ON THE SEQUENTIAL CHI-SQUARE TEST

M. P. Savelov

Lomonosov Moscow State University
Moscow, RUSSIA

e-mail: savelovmp@gmail.com

Abstract

Chi-square test based on the Pearson statistics is used to check whether frequencies of the finite number of outcomes correspond with their hypothetical probabilities. A sequential version of this test is based on several Pearson statistics computed for nested samples; this version was considered in several papers. Formulas for the joint distributions of the Pearson statistics for nested samples are very cumbersome. Here we present exact formulas for the covariance of two Pearson statistics computed for nested samples and asymptotic relations connecting the error probabilities of one- and two-dimensional chi-square tests.

1 Introduction

Suppose that independent identically distributed trials with \( m \) outcomes having probabilities \( p_1, \ldots, p_m \) are performed. Denote by \( \nu_j(n) \) the frequency of the \( j \)-th outcome in the first \( n \) trials. Pearson statistics \( \chi^2(n) := \sum_{j=1}^m \frac{(\nu_j(n) - np_j)^2}{np_j} \) is widely used to test the hypothesis \( H(p) : \) “outcomes have law \( p = (p_1, \ldots, p_m) \)”, because \( \chi^2(n) \) converges in distribution to the standard \( \chi^2_{m-1} \) at \( n \to \infty \). So, if \( \pi_{m-1}(\alpha) \) is the \((1-\alpha)\)-quantile of \( \chi^2_{m-1} \), then the rule, accepting \( H(p) \), if \( \chi^2(n) < \pi_{m-1}(\alpha) \), has type I error \( \approx \alpha \).

A sequential version of the chi-square test is based on the statistics \( (\chi^2(n_1), \ldots, \chi^2(n_r)) \) for \( n_1 < \cdots < n_r \); it was studied by Zakharov et al. [5] and others [2, 3]. In the sequential version \( H(p) \) is rejected if and only if \( \chi^2(n_k) > \pi_{m-1}(\alpha_k) \) for all \( k = 1, \ldots, r \).

For true hypothesis \( H(p) \) we find (Theorems 1, 2) the covariance \( \text{cov}(\chi^2(n_1), \chi^2(n_2)) \) and an asymptotic relation between error probabilities \( P\{A_1\} \approx \alpha_1, P\{A_2\} \approx \alpha_2 \) and \( P\{A_1A_2\}, A_j = \{\chi^2(n_j) > \pi_{m-1}(\alpha_j)\} \), as \( n_1, n_2 \to \infty, n_1/n_2 \to \text{const}, \alpha_1, \alpha_2 \to 0 \).

2 Main results

Theorem 1. If \( n_1 \leq n_2 \), then

\[
\text{cov}(\chi^2(n_1), \chi^2(n_2)) = \frac{(2(n_1 - 1)(m - 1) - m^2 + \sum_{k=1}^m \frac{1}{p_k})}{n_2}.
\]

Corollary 1. \( \text{cov}(\chi^2(n_1), \chi^2(n_2)) \) is nonincreasing in \( n_2 \) for fixed \( n_1 \leq n_2 \).

Note that if \( \frac{n_1}{n_2} \to \infty \), then the covariance tends to 0. It follows from the Theorem 1 that the variance \( D\chi^2(n) = 2(m - 1) + \frac{1}{n} \left(2 - 2m - m^2 + \sum_{k=1}^m \frac{1}{p_k}\right) \), this expressions was obtained in [4].
Zakharov, Sarmanov and Sevastyanov [5] have derived asymptotic formulas for the error probabilities of the sequential chi-square test in the form of integrals containing Infeld functions and exponential functions. By means of these formulas we obtain explicit asymptotic expressions for the error probability in the case \( r = 2 \).

**Theorem 2.** Let \( m \geq 3, n_1, n_2 \to \infty, n_1/n_2 \to c^2, c \in (0, 1) \), \( \alpha_1 := \lim P\{\chi^2(n_1) > \pi_{m-1}(\alpha_1)\}, \alpha_2 := \lim P\{\chi^2(n_2) > \pi_{m-1}(\alpha_2)\} \) and \( \alpha = \lim P\{\chi^2(n_1) > \pi_{m-1}(\alpha_1), \chi^2(n_2) > \pi_{m-1}(\alpha_2)\} \). If \( \alpha_1, \alpha_2 \to +0 \) and \( \sqrt{\ln \alpha_2/\ln \alpha_1} = P \in (c, 1/c) \), then

\[
\alpha = \frac{(1 - c^2)^{1/2} \cdot P^{n_2-1} \cdot (-\ln \alpha_1)^n_2-2 \cdot Q^{-1/2(1-c^2)}}{2c^{m-1/2} \sqrt{\pi} \Gamma\left(\frac{m-1}{2}\right)(P - c)(1 - cP)} \cdot (1 + o(1)),
\]

where

\[
Q = \frac{P^{2(m-3)(1-P)}}{(\Gamma\left(\frac{m-1}{2}\right))^4(1-P)^7} \cdot (-\ln \alpha_1)^{m-3}(2-c(P+P^{-1})) \cdot \alpha_1^{-2(P^2-2P+c+1)}.
\]

**Corollary 2.** If conditions of the Theorem 2 hold and \( \alpha_2 = \alpha_1 \) (i.e. \( P = 1 \)), then

\[
\alpha = \frac{(1 - c^2)^{1/2} \cdot (P^{n_2-1})^{1+c}}{2c^{m-1/2} \sqrt{\pi}(1 - c)^2} \cdot (-\ln \alpha_1)^{m-3}(2-c)^2(\alpha_1)^{2+c} \cdot (1 + o(1)).
\]

Note that \( \frac{2}{1+c} \in (1, 2) \) since \( c \in (0, 1) \).

### 3 Acknowledgments

The author is grateful to A.M. Zubkov for problem statement and constant attention.

### References


