

Some Problems Connected With Rayleigh Distributions

M. M. Siddiqui

Contribution from Boulder Laboratories, National Bureau of Standards, Boulder, Colo.

(October 19, 1961)

This is an expository paper presenting the following: (1) the origin, and (2) the properties of the Rayleigh distribution; (3) the most efficient estimators of its parameters; (4) a test of the hypothesis that a set of observations is from a Rayleigh distribution; (5) the distribution of the ratio of two independent Rayleigh variates; and (6) the Rayleigh process derived from a normal process.

1. Introduction

In some problems of physics and engineering, such as the diffusion process, or the distribution of the power (or amplitude) of electromagnetic waves received through a scattering medium, we are concerned with the resultant of many (two- or three-dimensional) random vectors. Such problems are treated in probability theory under the heading of "random walk" in one, two, three, or, in general, k dimensions. Here, we will confine our attention to two-dimensional random walk problems which asymptotically give rise to the so-called Rayleigh distribution. Unfortunately, in radiowave propagation literature the name "Rayleigh distribution" is used indiscriminately both for the distribution of power and of amplitude. To distinguish between the two, we will call them *Rayleigh power distribution* and *Rayleigh amplitude distribution*, respectively. In statistical literature the distinction is clear: the Rayleigh power is a multiple of χ^2 (with 2 degrees of freedom), whereas the Rayleigh amplitude is a multiple of χ . Since the properties of the Rayleigh power distribution are much neater and better suited for statistical estimation and for testing of hypotheses than the properties of the Rayleigh amplitude distribution, we will study only the Rayleigh power distribution. Important characteristics of the amplitude distribution will be obtained from the relationship amplitude = (power)^{1/2}. The advantages of power over amplitude are roughly the same as those of variance over standard deviation, or of spectral density over its square root. For example, the powers are additive and amplitudes are not; the sum of two independent χ^2 variates is again a χ^2 variate with added degrees of freedom, but the sum of two χ variates is not a χ variate. It is, therefore, suggested that the name "Rayleigh distribution," if used without any other qualification, be reserved for Rayleigh power distribution, and "Rayleigh variate" for a multiple of χ^2 with 2 degrees of freedom.

2. Notation

Ordinarily the same letter will be used to designate both a random variable and the argument of its probability density or distribution function. Thus we may say: x is a random variable with probability density function $p(x)$ and distribution function $P(x)$. If it becomes necessary to distinguish between the two to avoid confusion or ambiguity, capital letters will be used for the random variables and lower case letters for the real variables. Thus we may say: X is a random variable and $Pr(x \leq X \leq x + dx) = p(x)dx$.

The function P , with one or more arguments, will be used generically for any distribution function, and p for any density function. Thus $P(x)$ and $P(y)$, in general, will not be the same functions. The former denotes the distribution function of x and the latter that of y , and the two may be different.

If x and y are random variables, Ex will stand for the expected value of x , and $E(x|y)$ for the conditional expected value of x given the value of y . Thus

$$Ex = \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x,y)dy dx,$$

$$E(x|y) = \int_{-\infty}^{\infty} x \frac{p(x,y)}{p(y)} dx.$$

It is easy to see that, if $Ex < \infty$,

$$\begin{aligned} E[E(x|y)] &= \int_{-\infty}^{\infty} dy p(y) \int_{-\infty}^{\infty} x \frac{p(x,y)}{p(y)} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x,y)dy dx = Ex. \end{aligned}$$

The characteristics of a theoretical distribution, i.e., its population parameters, will be designated by Greek letters such as μ , α , σ . Their estimates may be denoted by corresponding Latin letters, with or without subscripts or primes. Thus if μ denotes the

mean (which is also the median) of a normal distribution, m may denote the sample mean and m_1 the sample median.

3. Asymptotic Distribution of the Resultant

Let $\vec{r}_i = (r_i \cos a_i, r_i \sin a_i)$, $i=1, 2, \dots, n$ be independent random coplanar vectors where all values of a_i ($i=1, 2, \dots, n$) between 0 and 2π are equally likely. Let r_i be a series of positive constants, or else be positive random variables distributed independently of a_i . Then

$$Er_i \cos a_i = Er_i E \cos a_i = Er_i \frac{1}{2\pi} \int_0^{2\pi} \cos a_i da_i = 0, \quad (3.1)$$

$$\sigma_i^2 \equiv Er_i^2 \cos^2 a_i = Er_i^2 E \cos^2 a_i = \frac{1}{2} Er_i^2. \quad (3.2)$$

Similarly,

$$Er_i \sin a_i = 0, \quad Er_i^2 \sin^2 a_i = \frac{1}{2} Er_i^2, \quad Er_i^2 \sin a_i \cos a_i = 0, \quad (3.3)$$

so that the components of a vector are uncorrelated. This does not mean, however, that they are statistically independent. We note that if r_i is a constant $Er_i = r_i$, $Er_i^2 = r_i^2$.

The resultant vector $\vec{r} = (x, y)$ is given by

$$x = \sum_{i=1}^n r_i \cos a_i, \quad y = \sum_{i=1}^n r_i \sin a_i. \quad (3.4)$$

We immediately have

$$Ex = Ey = 0, \quad \sigma^2 = Ex^2 = Ey^2 = \sum_{i=1}^n \sigma_i^2. \quad (3.5)$$

Applying the Central Limit Theorem for vector variates [Cramér, 1951, pp. 215–216, 285–286], we obtain the following theorem.

THEOREM. Suppose that $\rho_i^3 = Er_i^3 < \infty$ for all i , and write

$$\rho^3 = \rho_1^3 + \dots + \rho_n^3.$$

If the condition

$$\lim_{n \rightarrow \infty} \rho/\sigma = 0$$

is satisfied, then, noting that x and y are uncorrelated, we obtain the asymptotic distribution of (x, y) as

$$dP(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy, \quad -\infty \leq x, \quad y \leq \infty.$$

Thus x/σ and y/σ are asymptotically independent standard normal variates.

We note that if r_i have identical distributions and $Er_i^3 < \infty$, or if $r_i = r$, a constant for all i , $\rho/\sigma = O(n^{-1/6})$, which tends to zero as $n \rightarrow \infty$.

If in the equation above we make the transformations

$$x = r \cos a, \quad y = r \sin a, \quad (3.6)$$

the Jacobian of the transformation is r , and the distribution of (r, a) is given by

$$dP(r, a) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr \cdot \frac{da}{2\pi}, \quad 0 \leq r \leq \infty, \quad 0 \leq a \leq 2\pi. \quad (3.7)$$

Hence the resultant phase, a , and the resultant amplitude, r , are independently distributed. We will write $\gamma = 2\sigma^2$, so that the distribution of r is

$$dP(r) = \frac{2r}{\gamma} e^{-\frac{r^2}{\gamma}} dr, \quad 0 \leq r \leq \infty. \quad (3.8)$$

This distribution is known as *Rayleigh amplitude distribution*. Making further the transformation

$$z = r^2 \quad (3.9)$$

we obtain the distribution of z , called the *Rayleigh power distribution*, as

$$dP(z) = \frac{1}{\gamma} e^{-\frac{z}{\gamma}} dz, \quad 0 \leq z \leq \infty. \quad (3.10)$$

We note that $z = x^2 + y^2$ is the sum of two independent normal variates, x and y , each with mean zero and variance $\gamma/2$. Thus $2z/\gamma$ is a χ^2 variate with 2 degrees of freedom [Cramér 1951, p. 233].

In case all $r_i = 1$ in (3.4), Kluyver [1906] showed that

$$P(r) = r \int_0^\infty [J_0(t)]^n J_1(rt) dt. \quad (3.11)$$

If we make the transformation

$$t = 2n^{-1/2}y$$

and observe that

$$\lim_{n \rightarrow \infty} [J_0(2n^{-1/2}y)]^n = e^{-y^2},$$

an asymptotic approximation to (3.11) is obtained as

$$P(r) \cong 1 - e^{-\frac{r^2}{n}}.$$

This formula is originally due to Rayleigh [1919].

In general, if we start with k dimensional vectors $\vec{r}_i = (x_{i1}, \dots, x_{ik})$, and $\vec{r} = \sum_{i=1}^n \vec{r}_i = (x_1, \dots, x_k)$, then, under the conditions of the Central Limit Theorem, x_1, \dots, x_k are asymptotically independent normal variates, each with mean zero, and the same variance, $\gamma/2$. Let $z = x_1^2 + \dots + x_k^2$. Then $2z/\gamma$ is a χ^2 variate with k degrees of freedom [Cramér 1951, p. 233]. Thus the asymptotic distribution of z is given by

$$dP(z) = \left[\gamma^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) \right]^{-1} e^{-\frac{z}{\gamma}} z^{\frac{k}{2}-1} dz; \quad 0 \leq z < \infty.$$

4. Rayleigh Power Distribution

4.1. Distribution Characteristics

In this section we will study some population characteristics of the Rayleigh power distribution. In the following sections the problems of estimation and of testing of hypotheses concerning this distribution will be considered.

The probability density function of z is

$$p(z) = \frac{1}{\gamma} e^{-z/\gamma}, \text{ if } z \geq 0; 0, \text{ otherwise,} \quad (4.1)$$

which has a maximum at $z=0$ and decreases monotonically as z increases.

The distribution function, $Pr(Z \leq z)$, is given by

$$P(z) = \frac{1}{\gamma} \int_0^z e^{-\frac{x}{\gamma}} dx = 1 - e^{-\frac{z}{\gamma}}. \quad (4.2)$$

If q is a number between 0 and 1, the 100 q percentile point, ζ_q , is defined by the equation

$$P(\zeta_q) = q,$$

which has the solution

$$\zeta_q = -\gamma \ln(1-q). \quad (4.3)$$

In particular the median

$$\zeta_{.5} = \gamma \ln 2 = 0.69315 \gamma.$$

The values of ζ_q/γ for $q=0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99$, are given in table 1.

TABLE 1. ζ_q/γ for selected q

q	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
ζ_q/γ	0.01005	0.05129	0.10536	0.28768	0.69315	1.38629	2.30259	2.99573	4.60517

If k is any real number greater than -1

$$\mu_k(z) = Ez^k = \frac{1}{\gamma} \int_0^\infty z^k e^{-\frac{z}{\gamma}} dz = \gamma^k \Gamma(k+1). \quad (4.4)$$

Thus the mean, $\mu_1(z)$, and the variance, $\sigma^2(z)$, are

$$\mu_1(z) = \gamma, \quad \sigma^2(z) = \mu_2(z) - \mu_1^2(z) = \gamma^2. \quad (4.5)$$

Also the mean and variance of $r = z^{1/2}$ are given by

$$Er = \mu_{1/2}(z) = \gamma^{1/2} \Gamma(3/2) = \frac{1}{2}(\pi\gamma)^{1/2},$$

$$\sigma^2(r) = \mu_1(z) - \mu_{1/2}^2(z) = \gamma \frac{(4-\pi)}{4}. \quad (4.5')$$

4.2. Estimation

The Rayleigh distribution is completely specified if the parameter γ is known. If z_1, z_2, \dots, z_N is a sample of N independent observations from this distribution, the likelihood function (the joint probability density of z_1, \dots, z_N considered as a function of γ) is given by

$$L(\gamma) = \frac{1}{\gamma^N} e^{-\frac{\sum z_i}{\gamma}} = \frac{1}{\gamma^N} e^{-\frac{Nc}{\gamma}}, \quad (4.6)$$

where

$$c = \frac{1}{N} \sum_{i=1}^N z_i. \quad (4.7)$$

This shows that c is sufficient for γ . Since $Ec = \gamma$, c is unbiased. We know that if an unbiased sufficient estimator exists, it is automatically the most efficient estimate, also the maximum likelihood estimate. The variance of c is, of course,

$$\text{var } c = \frac{\gamma^2}{N}. \quad (4.8)$$

If t is any other unbiased estimate of γ , then $\text{var } c \leq \text{var } t$.

Since $2z_i/\gamma$ is a χ^2 variate with 2 degrees of freedom, and z_1, \dots, z_N are independent

$$2Nc/\gamma = 2 \sum_{i=1}^N z_i/\gamma,$$

is a χ^2 variate with $2N$ degrees of freedom. Thus the distribution of c is

$$dP(c) = \frac{N^N}{\gamma^N \Gamma(N)} e^{-\frac{Nc}{\gamma}} c^{N-1} dc, \text{ if } c \geq 0; = 0, \text{ otherwise.} \quad (4.9)$$

To set up confidence limits on γ with a confidence coefficient $1-\alpha$, we determine from χ^2 tables two numbers χ_1^2 and χ_2^2 , corresponding to $2N$ degrees of freedom, such that

$$Pr(\chi^2 \leq \chi_1^2) = \alpha/2, \quad Pr(\chi^2 \leq \chi_2^2) = 1 - \alpha/2.$$

Then

$$\chi_1^2 \leq 2Nc/\gamma \leq \chi_2^2,$$

has probability $1-\alpha$ and

$$\frac{2Nc}{\chi_2^2} \leq \gamma \leq \frac{2Nc}{\chi_1^2}, \quad (4.10)$$

will be 100(1- α) percent confidence limits for γ .

Example 1. Let $N=3$, $1-\alpha=0.98$. From χ^2 tables we find, for 6 degrees of freedom, $Pr(\chi^2 \leq 0.872) = 0.01$, $Pr(\chi^2 \leq 16.812) = 0.99$. Hence

$$\frac{6}{16.812} c \leq \gamma \leq \frac{6}{0.872} c$$

or

$$0.357c \leq \gamma \leq 6.88c$$

are 98 percent confidence limits for γ . Similarly

$$0.476c \leq \gamma \leq 3.67c$$

are 90 percent confidence limits.

If $N > 15$ so that the number of degrees of freedom $f = 2N > 30$,

$$\sqrt{2x^2} - \sqrt{2f-1} = \sqrt{\frac{4Nc}{\gamma}} - \sqrt{4N-1}$$

is approximately a standard normal variate [Cramér, 1951, p. 251]. If x is a number such that

$$\frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{y^2}{2}} dy = 1 - \alpha, \quad (4.11)$$

then

$$\frac{4Nc}{(\sqrt{4N-1} + x)^2} \leq \gamma \leq \frac{4Nc}{(\sqrt{4N-1} - x)^2} \quad (4.12)$$

are $100(1 - \alpha)$ percent confidence limits for γ . When N is large, say $N > 100$, (4.12) is approximated by

$$\frac{\sqrt{Nc}}{\sqrt{N} + x} \leq \gamma \leq \frac{\sqrt{Nc}}{\sqrt{N} - x} \quad (4.13)$$

Example 2. Let $N = 100$, $1 - \alpha = 0.95$. We find $x = 1.96$. From (4.12)

$$0.831c \leq \gamma \leq 1.23c.$$

If, on the other hand, we use (4.13), we get

$$0.836c \leq \gamma \leq 1.24c.$$

Since c is the sufficient statistic for the Rayleigh distribution, if it is known that the sample has come from a Rayleigh distribution, we simply evaluate c and estimate the entire distribution by

$$\hat{P}(z) = 1 - e^{-\frac{z^2}{c}}, \quad 0 \leq z \leq \infty. \quad (4.14)$$

The usual practice of estimating the population percentiles by the sample percentiles is utterly unjustified. The distribution percentile, ζ_q , should be estimated by

$$z_q = -c \ln(1 - q), \quad (4.15)$$

which is an unbiased and minimum variance estimator of ζ_q . Thus, for example, the median, $\zeta_{.5}$, will be estimated by

$$z_{.5} = c \ln 2 = 0.69315c,$$

which has the variance

$$\text{var } z_{.5} = (\ln 2)^2 \text{ var } c = \frac{0.480\gamma^2}{N}$$

On the other hand, the asymptotic variance of the sample median, $z'_{.5}$, is given by

$$\text{var } z'_{.5} \approx \frac{1}{4Np^2(\zeta_{.5})} = \frac{\gamma^2}{N}$$

Thus

$$\frac{\text{var } z_{.5}}{\text{var } z'_{.5}} \approx 0.480,$$

i.e., the sample median is only 48 percent as efficient as $z_{.5}$.

To take another example: the first decile, $\zeta_{.1}$, is most efficiently estimated by

$$z_{.1} = -c \ln(.90) = 0.10536c,$$

which is unbiased and has variance

$$\text{var } z_{.1} = 0.0111 \frac{\gamma^2}{N}$$

The sample 1st decile, $z'_{.1}$, is biased and has asymptotic variance

$$\text{var } z'_{.1} = \frac{0.1 \times 0.9}{Np^2(\zeta_{.1})} = \frac{\gamma^2}{9N}$$

so that

$$\frac{\text{var } z_{.1}}{\text{var } z'_{.1}} \approx 9 \times 0.0111 \approx 0.1,$$

i.e., the efficiency of $z'_{.1}$ is only 10 percent.

4.2.1. Estimation of the Amplitude Characteristics

We know that if sufficient statistics are available for a distribution, its characteristics are most efficiently estimated in terms of these statistics. Thus the characteristics of the amplitude distribution are estimated in terms of c .

For example to estimate the mean amplitude, $(\pi\gamma)^{1/2}/2$, we need only to find a function of c which is an unbiased estimator of this quantity.

Now

$$\begin{aligned} Ec^{\frac{1}{2}} &= E(\gamma + c - \gamma)^{\frac{1}{2}} \\ &= \gamma^{\frac{1}{2}} E \left[1 + \frac{1}{2} \frac{c - \gamma}{\gamma} - \frac{1}{8} \frac{(c - \gamma)^2}{\gamma^2} + \dots \right] \\ &\approx \gamma^{\frac{1}{2}} \left[1 - \frac{1}{8N} \right]. \end{aligned}$$

Thus, to the order N^{-1} , $[1 + 1/(8N)]c^{1/2}$ is an unbiased estimate of $\gamma^{1/2}$. The mean of r is then estimated by

$$m = \frac{1}{2} \left(1 + \frac{1}{8N} \right) \sqrt{\pi c}. \quad (4.16)$$

Since

$$\begin{aligned} \text{var } c^{\frac{1}{2}} &= Ec - (Ec^{\frac{1}{2}})^2 \\ &\approx \gamma - \gamma \left(1 - \frac{1}{8N}\right)^2 \approx \frac{\gamma}{4N}, \\ \text{var } m &\approx \frac{\pi\gamma}{16N}. \end{aligned} \quad (4.17)$$

On the other hand, let $r_i = z_i^{\frac{1}{2}}$, and

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N r_i.$$

We have

$$\text{var } \bar{r} = \frac{\text{var } r_i}{N} = \frac{\gamma(4-\pi)}{4N}.$$

Thus

$$\frac{\text{var } m}{\text{var } \bar{r}} = \frac{\pi}{4(4-\pi)} \approx 0.91,$$

i.e., the sample mean, \bar{r} , is only 91 percent as efficient as the unbiased estimator based on the sufficient statistic c .

Similar considerations apply to estimating other characteristics of the amplitude distribution, and hence there is no need for us to go into details.

4.2.2. Estimation When the Data Are Truncated

Sometimes, due to the limitations of the recording instrument, observations above a certain level x are not recorded. However, the prior information is there that the unrecorded observations will be greater than x , and the recorded ones will be less than x . This prior information together with the actual values of the recorded observations should be utilized to obtain the most efficient estimator of γ .

In a sample of N independent observations let $N-k$ values exceed x , and let the remaining k observations be z_1, z_2, \dots, z_k . Obviously, k itself is a random variable with possible values $0, 1, 2, \dots, N$. Also the probability density function of z_i is now conditioned upon the knowledge that $0 \leq z_i \leq x$.

The conditional probability density function is

$$p(z|0 \leq z \leq x) = \frac{1}{\gamma p} e^{-z/\gamma}, \quad 0 \leq z \leq x,$$

where $p = 1 - e^{-x/\gamma}$. The probability distribution of k is

$$Pr(K=k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k=0, 1, \dots, N.$$

Hence the likelihood function is

$$L(\gamma) = \begin{cases} e^{-Nx/\gamma}, & \text{if } k=0, \\ \binom{N}{k} \frac{1}{\gamma^k} \exp \left[-\frac{1}{\gamma} \left\{ (N-k)x + \sum_{i=1}^k z_i \right\} \right], & \\ 0 \leq z_1, \dots, z_k \leq x, & k=1, 2, \dots, N. \end{cases} \quad (4.18)$$

Obviously, when $k=0$ no point estimate of γ is possible, since the only information available is that all observations are greater than x . In this case, which has probability $e^{-Nx/\gamma}$, we can only infer that $x < \gamma < \infty$. When $k \neq 0$, the maximum likelihood estimate of γ is

$$c_1 = \frac{(N-k)x + \sum_{i=1}^k z_i}{k}, \quad (4.19)$$

The properties of c_1 are not as easy to establish as those of c in section 4.2. Using the well known properties of the maximum likelihood estimators, however, we can say that for large N , c_1 will be normally distributed with mean Ec_1 , and variance, $\text{var } c_1$, where

$$\begin{aligned} Ec_1 &\approx \gamma + \frac{x(1-p)}{Np^2}, \\ \text{var } c_1 &\approx \frac{2\gamma^2}{Np} - \frac{(4p-3)x^2}{Np^3}, \quad p = 1 - e^{-x/\gamma}. \end{aligned} \quad (4.20)$$

4.3. Is the Variate Rayleigh Distributed?

Let z_1, z_2, \dots, z_N be independent observations on a non-negative variate z . We wish to test whether z is Rayleigh distributed. Proceed in the following manner.

Choose an integer m , not too small (at least 5), such that N/m is greater than 5, preferably greater

than 10. Calculate $c = \sum_{i=1}^N z_i / N$. Determine numbers x_1, \dots, x_{m-1} , such that

$$\hat{P}(x_i) = 1 - \frac{x_i}{e^c} = \frac{i}{m}, \quad i=1, 2, \dots, m-1,$$

i.e.,

$$x_i = c \ln \left(\frac{m}{m-i} \right), \quad i=1, 2, \dots, m-1.$$

The range $(0, \infty)$ of z is then divided into m nonoverlapping intervals $I_1 = (0, x_1)$, $I_2 = (x_1, x_2)$, \dots , $I_m = (x_{m-1}, \infty)$, such that the expected number of observations, under the hypothesis $\hat{P}(z)$, in each interval is N/m .

Let the number of actual observations falling in these intervals be f_1, f_2, \dots, f_m . Then

$$\chi_0^2 = \sum_{i=1}^m \frac{(f_i - N/m)^2}{N/m} = \frac{m}{N} \sum_{i=1}^m f_i^2 - N \quad (4.21)$$

is approximately a χ^2 variate with $m-2$ degrees of freedom [Cramér, 1951, pp. 424-426]. Pre-assign some critical probability level α ($=0.05$, or 0.01 , say) and let χ_α^2 be the number such that

$$Pr(\chi^2 \geq \chi_\alpha^2) = \alpha.$$

If the observed value $\chi_0^2 < \chi_\alpha^2$, accept the hypothesis that z is a Rayleigh variate, and estimate its dis-

tribution function by $\hat{P}(z)$. If $\chi_0^2 \geq \chi_\alpha^2$, reject the hypothesis of Rayleigh distribution.

If the sample is truncated, i.e., if the maximum value which could be recorded is x , we estimate γ by

$$c_1 = \left[(N-k)x + \sum_{i=1}^k z_i \right] / k, \text{ where } N-k \text{ values are}$$

greater than x and unrecorded. We proceed as above, changing the last interval (or intervals), if necessary, to read (x, ∞) and determining the expected number of observations in this interval by

$$Ne^{-x/c_1}; \text{ and calculating } \chi_0^2 = \sum \frac{(f_o - f_e)^2}{f_e}; f_o = \text{ob-}$$

served, $f_e = \text{expected frequency.}$

Example 3. The following is a systematic sample of 80 observations (read at 5 sec intervals) of received field intensity in (microvolts)². The following are the transmission parameters:

Frequency:	400 Mc/s
Transmitter power:	2 kw
Angle:	zero
Path length:	300 milliradians (165 miles)
Antennas:	60 ft diam parabolic reflectors having a gain of 38 db over an isotropic radiator
Loss:	The measured hourly median basic transmission loss \cong 209 db
Time:	0425 a.m., July 27, 1961

Observed values of received power in $(\mu v)^2$ (read left to right):

0.20, 0.71, 0.06, 0.05, 0.76, 0.32, 0.96, 0.63, 0.09, 0.18, 0.25, 0.45, 0.26, 0.10, 0.95, 0.01, 0.50, 1.26, 1.99, 0.32, 0.51, 0.01, 0.16, 0.56, 3.16, 1.27, 2.24, 1.00, 0.81, 1.29, 0.28, 0.21, 0.35, 0.20, 0.39, 0.89, 1.24, 0.08, 0.98, 1.01, 0.49, 0.90, 1.90, 1.42, 1.56, 1.32, 1.20, 1.59, 2.40, 2.24, 0.80, 0.56, 1.45, 0.18, 0.02, 0.28, 0.81, 0.18, 1.31, 0.64, 1.95, 0.48, 0.55, 0.44, 0.28, 0.07, 0.71, 0.48, 0.40, 0.06, 0.79, 1.01, 0.51, 0.70, 0.14, 0.16, 0.01, 0.06, 0.03, 0.01.

We have $N=80$, $c=0.71$. Take $m=8$, $\alpha=0.05$. For 6 degrees of freedom the critical value of χ^2 is 12.592.

$$\text{Calculate } x_i = -0.71 \ln \frac{8-i}{8}, \quad i=1, 2, \dots, 7.$$

The resulting class intervals and the observed frequencies are as follows.

Class intervals	f_o
0 - 0.095	13
0.095 - .204	9
.204 - .334	8
.334 - .492	8
.492 - .696	8
.696 - .984	13
.984 - 1.476	12
1.476 - ∞	9

The expected frequency in each interval is 10. We have

$$\chi_0^2 = \frac{8}{80} \sum f^2 - 80 = 3.6 < 12.592.$$

The data are consistent with the hypothesis that z is Rayleigh distributed. In fact, the probability that such a sample or worse, as measured by χ^2 , comes from a Rayleigh distribution is more than 70 percent. The entire distribution is estimated by

$$\hat{P}(z) = 1 - e^{-\frac{z}{0.71}}, \quad 0 \leq z \leq \infty.$$

4.4. Two Independent Samples

Sometimes we wish to test the hypothesis that two samples, which are known to be from Rayleigh distributions, are from the same distribution, i.e., $\gamma_1 = \gamma_2$. For example, to test the assumption of stationarity of a Rayleigh process, we may take two samples from the record (sufficiently far apart in time to insure independence), and test the hypothesis that $\gamma_1 = \gamma_2$. Or, we may be recording a signal, not continuously, but after breaks in between recordings, and may wish to test whether the characteristics of the received signal have changed from record to record.

Let it be known that two independent samples, z_1, z_2, \dots, z_N , and z'_1, z'_2, \dots, z'_M , are from Rayleigh distributions. To test the hypothesis that $\gamma_1 = \gamma_2$, it is sufficient to note that

$$F(2N, 2M) = \frac{\sum_{i=1}^N z_i}{2N} \cdot \frac{2M}{\sum_{i=1}^M z'_i} = \frac{c_1}{c_2}$$

is a Fisher-Snedecor F variate with the indicated degrees of freedom. Here, c_1 and c_2 denote the sample means of the first and second sample respectively. We assume that $c_1 \geq c_2$; if not we simply invert the ratio and interchange the degrees of freedom. Pre-assign a significance level α , and test for the significance of the calculated F .

Example 4. Consider the data in example 3. Let us test the hypothesis that the first 25 observations are from the same Rayleigh distribution as the last 25 observations. Let the significance level be taken as 0.05. From F tables the 5 percent significant value, for 50 and 50 degrees of freedom, is 1.60. Denoting by c_1 and c_2 the mean values of the first 25 and the last 25 observations, respectively, we have

$$c_1 = 0.578, \quad c_2 = 0.482.$$

Hence, the calculated value of F is

$$F_c = \frac{578}{482} = 1.20,$$

which is less than 1.60. Hence we accept the hypothesis that $\gamma_1 = \gamma_2$.

5. Distribution of the Ratio of Two Rayleigh Variates

Many problems in radio engineering require the knowledge of the distribution of the ratio of two independent Rayleigh variates, such as the signal power and the noise power, or the powers of two interfering signals.

Let z_1 and z_2 be two independent Rayleigh variates with parameters γ_1 and γ_2 . Let $u = z_1/z_2$, and $\epsilon = \gamma_1/\gamma_2$. Since $2z_1/\gamma_1$ and $2z_2/\gamma_2$ are independent χ^2 variates each with 2 degrees of freedom,

$$u_1 = \frac{z_1/\gamma_1}{z_2/\gamma_2} = \frac{z_1}{\epsilon z_2}$$

is a Fisher-Snedecor F variate with 2 and 2 degrees of freedom [Cramér, 1951, pp. 241-242]. Hence the distribution of $u = \epsilon u_1$ is given by

$$dP(u) = \frac{\epsilon du}{(\epsilon + u)^2}, \quad 0 \leq u \leq \infty; \quad (5.1)$$

$$1 - P(u) = Pr(z_1/z_2 \geq u) = \frac{\epsilon}{\epsilon + u}. \quad (5.2)$$

Example 5. Let the average signal power be 20 db above the average noise power so that $\epsilon = 100$. Let satisfactory service require signal power 10 db above the noise power. To calculate the probability of satisfactory service we set $\epsilon = 100$, $u = 10$ in (5.2) and obtain $Pr(z_1/z_2 \geq 10) = 0.91$.

The distribution of u involves only one parameter $\epsilon = \gamma_1/\gamma_2$, which is the median of the distribution. We note that the mean, variance, and higher moments of u do not exist.

Let u_1, \dots, u_N be independent observations on u . The logarithm of the likelihood function is

$$\ln L(\epsilon) = N \ln \epsilon - 2 \sum_{i=1}^N \ln(\epsilon + u_i), \quad (5.3)$$

which shows that no sufficient estimator exists for ϵ . The maximum likelihood estimator, e , of ϵ is the solution of the equation

$$\left. \frac{\partial \ln L}{\partial \epsilon} \right|_{\epsilon=e} = 0,$$

i.e., of

$$\frac{N}{e} - \sum_{i=1}^N \frac{2}{e + u_i} = 0. \quad (5.4)$$

This equation is quite difficult to solve if $N > 3$. We, therefore, find some less efficient but easily available estimator. The Cramér-Rao greatest lower bound for the variance of any unbiased estimator is found to be

$$v = \left[E \left(\frac{\partial \ln L}{\partial \epsilon} \right)^2 \right]^{-1} = \frac{3\epsilon^2}{N}, \quad (5.5)$$

which is also the asymptotic variance of e . The asymptotic variance of the sample median, $u'_{.5}$, is given by

$$\text{var } u'_{.5} \approx \frac{1}{4Np^2(\epsilon)} = \frac{4\epsilon^2}{N}, \quad (5.6)$$

so that the efficiency of the sample median is 75 percent. Thus, for a quick estimate of ϵ , the median may be used. *In any case the sample mean should not be taken as an estimate of ϵ .*

6. A Rayleigh Process Derived From a Normal Process

Referring back to the theoretical framework of section 3, we assume now that the resultant vector \vec{r} is a function of time, i.e., $\vec{r}(t)$ is a random vector process. The components (x, y) will now be written $(x(t), y(t))$ and it will be assumed that $x(t)$ and $y(t)$ are independent stationary normal processes with the same variance σ^2 and the same autocorrelation function $\rho(s)$. That is to say, for all t and s

$$Ex^2(t) = Ey^2(t) = \sigma^2,$$

$$Ex(t)x(t+s) = Ey(t)y(t+s) = \sigma^2\rho(s). \quad (6.1)$$

Set

$$z(t) = x^2(t) + y^2(t), \quad \gamma = 2\sigma^2. \quad (6.2)$$

Then, for all t and s ,

$$Ez(t) = \gamma,$$

$$Ez(t)z(t+s) = \gamma^2(1 + \rho^2(s)). \quad (6.3)$$

Writing $\alpha(s)$ for the autocorrelation of the $z(t)$ process, we have

$$\alpha(s) = \frac{\text{cov}(z(t), z(t+s))}{\gamma^2} = \rho^2(s). \quad (6.4)$$

The process is completely determined if γ and $\alpha(s) = \rho^2(s)$ are known.

As before we use the sample mean now given by

$$c = \frac{1}{T} \int_0^T z(t) dt, \quad (6.5)$$

to estimate γ . c is unbiased and has variance

$$\text{var } c = \frac{2\gamma^2}{T} \int_0^T \left(1 - \frac{s}{T}\right) \alpha(s) ds. \quad (6.6)$$

We note that in the case of a discrete sample $z(1), z(2), \dots, z(N)$, (6.6) will read

$$\text{var } c = \frac{\gamma^2}{N} + \frac{2\gamma^2}{N} \sum_{s=1}^{N-1} \left(1 - \frac{s}{N}\right) \alpha(s). \quad (6.6')$$

The exact distribution of c is unknown. However, let N' be defined by the equation

$$\text{var } c = \gamma^2/N', \text{ i.e., } N' = \gamma^2/\text{var } c, \quad (6.7)$$

where $\text{var } c$ is given by (6.6) or (6.6'). N' will be called *equivalent random sample size for estimating γ* . The results of sections 4.2-4.4, then, will approximately hold with N and M replaced by equivalent random sample sizes N' and M' . N' , in general, will not be an integer, but this presents no difficulty either in theory or in calculations. If $[2N']$ is the largest integer in $2N'$, we will simply interpolate between the percentiles of χ^2 for $[2N']$ and $[2N'] + 1$ degrees of freedom to obtain the percentile points of $2N'c/\gamma$. Similarly, for $F(2N', 2M')$ of section 4.4.

Example 6. Let $\alpha(s) = \alpha(-s) = e^{-\mu s}$, $s \geq 0$, $\mu > 0$. Also, let $\rho = e^{-\mu}$, so that $\alpha(s) = \rho^s$. From (6.6)

$$\text{var } c = \frac{2\gamma^2}{\mu T} \left[1 - \frac{1 - e^{-\mu T}}{\mu T} \right] \cong \frac{2\gamma^2}{\mu T}$$

if μT is large. Hence, the equivalent random sample size $N' \cong \mu T/2$.

In case of a discrete sample, we obtain from (6.6')

$$\text{var } c \cong \frac{\gamma^2}{N} \left(\frac{1 + \rho}{1 - \rho} \right),$$

so that $N' \cong N(1 - \rho)/(1 + \rho)$.

For example, if $e^{-\mu} = \rho = 0.5$, $\mu = 0.69315$, $2N' = 0.69315T$ in the continuous sample case, and $2N' = 2N/3$ in the discrete sample case.

Let

$$C(s) = \frac{1}{T-s} \int_0^{T-s} z(t) z(t+s) dt. \quad (6.8)$$

Then

$$a(s) = \frac{C(s)}{c^2} - 1 \quad (6.9)$$

is a consistent estimate of $\alpha(s)$.

From the joint distribution of $z(t)$ and $z(t+s)$, obtained from Siddiqui's [1961] equations (5.8) and (5.9) by setting $n=2$, we also find the following:

$$E(z(t+s)|z(t)) = \alpha(s)z(t) + \gamma(1 - \alpha(s)), \quad (6.10)$$

so that the regression of $z(t+s)$ on $z(t)$ is linear. However,

$$\text{var } (z(t+s)|z(t)) = \gamma^2(1 - \alpha(s))^2 + 2\gamma\alpha(s)(1 - \alpha(s))z(t), \quad (6.11)$$

is not independent of $z(t)$.

7. References

- Cramér, H., *Mathematical methods of statistics* (Princeton University Press, Princeton, 1951).
 Kluyver, J. C., A local probability problem, *Proc. Nederl. Akad. Wetensch.* **8**, 341-350 (1906).
 Rayleigh, J. W. Strutt, Lord, On the problem of random vibrations and of random flights in one, two, and three dimensions, *Phil. Mag.* **37**, 321-347 (1919).
 Siddiqui, M. M., Some properties of the empirical distribution function of a random process, *J. Research NBS* **65B** (Math. and Math. Phys.), No. 2, 117-127 (April-June 1961).

(Paper 66D2-184)