*The Annals of Statistics*
1977, Vol. 5, No. 3, 583-592

A. M. KAGAN, YU. V. LINNIK AND C. RADHAKRISHNA RAO, *Characterization Problems in Mathematical Statistics* (translated from the Russian by B. Ramachandran). John Wiley and Sons, New York, 1973, xii+499 pages, $26.25.

Review by P. DIACONIS, I. OLKIN and S. G. GHURYE
*Stanford University*; *Stanford University*; *University of Alberta and University of Victoria*

**Introduction.** The study of characterizations of probability distributions has had a long history. Indeed, Gauss (1809) proved (under certain restrictions) that the maximum likelihood estimator of the location parameter of a distribution is the sample mean if and only if the distribution is normal.

In general, a characterization problem takes the following form: Suppose that for a random vector $X$, there is a family $\mathscr{F}$ of distributions such that $\mathscr{L}(X) \in \mathscr{F}$ implies that $X$ has a certain property $\mathscr{P}$. The characterization problem is the converse, namely, to show that if the random vector $X$ exhibits property $\mathscr{P}$, then $\mathscr{L}(X) \in \mathscr{F}$.

There are two ingredients in a characterization problem: the family of distributions $\mathscr{F}$, and the property $\mathscr{P}$. The list of possible families includes the normal, exponential, gamma, beta, geometric, Poisson, Cauchy, and Wishart distributions, but the normal distribution receives most of the attention. The various possible properties cannot be described very succinctly, but the following brief topic headings provide some loose indications: identical distributions of specified functions of $X$, independence of specified functions of $X$, convolutions of the distribution of $X$, functions of $X$ having specified regression on other functions of $X$, functions of $X$ being maximum likelihood estimators (or admissible estimators) of an unknown parameter of the distribution of $X$.

Although Pólya proved an elegant result in 1923 characterizing the normal distribution by the identical distribution of two linear statistics, characterization problems did not begin to attract serious attention until 1935 when Paul Lévy conjectured that if $X + Y$ is normal,[1] then $X$ and $Y$ are normally distributed, that characterization problems began to attract some interest. Cramér (1936) proved Lévy's conjecture, while Raikov (1937) proved a similar result for the Poisson distribution. Marcinkiewicz (1939) proved a result related to that of Pólya mentioned above, and Kac (1939) and Bernstein (1941) proved that $X + Y$ independent of $X - Y$ implies that $X$ and $Y$ are normal. There was a modest amount of activity in the 1940's, followed by a rapid growth of activity in the 1950's.

There have been few expository works covering this material; of import are the review paper by Lukacs (1956), and the monograph by Lukacs and Laha

Received February 1976; revised July 1976.

[1] In the remainder of this review it will be tacitly assumed that all random variables are independently distributed.

(1964). The book under review is the first extensive work on the subject. It is an outgrowth of a set of lectures originally presented in Russian; the present volume is a translation (by B. Ramachandran) of the Russian text. We have not checked the accuracy of the translation. The book does not purport to be a complete review of the field. As the authors state in the preface, they have collected together many results and formulated problems which appear to them to be of interest and importance. As a consequence, some topics and references that would have been included in a comprehensive review of the field are missing from the book.

A comment about the bibliography is in order. Because the book is a translation, the references are listed alphabetically according to the Cyrillic alphabet. This leads to a sequence such as Bernstein, Blackwell, Van der Waerden, Weber, Wolfowitz, Hajek, Giri. In addition, there are some late additions at the end of the bibliography. This word of caution should help the reader to locate specific references. Perhaps a more serious deficiency for the historically minded reader is that the use of Jones [5] rather than Jones (1950) detracts considerably from trying to follow the chronology of results. The use of years would also have permitted an easy rearrangement of authors.

The book contains thirteen chapters, plus one chapter of unsolved problems and two addenda. We now describe these in some detail.

**1. Review of principal tools.** The mathematical tools needed in some areas of the field will be alien to many probabilists and statisticians; Chapter 1 contains a welcome review listing some of the mathematical tools used in the book. The techniques least likely to be known include some basic results of algebraic geometry and the detailed analytic study of the solutions of certain functional, integral and differential equations. A functional equation (or variants thereof) that often arises in characterization problems is $\psi_1(u + b_1 v) + \cdots + \psi_r(u + b_r v) = A(u) + B(v) + P_k(u, v)$, where $P_k$ is a polynomial, $\psi_i$, $A$ and $B$ are complex valued functions subject to various regularity conditions, and $u$, $b_i$ and $v$ are (possibly matrix valued) variables. This equation is considered in some detail.

It was first noticed by Laha (1953) that some characterization results do not use full independence $E\exp(itX + isY) = (E\exp itX)(E\exp isY)$, but only (*) $E(Y\exp itX) = (EY)(E\exp itX)$. Further, $Y$ has constant regression on $X$, i.e., $E(Y|X) = EY$, if and only if (*) holds for all real $t$. A number of techniques are needed to carry through the arguments after weakening the assumption of independence to that of constancy of regression. These techniques are discussed in Chapter 1.

This chapter also contains a straightforward review of facts about characteristic functions, infinitely divisible laws and related topics. A particularly nice result proved in detail is Zinger's theorem on the existence of the moments in terms of the convergence of the integral $\int_0^\infty [1 - F(ax)]^{-\Delta}\, dF(x)$, $\Delta > 0$.

**2. Linear statistics.** The second chapter brings up to date the first modern characterization result, namely that of Pólya (1923): If $X$ and $Y$ are independent identically distributed, have moments of all orders, and are such that $X$ and $aX + bY$ $(a, b \neq 0)$ have identical distributions, then $X$ and $Y$ are normal. In 1938 Marcinkiewicz extended this to the same conclusion if $aX + bY$ and $cX + dY$ are identically distributed. Now, if $X$, $Y$ are independent with the same distribution $F$, then $aX + bY$ and $cX + dY$ are identically distributed for a suitable choice of $a$, $b$, $c$, $d$ if $F$ is any symmetric stable law. Consequently, additional conditions are needed in order for $F$ to be normal. The main thrust of the research presented here is to extend the result to two linear statistics in any finite number of summands and to weaken the original moment restrictions.

Readers should be warned that Chapter 2 is the most technically intricate chapter in the book. It presents Zinger's simplifications and extensions of Linnik's 1953 work on this problem. This work is too complex to be adequately described in the space of this review, and we give only a sketch of the ideas.

Let $X_1, X_2, \cdots, X_n$ be i.i.d., then the identical distribution of $Y_1 = \sum_1^n a_i X_i$ and $Y_2 = \sum_1^n b_i X_i$ implies that each $X_i$ is normal, provided that certain conditions defined in terms of the $a_i$ and $b_i$ are satisfied. If $W$ is the log of the characteristic function of the law of $X_1$, then $Y_1$ and $Y_2$ have the same distribution if and only if $\sum_1^n W(a_j t) = \sum_1^n W(b_j t)$. The essence of the proof is to introduce a transform of $W$ and to study the properties of $W$ through the behaviour of this transform in the complex plane. This leads to a study of the complex zeros of $\sigma(z) = \sum_1^n (a_i^z - b_i^z)$ in restricted parts of the plane, the results of which provide a proof of the theorem.

Using very elaborate machinery, Linnik (1953) proved the result stated at the beginning of the previous paragraph. Zinger (1969) provided considerable simplification of this proof, but even this "simplified" proof contains substantial difficult analysis. Typical of the work on this problem is the following result due to Linnik: Suppose $\max(|a_1|, \cdots, |a_n|) \neq \max(|b_1|, \cdots, |b_n|)$; let $\gamma$ be the largest real zero of $\sigma(z)$, and suppose the moment of order $\gamma + 2$ of $X_1$ is finite, then $X_1$ is normal.

This chapter also has two brief extensions of the theorem characterizing the normal law by the identical distribution of a pair of linear forms. One extension is to multivariate distributions, and the other uses a clever duality argument to obtain a characterization result in terms of certain identically distributed maxima.

One of the simplest characterizing properties of the normal distribution is the independence of a pair of linear forms in independent random variables. The result, that if $X$, $Y$ are independent random variables such that $X + Y$ and $X - Y$ are independent, then $X$ and $Y$ are normally distributed, was first essentially proved independently by M. Kac (1939) and S. N. Bernstein (1941). A complete proof of the extension of this characterization result to an arbitrary number of random variables was first given by V. P. Skitovič (1953), who used the method of exhaustion of possible cases (perhaps even of author and reader); an elegant proof was provided by A. Zinger and Yu. V. Linnik (1955). Various extensions (to an infinite number of multi-dimensional random variables, etc.) have since been obtained.

Chapter 3 deals with these topics. Section 3.1 presents the proof by Zinger and Linnik (without explicit mention of the authors) of the fact that if two linear forms in a finite number of independent random variables are independent, then each random variable which appears in both forms is normally distributed. Section 3.2 deals with multivariate extensions of this result, and Section 3.3 with countably infinite dimensional linear forms. In Sections 3.4 and 3.5 modifications to "relativistic" linear statistics and applications to statistical physics are discussed.

These last two sections contain material which might be particularly interesting to the general reader of the book, since they establish a connection with areas such as relativistic and molecular physics and cosmology. Unfortunately there are no precise references to help guide the reader. For example, on page 96 (line 5 from below), the authors mention that the relation between the addition law for the hyperbolic tangent and Einstein's relativistic addition of collinear velocities was first pointed out by A. Sommerfeld (1969); but the original reference is omitted. For the convenience of the original Russian edition of the book, the authors refer to a Russian book on relativity by Fok; the convenience of readers of the English edition might have been served better by the English translation of Fok's book which had existed for some years before the book under review was written. Actually, the reviewers have been unable to locate any reference to Sommerfeld in Fok's book. On the other hand, Henri Arzeliès, (1966) gives an excellent historical discussion and the reference to Sommerfeld's paper.

Some results (due to Khatri, Kotlarski, Rao), dealing with characterizations through properties of linear functions, would normally be included in Chapter 3. These were not available when the original Russian edition went to press, and are treated in Appendix A.

In 1936, R. C. Geary proved, under the restrictive assumption of the existence of the moment generating function, that the stochastic independence of the mean and variance of a random sample from a population implies normality of the underlying distribution. The restrictions were eliminated by T. Kawata and H. Sakamoto in 1949, and independently by A. Zinger in 1951. These results have been extended in several directions to various special cases of the independence of a linear and a nonlinear polynomial statistic, of two nonlinear polynomial statistics, independence of "quasipolynomial" statistics, and finally independence of pairs of linear forms whose coefficients are random variables.

Chapter 4 deals with the topics listed in the previous paragraph. Section 4.2 is devoted to the independence of the mean and variance of a sample of i.i.d. random variables, and to extensions of this problem in which the mean is replaced by a more general linear form and the variance by a more general quadratic form. Sections 4.3 and 4.4 are concerned with the consequences of independence of "quasipolynomial" statistics in independent (but not necessarily identically distributed) random variables. A "quasipolynomial" statistic $S$ is a function with the property that there exists a continuous function $\varphi$ and polynomials $p$ and $q$, both of the same degree, such that $p(x) \leqq \varphi(S(x)) \leqq q(x)$ for all $x$.

**3. Regression and structural properties.** Chapter 5 begins with the result that if $X_1, \cdots, X_n$ $(n \geqq 3)$ are i.i.d. with $EX_i = 0$ and $E(\Sigma X_i \mid X_1 - \bar{X}, \cdots, X_n - \bar{X}) = 0$, then the $X_i$ are normally distributed. The result is then generalized to permit other linear combinations instead of $X_i - \bar{X}$. A complete analysis is provided for the case of $n = 2$. The condition $E(X_1 - \alpha X_2 \mid X_1 + \beta X_2) = 0$ for $|\beta| \leqq 1$, and more generally $E(\Sigma a_i X_i \mid \Sigma b_i X_i) = 0$, is a weakening of the Kac–Bernstein assumption by replacing independence with constant regression. The proof depends on the solution of the functional equation $f(t) = \prod_1^{n-1} [f(\beta_i t)]^{\gamma_i}$, where $f$ is the characteristic function of the $X_i$. A final extension to infinite sums is then provided. In this latter case, the solution of the functional equation (*) $f(t) = \prod_1^\infty [f(\pm\beta_j t)]^{\gamma_j}$, $0 < \beta_j < 1$, $0 < \gamma_j$, $j = 1, 2, \cdots$, is needed, where $f$ is a nonvanishing characteristic function not identically equal to 1. A first step is to show that if the $\gamma_j$ are positive integers, then (*) implies that $f$ cannot vanish. Once this is proved, it is shown that $f$ is the characteristic function of a normal law if and only if $\sum_1^\infty \gamma_j \beta_j^2 = 1$.

Characterizations by the constancy of regression for nonlinear statistics is the topic of Chapter 6. In 1937, E. J. G. Pitman showed that if $X_1, \cdots, X_n$ are i.i.d. having a gamma distribution, $\sum X_i$ is independent of every scale-invariant statistic $S(X_1, \cdots, X_n)$, i.e., for every $S$ such that $S(X_1, \cdots, X_n) = S(aX_1, \cdots, aX_n)$ for all real $a \neq 0$. This property is suggestive of a number of characterizations. The first result proved (by Khatri and Rao (1968)) is that if $X_1, \cdots, X_n$ $(n \geqq 3)$ are positive random variables (not necessarily identically distributed) with finite expectations, and if the regression of $\sum X_i$ on the vector $(X_2/X_1, \cdots, X_n/X_1)$ is constant, then each $X_j$ has a gamma distribution with the same scale parameter. This result is extended in a variety of ways. There is little motivation or discussion of the development of these results.

Chapter 10 concerns multivariate versions of some results presented in earlier chapters. For example, let $A$ and $B$ be two constant matrices with the property that no column of $A$ is proportional to any column of $B$. If $X$ is a $p$-dimensional random vector $(p \geqq 2)$ having the representations, $X = AU$ and $X = BV$, where $U$ is an $r$-dimensional random vector and $V$ is an $s$-dimensional random vector, then the $X_i$'s have a normal distribution. A discussion of factor analytic models and other structural equations models is included.

**4. Statistical properties.** Chapter 7 is entitled "Characterizations of Distributions Through the Properties of Admissibility and Optimality of Certain Estimators." The results in this chapter are different in flavor from the previous results in the book. The underlying idea is to translate statistical optimality properties to properties of constant regression. Suppose that $X_1, \cdots, X_n$ $(n \geqq 3)$ are independent random variables whose distributions depend on a location parameter $\theta$, i.e., $P_\theta\{X_j \leqq x\} = F_j(x - \theta)$, $j = 1, \cdots, n$, and having $EX_j = 0$, $0 < EX_j^2 = \sigma_j^2 < \infty$. The optimal linear unbiased estimator for $\theta$ (with quadratic loss function) is $L = \sum c_j X_j$, where $c_j = \sigma_j^{-2}/\sum \sigma_i^{-2}$. If we start with a normal family, then it is well known that $L$ is admissible (in the class, $\mathscr{U}$, of all unbiased estimators of $\theta$). The new feature is that if $L$ is admissible in $\mathscr{U}$, then the df's must be normal. The condition $n \geqq 3$ is shown to be essential; the admissibility of $L$ in $\mathscr{U}$ is not a characterizing property of the normal distribution for $n = 2$.

Once it is shown that an admissibility condition implies a constancy of regression, there are a myriad of possible extensions. For example, consider a Gauss–Markov model in which $Y_1, \cdots, Y_n$ are independent, $EY_i = \theta_i$, and $\theta = C\beta$, where $\theta' = (\theta_1, \cdots, \theta_n)$, $C$ is a known $n \times m$ matrix and $\beta$ is an unknown $m \times 1$ vector. Suppose we wish to estimate $a'\beta$, where $a$ is prescribed vector. Let $L$ be that linear combination of the $Y_j$ that is the least squares estimator of $a'\beta$. Then under conditions on the rank of $C$, and using quadratic loss, the admissibility of $L$ in the class of unbiased estimators of $a'\beta$ is again a characterizing property of the normal distribution.

Section 7.8 deals with the case of dependent observations, where an autoregressive model is assumed, and provides a very nice extension of the admissibility result concerning the linear statistic $L$. In the remainder of the chapter, other loss functions (e.g., the absolute value) and other families of distributions are discussed. Some previously unpublished results characterizing families of distributions by means of properties of Bayes estimators are contained in Appendix B.

Chapter 8 contains a straightforward presentation of well-known results about sufficiency. The material covered includes definitions, the Halmos–Savage version of the factorization criterion, and finally Dynkin's derivation of the Koopman–Pitman–Darmois result that the only family of distributions with range independent of the parameter that admits a nontrivial sufficient statistic is the exponential family. Significant contributions to Dynkin's treatment were made by Brown (1964).

The chapter also contains some of the less well-known results on sufficiency. If location (scale) families admit the sample mean as a sufficient statistic, then the underlying distribution is normal (gamma).

The final section of Chapter 8 gives a clear, short treatment of Kagan's notions of sufficient subspaces. At the present time only very simple cases are well understood, but the notion could lead to a significant generalization of the exponential families. No mention is made of the work of Dynkin (1961) and others on characterizations of nonregular families admitting sufficient statistics.

**5. Some related and unrelated results.** Chapter 11 contains no characterization results; it is devoted to the study of various properties of functions of polynomial statistics from a multivariate normal distribution. Two of the most useful facts about $n$-dimensional normal vectors are that they can be related to $n$-dimensional vectors with independent standard normal coordinates by linear transformations, and that the distribution of vectors with independent normal coordinates with the same variance remains unchanged by orthogonal linear transformations. Chapter 11 deals with the question of what happens when linear transformations are replaced by polynomials, rational functions, entire functions or meromorphic functions. For example, if $X$ and $Y$ are independent standard normal, then $2XY/(X^2 + Y^2)^{\frac{1}{2}}$ and $(X^2 - Y^2)/(X^2 + Y^2)^{\frac{1}{2}}$ are independent standard normal. The authors initiate a classification of such phenomena, but the problems are far from simple. For example, it is not known if the distribution of a general polynomial of degree 3 in $n$ standard normal variables is determined by its moments.

The bulk of this chapter is devoted to discussing the converse of the following proposition: Let $X = (X_1, \cdots, X_n)$ be a vector of independent and identically distributed normal random variables, and $P$ and $Q$ be functions of $n$ variables, such that $P(X)$ and $Q(X)$ are independent. If $\Gamma$ is any orthogonal transformation, then it is immediate that $P(X\Gamma)$ and $Q(X\Gamma)$ are independent. In particular, if $P$ and $Q$ depend on disjoint subsets of the coordinates, they will be independent. The authors consider the converse of this proposition for the case where $P$ and $Q$ are polynomials; that is, they try to establish the truth of the conjecture that every pair of independent polynomial statistics can be "unlinked" in the sense that there exists an orthogonal transformation $\Gamma$ such that $P(X\Gamma)$ and $Q(X\Gamma)$ are functions of disjoint subsets of coordinates in the transformed variables $X\Gamma$.

The results to date are far from complete, and the methods developed are fairly complex. The most important cases in which the converse has been verified all have a symmetry assumption attached to them; although Zinger and Linnik (1967) have shown that the unlinking conjecture holds for a "large proportion" of polynomial pairs $P$ and $Q$. The definition of "large proportion" has to do with the topological dimension of the spaces of coefficients associated with $P$ and $Q$. The proofs make heavy use of the algebraic properties of the variety in the coefficient space generated by the relations $E(P^nQ^m) - E(P^n)E(Q^m) = 0$.

The material of Chapter 12 is different from the major part of the book in that what are being characterized are sequential estimation plans rather than distribution functions. The key result is that for estimating the parameter, $p$, of a binomial process, a complete finite sequential plan of size $n$ is completely determined by the values of the mean stopping time, for any $(n + 1)$ distinct values of $p$. The proofs used in this chapter are similar to the proofs of Girshick, Mosteller, and Savage (1946) and DeGroot (1959).

**6. Miscellaneous characterizations.** Chapter 13 contains miscellaneous results, some of which are deep and complex. Various members of the exponential family are characterized by the property of having minimum Fisher information or minimum entropy, subject to constraints on the expectation of a fixed finite set of functions. Such properties have been proposed to justify the use of particular prior distributions in a Bayesian context [e.g., Jaynes (1968)]. In this context, maximum entropy has been used by Posner (1975) to choose a particular multivariate distribution with given marginals.

We illustrate the usefulness of Section 13.4 by means of an example. R. A. Fisher's approach toward determining the correct number of degrees of freedom in the chi-squared goodness of fit test for the multinomial distribution was to use the fact that if $n$ is chosen from a Poisson distribution and then a sample of size $n$ is drawn by flipping a coin $n$ times, the unconditional number, $s$, of heads has a Poisson distribution and is independent of the number of tails. One easily sees that if the Poisson distribution is replaced by the geometric distribution, then $s$ has a geometric distribution. Unfortunately, in this case $s$ is not independent of the number of tails. Theorem 13.4.4 shows that if $s$ is independent of the number of tails, then $n$ has a Poisson distribution. This shows that useful variations of Fisher's result are not possible.

Section 13.5 deals with the problem of characterizing families of distributions by means of the distributions of sample statistics. The problem is formulated more precisely as follows: Let $\mathscr{P}$ be a family of probability laws, and let $X_1, \cdots, X_n$ be independent random variables (or vectors) all having the same distribution $P \in \mathscr{P}$. Let $T$ be a function of $(X_1, \cdots, X_n)$ with the property that its distribution $F_{\mathscr{P}}$ is the same for all $P \in \mathscr{P}$; then what pairs $(\mathscr{P}, T)$ do there exist with the property that $T$ has the distribution $F_{\mathscr{P}}$ only if the common distribution of the $X_j$ is an element of $\mathscr{P}$? For each of the three location-and-scale families generated by the standard normal, the standard exponential and the standard uniform, there exists a characterizing statistic whose distribution is uniform over an appropriate surface.

When several samples are available which are known to have come from the same family (though perhaps not the same distribution), tests of uniformity of distribution can then be used to identify the family. In the case of the normal family, the surface of uniformity is a sphere, and the procedure for testing for uniformity is relatively simple. On the other hand, the statistical problem is more difficult for the exponential and uniform distributions. A discussion of this problem is given by Giné (1975).

**7. Scope for further work.** Whereas most work in the area of characterization of distributions deals with the precise characterization of distributions, from the practical point of view it is often adequate to know whether a particular property characterizes a distribution sufficiently closely. To make this notion precise, let us consider the Lévy–Cramér theorem, which states that if $X$, $Y$ are independent and $X + Y = Z$ is normal, then $X$ and $Y$ are normal. Now, we ask ourselves: What can be said about the distributions of $X$, $Y$ if $Z$ is $\varepsilon$-normal in the sense that its cdf is within a distance $\varepsilon$ of the standard normal cdf (distance being measured in the sup norm)?

The short Chapter 9 deals with such questions; its shortness is due to the paucity of work on such problems, which in turn is due to the fact that most of these questions lead immediately to quite difficult problems.

In the opinion of the reviewers, this area provides opportunity for new and worthwhile contributions to the field of characterization. These would stem from the obverse approach to the problem: whereas the attitude of Chapter 9 is to regard a characterization as relatively "stable" if a small perturbation in the characterizing property leads to only a small variation in the admissible distribution, one could regard a characterizing property as being relatively "precise" if a small perturbation in it permits a large class of underlying distributions. It appears clear that powerful mathematical tools will be required to obtain results in this general area.

Chapter 14 lists unsolved problems which are of interest and importance.

**8. Conclusion.** To summarize, there is a wealth of information in this book, some of which was previously unavailable in English. The writing tends to be in the theorem-proof format. Although some discussions are provided, they do not suffice for the general occasional reader, and the novice will not find the book immediately readable. However, the monograph was written for the researcher in the field, for whom the book will be exceedingly useful. The sheer amount of material covered by the authors makes this an essential and indispensable book to the specialist.

## REFERENCES

[1] Arzeliès, Henri (1966). *Relativistic Kinemetrics* (English translation). Pergamon, Oxford, New York.

[2] Bernstein, S. N. (1941). Sur une propriété caractéristique de la loi de Gauss. *Trans. Leningrad Polytechn. Inst.* **3** 21–22.

[3] Brown, L. D. (1964). Sufficient statistics in the case of independent random variables. *Ann. Math. Statist.* **35** 1456–1474.

[4] Cramér, H. (1936). Über eine Eigenschaft der normalen verteilungsfunktion. *Math. Z.* **41** 405–414.

[5] DeGroot, M. H. (1959). Unbiased sequential estimation for binomial populations. *Ann. Math. Statist.* **30** 80–101.

[6] Dynkin, E. B. (1951). Necessary and sufficient statistics for a family of probability distributions. *Uspehi Matem. Nauk* (*N.S.*) **6** No. 1(41), 68–90. (In *Selected Translations in Mathematical Statistics and Probability*, **1** 1961, 23–40. American Mathematical Society.)

[7] Gauss, C. F. (1809). Theoria motus corporum coelestium. Liber II, Section III, 240–244. Gauss Werke Band VII.

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

[9] Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266.

[10] Girshick, M. A., Mosteller, F. and Savage, L. J. (1946). Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statist.* **17** 13–23.

[11] JAYNES, E. T. (1968). Prior probabilities. *IEEE Trans. Systems Sci. Cybernet.* **4** 227–241.
[12] KAC, M. (1939). On a characterization of the normal distribution. *Amer. J. Math.* **61** 726–728.
[13] KAWATA, T. and SAKAMOTO, H. (1949). On the characterization of the normal population by the independence of the sample mean and the sample variance. *J. Math. Soc. Japan* **1** 111–115.
[14] KHATRI, C. G. and RAO, C. R. (1968). Some characterizations of the gamma distribution. *Sankhyā Ser. A* **30** 157–166.
[15] LAHA, R. G. (1953). On an extension of Geary's Theorem. *Biometrika* **40** 228–229.
[16] LINNIK, YU V. (1953). Linear forms and statistical criteria I, II. *Ukrain. Mat. Ž.* **5** 207–243; 247–290.
[17] LUKACS, E. (1956). Characterization of populations by properties of suitable statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 195–214, Univ. of California Press.
[18] LUKACS, E. and LAHA, R. G. (1964). *Applications of Characteristic Functions.* Griffith Statistical Monographs, No. 14. Hafner, New York.
[19] MARCINKIEWICZ, J. (1939). Sur une propriété de la loi de Gauss. *Math. Z.* **44** 612–618.
[20] PITMAN, E. J. G. (1937). The (closest estimates) of statistical parameters. *Proc. Cambridge Philos. Soc.* **33** 212–222.
[21] PÓLYA, G. (1923). Herleitung des Gauszschen Fählergesetzes aus einer Funktionalgleichung. *Math. Z.* **18** 96–108.
[22] POSNER, E. G. (1975). Mutual information for constructing joint distributions. *Utilitas Math.* **7** 3–23.
[23] RAĬKOV, D. A. (1937). On the decomposition of Poisson laws. *Dokl. Akad. Nauk SSSR* **14** 9–11.
[24] SKITOVIČ, V. P. (1953). On a property of the normal distribution. *Dokl. Akad. Nauk SSSR* **89** 217–219.
[25] SOMMERFELD, A. (1909). Über die Zusammensetzung der geschwindigkeiten in der Relativtheorie. *Phys. Z.* **10** 826–829.
[26] ZINGER, A. (1951). On independent samples from a normal population. *Uspehi Mat. Nauk* **6** 172–175.
[27] ZINGER, A. (1969). *Investigations into Analytical Statistics and Their Applications to Limit Theorems of Probability Theory.* Ph. D. dissertation, Leningrad Univ.
[28] ZINGER, A. and LINNIK, YU V. (1955). On an analytic generalization of a theorem of Cramér and its application. *Vestnik Leningrad. Univ.* **10** No. 11, 51–56.
[29] ZINGER, A. and LINNIK, YU V. (1967). Polynomial statistics of a normal sample. *Dokl. Akad. Nauk SSSR* **176** 766–767.

P. DIACONIS
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

S. G. GHURYE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, B.C. VGT 1W5
CANADA

I. OLKIN
IMPERIAL COLLEGE
DEPARTMENT OF MATHEMATICS
QUEENS GATE
LONDON SW7, ENGLAND