A SET OF ASYMPTOTICALLY INDEPENDENT STATISTICS OF POLYNOMIAL FREQUENCIES CONTAINING THE PEARSON STATISTIC

M.P. Savelov
Lomonosov Moscow State University
Moscow, RUSSIA
e-mail: savelovmp@gmail.com

Abstract

Consider a multinomial scheme with N outcomes. We suggest a set of $N-2$ new statistics which along with the Pearson statistic are jointly asymptotically independent. Limit distributions of these statistics are found.

Keywords: data science, Pearson statistic, multinomial scheme

1 Introduction

Suppose that independent identically distributed trials with N outcomes having probabilities $p_1, \ldots, p_N$ ($p_1 + \ldots + p_N = 1$, $p_i > 0$) are performed. Denote by $\nu_i = \nu_i(n)$ the frequency of the j-th outcome in the first n trials. The Pearson statistics $X(n) := \sum_{i=1}^{N} \frac{(\nu_i - np_i)^2}{np_i}$ is widely used to test the hypothesis $H()$: "probabilities of outcomes constitute a vector $= (p_1, \ldots, p_N)$", because the distribution of $X(n)$ for $n \to \infty$ converges to the standard chi-square distribution with $N-1$ degrees of freedom denoted here by χ^2_{N-1} [1]. We suggest a set of $N-2$ new statistics (polar coordinates of the vector of frequencies in some basis) which along with the Pearson statistic are jointly asymptotically independent. Limit distributions of these statistics are found.

2 Main results

Theorem 1. *Suppose that vector* $p = (\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_N})$ *and* $v_1, v_2, \ldots, v_{N-1}$ *form an orthonormal basis of* $\mathbb{R}^N$. *Put*

$$Y_i = \left(\left(\frac{\nu_1 - np_1}{\sqrt{np_1}}, \frac{\nu_2 - np_2}{\sqrt{np_2}}, \ldots, \frac{\nu_N - np_N}{\sqrt{np_N}} \right), v_i \right), 1 \le i \le N-1$$

$$\alpha_{N-2}(n) = \operatorname{arcctg} \frac{Y_{N-1}}{\sqrt{Y_1^2 + \ldots + Y_{N-2}^2}} = \arccos \frac{Y_{N-1}}{\sqrt{Y_1^2 + \ldots + Y_{N-1}^2}},$$

$$\alpha_{N-3}(n) = \operatorname{arcctg} \frac{Y_{N-2}}{\sqrt{Y_1^2 + \ldots + Y_{N-3}^2}} = \arccos \frac{Y_{N-2}}{\sqrt{Y_1^2 + \ldots + Y_{N-2}^2}},$$

$$\alpha_{N-4}(n) = \operatorname{arcctg} \frac{Y_{N-3}}{\sqrt{Y_1^2 + \ldots + Y_{N-4}^2}} = \arccos \frac{Y_{N-3}}{\sqrt{Y_1^2 + \ldots + Y_{N-3}^2}},$$

$$\ldots$$

$$\alpha_2(n) = \operatorname{arcctg}\frac{Y_3}{\sqrt{Y_1^2+Y_2^2+Y_3^2}} = \arccos\frac{Y_3}{\sqrt{Y_1^2+\ldots+Y_4^2}},$$

$$\alpha_1(n) = 2\operatorname{arcctg}\frac{Y_2+\sqrt{Y_1^2+Y_2^2}}{Y_2} = \begin{cases} \arccos\frac{Y_2}{\sqrt{Y_1^2+Y_2^2}}, & Y_1 \geq 0, \\ 2\pi - \arccos\frac{Y_2}{\sqrt{Y_1^2+Y_2^2}}, & Y_1 < 0, \end{cases}$$

and if the denominator of the formula for α_i is zero we put $\alpha_i = 0$. Then the limit distribution of $(X(n), \alpha_1(n), \ldots, \alpha_{N-2}(n))$ is a distribution of a random vector $\rho^2, \alpha_1, \ldots, \alpha_{N-2}$ with independent components such that $\rho^2 \sim \chi^2_{N-1}$, $\alpha_1 \sim U[0, 2\pi]$ and for all $i \in \{2, \ldots, N-2\}$ a density $p_{\alpha_i}(x)$ of a random variable α_i distribution is defined by

$$p_{\alpha_i}(x) = \begin{cases} \frac{\Gamma\left(\frac{i+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{i}{2}\right)} \sin^{i-1} x, & 0 \leq x \leq \pi. \\ 0. \end{cases}$$

For $N \geq 3$ put

$$T_1(n) = \operatorname{arctg}\left(\frac{\sqrt{p_3}(p_1+p_2+p_3)}{\sqrt{p_1 p_2}} \cdot \frac{\nu_2 p_1 - \nu_1 p_2}{(\nu_1+\nu_2)p_3 - (p_1+p_2)\nu_3}\right),$$

$$T_2(n) = \arccos\frac{\nu_N - np_N}{(1-p_N)\sqrt{np_N X(n)}}.$$

Example 1. If $N \geq 3$ then α_1 is defined and $T_1(n) = \arctan(\tan(\alpha_1(n))$. Hence the limit distribution of a vector $(X(n), T_1(n))$ is a distribution of a vector (ζ_1, ζ_2) with independent components such that $\zeta_1 \sim \chi^2_{N-1}$, $\zeta_2 \sim U[-\frac{\pi}{2}, \frac{\pi}{2}]$. If $N \geq 4$ then α_1 and $\alpha_{N-2} = \pi - T_2(n)$ are defined, and the limit distribution of a vector $(X(n), T_1(n), T_2(n))$ is a distribution of a vector $(\zeta_1, \zeta_2, \zeta_3)$ with independent components such that $\zeta_1 \sim \chi^2_{N-1}$, $\zeta_2 \sim U[-\frac{\pi}{2}, \frac{\pi}{2}]$. The density of ζ_3 is equal to $\frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{N-2}{2}\right)} \sin^{N-3} x$ for $x \in [0, \pi]$ and 0 otherwise.

3 Acknowledgments

The author is grateful to A.M. Zubkov for the constant attention.

References

[1] Borovkov A.A. (1984). *Mathematical statistics (in Russian)*. Nauka, Moscow.