore, under $H_0$ hypothesis, the random vector $\mathbf{N} = (N_1, N_2, \ldots, N_k)$ has multinomial distribution having mean values vector

$$\mathbf{m} = n(p_{0,1}, p_{0,2}, \ldots, p_{0,k}),$$

and covariance matrix

$$\Sigma = (\sigma_{i,j})_{i,j=1,2,\ldots,k}$$

with

$$\sigma_{i,j} = \begin{cases} np_{0,i}(1 - p_{0,j}), & \text{if } j = i, \\ -np_{0,i}p_{0,j}, & \text{if } j \neq i. \end{cases}$$

See [4, formulas (vii.) and (viii.)].

3) It holds $\sum_{i=1}^{k} N_i = n$.

Refer to the position $e = m' - \mathbf{m}$ and the successive formula $e_1 + e_2 + \cdots + e_{n+1} = 0$ in [4, p. 160].

Then, the matrix $\Sigma$ has rank $k - 1$ so that $|\Sigma| = 0$ and therefore $\Sigma$ is not invertible. However, it is possible to leave out one of the $N_i$, for example $N_k$.

Pearson in [4, pp. 160–161] wrote: “Hence only $n$ of the $n + 1$ errors are variables; the $n + 1$ is determined when the first $n$ are known, and using formula (ii.) we treat only of $n$ variables.”

After that, the random non-singular vector $\mathbf{N}^* = (N_1, N_2, \ldots, N_{k-1})$ is considered having multinomial distribution with mean values vector $\mathbf{m}^*$ obtained from $\mathbf{m}$ omitting the last component and whose covariance matrix $\Sigma^*$ is obtained from $\Sigma$ omitting the last row and the last column.
4) Let \( A := (p_{0,k}^{-1})_{i,j=1,2,...,k-1} \) and \( B = (b_{i,j})_{i,j=1,2,...,k-1} \)
given by:
\[
b_{i,j} := \begin{cases} p_{0,i}^{-1}, & \text{if } j = i, \\ 0, & \text{if } j \neq i. \end{cases}
\]

Pearson proved that:
\[
(\Sigma^*)^{-1} = \frac{1}{n} (A + B).
\]

See [4, formulas (xiii.) and (xiv)].

5) The quadratic form
\[
Q := (N^* - m^*)(\Sigma^*)^{-1}(N^* - m^*)^t
\]

coincides with the statistic \( Q_{k,n} \) defined in the item c) of Section 1.

Indeed:
\[
Q = \sum_{i=1}^{k-1} \frac{(N_i - np_{0,i})^2}{np_{0,i}} + \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \frac{(N_i - np_{0,i})(N_j - np_{0,j})}{\sqrt{np_{0,k}} \sqrt{np_{0,k}}}
\]
\[
= \sum_{i=1}^{k-1} \frac{(N_i - np_{0,i})^2}{np_{0,i}} + \left[ \sum_{i=1}^{k-1} \frac{(N_i - np_{0,i})}{\sqrt{np_{0,k}}} \right]^2
\]
\[
= \sum_{i=1}^{k-1} \frac{(N_i - np_{0,i})^2}{np_{0,i}} + \frac{1}{np_{0,k}} \left[ \sum_{i=1}^{k-1} (N_i - np_{0,i}) \right]^2
\]
\[
= \sum_{i=1}^{k} \frac{(N_i - np_{0,i})^2}{np_{0,i}} \equiv Q_{k,n}.
\]

See [4, formula (xv.) and the subsequent sentence “where the summation is now to extend to all \((n+1)\) errors, and not merely to the first \(n\)”, p. 163].

6) Applying the multidimensional Central Limit Theorem, one can state that
\[
N^* - m^* \xrightarrow{n \to \infty} N_{k-1}(0, \Sigma^*).
\]

In [4] this result is not explicitly mentioned, but it is clear from the list of “stages of our investigation” on p. 164 and from the illustrations that appear at the end of the paper.

The result of the previous item 1) is then applied, asymptotically with respect to \( n \), to the vector \( N^* - m^* \), leading to \( Q_{k,n} \xrightarrow{n \to \infty} \chi^2(k-1) \).
3 The Fisher’s proof

In this section, following the lines of [3], we recall the proof given by Ronald Aylmer Fisher in [1].

Let \( r \) be an integer, \( I_r \) the identity matrix of order \( r \) and let \( Z = (Z_1, Z_2, \ldots, Z_r) \) be a random vector with multinormal distribution \( \mathcal{N}_r(0, I_r) \) of order \( r \) (note that the components of \( Z \) are independent). The purpose is to use a geometric argument to determine the distribution of the random variable

\[
U := Z_1^2 + Z_2^2 + \cdots + Z_r^2.
\]

We may imagine the values \( Z_1, Z_2, \ldots, Z_r \) of any given sample of \( Z \) as the co-ordinates of a point \( P \) in the \( r \)-dimensional Euclidean hyperspace \( \mathbb{R}^r \). In such a way \( U \) represents the square of the distance of \( P \) from the origin \( O \) and \( U = |OP|^2 \) is therefore unchanged by any rotation of the co-ordinate orthonormal axes. Furthermore, the joint probability density function of the components of \( Z \) is proportional to \( e^{-|OP|^2/2} \) and it remains constant on the surface of the \( r \)-dimensional hypersphere with radius \( \sqrt{u} = |OP| \); therefore, \( f_U(u) \) is obtained by integrating of such a constant between the hyperspheres located in \( u \) and in \( u + du \):

\[
f_U(u) = c e^{-|OP|^2/2} \frac{d}{d|OP|} |OP|^r = c e^{-u/2} \frac{d}{du} u^{r/2},
\]

with \( c \) a suitable constant. Applying the derivative operator and imposing the normalization condition, the following result is easily obtained:

\[
f_U(u) = \frac{1}{2^{r/2} \Gamma(r/2)} e^{-u/2} u^{r/2-1}.
\]

Hence, \( U \sim \chi^2(r) \).

We now consider the case in which the components of \( Z \) are subject to \( s \leq r \) linear (independent) constraints. (A linear constraint for \( Z \) is generally represented by the equation \( a_1Z_1 + a_2Z_2 + \cdots + a_rZ_r = 0 \), with \( (a_1, a_2, \ldots, a_r) \in \mathbb{R}^r \).) Each of these constraints defines a hyperplane of \( \mathbb{R}^r \) containing \( O \), say \( \pi \), and \( Z \) is forced to belong to it. Of course, the intersection of the generic hypersphere of \( \mathbb{R}^r \) with \( \pi \) is a hypersphere of the space \( \mathbb{R}^{r-1} \). The result of the \( s \) linear constraints will be to achieve a hypersphere of \( \mathbb{R}^{r-s} \) from a generic hypersphere of \( \mathbb{R}^r \). Therefore, the probability density function of \( U \) is immediately obtained from (1) by replacing \( r \) with \( r - s \). In other words, if there are \( s \leq r \) linear constraints for \( Z \), one has

\[
U = \sum_{i=1}^{r} Z_i^2 \sim \chi^2(r - s).
\]
We now consider $V = (V_1, V_2, \ldots, V_k)$ consisting of $k$ independent random variables having Poisson distribution with parameter $np_0,i$, for $i = 1, 2, \ldots, k$: $V_i \sim \Pi(np_0,i)$. With $(n_1, n_2, \ldots, n_k, n) \in \mathbb{N}^{k+1}$ and $N = \sum_{i=1}^{k} V_i$, at first, one has:
\[
\begin{align*}
\mathbb{P}(V_1 = n_1, V_2 = n_2, \ldots, V_k = n_k, N = n) &= \prod_{i=1}^{k} \frac{(np_0,i)^{n_i}}{n_i!} e^{-np_0,i} \\
&= e^{-n} n^k \prod_{i=1}^{k} \frac{(p_0,i)^{n_i}}{n_i!}.
\end{align*}
\] (3)

Note that, although in the left-hand side of (3) $N$ appears as a random variable, the joint distribution is singular because $n = \sum_{i=1}^{k} n_i$. On the other hand, the marginal distribution of $N$ is easy to determine because $N$ is the sum of $k$ independent random variables having Poisson distribution and $\mathbb{E}(N) = n$, whereby $N \sim \Pi(n)$. From this and from (3), one obtains:
\[
\begin{align*}
\mathbb{P}(V_1 = n_1, V_2 = n_2, \ldots, V_k = n_k | N = n) &= \frac{n!}{n_1! n_2! \cdots n_k!} \prod_{i=1}^{k} (p_0,i)^{n_i}.
\end{align*}
\]

In this way, one can see that the distribution of $V$, under the condition $N = n$, is the same of $N$ defined in the item 2) of Section 2.

Now, the vector $\tilde{Z} = (\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_k)$ with
\[
\tilde{Z}_i := \frac{N_i - np_0,i}{\sqrt{np_0,i}},
\]
for $i = 1, 2, \ldots, k$, is such that:
\[
\sum_{i=1}^{k} \sqrt{np_0,i} \tilde{Z}_i = 0. \quad (4)
\]

So, applying the multidimensional Central Limit Theorem, with $r = k$, we have that $\tilde{Z} \overset{n \rightarrow \infty}{\sim} Z$ and, taking into account the linear constraint (4), from (2) one has:
\[
\sum_{i=1}^{k} \tilde{Z}_i^2 = \sum_{i=1}^{k} \frac{(N_i - np_0,i)^2}{np_0,i} \equiv Q_{k,n} \overset{n \rightarrow \infty}{\sim} \chi^2(k-1). \]

The power of this proof lies also on the circumstance that if $s < k-1$ additional linear constraints are imposed to the vector $\tilde{Z}$, as previously argued, each of ones acts decreasing the number of freedom degrees of one unity and, without further investigations, one can write:
\[
\sum_{i=1}^{k} \frac{(N_i - np_0,i)^2}{np_0,i} \equiv Q_{k,n} \overset{n \rightarrow \infty}{\sim} \chi^2(k-1-s). \]
This is the case of the composite null hypothesis when the estimation of $s$ unknown parameters $\theta_1, \theta_2, \ldots, \theta_s$ of the distribution $F$ assigned to $Y$ is required by means of the same sample. In such a case, as highlighted by Fisher in [2], it is appropriate to use the maximum likelihood estimators applied to the vector $N$, i.e. solving, for $j = 1, 2, \ldots, s$, the following equations

$$
\sum_{i=1}^{k} \frac{1}{p_{0,i}} \frac{\partial p_{0,i}}{\partial \theta_j} N_i = 0 \iff \sum_{i=1}^{k} \frac{1}{\sqrt{p_{0,i}}} \frac{\partial p_{0,i}}{\partial \theta_j} \tilde{Z}_i = 0,
$$

being each one a linear constraint for $\tilde{Z}$.

4. An alternative proof

In this section a proof based on the Principle of mathematical induction is given. For this purpose, it is useful to state some results.

**Lemma 4.1.** Let $k > 2$. Using the same positions given in Section 1, we set

$$L := N_{k-1} + N_k \quad \text{and} \quad M := p_{0,k-1}N_{k-1} - p_{0,k-1}N_k.$$  

Then, asymptotically with respect to $n$, we obtain:

(i) the random vector

$$\tilde{N} = (N_1, N_2, \ldots, N_{k-2}, L, M)$$

has multinormal distribution;

(ii) $M$ is independent on each component of the random vector

$$N' = (N_1, N_2, \ldots, N_{k-2}, L);$$

(iii) $M$ is independent on the random vector $N'$.

**Proof.** By means of the De Moivre's Theorem, asymptotically with respect to $n$, each component of the vector $N$ has normal distribution; in the same way, asymptotically with respect to $n$, the random variables $L$ and $M$ have normal distribution, being each one a linear combination of $N_{k-1}$ and $N_k$.

Now, denoted by $O$ the null rectangular matrix of order $[2, k-2]$, we set:

$$C := \begin{pmatrix} 1 & p_{0,k} \\ 1 & -p_{0,k-1} \end{pmatrix} \quad \text{and} \quad D := \begin{pmatrix} I_{k-2} & O \\ O & C \end{pmatrix}.$$  

Since $\tilde{N} = ND$, then (i) is a direct consequence of De Moivre's Theorem applied to $N$. 
We now prove (ii). Denote \(\mathbb{C}(\cdot, \cdot)\) the covariance between two random variables; for \(i = 1, 2, \ldots, k - 2\) one has:
\[
\mathbb{C}(M, N_i) = \mathbb{C}(p_{0,k}N_{k-1} - p_{0,k-1}N_k, N_i) \\
= p_{0,k}\mathbb{C}(N_{k-1}, N_i) - p_{0,k-1}\mathbb{C}(N_k, N_i) \\
= -np_{0,k}p_{0,k-1}p_{0,i} + np_{0,k-1}p_{0,k}p_{0,i} \\
= 0.
\]

Therefore:
\[
\mathbb{C}(M, L) = \mathbb{C}(M, N_{k-1}) + \mathbb{C}(M, N_k) \\
= \mathbb{C}(p_{0,k}N_{k-1} - p_{0,k-1}N_k, N_{k-1}) + \mathbb{C}(p_{0,k}N_{k-1} - p_{0,k-1}N_k, N_k) \\
= p_{0,k}[\mathbb{C}(N_{k-1}, N_{k-1}) + \mathbb{C}(N_{k-1}, N_k)] \\
- p_{0,k-1}[\mathbb{C}(N_k, N_{k-1}) + \mathbb{C}(N_k, N_k)] \\
= np_{0,k}p_{0,k-1}[(1 - p_{0,k}) - p_{0,k}] - np_{0,k-1}p_{0,k}[-p_{0,k-1} + (1 - p_{0,k})] \\
= 0.
\]

Then (ii) follows recalling that two uncorrelated normal variables are also independent.

Part (iii) is a direct consequence of (ii) by recalling that, asymptotically with respect to \(n\), \(M\) and the vector \(\mathbf{N}\) have normal distributions.

Note that, denoting by \(\mathbb{E}(\cdot)\) and \(\mathbb{D}^2(\cdot)\) the mean value and the variance of a random variable, respectively, \(M\) is such that:
\[
\mathbb{E}(M) = p_{0,k}nN_{k-1} - p_{0,k-1}N_k \\
= p_{0,k}np_{0,k-1} - p_{0,k-1}np_{0,k} = 0, \quad (5)
\]

and
\[
\mathbb{D}^2(M) = p_{0,k}^2\mathbb{D}^2(N_{k-1}) + p_{0,k-1}^2\mathbb{D}^2(N_k) - 2p_{0,k-1}p_{0,k}\mathbb{C}(N_{k-1}, N_k) \\
= np_{0,k-1}p_{0,k}(p_{0,k-1} + p_{0,k}). \quad (6)
\]

**Lemma 4.2.** Setting
\[
S := \left[\frac{(N_{k-1} - np_{0,k-1}) + (N_k - np_{0,k})}{n(p_{0,k-1} + p_{0,k})}\right]^2 - \sum_{i=k-1}^{k} \frac{(N_i - np_{0,i})^2}{np_{0,i}} \\
\equiv \sum_{i=k-1}^{k} \left[ \frac{(N_i - np_{0,i})^2}{n(p_{0,k-1} + p_{0,k})} - \frac{(N_i - np_{0,i})^2}{np_{0,i}} \right] \\
+ 2\frac{(N_{k-1} - np_{0,k-1})(N_k - np_{0,k})}{n(p_{0,k-1} + p_{0,k})},
\]
one has
\[
\left[\frac{M}{\sqrt{\mathbb{D}^2(M)}}\right]^2 = -S.
\]
Proof. By a direct calculation, we obtain:

\[
S = - \left[ \sum_{i=k-1}^{k} \frac{p_{0,2k-1-i} (N_{i} - np_{0,i})^2}{p_{0,i} n (p_{0,k-1} + p_{0,k})} - 2 \frac{(N_{k-1} - np_{0,k-1}) (N_{k} - np_{0,k})}{n (p_{0,k-1} + p_{0,k})} \right]
\]

\[
= - \left[ \sum_{i=k-1}^{k} \sqrt{\frac{p_{0,2k-1-i}}{p_{0,i}}} \frac{(-1)^i (N_{i} - np_{0,i})^2}{\sqrt{n (p_{0,k-1} + p_{0,k})}} \right] \]

\[
= - \left[ \frac{p_{0,k} N_{k-1} - p_{0,k-1} N_k}{\sqrt{np_{0,k-1} p_{0,k} (p_{0,k-1} + p_{0,k})}} \right]^2 \]

\[
= \left[ \frac{M}{\sqrt{\delta^2(M)}} \right]^2.
\]

We are now able to give the inductive proof.

**Theorem 4.3.** Using the same positions given in Section 1, let \( k > 1 \) the number of classes partitioning the range of \( Y \). One has:

\( Q_{k,n} \xrightarrow{n \to \infty} \chi^2(k-1) \).

Proof. The basis of induction is obtained for \( k = 2 \) and it is a known result. Indeed, from

\[
N_2 = n - N_1 \quad \text{and} \quad p_{0,2} = 1 - p_{0,1},
\]

follows

\[
Q_{2,n} = \frac{(N_1 - np_{0,1})^2}{np_{0,1}} + \frac{(N_2 - np_{0,2})^2}{np_{0,2}} = \left[ \frac{N_1 - np_{0,1}}{\sqrt{np_{0,1} (1 - p_{0,1})}} \right]^2.
\]

Being \( N_1 \sim \text{Bin}(n, p_{0,1}) \), then \( Q_{2,n} \xrightarrow{n \to \infty} \chi^2(1) \) is a consequence of De Moivre’s Theorem recalling that the square of a standard normal random variable has chi-squared distribution with one freedom degree.

Assume now that \( Q_{k-1,n} \xrightarrow{n \to \infty} \chi^2(k-2) \). We set

\[
E'_i := \begin{cases} E_i, & \text{for } i = 1, 2, \ldots, k-2, \\ E_{k-1} \cup E_k, & \text{for } i = k-1, \end{cases}
\]

and

\[
p'_{0,i} := \begin{cases} p_{0,i}, & \text{for } i = 1, 2, \ldots, k-2, \\ p_{0,k-1} + p_{0,k}, & \text{for } i = k-1. \end{cases}
\]

By using \( N'_i \) given in (ii) of Lemma 4.1 and noting that

\[
\sum_{i=1}^{k-1} N'_i = n \quad \text{and} \quad \sum_{i=1}^{k-1} p'_{0,i} = 1,
\]
the induction hypothesis allows us to state that:

$$\sum_{i=1}^{k-1} \frac{(N_i' - np_{0,i}')}{{np_{0,i}'}}^2 \xrightarrow{n \to \infty} \chi^2(k - 2).$$  \tag{7}$$

On the other hand, with $S$ given in Lemma 4.2, we have

$$\sum_{i=1}^{k-1} \frac{(N_i' - np_{0,i}')}{{np_{0,i}'}}^2 = \sum_{i=1}^{k-2} \frac{(N_i - np_{0,i})^2}{{np_{0,i}}^2} + \frac{(N_{k-1}' - np_{0,k-1}')^2}{{np_{0,k-1}'}^2}$$

$$= \sum_{i=1}^{k} \frac{(N_i - np_{0,i})^2}{{np_{0,i}}^2} + S,$$

and by making use of Lemma 4.2, finally we can write:

$$\sum_{i=1}^{k} \frac{(N_i - np_{0,i})^2}{{np_{0,i}}^2} = \left[ \frac{M}{\sqrt{D_2(M)}} \right]^2 + \sum_{i=1}^{k-1} \frac{(N_i' - np_{0,i}')^2}{{np_{0,i}'}^2}.$$

Hence, taking into account (i) and (iii) of Lemma 4.1 and (5), (6) and (7), applying the additivity Theorem for independent chi-squared random variables, we obtain:

$$\sum_{i=1}^{k} \frac{(N_i - np_{0,i})^2}{{np_{0,i}}^2} \equiv Q_{k,n} \xrightarrow{n \to \infty} \chi^2(k - 1).$$

The next lemma is useful for obtaining the generalization to the case of composite hypothesis.

**Lemma 4.4.** Let $r$ be an integer, $I_r$ the identity matrix of order $r$, $Z = (Z_1, Z_2, \ldots, Z_r)$ a random vector with multinormal distribution $N_r(0, I_r)$ of order $r$ (note that the components of $Z$ are independent). If the components of $Z$ are subject to $r - 1$ (independent) linear constraints one has:

$$U := Z_1^2 + Z_2^2 + \cdots + Z_r^2 \sim \chi^2(1).$$

**Proof.** Consider an orthonormal system of axes $Oz_1 \ldots z_r$ in the space $\mathbb{R}^r$. The vector $Z$ locates a point $P$ of $\mathbb{R}^r$ and $U$ is the square of its distance from the origin. We now denote with $\alpha_1$ the hyperplane defined by the first linear constraint which $Z$ is subject, and with $B_r(u) \subset \mathbb{R}^r$, $u > 0$, the hypersphere centered in $O$ with radius $\sqrt{u}$. Then $\alpha_1 \cap B_r(u)$ is an hypersphere $B_{r-1}(u) \subset \mathbb{R}^{r-1}$ and the probability density function of $|OP|^2$ evaluated in $u$ is obtained exclusively by means of the points of $B_{r-1}(u)$. Repeating the same arguments for everyone of remaining $r - 2$ linear constraints, the sphere $B_r(u)$ finally reduces to the set of two points having the distance $\sqrt{u}$ from the origin and belonging to a straight line containing the origin. The probability density function of $U$ is the same of that of $|OP|^2$ and, taking into
account the distribution form of $Z$, it is constant over every such hypersphere. On the other hand such a constant value is preserved intersecting $B_r(u)$ with hyperplanes $Z_2 = 0, Z_3 = 0, \ldots, Z_r = 0$; therefore the distribution of $U$ coincides with that of $Z^2_1$.

**Remark 4.5.** Removing $l \leq r - 1$ linear constraints leads to a distribution $\chi^2(l + 1)$. Indeed, proceeding in the same way of the proof of Lemma 4.4 one is able to claim that $U$ has the same distribution of $Z^2_1$.

**Theorem 4.6.** Using the same positions as in Section 1, let $s$ the number of unknown parameters in $F$ and $k > s + 1$ the number of classes partitioning the range of $Y$. If maximum likelihood estimators are used, then

$$Q_{k,n} \underset{n \rightarrow \infty}{\sim} \chi^2(k - 1 - s). \quad (8)$$

**Proof.** Consider the case $k = s + 2$.

For any, but assigned, $p_1, \ldots, p_{s+2}$ such that $\sum_{i=1}^{s+2} p_i = 1$, Theorem 4.3 allows us to write:

$$Q_{s+2,n} \underset{n \rightarrow \infty}{\sim} \chi^2(s + 1).$$

On the other hand setting, for $i = 1, \ldots, s + 2$,

$$\hat{Z}_i := \frac{N_i - np_i}{\sqrt{np_i(1 - p_i)}} \underset{n \rightarrow \infty}{\sim} N(0, 1),$$

without further conditions on $\sum_{i=1}^{s+2} N_i$ the independence of $(N_1, \ldots, N_{s+2})$ is ensured.

In this context we have

$$\hat{Q}_{s+2,n} = \sum_{i=1}^{s+2} \frac{(N_i - np_i)^2}{np_i(1 - p_i)} = \sum_{i=1}^{s+2} \hat{Z}_i^2 \underset{n \rightarrow \infty}{\sim} \chi^2(s + 2).$$

Remark 4.5 and the linear constraint

$$\sum_{i=1}^{s+2} N_i = n \iff \sum_{i=1}^{s+2} \sqrt{np_i(1 - p_i)} \hat{Z}_i = 0$$

imply that $\hat{Q}_{s+2,n} \underset{n \rightarrow \infty}{\sim} \chi^2(s + 1)$. Hence $Q_{s+2,n} \underset{n \rightarrow \infty}{\sim} \hat{Q}_{s+2,n}$. Lemma 4.4 and, for $j = 1, \ldots, s$, the linear constraints

$$\sum_{i=1}^{s+2} \sqrt{1 - p_{0,i} \frac{\partial p_{0,i}}{\partial \theta_j}} \hat{Z}_i = 0$$

by which the probabilities of the classes are obtained according to the null hypothesis, allow us to claim that:

$$Q_{s+2,n} \underset{n \rightarrow \infty}{\sim} \hat{Q}_{s+2,n} \underset{n \rightarrow \infty}{\sim} \chi^2(s + 1 - s) \equiv \chi^2(1).$$

Hence the (8) is valid and it can be assumed as the induction basis.

For $k > s + 2$ the proof follows by induction along the lines of Theorem 4.3. \qed
References


[4] K. Pearson, On a Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables in such that it can be reasonably supposed to have arisen in Random Sampling, *Philosophical Magazine Series 5*, **50**(302) (1900), 157–175.


Received: April 15, 2014