KENNETH S. MILLER, *Multidimensional Gaussian Distributions*. John Wiley and Sons, Inc., New York, 1964. $9.50. viii + 129 pp.

Review by HAROLD RUBEN

University of Minnesota

Two things should be said about this book. (1) The title is a misnomer. Of a total of 112 pages of text, 42 pages are devoted to a very special distribution associated with the multivariate Gaussian distribution, namely, the multivariate Rayleigh distribution. This is clearly the heart of the book. Only 36 pages, of which 17 are concerned nominally with Gaussian noise, actually deal with the subject named in the title. The remaining pages form an unclassifiable group concerned with various elementary topics which are not inherently connected with Gaussian distributions. A breakdown of this group results in the following list: diagonalization of a quadratic form, covariance matrices, inversion of a partitioned matrix, transformation of Cartesian to polar coordinates in n-space, the Fourier inversion formula and, finally, least squares estimation. (2) The discussion on the Gaussian distribution is far from connected, either spatially or logically. The same applies with yet greater force to the discussion on least squares estimation which does not always appear under that title and is in addition unduly repetitious, not to say downright muddled.

The p-variate Rayleigh distribution, as defined by the author, is the joint distribution of p correlated χ (not χ^2) variates, forming a Rayleigh random vector. More precisely, it is the distribution of the vector $(r_1, \cdots, r_p)$, where r_α^2 is the αth diagonal element of the matrix[1] $\mathbf{A} = \sum_1^n \mathbf{x}_\alpha \mathbf{x}_\alpha'$ and the $\mathbf{x}_\alpha$ are independent p-component normal vectors with arbitrary expectation vectors $\boldsymbol{\mu}_\alpha$ and a common positive definite covariance matrix $\boldsymbol{\Sigma}$. Some light is thrown on the distribution by noting that r_α^2 is a multiple of a (in general non-central) χ^2 variate, and further that in principle (but presumably only in principle) the joint distribution of the r_α^2 could be obtained by integrating out the "crossproduct" variates in the non-central Wishart distribution of $\mathbf{A}$, first studied apparently by T. W. Anderson [1], [2]. (A warning is in order here. The author's definition of the multivariate Rayleigh distribution (p. 27) is sloppy and strictly speaking vacuous. That 'definition' imposes normality on each of the p n-component vectors obtained by taking corresponding components of the $\mathbf{x}_\alpha$, and imposes only independence on the $\mathbf{x}_\alpha$ themselves.) The density functions of the distribution for $p = 1, 2, 3$ and for special values of the $\boldsymbol{\mu}_\alpha$ (mostly $\boldsymbol{\mu}_\alpha = \mathbf{0}$), as well as for general p with $\boldsymbol{\Sigma}$ a continuant and $\boldsymbol{\mu}_\alpha = \mathbf{0}$, are given in six theorems. A seventh theorem gives the density in symbolic form for arbitrary p, $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\Sigma}$. Three theorems deal respectively with the densities of the product of norms, the inner product and the angle between two correlated Gaussian vectors, and a further three theorems

[1] Here, as elsewhere in this review, I am adapting the author's notation to conform to common statistical usage.

deal with the quotients of norms of such vectors. There are also some additional theorems concerned mainly with the sum and difference of the squared norms of two (real-valued) Rayleigh variates, and some results on moments. Many of these distributional results, under a slightly different guise, are classical, and a few others will probably have been derived by most readers of this journal at some point in their statistical careers. Examples of these follow: Rayleigh distribution for $p = 1$ (essentially, the non-central χ^2 distribution), distribution of the ratio of two nonindependent Rayleigh variates (essentially, the distribution of the ratio of variances in samples from a bivariate normal population, due to Bose [5] and Finney [7]), distribution of the angle between two correlated Gaussian vectors (essentially, Hotelling's variant[2] of the distribution of the correlation coefficient in normal samples [9]), distributions of the weighted sum and difference of two independent χ^2 variates (expressed in terms of confluent hypergeometric and modified Bessel functions), distribution of the ratio of two independent non-central Rayleigh variates (essentially the distribution of a doubly non-central F, expressed in Tang's form [21] as an infinite series of Snedecor F-densities, of which the non-central F in analysis of variance applications is an even more familiar case). Other distributions derived may not be generally known, and these are often rather complicated in structure, being typically expressed as multiple infinite series in which each term involves products of Bessel functions of various kinds. The value of such results appears dubious from a numerical point of view. In a publication in the SIAM series one would have expected a discussion of the practical value of the results (rapidity of convergence of the series, etc.), and also of possible applications in physics and technology, in order to justify the rather dull algebra invoked in the derivation.

The author's failure to point out that in dealing with the generalized Rayleigh distribution one is essentially concerned with the joint distribution of quadratic forms of normal variates has the effect of obscuring the true significance of that distribution. Related to this is a failure to mention the important work of Robbins [16] and Robbins and Pitman [17] in which the distribution of a single such quadratic form is expressed as a mixture (either improper, i.e., arbitrary linear combination, or proper, i.e., convex combination) of scaled χ^2-distributions. (See also [18] and [19].) Such a tie-up would suggest, for example, that the general

[2] Define $\varphi = \operatorname{arc\,cos} [\sum_1^n x_\alpha y_\alpha / \{\sum_1^n x_\alpha^2 \sum_1^n y_\alpha^2\}^{\frac{1}{2}}]$, where the (x_α, y_α) are independent bivariate Gaussian vectors, each with zero expectation and positive definite covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})$, $i, j = 1, 2$. Theorem 2 (p. 45) gives

$$[(n - 1)\Gamma(n)/|\mathbf{\Sigma}|^{\frac{1}{2}n}\Gamma(\tfrac{1}{2})\Gamma(n + \tfrac{1}{2})][\sin^{n-2} \varphi / \{(\sigma^{11}\sigma^{22})^{\frac{1}{2}} + \sigma^{12} \cos \varphi\}^n]$$
$$\cdot F(n, \tfrac{1}{2}; n + \tfrac{1}{2}; [\sigma^{12} \cos \varphi - (\sigma^{11}\sigma^{22})^{\frac{1}{2}}/\sigma^{12} \cos \varphi + (\sigma^{11}\sigma^{22})^{\frac{1}{2}}])$$

as the density of φ ($(\sigma^{ij}) = \mathbf{\Sigma}^{-1}$). On setting $r = \cos\varphi$, and using the Gaussian transformation $F(\alpha, \beta; \gamma; x) = (1 - x)^{-\beta} F(\beta, \gamma - \alpha; \gamma; x/(x - 1))$ we obtain the density of r, the sample correlation coefficient in random samples of size $n + 1$ form an arbitrary non-singular bivariate normal population, in the form given by Hotelling. This again involves the hypergeometric function F, but the argument is $(1 + \rho r)/2$, where $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{\frac{1}{2}}$ is the population correlation.

p-variate Rayleigh distribution could probably be expressed, relatively simply, as a p-tuply infinite series of scaled χ^2-distributions. A further related point should be made here. The assertion (p. 31) that the determination of the density function of the norm of an arbitrary Gaussian vector appears to be extremely difficult is false. Actually, the density referred to can be obtained by an elementary conditional probability argument [19]. Incidentally, this density has a rather simple form ([18], [19]).

The discussion on finite-dimensional Gaussian distributions presents, in scattered form (pp. 16–19, 20–26, 71–72, 75–76, 83–84, 88–92), some basic and elementary properties of such distributions (marginal and conditional distributions, linear transformations of a Gaussian vector, moment generating and characteristic functions, linearity of regression). Singular Gaussian distributions are also discussed. This list is noteworthy for its brevity. One would normally expect to find a similar degree of coverage in the first post-introductory chapter of any text book on multivariate analysis (e.g. Anderson [3]). No mention is made of, *inter alia*, quadratic functions of normal variates (including independence of such functions and Cochran's decomposition theorem), characterization properties (in particular, the projection property), the optimum entropy property of the multivariate Gaussian distribution and the multivariate central limit theorem. One might add that since the area of "multidimensional Gaussian distributions" has been most intensively explored by statisticians, any book so entitled, even if aimed primarily (as is apparently this book) at mathematicians and engineers should certainly contain, for the sake of general orientation, a synoptic review, however brief, of sampling theory in its application to multivariate statistical analysis. The treatment would have gained greatly in naturalness and informational content from a geometrical (coordinate-free) approach, and from a definition of the multivariate Gaussian distribution as that of the joint distribution of linear combinations of independent normal variates. Many of the basic properties are then immediate consequences of the definition. A further advantage is that the singular Gaussian distribution arises automatically and in a very natural manner. [The derivation of the "density" of a singular Gaussian distribution in terms of delta functions (pp. 88–92) is both unconvincing and unappealing.] The derivation of the linear regression property (pp. 83–84) is clumsy and possibly misleading: linearity of regression of the vector $\mathbf{y}$ on the vector $\mathbf{x}$ implies that the generalized mean-square linear regression of $\mathbf{y}$ on $\mathbf{x}$ coincides with the former regression (irrespective of whether the composite vector $(\mathbf{x}, \mathbf{y})$ is Gaussian or not), a result which follows directly from an obvious conditional argument and the minimal property of the mean (i.e., $E(z - c)^2$ is minimized for $c = Ez$).

The treatment of least squares estimation, again in highly disjointed form, must be bewildering to the reader with no previous knowledge of the subject. The Gauss-Markov theorem (the description Gauss-Markov is never used) in the form of minimum variance unbiased estimation of a linear parametric function is given on pp. 87–88. (Maximal rank is implied.) On pp. 112–113, Section 4.6, the author considers the estimation of the regression parameters from a finite set of observations when the regression function is a linear combination of time-functions (continuity of these functions is unnecessarily imposed) and the errors are correlated. Why this should appear in a chapter entitled "Some Applications to Gaussian Noise" is a mystery. Next, optimization in the class of linear unbiased estimators is sought on the basis of the odd criterion that the sum of variances of the regression coefficient estimators be minimized. There is no hint that the correlated-errors model is only a slight modification of, and is easily reduced to, the more usual model in the Gauss-Markov theorem, and that the estimation problems (efficient estimation of a linear compound of the parameters) for the two cases are essentially the same (simultaneous minimization of the variances of the regression coefficient estimators being then achieved trivially by taking special linear compounds). The reader's bewilderment will not be lessened by the fact that special cases of the correlated-errors Gauss-Markov theorem (namely, constant regression on pp. 85–86 and a single regression coefficient on p. 97, this last case under the guise of maximum likelihood estimation, the errors being assumed jointly normal) have appeared previously in the book and by the vague statement that the result of Section 4.6 may be regarded "in certain respects" as a generalization of the earlier results.

Potentially, much the most interesting topic in the book is that of Gaussian noise (Chapter 4). However, one's hopes are soon dashed. Actually only about one-quarter of Chapter 4 is devoted to *infinite*-dimensional processes. What there is of this is useful, but it is little indeed. Essentially it consists of (i) a proof that (analogously to a finite-dimensional distribution) the normality of a continuous-time process is preserved under a linear transformation ('filter'), and (ii) the determination of the covariance function of output noise in terms of that of input noise, assumed to be wide-sense stationary, under a time-invariant linear filter with given impulsive response function, or, equivalently, in terms of the spectral density of the input noise and the transfer function of the filter. For (ii) normality is irrelevant, and one wonders again what this is doing in a chapter called "Some Applications to Gaussian Noise." (The justification that a Gaussian process is determined by the mean and covariance functions is lame.) The remainder of the chapter consists of (iii) Section 4.6, already referred to previously, and (iv) maximum likelihood estimation of a parameter entering into the specification of a signal contaminated by Gaussian noise from a finite set of observations of the distorted signal. In half of (iv), namely, when the noise is additive, estimation by maximum likelihood is equivalent to estimation by generalized least squares (i.e., using the correlated-errors Gauss-Markov theorem), though the connection is not brought to light. Omission of the increasingly important topic of inference based on continuous time records (requiring methods for associating probability densities and likelihood ratios to specified waveforms) is particularly regrettable from the point of view of the communications engineer concerned with such estimation and discriminating problems as evaluation of a signal strength or testing for the existence of a weak signal in the presence of strong background noise. (As already mentioned previously, the estimation problems considered by the author are not intrinsically associated with infinite-dimensional processes.) One general and rather powerful line of attack is provided by the Karhunen-Loève representation [10] [12] (series expansion of a mean-square continuous random function on an arbitrary time interval in terms of normalized eigenfunctions and eigenvalues of a homogeneous linear integral equation with the covariance function as kernel)—which is also of fundamental theoretical importance in its own right—or some equivalent formulation (such as that of Parzen's reproducing correlation function kernels in Hilbert space [14]), the problem being thereby reduced to a specification of the joint density of the countable set of (observable) random coefficients in the K-L expansion [8] [20]. The later coefficients are uncorrelated and, in the case of Gaussian noise, jointly normal and therefore also independent. To mention a few more fairly obvious lacunae: there is no mention of the Gaussian limit distribution in shot noise processes, of Wiener processes, of maxima and zero-crossings, and (if the text is not to be devoted mainly to Gaussian distributions) of prediction or design (optimum allocation of observations).

The Fourier inversion formula as stated on p. 76 is incorrect, unless the integral is to be interpreted as a Cauchy principal value. (In the proof, $\lim_{\alpha_k \to \infty} \int_{-\alpha_k}^{\alpha_k} dt_k$ is replaced indifferently by $\int_{-\infty}^{\infty} dt_k$.) Ψ_n, referred to after Equation 2 on p. 95, is a covariance matrix only (and not a correlation matrix in the usual statistical sense). The term 'regression curve' (p. 83) for the regression of one random vector on another is unfortunate. The statement that the Cramér-Rao bound provides an upper bound for the efficiency of an estimator (p. 96) is misleading (the CR bound being, in general, unattainable). In line 8 from the bottom of p. 101, $W_n \partial W_n / \partial \alpha$ should read $(W_n \partial W_n / \partial \alpha)^2$. The formula for the fourth moment of a quadratic function of a $N(\mathbf{0}, \mathbf{I})$ vector (Equation (11) on p. 105 and Equation (4.4.11) on p. 120) is incorrect, and the coefficients 14, 28, 52, 10 should read 12, 32, 48, 12. (Incidentally, all the moments of such a function are computed easily from Lancaster's elegant formula [11] for the cumulants of the function.) The term "reproducing properties" (p. 24) is used in altogether too wide a sense.[3]

To sum up: the author's stated objective of presenting "the basic facts concerning multidimensional Gaussian distributions in a concise, crisp and we hope elegant form" has not been met. Only a very few of the most elementary of such facts have actually been presented, and for these the enquiring reader would do better to refer to, say, Chapter 2 of Anderson [3] or to the appropriate sections in any good statistics textbook, e.g., Cramér [6], where the treatment is superior,

---

[3] The fact the $\mathbf{x} + \mathbf{y}$ is normal if $\mathbf{x}, \mathbf{y}$ are independent and normal—a property which is a consequence of the fact that a linear function of a normal vector is itself normal— is not mentioned, while the "reproducing property"

$$\int_{-\infty}^{\infty} n(\mathbf{x}, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) n(\mathbf{x}, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\, d\mathbf{x} = n(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2),$$

$(n(\cdot, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ being the density of a $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ positive definite) is allowed to take up two pages of proof via elementary matrix algebra.

better motivated and at the same time often more compact. So far as infinite-dimensional Gaussian processes are concerned, the reader will surely gain greater insight from even a cursory reading of judiciously chosen sections in (say) [15] and [13], supplemented on the more statistical side by [4] and [8]. A monograph on Rayleigh distributions, suitably extended and modified along the lines mentioned previously, would in fact have been more appropriate and perhaps more honest. As it is, one gets an overall impression of padding for the sake of a respectably sized book. This impression is strengthened by the peculiar unevenness of level. The reader for whom the book is intended (described as one "familiar with the elementary facts concerning linear algebra" and who has "some acquaintance with advanced calculus and probability") is, to quote one example, on the one hand given a description of the mechanics of reducing a quadratic form to a sum of squares, and on the other is assumed to require no explanation of (or even reference to) Fubini's integral theorem (used on p. 103) or the Wiener-Khintchine relations (quoted on p. 106). I fear that after struggling with this book the earnest but uninitiated reader will, like the Persian poet-philosopher, come out by the same door as in he went.

Not a few statistics books have appeared recently which are insufficiently motivated, poorly organized, substantively thin and bereft of all cultural depth or historical perspective. Perhaps one should hardly expect even reputable publishing firms to be very much concerned about this depressing phenomenon. The primary concern of a publication firm, like every commercial organization, is, after all, to make a profit, and, with the rapid and universal growth in the number and size of university departments, technical institutes, research groups, etc., a publisher can now reasonably expect, at the very least, to break even with any scientific book from sales to libraries alone. Dare one hope that writers and would-be writers of scientific books will, for their part, return to an old scholarly tradition by refraining from rushing into print unless and until they have something worthwhile and substantial to say?

REFERENCES

[1] ANDERSON, T. W. (1945). The non-central Wishart distribution and its application to problems in multivariate statistics. Doctoral dissertation, Princeton University Library.

[2] ANDERSON, T. W. (1946). The non-central Wishart distribution and certain problems of multivariate statistics. *Ann. Math. Statist.* **17** 409–431.

[3] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[4] BARTLETT, M. S. (1956). *An Introduction to Stochastic Processes.* Cambridge Univ. Press.

[5] BOSE, S. S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population. *Sankhyā* **2** 65–72.

[6] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[7] FINNEY, D. J. (1938). The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population. *Biometrika* **30** 190–192.

[8] GRENANDER, ULF (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195.

[9] HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *J. Royal Statist. Soc. B* **15** 193–225.

[10] KARHUNEN, KARI (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn. Ser. A. I.* **37**.

[11] LANCASTER, H. O. (1954). Traces and cumulants of quadratic forms in normal variables. *J. Royal Statist. Soc. B* **16** 247–254.

[12] LOÈVE, MICHEL (1948). Fonctions aléatoires de second ordre. [In Paul Lévy, (1948). *Processus Stochastiques et Mouvement Brownien.* Gauthiers-Villars, Paris.]

[13] PAPOULIS, ATHANASIOS (1965). *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, New York.

[14] PARZEN, EMANUEL (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[15] PARZEN, EMANUEL (1962). *Stochastic Processes.* Holden-Day, Inc., San Francisco.

[16] ROBBINS, HERBERT (1948). The distribution of a definite quadratic form. *Ann. Math. Statist.* **19** 266–270.

[17] ROBBINS, HERBERT and PITMAN, E. J. G. (1949). Application of the method of mixtures to quadratic forms in normal variables. *Ann. Math. Statist.* **20** 552–560.

[18] RUBEN, HAROLD (1962). Probability content of regions under spherical normal distributions IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. *Ann. Math. Statist.* **33** 542–570.

[19] RUBEN, HAROLD (1963). A new result on the distribution of quadratic forms. *Ann. Math. Statist.* **34** 1582–1584.

[20] SELIN, IVAN (1962). The sequential estimation and detection of signals in normal noise. Memorandum RM-2994-PR, Rand Corporation, Santa Monica, California.

[21] TANG, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Statist. Res. Mem.* **2** 126–157.