Sankhyā : The Indian Journal of Statistics
2015, Volume 77-A, Part 2, pp. 330-336

A Characterisation of the Gaussian Distribution through the Sample Variance

Nina N. Golikova
Moscow Regional State University, Moscow, Russia
Victor M. Kruglov
Moscow State University, Moscow, Russia

Abstract

A classical result states that the sample variance of a standard Gaussian sample has the chi-square distribution. In this note, a partial reverse of this result is proved for independent infinitely divisible random variables $X_1, \ldots, X_n, n \geq 2$. If $n \geq 3$, $\mathbb{E}X_1 = \cdots = \mathbb{E}X_n$ and the random variable $nS_n^2 = (X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2$ where $\overline{X} = (X_1 + \cdots + X_n)/n$, has the chi-square distribution with $n-1$ degrees of freedom then $X_1, \ldots, X_n$ are Gaussian random variables with $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \cdots = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = 1$. In the case $n = 2$, the random variable $2S_2^2$ has the chi-square distribution with 1 degree of freedom if and only if X_1 and X_2 are Gaussian random variables with $\mathbb{E}X_1 = \mathbb{E}X_2$ and $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 + \mathbb{E}(X_2 - \mathbb{E}X_2)^2 = 2$.

AMS 2000 subject classifications. Primary 60E07; Secondary 62E10.
Keywords and phrases. Gaussian distribution, chi-squared distribution, distribution theory.

Let independent random variables $X_1, \ldots, X_n$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The random variables

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

are called the sample mean and the sample variance. A well known (Hogg and Craig 1970, p. 108) classical result states that the random variable $nS_n^2, n \geq 2$, has the chi-square distribution with $n-1$ degrees of freedom

$$\mathbb{P}\{nS_n^2 < x\} = \int_0^x \frac{u^{(n-3)/2}}{2^{(n-1)/2}\Gamma((n-1)/2)} e^{-u/2} du, x \geq 0,$$

Research supported by the Russian Scientific Foundation, project 14-11-00362.

if $X_1, \dots, X_n$ are Gaussian random variables with

$$\mathbb{E}X_1 = \dots = \mathbb{E}X_n, \mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \dots = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = 1.$$

Harold Ruben (1974, 1975) discovered an important characterisation of the Gaussian distribution. His result can be exposed as follows. If symmetric independent random variables $X_1, \dots, X_n, n \geq 2$, are identically distributed and the random variable nS_n^2 has the chi-square distribution with $n-1$ degrees of freedom then $X_1, \dots, X_n$ are standard Gaussian random variables. The assumption on symmetry can be removed under some additional condition. Let the sample variance S_m^2 be constructed with the help of random variables $X_1, \dots, X_m$ for some $1 < m < n$. If the random variables nS_n^2 and mS_m^2 have the chi-square distributions with $n-1$ and $m-1$ degrees of freedom then $X_1, \dots, X_n$ are Gaussian random variables with $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = 1$. The proof of the Ruben theorem relies heavily on supposition of symmetry of random variables and, what is more important, that the random variables are identically distributed. Is it possible to reject the hypothesis that the random variables are symmetric and identically distributed? In this note we give a partial answer to this question.

THEOREM. *Let $X_1, \dots, X_n, n \geq 2$, be independent infinitely divisible random variables. The random variable $2S_2^2$ has the chi-square distribution with 1 degree of freedom iff X_1 and X_2 are Gaussian random variables with $\mathbb{E}X_1 = \mathbb{E}X_2$ and $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 + \mathbb{E}(X_2 - \mathbb{E}X_2)^2 = 2$. In general for $n \geq 3$, if $\mathbb{E}X_1 = \dots = \mathbb{E}X_n$ and the random variable nS_n^2 has the chi-square distribution with $n-1$ degrees of freedom then $X_1, \dots, X_n$ are Gaussian random variables with $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \dots = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = 1$.*

The proof is based on some properties of infinitely divisible distributions. Recall (Loève 1977, p. 343) that a random variable X is infinitely divisible if and only if its characteristic function can be represented in the following form

$$\mathbb{E}e^{itX} = \exp\left\{ i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) dH(x) \right\}, t \in \mathbb{R},$$

where $\gamma, \sigma \in \mathbb{R}, \sigma \geq 0$, the left-continuous function H defined on $\mathbb{R} \setminus \{0\}$ is nondecreasing on $(-\infty, 0)$ and $(0, \infty)$ with $H(\pm\infty) = 0$ and $\int_{(-1,0)\cup(0,1)} x^2 dH(x) < \infty$.

One can easily verify that the sum of a finite number of independent infinitely divisible random variables is an infinitely divisible random variable.

Kruglov (1970, 1972) discovered close relations between the asymptotic behaviour of the distribution function $\mathbb{P}\{X < x\}, x \in \mathbb{R}$, and the Lévy function $H(x)$, $x \in \mathbb{R} \setminus \{0\}$. Some important additions were made by Rossberg et al. (1985), Ken-Iti (1999), Yakymiv (2005), and Ssörgő (2005). With the help of these relations one can establish new characterisations of Poisson and Gaussian distributions in the set of infinitely divisible distributions, for example, see Kruglov (2003, 2010, 2012, 2013). The proof of our theorem is based on the following two lemmas.

LEMMA 1. *An infinitely divisible random variable X has a Gaussian distribution, may be with zero variance, if*

$$\lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X| > r\}}{r \ln r} = -\infty.$$

PROOF. can be found in Kruglov (1972) or in Kruglov and Antonov (1982). It can also be found in Rossberg et al. (1985) and Ken-Iti (1999).

LEMMA 2. *Let ξ be a random variable with the standard Gaussian distribution. Then for any real number λ the random variable $(\xi+\lambda)^2$ has the moment-generating function*

$$\mathbb{E}e^{t(\xi+\lambda)^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(x+\lambda)^2} e^{-x^2/2} dx = \frac{1}{\sqrt{1-2t}} e^{\lambda^2 t/(1-2t)}, t \in (-\infty, 1/2).$$

PROOF. can be found in textbooks on Probability and Statistics (for example, see Hogg and Craig 1970, p. 288–289).

PROOF OF THE THEOREM. For any $r > 0$ the inequality

$$\begin{aligned} \mathbb{P}\{|X_k - \overline{X}| > r\} &= \mathbb{P}\{|X_k - \overline{X}|^2 > r^2\} \le \mathbb{P}\{nS_n^2 > r^2\} \\ &= \int_{r^2}^{\infty} \frac{x^{(n-3)/2}}{2^{(n-1)/2}\Gamma((n-1)/2)} e^{-x/2} dx \end{aligned}$$

holds. If $n = 3$, then one can get

$$\lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X_k - \overline{X}| > r\}}{r \ln r} \le \lim_{r\to\infty} \frac{1}{r \ln r} \ln \int_{r^2}^{\infty} \frac{1}{2\Gamma(1)} e^{-x/2} dx = -\infty.$$

For $n \ge 2, n \ne 3$, with the help of de L'Hospital rule (apply twice) one can verify that

$$\begin{aligned} \lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X_k - \overline{X}| > r\}}{r \ln r} &\le \lim_{r\to\infty} \frac{1}{r \ln r} \ln \int_{r^2}^{\infty} \frac{x^{(n-3)/2}}{2^{(n-1)/2}\Gamma((n-1)/2)} e^{-x/2} dx \\ &= \lim_{r\to\infty} \frac{-2r^{n-2} e^{-r^2/2}}{(1+\ln r) \int_{r^2}^{\infty} x^{(n-3)/2} e^{-x/2} dx} = -\infty. \end{aligned}$$

By Lemma 1 the infinitely divisible random variable $X_k - \overline{X} = (1-1/n)X_k - \sum_{j\neq k} X_j/n$ is Gaussian. It then follows from Cramér's theorem (Loève 1977, p. 283) on the decomposition of the Gaussian distribution that $X_1, \ldots, X_n$ are Gaussian random variables with some variances $\sigma_1^2 = \mathbb{E}(X_1 - \mathbb{E}X_1)^2, \ldots, \sigma_n^2 = \mathbb{E}(X_n - \mathbb{E}X_n)^2$. For any k and $j, 1 \leq k \neq j \leq n$, the difference $X_k - X_j$ is a Gaussian random variable with variance $\sigma_k^2 + \sigma_j^2$. In accordance with a known (Rossberg et al. 1985, p. 188) classical result one can get

$$\lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X_k - X_j| > r\}}{r^2} = -\frac{1}{2(\sigma_k^2 + \sigma_j^2)}.$$

Note that $|X_k - X_j|^2 = |(X_k - \overline{X}) - (X_j - \overline{X})|^2 \leq 2(|X_k - \overline{X}|^2 + |X_j - \overline{X}|^2) \leq 2nS_n^2$. If $n = 3$ one can verify that

$$-\frac{1}{2(\sigma_k^2 + \sigma_j^2)} = \lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X_k - X_j| > r\}}{r^2} \leq \lim_{r\to\infty} \frac{\ln \mathbb{P}\{2nS_n^2 > r^2\}}{r^2}$$

$$= \lim_{r\to\infty} r^{-2} \ln \int_{r^2/2}^{\infty} \frac{1}{2\Gamma(1)} e^{-x/2} dx = -\frac{1}{4}.$$

For $n \geq 2, n \neq 3$, again, by de L'Hospital rule (apply twice) one can get

$$\begin{aligned} -\frac{1}{2(\sigma_k^2 + \sigma_j^2)} &= \lim_{r\to\infty} \frac{\ln \mathbb{P}\{|X_k - X_j| > r\}}{r^2} \leq \lim_{r\to\infty} \frac{\ln \mathbb{P}\{2nS_n^2 > r^2\}}{r^2} \\ &= \lim_{r\to\infty} r^{-2} \ln \int_{r^2/2}^{\infty} \frac{x^{(n-3)/2}}{2^{(n-1)/2}\Gamma((n-1)/2)} e^{-x/2} dx \\ &= \lim_{r\to\infty} \frac{2(n-3)r^{n-4} - r^{n-2}}{4r^{n-2}} = -\frac{1}{4}. \end{aligned}$$

It follows that $0 \leq \sigma_k^2 + \sigma_j^2 \leq 2$. If $\mathbb{E}X_1 = \cdots = \mathbb{E}X_n$ then one can easily verify that

$$\mathbb{E}(X_k - \overline{X})^2 = \frac{n-2}{n}\sigma_k^2 + \frac{1}{n^2}\sum_{j=1}^{n} \sigma_j^2,$$

$$n - 1 = \mathbb{E}(nS_n^2) = \sum_{k=1}^{n} \mathbb{E}(X_k - \overline{X})^2 = \frac{n-1}{n}\sum_{k=1}^{n} \sigma_k^2.$$

If $n = 2m$ is an even integer for some $m \geq 2$ then $\sigma_1^2 = \cdots = \sigma_n^2 = 1$. To prove this, we note that for any permutation π of integers $1, \ldots, n$ the equalities

$$\sum_{k=1}^{m}(\sigma^2_{\pi(2k-1)} + \sigma^2_{\pi(2k)}) = \sum_{k=1}^{n} \sigma_k^2 = n = 2m$$

hold. It was proved above that $\sigma_k^2 + \sigma_j^2 \leq 2$ for any $1 \leq k \neq j \leq n$. It follows that

$$\sum_{k=1}^{m}(2 - (\sigma^2_{\pi(2k-1)} + \sigma^2_{\pi(2k)})) = 0$$

and $\sigma_k^2 + \sigma_j^2 = 2$ for any $1 \leq k \neq j \leq n$, and hence $\sigma_1^2 = \cdots = \sigma_n^2 = 1$. Now suppose that $n = 2m + 1$ is an odd integer with some $m \geq 1$. Again, the equality

$$\sum_{k=1}^{m}(\sigma^2_{\pi(2k-1)} + \sigma^2_{\pi(2k)} - 2) + \sigma^2_{\pi(n)} = 1$$

holds. Since $\sigma^2_{\pi(2k-1)} + \sigma^2_{\pi(2k)} - 2 \leq 0$, it follows that $\sigma^2_{\pi(n)} \geq 1$. It means that $\sigma_k^2 \geq 1$ for any $k = 1, \ldots, n$, since π is an arbitrary permutation of integers $1, \ldots, n$. The inequalities $2 \leq \sigma_k^2 + \sigma_j^2 \leq 2$ for any $1 \leq k \neq j \leq n$ involve the equalities $\sigma_k^2 + \sigma_j^2 = 2$ and $\sigma_1^2 = \cdots = \sigma_n^2 = 1$. This proves the second part of the theorem. *If $X_1, \ldots, X_n, n \geq 3$, are independent infinitely divisible variables, $\mathbb{E}X_1 = \cdots = \mathbb{E}X_n$ and the random variable nS_n^2 has the chi-square distribution with $n-1$ degrees of freedom then $X_1, \ldots, X_n$ are Gaussian random variables with $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \cdots = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = 1$.*

To prove the first part of the theorem, it should be noted that $2S_2^2 = (X_1 - X_2)^2/2$. Suppose that independent random variables X_1 and X_2 are infinitely divisible and the random variable $2S_2^2$ has the chi-square distribution with 1 degree of freedom. It was proved above that X_1 and X_2 are Gaussian random variables with some means $\mathbb{E}X_1$ and $\mathbb{E}X_2$ and some variances σ_1^2 and σ_2^2. The random variable $2S_2^2$ can be written in the following way

$$2S_2^2 = \frac{\sigma_1^2 + \sigma_2^2}{2}(\xi + \lambda)^2 \text{ with } \xi = \frac{(X_1 - \mathbb{E}X_1) - (X_2 - \mathbb{E}X_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \lambda = \frac{\mathbb{E}X_1 - \mathbb{E}X_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

The random variable ξ is Gaussian with $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$. By Lemma 2 the following equalities

$$\frac{1}{\sqrt{1-2t}} = \mathbb{E}e^{t2S_2^2} = \mathbb{E}\exp\left\{\frac{t(\sigma_1^2 + \sigma_2^2)}{2}(\xi + \lambda)^2\right\}$$

$$= \frac{1}{\sqrt{1-(\sigma_1^2+\sigma_2^2)t}} \exp\left\{\frac{t\lambda^2(\sigma_1^2+\sigma_2^2)}{2(1-(\sigma_1^2+\sigma_2^2)t)}\right\}$$

hold for $t \in (-\infty, 1/2)$. It was proved above that $\sigma_1^2 + \sigma_2^2 \leq 2$. If this inequality is strong, then one can get an impossible equality of the form $\infty = a$ for some real number a, by letting $0 < t \uparrow 1/2$. It follows that $\sigma_1^2 + \sigma_2^2 = 2$ and $\lambda = 0$.

Now assume that that X_1 and X_2 are independent Gaussian random variables with $\mathbb{E}X_1 = \mathbb{E}X_2$ and $\sigma_1^2 + \sigma_2^2 = 2$. One can easily verify that the random variable $2S_2^2$ has the chi-square distribution with 1 degree of freedom. The theorem is proved.

REMARK. It may seem the case with respect to the second part of the theorem proved above that the result is not a 'characterisation' not being an 'if and only if' statement. To substantiate the title of the note one can rewrite the second part of the theorem in the following way.

Let $X_1, \ldots, X_n, n \geq 3$, *be independent infinitely divisible random variables such that* $\mathbb{E}X_1 = \cdots = \mathbb{E}X_n$. *Then they are Gaussian random variables with* $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = \cdots = \mathbb{E}(X_n - \mathbb{E}X_n)^2 = 1$ *if and only if the random variable* nS_n^2 *has the chi-squad distribution with* $n-1$ *degrees of freedom.*

Indeed, if independent random variables $X_1, \ldots, X_n$ have the same Gaussian distribution with $\mathbb{E}(X_1 - \mathbb{E}X_1)^2 = 1$ then the random variable nS_n^2 has the chi-squad distribution with $n - 1$ degrees of freedom. This result is recorded in many textbooks on Probability theory and it was mentioned in the beginning of the note. It is also known that all Gaussian distributions are infinitely divisible.

Acknowledgments. The authors are grateful to anonymous reviewers for their useful remarks which helped to improve the presentation.

References

HOGG, R.V. and CRAIG, A.T. (1970). *Introduction to mathematical statistics, 4th edn.* Macmillan Publ. Co. Inc., New York.

KRUGLOV, V.M. (1970). A note on the theory of infinitely divisible laws. *Theory Probab. Appl.* **15**, 319–324.

KRUGLOV, V.M. (1972). Integrals with respect to infinitely divisible distributions in a Hilbert space. *Math. Notes* **11**, 407–411.

KRUGLOV, V.M. and ANTONOV, C.N. (1982). On asymptotic behaviour of infinitely divisible distributions in Banach space. *Theory Probab. Appl.* **27**, 626–642.

KRUGLOV, V.M. (2003). Normal and Poisson convergence. *Theory Probab. Appl.* **48**, 392–398.

KRUGLOV, V.M. (2010). A characterization of the Poisson distribution. *Stat. Probab. Lett.* **80**, 2032–2034.

KRUGLOV, V.M. (2012). A characterization of the convolution of Gaussian and Poisson distributions. *Sankhyā Series A* **74**, 1–9.

KRUGLOV, V.M. (2013). A characterization of the Gaussian distribution. *Stoch. Anal. Appl.* **31**, 872–875.

LOÈVE, M. (1977). Probability theory, I, 4th edn. Springer-Verlag.

ROSSBERG, H.-J., JESIAK, B. and SIEGEL, G. (1985). Analytical methods of probability theory. Berlin, Akademie-Verlag.

RUBEN, H. (1974). A new characterisation of the normal distribution through the sample variance. *Sankhyā Series A* **36**, 379–388.

RUBEN, H. (1975). A further characterisation of normality through the sample variance. *Sankhyā Series A* **37**, 72–81.

KEN-ITI, S. (1999). *Levy processes and infinitely divisible distributions.* University Press, Cambridge.

SSÖRGŐ, A. (2005). A probabilistic proof of Kruglov's theorem on the tails of infinitely divisible distributions. *Acta Sci. Math.* **71**, 405–415.

YAKYMIV, A.L. (2005). Probabilistic applications of Tauberian theorems. Leiden-Boston, VSP.

Nina N. Golikova
Faculty of Physics and Mathematics
Moscow Regional State University
Radio st. 10a, 105005
Moscow, Russia
E-mail: nina.golikova@mail.ru

Victor M. Kruglov
Department of Statistics
Faculty of Computational
Mathematics and Cybernetics
Moscow State University
Vorobyovy Gory, GSP-1
119992 Moscow, Russia
E-mail: krugvictor@gmail.com

Paper received: 24 March 2014; revised: 14 June 2014.