THE DISTRIBUTION OF LENGTH AND COMPONENTS OF THE SUM OF n RANDOM UNIT VECTORS[1]

By J. Arthur Greenwood and David Durand

Manhattan Life Insurance Company and Massachusetts Institute of Technology

0. Summary. Statistical problems involving angular observations may arise in diverse scientific fields, either from direct measurement of angles—say the direction of winds or of glacial pebbles or of fracture planes—or they may arise from the measurement of times reduced modulo some period and converted into angles—say time of day when train wrecks occur. Specifically, we consider a set of n points ξ_ν situated on a unit circle and assumed to constitute a sample from a distribution having the p.d.f. $g(\xi)$, where $0 \leqq \xi < 2\pi$. Let the n random unit vectors thus defined have the components $\sin \xi_\nu$ and $\cos \xi_\nu$, and set

$$V = \sum \cos \xi_\nu, \qquad W = \sum \sin \xi_\nu, \qquad R = \sqrt{V^2 + W^2}. \tag{0.1}$$

Let $P(r, n)$ be the probability that $R \leqq r$, and let $Q = 1 - P(r, n)$.

This paper shows how the statistics V and R provide tests for the uniform distribution $g(\xi) = 1/2\pi$. The distribution of R on the hypothesis of uniformity was derived by Kluyver as a solution to Pearson's random walk problem, and is tabulated here for use in significance tests. The distribution of V is derived here, but has not been calculated.

To illustrate the type of tests that might employ the statistics R and V, consider a carnival wheel—first from the standpoint of the punter who suspects bias, second from the standpoint of the mechanic who has attempted to introduce bias. The punter, by studying the performance of the wheel, might wish to answer two questions: first, does the wheel differ credibly from an unbiased wheel? second, what is the direction and extent of the bias, if any? An answer to the first question is obtainable from the distribution of R, and an answer to the second has been provided by Mises. The mechanic, on the other hand—because he knows the direction of the bias, if there is a bias—might better use the statistic V as a test of his success, and he might appropriately modify the Mises approach in estimating the extent of the bias.

1. Pearson's random walk. In 1905 Pearson [13] posed the following problem: "A man starts from a point O and walks a distance a in a straight line; he then turns through any angle whatever and walks a distance a in a second straight line. He repeats this process n times.

Received April 6, 1954.

[1] This is a revised version of a paper presented at the Kingston meeting of the Institute of Mathematical Statistics in September 1953, under the title "The Integral Solution of Pearson's Random Walk Problem and Related Matters."

"I require the probability that after these n stretches he is at a distance between r and $r + \delta r$ from his starting point, O."

Thus, Pearson was assuming a sample of n random angles ξ_ν from the uniform distribution $g(\xi) = 1/2\pi$, and his required probability is the differential.

$$\frac{d}{dr} P(r, n)\ \delta r. \tag{1.1}$$

After Pearson's statement of the random walk problem, Rayleigh [16] observed an analogy to the theory of vibrations and deduced the asymptotic formula

$$Q(r, n) = \exp[-r^2/n]. \tag{1.2}$$

Then Kluyver [7] obtained a solution, showing that the probability of $R \leqq r$ is

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx. \tag{1.3}$$

This formal result can easily be found by bivariate characteristic functions, but the transformation from rectangular to polar coordinates is difficult to make rigorous.

Both Kluyver and Pearson were loath to attempt direct quadrature of (1.3) or its p.d.f. (see [14], p. 5), and Pearson devoted his efforts toward developing an asymptotic approximation (Section 6). We, however, felt that Pearson's series could best be checked by quadratures and that, with the advent of extended tables of the Bessel functions and improved calculating equipment, this was now feasible. Table 1 presents quadrature values of Kluyver's integral. Table 2 presents interpolated 5 percent and 1 percent points for r, and for several functions of r that may be useful in making significance tests. A discussion of the calculation procedure appears in Sections 4 and 5.

2. The problem of Mises. In another problem involving the summation of random unit vectors, Mises [12] considered vectors distributed according to the p.d.f.

$$g(\xi) = \frac{\exp\{k \cos(\xi - \alpha)\}}{2\pi I_0(k)}, \tag{2.1}$$

of which the uniform distribution is a degenerate case ($k = 0$). He then showed that a joint maximum likelihood estimate of the concentration parameter k and the modal angle α is obtainable from the vector sum by means of the relations

$$R \cos \alpha = \sum \cos \xi_\nu = V, \qquad R \sin \alpha = \sum \sin \xi_\nu = W. \tag{2.2}$$

Recently Gumbel, Greenwood, and Durand [6] christened (2.1) the "circular normal distribution," tabulated the integral thereof, and calculated a table for converting $R/n = \bar{a}$ into $\hat{k}$, the maximum likelihood estimate[2] of k.

[2] According to Mises ([12], eq. 16), $R/n = I_1(k)/I_0(k)$, which is tabulated by Gumbel et al. ([6], p. 140).

TABLE 1

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx$$ *(See Section 5 before attempting to interpolate in this table)*

r	$n = 6$	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0.5	.03885	.03229	.02911	.02588	.02353	.02148	.01979	.01834	.01709	.01599	.01503	.01418	.01342	.01274	.01212	.01156	.01105	.01058	.01015
1.0	.14286	.12500	.11111	.10000	.09091	.08333	.07692	.07143	.06667	.06250	.05882	.05556	.05263	.05000	.04762	.04545	.04348	.04167	.04000
1.5	.29324	.26141	.23345	.21182	.19348	.17821	.16511	.15383	.14398	.13533	.12765	.12080	.11464	.10908	.10404	.09944	.09523	.09137	.08780
2.0	.46782	.41782	.37907	.34622	.31884	.29535	.27512	.25746	.24193	.22816	.21587	.20483	.19487	.18583	.17759	.17005	.16312	.15673	.15083
2.5	.63165	.57455	.52795	.48750	.45291	.42271	.39626	.37287	.35207	.33344	.31666	.30149	.28769	.27509	.26354	.25292	.24312	.23405	.22563
3.0	.76722	.71404	.66424	.62132	.58281	.54868	.51809	.49063	.46583	.44336	.42290	.40421	.38708	.37131	.35676	.34330	.33080	.31917	.30833
3.5	.87020	.82278	.77819	.73679	.69872	.66381	.63182	.60249	.57555	.55077	.52791	.50679	.48723	.46906	.45216	.43640	.42168	.40789	.39496
4.0	.93755	.90039	.86465	.82871	.79452	.76189	.73117	.70230	.67524	.64989	.62616	.60393	.58309	.56353	.54516	.52788	.51160	.49625]	.48176
4.5	.97321	.95066	.92400	.89635	.86805	.84009	.81276	.78635	.76099	.73675	.71364	.69165	.67074	.65088	.63201	.61409	.59706	.58086	.56545
5.0	.99121	.97864	.96149	.94198	.92067	.89848	.87590	.85333	.83105	.80923	.78800	.76743	.74756	.72842	.71001	.69231	.67533	.65902	.64337
5.5	.99858	.99205	.98280	.97022	.95562	.93932	.92196	.90389	.88547	.86694	.84847	.83020	.81224	.79466	.77751	.76081	.74459	.72886	.71362
6.0	1.00000	.99788	.99329	.98626	.97705	.96604	.95359	.94005	.92570	.91080	.89555	.88012	.86463	.84920	.83390	.81879	.80393	.78934	.77506
6.5		.99976	.99783	.99442	.98916	.98233	.97403	.96451	.95397	.94263	.93067	.91824	.90549	.89251	.87942	.86629	.85318	.84014	.82723
7.0		1.00000	.99953	.99804	.99542	.99151	.98640	.98014	.97285	.96468	.95576	.94621	.93616	.92570	.91494	.90395	.89281	.88157	.87029
7.5			.99996	.99945	.99830	.99629	.99338	.98954	.98481	.97924	.97291	.96591	.95833	.95025	.94174	.93289	.92375	.91439	.90486
8.0			1.00000	.99990	.99946	.99855	.99704	.99485	.99197	.98838	.98412	.97923	.97376	.96776	.96130	.95442	.94719	.93965	.93185
8.5				1.00000	.99987	.99951	.99880	.99765	.99601	.99383	.99111	.98786	.98408	.97981	.97508	.96994	.96441	.95853	.95235
9.0					.99998	.99986	.99957	.99902	.99815	.99691	.99527	.99320	.99071	.98780	.98448	.98076	.97668	.97225	.96751
9.5					1.00000	.99997	.99986	.99963	.99920	.99855	.99761	.99637	.99480	.99290	.99065	.98807	.98516	.98193	.97840
10.0						1.00000	.99997	.99988	.99969	.99936	.99886	.99816	.99722	.99603	.99457	.99284	.99084	.98856	.98601
10.5							.99999	.99996	.99989	.99974	.99949	.99911	.99858	.99787	.99696	.99585	.99452	.99296	.99118
11.0							1.00000	.99999	.99997	.99990	.99979	.99960	.99931	.99890	.99836	.99768	.99682	.99580	.99460
11.5								1.00000	.99999	.99997	.99992	.99983	.99968	.99946	.99916	.99875	.99822	.99757	.99678
12.0									1.00000	.99999	.99997	.99993	.99986	.99975	.99958	.99935	.99904	.99864	.99814
13.0										1.00000	1.00000	.99999	.99998	.99995	.99991	.99985	.99975	.99962	.99944
14.0												1.00000	1.00000	.99999	.99999	.99997	.99995	.99991	.99985
15.0														1.00000	1.00000	1.00000	.99999	.99998	.99997
16.0																	1.00000	1.00000	.99999
17.0																			1.00000

TABLE 2

5 and 1 percent points for r and derived functions

n	5 percent — $P(r, n) = .95$				1 percent — $P(r, n) = .99$			
	r	r/n	r^2	$z = r^2/n$	r	r/n	r^2	$z = r^2/n$
6	4.141	.6901	17.14	2.8573	4.951	.8251	24.51	4.0851
7	4.491	.6416	20.17	2.8819	5.394	.7705	29.09	4.1561
8	4.818	.6022	23.21	2.9014	5.797	.7246	33.61	4.2007
9	5.118	.5686	26.19	2.9102	6.185	.6872	38.25	4.2504
10	5.402	.5402	29.19	2.9187	6.550	.6550	42.90	4.2899
11	5.674	.5158	32.19	2.9262	6.893	.6266	47.51	4.3195
12	5.932	.4943	35.18	2.9320	7.220	.6017	52.13	4.3441
13	6.179	.4753	38.18	2.9370	7.533	.5795	56.75	4.3651
14	6.417	.4584	41.18	2.9413	7.833	.5595	61.36	4.3829
15	6.646	.4431	44.18	2.9450	8.122	.5415	65.97	4.3983
16	6.868	.4293	47.17	2.9482	8.402	.5251	70.59	4.4116
17	7.083	.4166	50.17	2.9511	8.672	.5101	75.20	4.4234
18	7.291	.4051	53.16	2.9536	8.934	.4963	79.81	4.4338
19	7.494	.3944	56.16	2.9558	9.188	.4836	84.42	4.4430
20	7.691	.3846	59.16	2.9579	9.435	.4718	89.03	4.4514
21	7.884	.3754	62.15	2.9597	9.677	.4608	93.64	4.4589
22	8.072	.3669	65.15	2.9613	9.912	.4505	98.24	4.4657
23	8.255	.3589	68.15	2.9629	10.142	.4409	102.85	4.4719
24	8.435	.3514	71.14	2.9642	10.366	.4319	107.46	4.4775
∞				2.9957				4.6052

Accordingly, if the punter wishes a joint maximum likelihood estimate of the direction and extent of bias in the carnival wheel, he may easily obtain this if he is willing to assume that the wheel is biased according to the Mises distribution law. Moreover, if he wishes to test whether the observed performance of the wheel is consistent with the uniform hypothesis $k = 0$, he may do so by a simple extension of Mises' argument. In fact, the likelihood ratio for testing $k = 0$ against the alternative $k = \hat{k}$ and $\alpha = \hat{\alpha}$ is

$$\frac{(2\pi)^{-n}}{[2\pi\, I_0(\hat{k})]^{-n} \exp[\hat{k}\Sigma \cos(\xi_\nu - \hat{\alpha})]} = \frac{[I_0(\hat{k})]^n}{\exp[\hat{k}R]}.$$

From the fact that the second derivative of $\ln I_0(k)$ is the variance of the linear distribution of $\cos \xi$, and therefore essentially positive, it can be shown that the above ratio, for fixed n, is a decreasing function of R alone, where R has the distribution (1.3) on $k = 0$. The test of $k = 0$ then consists in comparing R or R^2 with the tabled values of r or r^2 (Table 2). An example will be given in Section 8.

The above hypothesis $k = \hat{k}$ and $\alpha = \hat{\alpha}$—the punter's alternative—is but one of several against which uniformity might be tested. In fact, twelve possible hypotheses can easily be formed by coupling one of the three α-hypotheses with one of the four k-hypotheses:

$$0 \leqq \alpha < 2\pi, \qquad \alpha_1 \leqq \alpha \leqq \alpha_2, \qquad \alpha_1 = \alpha;$$

$$k = k^* > 0, \ (k^* \text{ unknown}), \qquad k = k_1, \qquad k \geqq k_1, \qquad k_2 \geqq k \geqq k_1.$$

In particular, $\alpha = \alpha_1$ and $k = k^* > 0$ is the appropriate alternative for the mechanic who has attempted to bias the carnival wheel; the mechanic, unlike the punter, knows the position of the mode if there is a mode. And if he rejects uniformity, the mechanic needs to estimate only the parameter k.

The mechanic's problem is of special interest because of its relation to the work of Mises, who formulated his "Kriterium der Ganzzahligkeit" ([12], p. 495 ff.) in hopes of showing that atomic weights are integers subject to error. The observed atomic weights reduced modulo 1 and transformed into angles are subject to a test of the uniformity hypothesis against the alternative $\alpha = 0$ and $k = k^* > 0$. In making this test, Mises proceeded to estimate k from the statistic R; but, on the alternative stated, V and not R furnishes a maximum likelihood estimate of k.

The likelihood $\ln L = -n \ln I_0(k) + kV - n \ln 2\pi$. Setting

$$\frac{d}{dk} \ln L = -n \frac{I_1(k)}{I_0(k)} + V = 0,$$

we find that $V/n = I_1(k)/I_0(k)$, so that the recent table ([6], p. 140) applies. The distribution of V is derived from the characteristic function of $\cos \xi$,

$$\phi(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix\cos\xi} \, d\xi = J_0(x).$$

Therefore, the sum of n independent cosines has the characteristic function $[J_0(x)]^n$ and the distribution function

$$C + \frac{1}{2\pi} \int_{-\infty}^{\infty} [J_0(x)]^n \frac{1 - e^{-ixv}}{ix} \, dx = C + \frac{1}{\pi} \int_0^{\infty} [J_0(x)]^n \frac{\sin vx}{x} \, dx.$$

When $v = 0$, the integral vanishes; therefore $C = \frac{1}{2}$ and the probability that $V \leqq v$ takes the form

$$\frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} [J_0(x)]^n \frac{\sin vx}{x} \, dx \tag{2.3}$$

on the hypothesis $k = 0$. In the limit, V is normally distributed with mean zero and variance $n/2$.

3. Nonzero k and the power function. Under some circumstances the mechanic may feel sanguine that he has effectively biased the wheel and yet have doubts whether the direction of bias is exactly zero, as designed. Then, an entirely different sort of test would be needed—in fact one analogous to the linear t-test for the difference between an observed mean and its hypothetical value. But the construction of this test—and several others that might be useful—requires investigation of the distribution of R or of $\hat{\alpha} - \alpha$, either jointly or severally, on the hypothesis $k = k_1 > 0$.

On this hypothesis the joint distribution of V and W (or of R and $\hat{\alpha}$), the distribution of R, and the distribution of V can be obtained by an artifice due to Fisher[3] [4], which derives these distributions painlessly from those on the hypothesis $k = 0$. Fisher seems to effect his derivation by means of the following

LEMMA. *Let $F(x, y)$ and $G(x, y)$ be two distribution functions such that*

$$dF(x,y) = Ae^{ax+by}\, dG(x, y), \qquad A^{-1} = \iint e^{ax+by}\, dG(x, y). \tag{3.1}$$

Denote the distributions of the n-fold convolutions of $F(x, y)$ and $G(x, y)$ respectively by $F_n(v, w)$ and $G_n(v, w)$. Then

$$dF_n(v, w) = A^n e^{av+bw}\, dG_n(v, w). \tag{3.2}$$

To prove this let

$$\phi(t, u) = \iint e^{i(tx+uy)}\, dF(x, y),$$

$$\psi(t, u) = \iint e^{i(tx+uy)}\, dG(x, y),$$

where the integrals are Stieltjes integrals extended over the x, y-plane. Then

$$\phi^n(t, u) = \iint e^{i(tv+uw)}\, dF_n(v, w),$$

$$\psi^n(t, u) = \iint e^{i(tv+uw)}\, dG_n(v, w).$$

By hypothesis (3.1)

$$\phi(t, u) = \iint e^{i(tx+uy)}\, Ae^{ax+by}\, dG(x, y)$$

$$= A\psi[(t - ia), (u - ib)].$$

The characteristic functions of the left and right sides of (3.2) are, respectively,

$$\phi^n(t, u) = A^n\psi^n[(t - ia), (u - ib)];$$

$$\iint e^{i(tv+uw)} A^n e^{av+bw}\, dG_n(v, w) = A^n\psi^n[(t - ia), (u - ib)].$$

The equality (3.2) follows from the equality of the characteristic functions.

---

[3] Fisher's work is concerned with the spherical analogue of the Mises distribution (2.1). Although the trigonometry is more involved in three dimensions than in two, the algebra is much simpler. See Rayleigh [17], p. 338 ff.

On the hypothesis $k = 0$, the differential of (2.3) is

$$dP = \frac{dv}{\pi} \int_0^\infty [J_0(x)]^n \cos vx \, dx.$$

Therefore, by (3.2) the differential of probability on the hypothesis $k = k_1$ and $a = 0$ is

$$dP = \frac{e^{k_1 v} \, dv}{2\pi [I_0(k_1)]^n} \int_{-\infty}^\infty [J_0(x)]^n e^{-ivx} \, dx,$$

so that

$$P(V \leqq v) = \frac{1}{2\pi [I_0(k_1)]^n} \int_{-\infty}^\infty [J_0(x)]^n \frac{e^{(k_1 - ix)v}}{k_1 - ix} \, dx.$$

This distribution is the power function of V against the Mises alternative $\alpha = 0$ and $k = k_1 > 0$.

To obtain the distributions of W and R, one first writes down the differential of the joint probability of V and W on the hypothesis $k = 0$. The probability density function corresponding to (1.3) is circularly symmetric and has the form

$$\text{(3.4)} \qquad \frac{1}{2\pi r} \frac{d}{dr} r \int_0^\infty [J_0(x)]^n J_1(rx) \, dx = \frac{1}{2\pi} \int_0^\infty [J_0(x)]^n J_0(rx) \, x \, dx,$$

so that

$$dP(V \leqq v, \quad W \leqq w) = \frac{dv \, dw}{2\pi} \int_0^\infty [J_0(x)]^n J_0(x\sqrt{v^2 + w^2}) x \, dx.$$

It will be convenient to use Pearson's notation of $\phi_n(v^2 + w^2)$ for the integral above. Then, on the hypothesis $\alpha = 0$ and $k = k_1 > 0$

$$\text{(3.5)} \quad dP(V \leqq v, \quad W \leqq w) = (2\pi)^{-1} [I_0(k_1)]^{-n} e^{k_1 v} \phi_n(v^2 + w^2) \, dv \, dw.$$

This differential does not readily yield the marginal distribution of W. Probably a convenient approximation is to substitute Pearson's Laguerre function expansion for $\phi_n(v^2 + w^2)$ and integrate v out (see Section 6).

The distribution of R can be obtained by transforming (3.5) into polar coordinates,

$$\text{(3.6)} \qquad dP(V \leqq r \cos \beta, \quad W \leqq r \sin \beta) = \frac{e^{k_1 r \cos \beta} \phi_n(r^2)}{2\pi [I_0(k_1)]^n} \, r \, dr \, d\beta;$$

whence $dP(R \leqq r) = [I_0(k_1)]^{-n} I_0(k_1 r) \, \phi_n(r^2) \, r \, dr$. Integration by parts then gives

$$P(R \leqq r) = [I_0(k_1)]^{-n} \left[I_0(k_1 r) P(r, n) - \int_0^r k_1 I_1(k_1 s) P(s, n) \, ds \right],$$

where $P(r, n)$ is the function of (1.3). This last distribution is, in fact, the power function of R against the Mises alternative $\alpha = 0$ and $k = k_1 > 0$. It is also the power function against the Mises alternative $\alpha = \alpha^*$ (unknown) and $k = k_1 > 0$, because R is unaffected by increasing all the ξ_ν by an arbitrary constant. The distribution of $\hat{\alpha}$ can be evaluated after expanding (3.6) in Laguerre functions.

4. Calculation of Kluyver's integral. The integral has been evaluated for $n = 6(1)24$ and $r = 0(.1)6(.5)12(1)n$, and for such other values at intervals of .05 or .1 as were necessary to embrace the 5 percent and 1 percent points of R. Table 1, as published here, is an abridgment. Evaluation of the integral for $n = 3$, 4, or 5 would have required special treatment, such as: (A) tabulating J_1 to values much greater than the present upper tabled limit of 100; (B) replacing J_0 and J_1 by their asymptotic series and then expanding the integral into a series of sine integrals for $n = 3$ or 5, or a series of Fresnel integrals for $n = 4$; (C) an extension by quadratures as performed by Pearson and Blakeman ([14], pp. 16–20), but carried on digitally instead of graphically. For $n = 2$ the distribution of R is simply $(2/\pi)$ arc sin $r/2$.

The integration formula used was the integral of Bessel's interpolation formula carried to 8 differences. For an interval of integration $h = .04$ was used for $r \leqq 10$, and $h = .02$ for $r \geqq 10$; agreement at $r = 10$ was good.

There were three chief sources of error in the integration:

(A) The integration was truncated at $rx = 100$, the upper limit of the J_1 table. This error is larger than that explained in (B) following for $n = 6$, 7, and 8; for $n = 4$ or 5 it would have been prohibitive. For $n = 7$ arguments for J_1 above 100 were supplied.

(B) The integration was often truncated for $|J_0(x)|^n$ less than 10^{-5}. This was by no means systematic, however, since for many of the integrations additional arguments were conveniently included.

(C) The remainder term, involving ninth differences and greater, was neglected.

The values $P(1, n) = 1/(n + 1)$ and $P(n, n) = 1$, known by elementary analysis, furnish a partial check on the accuracy of the integration. In most of the table discrepancies were less than 10^{-6}, and nowhere were they as great as 10^{-5}. We believe that Table 1 is accurate to 5 decimals as given, with the exception of $n = 6$, where an error of 1 in the fifth place can easily have occurred.

5. Interpolation of percentage points. Because of discontinuities in its derivatives with respect to r, Kluyver's integral (1.3) presents problems of interpolation, both direct and inverse. These discontinuities have been discussed by Rayleigh [17] who investigated the behavior of the leading term in the asymptotic expansion of the Bessel function product. He found that the derivatives of order $\frac{1}{2}(n - 2)$ for n even and $\frac{1}{2}(n - 3)$ for n odd are discontinuous[4] at the points $n - 2k$, for $k = 0, 1, 2, 3$, etc. Although the integral

$$\int_0^X [J_0(x)]^n J_1(rx)\, dx,$$

[4] For a discussion of discontinuities in the sum of n rectangular variables, see Cramér [3], p. 245, and Fisher [4], p. 298.

where the upper limit of quadrature X is a large number, has continuous derivatives of all orders, these derivatives become large for the same points $n - 2k$.

For the problem of inversely interpolating the 5 percent and 1 percent points (Table 2), the singularities of (1.3) were not particularly troublesome because none of the desired points happen to fall uncomfortably close to a singularity. Thus, for the interpolation of $Q(r, 6) = .05$, the function was evaluated for the five arguments 4.10, 4.15, 4.20, 4.25, and 4.30, which do not include the singularity $r = 4.00$. Although singularities are less dangerous for large n—since the order of the first discontinuous derivative increases—they were avoided whenever convenient, and in no case did an interpolation use points on both sides of a singularity. Table 2 presents interpolated 5 percent and 1 percent points for r, and for several functions of r that may be useful in making significance tests. The first three of these functions—namely, r, r/n, and r^2—should all be correct to four significant figures, as tabled. The function z, given to five significant figures for comparisons in Section 7, may possibly contain an error of about 2 in the last figure.

6. Pearson's asymptotic expansion. Pearson derived an asymptotic expansion for the probability density function (3.4) corresponding to (1.3). This expansion is, in effect, a series of Laguerre functions

$$f(r^2) = e^{-kr^2} \sum_{i=0}^{\infty} c_i L_i(kr^2), \qquad L_i(x) = e^x \frac{d^i}{dx^i}(x^i e^{-x}) = \sum_{t=0}^{i} \frac{i!\, i!(-x)^t}{t!\, t!(i-t)!}.$$

To derive the c_i, Pearson essentially used $[J_0(x)]^n$ as a moment generating function, thus obtaining the expansion

$$\frac{1}{2\pi}\int_0^\infty [J_0(x)]^n J_0(rx)x\,dx = \frac{1}{\pi n} e^{-z} \sum_{i=0}^{\infty} c_i L_i(z) \tag{6.1}$$

or, with $z = r^2/n$,

$$2\pi r\,dr\,\frac{1}{2\pi}\int_0^\infty [J_0(x)]^n J_0(rx)x\,dx = dz\,e^{-z} \sum_{i=0}^{\infty} c_i L_i(z). \tag{6.2}$$

Each c_i ([14], p. 9) is of order $n^{-\beta}$, where β is the integral part of $\frac{1}{2}(1 + i)$. Pearson derived the c_i through c_6, which contains terms in n^{-5}.

From the well-known relation

$$\int_z^\infty e^{-z} L_i(z)\,dz = e^{-z}[L_i(z) - iL_{i-1}(z)],$$

the series on the right of (6.2) may be integrated term by term as

$$Q(r, n) = e^{-z}\left\{1 + \sum_{i=0}^{\infty} c_i[L_i(z) - iL_{i-1}(z)]\right\}. \tag{6.3}$$

Rearranging (6.3) in descending powers of n and omitting contributions of order n^{-4} and n^{-5}, which appear in c_5 and c_6, we get

$$Q(r, n) = e^{-z}\left[1 + \frac{1}{2n}\left(z - \frac{z^2}{2!}\right) + \frac{1}{12n^2}\left(-z + \frac{11z^2}{2!} - \frac{19z^3}{3!} + \frac{9z^4}{4!}\right) + \frac{1}{24n^3}\left(-2z - \frac{4z^2}{2!} + \frac{69z^3}{3!} - \frac{163z^4}{4!} + \frac{145z^5}{5!} - \frac{45z^6}{6!}\right)\right]. \tag{6.4}$$

This series is more convenient for computation than (6.3). Table 3 compares for $n = 7$ and 14 the value of $P(r, n) = 1 - Q(r, n)$ obtained by quadratures with Rayleigh's approximation $P(r, n) = 1 - e^{-z}$, Pearson's approximation (6.3) through c_6, and the approximation (6.4) through n^{-3}. It appears that either (6.3) or (6.4) gives roughly three-decimal accuracy for $n = 7$, and that, if anything, (6.4) is a little better than (6.3). Although three reliable decimal places provide a fair degree of accuracy over most of the curve, they are clearly insufficient for estimating percentage points in the extreme tail.

7. Series expansion of the percentage points. Having obtained $Q(r, n)$ as an expansion in powers of z and n^{-1}, we now attempt to obtain an expansion for z in terms of $Q(r, n)$ and n^{-1}. Since the Rayleigh approximation above obviously leads to $z = -\ln Q(r, n)$, it is convenient to set

$$y = -\ln Q(r, n) \tag{7.1}$$

and expand z in powers of y. To do this, set

$$z = y + \sum a_i n^{-i}. \tag{7.2}$$

Then $e^z Q(r, n) = \exp \sum a_i n^{-i}$, whence we obtain

$$a_1 = \tfrac{1}{4}(2y - y^2), \qquad a_2 = \tfrac{1}{72}(12y - 3y^2 - y^3),$$

$$a_3 = \tfrac{1}{288}(-12y + 42y^2 - 8y^3 - y^4).$$

For $n = 7$, 14, and 21, Table 4 gives the successive approximations to the 1 percent and 5 percent points and compares them with the values obtained by inverse interpolation. Even when four terms (through a_3) are used, good accuracy is not achieved for $n = 7$. Three terms give fair three-place accuracy for $n = 14$ and four-place accuracy for $n = 21$, so that three terms suffice to extend Table 2 beyond $n = 24$. However, three terms may be inadequate to obtain points in the extreme tail, such as the 0.1 percent point, if these are desired. For example, $Q(12, 21) = .000{,}648{,}77$ by quadratures. The three-term approximation for z is 6.8590 against the correct value $144/21 = 6.8571$. The fourth term is $-.0015$, and the four-term approximation is therefore 6.8575.

8. An example of the *R*-test. Forrest [5] gives 651 observed times of breakdown due to lightning in the British electrical grid system for the eight years 1940–47. The observations fall into two groups: 588 for the summer months (April–September) and 63 for the winter months (October–March). The summer

TABLE 3

Various calculations of Kluyver's integral for $n = 7$ and 14

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx$$

r	$P(r, n)$ by quadrature	$1 - e^{-r^2/n}$ (Rayleigh)	(6.3) through c_6 (Pearson)	(6.4) through n^{-3}
		$n = 7$		
0.5	.03229	.03508	.03284	.03272
1.0	.12500	.13312	.12430	.12500
1.5	.26141	.27489	.26106	.26078
2.0	.41782	.43528	.41819	.41819
2.5	.57455	.59052	.57469	.57497
3.0	.71404	.72355	.71305	.71339
3.5	.82278	.82623	.82279	.82289
4.0	.90039	.89830	.90095	.90082
4.5	.95066	.94458	.95065	.95046
5.0	.97864	.97188	.97857	.97846
5.5	.99205	.98672	.99219	.99219
6.0	.99788	.99416	.99779	.99785
6.5	.99976	.99761	.99962	.99968
7.0	1.00000	.99909	1.00003	1.00006
		$n = 14$		
0.5	.01709	.01770	.01709	.01709
1.0	.06667	.06894	.06668	.06667
1.5	.14398	.14846	.14401	.14399
2.0	.24193	.24853	.24196	.24194
2.5	.35207	.36009	.35208	.35207
3.0	.46583	.47421	.46583	.46583
3.5	.57555	.58314	.57569	.57555
4.0	.67524	.68109	.67521	.67524
4.5	.76099	.76459	.76098	.76099
5.0	.83105	.83232	.83104	.83105
5.5	.88547	.88476	.88548	.88548
6.0	.92570	.92357	.92571	.92570
6.5	.95397	.95109	.95399	.95397
7.0	.97285	.96980	.97286	.97285
7.5	.98481	.98201	.98481	.98481
8.0	.99197	.98966	.99196	.99196
8.5	.99601	.99427	.99600	.99601
9.0	.99815	.99693	.99814	.99815
9.5	.99920	.99841	.99920	.99921
10.0	.99969	.99921	.99969	.99969
11.0	.99997	.99982	.99997	.99997
12.0	1.00000	.99997	1.00000	1.00000

TABLE 4

Various calculations of the 5 and 1 percent points for $n = 7, 14,$ and 21

	5 percent points $P(r, n) = .95$	1 percent points $P(r, n) = .99$
y	+2.9957 3227	+4.6051 7019
a_1	−0.7457 3683	−2.9993 1302
a_2	−0.2480 4699	−1.4725 7372
a_3	+0.1574 8945	−1.3736 8651
$n = 7$		
$y + a_1/7$	2.8892	4.1767
$y + a_1/7 + a_2/49$	2.8841	4.1466
$y + a_1/7 + a_2/49 + a_3/343$	2.8846	4.1426
z, by interpolation	2.8819	4.1561
$n = 14$		
$y + a_1/14$	2.9425	4.3909
$y + a_1/14 + a_2/196$	2.9412	4.3834
$y + a_1/14 + a_2/196 + a_3/2744$	2.9413	4.3829
z, by interpolation	2.9413	4.3829
$n = 21$		
$y + a_1/21$	2.9602	4.4623
$y + a_1/21 + a_2/441$	2.9597	4.4590
$y + a_1/21 + a_2/441 + a_2/9261$	2.9597	4.4589
z, by interpolation	2.9597	4.4589

breakdown times have a rough resemblance to a Mises distribution with a conspicuous mode at about 1700 hours, though the fit is not first-rate, owing to an apparent secondary mode in the early morning. The winter breakdown times are given in the table below, and it is by no means certain that this distribution differs significantly from the uniform.

We propose to test the hypothesis $k = 0$ against the alternative $k = k^* > 0$ and $0 \leqq \alpha < 2\pi$. To this end, we reduce the midpoints of the hourly time intervals to angles (0–1: 7°.5, etc.) and include the corresponding sines and cosines in Forrest's table.

G.M.T.	Freq.	Sines	Cosines	G.M.T.	Freq.	Sines	Cosines
0–1	1	.1305	.9914	12–13	4	−.1305	−.9914
1–2	1	.3827	.9239	13–14	3	−.3827	−.9239
2–3	1	.6088	.7934	14–15	1	−.6088	−.7934
3–4	3	.7934	.6088	15–16	4	−.7934	−.6088
4–5	1	.9239	.3827	16–17	7	−.9239	−.3827
5–6	2	.9914	.1305	17–18	4	−.9914	−.1305
6–7	1	.9914	−.1305	18–19	2	−.9914	.1305
7–8	3	.9239	−.3827	19–20	4	−.9239	.3827
8–9	0	.7934	−.6088	20–21	4	−.7934	.6088
9–10	7	.6088	−.7934	21–22	1	−.6088	.7934
10–11	0	.3827	−.9239	22–23	6	−.3827	.9239
11–12	1	.1305	−.9914	23–24	2	−.1305	.9914

From this layout, it is a simple matter to calculate the statistics

$$W = -13.3393, \qquad V = -3.2652, \qquad R^2 = 188.59.$$

Since the corresponding percentage points for r^2 have not been tabulated for $n > 24$, it is necessary to calculate an approximate value for z after the method of Section 7. From the two-term formula (7.2), specifically $y + a_1/63$, the 5 percent approximation 2.9835 is obtained, which compares with $R^2/n = 2.993$.

9. Unsolved problems. We have indicated in Section 3 how the distribution of $\hat{\alpha} - \alpha$ may be obtained for known k. To test the significance of an observed discrepancy with no a priori information about k is an obvious extension of the problem in linear statistics solved by the t-test. For large n one can replace k by $\hat{k}$ and treat it as known; for small n the appropriate test is unknown.

On the assumption of discrete sample points measured without error, (1.3) is a suitable mathematical model. Forrest's data, to which we applied (1.3), are sparse data grouped at the midpoints of class intervals. We do not know the effect of such grouping on V, W, or R, or on any estimators derived from them.

Although we have shown that R provides a likelihood ratio test for $k = 0$ against the composite Mises alternative $0 \leqq \alpha < 2\pi$, and $k = k^* > 0$, and that V provides a likelihood ratio test against the Mises alternative $\alpha = \alpha_1$, and $k = k^* > 0$, we feel that these results are by no means conclusive because of the existence of other unimodal, symmetrical p.d.f.s. For example, Levy [9], [10], Marcinkiewicz [11], Perrin ([15], p. 20), Wintner [18], [19], and Zernike ([20], pp. 477–478) have discussed wrapped-up normal and Cauchy distributions —that is, the angle ξ is assumed to have the normal or Cauchy distribution but the angles ξ, $\xi \pm 2\pi$, $\xi \pm 4\pi$, $\cdots$, are indistinguishable. Arnold ([1], p. 6) has shown that V and W jointly provide consistent estimates of the parameters of this distribution. Brooks and Carruthers ([2], p. 199) have translated a bivariate normal distribution into polar coordinates with the origin removed from the center of the distribution and have integrated out with respect to the radius.

Arnold and Krumbein have indicated problems wherein angles are identifiable only modulo π. A workable approach is to use the distribution $Ce^{k\cos 2\xi}$ and the statistics

$$V_2 = \sum \cos 2\xi_\nu, \qquad W_2 = \sum \sin 2\xi_\nu, \qquad R_2 = \sqrt{V_2^2 + W_2^2}.$$

This solution is not unique, and the merits of alternative solutions should be studied.

Spherical distribution problems have received some attention, both as generalizations of circular problems and on their own merits. A few examples might be mentioned. Arnold [1] has extended the Mises distribution and the wrapped-up normal to the surface of the sphere. Rayleigh [17] has shown that the spherical analogue of R has for distribution a set of elementary functions abutting with discontinuous derivatives as in Section 5, and Fisher has given an explicit formula for this distribution. The spherical analogue of the V statistic is distributed as the sum of n rectangular variables (Fisher [4], p. 296).

10. Acknowledgments. We are indebted to E. J. Gumbel, who suggested the problem in 1951 during the joint efforts leading to Gumbel et al. [6], and to Robert Schiffman for indispensable aid in assembling a punched card table of the Bessel function J_1. The calculations were performed with punched cards and equipment contributed by The Manhattan Life Insurance Company and The National Bureau of Economic Research.

REFERENCES

[1] K. J. ARNOLD, "Spherical probability distributions," Dissertation, Massachusetts Institute of Technology, 1941.

[2] C. E. P. BROOKS AND N. CARRUTHERS, *Handbook of Statistical Methods in Meteorology*, London, H. M. Stationery Office, 1953.

[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] SIR RONALD FISHER, "Dispersion on a sphere," *Proc. Roy. Soc.* London. Ser. A, Vol. 217 (1953), pp. 295–305.

[5] J. S. FORREST, "Variations in thunderstorm severity in Great Britain," *Quart. J. Roy. Meteor. Soc.*, Vol. 76 (1950), pp. 277–286.

[6] E. J. GUMBEL, J. A. GREENWOOD, AND D. DURAND, "The circular normal distribution: theory and tables," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 131–152.

[7] J. C. KLUYVER, "A local probability problem," *Nederl. Akad. Wetensch., Proc.* Vol. 8 (1906), pp. 341–350.

[8] W. C. KRUMBEIN, "Preferred orientation of pebbles in sedimentary deposits," *J. Geol.*, Vol. 47 (1939), pp. 673–706.

[9] P. LÉVY, "Sur l'arithmétique des lois de probabilités enroulées," *C. R. Soc. Math. France*, 1938, pp. 32–34.

[10] P. LÉVY, "L'addition des variables aléatoires définies sur une circonférence," *Bull. Soc. Math. France*, Vol. 67 (1939), pp. 1–41.

[11] J. M. MARCINKIEWICZ, "Sur les variables aléatoires enroulées," *C. R. Soc. Math. France*, 1938, pp. 34–36.

[12] R. v. MISES, "Über die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen," *Physikal. Zeitschr.*, Vol. 19 (1918), pp. 490–500.

[13] K. PEARSON, "The problem of the random walk," *Nature*, Vol. 72 (1905), pp. 294, 302.

[14] K. PEARSON, "A mathematical theory of random migration," *Drapers' Company Research Memoirs*, Biometric Series, No. 3, 1906.

[15] F. PERRIN, "Étude mathématique du mouvement brownien de rotation," *Ann. Ecole Norm. Sup.*, Vol. 45 (1928), pp. 1–51.

[16] J. W. STRUTT, LORD RAYLEIGH, "The problem of the random walk," *Nature*, Vol. 72 (1905), p. 318.

[17] J. W. STRUTT, LORD RAYLEIGH, "On the problem of random vibrations and of random flights in one, two, and three dimensions," *Philos. Mag.* (6), Vol. 37 (1919), pp. 321–347.

[18] A. WINTNER, "On the stable distribution laws," *Amer. J. Math.*, Vol. 55 (1933), pp. 335–339.

[19] A. WINTNER, "On the shape of the angulár case of Cauchy's distribution curves," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 589–593.

[20] F. ZERNIKE, "Wahrscheinlichkeitsrechnung und mathematische Statistik," *Handbuch der Physik*, Berlin, Springer, Vol. 3 (1928), pp. 419–492.