ON THE DISTRIBUTION OF THE HURST RANGE OF INDEPENDENT NORMAL SUMMANDS

A.A. Anis* and E.H. Lloyd**

RR-77-16
July 1977

This work was carried out at the University of Lancaster under the auspices of the Science Research Council, London, and at IIASA.

*Ain Shams University, Cairo, Egypt.
**University of Lancaster, UK.

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

PREFACE

As new techniques evolve for the use of water and as human needs increase, the management of water resources becomes a task of growing importance.

From 1974 to 1976, stochastic reservoir theory was one of the major research fields of the IIASA Water Project (now the Water Group of the Resources and Environment Area). A number of research reports were published on different aspects of this problem. The present report by two leading authorities on stochastic reservoir theory deals with the Hurst range which in a hydrological context is of special relevance to reservoir design and the simulation of river flows. It is also relevant to the analysis of the structure of stochastic processes and time series.

SUMMARY

The well-known empirical findings of H.E. Hurst have inspired a vast quantity of research on the ability of various mathematical models of geophysical time series to reproduce the Hurst effect. Anis and Lloyd have earlier obtained an exact result for the *expected value* of the Hurst range when the inflow process is a sequence of independent and identically distributed normal variables. The research described in the present report is an attempt to extend this so as to obtain an explicit formula for the *probability distribution* of the Hurst range for this process.

A general method is developed, and explicit results are obtained for the cases $n \leq 4$ (n denoting the number of summands considered). These distributions are of an unusual form, the probability density functions showing discontinuities and corners. Graphs are provided showing the quantitative behavior of these functions.

It is surmised that the problem of deriving exact explicit formulae for the distribution of Hurst's range for general values of n is of unmanageable complexity.

On the Distribution of the Hurst Range of Independent Normal Summands

INTRODUCTION

Hurst's range is that of accumulated sums of deviations from the mean of a set of identically distributed random variables, expressed in units of the sample standard deviation of the variables in question [1]. The concept of this range first arose in a hydrological context where it has special relevance to reservoir design and to the simulation of river flows. It is also relevant to the analysis of the structure of stochastic processes and time series.

While the Hurst range has been intensively studied using simulatory methods, the analytical theory has so far proved somewhat intractable. Progress toward an understanding of the distribution of the Hurst range of autocorrelated and not necessarily normal random variables has so far advanced no further than the expected value of a set of mutually independent normal variables (or, further, that of symmetrically correlated normal variables [2]; but since symmetrical correlation is not a physically possible form of geophysical autocorrelation, this generalization does not signify much progress). The expected values are discussed in this report.

A general solution to the distribution problem would have practical and theoretical value, for example with regard to the design of tests of significance of some river flow models.

The present work is an attempt to move toward a solution to this distribution problem. Restricting ourselves to mutually independent normal variables, we provide an appropriate formulation of the Hurst range for a set of n such variables, and obtain some general results relative to the distribution of this range. We summarize our earlier work on the expected value of the Hurst range, and develop a method for obtaining the required probability distribution. We have applied this method to obtain explicit formulae for the probability distribution for the cases where $n \leq 4$. It turns out that the density function is of a complicated and interesting type, exhibiting continuity intervals, corners, and discontinuities.

DEFINITION OF THE HURST RANGE

Let $X_1, X_2, \ldots, X_n$ denote a set of mutually independent random variables, with a common normal distribution, which we may without loss of generality take as standardized (expectation = 0, variance = 1). Let

$$\bar{X} = \sum_{1}^{n} Xr/n \tag{1}$$

denote the sample mean, and

$$D_n = \sqrt{\left\{\sum_{1}^{n} (X_r - \bar{X})^2 / n\right\}} \tag{2}$$

the sample standard deviation.

Hurst's range relates to the cumulative sums of the "adjusted variables" $X_r - \bar{X}$, $r = 1,2,\ldots,n$, scaled by the divisor D_n. Thus we require the "partial sums" of the scaled adjusted variables

$$Y_j = (X_j - \bar{X})/D_n \quad , \qquad j = 1,2,\ldots,n \quad , \tag{3}$$

namely

$$S_1 = Y_1 \quad , \quad S_2 = Y_1 + Y_2, \ldots, S_{n-1} = Y_1 + \cdots + Y_{n-1} \quad ,$$

and

$$S_n = Y_1 + \cdots + Y_n (=0) \quad . \tag{4}$$

Note: an unambiguous notation would require the replacement of our $\bar{X}$ by $\bar{X}_n$, and our S_r = by ${}_nS_r$. It is hoped that no confusion will arise from our simplification. Let

$$M_n = \max\ (S_1, S_2, \ldots, S_{n-1}, 0) \tag{5}$$

denote the largest of the nonnegative partial sums, and

$$L_n = \min\ (S_1, S_2, \ldots, S_{n-1}, 0) \tag{6}$$

the numerically largest of the nonpositive values, so that

$$M_n \geq 0 \text{ and } L_n \leq 0.$$

The Hurst range is then

$$W_n = M_n - L_n \quad , \qquad n = 2,3,\ldots, \tag{7}$$

(W_1 being undefined).

We note at this point that, in addition to the relation

$$Y_1 + Y_2 + \cdots + Y_n = 0 \quad , \tag{8}$$

the Y_r also satisfy the quadratic relation

$$Y_1^2 + \cdots + Y_n^2 = n \quad . \tag{9}$$

We may eliminate Y_n between these two, resulting in

$$2\left\{\sum_{r=1}^{n-1} Y_r^2 + \sum_{r=1}^{n-1}\sum_{s=1}^{n-1} Y_r Y_s\right\} = n \quad . \tag{10}$$

The Y_r are not mutually independent, but they are identically distributed and their joint distribution is such that

$$\text{distr }(Y_1,Y_2,\ldots,Y_n) = \text{distr }(Y_\alpha,Y_\beta,\ldots,Y_\nu) \tag{11}$$

for every permutation $(\alpha,\beta,\ldots,\nu)$ of the ordered set $(1,2,\ldots,n)$. Such variables are called *exchangeable*. This exchangeability is a consequence of the fact that the Y_r are defined in terms of the X_s, which are mutually independent and therefore exchangeable, and of $\bar{X}$ and D_n, which are symmetric functions of the X_j, whose values are unaffected by any rearrangement of the order of the X_j.

These considerations have simplifying consequences in studying the joint distribution of the S_r.

Further, since $Y_1 + \cdots + Y_n = 0$, it follows that

$$- Y_n = Y_1 + \cdots + Y_{n-1}$$

but

$$\text{distr}\ (-Y_n) = \text{distr}\ (Y_1)$$

whence

$$\text{distr}\ (S_1) = \text{distr}\ (S_{n-1}) \quad .$$

Similarly

$$\text{distr}\ (S_r) = \text{distr}\ (S_{n-r}) \quad , \qquad r = 1,2,\ldots,n-1 \quad .$$

THE EXPECTED VALUE OF THE HURST RANGE

The expected value of the Hurst range $E(W_n)$ has been shown by Anis and Lloyd [2] to be

$$E(W_n) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\sqrt{\pi}\ \Gamma\{\frac{1}{2}n\}} \sum_{r=1}^{n-1} \sqrt{\left(\frac{n-r}{r}\right)} \quad , \qquad n = 2,3,\ldots, \quad . \tag{12}$$

THE DISTRIBUTION OF THE HURST RANGE

The most promising direct procedure would seem to be to obtain the joint distribution of the scaled partial sums S_r, and then use the usual theory of the order statistics of correlated variables to obtain the range W_n. It is more convenient to work in terms of the Y_r, since these are exchangeable, where

$$Y_r = (X_r - \bar{X})/D_n \quad , \qquad r = 1,2,\ldots,n \tag{13}$$

and D_n is defined in (2).

All the Y_r can be expressed in terms of any subset of size n-2 of them, since

$$\sum_{1}^{n} Y_r = 0 \tag{14}$$

and

$$\sum_1^n Y_r^2 = n \quad . \tag{15}$$

We shall in fact retain n-1 of the Y's, say $Y_1, Y_2, \ldots, Y_{n-1}$, eliminating Y_n by using (14) from which

$$Y_n = - \sum_1^{n-1} Y_r \quad ;$$

the n-1 retained variables are then subject to the restraint

$$\sum_1^{n-1} Y_r^2 + \left(\sum_1^{n-1} Y_r \right)^2 = n \quad .$$

This may be expressed concisely in terms of the vector

$$\eta = (Y_1, Y_2, \ldots, Y_{n-1})'$$

(of order n-1), in the form

$$\eta'\eta + (1'\eta)^2 = n$$

or

$$\eta' A \eta = n \quad , \tag{16}$$

where

$$A = I + 11' \tag{17}$$

is a symmetric matrix of order (n-1). The vector η may then be regarded as the position vector, in (n-1)-space, of an appropriately distributed random point which lies on the surface of the ellipsoid (16).

This attractive geometrical picture can be improved by a further transformation, which will send the ellipsoid into a sphere. Accordingly, we introduce new random variables $Z_1, Z_2, \ldots, Z_{n-1}$, in the form of a vector

$$\zeta = (Z_1, Z_2, \ldots, Z_{n-1})' \quad ,$$

by means of a transformation

$$\eta = B\zeta \tag{18}$$

such that (16) becomes

$$\zeta'\zeta = 1 \quad . \tag{19}$$

Now, by (16) and (18), we have

$$\eta'A\eta = n = \zeta'B'A'B\zeta$$

whence the requirement of (19), namely that $\zeta'\zeta = 1$, can be satisfied by choosing the matrix B such that

$$B'AB = nI$$

or

$$BB' = nA^{-1} \quad .$$

Since A (of order n-1) has the simple form (17), its inverse A^{-1} is

$$A^{-1} = I - \frac{1}{n} 11' \quad ,$$

and we may take B to be the unique lower triangular matrix with positive diagonal elements such that

$$\begin{aligned} BB' &= nA^{-1} \\ &= nI - 11' \quad . \end{aligned} \tag{19a}$$

Then

$$\zeta = B^{-1}\eta \quad .$$

This "triangular resolution" is a standard matrix procedure, and the explicit form for B is readily found. For example, when n = 3, we have

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

and

$$3A^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} \surd 2 & 0 \\ -1/\surd 2 & \surd(3/2) \end{pmatrix} \begin{pmatrix} \surd 2 & -1/\surd 2 \\ 0 & \surd(3/2) \end{pmatrix}$$

whence

$$B = \begin{pmatrix} \surd 2 & 0 \\ -1/\surd 2 & \surd(3/2) \end{pmatrix} \tag{20}$$

$$= \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \cdot \operatorname{diag}\left\{\sqrt{\frac{3}{3.2}} \, , \sqrt{\frac{3}{2.1}}\right\} \quad .$$

Similarly when n = 4 we have

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

and

$$4A^{-1} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

so that

$$B = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ -1/\sqrt{3} & \sqrt{(2/3)} & 0 \\ -1/\sqrt{3} & -\sqrt{(2/3)} & \sqrt{2} \end{pmatrix} \tag{21}$$

$$= \begin{pmatrix} 3 & 0 & 0 \\ -1 & 2 & 0 \\ -1 & -1 & 1 \end{pmatrix} \text{diag}\left(\sqrt{\frac{4}{4.3}}\ ,\ \sqrt{\frac{4}{3.2}}\ ,\ \sqrt{\frac{4}{2.1}}\right) .$$

For a general value of n we find

$$B = B_1 B_2 \tag{22}$$

where

$$B_1 = \begin{bmatrix} n-1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & n-2 & 0 & \cdots & 0 & 0 & 0 \\ -1 & -1 & n-3 & & \vdots & \vdots & \vdots \\ -1 & -1 & -1 & & \vdots & \vdots & \vdots \\ \vdots & & & & \vdots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & -1 & 2 & 0 \\ -1 & -1 & -1 & \cdots & -1 & -1 & 1 \end{bmatrix} \tag{23}$$

and

$$B_2 = \text{diag}\left[\sqrt{\{n/(n - r)(n - r - 1)\}}\ ,\quad r = 0,1,\ldots,n-2\right] . \tag{24}$$

Since the n-1 variables $Z_1, Z_2, \ldots, Z_{n-1}$ are constrained to lie on the (n-1)-sphere

$$\sum_{1}^{n-1} Z_i^2 = 1 \quad ,$$

a sensible way of expressing their joint pdf (probability density function) is to give the pdf of any subset of them of size n-2, say the subset $Z_1, Z_2, \ldots, Z_{n-2}$. It turns out that this is

$$g\left(Z_1, Z_2, \ldots, Z_{n-2}\right) = a_n\left(1 - Z_1^2 - Z_2^2 - \cdots - Z_{n-2}^2\right)^{-\frac{1}{2}} \quad ,$$

$$Z_1^2 + Z_2^2 + \cdots + Z_{n-2}^2 \leq 1 \quad , \tag{25}$$

where

$$a_n = \Gamma\{\tfrac{1}{2}(n-1)\}/\Gamma(\tfrac{1}{2}n) \quad , \qquad n = 2,3,\ldots \quad .$$

The proof will be deferred (see Appendix).

Returning now to the main problem, we have

$$\begin{aligned} W_n &= (\max - \min)(S_1, S_2, \ldots, S_{n-1}, 0) \\ &= (\max - \min)(Y_1, Y_1 + Y_2, \ldots, Y_1 + \cdots + Y_{n-1}, 0) \\ &= (\max - \min)(b_1'\zeta, b_2'\zeta, \ldots, b_{n-1}'\zeta, 0) \end{aligned} \tag{26}$$

where, for r = 1,2,...,n-1, we have expressed $Y_1 + \cdots + Y_r$ as a linear function $b_r'\zeta$ of the Z_i by means of the transformation (18), with B as in (21), where the point S lies on the sphere $\zeta'\zeta = 1$, and the distribution of ζ is given by (25).

Then the pdf $g_n(w)$ of W_n is obtained by using the following argument:

$$g_n(w) = \sum_{\substack{i=1 \\ (i \neq j)}}^{n} \sum_{j=1}^{n} \pi_{ij}(w) \quad , \qquad n = 2,3,\ldots \tag{27}$$

where

$$\pi_{ij}(w)\, dw = P\{\text{max is } S_i, \text{ min is } S_j, \text{ and } S_i - S_j \in dw\} \quad .$$

Here "max" means "$\max(S_1, S_2, \ldots, S_{n-1}, 0)$", and similarly for "min". The abbreviation "$S_1 - S_j \in dw$" means "$w \leq S_i - S_j < w + dw$". Thus

$$\pi_{ij}(w) = P\{\text{max is } S_i, \text{ min is } S_j | S_i - S_j = w\}\, f_{ij}(w) \tag{28}$$

where $f_{ij}(w)$ is the pdf of $S_i - S_j$ at w.

In evaluating (19) we express the S_r in terms of ζ and use the distribution (25) of that vector. There are n(n-1) terms in the sum (27) but not all are distinct. In the first place we have the obvious symmetry relations

$$\pi_{ij}(w) = \pi_{ji}(w) \quad , \qquad i,j = 1,2,\ldots,n \tag{29}$$

which we abbreviate to

$$\pi_{ij} = \pi_{ji} \quad .$$

In addition we have

$$\pi_{12} = \pi_{23} = \cdots = \pi_{n-1,n} = \pi_{n1} \quad ,$$

$$\pi_{13} = \pi_{24} = \cdots = \pi_{n-2,n} = \pi_{n-1,1} = \pi_{n,2}$$

$$\pi_{14} = \pi_{25} = \cdots = \pi_{n-3,n} = \pi_{n-2,1} = \pi_{n-1,2} = \pi_{n,3}$$

and so on, or, in general

$$\pi_{ij} = \pi_{i+h,j+h} \quad , \qquad h = 1,2,\ldots \tag{30}$$

where addition in the subscripts is modulo n. This follows from the symmetry of the S_r and the exchangeability of the Y_r.

EXPLICIT EVALUATION OF Distr (W_n), n = 3

We now apply the foregoing to evaluate distr (W_n) in the cases n = 2,3,4. The case n = 2 is trivial since, in view of the restraints (8) and (9), namely

$$Y_1 + Y_2 = 0 \quad , \qquad Y_1^2 + Y_2^2 = 2 \quad ,$$

the Hurst range must be a constant. In fact, it is easy to see from first principles that W_2 is identically equal to unity; for

$$\bar{X} = \tfrac{1}{2}(X_1 + X_2) \quad ,$$

$$X_1 - \bar{X} = \tfrac{1}{2}(X_1 - X_2) \quad , \quad X_2 - \bar{X} = \tfrac{1}{2}(X_2 - X_1) \quad ,$$

$$D_2^2 = \tfrac{1}{4}(X_1 - X_2)^2 \quad , \quad D_2 = \tfrac{1}{2}|X_1 - X_2| \quad .$$

Then

$$\begin{aligned} M_2 &= \max\{X_1 - \bar{X}, 0\}/D_2 \\ &= \max\{\tfrac{1}{2}(X_1 - X_2), 0\}/\tfrac{1}{2}|X_1 - X_2| \\ &= \begin{cases} 1 \quad , & \text{if } X_1 > X_2 \\ 0 & \text{otherwise} \end{cases} \quad . \end{aligned}$$

Similarly

$$\begin{aligned} L_2 &= \min\{X_1 - \bar{X}, 0\}/D_2 \\ &= \begin{cases} 0 \quad , & \text{if } X_1 > X_2 \\ -1 & \text{otherwise} \end{cases} \quad . \end{aligned}$$

Then

$$W_2 = M_2 - L_2 = \begin{cases} 1 - 0 & , \text{ if } X_1 > X_2 \\ 0 - (-1) & , \text{ otherwise} \end{cases}$$

$$= 1 \quad , \quad \text{in either case} \quad .$$

The case of $n = 3$ is less trivial but still fairly simple. From (27) we have

$$g_3(w) = \sum_{i=1}^{3} \sum_{\substack{j=1 \\ (i \neq j)}}^{3} \pi_{ij}(w)$$

$$= \pi_{12} + \pi_{21} + \pi_{13} + \pi_{31} + \pi_{23} + \pi_{32} \quad .$$

On using the symmetry relations (29) and (30) it will be seen that all six terms have a common value, which it is convenient to take as π_{13}. Hence

$$g_3(w) = 6\pi_{13}(w) \quad . \tag{31}$$

In $n = 3$, the transforming matrix B is given by (20), whence

$$Y_1 = Z_1\sqrt{2} \quad , \qquad Y_2 = -Z_1/\sqrt{2} + Z_2\sqrt{(3/2)} \quad ,$$

with

$$Z_1^2 + Z_2^2 = 1 \quad .$$

The pdf of Z_1 at z is given by (25) as

$$g(z) = 1/\pi(1 - z^2)^{\frac{1}{2}} \quad , \qquad z^2 \leq 1 \quad . \tag{32}$$

To evaluate $\pi_{13}(w)$ we note that

$$\pi_{13}(w) = P\{\text{max is } S_1, \text{ min is } S_3(=0) | S_1 = w\}\ f_{13}(w) \quad ,$$

$$= P\{\text{max is } Y_1, \text{ min is } 0 | Y_1 = w\}\ f_{13}(w) \quad . \tag{33}$$

The factor $f_{13}(w)$ is the pdf at w of $S_1 - S_3 = S_1 = Y_1 = Z_1\sqrt{2}$. The pdf of Z_1 at z is given by (32); thus the pdf of $Z_1\sqrt{2}$ at w is

$$f_{13}(w) = 2^{-\frac{1}{2}} g(w/\sqrt{2}) \quad , \quad w^2/2 \le 1$$

$$= 1/\pi(2 - w^2)^{\frac{1}{2}} \quad , \quad w^2 \le 2 \quad . \tag{34}$$

The other factor in (33) is

$$P\left(0 < Y_1 + Y_2 < Y_1 | Y_1 = w\right) \tag{35}$$

since the proposition "max is S_1, min is $S_3(=0)$" is equivalent to

$$S_1 > S_2, S_1 > S_3(=0) \quad , \quad S_2 > S_3(=0)$$

that is, to

$$0 < S_2 < S_1$$

or to

$$0 < Y_1 + Y_2 < Y_1 \quad .$$

The expression (35) may be developed as

$$P\left(0 < Y_1 + Y_2 < Y_1 | Y_1 = w\right)$$

$$= P\left(0 < Z_1/\sqrt{2} + Z_2\sqrt{(3/2)} < Z_1\sqrt{2} | Z_1\sqrt{2} = w\right)$$

$$= P\left(0 < \tfrac{1}{2}w + Z_2\sqrt{(3/2)} < \tfrac{1}{2}w | Z_1\sqrt{2} = w\right) \tag{36}$$

$$= P\left(-\tfrac{1}{2}\sqrt{\tfrac{2}{3}}w < Z_2 < \tfrac{1}{2}\sqrt{\tfrac{2}{3}}w \,\middle|\, Z_1 = w/\sqrt{2}\right)$$

where $Z_1^2 + Z_2^2 = 1$. In geometrical terms, we have to consider the unit circle $Z_1^2 + Z_2^2 = 1$ and, for various values of w, the probability that for a point (Z_1,Z_2) on this circle, the co-ordinate Z_2 will lie between $\tfrac{1}{2}\sqrt{(3/2)}w$ and $-\tfrac{1}{2}\sqrt{(3/2)}w$ when the other coordinate Z_1 is defined by $Z_1 = w/\sqrt{2}$ (see Figure 1).

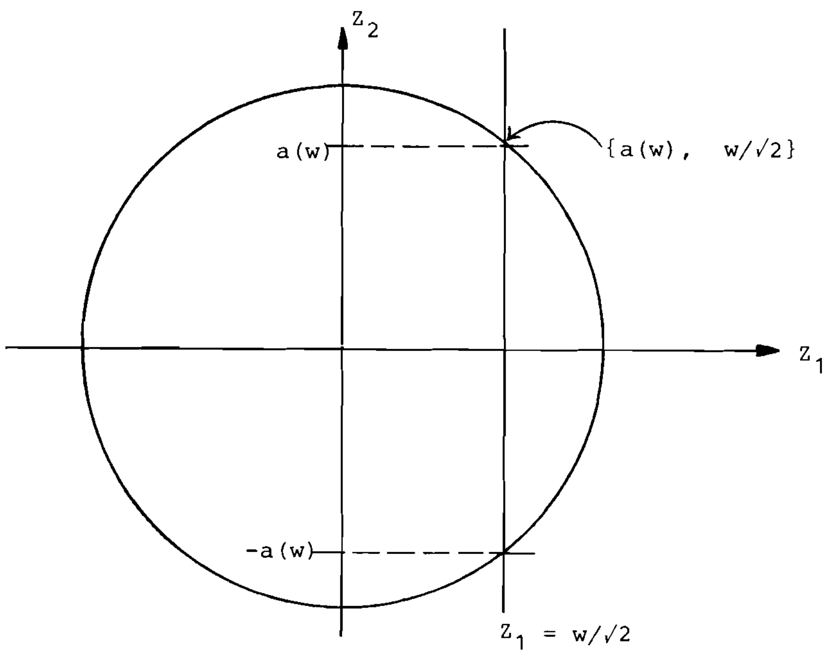

Figure 1.

Clearly the value $Z_1 = w/\sqrt{2}$ defines two values of Z_2, say $Z_2 = a(w)$ and $Z_2 = -a(w)$, where $\{a(w), w/\sqrt{2}\}$ is a point on the unit circle, so that $a(w) = +\sqrt{(1 - \tfrac{1}{2}w^2)}$. When

$$0 < a(w) < \tfrac{1}{2}\sqrt{\tfrac{2}{3}}w \tag{37}$$

the required conditional probability (36) is equal to 1, and in all other cases it is equal to zero. Now (37) is satisfied if and only if

$$1 - \tfrac{1}{2}w^2 < \tfrac{1}{6}w^2$$

that is

$$w^2 > \frac{2}{3} \quad . \tag{38}$$

(It is because of this geometrical interpretation that we refer to (38), and its analogies for general values of n, as the "geometrical factor".)

Combining these results we finally obtain from (31) and (33) the following expression for the pdf $g_3(w)$ of the Hurst range for $n = 3$:

$$\begin{aligned} g_3(w) &= 6/\pi\sqrt{(2 - w^2)} \quad , \quad \sqrt{(3/2)} < w < \sqrt{2} \\ &= 0 \quad , \quad \text{otherwise} \quad . \end{aligned} \tag{39}$$

The shape of this is illustrated by Figure 2. Note the corner at $w = \sqrt{(3/2)}$.

The expected value of W_3 is

$$\begin{aligned} E(W_3) &= \int_{-\infty}^{\infty} wg_3(w)\, dw = (6/\pi) \int_{\sqrt{(3/2)}}^{\sqrt{2}} w(2 - w^2)^{-\frac{1}{2}}\, dw \\ &= 3\sqrt{2}/\pi \end{aligned} \tag{40}$$

in agreement with the value given by (12), for $n = 3$.

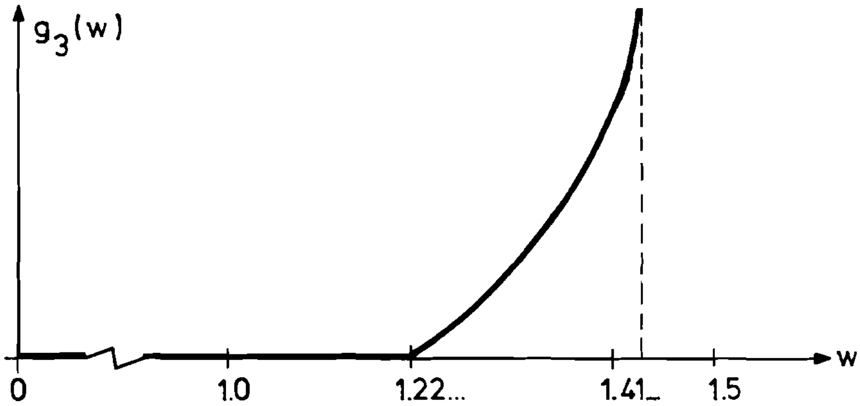

Figure 2.

Similarly one finds

$$\text{var}\ (W_3) = 1 + 3\sqrt{3}/2\pi - 18/\pi^2 \quad .$$

EXPLICIT EVALUATION OF Distr (W_n): (n = 4)

While in principle the general method outlined and illustrated for n = 3 may be applied for any value of n; the practical complexities increase rapidly with n. We have completed the calculations for n = 4, which proceed as follows.

By (25),

$$g_4(w) = \sum_{i=1}^{4} \sum_{\substack{j=1 \\ (i \neq j)}}^{4} \pi_{ij}(w) \quad . \tag{41}$$

Among the twelve terms π_{ij} we have the relations (29):

$$\pi_{12} = \pi_{21} = \pi_{23} = \pi_{32} = \pi_{34} = \pi_{43} = \pi_{41} = \pi_{14} \quad (= \pi_a,\ \text{say})$$

and

$$\pi_{13} = \pi_{31} = \pi_{24} = \pi_{42} \qquad (= \pi_b, \text{ say}) \quad .$$

The expression (41) thus reduces to

$$g_4(w) = 8\pi_a + 4\pi_b \quad . \tag{41a}$$

We consider separately the evaluation of π_a and π_b.

We shall need the transforming matrix B of (22), which for $n = 4$ gives

$$Y_1 = Z_1\surd 3$$

$$Y_2 = -Z_1/\surd 3 + 2Z_2\surd(2/3)$$

$$Y_3 = -Z_1/\surd 3 - Z_2\surd(2/3) + Z_3\surd 2 \quad ,$$

where

$$Z_1^2 + Z_2^2 + Z_3^2 = 1 \quad . \tag{41b}$$

In terms of the Z_i we find that

$$S_1 = Z_1\surd 3$$

$$S_2 = 2Z_1/\surd 3 + 2Z_2\surd(2/3) \tag{42}$$

$$S_3 = Z_1/\surd 3 + Z_2\surd(2/3) + Z_3\surd 2 \quad ,$$

where the joint pdf of any two of the Z's, say Z_1 and Z_2, is

$$\frac{1}{2\pi\surd(1 - Z_1^2 - Z_2^2)} \quad , \qquad Z_1^2 + Z_2^2 \leq 1 \quad ; \tag{43}$$

consequently the *common distribution of Z_1 and of Z_2 is uniform $(-1,1)$.*

It is a simple matter to verify from (43) that S_2 is also uniformly distributed, on (-2,2).

<u>Evaluation of $\pi_a(w)$</u>

We have

$$\pi_a(w) = \pi_{21}(w) = P\{S_2 \text{ is max}, S_1 \text{ is min} | S_2 - S_1 = w\} f_{21}(w) \tag{44}$$

where $f_{21}(w)$ is the pdf at w of $S_2 - S_1 = Y_1 = Z_1\sqrt{3}$. Since Z_1 is uniformly distributed on (-1,1), it follows that $Z_1\sqrt{3}$ is uniform on $(-\sqrt{3},\sqrt{3})$, with pdf

$$f_{21}(w) = \begin{cases} 1/2\sqrt{3} \quad , & -\sqrt{3} < w < \sqrt{3} \\ 0 \quad , & \text{otherwise} \end{cases} \quad . \tag{45}$$

As regards the "geometrical" factor in (44), that is the conditional probability term, we note that the proposition

"S_2 is max, S_1 is min"

is equivalent to the following:

$$S_2 > S_3 \quad , \quad S_2 > S_4(=0) \quad ,$$

$$S_1 < S_2 \quad , \quad S_1 < S_3 \quad , \quad S_1 < S_4(=0) \quad ,$$

that is, to

$$S_1 < 0 \quad , \quad S_2 > 0 \quad , \quad S_1 < S_3 < S_2 \quad .$$

The conditional probability term in (44) therefore reduces to

$$P\{S_1 < 0, S_2 > 0, S_1 < S_3 < S_2 | S_2 - S_1 = w\}$$

which, on using the relations (42), and replacing Z_1^2 by $1 - Z_2^2 - Z_3^2$ in accordance with (41a), reduces to

$$P\{a_1(w) < \ell_1(Z_2,Z_3) < b_1(w), a_2(w) < \ell_2(Z_2,Z_3) < b_2(w) \mid Z_2^2 + Z_3^2 = 1 - w^2/3\} \tag{46}$$

where $\ell_1(Z_2,Z_3)$ and $\ell_2(Z_2,Z_3)$ are linear functions of Z_2 and Z_3. To be precise:

$$a_1(w) = -w/\sqrt{6} \quad , \quad b_1(w) = w/2\sqrt{6} \quad , \quad a_2(w) = -2w/\sqrt{6} \quad ,$$
$$b_2(w) = w/\sqrt{6} \quad , \quad \ell_1(Z_2,Z_3) = Z_2 \quad , \quad \ell_2(Z_2,Z_3) = -Z_2 + Z_3\sqrt{3} \quad . \tag{47}$$

In geometrical terms, (46) represents the probability content of that portion (if any) of the circumference of the circle $Z_2^2 + Z_3^2 = 1 - w^2$ that lies between the lines

$$\ell_1(Z_2,Z_3) = a_1(w)$$

$$\ell_1(Z_2,Z_3) = b_1(w)$$

and also between the lines

$$\ell_2(Z_2,Z_3) = a_2(w)$$
$$\ell_2(Z_2,Z_3) = b_2(w) \quad . \tag{48}$$

Figure 3 gives a representation of a typical situation, the relevant parts of the circumference (shown hatched as arcs (ab) and (cd)) being those inside the parallelogram represented by (48).

Since by (45) $f_{21}(w)$ is a uniform distribution, it follows that p(w) will be proportional to the total length of the enclosed arcs (ab), (cd), and therefore, on normalizing, numerically equal to the ratio of the total arc length to the circumference of the circle.

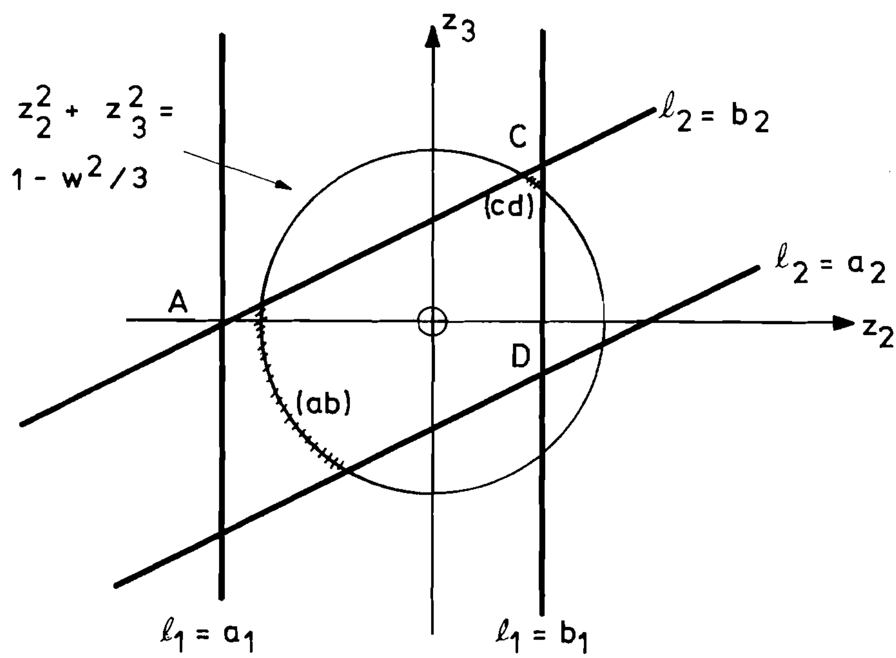

Figure 3.

For sufficiently small values of w, all vertices A, B, C, D, of the parallelogram lie inside the circle; as w is allowed to increase, the vertices move outwards. On consideration, it will be seen that four cases have to be considered, namely

$0 < w \leq 1$:	all vertices inside; no arcs.
$1 < w \leq \sqrt{2}$:	vertex B outside, others inside; arc (ab) exists; no arc (cd).
$\sqrt{2} < w \leq \sqrt{(8/3)}$:	all vertices outside, but circle not entirely included in the parallelogram; arcs (ab), (cd) both exist.
$\sqrt{(8/3)} < w \leq \sqrt{3}$:	all vertices outside; circle completely contained.

The corresponding values of p(w) are found to be:

$$p_1(w) = 0$$

$$p_2(w) = \frac{1}{\pi} \text{ arc sin}\left\{\sqrt{\frac{3}{8}}\,\frac{w - \surd(2 - w^2)}{\surd(3 - w^2)}\right\}$$

$$p_3(w) = \frac{1}{3} + \frac{2}{\pi} \text{ arc sin}\left\{\sqrt{\frac{3}{32}}\,\frac{w - \surd(8 - 3w^2)}{\surd(3 - w^2)}\right\}$$

$$p_4(w) = 1 \quad .$$

Taking into account the factor $f_{21}(w)$ of (45), we finally have the expression for $\pi_a(w)$:

$$\pi_a(w) = \begin{cases} 0 & , \quad 0 < w \le 1 \\ p_2(w)/2\surd 3 & , \quad 1 < w \le \surd 2 \\ p_3(w)/2\surd 3 & , \quad \surd 2 < w \le \surd(8/3) \\ 1/2\surd 3 & , \quad \surd(8/3) < w \le \surd 3 \\ 0 & , \quad \text{otherwise} \end{cases} \quad . \qquad (49)$$

Evaluation of $\pi_b(w)$

Taking $\pi_b(w)$ as $\pi_{24}(w)$, arguments of the above kind lead to the following:

$$\pi_{24}(w) = P(S_2 > S_1 > 0, S_2 > S_3 > 0 | S_2 = w) \; f_{24}(w) \quad . \qquad (50)$$

Here f_{24} is the pdf at w of $S_2 - S_4 = S_2$, which in accordance with (43) is uniformly distributed on (-2,2), so that

$$f_{24}(w) = \begin{cases} \frac{1}{4} & , \quad -2 < w < 2 \\ 0 & , \quad \text{otherwise} \end{cases} \quad .$$

The geometrical factor in (50) turns out to be

$$P\{0 < Z_1 < w/\surd 3, Z_1^2 + Z_2^2 > 1 - w^2/8 | Z_1 + Z_2\surd 2 = \surd 3w/2\} \qquad (51)$$

where (since $Z_1^2 + Z_2^2 + Z_3^2 = 1$) we have non-zero probabilities only when $Z_1^2 + Z_2^2 \leq 1$. We are therefore concerned with the probability content of that region of (Z_1, Z_2) space which lies in the intersection of the slab $0 < Z_1 < w/\sqrt{3}$, the oblique line $Z_1 + Z_2\sqrt{2} = w\sqrt{3}/2$, and the annulus

$$1 - w^2/8 \leq Z_1^2 + Z_2^2 \leq 1 \quad .$$

Figure 4 illustrates a typical situation, the hatched region being the intersection in question.

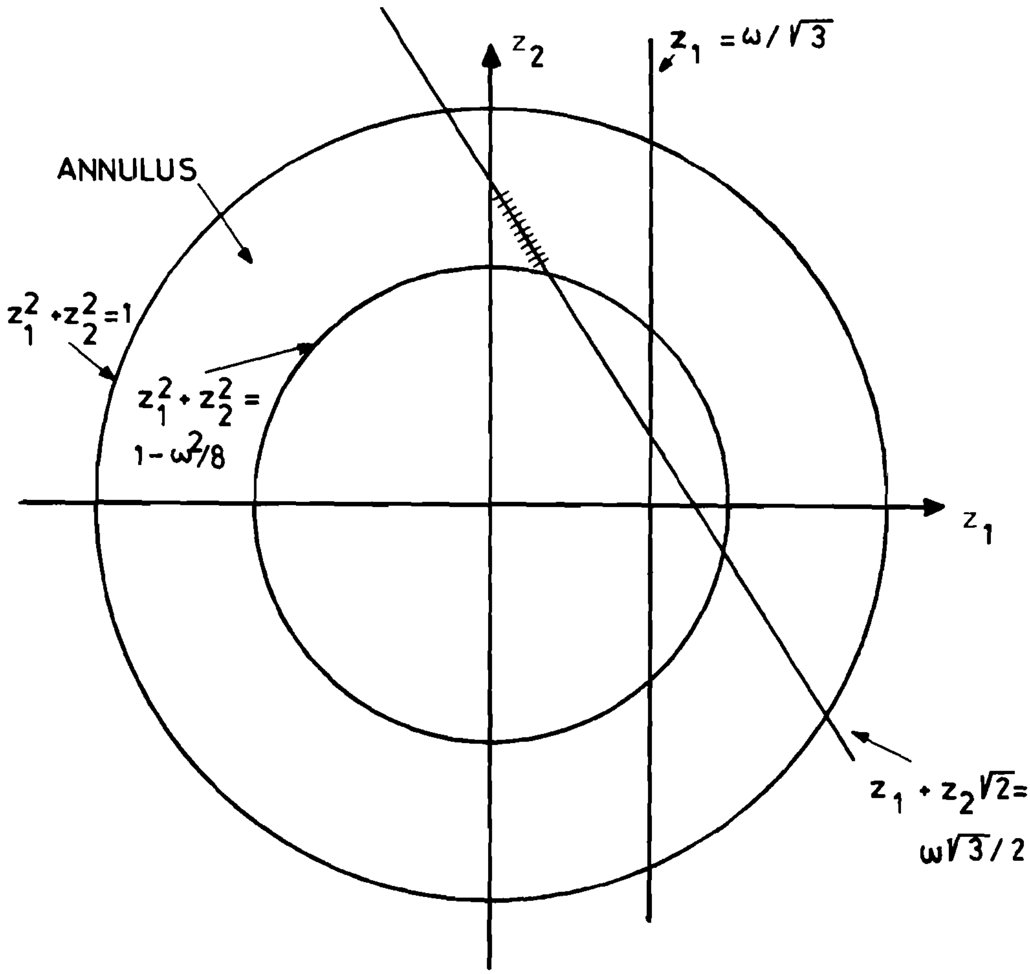

Figure 4.

By methods similar to those used for $\pi_a(w)$ we find

$$\pi_b(w) = \begin{cases} \frac{1}{2\pi}\left\{2 \text{ arc sin } \frac{w}{\sqrt{(8 - 2w^2)}} - \tfrac{1}{2}\pi\right\}, & \sqrt{2} < w \le \sqrt{(8/3)} \\ \tfrac{1}{4}, & \sqrt{(8/3)} < w \le \sqrt{2} \\ 0, & \text{otherwise} \end{cases} \qquad (52)$$

The pdf of the Range W_4

Combining these results in accordance with (41a) we finally obtain the following expressions for the pdf of the Hurst range for n = 4:

$$g_4(w) = \begin{cases} \frac{4}{\pi\sqrt{3}} \text{ arc sin}\left\{\sqrt{\left(\frac{3}{8}\right)}\, \frac{w - \sqrt{(2 - w^2)}}{\sqrt{(3 - w^2)}}\right\}, & 1 < w \le \sqrt{2} \\ \frac{4}{3\sqrt{3}} - 1 + \frac{8}{\pi\sqrt{3}} \text{ arc sin}\left\{\sqrt{\left(\frac{3}{32}\right)}\, \frac{w - \sqrt{(8 - 3w^2)}}{\sqrt{(3 - w^2)}}\right\} & \\ \qquad + \frac{4}{\pi} \text{ arc sin}\left\{\frac{w}{\sqrt{(8 - 2w^2)}}\right\}, & \sqrt{2} < w \le \sqrt{(8/3)} \\ \frac{4}{\sqrt{3}} + 1, & \sqrt{(8/3)} < w \le \sqrt{3} \\ 1, & \sqrt{3} < w \le 2 \\ 0, & \text{otherwise} \end{cases} \qquad (53)$$

This expression is a correctly normalized pdf ($\int_{-\infty}^{\infty} g_4(w)\, dw = 1$), and the expectation $E(W_4) = \tfrac{1}{2}(1 + \frac{4}{\sqrt{3}})$ agrees with the value given by (12) for n = 4. The variance is

$$\text{var}(W_4) = \frac{1}{3} + \frac{8}{3\pi} \text{ arc tan } \sqrt{2} + \frac{4}{3\pi}(1 + 2\sqrt{2}) - \frac{1}{12}(8\sqrt{3} + 19) \quad .$$

Figure 5 shows a graph of $g_4(w)$.

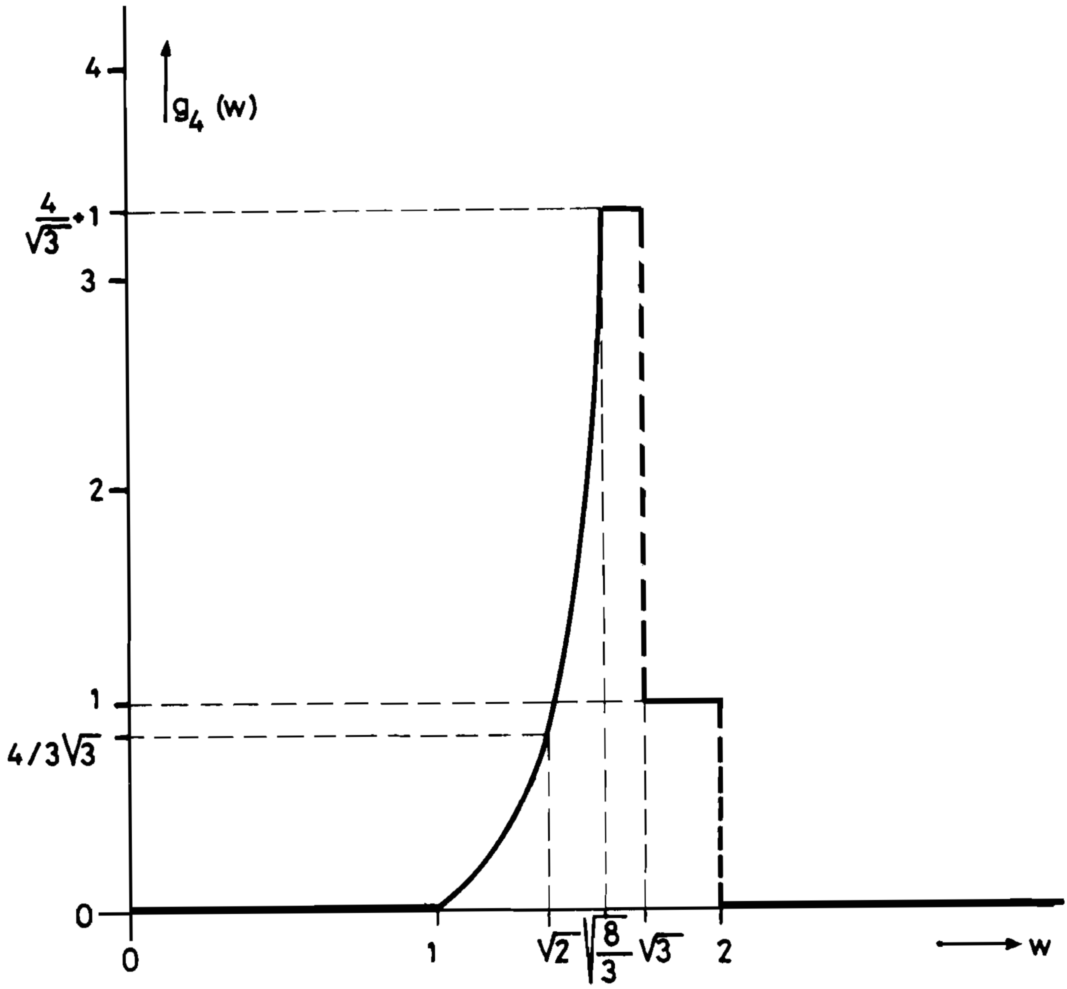
4
g_4 (w)
$\frac{4}{\sqrt{3}}+1$
3
2
1
$4/3\sqrt{3}$
0
0
1
$\sqrt{2}$
$\sqrt{\frac{8}{3}}$
$\sqrt{3}$
2
w

Figure 5.

Appendix

Derivation of the Joint Distribution (25)

The transformation to the "spherical" variables Z_j described in (18) is not only a geometrically attractive device: it leads to a simple derivation of the distributions in which we are interested.

We have

$$\zeta = B^{-1}\eta \quad ,$$

where

$$\zeta = (Z_1,\ldots,Z_{n-1}) \quad ,$$

B is a lower triangular matrix of order n-1 satisfying $BB' = nA^{-1}$, (A being defined in (17)), and

$$\begin{aligned} \eta &= (Y_1,Y_2,\ldots,Y_{n-1})' \\ &= (X_1 - \bar{X},X_2 - \bar{X},\ldots,X_{n-1} - \bar{X})' \; D_n \\ &= (T_1,T_2,\ldots,T_{n-1})/D_n \quad , \qquad \text{say} \\ &= \tau/D_n \quad . \end{aligned}$$

Since the T_j are linear functions of the X_i, with

$$\tau = C\xi$$

say, the vector τ is multivariate normal in n-1 variables. The matrix C is of order $(n-1) \times n$. It is the matrix defined by the first n-1 rows of the $(n \times n)$ matrix $I_n - \frac{1}{n} 1_n 1_n'$ (= N, say).

The dispersion matrix of τ is CC', which is the leading principal submatrix, of order $(n-1) \times (n-1)$ of NN'. But N is symmetric and idempotent, whence $NN' = N^2 = N$, and its leading principal submatrix, of order (n-1) × (n-1) is the matrix $I_{n-1} - \frac{1}{n} 1_{n-1}1'_{n-1}$, of order (n-1) × (n-1), namely the matrix A^{-1}. Hence

$$CC' = A^{-1} \quad .$$

It follows that τ is multivariate normal with

$$E(\tau) = 0 \text{ and } \mathcal{D}(\tau) = CC' = A^{-1} \quad .$$

We have expressed η as $\eta = \tau/D_n$. The quantity D_n itself may be expressed in the form

$$nD_n^2 = \sum_1^n (X_r - \bar{X})^2 = \sum_1^n T_j^2$$

$$= \sum_1^{n-1} T_j^2 + \left(\sum_1^{n-1} T_j\right)^2 \quad , \qquad \text{since } \sum_1^n T_j = 0$$

$$= \tau'\tau + (1'\tau)^2$$

$$= \tau'(I + 11')\,\tau = \tau'A\tau \quad .$$

Thus

$$\eta = \tau/\surd\{\tfrac{1}{n}\,\tau'A\tau\} \quad ,$$

and we may now express the vector ζ (the elements of which are the spherical variables $Z_1,\ldots,Z_{n-2}$) in terms of τ:

$$\zeta = B^{-1}\eta$$

$$= B^{-1}\tau/\surd\{\tau'(BB')^{-1}\,\tau\}$$

$$= \omega/\surd(\omega'\omega)$$

say, where

$$\omega = B^{-1}\tau \quad .$$

Since τ is multivariate normal, with dispersion matrix $\mathcal{D}(\tau) = A^{-1} = \frac{1}{n}(BB')$, it follows that ω is multivariate normal with dispersion matrix

$$\mathcal{D}(\omega) = B^{-1}\,\mathcal{D}(\tau)(B^{-1})' = \frac{1}{n}\,I \quad .$$

Hence the elements $W_1,\ldots,W_{n-1}$ if ω are mutually independent normal variables with common variance $1/n$, and

$$Z_j^2 = W_j^2 / \sum_1^{n-1} W_r^2 \quad ,$$

$$= \tfrac{1}{2}nW_j^2 / \sum_1^{n-1} \tfrac{1}{2}nW_j^2 \quad , \qquad j = 1,2,\ldots,n-1$$

where the variables $\frac{1}{2}nW_r^2$ are mutually independent gamma variables, with common exponent $\frac{1}{2}$.

By a standard theorem [3] the joint distribution of $Z_1^2, Z_2^2, \ldots, Z_{n-2}^2$ is a Dirichlet distribution. If we denote Z_j^2 by R_j, $j = 1,2,\ldots,n-2$, the joint pdf of $R_1,R_2,\ldots,R_{n-2}$ at $(r_1,r_2,\ldots,r_{n-2})$ is given by the theorem as

$$f(r_1,r_2,\ldots,r_{n-2}) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\pi^{\frac{1}{2}(n-1)}}\, r_1^{-\frac{1}{2}} r_2^{-\frac{1}{2}} \cdots r_{n-2}^{-\frac{1}{2}}$$

$$(1 - r_1 - r_2 - \cdots - r_{n-2})^{-\frac{1}{2}} \quad ,$$

$$r_1 > 0,\ldots,r_{n-2} > 0 \quad , \quad \sum r_j \leq 1 \quad ,$$

$$= f(r) \quad , \qquad \text{say}$$

and hence the joint pdf of $Z_1, Z_2, \ldots, Z_{n-2}$ at $(Z_1, Z_2, \ldots, Z_{n-2})$ is

$$\phi(z) = 2^{-(n-2)} f(r)\,|\partial(r)/\partial(z)|$$

$$= \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\pi^{\frac{1}{2}(n-1)}} (1 - z_1^2 - z_2^2 - \cdots - z_{n-2}^2)^{-\frac{1}{2}} \quad , \quad (\sum z_i^2 \le 1) \quad .$$

This is the result quoted in (25).

In fact the theorem referred to goes further: the joint distribution of every subset of size k of the set $(Z_1^2, Z_2^2, \ldots, Z_{n-2}^2)$, $k \le n-2$, is also of the Dirichlet form. For example, the joint pdf of $R_1 = Z_1^2, R_2 = Z_2^2, \ldots, R_k = Z_k^2$ is

$$g_k(r_1, r_2, \ldots, r_k) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\pi^{\frac{1}{2}k}\,\Gamma\{\frac{1}{2}(n-k-1)\}} r_1^{-\frac{1}{2}} \cdots r_k^{-\frac{1}{2}} (1 - r_1 - \cdots - r_k)^{\frac{1}{2}(n-k-3)} \quad .$$

In particular, the pdf of $R_1 = Z_1^2$ is

$$g_1(r_1) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\sqrt{\pi}\,\Gamma\{\frac{1}{2}(n-2)\}} r_1^{-\frac{1}{2}} (1 - r_1)^{\frac{1}{2}(n-4)} \quad , \quad (0 < r_1 \le 1) \quad .$$

Making the appropriate transformations we obtain the joint pdf of $(Z_1, Z_2, \ldots, Z_k)$, $k \le n-2$, as

$$\phi_k(z_1, \ldots, z_k) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\pi^{\frac{1}{2}k}\,\Gamma\{\frac{1}{2}(n-k-1)\}} (1 - z_1^2 - z_2^2 - \cdots - z_k^2)^{\frac{1}{2}(n-k-3)} \quad , \quad (\sum z_j^2 \le 1)$$

and, in particular, the pdf of Z_1 is

$$\phi_1(z_1) = \frac{\Gamma\{\frac{1}{2}(n-1)\}}{\sqrt{\pi}\ \Gamma\{\frac{1}{2}(n-2)\}} (1 - z_1^2)^{\frac{1}{2}(n-4)} \quad , \qquad z_1^2 \leq 1 \quad .$$

REFERENCES

[1] Hurst, H.E., Methods of Using Long-Term Storage in Reservoirs, *Proc. Inst. Civil Engrs.*, 5 (1956), 519-590.

[2] Anis, A.A., and E.H. Lloyd, The Expected Value of the Adjusted Rescaled Hurst Range of Independent Normal Summands, *Biometrika*, 63 (1976), 111-116.

[3] Wilks, S.S., *Mathematical Statistics*, Wiley, New York, 1962.