

Statistical Inference for Rayleigh Distributions

M. M. Siddiqui¹

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

(Received December 6, 1963; revised May 7, 1964)

The main inference problems related to the Rayleigh distribution are the estimation of its parameter and the test of the hypothesis that a given set of observations is from such a distribution. It is shown that (in case of radio signals) the most efficient estimate of the parameter is obtained using the sample mean power. Complications may arise when data are missing or are autocorrelated. Methods are given to deal with such complications also.

1. Introduction

In a previous paper [Siddiqui, 1962] some problems of estimation and testing of hypotheses were discussed in connection with the Rayleigh distribution. In the present survey article, for the sake of completeness, some parts of the previous paper will be repeated. However, an attempt will be made to condense the material already covered.

1.1. Notation

The function P 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. If X is a random variable EX will stand for the expected value of X and $\text{var } X$ for the variance of X . Thus

$$P(x) = Pr(X \leq x), EX = \int_{-\infty}^{\infty} x dP(x),$$

$$\text{var } X = EX^2 - (EX)^2,$$

and if $P(x)$ is differentiable, $P'(x) = p(x)$.

2. General Properties of Rayleigh Distributions

When an incoming plane wave (electromagnetic, sound, or some other kind) passes through a scattering medium, which at each instant of time may be considered as a sample from an ensemble of media with stable statistical characteristics, the outgoing wave $\vec{R} = (R \cos A, R \sin A)$ may be thought of as composed of many independent random wavelets $\vec{R}_i = (R_i \cos A_i, R_i \sin A_i)$, $i = 1, 2, \dots, n$. The problem in its most generality will impose no restrictions on the distributions of \vec{R}_i . However, under certain special circumstances, it may be reasonable to assume the following:

(1) *Complete incoherence*, which means that each of A_i is distributed uniformly over the interval $(0, 2\pi)$.

(2) *Independence of phase and amplitude*, i.e., R_i and A_i are independent random variables.

(3) *Absence of a dominant vector*, defined by the condition of the Central Limit Theorem [Cramér, 1951, pp. 215-216, 185-186]. If the first two conditions are satisfied, this third condition will be satisfied if: $(E \sum_1^n R_i^3)^{1/3} (E \sum_1^n R_i^2)^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$. Thus, in particular, if R_i have identical distributions or if $R_i = r$, a constant for all i , then (3) is satisfied.

The condition (3) will not be satisfied if, say, a dominant vector is present. Under such circumstances the Nakagami-Rice distribution applies [Nakagami, 1940; Rice, 1944, and 1945].

If the above assumptions hold $R \cos A$ and $R \sin A$ are asymptotically independent normal variates, each with mean zero and variance $(1/2)\sigma^2$, where $\sigma^2 = ER^2$. The distribution of (R, A) is approximately given by

$$dP(r, a) = \frac{2r}{\sigma^2} e^{-r^2/\sigma^2} dr \frac{da}{2\pi}, 0 \leq r < \infty, 0 \leq a < 2\pi,$$

and R and A are asymptotically independent. We, therefore, consider a random variable R , which has the probability density and distribution function, respectively

$$p(r) = 2\sigma^{-2} r e^{-r^2/\sigma^2}, P(r) = 1 - e^{-r^2/\sigma^2}, 0 \leq r < \infty.$$

This distribution is known as the *Rayleigh amplitude distribution*. We immediately have $p(0) = p(\infty) = 0$, and $p(r)$ has a unique maximum at the mode $r_0 = 2^{-1/2}\sigma = 0.707\sigma$. The *quantile function*, $r(p)$, that is, the inverse function of $P(r)$ is given by

$$r(p) = \sigma[-\ln(1-p)]^{1/2}, 0 \leq p < 1, \quad (2.2)$$

and gives the *100 p percentile*, $r(p)$, of the distribution (2.1). Thus if $[-\ln(1-p)]^{1/2}$ is plotted horizontally against $r(p)$ vertically we obtain a straight line through the origin with slope σ . The values

¹ Now at Colorado State University, Fort Collins, Colo.

of $r(p)/\sigma$ are given in table 1 column 2, in particular, the median

$$r(0.5) = \sigma(\ln 2)^{1/2} = 0.832\sigma.$$

For $k > -2$

$$ER^k = 2\sigma^{-2} \int_0^\infty r^{k+1} e^{-r^2/\sigma^2} dr = \sigma^k \Gamma(1+k/2). \quad (2.3)$$

The mean, ER , and the variance, $\text{var } R$, are

$$ER = (1/2)\pi^{1/2}\sigma = 0.866\sigma, \quad \text{var } R = (1 - \pi/4)\sigma^2 = 0.2146\sigma^2.$$

Note that $ER > r(0.5) > r_0$, which indicates that the right-hand tail of the probability density function is much longer than the left-hand one (positive-skewness).

Set $Z = R^2$, $\gamma = \sigma^2$. If R is identified as the amplitude of a wave, then Z is its power and $\gamma = EZ$ the mean power. The probability density and distribution functions of Z are, respectively,

$$p(z) = \gamma^{-1} e^{-z/\gamma}, \quad P(z) = 1 - e^{-z/\gamma}, \quad 0 \leq z < \infty. \quad (2.4)$$

This distribution is known as *Rayleigh power distribution* or *exponential distribution*. The density function has its maximum at $z=0$ and decreases monotonically to zero as z increases to infinity. If $z(p)$ is the quantile function, then $z(p) = r^2(p)$, i.e.,

$$z(p) = -\gamma \ln(1-p), \quad 0 \leq p < 1. \quad (2.5)$$

If $-\ln(1-p)$ is plotted horizontally against $z(p)$ vertically we obtain a straight line through the origin with slope γ . The values of $z(p)/\gamma$ are given in table 1 column 3. If $k > -1$

$$E(Z/\gamma)^k = \int_0^\infty z^k e^{-z} dz = \Gamma(k+1). \quad (2.6)$$

Thus $EZ = \gamma$, $\text{var } Z = \gamma^2$. Differentiating (2.6) j times with respect to k and then setting $k=0$, we obtain

$$E[\ln(Z/\gamma)]^j = \int_0^\infty (\ln z)^j e^{-z} dz = \Gamma^{(j)}(1), \quad j=0, 1, \dots$$

Thus

$$E(\ln Z) = \ln \gamma + \psi(1) = \ln \gamma - 0.577$$

$$\text{var}(\ln Z) = \psi'(1) = \pi^2/6 = 1.64493,$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$. It is customary to measure power in decibels. Thus, writing $Q = 10 \log_{10} Z = 20 \log_{10} R$, we have

$$EQ = (10 \log_{10} e) E(\ln Z) = 4.343 (10 \log_{10} \gamma - 0.577)$$

$$\text{var } Q = (10 \log_{10} e)^2 \text{var}(\ln Z) = 31.0 \quad (2.7)$$

We note that the variance and also other central moments of Q are independent of the parameter γ . This is not unexpected since the distribution of (Z/γ) , hence that of $\ln Z - \ln \gamma$, is free of parameters.

The quantile function, $q(p)$, of Q is easily evaluated without explicitly writing down the distribution function of Q . In fact, the events " $Q \leq q(p)$ " and " $Z \leq z(P)$ " have the same probability if $q(p) = 10 \log_{10} z(p)$.

Thus

$$q(p) = 10 \log_{10} \gamma + 10 \log_{10}(z(p)/\gamma).$$

Hence $q(p) - q(0.5)$ is free of the parameter γ . This quantity, which measures the distance of the 100 p percentile from the median, is tabulated in table 1 column 4.

TABLE 1. Percentiles of R, Z, and Q

(1) 100 p	(2) r(p)/σ	(3) z(p)/τ	(4) q(p) - q(0.5)
0.1	0.0316	0.00100	-28.4
1	.100	.0101	-18.4
5	.227	.0513	-11.3
10	.324	.105	-8.18
25	.537	.288	-3.82
50	.832	.693	0.00
75	1.18	1.39	3.01
90	1.52	2.30	5.21
95	1.73	3.00	6.36
99	2.14	4.61	8.22
99.9	2.63	6.91	9.98

A comparison of (2.4) with (2.1), or (2.5) with (2.2) indicates that the properties of the Rayleigh power distribution are much neater and better suited for statistical estimation and for testing of statistical hypotheses. In fact, if we even start with (2.1), we find that the maximum likelihood (also the unbiased sufficient) estimator of σ^2 from random sample R_1, \dots, R_N is $N^{-1} \sum_1^N R_i^2 = N^{-1} \sum_1^N Z_i$, which is readily understood as (sample) "mean power" rather than "mean of amplitude square." Similarly, the reparametrization, with γ replacing σ^2 , simplifies our estimates. From here onwards, therefore, we will mainly study the Rayleigh power distribution as given in (2.4).

3. Estimation

3.1. Estimation From a Random Sample

The only essential parameter to be estimated in a Rayleigh power distribution is the mean γ . All other characteristics of the distribution such as the distribution function, moments and percentiles are functions of γ . Let Z_1, \dots, Z_N denote a complete sample of N independent observations from the distribution. The likelihood function (the joint probability density of Z_1, \dots, Z_N considered as a function of γ) is given by

$$L(\gamma) = \gamma^{-N} \exp(-\sum Z_i/\gamma).$$

This shows that $\sum Z_i$ a sufficient statistic for γ

Since $E\Sigma Z_i = N\gamma$, the unbiased sufficient (also the maximum likelihood) estimator of γ is the sample mean,

$$c = N^{-1}\Sigma Z_i \quad (3.1)$$

The variance of c is, of course, $\text{var } c = \gamma^2/N$, and if c' is any other unbiased estimator of γ then $\text{var } c < \text{var } c'$. Since $2Z_i/\gamma$ is a χ^2 variate with 2 deg of freedom and Z_1, \dots, Z_N are independent,

$$2Nc/\gamma = 2\Sigma Z_i/\gamma$$

is a χ^2 variate with $2N$ deg of freedom. Thus

$$X = Nc/\gamma \quad (3.2)$$

has the distribution

$$dP(x) = [\Gamma(N)]^{-1} e^{-x} x^{N-1} dx, \quad 0 \leq x < \infty, \quad (3.3)$$

from which the distribution of c is easily derived by substitution from (3.2). We note that if $t > -N$

$$EX^t = [\Gamma(N)]^{-1} \int_0^\infty x^{t+N-1} e^{-x} dx = \Gamma(t+N)/\Gamma(N). \quad (3.4)$$

Since c is the sufficient statistic for the family of Rayleigh distributions $\{P(z): 0 \leq \gamma < \infty\}$, all other characteristics of the distribution should be estimated in terms of c . For example

$$\hat{P}(z) = 1 - e^{-z^2/c}, \quad 0 \leq z < \infty, \quad (3.5)$$

is the maximum likelihood estimator of $P(z)$ and

$$\hat{z}(p) = -c \ln(1-p), \quad 0 \leq p < \infty, \quad (3.6)$$

that of $z(p)$. In general, if a function $f(\gamma)$ has an unbiased estimator, it also has an unbiased estimator based on the sufficient statistic c , and this estimator is the most efficient estimator of $f(\gamma)$. Also let $h(x)$ be a function which in some neighborhood of the point $x = \gamma$ is continuous and has continuous derivatives of the first and second order. Then, if $|h(x)| < e^{Kx}$ for all $x \geq 0$ where K is some constant independent of x , it can be shown that, for sufficiently large N ,

$$\begin{aligned} Eh(c) &= h(\gamma) + O(N^{-1}) \\ \text{var } h(c) &= N^{-1} \gamma^2 h'^2(\gamma) + O(N^{-3/2}). \end{aligned} \quad (3.7)$$

For example if we wish to estimate the mean amplitude $ER = (\pi\gamma)^{1/2}/2$, we consider $m = (\pi c)^{1/2}/2$ and from (3.7)

$$Em = (\pi\gamma)^{1/2}/2 + O(N^{-1}), \quad \text{var } m = \frac{\pi\gamma}{16N} + O(N^{-3/2}).$$

However, in this case, exact evaluations of $Ee^{1/2}$

and $\text{var } e^{1/2}$ are possible. In fact, from (3.2) and (3.3)

$$Ee^{1/2} = \frac{\Gamma(N+1/2)}{N^{1/2}\Gamma(N)} \gamma^{1/2}, \quad \text{var } e^{1/2} = \left[1 - \frac{\Gamma^2(N+1/2)}{N\Gamma^2(N)}\right] \gamma.$$

Thus an unbiased estimate of ER and its variance are

$$m^* = \frac{(\pi N)^{1/2}\Gamma(N)}{2\Gamma(N+1/2)} e^{1/2}, \quad \text{var } m^* = \frac{\pi}{2} \left(\frac{N\Gamma^2(N)}{\Gamma^2(N+1/2)} - 1 \right) \gamma. \quad (3.8)$$

For small N these exact expressions can be used. For large N ,

$$\begin{aligned} m^* &= 1/2 \left(1 + \frac{1}{8N} \right) (\pi c)^{1/2} + O(N^{-2}), \\ \text{var } m^* &= \frac{\pi\gamma}{16N} + O(N^{-2}), \end{aligned}$$

which is an improvement over the biased estimate m . On the other hand, let $\bar{R} = N^{-1}\Sigma_1^N R_i$, then $E\bar{R} = ER$,

$$\text{var } \bar{R} = \frac{4-\pi}{4N} \gamma.$$

Thus the relative asymptotic efficiency, $e(\bar{R})$, of \bar{R} is

$$e(\bar{R}) = \frac{\text{var } m^*}{\text{var } \bar{R}} = \frac{\pi}{4(4-\pi)} = 0.91,$$

i.e., the sample mean \bar{R} is a less efficient estimator of the population mean ER than the estimator m^* based on the sufficient statistic c .

If the measurements are in decibels, that is, if our sample consists of $Q_i = 10 \log_{10} Z_i$, $i = 1, 2, \dots, N$, then we note from (2.7) that an estimation of EQ , which involves $\ln \gamma$, is required. We can directly evaluate $E(\ln c)$ and $\text{var}(\ln c)$ by differentiating (3.4) once and twice with respect to t and then setting $t=0$. We note that

$$\begin{aligned} E(\ln X)^j &= \Gamma^{(j)}(N)/\Gamma(N), \quad j=1, 2, \dots, \\ \psi(N) &= \Gamma'(N)/\Gamma(N) = \psi(1) + \sum_{r=1}^{N-1} r^{-1}, \\ \psi'(N) &= \sum_{r=0}^{\infty} (N+r)^{-2}. \end{aligned}$$

Thus, since $\ln c = \ln X + \ln \gamma - \ln N$,

$$\begin{aligned} E(\ln c) &= \ln \gamma - \ln N + \psi(N) = \ln \gamma + O(N^{-1}) \\ \text{var}(\ln c) &= \text{var}(\ln X) = \psi'(N). \end{aligned}$$

Thus, to order N^{-1} , the unbiased estimator of EQ based on c is

$$Q^* = 4.343 (\ln c - 0.577), \quad \text{var } Q^* = (4.343)^2 \psi'(N). \quad (3.9)$$

On the other hand, if $\bar{Q} = N^{-1} \sum Q_i$, $E\bar{Q} = EQ$ and

$$\text{var } \bar{Q} = \frac{(4.343\pi)^2}{6N}.$$

Thus the efficiency $e_N(\bar{Q})$ of \bar{Q} relative to Q^* is

$$e_N(\bar{Q}) = \frac{\text{var } Q^*}{\text{var } \bar{Q}} = \frac{6N}{\pi^2} \sum_{r=0}^{\infty} (N+r)^{-2}.$$

For $N=1$, $e_N(\bar{Q})=1$, otherwise $e_N(\bar{Q}) < 1$. Using the Euler-Maclaurin formula

$$\begin{aligned} \sum_{r=0}^{k-1} f(r) &= \int_0^k f(x) dx - 1/2[f(k) - f(0)] \\ &+ \frac{1}{12} [f^{(1)}(k) - f^{(1)}(0)] - \dots \end{aligned} \quad (3.10)$$

to sum the series in $e_N(\bar{Q})$ we obtain, for large $N (\geq 10)$,

$$e_N(\bar{Q}) \cong \frac{6}{\pi^2} = 0.608$$

Thus the efficiency of \bar{Q} is approximately 61 per cent.

We have thus established the fact that when a complete random sample is available, one should invariably use estimates of population characteristics using appropriate functions of the sufficient statistic c . The use of the other estimators mentioned above involves considerable loss of efficiency.

3.2. Confidence Limits

We noted that $2Nc/\gamma$ is a χ^2 variate with $2N$ deg of freedom. To set up confidence limits on γ with a confidence coefficient $1-\alpha$ we determine from χ^2 tables two numbers a and b , corresponding to $2N$ deg of freedom, such that

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

Then $a \leq 2Nc/\gamma \leq b$ has probability $1-\alpha$ and

$$2Nc/b \leq \gamma \leq 2Nc/a \quad (3.11)$$

are the desired confidence limits. If h is a monotone function from $(0, \infty)$, then $h(2Nc/b)$ and $h(2Nc/a)$ will be $100(1-\alpha)$ percent confidence limits for $h(\gamma)$. Thus, for example, EQ and $P(z)$ are monotone functions of γ , and confidence limits for them can be easily written down.

If $N \geq 15$, $x = (4Nc/\gamma)^{1/2} - (4N-1)^{1/2}$ is approximately a standard normal variate [Cramér, 1951, p. 251]. If x_α is the number such that

$$(2\pi)^{-1/2} \int_{-x_\alpha}^{x_\alpha} e^{-(1/2)x^2} dx = 1 - \alpha,$$

then $4Nc[(4N-1)^{1/2} \pm x_\alpha]^{-2}$ are $100(1-\alpha)$ percent

confidence limits for γ . If $N \geq 100$, these limits are approximately $N^{1/2}c(N^{1/2} \pm x_\alpha)^{-1}$.

3.3. Estimation From Order Statistics

An estimate of γ can be made with reasonably high efficiency from a few order statistics. Let Z_1, \dots, Z_N denote, as before, a sample of N independent observations on Z and let $Y_1 \leq Y_2 \leq \dots \leq Y_N$ be their ordered values, e.g., $Y_1 = \min(Z_1, \dots, Z_N)$, $Y_N = \max(Z_1, \dots, Z_N)$. The probability density function of Y_k is given by

$$p(y) = \frac{N!}{(k-1)!(N-k)! \gamma^N} \exp[-(N-k+1)y/\gamma] [1 - \exp(-y/\gamma)]^{k-1}, \quad 0 \leq y < \infty.$$

The moment generating function of Y_k is

$$\Phi_k(u) = \int_0^\infty e^{uy} p(y) dy = \prod_{j=1}^k (N-k+j)(N-k+j+\gamma u)^{-1}.$$

From $\Phi_k(u)$ we easily evaluate, with $a_i = (N-i+1)^{-1}$,

$$EY_k = \gamma \sum_{i=1}^k a_i, \quad \text{var } Y_k = \gamma^2 \sum_{i=1}^k a_i^2. \quad (3.12)$$

Using the Euler-Maclaurin formula (3.10) to approximate these summations, I have shown elsewhere [Siddiqui, 1963] that the optimum unbiased estimator from a single order statistic corresponds to

$$k \cong 0.79681(N+1) - 0.39841 + 1.16312(N+1)^{-1} \quad (3.13)$$

where the value of k thus obtained is rounded off to the nearest integer. For this k ,

$$c_1 = Y_k / \sum_{i=1}^k a_i \cong 0.6275 Y_k \quad (3.14)$$

is an unbiased estimator of γ with efficiency

$$e_N(c_1) = \frac{\text{var } c}{\text{var } c_1} = 0.65 + O(N^{-1})$$

Comparing the efficiencies $e_N(c_1)$ and $e_N(\bar{Q})$, we notice that even a single order statistic, properly chosen, is more efficient than the mean of all the decibel values for estimating the parameter of a Rayleigh distribution. It may also be noted that the efficiency of the estimator based on the median, i.e., the order statistic corresponding to $k \cong (1/2)(N+1)$, is only 48 percent.

The asymptotically ($N \rightarrow \infty$) optimum unbiased estimator of γ from two order statistics is given by

$$c_2 = 0.523 Y_m + 0.179 Y_n, \quad m \cong 0.639(N+1), \\ n \cong 0.927(N+1), \quad (3.15)$$

with asymptotic efficiency $e(c_2) = 0.82$. Sarhan,

Greenberg, and Ogawa [1963] have given asymptotically optimum unbiased estimators of γ from 1 to 15 order statistics. A linear combination of as few as five properly chosen order statistics provides an estimator of γ whose efficiency exceeds 94 percent.

A big advantage of the estimators based on order statistics is that only a partial knowledge of the sample is required. They are especially good when low values of the sample are poorly recorded, not recorded or missing. Let us say that the lowest 60 percent of the sample is missing. We can still construct the estimators c_1 and c_2 , although the sample mean (even median) can not be determined. The optimum estimators from more than one order statistic cannot be constructed if, say the highest 8 percent or more values in the sample are missing. If only the highest values are missing so that we are left with $Y_1, \dots, Y_m, m \leq N$, the most efficient unbiased estimator of γ from these order statistics is

$$c_m = m^{-1} \left[\sum_1^m Y_i + (N-m) Y_m \right] \quad (3.16)$$

and has efficiency $e(c_m) = m/N$. In fact, $2mc_m/\gamma$ is a χ^2 variate [Epstein and Sobel 1954] with $2m$ deg of freedom. Confidence limits based on c_m can be constructed in exactly the same way as with c .

3.4. Estimation From Correlated Observations

In this section we consider $Z=Z(t)$ to be a stationary Rayleigh process and assume that for all t and s

$$EZ(t) = \gamma, EZ(t)Z(t+s) = \gamma^2(1+\alpha(s)), \quad (3.17)$$

so that $\alpha(s)$ is the correlation function of the process. Note that $\alpha(0)=1, \alpha(-s)=\alpha(s)$. Let $Z(t)$ be observed over an interval of time $0 \leq t \leq T$. From the analogy of the random sample estimate, we use the sample mean

$$c_T = T^{-1} \int_0^T Z(t) dt \quad (3.18)$$

to estimate γ . c_T is unbiased and has variance

$$\text{var } c_T = 2\gamma^2 T^{-1} \int_0^T (1-s/T) \alpha(s) ds. \quad (3.19)$$

We note that if we choose to take a discrete set of equally spaced observations $Z(h), Z(2h), \dots, Z(Nh)$, then the mean, c_N , will have the variance

$$\text{var } c_N = \frac{\gamma^2}{N} + \frac{2\gamma^2}{N} \sum_1^{N-1} \left(1 - \frac{s}{N}\right) \alpha(sh). \quad (3.20)$$

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

$$\text{var } c = \gamma^2/N', \text{ i.e., } N' = \gamma^2/\text{var } c, \quad (3.21)$$

where $\text{var } c$ is either (3.19) or (3.20). Then $2N'c_T/\gamma$ or $2N'c_N/\gamma$ is approximately a χ^2 type variate with $2N'$ deg of freedom.

Generally, the correlation function $\alpha(s)$ will be unknown and will have to be estimated from the sample. Let

$$C(s) = (T-s)^{-1} \int_0^{T-s} Z(t)Z(t+s) dt.$$

Then

$$a(s) = C(s)/c_T^2 - 1$$

is a consistent estimator of $\alpha(s)$. Usually, it is desirable to fit to $\alpha(s)$ some mathematically specified function such as $\exp(-\mu|s|), \exp(-\mu s^2), \mu(\mu+s^2)^{-1}$. Before choosing an approximating function, however, a more careful investigation of $a(s)$ near $s=0$ is necessary. For example, if we take $\exp(-\mu|s|)$ to represent $\alpha(s)$ then, since this function is not differentiable at $s=0$, we will be compelled to conclude that the process $Z(t)$ is not differentiable, and this may not be a very desirable state of affairs.

For the sake of simplicity we take the unit of time such that $T=1$. Let

$$y_i = Z\left(\frac{i}{N}\right) - Z\left(\frac{i-1}{N}\right), \quad i=1, 2, \dots, N,$$

where N is given different values, say, 25, 50, 100, etc. We calculate

$$S_N = \sum_1^N y_i y_{i-1}, \text{ and } NS_N.$$

It can be shown that

$$\text{plim}_{N \rightarrow \infty} S_N = \gamma^2[\alpha'(-0) - \alpha'(0)]$$

where $\alpha'(-0)$ is the left-hand and $\alpha'(0)$ is the right-hand derivative of $\alpha(s)$ at $s=0$. If $\alpha(s)$ is differentiable at zero then this limit is zero, otherwise not equal to zero. Thus the behavior of S_N as N increases will tell us whether $\alpha(s)$ is differentiable at $s=0$ or not. Now supposing $\alpha(s)$ is twice differentiable at $s=0$ with $\alpha''(0) \neq 0$, we must have

$$\text{plim } S_N = 0, \text{ plim } NS_N = -\gamma^2 \alpha''(0).$$

We already know, in this case, that $\alpha'(0)=0$. If $\exp(-\mu s^2)$ seems to represent $\alpha(s)$, then an estimate of μ is obtained as $m = NS_N/(2c_T^2)$.

4. Tests of Statistical Hypotheses

4.1. Tests for the Distribution Function

Let Z_1, \dots, Z_N denote independent observations on a nonnegative variate Z . We wish to test the statistical hypothesis that Z has Rayleigh power distribution. We employ the χ^2 goodness-of-fit test for this purpose. We first calculate $c = N^{-1} \sum_1^N Z_i$; then the numbers $x_i = -c \ln(1-p_i), p_i = i/m, i=1, 2, \dots, m-1$, where $m \geq 5$. From (3.6) we see that x_i are $100p_i$ percentiles of the Rayleigh power distribution with $\gamma=c$. The expected number of observa-

tions in each of the intervals $(0, x_1), (x_1, x_2), \dots, (x_{m-1}, \infty)$ is N/m . Let f_1, \dots, f_m be the number of observed Z_i 's falling in these intervals. The statistic

$$\chi^2 = \sum_1^m \frac{(f_i - N/m)^2}{N/m} = \frac{m}{N} \sum_1^m f_i^2 - N \quad (4.1)$$

is asymptotically a χ^2 variate with $m-2$ deg of freedom under the hypothesis that Z is a Rayleigh variate. We preassign some critical probability level $\alpha (= 0.05$ or 0.01 , say) and from χ^2 tables determine the number χ_α^2 such that $Pr(\chi^2 \geq \chi_\alpha^2) = \alpha$. If the observed $\chi^2 < \chi_\alpha^2$, we accept the hypothesis that Z is a Rayleigh variate, otherwise we reject it.

If the sample is truncated at one or both ends, we use order statistics to estimate γ and proceed as above changing the first and last intervals, if necessary, to read $(0, x)$ and (y, ∞) where x is a number slightly less than the lowest recorded observation and y is a number slightly greater than the highest recorded observation. The expected number of observations in $(0, x)$ is $N(1 - e^{-x/c})$ and in (y, ∞) is $Ne^{-y/c}$. We calculate

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

where f_0 = observed, f_e = expected frequency.

In case of correlated observations it is necessary to multiply the value of χ^2 thus obtained by N''/N where N'' is the equivalent number of independent observations. Thus $\chi_0^2 = (N''/N) \sum \frac{(f_0 - f_e)^2}{f_e}$ should be considered as an approximate χ^2 variate with $m-2$ degrees of freedom. It is difficult to determine N'' . However, if $\alpha(s)$ is the correlation function of $Z(t)$ process and the sample consists of $Z(h), \dots, Z(Nh)$,

$$N''/N \cong \left[\sum_{j=-\infty}^{\infty} \alpha^2(jh) \right]^{-1} = \left[1 + 2 \sum_1^{\infty} \alpha^2(jh) \right]^{-1},$$

appears to be a reasonably good approximation to N''/N .

4.2. Comparison of Two Samples

Sometimes we wish to test the hypothesis that two independent samples, which are known to be from Rayleigh power distributions, are from the

same distributions. Let c_1 and c_2 be the means of the samples of sizes N_1 and N_2 from Rayleigh distributions with means γ_1 , and γ_2 , respectively. Then, under the hypothesis $\gamma_1 = \gamma_2$,

$$F(2N_1, 2N_2) = c_1/c_2$$

is a Fisher-Snedecor F variate with the indicated degrees of freedom. We assume that $c_1 \geq c_2$; if not we simply invert the ratio and interchange the degrees of freedom, preassign a significance level α and test for the significance of the calculated F by comparing it with the upper $100\alpha/2$ percentage point of the F distribution. If the observations in each sample are autocorrelated, we modify the degrees of freedom from N to N' according to (3.21). If the F -test indicates acceptance of the hypothesis $\gamma_1 = \gamma_2 = \gamma$, the common mean, γ , is estimated by $c = N^{-1}(N_1 c_1 + N_2 c_2)$, $N = N_1 + N_2$.

5. References

- Cramer, H. (1951), *Mathematical methods of statistics* (Princeton University Press, Princeton).
 Epstein, B., and M. Sobel (1954), Some theorems relevant to life testing from an exponential distribution, *Annals of Mathematical Statistics* **25**, 373-381.
 Nakagami, Minoru (1940), Study on the resultant amplitude of many vibrations whose phases and amplitudes are random, *Nippon Elec. Comm. Eng. No.* 22, 69-92.
 Rice, S. O. (1944), Mathematical analysis of random noise, *Bell System Tech. J.* 23, 282-332. (Also published as *Bell Telephone Monograph B 1589* and included in selected papers on noise and stochastic processes, N. Wax, ed. Dover Publications, New York, N.Y. 1954.)
 Rice, S. O. (1945), Mathematical analysis of random noise, *Bell System Tech. J.* 24, 46-156.
 Sarhan, A. E., B. G. Greenberg, and J. Ogawa (1963), Simplified estimates for the exponential distribution, *Annals of Mathematical Statistics*, 34, 102-116.
 Siddiqui, M. M. (1962), Some problems connected with Rayleigh distributions, *J. Res. NBS 66D (Radio Prop.)*, No. 2, 167-174.
 Siddiqui, M. M. (1963), Optimum estimators of the parameters of negative exponential distributions from one or two order statistics, *Annals of Mathematical Statistics*, **34**, 117-121.

(Paper 68D9-400)