PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 14, Fasc. 2 (1993), pp. 257–263

ON THREE CHARACTERIZATIONS OF THE NORMAL DISTRIBUTION

BY

M. P. QUINE (SYDNEY)

Abstract. Of all the characterizations of the normal distribution, three landmarks are the theorems of Bernstein and Skitovitch concerning independence of linear forms and the theorem of Geary concerning independence of sample mean and variance. In this note, ideas from several proofs of these theorems are distilled to give unified proofs which depend mostly on simple probability and characteristic function concepts.

1. Introduction. In this note we present proofs of three characterization theorems: those of Bernstein and of Skitovitch concerning independence of linear forms and Geary's Theorem which concerns independence of sample mean and variance. Part of the motivation for this work was the realization that these three important theorems seem not to be very well known amongst statisticians, perhaps because of the rather patchy coverage in probability textbooks. Feller [3] seems to provide the best coverage with a proof of Bernstein's Theorem and careful statements of the Theorems of Skitovitch and Geary. Other probability textbooks I have examined give (if they mention the topic at all) versions of Bernstein's Theorem or Geary's Theorem, often with restrictive moment conditions imposed.

Early proofs of these results involved such restrictive moment assumptions or used cumbersome finite difference methods (see, e.g., Feller [3], especially p. 79). Zinger [15] and Lancaster [8] proved the same results by showing that the independence of certain functions of independent variables implies the existence of moments of all orders and then using cumulant arguments. The proofs presented here are *unified*, in that they all involve first adapting Lancaster's argument to establish finite second (or higher) moments, and then adapting a characteristic function (cf) argument of Lukacs [10], and *elementary* in that mostly only simple probability and cf concepts are used. It is hoped that this approach will make the theory accessible to a wider audience.

2. Bernstein's Theorem. The following result was first proved under the assumption of equal (and finite) variances by Bernstein [1].

THEOREM 1. *Suppose X_1 and X_2 are independent random variables (rv's). Put $Y_1 = X_1+X_2$ and $Y_2 = X_1-X_2$. If Y_1 and Y_2 are independent, then X_1 and X_2 are normally distributed with the same variance.*

Proof. We prove this theorem using two lemmas, which are proved at the end of this section. We put $\mu_i = \mathbf{E}(X_i)$ for $i = 1, 2$.

LEMMA 1. *The assumptions of Theorem* 1 *imply that X_1 and X_2 have finite variances.*

LEMMA 2. *Under the assumptions of Theorem* 1, *$X_1-\mu_1$ and $X_2-\mu_2$ have the same distribution.*

By Lemma 2 we may suppose without loss of generality that $\mu_1 = \mu_2 = 0$ and write ϕ for the common cf of X_1 and X_2. Let $Y_3 = Y_2^2 = X_1^2+X_2^2 -2X_1X_2$ and let γ be the joint cf of Y_1 and Y_3:

$$\gamma(s, t) = \mathbf{E}(\exp[isY_1+itY_3]).$$

We have

$$\mathbf{E}|Y_3| = \mathbf{E}(X_1-X_2)^2 \leqslant 2\mathbf{E}(X_1^2)+2\mathbf{E}(X_2^2) < \infty$$

from Lemma 1, so that, using dominated convergence,

$$\begin{aligned} h^{-1}|\gamma(s, t+h)-\gamma(s, t)| &= |\mathbf{E}(\exp[isY_1+itY_3])(\exp[ihY_3]-1)/h| \\ &\to i\mathbf{E}Y_3\exp[isY_1+itY_3] \quad \text{as } h\to 0. \end{aligned}$$

Thus the partial derivative $\partial\gamma/\partial t$ exists. Therefore, using independence and Lemma 2, we have on the one hand

$$\begin{aligned} \left.\frac{\partial\gamma}{\partial t}\right|_{t=0} &= 2i\mathbf{E}(X_1^2\exp[isX_1])\mathbf{E}(\exp[isX_2])-2i\mathbf{E}^2(X_1\exp[isX_1]) \\ &= -2i\phi''(s)\phi(s)+2i(\phi'(s))^2, \end{aligned}$$

and on the other hand

$$\left.\frac{\partial\gamma}{\partial t}\right|_{t=0} = \mathbf{E}(\exp[isY_1])\frac{\partial}{\partial t}\mathbf{E}(\exp[itY_3])|_{t=0} = 2i\sigma^2(\phi(s))^2,$$

where σ^2 is the common variance of X_1 and X_2. Equating these two expressions gives

$$-\phi\phi''+(\phi')^2 = \sigma^2\phi^2, \tag{1}$$

which is easily solved to give $\phi(s) = \exp[-\sigma^2 s^2/2]$ and the theorem follows. ∎

We complete this section with proofs of the lemmas.

Proof of Lemma 1. Take any $\varepsilon < 1/2$ and choose $A\in(0, \infty)$ so that

$$\mathbf{P}(|X_i| > A) < \varepsilon \quad \text{and} \quad \mathbf{P}(|Y_i| > A) < \varepsilon \quad \text{for } i = 1, 2.$$

Using independence and the inequalities $|Y_i| \geqslant |X_1|-|X_2|$ for $i=1, 2$, we obtain

$$(1-\varepsilon)\mathbf{P}(|X_1|>3A) \leqslant \mathbf{P}(|X_1|>3A, |X_2| \leqslant A)$$
$$\leqslant \mathbf{P}(|Y_1| \geqslant 2A, |Y_2| \geqslant 2A) \leqslant \varepsilon^2.$$

It follows that $\mathbf{P}(|X_i|>3A)<2\varepsilon^2$ for $i=1$; the inequality can be proved in the same way for $i=2$. Iterating this result leads to

$$\mathbf{P}(|X_i|>3^k A)<(2\varepsilon)^{2^k}/2, \quad i=1, 2;\ k \geqslant 0. \tag{2}$$

Now

$$\mathbf{E}\left(\frac{X_i^2}{A^2}\right)=\int_0^\infty \mathbf{P}(X_i^2>A^2 x)\,dx \leqslant \sum_{n=0}^\infty \mathbf{P}(|X_i|>A\sqrt{n}) \tag{3}$$
$$\leqslant 1+\sum_{k=0}^\infty \sum_{3^k \leqslant \sqrt{n}<3^{k+1}} \mathbf{P}(|X_i|>3^k A)$$
$$\leqslant 1+\sum_{k=0}^\infty 3^{2k+2}\mathbf{P}(|X_i|>3^k A)$$
$$\leqslant 1+\sum_{k=0}^\infty 3^{2k+2}(2\varepsilon)^{2^k}/2<\infty. \blacksquare$$

Proof of Lemma 2. Let the joint cf of Y_1 and Y_2 be

$$\psi(s, t)=\mathbf{E}(\exp[isY_1+itY_2]).$$

The partial derivative $\partial\psi/\partial t$ exists for the same kind of reasons that $\partial\gamma/\partial t$ does, and

$$\left.\frac{\partial\psi}{\partial t}\right|_{t=0}=\mathbf{E}(iX_1\exp[isX_1])\mathbf{E}(\exp[isX_2])-\mathbf{E}(iX_2\exp[isX_2])\mathbf{E}(\exp[isX_1])$$
$$=\phi_1'(s)\phi_2(s)-\phi_1(s)\phi_2'(s),$$

where ϕ_i is the cf of X_i for $i=1, 2$. But we also have

$$\left.\frac{\partial\psi}{\partial t}\right|_{t=0}=\mathbf{E}(\exp[isY_1])\frac{\partial}{\partial t}\mathbf{E}(\exp[itY_2])|_{t=0}=i\phi_1(s)\phi_2(s)(\mu_1-\mu_2).$$

Thus

$$\phi_1'/\phi_1-\phi_2'/\phi_2=i(\mu_1-\mu_2),$$

from which it follows that

$$\phi_1(s)=\phi_2(s)\exp[is(\mu_1-\mu_2)],$$

which implies Lemma 2. $\blacksquare$

3. Skitovitch's Theorem. In this section we show how to extend the methods of Section 2 to prove a generalisation of Bernstein's Theorem due to Skitovitch [13] (see also [14]). We work with cf's again and avoid cumulants and Marcinkiewicz's Theorem (see [11]).

THEOREM 2. *Suppose that $X_1, \ldots, X_n$ are independent rv's, where $n \geqslant 2$, and that $Y_1 = a_1 X_1 + \ldots + a_n X_n$ and $Y_2 = b_1 X_1 + \ldots + b_n X_n$ are also independent, where $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are non-zero constants. Then each X_i is normally distributed.*

Proof. It is easy to see that we can assume $a_1, \ldots, a_n$ all equal 1. Lemma 1 can be extended in a straightforward way to establish the finiteness of the variances of $X_1, \ldots, X_n$ as follows. Let $\alpha = \min_i |b_i|$ and $\beta = \max_i |b_i|$. Choose $\varepsilon < 1 - 2^{-1/(n-1)}$ and choose A so that

$$\mathbf{P}(|Y_1| > nA) < \varepsilon, \qquad \mathbf{P}(|Y_2| > n\beta A) < \varepsilon$$

and

$$\mathbf{P}(|X_i| > A) < \varepsilon \quad \text{for } i = 1, \ldots, n.$$

Put $\gamma = (2n-1)\beta/\alpha (\geqslant 3)$. Then it follows as in Section 2 (see also (13) below) that for $i = 1, \ldots, n$

$$\mathbf{P}(|X_i| > \gamma A) < 2\varepsilon^2 \tag{4}$$

and $\mathbf{E}(X_i^2) < \infty$ holds as in the proof of Lemma 1.

Extending the notation of the last section in an obvious way we have

$$\left.\frac{\partial \psi}{\partial t}\right|_{t=0} = \mathbf{E}\left(\sum_j i b_j X_j \exp\left[is \sum_k X_k\right]\right) = \sum_j b_j \phi_j'(s) \prod_{k \neq j} \phi_k(s)$$

and

$$\left.\frac{\partial \psi}{\partial t}\right|_{t=0} = \prod_j \phi_j(s) \frac{\partial}{\partial t} \mathbf{E}(\exp[itY_2])|_{t=0} = \prod_j \phi_j(s) \sum_k i b_k \mu_k.$$

Combining these equations and integrating we obtain

$$\prod_j (\phi_j(s) \exp[-is\mu_j])^{b_j} = 1. \tag{5}$$

Using the fact that $\sum_j b_j \sigma_j^2 = 0$, which follows from $\operatorname{cov}(Y_1, Y_2) = 0$, we can write (5) as

$$\prod_j (\phi_j(s) \exp[-is\mu_j + s^2 \sigma_j^2/2])^{b_j} = 1. \tag{6}$$

In fact, this equation also holds with b_j replaced by b_j^2. To see this, put

$$Y_3 = Y_2^2 = \sum_j b_j^2 X_j^2 + \sum_{j \neq k} b_j b_k X_j X_k.$$

Then with γ as in Section 2 we get

(7) $$\left.\frac{\partial \gamma}{\partial t}\right|_{t=0} = -i\Big(\sum_j b_j^2 \phi_j''(s) \prod_{k \neq j} \phi_k(s) + \sum_{j \neq k} b_j b_k \phi_j'(s) \phi_k'(s) \prod_{l \neq j,k} \phi_l(s)\Big)$$

and

(8) $$\left.\frac{\partial \gamma}{\partial t}\right|_{t=0} = i\varkappa \prod_j \phi_j(s),$$

where $\varkappa = \sum_j b_j^2 \sigma_j^2 + (\sum_j b_j \mu_j)^2$. Therefore, (7) and (8) lead to the equation

(9) $$\sum_j b_j^2 \frac{\phi_j''}{\phi_j} + \sum_{j \neq k} b_j b_k \frac{\phi_j' \phi_k'}{\phi_j \phi_k} = -\varkappa$$

(it is clear from (5) that no $\phi_j(s)$ vanishes). We define $K_j(s) = \log \phi_j(s)$, in terms of which (9) together with (6) give

(10) $$\sum_{j=1}^{n} b_j^2 K_j''(s) = -\sum_{j=1}^{n} b_j^2 \sigma_j^2$$

which integrates to give

(11) $$\prod_j \big(\phi_j(s) \exp[-is\mu_j + s^2 \sigma_j^2/2]\big)^{b_j^2} = 1.$$

If the b_j's are integers, then each $\phi_j^{b_j^2}$ is a cf, and so the normality of the X_j's follows from Cramér's Theorem which states that the only independent decomposition of a normal rv is into normal components (see, e.g., Linnik [9], who establishes the equivalence of this special case of Theorem 2 and Cramér's Theorem).

However, it is not difficult to adapt Cramér's proof ([2], pp. 55–56) to prove Theorem 2 in general. Choose $1 \leqslant j \leqslant n$. It follows from (4) that $\mathbf{E}(\exp[AX_j^2]) < \infty$ for any $A > 0$ (cf. (3)). This guarantees that $\phi_j(s)$ is an entire function for all complex values of s. Also, it follows from (11) that

$$\prod_j |\phi_j(s)|^{b_j^2} \leqslant \exp\big[\tfrac{1}{2}|s|^2 \sum_j b_j^2 \sigma_j^2 + |s| \sum_j b_j^2 |\mu_j|\big],$$

so that each $\phi_j^{b_j^2}$ is of order $\leqslant 2$. Since no ϕ_j has a zero, it follows from Hadamard's factorization theorem (see, e.g., [3], p. 525) that $b_j^2 \log \phi_j(s)$ is a quadratic polynomial so that X_j is normally distributed. ∎

4. Geary's Theorem. Geary [4] proved the following result assuming all moments are finite.

THEOREM 3. *Let $X_1, \ldots, X_n$ be independent and identically distributed rv's and put*

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}.$$

If $\bar{X}$ and S^2 are independent, then X_1 is normally distributed.

Proof. Our proof follows Lukacs's in showing that the cf ϕ of X_1 satisfies the differential equation (1), by differentiating with respect to t and setting $t = 0$ in the identity

$$\mathbf{E}\exp[is\bar{X}+itS^2] = \phi^n(s/n)\varkappa(t), \tag{12}$$

where $\varkappa$ is the cf of S^2; we omit details of this step, which may also be found on p. 363 of [7]. Use is made of the interesting identity (in terms of our unbiased S^2)

$$S^2 = \frac{(n-1)\sum_{i=1}^{n} X_i^2 - \sum\sum_{i\neq j} X_i X_j}{n(n-1)}.$$

The validity of this differentiation of (12) depends of course on the finiteness of $\operatorname{var}(X_1)$, which Lukacs assumed, but which is actually a consequence of independence and can be verified along the lines of Lemma 1:

LEMMA 3. *Under the conditions of Theorem* 3, $\mathbf{E}(X_1^2) < \infty$.

Proof. Take ε as in the last section and choose A such that $\mathbf{P}(|X_i| > A) < \varepsilon$ for $i = 1, \ldots, n$ and $\mathbf{P}(S^2 \geqslant 4(n-1)^3 A^2/n^2) < \varepsilon$. Then

$$\begin{aligned}(1-\varepsilon)^{n-1}\mathbf{P}(|X_1| > (2n-1)A) &\leqslant \mathbf{P}(|X_1| > (2n-1)A,\ |X_j| \leqslant A,\ j = 2, \ldots, n) \\ &\leqslant \mathbf{P}\left(|\bar{X}| > A,\ S^2 \geqslant \frac{4(n-1)^3 A^2}{n^2}\right) < \varepsilon^2,\end{aligned} \tag{13}$$

where in the second inequality we use in particular

$$\sqrt{(n-1)S^2} \geqslant |X_1 - \bar{X}| \geqslant (1-1/n)|X_1| - \sum_{2}^{n} |X_i|/n \geqslant 2(n-1)^2 A/n.$$

It follows that $\mathbf{P}(|X_1| > (2n-1)A) < 2\varepsilon^2$ and the rest of the proof is similar to that of Lemma 1. ∎

REFERENCES

[1] S. Bernstein, *On a characteristic property of the normal law*, Trud. Leningrad Poly. Inst. 3 (1941), pp. 21–22. *Collected Works* 4 (1964), pp. 314–315. (In Russian.)

[2] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, Cambridge 1970.

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York 1971.

[4] R. C. Geary, *Distribution of Student's ratio for non-normal samples*, J. Roy. Statist. Soc. Suppl. 3 (1936), pp. 178–184.

[5] A. M. Kagan, Y. V. Linnik and C. Radakrishna Rao, *Characterization Problems in Mathematical Statistics*, Wiley, New York 1978.

[6] T. Kawata and H. Sakamoto, *On the characterisation of the normal population by the independence of the sample mean and the sample variance*, J. Math. Soc. Japan 1 (1949), pp. 111–115.

[7] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin, London 1969.

[8] H. O. Lancaster, *The characterization of the normal distribution*, J. Austral. Math. Soc. 1 (1960), pp. 247–254.

[9] Y. V. Linnik, *A remark on Cramér's theorem on the decomposition of the normal law*, Theory Probab. Appl. 1 (1956), pp. 435–436.

[10] E. Lukacs, *A characterization of the normal distribution*, Ann. Math. Statist. 13 (1942), pp. 91–93.

[11] J. Marcinkiewicz, *Sur une propriété de la loi de Gauß*, Math. Z. 44 (1938), pp. 612–618.

[12] P. A. P. Moran, *An Introduction to Probability Theory*, Clarendon, Oxford 1968.

[13] V. P. Skitovitch, *On a property of the normal distribution*, Dokl. Akad. Nauk SSSR (N.S.) 89 (1953), pp. 217–219.

[14] – *Linear forms in independent random variables and the normal distribution*, Izv. Akad. Sci. SSSR Mat. Ser. 18 (1954), pp. 185–200. (Selected Translations in Math. Statist. and Probab. 2 (1962), pp. 211–228.)

[15] A. A. Zinger, *Independence of quasi-polynomial statistics and analytic properties of distributions*, Theory Probab. Appl. 3 (1958), pp. 247–265.

School of Mathematics and Statistics
University of Sydney
NSW 2006, Australia

Received on 20.12.1993