functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 1951, pp. 53–73.

[6] I. J. Good, *Probability and the Weighing of Evidence*, Charles Griffin & Company Limited, London, 1950.

[7] I. J. Good, "Some terminology and notation in information theory," *Monograph No. 155R*, The Institution of Electrical Engineers, London, 1955.

[8] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.

[9] S. Kullback, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.

[10] S. Kullback, "Certain inequalities in information theory and the Cramér-Rao inequality," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 745–751.

[11] S. Kullback, "An application of information theory to multivariate analysis, II," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 122–146.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 79–86.

[13] M. Kupperman, "Further applications of information theory to multivariate analysis and statistical inference, (preliminary report)," (abstract), *Ann. Math. Stat.*, Vol. 27 (1956), p. 1184.

[14] M. Kupperman, "Further applications of information theory to multivariate analysis and statistical inference," Ph.D. dissertation on file at the library of The George Washington University, February 1957.

[15] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, Vol. 27 (1948), pp. 379–423 and 623–656.

[16] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc., New York, 1937.

[17] N. Wiener, *Cybernetics*, John Wiley & Sons, Inc., New York, 1948.

ON THE DISTRIBUTION OF 2 × 2 RANDOM NORMAL DETERMINANTS[1]

By W. L. Nicholson[2]

Princeton University

1. Summary. The c.d.f. of a 2 × 2 random determinant with mutually independent normally distributed entries is derived as an infinite series. Error functions that bound the tail of this series facilitate numerical calculation. Conditions are imposed on four variable quadratic forms for this distribution to apply. A normal approximation to the distribution is suggested.

2. Introduction. Let X_1, X_2, X_3 and X_4 be mutually independent random variables, each normally distributed, with means μ_1, μ_2, μ_3 and μ_4, and common variance σ^2. Let D be the random determinant,

$$D = \begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix} = X_1 X_4 - X_2 X_3.$$

Received June 18, 1956; revised December 9, 1957.

[1] Work sponsored by the Office of Naval Research.

[2] Now employed by the General Electric Company, Richland, Washington.

If all the μ_i vanish the p.d.f. of D/σ^2 is easily calculated to be the Laplace distribution [5],

$$\tfrac{1}{4} \exp \{-\tfrac{1}{2}|x|\}.$$

When the μ_i are not zero the distribution is, in general, skewed and not expressible in a simple closed form.

Craig [1] derived the p.d.f. of the product of two normal variables (not necessarily independent) as an infinite series of Bessel functions. Theoretically, his result plus the convolution formula for density functions determines the p.d.f. of D. However, the form of such an answer is not particularly adapted to numerical work. Most methods for handling the distribution problems connected with normal variable quadratic forms are not applicable here. The reasons for this are, first, that D is not a definite form, and, second, that it cannot be represented as a linear combination of central Chi-Square variables. The former obstacle can be overcome to a measured degree by several different procedures; e.g., Pitman's and Robbins' method of mixtures [6] and Gurland's Laguerrian expansions [3]. The latter causes more difficulty. There does not seem to be an adequate technique available to handle linear combinations of non-central Chi-Square variables.

Our approach is basically a brute force method consisting of straightforward inversion of the characteristic function of D. The independence and homoscedasticity assumptions cannot be relaxed without greatly complicating this inversion problem. In the process a single integration leads to the c.d.f. of D. Percentage points are immediately available without resorting to quadratures.

In the sequel $\sigma^2 = 1$. There is no loss of generality in this simplification, since σ^2 appears as a scale parameter in the distribution of D; i.e., we derive the distribution of the normalized variable D/σ^2.

The characteristic function of D is easily calculated to be

$$\phi_D(t) = E(e^{itD}) = (1 + t^2)^{-1} \exp\left\{\frac{-\Lambda t^2 + 2i\Delta t}{2(1 + t^2)}\right\} \tag{2.1}$$

where

$$\begin{aligned} \Lambda &= \mu_1^2 + \mu_2^2 + \mu_3^2 + \mu_4^2, \qquad 0 \leqq \Lambda < +\infty, \\ \Delta &= \begin{vmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{vmatrix} = \mu_1 \mu_4 - \mu_2 \mu_3, \qquad -\frac{\Lambda}{2} \leqq \Delta \leqq \frac{\Lambda}{2}. \end{aligned} \tag{2.2}$$

Thus, we see that the distribution of D depends on the means only in the form Λ and Δ. Expand $\log \phi_D(t)$ in powers of t to get the semi-invariants of D as

$$\begin{aligned} \alpha_{2k} &= (2k)!\left(\frac{\Lambda}{2} - \frac{1}{k}\right) \qquad k = 1, 2, \cdots; \\ \alpha_{2k+1} &= (2k + 1)!\,\Delta \qquad k = 0, 1, 2, \cdots. \end{aligned} \tag{2.3}$$

The mean and variance of D are

$$\mu_D = \alpha_1 = \Delta, \qquad \sigma_D^2 = \alpha_2 = \Lambda + 2. \tag{2.4}$$

The coefficients of skewness and excess are

$$
(2.5) \qquad
\begin{aligned}
\gamma_1 &= \frac{\alpha_3}{\alpha_2^{3/2}} = \frac{6\Delta}{(\Lambda + 2)^{3/2}}, \\
\gamma_2 &= \frac{\alpha_4}{\alpha_2^2} = \frac{12(\Lambda + 1)}{(\Lambda + 2)^2}.
\end{aligned}
$$

The distribution is skewed if and only if $\Delta \neq 0$. From (2.5) this skewness is never great. In fact $|\gamma_1| \leqq 2\sqrt{6}/3$ and $\gamma_1 = 0(\Lambda^{1/2})$ for large Λ. The excess γ_2 is monotone decreasing in Λ with a maximum value of three for $\Lambda = 0$. Also, $\gamma_2 = 0(\Lambda)$ for large Λ. Thus, if at least one μ_i is large the distribution is almost symmetric and of approximately the same peakedness as that of the normal. In Section 4 we show that D (appropriately normalized) is approximately normally distributed for large Λ.

3. Exact Distribution. The functional form of the characteristic function (2.1) indicates that the p.d.f., $f_{\Lambda,\Delta}$, and the c.d.f., $F_{\Lambda,\Delta}$, of D satisfy for all real x

$$
(3.1) \qquad f_{\Lambda,\Delta}(x) = f_{\Lambda,-\Delta}(-x), \qquad F_{\Lambda,\Delta}(x) = 1 - F_{\Lambda,-\Delta}(-x).
$$

Hence we need only consider the distribution of D for negative argument. In the remainder of the paper x always satisfies $x \leqq 0$. The c.d.f. of D is not expressible in a simple closed form (unless $\Delta = \Lambda/2$.). Introduction of an appropriate error term does make it possible to represent it as a damped polynomial in $|x|$ with coefficients that are elementary functions of Λ and Δ. Let R be any set of non-negative integers. The c.d.f. of D can be written as (see Sec. 5)

$$
(3.2) \qquad F_{\Lambda,\Delta}(x) = \sum_{r \in R} \sum_{t=0}^{r} h(r, t) g(r, t \mid \Delta, \Lambda/2, |x|) + L,
$$

where L satisfies

$$
(3.3) \qquad
\begin{aligned}
0 &\leqq L < \frac{1}{2} \sum_{r \in R} e^{-\Lambda/2} \frac{\left(\frac{\Lambda}{2}\right)^r}{r!} & \Delta \geqq 0, \\
0 &\leqq L < \frac{1}{2} \sum_{r \in R} e^{-\Lambda/2} \frac{\left(\frac{\Delta + |\Delta|}{2}\right)^r}{r!} & \Delta < 0.
\end{aligned}
$$

The auxiliary functions h and g are defined by

$$
(3.4) \qquad
\begin{aligned}
h(r, t) &= \sum_{j=0}^{r-t} \binom{2r - t + 1}{j} \left(\frac{1}{2}\right)^{2r - t + 1}, \\
g(r, t \mid a, b, c) &= e^{-(b+c)} \sum_{j=0}^{t} \frac{a^j (b - a)^{r-j} c^{t-j}}{j!(r - j)!(t - j)!}.
\end{aligned}
$$

Here $h(r, t)$ is just the probability of not more than $r - t$ heads on $2r - t + 1$ flips of an unbiased coin. Several tables of h are available; e.g., [7]. The function

g satisfies a number of recursion formulae. The most useful of these for computational purposes is

(3.5) $tg(r, t \mid a, b, c) = ag(r - 1, t - 1 \mid a, b, c) + cg(r, t - 1 \mid a, b, c).$

The boundary condition,

$$g(r, 0 \mid a, b, c) = e^{-(a+c)} \left\{ e^{-(b-a)} \frac{(b-a)^r}{r!} \right\}, \tag{3.6}$$

and (3.5) provide a rapid method of generating a matrix of g values for any triple $(a, b, c) = (\Delta, \Lambda/2, \mid x \mid)$. The right side of (3.6) is most easily calculated as the product of two tabular values. The bracket term as a Poisson density is tabled (see [4]).

The bound (3.3) on L is quite good for $\Delta \geqq 0$. Numerical checks show, for example, that for values of Λ of the order of ten a bound of 0.01 on L adds only one integer to R over and above that necessary to give an error $\leqq 0.01$. For $\Delta < 0$ the bound is admittedly rather poor and certainly could be improved.

To minimize the calculation necessary to evaluate $F_{\Lambda,\Delta}(x)$ the set R should contain as few elements as possible. To accomplish this and still maintain a specified bound on the error R should consist only of integers in an appropriate interval including $\Lambda/2$ (at least when $\Delta \geqq 0$). However, from the standpoint of iterated computation of the g function the optimum R set is $\{0, 1, 2, \cdots, M\}$ for suitable M. In this case the bound (3.3) on L is, except possibly for an exponential factor, the tail area of a Poisson distribution; its value can be read directly from tables [4].

At least three values of Δ lead to extreme simplification in the formula (3.2). These values are the maximum and the minimum Δ value for fixed Λ, and $\Delta = 0$. The simplified forms make possible several quickly computed bounds on $F_{\Lambda,\Delta}(x)$. The simplifications are

$$\begin{aligned} F_{\Lambda,-\Lambda/2}(x) &= e^{-(\Lambda/2)-|x|} \sum_{r \in R} \sum_{t=0}^{r} \frac{2^{r-t+1}-1}{2^{r-t+1}} \frac{|x|^t}{t!} \frac{(\Lambda/2)^r}{r!} + L; \\ F_{\Lambda,0}(x) &= e^{-(\Lambda/2)-|x|} \sum_{r \in R} \sum_{t=0}^{r} h(r, t) \frac{|x|^t}{t!} \frac{(\Lambda/2)^r}{r!} + L; \\ F_{\Lambda,\Lambda/2}(x) &= \tfrac{1}{2} e^{-(\Lambda/4)-|x|}. \end{aligned} \tag{3.7}$$

The bound (3.3) on L (with $\Delta \geqq 0$) is also applicable for the first two lines of (3.7). Since $F_{\Lambda,\Delta}(x)$ for fixed Λ and x is a monotone decreasing function of $\Delta(-\Lambda/2 \leqq \Delta \leqq \Lambda/2)$, the following inequalities are immediately available.

$$\begin{aligned} F_{\Lambda,-\Lambda/2}(x) \geqq F_{\Lambda,\Delta}(x) &\geqq F_{\Lambda,0}(x) \qquad \Delta \leqq 0, \\ F_{\Lambda,0}(x) \geqq F_{\Lambda,\Delta}(x) &\geqq F_{\Lambda,\Lambda/2}(x) \qquad \Delta \geqq 0. \end{aligned} \tag{3.8}$$

A simple but interesting application of (3.8) is the following bound on the probability that D is negative.

$$\frac{1}{2}e^{-\Delta/4} \leqq \Pr\{D \leqq 0\} \leqq \frac{1}{2} \qquad \Delta \geqq 0,$$
(3.9)
$$\frac{1}{2} \leqq \Pr\{D \leqq 0\} \leqq 1 - \frac{1}{2}e^{\Delta/4} \qquad \Delta \leqq 0.$$

These are the best bounds possible that are functions of Δ only.

The quadratic form D is by no means representative of the class of all such forms in four variables. In general, the distribution of a four variable form is more complicated than that of D. There are certain special cases, however, when the distribution function (3.2) does apply. The following theorem gives a necessary and sufficient condition on four normal variables and on the accompanying form for the distribution to be that of this paper.

THEOREM. *Let $X = (X_1, X_2, X_3, X_4)$ be a random vector distributed by a multivariate normal law with mean vector μ and covariance matrix Σ. Let M be a 4×4 symmetric matrix. The quadratic form $XMX'/2d$ is distributed according to the law* (3.2) *if and only if the eigenvalues of the matrix $M\Sigma$ are d, d, $-d$ and $-d$. If such is the case $\Lambda = \mu\,\Sigma^{-1}\,\mu'$ and $\Delta = \mu M\mu'/2d$.*

A proof is easily constructed by identifying the characteristic function of $XMX'/2d$ with (2.1).

4. Normal Approximation. Let $\tilde{D} = D/(2 + \Lambda)^{\frac{1}{2}}$. Then, if Λ increases without bound in such a manner that $\Delta/(2 + \Lambda)^{\frac{1}{2}} \to a$, we have from (2.1) that $\phi_{\tilde{D}}(t) \to \exp(iat - t^2/2)$. So, by the continuity theorem for characteristic functions [2], for large Λ $\tilde{D}$ is approximately normal with mean $\Delta/(2 + \Lambda)^{\frac{1}{2}}$ and unit variance. The question of how large Λ must be for the approximation to render reasonable accuracy is quite difficult to answer. The following remarks are offered to give some insight into this problem. Clearly, the rapidity of the convergence depends upon Δ and $|x|$ in some fashion. For $\Delta = 0$ the approximation is very good since this is the symmetric case. With Λ fixed the accuracy decreases as $|x|$ increases. Numerical checks indicate that for $|x|$ less than three and Λ about 20 the relative error in the c. d. f. is less than 5%. For the general case with Δ not too far different from zero the accuracy seems to be roughly a monotone decreasing function of $|x| + \Delta$ for fixed Λ. With $|x| + \Delta$ less than four and Λ about 20 the relative error is less than 7%. For large numerical values of Δ the approximation is extremely poor. For example, if $\Delta = 0(\Lambda)$ and if $|x| > 0$, then the relative error approaches 100% as Λ increases.

5. Derivation of Exact Distribution of D. Since ϕ_D is Lebesque integrable the Lévy inversion formula [2] gives the p.d.f. of D as

(5.1)
$$f_{\Lambda,\Delta}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\phi_D(t)\,dt = \frac{e^{-\Lambda/2}}{2}\sum_{j,k=0}^{\infty}\frac{\Delta^j(\Lambda/2)^k}{j!k!(j+k)!} \times \frac{\partial^{2j+k}}{\partial x^j \partial z^{j+k}} Q(x, z)\Bigg|_{z=1}^{x=x}.$$

Here,

(5.2)
$$Q(x, z) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{e^{-ixt}}{z + t^2}\,dt = z^{-1/2}e^{-|x|z^{1/2}},$$

for $x \leqq 0$ and $z > 0$.

Evaluate the jth partial derivative of Q with respect to x first and integrate $f_{\Lambda,\Delta}(x)$ over the interval $(-\infty, x)$. Then (5.1) becomes

$$(5.3) \qquad F_{\Lambda,\Delta}(x) = \frac{e^{-\Lambda/2}}{2} \sum_{r \varepsilon R} \left(\frac{1}{r!}\right)^2 \frac{\partial^r}{\partial z^r} \left[z^{-1} \left(\Delta z^{\frac{1}{2}} - \frac{\Lambda}{2} \right)^r e^{-|x| z^{1/2}} \right|_{z=1} + L,$$

where R is any set of non-negative integers and L consists of the remainder of the series; i.e., those terms such that $r \not\varepsilon R$ (here $r = j + k$).

Rewrite the factor $(\Delta z^{1/2} - \Lambda/2)^r$ as an rth partial derivative of the appropriate exponential function of W evaluated at zero. Use Leibnitz's Rule for differentiating a product to compute the rth partial with respect to z of the resulting function after choosing one product factor as $z^{-1/2}$. Employ the identity,

$$(5.4) \qquad \frac{\partial^P}{\partial z^P} z^{-\frac{1}{2}} e^{-az^{1/2}} \Big|_{z=1} = (-\tfrac{1}{4})^P P! \, e^{-a} \sum_{i=0}^{P} \binom{2P - i}{P} \frac{(2a)^i}{i!},$$

with $a = |x| - \Delta W$, and complete the differentiation with respect to W. Considerable algebraic simplification involving routine summing of finite combinatorial type series gives the form (3.2) after the appropriate identifications have been made with the h and g functions defined by (3.4). The bounds (3.3) and the simplifications (3.7) result from straightforward algebra. Details are omitted.

6. Acknowledgement. The author expresses his appreciation to Dr. L. Marcus whose interest in this problem gave impetus for the paper. Thanks are also due to Professor J. W. Tukey for several helpful suggestions.

REFERENCES

[1] C. C. Craig, "On the frequency function of xy," *Ann. Math. Stat.*, Vol. 7 (1936), pp 1–15.
[2] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, Inc., New York, 1953.
[3] J. Gurland, "Distribution of quadratic forms and ratios of quadratic forms," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 416–427.
[4] E. C. Molina, *Poisson's Exponential Binomial Limit*, D. Van Nostrand Company, Inc., New York, 1942.
[5] H. Nyguist, S. O. Rice and J. Riordan, "The distribution of random determinants," *Quart. Appl. Math.*, Vol. 42 (1954), pp. 97–104.
[6] H. Robbins and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 20 (1949), ppl 552–560.
[7] *Tables of the Cumulative Binomial Probabilities*, U. S. Army Ordnance Corps, Ordnance Corps Pamphlet ORDP 20-1.