SIO REFERENCE SERIES

77-20

MOST PROBABLE EIGENVALUES OF A RANDOM COVARIANCE MATRIX

Rudolph W. Preisendorfer

University of California Scripps Institution of Oceanography

October 1977

UNIVERSITY OF CALIFORNIA, SAN DIEGO
SCRIPPS INSTITUTION OF OCEANOGRAPHY
LA JOLLA, CALIFORNIA

MOST PROBABLE EIGENVALUES OF A RANDOM COVARIANCE MATRIX

Rudolph W. Preisendorfer

This work was done while on leave to the Climate Research Group from the Pacific Marine Environmental Laboratory of the Environmental Research Laboratories, National Oceanic and Atmospheric Administration.

SIO Reference 77-20

October 1977

Abstract

When using principal component analysis to study causal or structural relations in a geophysical field (say sea surface temperature, or sea level pressure) the problem of the statistical significance of the relations arises. This leads one to consider the probability density function of the eigenvalues of the random covariance matrix Φ derived from a set of independent gaussian variates associated with the principal component analysis. The analytical intractibility of this probability density function stands in the way of a clean solution of the significance problem of the geophysical field. Therefore, numerical recourse via Monte Carlo simulations of Φ is indicated. These simulations produce two interesting phenomena: the degeneracy of the eigenvalues of the associated random covariance matrix Φ and the exponential decay (with index number) of its eigenvalues. These phenomena also arise in the principal component analysis of real data. In this note we study these phenomena directly by returning to the basic probability density function and deriving therefrom a polynomial whose roots are the most probable eigenvalues of the random covariance matrix Φ. In this way we provide a theoretical base for understanding the repeatedly observed phenomena of eigenvalue degeneracy and eigenvalue exponentiality which occur in geophysical applications of principal component analysis.

Table of Contents

0. Introduction

This study of the most probable eigenvalues of a random covariance matrix stems from the consideration of two phenomena arising in the numerical determination of the empirical eigenvectors associated with a real data set {F(t,x): t=1,2...,n; k=1,2,...,p} drawn from a population of F(t,x) values. These phenomena arise most often when the number p of spatial points x is specified and at the same time the sample size n of the data set falls either short of or much beyond p.

The first of these conditions occurs in the study, say, of geophysical data covering extensive regions of the earth's surface, but with a relatively sparse historical sample of such data, as for example in the meteorologic or oceanographic time series over the oceans of the world. The phenomenon encountered is as follows: For fixed p, as n decreases in value, some of the smaller eigenvalues of the data covariance matrix

$$\Phi(x',x) = (n-1)^{-1} \sum_{t=1}^{n} F(t,x') F(t,x)$$

appear to become degenerate, i.e., fall abruptly to values at or near zero. Since the sample size n is finite and contaminated with noise the numerical results are never quite decisive on this point. Practically, however, the generate eigenvalues* do seem to appear under wide conditions, i.e., when $n<p+3$; and when this inequality holds, there are apparently n-3 non degenerate larger eigenvalues and p-(n-3) degenerate smaller eigenvalues. We have studied this phenomenon in some detail (Preisendorfer and Barnett, 1977), using Monte Carlo simulations of the data field {F(t,x): t=1,...,n; x=1,...,p}. The fact that this form of degeneracy sets in is of practical importance; for in the empirical eigenvector analysis of, e.g., real climatological data, it is desirable to retain only those eigenvectors of Φ(x',x) that have the greatest eigenvalues. (Presumably these few largest eigenvalues relate in some way to the mechanics

---

* Strictly speaking, we must say 'most probable eigenvalues; see below.

of the system under study.) The largest eigenvalues are usually obtained by the process of eliminating the smaller ones according to some kind of 'stopping rule,' usually statistical in form (Preisendorfer and Barnett, 1977). This rule can be strengthened if it is known that a certain number of the smaller eigenvalues of $\Phi(x',x)$ are numerically degenerate. These degenerate eigenvalues can then be eliminated simply on algebraic (rather than statistical) grounds. In this study we shall indicate that the observed numerical degeneracy of the eigenvalues is related to an analogous property of the most probable eigenvalues of $\Phi(x',x)$.

The second phenomenon concerns the case when p is large and $n>p+3$; i.e., when the data sample is extensive spatially and free of the degeneracy encountered above. Meteorological data over continental U.S. and Europe areas are rich in high-n values because of extensive historical records. The phenomenon is most strikingly encountered however when one plots the eigenvalues λ_j of a randomly generated $\Phi(x',x)$ as a function of j. It appears that, after the first few of the larger λ_j are past, as one progresses along the j axis, the values of λ_j fall off exponentially; or equivalently, in the words of Craddock and Flood (1969), the λ_j decrease in the form of a 'geometrical progression.' If there is some theoretical basis to this empirical fact, then the stopping rule mentioned above has still further properties to help it estimate and discard the smaller eigenvalues most likely associated with noise rather than signal in the data. We shall adduce a simple physical argument applied to the theoretical equations governing the distribution of the most probable eigenvalues of a randomly generated Φ matrix to indicate the plausibility of the observed exponential fall-off of λ_j with j, for the smaller λ_j.

Before going into the theoretical analysis, we examine some evidence of these two phenomena. We consider first the degeneracy case. Figure 1 is a plot of $\ell n\ \overline{\ell}_j$ vs j for $j = 1,2,\ldots,9$, where $\overline{\ell}_j$ are the numerically-sampled estimates of the λ_j obtained as follows. We chose for p the value of 18 thereby defining 18 locations x, and at each of these eighteen points we built, with the help of random number generators, discrete time series $F(t,x)$ for $n = 11$ in such a way that the series were in the limit of infinite n, pairwise statistically independent and each of unit variance. (These, then, formed the ideal population of unit valued eigenvalues which the sampling would approximate.) From these we found the 18 x 18 matrix $\Phi(x',x)$ and its associated 18 eigenvectors and eigenvalues approximating the ideal λ_j. This kind of numerical experiment was done 100 times so that we obtained 100 samples of each of the eighteen eigenvectors. On averaging together the 100 realizations of the jth eigenvalue, we obtained each $\overline{\ell}_j$, j=1 ..., 18 along with its standard deviation s_j. In Figure 1 the plots are of the bounds $\ln(\overline{\ell}_j \pm s_j)$ and $\ln \overline{\ell}_j$ for the first nine values of j.

It can be seen that the more or less methodical decrease of $\overline{\ell}_j$ is abruptly accelerated at j = 9. Values of $\overline{\ell}_j$ for j = 10, and beyond to 18 even more abruptly decrease toward zero. The following Table I shows this behavior:

Table I $\overline{\ell}_j, s_j$ (p=18, n=11)

j	$\overline{\ell}_j$	s_j
1	5.649	0.6588
2	4.469	0.3407
3	3.584	0.3033
4	2.828	0.2762
5	2.246	0.2393
6	1.651	0.2476
7	1.184	0.2229
8	0.7546	0.1738
9	0.4329	0.1647
10	0.3549×10^{-10}	0.6817×10^{-11}

Another instance of this behavior is for p = 18, n = 6 as shown in Figure 2 and Table 2. The eigenvalues were generated in the same way as in the preceding case.

Now the degeneracy visibly sets in already at the fourth eigenvalue and decisively takes over in the fifth eigenvalue and beyond to the eighteenth.

Table 2 $\bar{\ell}_j$ s_j (p=18, n=6)

j	$\bar{\ell}_j$	s_j
1	17.41	1.761
2	12.54	1.134
3	9.176	1.128
4	5.846	1.234
5	0.3767×10^{-10}	0.7318×10^{-11}

It is from empirical data such as these that we formulated the rule enunciated above, i.e., that for n<p+3, there will be n-3 non degenerate and p-(n-3) degenerate most probable eigenvalues of the covariance matrix Φ.

We turn next to the evidence for the exponential (or geometric) decrease in eigenvalues λ_j with increasing j. In a Monte Carlo experiment of the kind already described, but now for p=18 and n=100 and with the basic experiment realized 100 times, we find the eighteen values $\bar{\ell}_j$ distributed on a semi log plot as shown in Figure 3. The values $\bar{\ell}_j$ and s_j are given in Table 3.

Table 3 $\bar{\ell}_j$, s_j, μ_j (p=18, n=100)

j	$\bar{\ell}_j$	s_j	μ_j
1	1.831	0.0985	1.802
2	1.655	0.0725	1.637
3	1.515	0.0593	1.506
4	1.405	0.0480	1.393
5	1.310	0.0412	1.291
6	1.225	0.0411	1.199
7	1.139	0.0398	1.114
8	1.063	0.0399	1.034
9	0.9926	0.0359	0.9583
10	0.9220	0.0355	0.8870
11	0.8476	0.0370	0.8190
12	0.7228	0.0367	0.7540
13	0.7158	0.0373	0.6912
14	0.6514	0.0364	0.6304
15	0.5935	0.0370	0.5708
16	0.5284	0.0367	0.5118
17	0.4646	0.0354	0.4518
18	0.3882	0.0379	0.3872

The essential linearity of the plot of the $\ln \bar{\ell}_j$ values is visually clear. Our experimentation with Monte Carlo simulation shows that the linearity becomes pronounced for the high values of n for a given p. An illustration of this is given elsewhere (Preisendorfer and Barnett, 1977) for the case of p = 36 and successive values of n = 38, 80, 100 and 200. Still further evidence for the exponential decay of the smaller eigenvalues is given in Figure 4, drawn from Craddock and Flood (1969). In this setting p = 130 and n = 1095.

The classical theory of eigenvalues of random covariance matrices, such as that given in Anderson (1958), suggests several approaches to the explanation of these phenomena. The basic equation for use in such a study would be that giving the joint probability densities $p(\lambda_1,\lambda_2,\ldots,\lambda_p)$ for the eigenvalues λ_j of the covariance matrix $A \equiv (n-1)\Phi$, namely:

$$p(\lambda_1,\lambda_2,\ldots,\lambda_p) = \frac{\pi^{\frac{1}{2}p}\left(\prod_{i=1}^{p}\lambda_i^{\alpha}\right)\exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\lambda_i\right\}\prod_{i<j}(\lambda_i-\lambda_j)}{2^{\frac{1}{2}p(n-1)}\prod_{i=1}^{p}\left\{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{p+1-i}{2}\right)\right\}} \qquad (0.1)$$

where

$$\alpha = \tfrac{1}{2}(n-p-2)$$

One approach to our solution would be to set p = 18 (say) and n = 100, to see if we could reproduce $\overline{\ell}_j$ in Table 3. An order of magnitude estimate can be made of the number of steps required to find that table's correspondents by numerically integrating (0.1) for the marginal probability density of λ_j. For example simply dividing the range of $\overline{\ell}_j$ values in Table 3 into seven parts, so as to get the crudest estimates via a discretized version of (0.1), we would have to perform about a hundred thousand steps for each of the seven values. To increase the accuracy, say to an m-part decomposition of the range of $\overline{\ell}_j$ (where m is now greater than 7), it can be shown* that, for a general number p (rather than 18), we would require

---

* Preisendorfer (1977)

(p+m-1)!/p!(m-1)! steps for each estimated value of $\bar{\ell}_j$. It can be seen that this number grows very fast with p and m. For example, in the case of p=36 cited above, if we let m = 7, then about five million steps are needed for the estimate of each $\bar{\ell}_j$.

In view of these sobering numerical facts, we have cast about for and found a more tractable route to the solution of our problem of explicating the two observed phenomena described above. The route leads not directly to the estimates of $\bar{\ell}_j$ but rather to somewhere sufficiently close to their vicinity so as to cast a sharp light on them. We have in mind the most probable eigenvalues associated with $A = (n-1)\Phi$, i.e., those eigenvalues of the random matrix A that maximize the probability density function $p(\lambda_1,\ldots,\lambda_p)$ in (0.1). It turns out that these most probable eigenvalues are accessible to relatively simple but exact analysis and may be represented as the roots of a certain polynomial which we now go on to derive.

1. The System of Algebraic Equations Governing the Most Probable Set of Eigenvalues

The set of λ's that maximizes $p(\lambda_1,\lambda_2,\ldots,\lambda_p)$ also maximizes $\ell n\, p(\lambda_1,\lambda_2,\ldots,\lambda_p)$, a quantity in this case more easily worked with. Thus, from (0.1) we find:

$$\begin{aligned}\ell n\, p = C &+ \alpha(\ell n\,\lambda_1 + \ell n\,\lambda_2 + \ldots + \ell n\,\lambda_p)\\ &- 1/2\,(\lambda_1 + \lambda_2 + \ldots + \lambda_p)\\ &+ \ell n\,(\lambda_1-\lambda_2) + \ell n\,(\lambda_1-\lambda_3) + \ell n\,(\lambda_1-\lambda_4) + \ldots + \ell n\,(\lambda_1-\lambda_p)\\ &\qquad + \ell n\,(\lambda_2-\lambda_3) + \ell n\,(\lambda_2-\lambda_4) + \ldots + \ell n\,(\lambda_2-\lambda_p)\\ &\qquad\qquad + \ell n\,(\lambda_3-\lambda_4) + \ldots + \ell n\,(\lambda_3-\lambda_p)\\ &\qquad\qquad + \ldots \qquad + \ell n\,(\lambda_{p-1}-\lambda_p)\end{aligned} \tag{1.1}$$

where C is the natural logarithm,

$$C = \ell n \left[\frac{\pi^{\frac{1}{2}p}}{2^{\frac{1}{2}p(n-1)} \prod_{i=1}^{p} \left\{ \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{p+1-i}{2}\right) \right\}} \right]$$

The necessary condition on the λ's for them to maximize $p(\lambda_1,\lambda_2,\ldots,\lambda_p)$ is

$$\frac{\partial p(\lambda_1,\lambda_2,\ldots,\lambda_p)}{\partial \lambda_j} = 0 \qquad j = 1,2,\ldots,p$$

Applying this criterion to (1.1) and rearranging the results, we find the requisite algebraic equations governing the most probable λ's:

$$\frac{\alpha}{\lambda_1} + \frac{1}{\lambda_1-\lambda_2} + \frac{1}{\lambda_1-\lambda_3} + \frac{1}{\lambda_1-\lambda_4} + \frac{1}{\lambda_1-\lambda_5} + \ldots + \frac{1}{\lambda_1-\lambda_p} = \tfrac{1}{2}$$

$$\frac{\alpha}{\lambda_2} + \frac{1}{\lambda_2-\lambda_1} + \frac{1}{\lambda_2-\lambda_3} + \frac{1}{\lambda_2-\lambda_4} + \frac{1}{\lambda_2-\lambda_5} + \ldots + \frac{1}{\lambda_2-\lambda_p} = \tfrac{1}{2}$$

$$\frac{\alpha}{\lambda_3} + \frac{1}{\lambda_3-\lambda_1} + \frac{1}{\lambda_3-\lambda_2} + \frac{1}{\lambda_3-\lambda_4} + \frac{1}{\lambda_3-\lambda_5} + \ldots + \frac{1}{\lambda_3-\lambda_p} = \tfrac{1}{2} \qquad (1.2)'$$

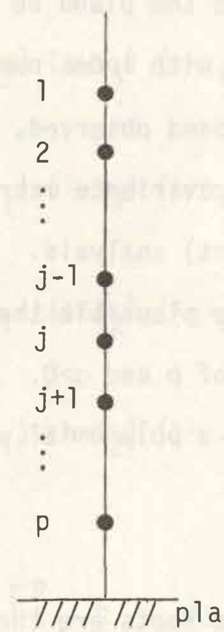
$$\vdots$$

$$\frac{\alpha}{\lambda_p} + \frac{1}{\lambda_p-\lambda_1} + \frac{1}{\lambda_p-\lambda_2} + \frac{1}{\lambda_p-\lambda_3} + \frac{1}{\lambda_p-\lambda_4} + \ldots + \frac{1}{\lambda_p-\lambda_{p-1}} = \tfrac{1}{2}$$

The structure of this set of equations is most interesting from both mathematical and physical points of view. Mathematically, they are fully symmetric: each equation centers on one of the eigenvalues λ_i, and uses all others as running variables. Thus the principal term in the λ_i equation is α/λ_i and the remaining p-1 terms on the left sides go methodically through the remaining p-1 λ's. The right sides of the equations are uniformly ½. Thus we can write the system (1.2)' as

$$\frac{\alpha}{\lambda_j} + \sum_{k=1}^{p}{}' \frac{1}{\lambda_j - \lambda_k} = \tfrac{1}{2} \qquad j = 1,2,\ldots,p. \tag{1.2}$$

where the prime denotes 'omit k=j in the sum'. When placed in this form, the summation term springs out as suggestive of a physical process; specifically the potential of a one-dimensional system of particles in a uniform gravitational field undergoing mutual repulsion by inverse square forces. We can thus imagine the p particles stacked vertically above an attracting plane



and if we rewrite (1.2) as

$$\frac{\alpha}{\lambda_j} + \sum_{k=j+1}^{p} \frac{1}{\lambda_j - \lambda_k} = \tfrac{1}{2} - \sum_{k=1}^{j-1} \frac{1}{\lambda_j - \lambda_k}$$

buoyant (or upward) potential on λ_j gravitational (or downward) potential on λ_j

we see that the j^{th} particle appears to be buoyed up by repulsion forces from the plane and the p-j particles below it, while it is pressed downward toward the plane by a constant gravitational potential of magnitude ½ and the repulsion of the j-1 particles above it. All particles, including the j^{th}, are in a static equilibrium of forces of mutual repulsion, and by 'gravitational' attraction. If we imagine the column of particles as but one of many standing side by side above the plane, they are seen to form an 'atmosphere of gas particles' which on the one hand is kept from collapsing in on itself by the mutually repulsive forces and on the other kept from going off into 'space' by the pervasive gravitational potential field of size ½. It would then seem plausible that, as p is increased to very large numbers, we would see the density of particles just above the plane be very great and such that it falls off exponentially with distance (i.e., with index number j) from the plane.* This phenomenon of eigenvalue spacing is indeed observed, as we noted above, in some studies of the distribution of eigenvalues of covariance matrices arising in empirical orthogonal function (principal component) analysis. These physical interpretations, aside from being amusing, serve to make plausible the existence of the most probable solutions λ_j of (1.2) for all choices of p and $\alpha>0$. We now show how the solution of (1.2) can be obtained via the roots of a polynomial which is naturally associated with the mathematical form of (1.2).

2. The Associated Polynominal whose Roots are the Most Probable Eigenvalues

The summation term in (1.2) is reminiscent of the logarithmic derivative of a product of p terms of the form $(\lambda_j-\lambda_k)$. Indeed, on tracing the route to (1.2) via (0.1) this observation is reinforced. Hence we consider the function

---

* We shall not pursue this suggestive analogy further here; but perhaps the present analysis could be extended to attain a derivation of the density of particles above the plane using the type of thermodynamic reasoning found in texts which derive the exponential lapse rate of an ideal-gas adiabatic atmosphere.

$$g(\lambda) = \prod_{i=1}^{p} (\lambda-\lambda_i),$$

which may be viewed as a pth order polynomial in λ, and form the derivative of $g(\lambda)$:

$$g'(\lambda) = \sum_{j=1}^{p} \prod_{k\neq j}^{p} (\lambda-\lambda_k)$$

For example, if p = 4, then

$$g(\lambda) = (\lambda-\lambda_1)\ (\lambda-\lambda_2)\ (\lambda-\lambda_3)\ (\lambda-\lambda_4)$$

and

$$\begin{aligned} g'(\lambda) &= (\lambda-\lambda_2)\ (\lambda-\lambda_3)\ (\lambda-\lambda_4) \\ &+ (\lambda-\lambda_1)\ (\lambda-\lambda_3)\ (\lambda-\lambda_4) \\ &+ (\lambda-\lambda_1)\ (\lambda-\lambda_2)\ (\lambda-\lambda_4) \\ &+ (\lambda-\lambda_1)\ (\lambda-\lambda_2)\ (\lambda-\lambda_3) \end{aligned}$$

Moreover,

$$\begin{aligned} g''(\lambda) &= \sum_{j=1}^{p} \left[\sum_{\ell=1}^{p} \prod_{k\neq j,\ell} (\lambda-\lambda_k) \right] \\ &= 2 \sum_{j<\ell} \prod_{k\neq j,\ell} (\lambda-\lambda_k) \end{aligned}$$

For example, for the case p = 4,

$$g''(\lambda) = (\lambda-\lambda_3)(\lambda-\lambda_4) + (\lambda-\lambda_2)(\lambda-\lambda_4) + (\lambda-\lambda_2)(\lambda-\lambda_3)$$

$$+ (\lambda-\lambda_3)(\lambda-\lambda_4) + (\lambda-\lambda_1)(\lambda-\lambda_4) + (\lambda-\lambda_1)(\lambda-\lambda_3)$$

$$+ (\lambda-\lambda_2)(\lambda-\lambda_4) + (\lambda-\lambda_1)(\lambda-\lambda_4) + (\lambda-\lambda_1)(\lambda-\lambda_2)$$

$$+ (\lambda-\lambda_2)(\lambda-\lambda_3) + (\lambda-\lambda_1)(\lambda-\lambda_3) + (\lambda-\lambda_1)(\lambda-\lambda_2)$$

i.e.,

$$\tfrac{1}{2}g''(\lambda) = (\lambda-\lambda_3)(\lambda-\lambda_4) + (\lambda-\lambda_2)(\lambda-\lambda_4) + (\lambda-\lambda_2)(\lambda-\lambda_3)$$

$$+ (\lambda-\lambda_1)(\lambda-\lambda_4) + (\lambda-\lambda_1)(\lambda-\lambda_3) + (\lambda-\lambda_1)(\lambda-\lambda_2)$$

Next, in both $g'(\lambda)$ and $g''(\lambda)$ set $\lambda = \lambda_j$. Then this suggests in general that

$$g'(\lambda_j) = \prod_{k\neq j}^{p} (\lambda_j-\lambda_k)$$

and

$$g''(\lambda_j) = 2\sum_{\ell=1}^{p} \prod_{k\neq j,\ell} (\lambda_j-\lambda_k)$$

whence

$$\frac{g''(\lambda_j)}{g'(\lambda_j)} = 2\sum_{k=1}^{p}{}' \left[\frac{1}{(\lambda_j-\lambda_k)}\right]$$

which may be verified in the special example above. Using this result in (1.2), we find

$$\frac{2\alpha}{\lambda_j} + \frac{g''(\lambda_j)}{g'(\lambda_j)} = 1$$

or

$$\lambda_j \, g''(\lambda_j) + (2\alpha-\lambda_j)g'(\lambda_j) = 0 .$$

This suggests that the polynomial

$$\lambda g''(\lambda) + (2\alpha-\lambda)g'(\lambda)$$

has the required λ_j, i.e., the most probable eigenvalues, as roots. Moreover, this polynomial is of order p. Hence it is identical to $g(\lambda)$ itself except perhaps for a scalar factor a. That is, for some a,

$$\lambda g''(\lambda) + (2\alpha-\lambda)g'(\lambda) = a\, g(\lambda) .$$

On matching the coefficients of λ^p on each side, we find that necessarily a = -p. In this way we arrive at a differential equation governing the polynomial $g(\lambda)$ whose roots are the required eigenvalues of $A = (n-1)\Phi$:

$$\boxed{\lambda g''(\lambda) + (2\alpha-\lambda)g'(\lambda) + p\, g(\lambda) = 0} \qquad (2.1)$$

We may solve this differential equation by writing out the polynomial form of $g(\lambda)$:

$$g(\lambda) = c_o + c_1\lambda + c_2\lambda^2 + \ldots + c_{p-1}\lambda^{p-1} + \lambda^p ,$$

substituting it in (2.1), and determining the necessary conditions on the c_j. For example, the terms involving the k^{th} power λ^k of λ are:

$$(k+1)\, k\, c_{k+1}\, \lambda^k \qquad \text{(via } \lambda g''(\lambda))$$

$$\left.\begin{array}{l} 2\alpha\, (k+1)\, c_{k+1}\, \lambda^k \\ -k\, c_k\, \lambda^k \end{array}\right\} \qquad \text{(via } (2\alpha-\lambda) g'(\lambda))$$

$$p\, c_k\, \lambda^k \qquad \text{(via } p\, g(\lambda)0$$

Since the right side of (2.1) may be viewed as a polynomial in the form $0 + 0\lambda + 0\lambda^2 + \ldots + 0\lambda^{p-1} + 0\lambda^p$, we conclude that the sum of the coefficients of λ^k displayed above is zero; which implies a necessary connection between c_{k+1} and c_k:

$$c_{k+1} = \frac{(k-p)}{(k+1)\,(k+2\alpha)}\, c_k \quad , \quad k = 0, 1, \ldots, p-1 \tag{2.2}$$

This is a recursion formula for the coefficients c_k. For if we fix c_0, all other coefficients follow at once. Leaving c_0 arbitrary for the moment and building up $c_1, c_2, \ldots, c_k$ recursively via (2.2) we find

$$c_k = \frac{p!}{k!\,(p-k)!} \cdot \frac{(-1)^k}{\prod_{j=0}^{k-1} (j+2\alpha)} \cdot c_0$$

Observe that $c_\ell=0$ for $\ell>p$, owing to the presence of the factor $(k-p)$ in (2.2).

One final adjustment is needed in our quest for the most probable eigenvalues. We recall that the λ_j are eigenvalues of a matrix A of the form $(n-1)\Phi$. Since Φ in the limit of large n-1 has unit eigenvalues, we see that the λ_j in the limit, go to n-1 rather than 1. We therefore introduce a new set $\mu_1, \mu_2, \ldots, \mu_p$ of eigenvalues for the original covariance matrix Φ by writing

$$'\mu_j' \text{ for } \lambda_j/(n-1)$$

Returning to the polynomial $g(\lambda)$ and using the connection $\lambda=(n-1)\mu$, we find

$$g(\lambda) = \sum_{k=0}^{p} c_k\lambda^k = \sum_{p=0}^{p} c_k(n-1)^k\mu^k$$

$$= c_0\sum_{k=0}^{p} \frac{p!}{k!(p-k)!} \cdot \frac{(-1)^k(n-1)^k}{\prod_{j=0}^{k-1}(j+2\alpha)} \mu^k$$

$$\equiv h(\mu) \equiv \sum_{k=0}^{p} c'_k \mu^k$$

We now require that the coefficient of μ^p be unity, i.e.,

$$c_0 \cdot \frac{(-1)^p(n-1)^p}{\prod_{j=0}^{p-1}(j+2\alpha)} = 1$$

whence

$$c_0 = (-1)^{-p}(n-1)^{-p}\prod_{j=0}^{p-1}(j+2\alpha) ,$$

and so for k = 0, 1, ..., p, the coefficients of the $h(\mu)$ polynomial are:

$$n^k c_k = c'_k = \frac{p!}{k!(p-k)!} \cdot (-1)^{p-k} \cdot \frac{\prod_{j=k}^{p-1}(j+2\alpha)}{(n-1)^{p-k}}$$

Here we observed that $(-1)^{p-k} = (-1)^{k-p}$ and that $\prod_{j=p}^{p-1} \equiv 1$.

Thus the most probable eigenvalues of the sample covariance matrix Φ are the roots of the polynomial

$$h(\mu) = \sum_{k=0}^{p} \binom{p}{k} \mu^k (-1)^{p-k} \cdot \frac{\prod_{j=k}^{p-1} (j+2\alpha)}{(n-1)^{p-k}} \qquad (2.3)$$

$$2\alpha = n-p-2$$

As $n \to \infty$, we see that

$$\frac{\prod_{j=k}^{p-1} (j+2\alpha)}{(n-1)^{p-k}} \to 1$$

and so $h(\mu)$ has the limiting form $(\mu-1)^p$, i.e., a polynomial with p unit roots, as expected.

3. The Sum and Products of the most Probable Eigenvalues

As a check on numerical experiments in finding the μ_j, we observe that the sum of the roots of (2.3) must be

$$\sum_{j=1}^{p} \mu_j = p\left(\frac{n-3}{n-1}\right), \quad n \geq p+3 \qquad (2.4)$$

and the product of the roots must be:

$$\prod_{j=1}^{p} \mu_j = \frac{\prod_{j=0}^{p-1} (j+2\alpha)}{(n-1)^p}, \quad 2\alpha \geq 1 \qquad (2.5)$$

For computation purposes, the product quotients in (2.3), (2.5) may be written

$$\frac{\prod_{j=k}^{p-1} (j+2\alpha)}{(n-1)^{p-k}} = \frac{1}{(n-1)^{p-k}} \cdot \frac{(n-3)!}{[(n-3)-(p-k)]!} \qquad (2.6)$$

For proof we note that $2\alpha = n-p-2$, and so

$$\prod_{j=k}^{p-1} (j+2\alpha) = \prod_{j=k}^{p-1} (n-p+j-2) = (n-p+k-2)(n-p+k-1) \ldots (n-3)$$

$$= \frac{(n-3)!}{[(n-3)-(p-k)]!}$$

4. The Special Cases p=2, p=3

Returning to (2.3) we set p=2 and find

$$h(\mu) = \mu^2 - \frac{2(1+2\alpha)}{(n-1)}\mu + \frac{(2\alpha)(1+2\alpha)}{(n-1)^2}$$

$$= \mu^2 - \frac{2(n-3)}{(n-1)}\mu + \frac{(n-3)(n-4)}{(n-1)^2}$$

where $2\alpha = n-p-2 = n-4$

and so $2\alpha+1 = n-3$

The two non degenerate roots of $h(\mu)$ are therefore

$$\mu_\pm = \frac{(n-3)}{n-1}\{1 \pm (n-3)^{-\frac{1}{2}}\} , \qquad n > p+3 = 5$$

From this we can see directly that μ_+ and μ_- go to 1 as $n \to \infty$. Moreover, the sum of the roots is

$$\mu_+ + \mu_- = \frac{2(1+2\alpha)}{n-1} = 2\,\frac{(n-3)}{n-2}$$

and product is

$$\mu_+\mu_- = \frac{(2\alpha)(1+2\alpha)}{(n-1)^2} = \frac{(n-3)(n-4)}{(n-1)^2}$$

Returning to (2.3) once again and setting p=3, we find

$$h(\mu) = \mu^3 - 3\frac{(n-3)}{(n-1)}\mu^2 + 3\frac{(n-3)(n-4)}{(n-1)^2}\mu - \frac{(n-3)(n-4)(n-5)}{(n-1)^3}$$

The three non degenerate roots of $h(\mu)$ then are given by

$$\mu_1 = y_1 + \frac{(n-3)}{(n-1)}$$

$$\mu_2 = y_2 + \frac{(n-3)}{(n-1)}$$

$$\mu_3 = y_3 + \frac{(n-3)}{(n-1)}$$

where

$$y_1 = 2\,\frac{(n-3)^{\frac{1}{2}}}{(n-1)}\cos\phi$$

$$y_2 = 2\,\frac{(n-3)^{\frac{1}{2}}}{(n-1)}\cos\left(\phi + \frac{2\pi}{3}\right)$$

$$y_3 = 2\,\frac{(n-3)^{\frac{1}{2}}}{(n-1)}\cos\left(\phi + \frac{4\pi}{3}\right)$$

and where

$$\phi = \frac{1}{3}\cos^{-1}(n-3)^{-\frac{1}{2}} \quad , \quad n \geq p+3 = 6$$

Once again we see directly that $\mu_j \to 1$ as $n \to \infty$. The sum of the roots is

$$\mu_1+\mu_2+\mu_3 = \frac{3(n-3)}{n-1)}$$

and

$$\mu_1\mu_2\mu_3 = \frac{(n-3)(n-4)(n-5)}{(n-1)^3}$$

Both special cases exhibit some general properties of the eigenvalues: In the latter case since the sums are of the form $3(1-\frac{2}{n'})$, and products are of the form $(1-\frac{2}{n'})(1-\frac{3}{n'})(1-\frac{3}{n'})$, where $n' = n-1$, we see that the averages of the most probable eigenvalues, for any finite n, tend to be smaller than 1. They are, in short, asymetrically distributed around 1 with a total predilection toward being below 1 than above. A visual examination of (0.1) also suggests that, because of the exponential factors, there is an asymmetry of the pdf along each μ_j axis so as to make the average value of μ_j slightly greater than its most probable value. These matters are of interest in that, if the mean and most probable values of each μ_j are workably identical, we can obviate the need for the extensive calculations otherwise needed to obtain the mean values of the μ_j. We would then have in the form of μ_j a satisfactory and easily obtained measure of the mean of each λ_j. Further observations in this direction are given in §5.

5. Numerical Results

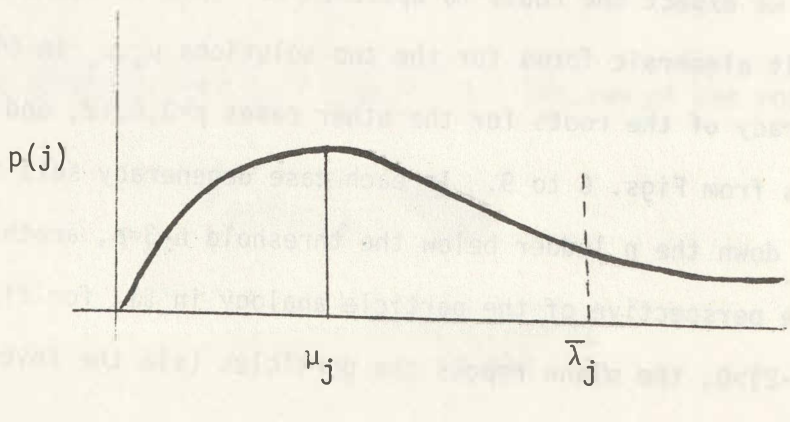
The roots of (2.3) were found for five values of p = 2,3,6,12 and 18 with sample size n running from 4 to 100. The results are depicted in Figures 5-9. Several features of these graphs are of interest.

First, we note the degeneracy of the roots whenever $n-3<p$. For example in the case of p=2, we have degeneracy when $n<5$, in this instance, when n=4. When n=3 both roots are zero, as we can see also from the special cases examined in §4. Moreover, as $n\to\infty$, we expect the roots to approach 1. This is borne out in Fig. 5 and in the explicit algebraic forms for the two solutions μ_+,μ_- in §4.

The degeneracy of the roots for the other cases p=3,6,12, and 18 are clearly seen in the graphs from Figs. 6 to 9. In each case degeneracy sets in when $n-3<p$, and for each step down the n ladder below the threshold $n-3=p$, another root collapses to zero. From the perspective of the particle analogy in §1, for fixed p, when we have $\alpha = \frac{1}{2}(n-p-2)>0$, the plane repels the particles (via the inverse square force which not only acts on particles but the plane and particle). When n decreases so that n-2-p = 0, the repulsion between the smallest λ_p and the plane vanishes and λ_p rests directly on the plane. When n decreases one more notch, so that n-2-p = -1, another particle, the next larger (i.e., λ_{p-1}) collapses onto the plane, and so on. Thus the 'atmosphere' of particles adjusts itself to the weight bearing down on the plane and the plane's ability to accommodate that weight.

Next, as $n \to \infty$, the λ_j converge toward 1, in every displayed case. As shown in the text just below (2.3), this is a general phenomenon, and harks back to the central law of the large samples in probability theory: The larger a sample we draw from the population with all $\lambda_j = 1$, the closer will be our estimates to those unit values.

If we plot the values $\mu_j(100)$ as a function of j, as given in Fig. 9, we would see that they fall off in essentially an exponential manner with j. Rather than plot these values, we have appended them to Table 3 as the third column of that table. Observe the closeness of the μ_j values, i.e., the most probable eigenvalues of Φ, to the empirically determined $\bar{\ell}_j$ values. It is this closeness of value that prompted recording our theoretical study of μ_j as practically acceptable substitutes for $\bar{\ell}_j$, providing n is large enough. For smaller n values (for a given p), degeneracy sets in and these μ_j and $\bar{\ell}_j$ must necessarily differ appreciably, since the marginal distributions of $p(\lambda, \ldots, \lambda_p)$ are highly skewed, as indicated in the sketch below.



This skewness manifests itself even in the closely placed $\overline{\ell}_j$ and μ_j values of Table 3, as may be seen by the necessary (skewness-induced) inequality between μ_j and $\overline{\ell}_j$ in the form $\overline{\ell}_j > \mu_j$.

6. Bibliographic Notes

It is possible to trace the eigenvalue degeneracy discussed above to an analogous degeneracy of certain volumes of linear subregions in p-space which arise in the derivation of (0.1) via the Wishart distribution (cf. e.g., Ch 7, Anderson, 1958). In this way the suggestive physical argument of §1 may be replaced by more complex considerations of geometry in euclidean p-space. Apparently one can see this degeneracy-difficulty coming by simply noting the possible singularities that may occur in $p(\lambda,...,\lambda_p)$, as given by (0.1), when $\alpha<0$. A similar phenomenon occurs in the theory of multiple regression analysis whenever $n<p$, for then the analysis is not even well defined, a situation which also requires further study (cf. p161 Dempster, 1969). It is interesting to note that the degeneracy problem is transparent for the case of $p=2$; for we generally need $n=5$ points in the plane to uniquely determine the scatter ellipsoid, and the relation $n=5=p+3$ is precisely at this threshold of uniqueness.

The phenomenon of algebraic degeneracy was encountered by Probert-Jones (1973), but apparently was not recognized as such. The phenomenon of exponential decay was also discussed in the same reference, but was not seen as being possibly a theorem in the asymptotic (large n) theory of eigenvalues of a random matrix. The latter theory, as exemplified by Anderson (1963), does not, however, broach this phenomenon. Apparently a rigorous proof of the asymptotic exponentiality of the λ_j for larger j, is still to be formulated. Our numerical experiments and the most probable eigenvalue theory of this note strongly indicate (cf. Table 3) the possibility of such a theorem.

Some interesting work on the distribution of eigenvalues of a random matrix has been done by atomic physicists in the past twenty-five years, or so. The work of Porter and Zweig (1960) is rich in numerical and theoretical examples of this subject. The study of Dyson (1953) has given rise to a considerable amount of work on random matrices and their eigenvalue distributions, as summarized in Mehta, (1967). The latter is a repository of many potentially useful techniques in random matrix analysis.* A particularly interesting approach to random-systems matrices which is reminiscent of invariant imbedding techniques (insofar as use of impedance-like quantities is made) may be found in Schmidt (1957). All of these studies give explicit expressions for the probability density function $p(\lambda_1,...,\lambda_p)$ or its close relatives. Yet the authors on the whole do not seem to be aware of the parallel work done in this area by the statisticians since the early thirties.

The population from which Φ (in §0) is drawn is such that ideally $\Phi = \sigma^2 I$, where I is the pxp identity matrix. An apparently definitive study of the distribution of the eigenvalues of a random matrix Φ drawn from a population which has a non identity (or nonspherical) covariance matrix was made by James (1960, 1964, 1966). Unfortunately, the results are inaccessible to simple numerical evaluation. There is also something still to be desired in the standard tests for sphericity of Φ (i.e., $= \sigma^2 I$), for the main references are Lawley (1956) and Mauchley (1940) and these are quite inadequate for geophysical purposes. Similarly work such as Chattopadhyay and Pillai (1973), like the James work, is inaccessible to practical workers. On the other hand, work such as Pillai (1956), while accessible, is inadequate because of the limited range of tabulated values. Work by Bartlett (1954) also gives practical tests for sphericity of the eigenvalue distribution, but before using it one should read the comments on the method as given in Kendall and Stuart (1976, p 299-300). Finally, we find the classical work by Girshick (1939) beautiful, but totally irrelevant to the robust needs of the general geophysicist, and climatologist in particular. Pretty theorems about the asymptotic properties of the

---

* For example, the derivation of the polynomial (2.3) was suggested by some ingenous arguments, originally due to Stieltjes, as summarized on p 191, Mehta (1967).

eigenvalues of a spherical Φ abound in Anderson (1958, 1963), but are inadequate to the geophysicist's needs, as explained in detail in Preisendorfer and Barnett (1977).

The largely negative findings of the preceding paragraph have been recorded principally to document how they have led us to experiment directly and numerically with the properties of the eigenvalues and eigenvectors of randomly generated covariance matrices. Our general conclusion is that one well designed numerical experiment, in this most difficult area of analysis, is worth more than the best of currently available theory.

7. Acknowledgments

The computations for the polynomial roots were done by Anthony Tubbs. The manuscript was typed by Grace Johnston. The problem of the most probable eigenvalues was developed in conversations with Tim Barnett. I am also grateful to Dr. Barnett for making available his computing facilities for this study. This work was partially supported by the Office for Climate Dynamics of the National Science Foundation under contract ATM75-18699, and as part of the NORPAX program which is jointly sponsored by National Science Foundation Office of International Decade of Ocean Exploration and Office of Naval Research. It was also partially supported by the Pacific Marine Environmental Laboratory of the Environmental Research Laboratories of the National Oceanic and Atmospheric Administration.

8. References

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, N.Y.

Anderson, T. W. (1963), 'Asymptotic Theory for Principal Component Analysis,' Ann. Math. Stat. 34, 122.

Bartlett, M. S. (1954), 'A Note on the Multiplying Factors for Various χ^2 Approximations,' J. R. Statist. Soc. 16B, 296.

Chattopadhyay, A. K., Pillai, K. C. S. (1973), 'Asymptotic Expansions for Distributions of Characteristic Roots when the Parameter Matrix has several Multiple Roots,' In 'Multivariate Analysis' Proc. Conf. on Multivariate Analysis, Dayton, Ohio, Ed. P. R. Krishnaiah, Academic Press, N.Y.

Craddock, J. M., Flood, C. R. (1969), 'Eigenvectors for Representing the 500 mb geopotential surface over the Northern Hemisphere,' Quart. J. Roy. Met. Soc. 95, 576.

Dempster, A. P. (1969), Elements of Continuous Multivariate Analysis, Addison-Wesley Pub. Co., Reading, Mass.

Dyson, F. (1953), 'The Dynamics of a Disordered Linear Chain.' Phys. Rev. 92, 1331.

Girshick, M. A. (1939), 'On the Sampling Theory of Roots of Determinantal Equations,' Ann. Math. Stat. 10, 203.

James, A. T. (1960), 'The Distribution of the Latent Roots of the Covariance Matrix,' Ann. Math. Stat. 31, 151.

James, A. T. (1964), 'Distributions of Matrix Variates and Latent Roots Derived from Normal Variates,' Ann. Math. Stat. 35, 475.

James, A. T. (1966), 'Inference on Latent Roots by Calculation of Hypergeometric Functions of Matrix Argument,' In 'Multivariate Analysis,' Proc. Conf. on Multivariate Analysis, Dayton, Ohio, Ed. P. R. Krishnaiah, Academic Press, N.Y.

Kendall, M., Stuart, A. (1976), The Advanced Theory of Statistics Volume 3 (3rd Ed.) Hafner Press, N.Y.

Lawley, D. N. (1956), 'Tests of Significance for the Latent Roots of Covariance and Correlation Matrices,' Biometrika 43, 128.

Mauchley, J. W. (1940), 'Significance Test for Sphericity of a Normal n-Variate Distribution,' Ann. Math. Stat. 11, 204.

Mehta, M. L. (1967), Random Matrices and the Statistical Theory of Energy Levels, Academic Press, N.Y.

Pillai, K. C. S. (1956), 'On the Distribution of the Largest or the Smallest Root of a Matrix in Multivariate Analysis,' Biometrika 43, 122.

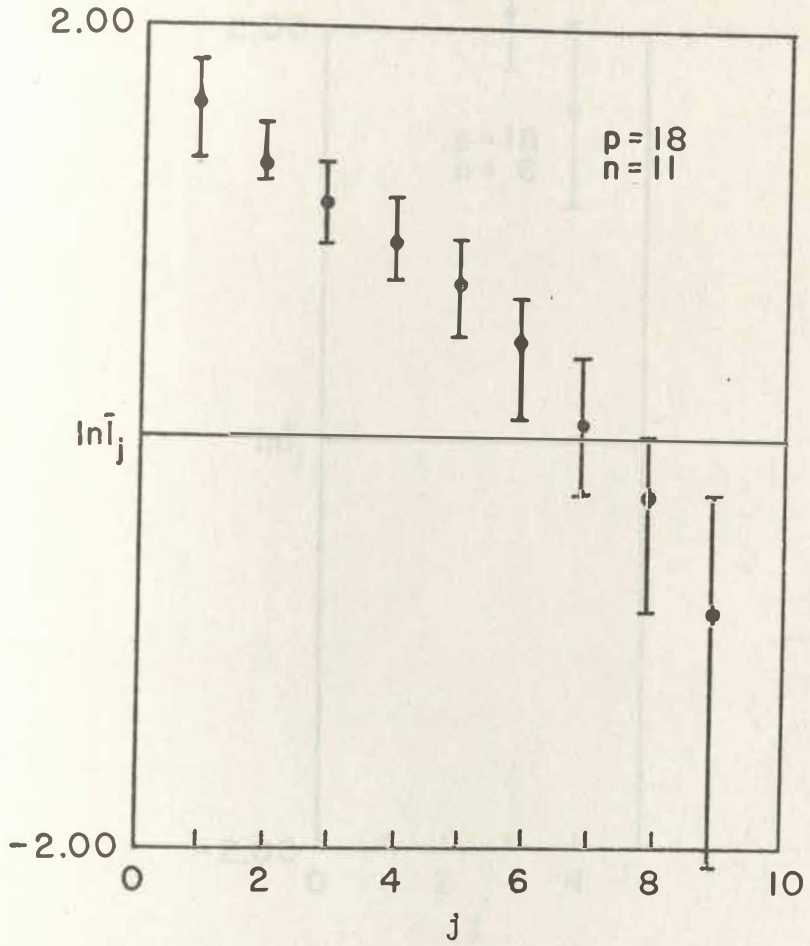
Porter, C., Zweig, N. (1960), 'Statistical Properties of Atomic and Nuclear Spectra,' Annales Academiae Scientarum Fennicae, Vol. VI. Physica, Helsinki.

Preisendorfer, R. W., Barnett, T. P. (1977), 'Significance Tests for Empirical Orthogonal Functions,' Fifth Conference on Probability and Statistics in Atmospheric Sciences, Nov. 15-18, 1977, Las Vegas, Nev. (Am. Met. Soc.).

Preisendorfer, R.W. (1977), 'Marginal Probability Distributions and Confidence Regions for Eigenvalues of Random Matrices,' in manuscript.

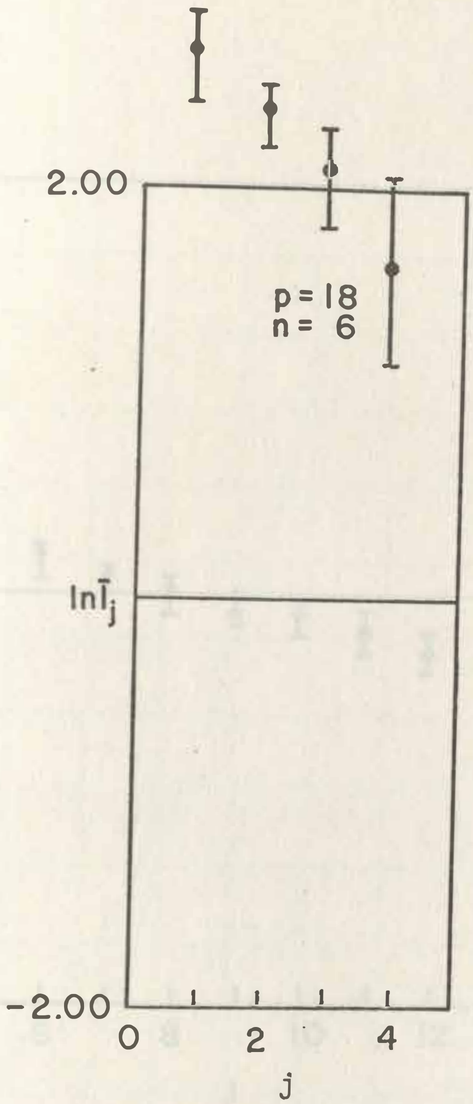
Probert-Jones, J. R. (1973), 'Orthogonal Pattern (Eigenvector) Analysis of Random and Partly Random Fields,' Third Conf. on Probability and Statistics in Atmospheric Sciences, June 19-22, 1973, Boulder, Col. (Am. Met. Soc.).

Schmidt, H. (1957), 'Disordered one-Dimensional Crystals,' Phys. Rev. 105, 425.



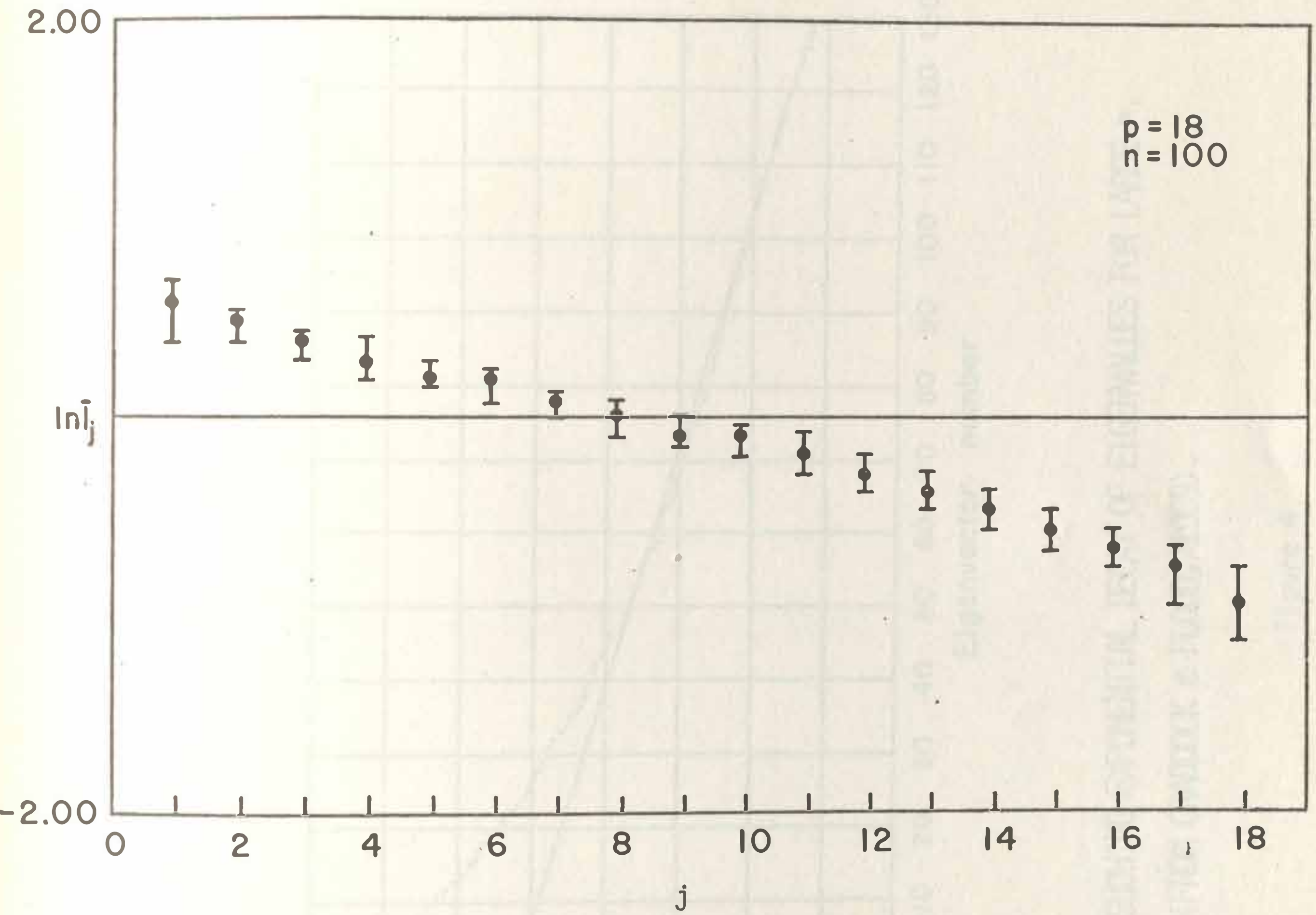
INDICATING DEGENERACY OF EIGENVALUES WHEN $n < p + 3$, GENERALLY THERE WILL BE $n - 3$ NON DEGENERATE MOST PROBABLE EIGENVALUES.

Figure 1



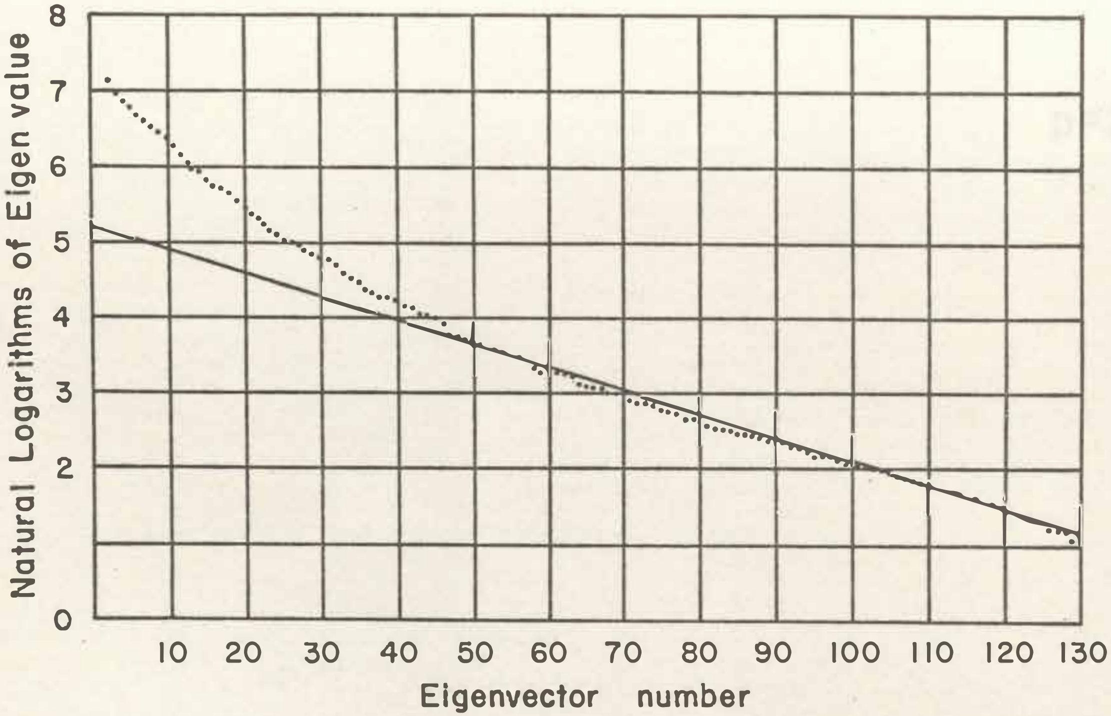
FOR P = 18 DATA POINTS AND SAMPLE SIZE N = 6 THERE ARE NOT 18 NONZERO AVERAGE EIGENVALUES, BUT ONLY FOUR, AS FOUND IN THIS MONTE CARLO EXPERIMENT OF 100 REALIZATIONS.

Figure 2



MONTE CARLO VERIFICATION OF EXPONENTIAL DECAY OF EIGENVALUES (100 REALIZATIONS).

Figure 3



SHOWING EXPONENTIAL DECAY OF EIGENVALUES FOR LARGE p.

(FROM CRADDOCK & FLOOD, 1969).

Figure 4

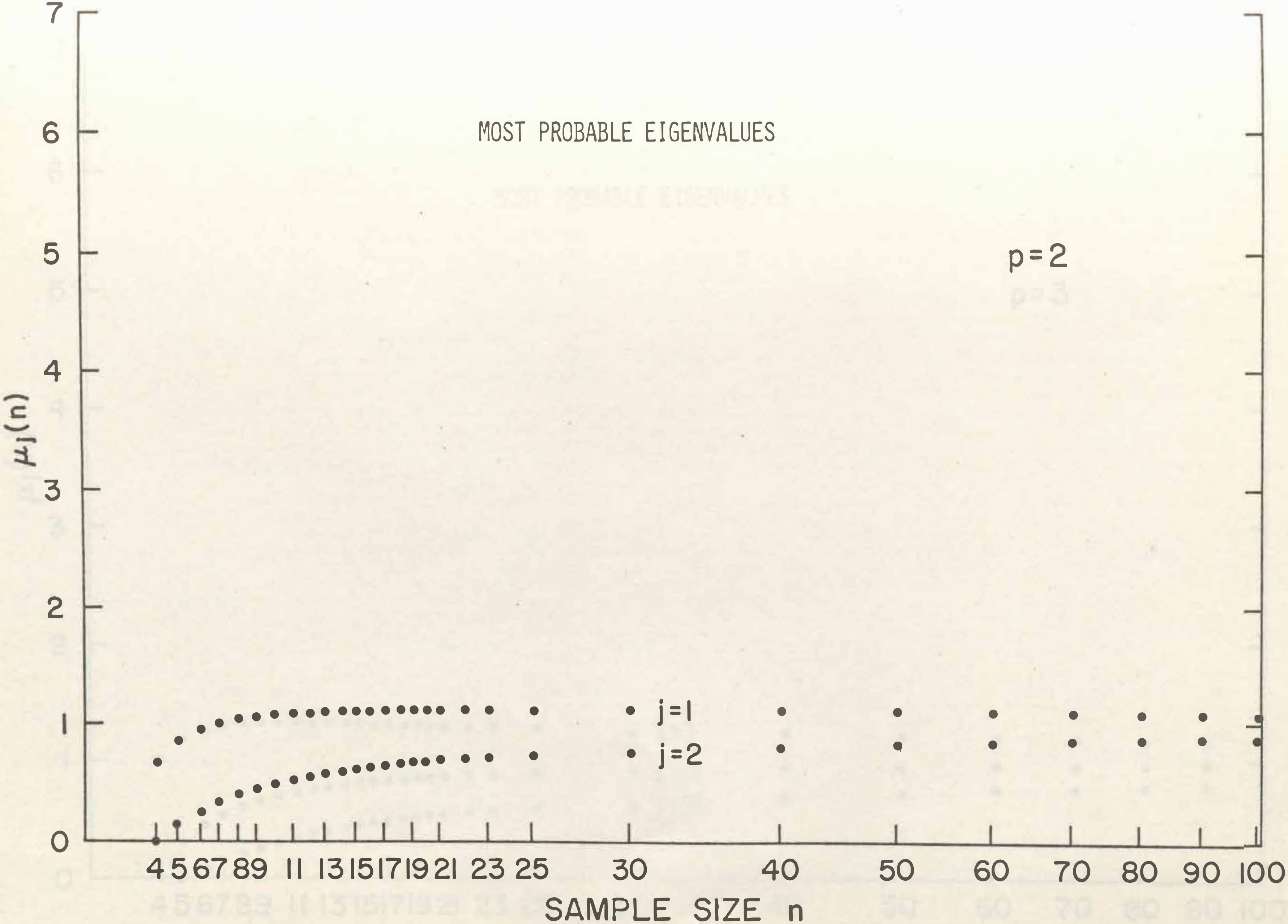

7
6
5
4
3
2
1
0
MOST PROBABLE EIGENVALUES
p=2
$\mu_j(n)$
j=1
j=2
4 5 6 7 8 9 11 13 15 17 19 21 23 25 30 40 50 60 70 80 90 100
SAMPLE SIZE n

Figure 5

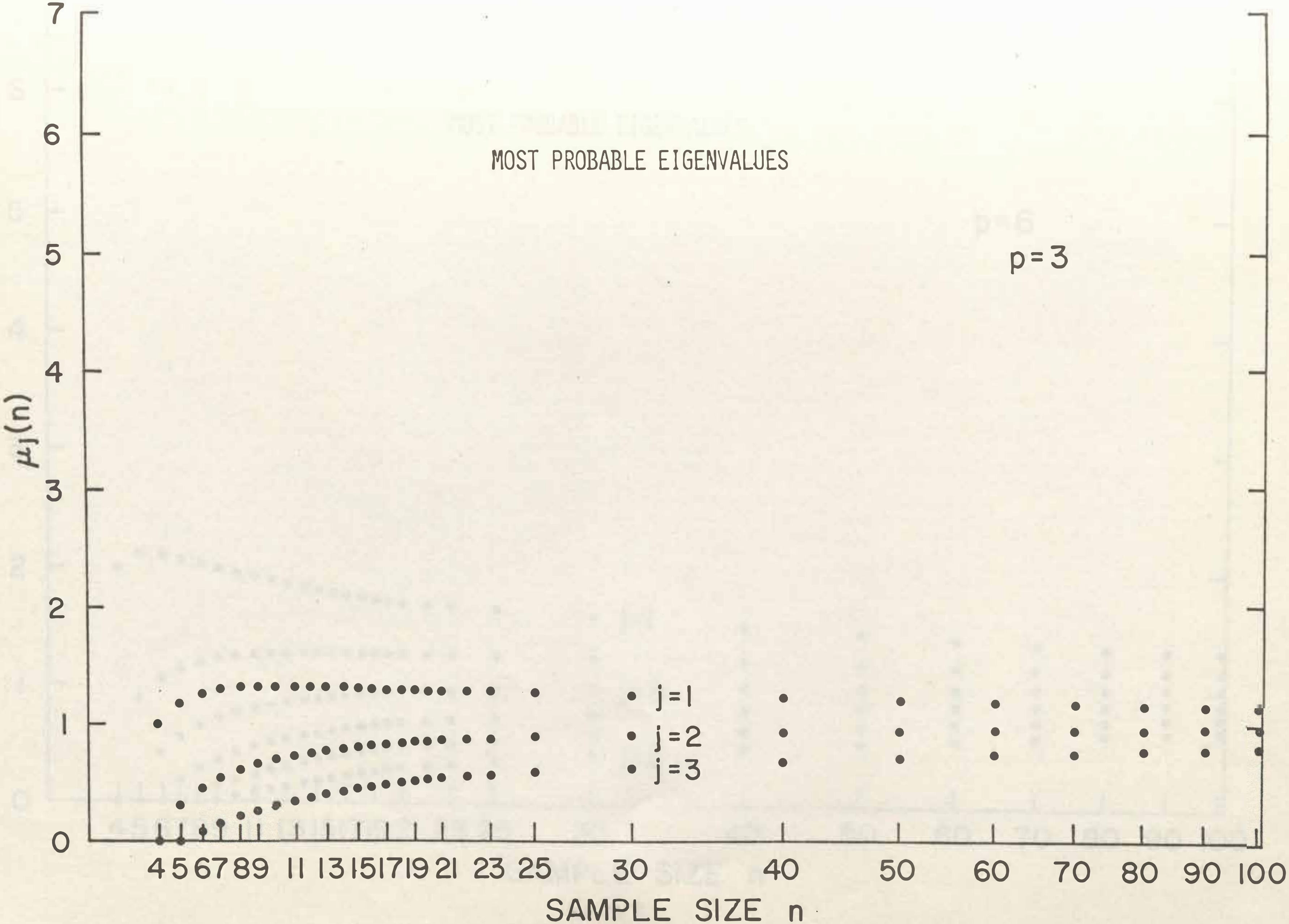
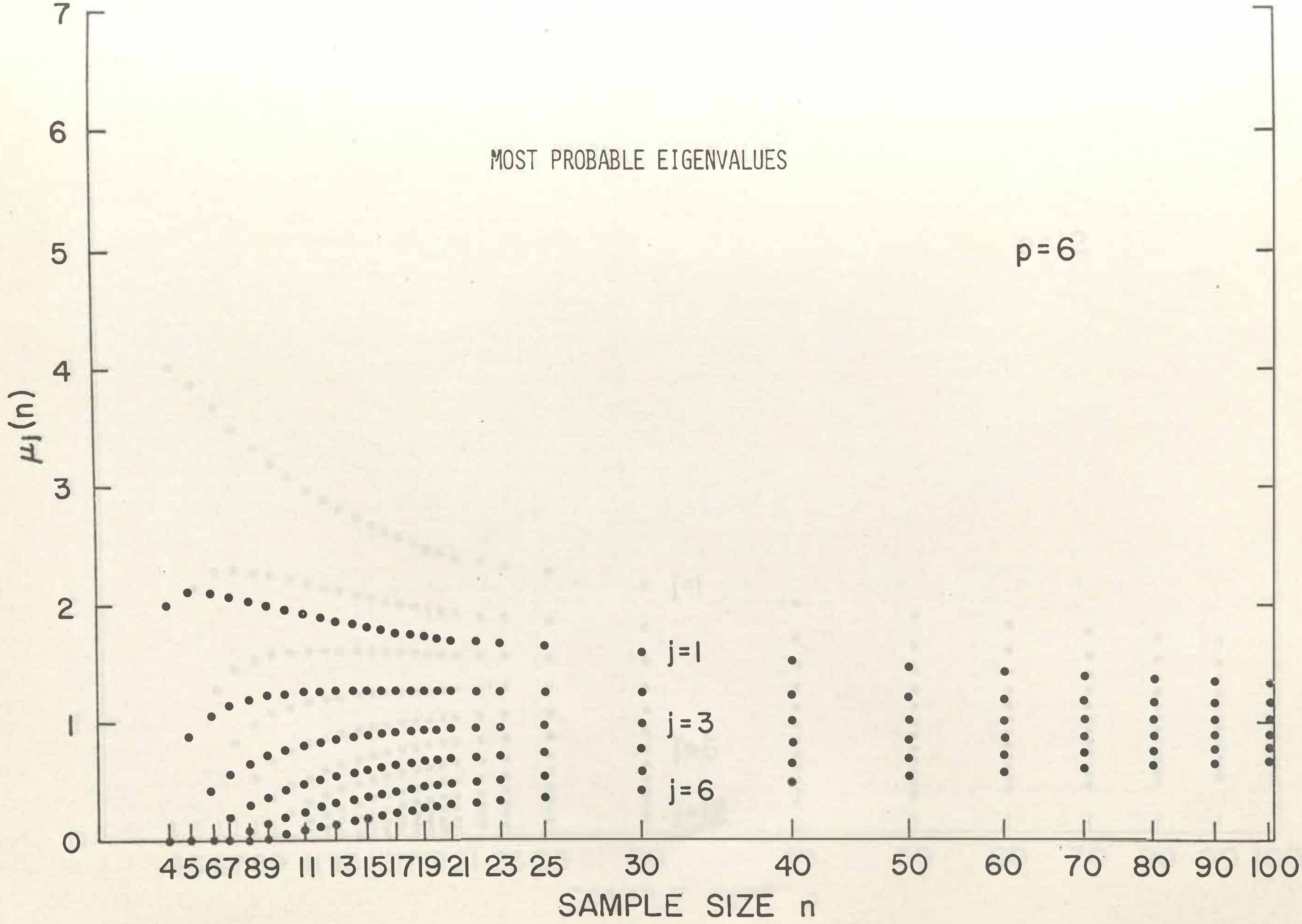

MOST PROBABLE EIGENVALUES
p = 3
$\mu_j(n)$
SAMPLE SIZE n
j = 1
j = 2
j = 3
0
1
2
3
4
5
6
7
4 5 6 7 8 9 11 13 15 17 19 21 23 25 30 40 50 60 70 80 90 100

Figure 6



MOST PROBABLE EIGENVALUES
p=6
j=1
j=3
j=6
$\mu_j(n)$
0
1
2
3
4
5
6
7
4 5 6 7 8 9 11 13 15 17 19 21 23 25 30 40 50 60 70 80 90 100
SAMPLE SIZE n

Figure 7

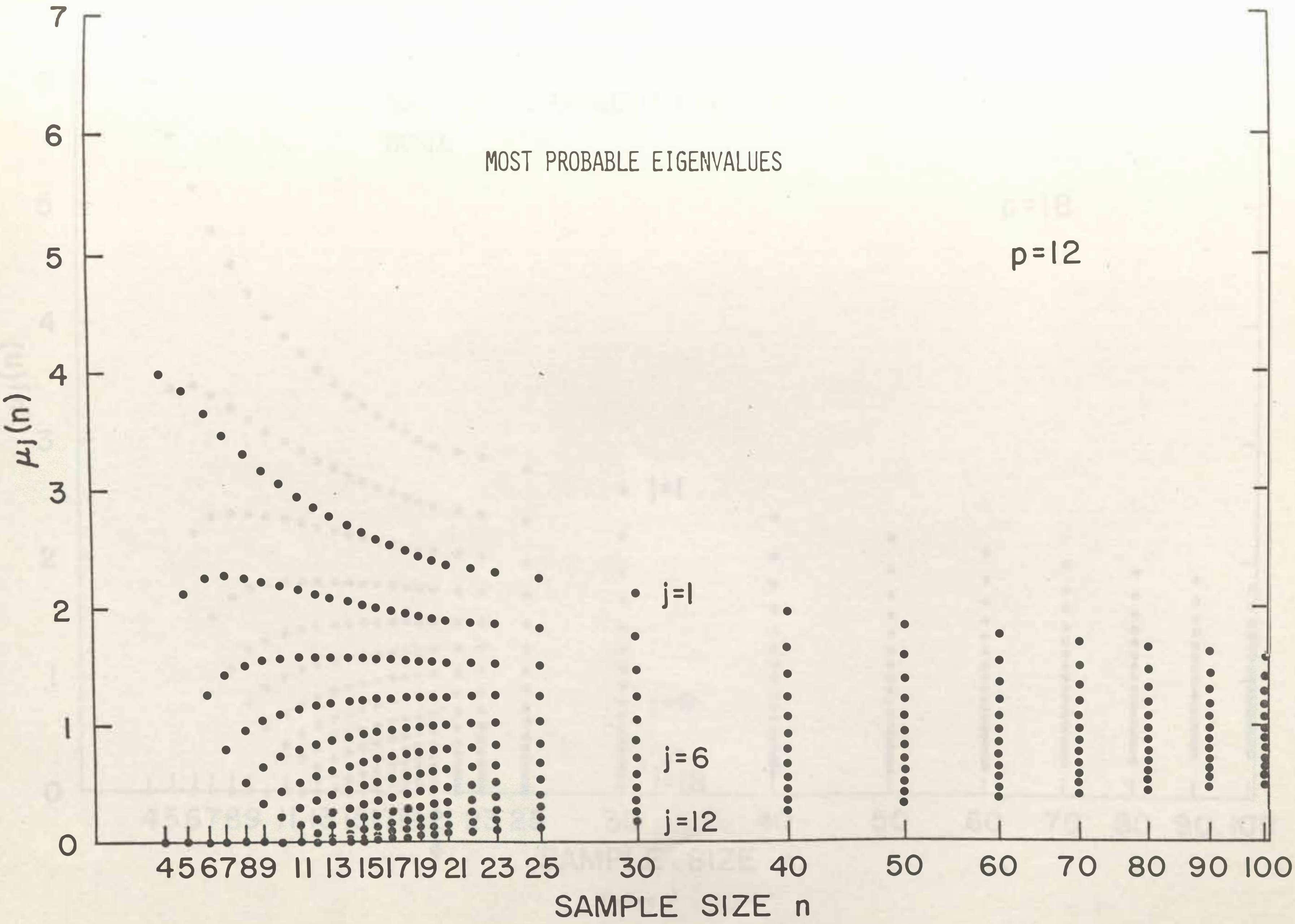

MOST PROBABLE EIGENVALUES
p=12
j=1
j=6
j=12
$\mu_j(n)$
0
1
2
3
4
5
6
7
4 5 6 7 8 9 11 13 15 17 19 21 23 25 30 40 50 60 70 80 90 100
SAMPLE SIZE n

Figure 8

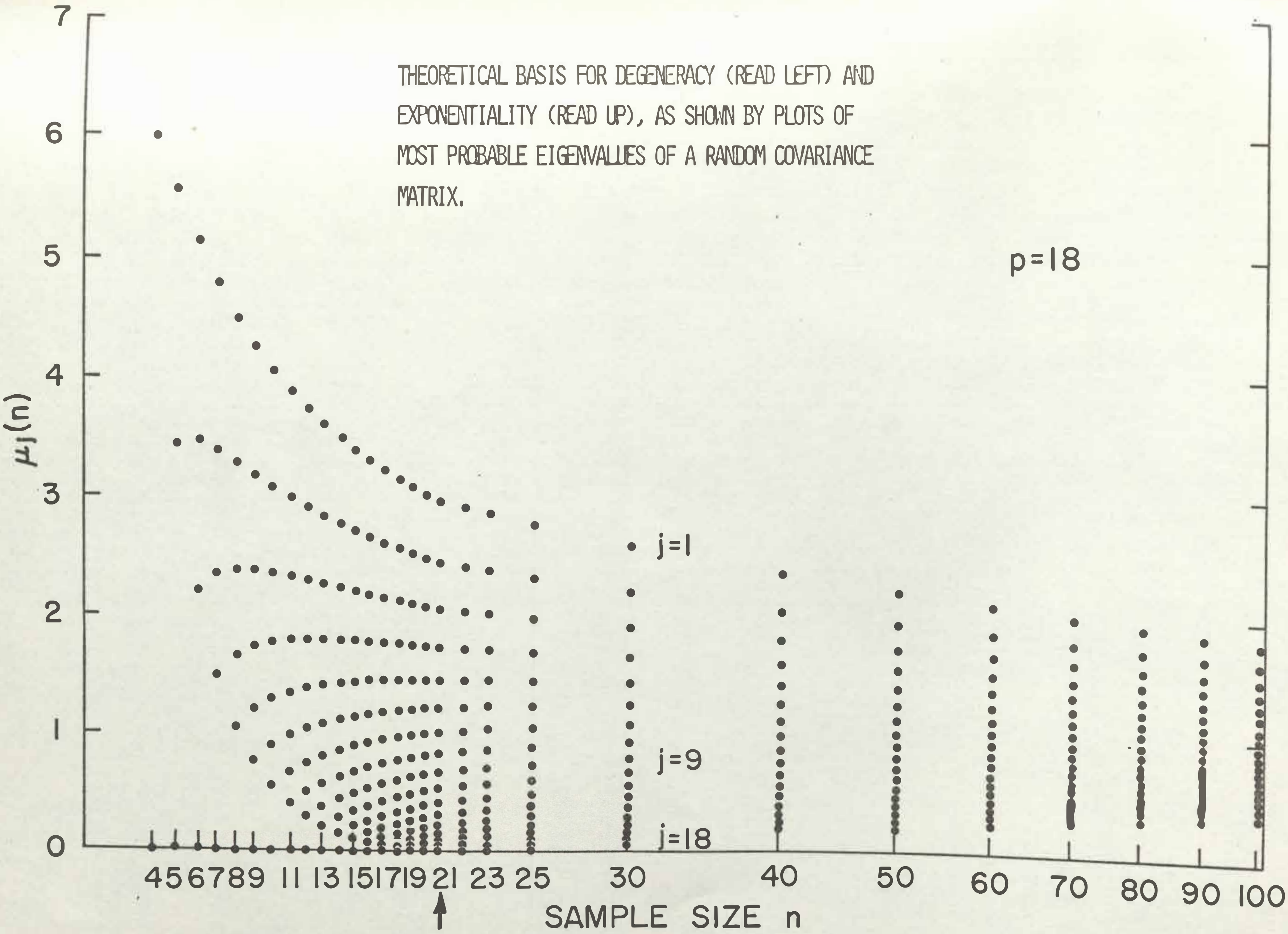

THEORETICAL BASIS FOR DEGENERACY (READ LEFT) AND EXPONENTIALITY (READ UP), AS SHOWN BY PLOTS OF MOST PROBABLE EIGENVALUES OF A RANDOM COVARIANCE MATRIX.
p=18
j=1
j=9
j=18
μj(n)
0 1 2 3 4 5 6 7
4 5 6 7 8 9 11 13 15 17 19 21 23 25 30 40 50 60 70 80 90 100
SAMPLE SIZE n

Figure 9