

# Some Theorems on the Asymptotic Distribution of the Statistic $\chi^2$

1970

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-197001.43001>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

UDC 519.2

**Yu. I. Medvedev**

## Some Theorems on the Asymptotic Distribution of the Statistic $\chi^2$

*(Presented by Academician Yu. V. Linnik on 3 XII 1969)*

1. Suppose that a sequence of  $n$  independent trials is performed, in each of which one and only one of  $N$  incompatible outcomes  $E_1, E_2, \dots, E_N$  can occur, and let  $\nu_1, \nu_2, \dots, \nu_N$

$$\left( \sum_{m=1}^N \nu_m = n \right)$$

be the frequencies of the corresponding outcomes. Denote by  $H_0$  the hypothesis that the probabilities of the outcomes do not depend on the number of the trial and are respectively equal to  $p_1, p_2, \dots, p_N$

$$\left( p_m > 0, m = 1, \dots, N, \sum_{m=1}^N p_m = 1 \right).$$

One of the most widespread methods for testing the agreement of the hypothesis  $H_0$  with the values  $\nu_m$  observed as a result of an experiment is Pearson's  $\chi^2$  test, based on the statistic

$$\chi^2 = \sum_{m=1}^N \frac{(\nu_m - np_m)^2}{np_m} = \sum_{m=1}^N \frac{\nu_m^2}{np_m} - n. \quad (1)$$

The limiting behavior of this statistic as the sample size  $n$  increases without bound and for fixed values of  $N, p_1, p_2, \dots, p_N$  is determined by the well-known theorem of K. Pearson <sup>(1)</sup>, according to which the statistic  $\chi^2$ , under the indicated conditions, has in the limit the  $\chi^2$  distribution with  $N - 1$  degrees of freedom. A generalization of this result to the case when, as  $n$  grows, the parameters  $N, p_1, p_2, \dots, p_N$  may vary simultaneously was obtained in <sup>(2)</sup>. There it was shown that in this case, under the condition

$$\min_{1 \leq m \leq N} np_m \rightarrow \infty, \quad (2)$$

the limiting distribution of the quantity  $\chi^2$  is the  $\chi^2$  distribution with  $N - 1$  degrees of freedom; moreover, if  $N \rightarrow \infty$ , then, uniformly in  $u$ ,

$$\mathbf{P}\{\chi^2 \leq N - 1 + u\sqrt{2(N - 1)}\} \rightarrow \Phi(u),$$

where  $\Phi(u)$  is the normal distribution function. Condition (2), under which this result was obtained, means that as  $n \rightarrow \infty$  the mean number of appearances of each possible outcome  $E_m$  must grow without bound. In this case the frequencies  $\nu_m$  ( $m = 1, \dots, N$ ) are distributed asymptotically normally, and the idea of the proof of the theorems in (1, 2) is essentially based on this.

The results obtained in (2) do not cover the needs arising in statistical practice. This is connected, first of all, with the circumstance that the restriction (2) often cannot be regarded as fulfilled. To fill the indicated gap, it is of interest to consider cases in which the quantities  $np_m$  remain bounded in the limit as  $n \rightarrow \infty$  for some values, or even for all values, of  $m$ . Also of interest is the question of the limiting behavior of the statistic  $\chi^2$  when  $np_m$  (all or some) tend to 0. It is natural to combine all these enumerated cases under the general name of the case of "small samples," understanding by this the failure of condition (2). Some results relating to these cases were obtained in (3). More will be said about them below.

In the present paper we present theorems on the limiting behavior of the statistic  $\chi^2$  under the null hypothesis for the case of small samples.

**2. Definition.** We shall say that a function of the frequencies  $f(\nu_1, \nu_2, \dots, \nu_N)$  belongs to the class of additively separable functions if it can be represented in the form

$$f(\nu_1, \nu_2, \dots, \nu_N) = \sum_{m=1}^N f_m(\nu_m),$$

where the function  $f_m(\nu_m)$  depends on  $\nu_m$  and  $p_m$ , and does not depend on  $\nu_i$  and  $p_i$  for  $i \neq m$  ( $i = 1, \dots, N$ ). The class of statistics introduced here arises, for example, in the application of such criteria as  $\chi^2$  and the likelihood-ratio criterion, for testing certain hypotheses concerning the probabilities  $p_i$  ( $i = 1, \dots, N$ ) in the multinomial scheme. The usefulness of the class introduced is explained by the fact that for the characteristic functions of this class there is an integral representation which serves as a good working tool for proving limit theorems for a comparatively broad spectrum of possible values of the parameters  $N, p_1, p_2, \dots, p_N$ .

**Theorem 1.** *The representation*

$$\mathbb{E}e^{itf} = \frac{n!}{N^n} \frac{1}{2\pi i} \oint e^{zN} \prod_{m=1}^N \left( \sum_{k=0}^{\infty} \pi_k(z a_m) e^{itf_m(k)} \right) \frac{dz}{z^{n+1}}, \quad (3)$$

is valid, where the contour of integration encircles the origin of the plane of the complex variable  $z$ ,  $\pi_k(\lambda) = \lambda^k e^{-\lambda}/k!$ ,  $k = 0, \dots, \infty$ ,  $a_m = Np_m$ ,  $m = 1, \dots, N$ .

The theorem is proved by introducing the generating function

$$\Phi(z, t) = \sum_{n=0}^{\infty} \frac{(zN)^n}{n!} \mathbb{E} e^{itf(\nu_1, \nu_2, \dots, \nu_N)}$$

and applying Cauchy's theorem for contour integrals.

Limit theorems on the behavior of the class of additively separable statistics  $f(\nu_1, \nu_2, \dots, \nu_N)$  as  $n \rightarrow \infty$  and under simultaneous variation of  $N, p_1, p_2, \dots, p_N$  are proved using the integral representation (3), mainly by the saddle-point method. In the general case, for arbitrary  $f$ , the formulations of the theorems and their proofs are rather cumbersome. For the statistic  $\chi^2$  they have a simpler form.

**3.** Let  $N \rightarrow \infty$ . Then it is natural to regard the probabilities  $p_m$  as functions of  $N$ , since

$$\sum_{m=1}^N p_m = 1.$$

We shall call a multinomial scheme **regular** if, for all  $m$ , the condition

$$p_m = a_m/N, \quad 0 < a \leq a_m \leq a' < \infty$$

is satisfied. In other words, in a regular multinomial scheme there are no very large probabilities ( $a_m \rightarrow \infty$ ) and no very small ones ( $a_m \rightarrow 0$ ). Put  $a_m = 1 + \varepsilon_m$  ( $m = 1, \dots, N$ ). If all  $\varepsilon_m = 0$ , then the scheme is called an **equiprobable multinomial scheme**. Introduce three measures of the deviation of the scheme from equiprobability:

$$\beta = \frac{1}{N^2} \sum_{m=1}^N \frac{1}{p_m} - 1, \quad \gamma = \frac{1}{N} \sum_{m=1}^N \varepsilon_m^2, \quad \delta = \frac{1}{N} \sum_{m=1}^N |\varepsilon_m|^3.$$

Within the class of regular schemes (r.s.) we single out in particular the class of quasi-equiprobable schemes (q.e.r.s.). We shall call a regular multinomial scheme **quasi-equiprobable** if the conditions

$$\gamma = O(1/\sqrt{N}), \quad \delta = o(\gamma)$$

are satisfied.

Let us note that for q.e.r.s.  $\beta \sim \gamma$ . The mathematical expectation and variance of  $\chi^2$  under the hypothesis  $H_0$  are known and equal to (2)

$$E\chi^2 = N - 1, \quad D\chi^2 = Nb - \frac{2N - 2}{n} \quad \left( b = 2 + \frac{N}{n}\beta \right).$$

4. The limit laws as  $N, n \rightarrow \infty$  for the statistic  $\chi^2$  in the case of regular polynomial schemes are the normal law, the Poisson law, and their mixture. It is convenient to describe the qualitative picture of the behavior of  $\chi^2$  in terms of the behavior of the parameter  $\alpha = n/N$ . Note that the behavior of  $\alpha$  for the class of regular schemes coincides with the behavior of the quantity  $np_m = \alpha a_m$ .

- 1) If  $\alpha$  is bounded away from 0 ( $\alpha \geq \alpha_0 > 0$ ), then  $\chi^2$  has, in the limit, a normal distribution for the entire class of regular schemes.
- 2) Let  $\alpha \rightarrow 0$ , but not too rapidly, namely so that  $n\alpha \rightarrow \infty$ . Then asymptotic normality holds for those regular schemes for which  $\delta$  is not too large, more precisely, for which  $\delta = o(n^2 b^{3/2}/N^{3/2})$ .
- 3) Let  $\alpha \rightarrow 0$  very rapidly, namely:  $n\alpha = 2\lambda$ ,  $\lambda < \infty$ .

Then:

- a) if the polynomial scheme is not very close to equiprobable in the measure  $\beta$  ( $n\beta \rightarrow \infty$ ) and, moreover,  $\delta = o(n^2 b^{3/2}/N^{3/2})$ , then the normal law is obtained in the limit for  $\chi^2$ ;
- b) if  $n\beta \rightarrow c$  ( $0 < c < \infty$ ) and  $\delta = o(\gamma)$ , then  $\chi^2$  is distributed in the limit according to a mixture of the normal law and the Poisson law;
- c) if  $n\beta \rightarrow 0$  and  $\delta = o(\gamma)$ , then  $\chi^2$  is distributed in the limit according to the Poisson law with parameter  $\lambda$ .

Cases b) and c) belong to the class of quasi-equiprobable schemes. We give exact formulations of the limit theorems. In doing so we shall assume that we are in the class of regular schemes, i.e., that the condition  $0 < d \leq Np_m \leq d' < \infty$  is satisfied, although for some assertions this class could be enlarged.

**Theorem 2.** *Let  $n, N \rightarrow \infty$ . Then, if one of the three conditions is satisfied:*

- 1)  $\alpha \geq \alpha_0 > 0$ ;
- 2)  $\alpha \rightarrow 0$ ,  $n\alpha \rightarrow \infty$ ,  $\delta = o(n^2 b^{3/2}/N^{3/2})$ ;
- 3)  $n\alpha \leq c < \infty$ ,  $n\beta \rightarrow \infty$ ,  $\delta = o(n^2 b^{3/2}/N^{3/2})$ ,

*then, uniformly in  $u$ ,*

$$\mathbf{P}\{\chi^2 < N + u\sqrt{bN}\} \rightarrow \Phi(u).$$

**Theorem 3.** Let  $n, N \rightarrow \infty$  in such a way that  $\lambda = n^2/2N$  remains bounded above, and, moreover,  $\gamma = o(1/\sqrt{N})$ ,  $\delta = o(\gamma)$ . Then the random variable

$$(\chi^2 + n - N)n/2N$$

is asymptotically distributed according to the Poisson law with parameter  $\lambda$ .

**Theorem 4.** Let  $n, N \rightarrow \infty$  in such a way that  $\lambda = n^2/2N$  remains bounded above, and, moreover,  $\gamma = \sigma^2/\sqrt{N}$ ,  $\sigma^2 > 0$ ,  $\delta = o(\gamma)$ . Then the limiting distribution of the random variable

$$(\chi^2 + n - N)n/2N$$

is the convolution of the Poisson law with parameter  $\lambda$  and the normal law with parameters  $(0, (2\lambda)^{1/4}\sigma/2)$ .

Obviously, Theorem 3 follows from Theorem 4, since under the conditions of Theorem 3 the parameter  $\sigma$  must tend to 0. As a consequence of the theorems given above, we obtain that, for an equiprobable polynomial scheme,  $\chi^2$ , under the corresponding normalization, is distributed in the limit as  $n, N \rightarrow \infty$  normally if  $n^2/2N \rightarrow \infty$ , and according to the Poisson law if  $n^2/2N \rightarrow \lambda < \infty$ .

We note that the result of Theorem 2 under condition 1) was obtained by another method in (3). There, the limiting Poisson law was also obtained for the equiprobable polynomial scheme when  $n^2/2N \rightarrow \lambda < \infty$ .

In conclusion, I express my gratitude to G. I. Ivchenko for his attention to the work and for useful advice.

Received  
28 XI 1969

## References

1. K. Pearson, *Phil. Mag.*, **6** (1901).
2. S. Kh. Tumanyan, *Theory of Probability and Its Applications*, **1** (1956).
3. G. P. Steck, *Univ. Calif. Publ. Stat.*, **2** (1957).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*