A Procedure for Improving the Estimate of the Mean for Weakly-Stationary Autoregressive Time Series

Richard Wiener

College of Engineering and Applied Science
University of Colorado at Colorado Springs
Colorado Springs, Colorado 80907

Jacob Rozmaryn

Department of Electrical Engineering
City College of New York
138 Street and Convent Avenue
New York, N.Y. 10031

Abstract

A procedure for improving the estimate of the mean (in the mean-square sense), compared to $\bar{x}$, for a weakly-stationary autoregressive time-series is presented. The improvement provided by the procedure is especially evident when only a relatively short sample time-series is available and the power spectrum of the underlying process is not flat (white). A detailed empirical evaluation of the new estimation procedure, based on a simulation analysis, is presented.

The new procedure is based on an estimator that would be BLUE (best linear unbiased estimator) if the order and coefficients of the underlying autoregressive process were known. The loss of efficiency caused by estimating the order and coefficients is shown to be small.

INTRODUCTION

Weakly-stationary autoregressive processes are used to model a wide variety of time-series [1]. A p^{th}-order autoregressive process, AR(p), may be represented by the difference equation $(X_t-\mu) + b_1(X_{t-1}-\mu) +\ldots+ b_p(X_{t-p}-\mu) = \varepsilon_t$, where X_t represents the state of the process at time t, where μ represents the mean of the process, where ε_t is an uncorrelated sequence of normally distributed random variables with mean zero and variance σ_e^2, and where, for stationarity, the roots of the characteristic equation $1 + b_1z^{-1} +\ldots+ b_pz^{-1} = 0$ must lie within the unit circle in the complex Z-plane.

The estimate of the mean, in addition to being an important feature in characterizing the process, is also used in the estimation of the autocovariance function. Thus, accurate estimates are quite important. Generally, the mean, μ, is estimated from a sample time-series $\{x_i:i = 1,N\}$ of length N using $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. In this paper, a procedure is developed for improving the estimate of μ (in the mean-square sense). The improvement obtained by using our procedure is especially evident when only a short sample time-series is available; indeed, in this case an improvement in the estimate of μ is most welcome since for short time-series samples, the estimate $\bar{x}$ generally has a large mean-square error. Furthermore, the degree of improvement afforded by our procedure is dependent on the location of the roots of the characteristic equation of the AR(p) within the unit circle in the complex Z-plane.

Although it is not assumed that the order p and the coefficients $b_1,\ldots,b_p$ are known for the AR(p) generating the sample time-series, our new estimation procedure for μ is based on an estimator that would be BLUE (best linear unbiased estimator) if the order p and coefficients $b_1,\ldots,b_p$ were known. Thus, the first step in this new procedure is to estimate p and $b_1,\ldots,b_p$. Then the estimator for μ that would be BLUE if $b_1,\ldots,b_p$ were known is employed using the estimated coefficients $a_1,\ldots,a_p$ in place of the

CH1437-3/79/0629-0637$00.75

1979 Winter Simulation Conference

required coefficients $b_1,\ldots,b_p$. Under conditions to be explored below, significant improvements in the estimate of μ (compared to $\bar{x}$) are possible.

After presenting the theoretical basis for our new estimation procedure, an empirical evaluation of the new estimator is presented using the results of an extensive simulation study. In particular, the conditions under which the greatest improvements may be obtained are explored.

THEORETICAL BASIS FOR NEW PROCEDURE

Our improved estimation procedure for μ is based on a linear estimator that would be BLUE if the order p and coefficients $b_1,\ldots,b_p$ were known. We establish first the following theorem:

Main Theorem: The BLUE estimator for the mean of a weakly-stationary AR(p) with coefficients $b_1,\ldots,b_p$ is $\hat{m} = k\underline{C}^T\underline{x}$, where

$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ represents the sample time-series, k is a scalar such that $k\underline{C}^T\underline{1} =$

$$1,\ \underline{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},\ \text{and } \underline{C} = \begin{bmatrix} 1 \\ 1+b_1 \\ \vdots \\ 1+b_1+\ldots+b_{p-1} \\ 1+b_1+\ldots+b_p \\ \vdots \\ 1+b_1+\ldots+b_p \\ 1+b_1+\ldots+b_{p-1} \\ \vdots \\ 1+b_1 \\ 1 \end{bmatrix}$$

The proof of the main theorem is expedited by first presenting some results concerning row-reversible matrices and reflective vectors.

Definition: An nxn matrix B is row-reversible, $B\varepsilon\phi^{RR}$ (ϕ^{RR} denotes the set of row-reversible matrices), if the elements of B, b_{ij}, satisfy, $b_{ij} = b_{n+1-i,\ n+1-j}$.

Lemma 1: All doubly symmetric matrices (matrices symmetric about the two main diagonals) are row-reversible.

Lemma 2: If nxn matrices D, E, and F satisfy the equation D = EF, and $D\varepsilon\phi^{RR}$, and $E\varepsilon\phi^{RR}$, then $F\varepsilon\phi^{RR}$. The proof is presented in Appendix A.

Lemma 3: If nxn matrix A is row-reversible then A^{-1} is row-reversible. The proof is presented in Appendix B.

Definition: A column vector $\underline{C}$ is reflective if the elements c_i of C satisfy, $c_i = c_{n+1-i}$.

Lemma 4: If R_n is the nxn autocovariance matrix of a weakly-stationary stochastic process, then $R^{-1}\underline{1}$ is a reflective vector. The proof is presented in Appendix C.

The proof of the main theorem stated above now follows. The weakly-stationary AR(p)*, $X_n + b_1X_{n-1} + \ldots + b_pX_{n-p} = \varepsilon_n$, may be represented in matrix form as

$$\underline{\varepsilon} = A\underline{x}, \tag{1}$$

where the nxn matrix A is given by

$$A = \begin{bmatrix} \not{Z}^{-1} & & & & & O & \\ b_p & b_{p-1} & \cdots & b_2 & b_1 & 1 & \\ & b_p & b_{p-1} & \cdots & b_2 & b_1 & 1 \\ & & \ddots & \ddots & & \ddots & \ddots \\ & & & b_p & b_{p-1} & \cdots & b_1 \quad 1 \end{bmatrix}$$

for $n \geq 2p$, where $\underline{\varepsilon} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, where $\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_2 \end{bmatrix}$,

*The mean of the process, μ, is assumed zero for notational convenience, without loss of generality.

and where Z is a pxp lower triangular matrix such that $ZZ^T = R_p$ (R_p is the pxp covariance matrix of the AR(p)). The Z matrix insures that the time-series $\underline{x}$ is covariance-stationary.

The autocovariance matrix of the vector $\underline{\varepsilon}$ is $I_n\sigma_e^2$, where I_n is the nxn identity matrix. Therefore,

$$I_n\sigma_e^2 = E(\underline{\varepsilon}\ \underline{\varepsilon}^T) = E(A\underline{x}\ \underline{x}^T A^T) = \qquad (2)$$
$$AE(\underline{x}\ \underline{x}^T)A^T = AR_nA^T$$

where $R_n = E(\underline{x}\ \underline{x}^T)$ is the nxn covariance matrix of the AR(p). From equation (2) it follows that

$$R_n^{-1} = \frac{1}{\sigma_e^2}A^TA. \qquad (3)$$

It is well known that for any weakly-stationary stochastic process, a BLUE estimator for the mean is [2],

$$\hat{m} = k\underline{C}^T\underline{X}, \qquad (4)$$

where

$$\underline{C} = cR_N^{-1}\underline{1}, \qquad (5)$$

c is an arbitrary scalar, and k is a scalar such that $k\underline{C}^T\underline{1} = 1$.

For a weakly-stationary AR(p) it follows from equations (3) and (5) that

$$C = \frac{c}{\sigma_e^2}A^TA\underline{1}. \qquad (6)$$

From the structure of the A matrix of the AR(p) it follows that

$$A\underline{1} = \begin{bmatrix} k_{11} \\ k_{21} \\ \vdots \\ k_{p1} \\ 1+b_1+\ldots+b_p \\ \vdots \\ 1+b_1+\ldots+b_p \end{bmatrix} = (1+b_1+\ldots+b_p)\begin{bmatrix} k_{12} \\ k_{22} \\ \vdots \\ k_{p2} \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad (7)$$

where the constants k_{i1};i = 1,p are the row sums of the Z^{-1} matrix and $k_{i2} = k_{i1}/(1 + b_1 +\ldots+ b_p)$.

Since

$$A^T = \begin{bmatrix} Z^{-1T} & & & \\ & b_p & & \\ & b_{p-1} & b_p & \\ & \vdots & b_{p-1} & \ddots \quad b_p \\ & b_1 & \vdots & \ddots \quad \vdots \\ & 1 & 1 & \vdots \\ & 0 & 0 & \ddots \quad 1 \\ & & & 0 \end{bmatrix}$$

it follows that

$$A^TA\underline{1} = \begin{bmatrix} k_{13} \\ \vdots \\ k_{p3} \\ 1+b_1+\ldots+b_p \\ \vdots \\ 1+b_1+\ldots+b_p \\ 1+b_1+\ldots+b_{p-1} \\ \vdots \\ 1+b_1 \\ 1 \end{bmatrix}(1+b_1+\ldots+b_p) \qquad (8)$$

where k_{i3};l = 1,p are determined from $(Z^{-1})^T$ and the constants k_{i2}.

Since $\frac{1}{\sigma_e^2}A^TA = R_n^{-1}$, it follows, from lemma 4, that $A^TA\underline{1}$ is a reflective vector.

Therefore,

$$A^T A\underline{1} = \begin{bmatrix} 1 \\ 1+b_1 \\ \vdots \\ 1+b_1+\ldots+b_{p-1} \\ 1+b_1+\ldots+b_p \\ \vdots \\ 1+b_1+\ldots+b_p \\ 1+b_1+\ldots+b_{p-1} \\ \vdots \\ 1+b_1 \\ 1 \end{bmatrix} (1+b_1+\ldots+b_p). \tag{9}$$

When the arbitrary scalar c, in equation (6), is chosen to be $c = \sigma_e^2/(1+b_1+\ldots+b_p)$, then from equations (4), (5), (6), and (9) it follows that the BLUE estimator, $\hat{m}$, is given by $\hat{m} = k\underline{C}^T\underline{x}$, where

$$\underline{C} = \begin{bmatrix} 1 \\ 1+b_1 \\ \vdots \\ 1+b_1+\ldots+b_{p-1} \\ 1+b_1+\ldots+b_p \\ \vdots \\ 1+b_1+\ldots+b_p \\ 1+b_1+\ldots+b_{p-1} \\ \vdots \\ 1+b_1 \\ 1 \end{bmatrix} \tag{10a}$$

and (10b)

$$k = \frac{1}{n(1+\ldots+b_p)-2pb_p-2(p-1)b_{p-1}-\ldots-2b_1}$$

NEW PROCEDURE FOR ESTIMATING THE MEAN

The main theorem stated and proved above provides the basis for the new procedure.

First, the order p and coefficients $b_1,\ldots,b_p$ of the AR(p) underlying the sample time-series are estimated by employing the method suggested by Durbin [3]. The estimate a_{rs} denotes the estimate of the s^{th} coefficient, b_s, when the order of AR(p) is r. From [3],

$$a_{rr} = -\frac{w_{r-1}}{v_{r-1}} \tag{11}$$

where

$$w_{r-1} = \sum_{s=0}^{r-1} a_{r-1,s}\hat{R}_{r-s}, \tag{12}$$

where

$$v_{r-1} = \sum_{s=0}^{r-1} a_{r-1,s}\hat{R}_s, \tag{13}$$

where

$$a_{rs} = 1; s = 0 \tag{14}$$

$$= a_{r-1,s} + (a_{r,r})(a_{r-1,r-s});\ r>1, s$$

$$= 1,\ldots,r-1,$$

where

$$a_{00} = 1$$

and where

$$R_s = \frac{1}{N}\sum_{i=1}^{N-s}(x_i-\bar{x})(x_{i+s}-\bar{x}).$$

The coefficients a_{rs} are obtained for values of r from 0 to n/2. Then, following [4], the residual variance $\hat{\sigma}_r^2 = \sum_{s=0}^{r} a_{rs}\hat{R}_s$ is computed for each r, and the value of r which minimizes $\hat{\sigma}_r^2$ is chosen as the estimate of the order of AR(p).

Based on equations (4) and (10), the estimate of the mean, μ, is given by

$$\hat{m}_1 = k\underline{C}^T\underline{x}, \tag{15}$$

where

$$\underline{C} = \begin{bmatrix} 1 \\ 1+a_{r1} \\ \vdots \\ 1+a_{r1}+\ldots+a_{r,r-1} \\ 1+a_{r1}+\ldots+a_{rr} \\ \vdots \\ 1+a_{r1}+\ldots+a_{rr} \\ 1+a_{r1}+\ldots+a_{r,r-1} \\ \vdots \\ 1+a_{r1} \\ 1 \end{bmatrix}, \qquad (16)$$

and k is computed so that $k\underline{C}^T\underline{1} = 1$.

EMPIRICAL EVALUATION OF THE NEW ESTIMATOR

The effectiveness of the new estimator for the mean, μ, was evaluated using simulation experiments. Each experiment involved the generation of 250 independent sample time-series, each of length 50, each initially in the steady-state (weakly-stationary), from an AR(p). The root configuration of the characteristic equation of the AR(p), $1+b_1z^{-1}+\ldots+b_pz^{-p} = 0$, was the feature used to classify each experiment. Experiments were conducted for different order AR(p)'s and, in particular, for different root configurations for a given order AR(p).

For each experiment 250 independent estimates of $\bar{x}$ and $\hat{m}_1$ were obtained. Each AR(p) sample throughout the experimentation was generated with $\mu = 0$ (with no loss of generality). The experimental mean-square errors $\sum_{i=1}^{250} (\bar{x}_i)^2$ and $\sum_{i=1}^{250} (\hat{m}_{1_i})^2$ were computed and compared for each experiment. The ratio of mean-square errors, $MSE(\hat{m}_1)/MSE(\bar{x})$, was used as a measure of effectiveness for the new estimator. In addition, the theoretical ratios of the mean-square errors for the BLUE estimator $\hat{m}$ and $\bar{x}$ (using the known coefficients $b_1,\ldots,b_p$)* were compared to the experimental ratios.

The results of the experimentation when p = 2, 4, and 6 are presented in Figures 1, 2, and 3, respectively. The experimental ratio of mean-square errors are shown, in parenthesis, above the theoretical ratio for each configuration of characteristic equation roots in the complex Z-plane.

It was occasionally found that the order of the process, r, estimated from the sample time-series of length 50 differed from the true order, p, of the AR(p) used to generate the sample. When the estimate of the mean, $\hat{m}_1$, obtained using the estimated order r and estimated coefficients $a_{r1},\ldots,a_{rr}$ was compared to the estimate $\hat{m}_1$ obtained by setting r equal to the known order p and using a new set of estimated coefficients $a_{p1},\ldots,a_{pp}$, our procedure (which uses $a_{r1},\ldots,a_{rr}$) most often yielded an estimate closer to the process mean than the estimate which used r = p. This may be explained by the observation that a particular sample time-series of relatively short length (e.g., 50) may display characteristics typical of an AR(r) of different order than the underlying process. Our procedure exploits the characteristics of the particular sample time-series and if the sample behaves like a representative from AR(r), the estimated coefficients $a_{r1},\ldots,a_{rr}$ are used to estimate μ.

RESULTS AND CONCLUSIONS

It is evident from the experimental results depicted in Figures 1, 2, and 3 that significant improvements in the estimation of the mean, μ, may be achieved using our procedure.

The potential degree of improvement afforded by our procedure may be seen by examining the theoretical ratios of mean-square error (obtained assuming the $b_1,\ldots,b_p$ are known) for different root configurations of the characteristic equation of the AR(p). Generally, the ratio of theoretical mean-square errors improves (becomes smaller) as the roots move radially outward (towards unit radius) and counterclockwise (towards 180 degrees). A root near the unit-circle produces a sharply defined peak in the power spectrum of the AR(p). As this root is moved counterclockwise towards 180 degrees, the frequency of the spectral peak increases. A root at the

*It is easily shown that $\frac{MSE(\hat{m})}{MSE(\bar{x})} = \frac{n^2}{\underline{1}^T R_n^{-1}\underline{1}\ \underline{1}^T R_n \underline{1}}$

center of the unit-circle corresponds to a flat (white) spectrum. Here, no improvement may be expected.

The experimental ratios of mean-square errors of the new estimator, $\hat{m}_1$, and the estimator $\bar{x}$, are seen to closely follow the theoretical ratios. Thus, the loss of efficiency caused by not knowing the order and coefficients of the underlying AR(p) is not very great.

Since our procedure may provide highly significant improvements in estimating μ when the spectrum of the underlying AR(p) possesses one or more sharply defined peaks, it is recommended that a screening procedure, based on the sample spectrum, be employed. For sample spectra that display narrow bandwidth peaks (preferably of higher frequency), a major improvement in estimating μ may be expected. When the sample spectrum resembles a flat (white) spectrum, our procedure will yield little or no benefit.

Appendix A

Proof of Lemma 2:

Since D = EF, it follows that

$$d_{ij} = \sum_{k=1}^{n} e_{ik} f_{kj}, \tag{A1}$$

where $\{e_{ik}\}$ and $\{f_{kj}\}$ are the elements of the matrices E and F.

It is useful to define

$$\begin{aligned} l &= n + 1 - i \\ m &= n + 1 - j \\ r &= n + 1 - k \end{aligned} \tag{A2}$$

Then

$$d_{lm} = \sum_{r=1}^{n} e_{lr} f_{rm}. \tag{A3}$$

Since $D \varepsilon \phi^{RR}$ and $E \varepsilon \phi^{RR}$, it follows that $d_{ij} = d_{lm}$ and $e_{ik} = e_{lr}$. Therefore,

$$d_{ij} = \sum_{k=1}^{n} e_{ik} f_{n+1-k,m}. \tag{A4}$$

But from equation A1,

$$d_{ij} = \sum_{k=1} e_{ik} f_{kj}.$$

Therefore,

$$f_{kj} = f_{n+1-k,m} = f_{n+1-k,n+1-j}$$

and $F \varepsilon \phi^{RR}$.

Appendix B

Proof of Lemma 3:

Since $AA^{-1} = I_n$ and $A \varepsilon \phi^{RR}$ and $I_n \varepsilon \phi^{RR}$, it follows from lemma 2 that $A^{-1} \varepsilon \phi^{RR}$.

Appendix C

Proof of Lemma 4:

R_n is a doubly symmetric matrix, and from lemma 1, $R_n \varepsilon \phi^{RR}$.

From lemma 3 it follows that $R_n^{-1} \varepsilon \phi^{RR}$. Then from the definition of a reflective vector, it follows that $R_n^{-1} \underline{1}$ is a reflective vector.

BIBLIOGRAPHY

1. Box, G.E.P., Jenkins, G.M., Time Series Analysis Forecasting and Control, Holden Day, 1970
2. Grenander, U., Szego, G., Toeplitz Forms and Their Applications, University of California Press, 1958
3. Durbin, J., "The Fitting of Time Series Models", Rev. Intern. Inst. Stat., Vol. 28, 1960, pp233-244
4. Jenkins, G.M., Watts, D.G., Spectral Analysis and Its Applications, Holden Day, 1968, pp197

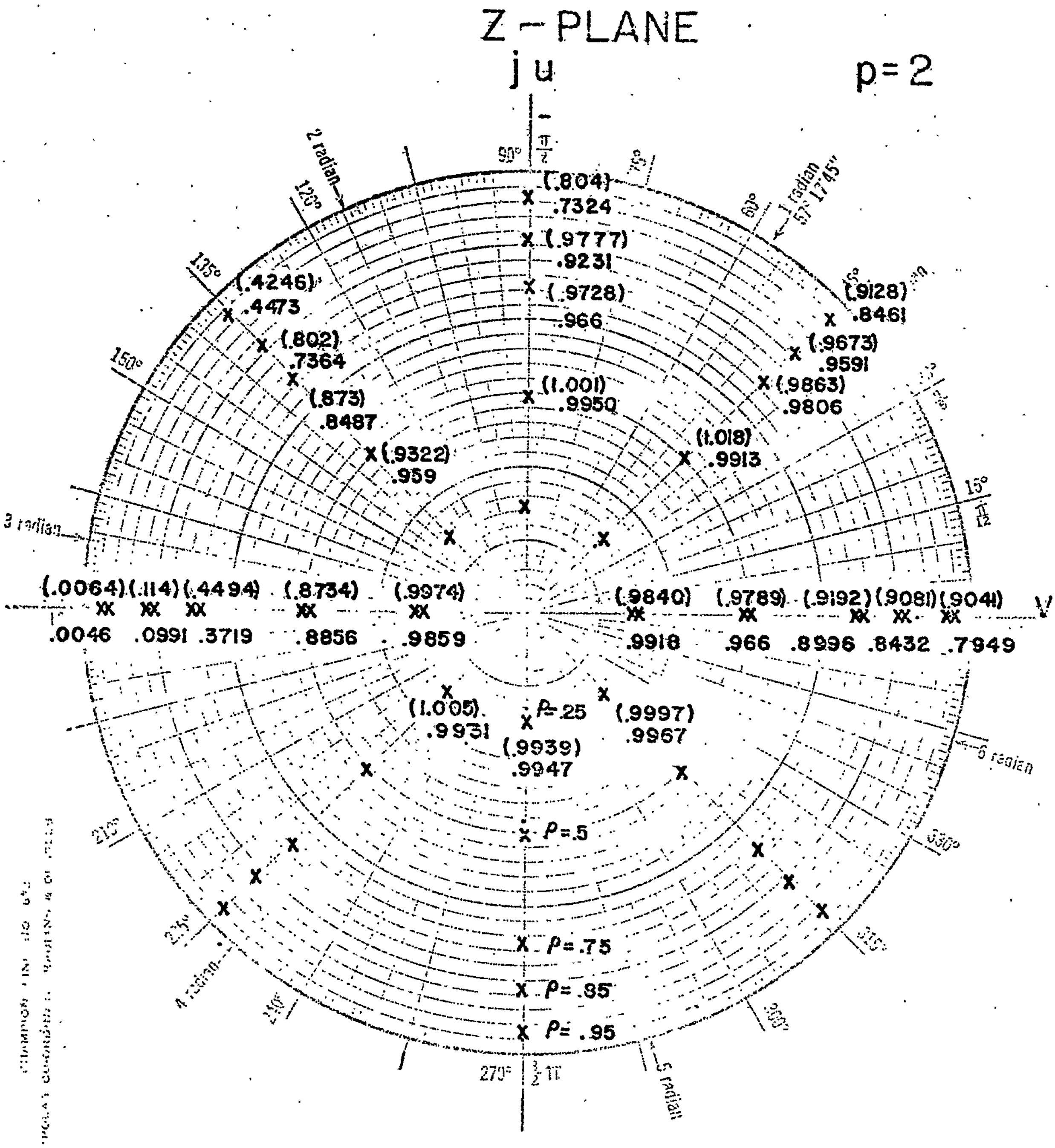
Z – PLANE
j u
p = 2
(.804)
.7324
(.9777)
.9231
(.9728)
.966
(1.001)
.9950
(.4246)
.4473
(.802)
.7364
(.873)
.8487
(.9322)
.959
(.9128)
.8461
(.9673)
.9591
(.9863)
.9806
(1.018)
.9913
(.0064)
.0046
(.114)
.0991
(.4494)
.3719
(.8734)
.8856
(.9974)
.9859
(.9840)
.9918
(.9789)
.966
(.9192)
.8996
(.9081)
.8432
(.9041)
.7949
(1.005)
.9931
(.9939)
.9947
(.9997)
.9967
ρ = .25
ρ = .5
ρ = .75
ρ = .85
ρ = .95
v

FIG. 1

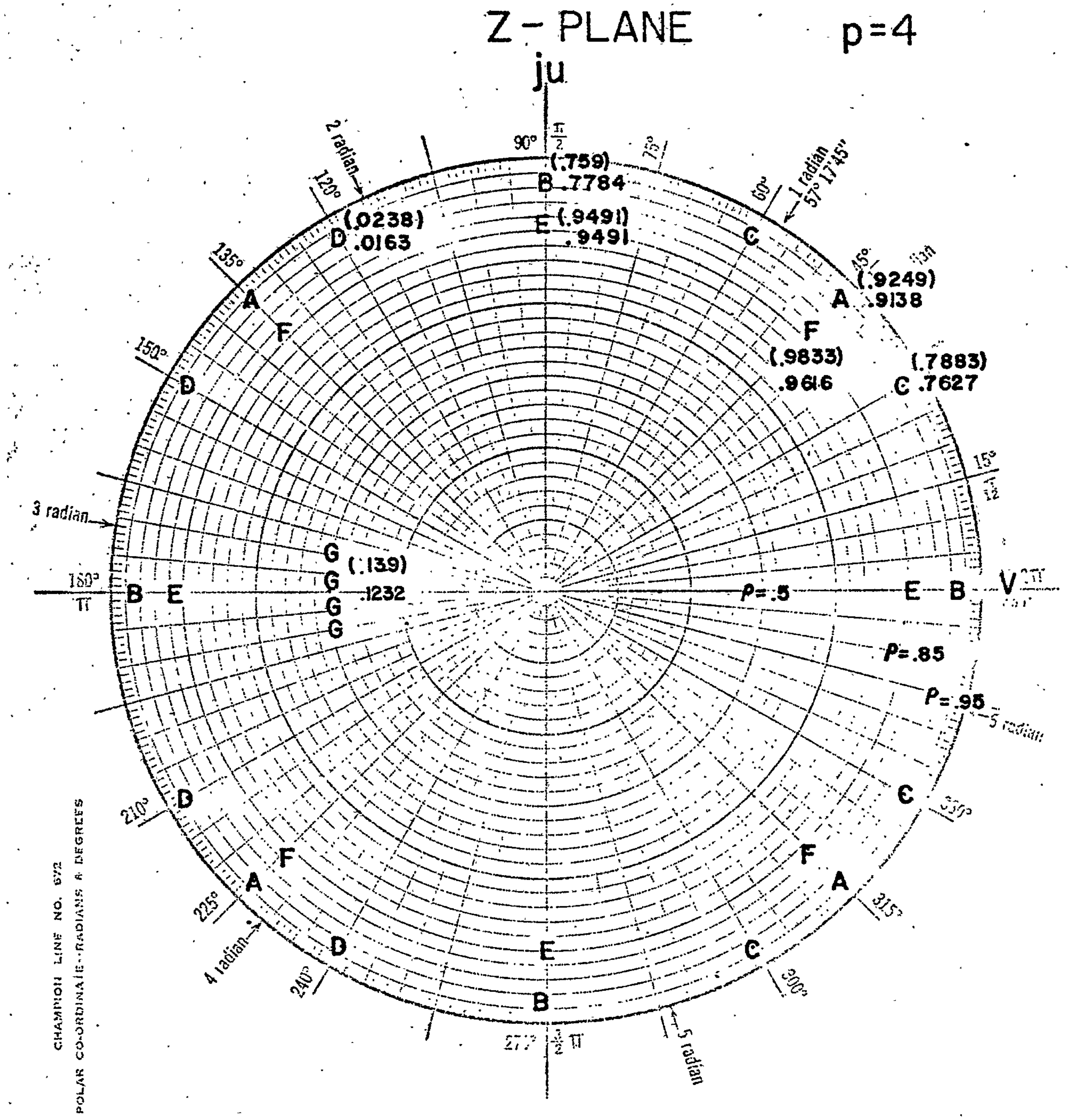
Z - PLANE
p=4
ju
(.759)
B .7784
(.9491)
E .9491
(.0238)
D .0163
C
A
F
(.9249)
A .9138
F
(.9833)
.9616
(.7883)
C .7627
D
G
G
G
G
(.139)
.1232
B
E
P=.5
E
B
V
P=.85
P=.95
C
D
F
A
F
A
D
E
C
B
2 radian
1 radian
57° 17'45"
3 radian
4 radian
5 radian
90°
120°
135°
150°
180°
210°
225°
240°
270°
300°
315°
330°
60°
45°
75°
15°
CHAMPION LINE NO. 672
POLAR CO-ORDINATE - RADIANS & DEGREES

FIG. 2

Z-PLANE

p=6

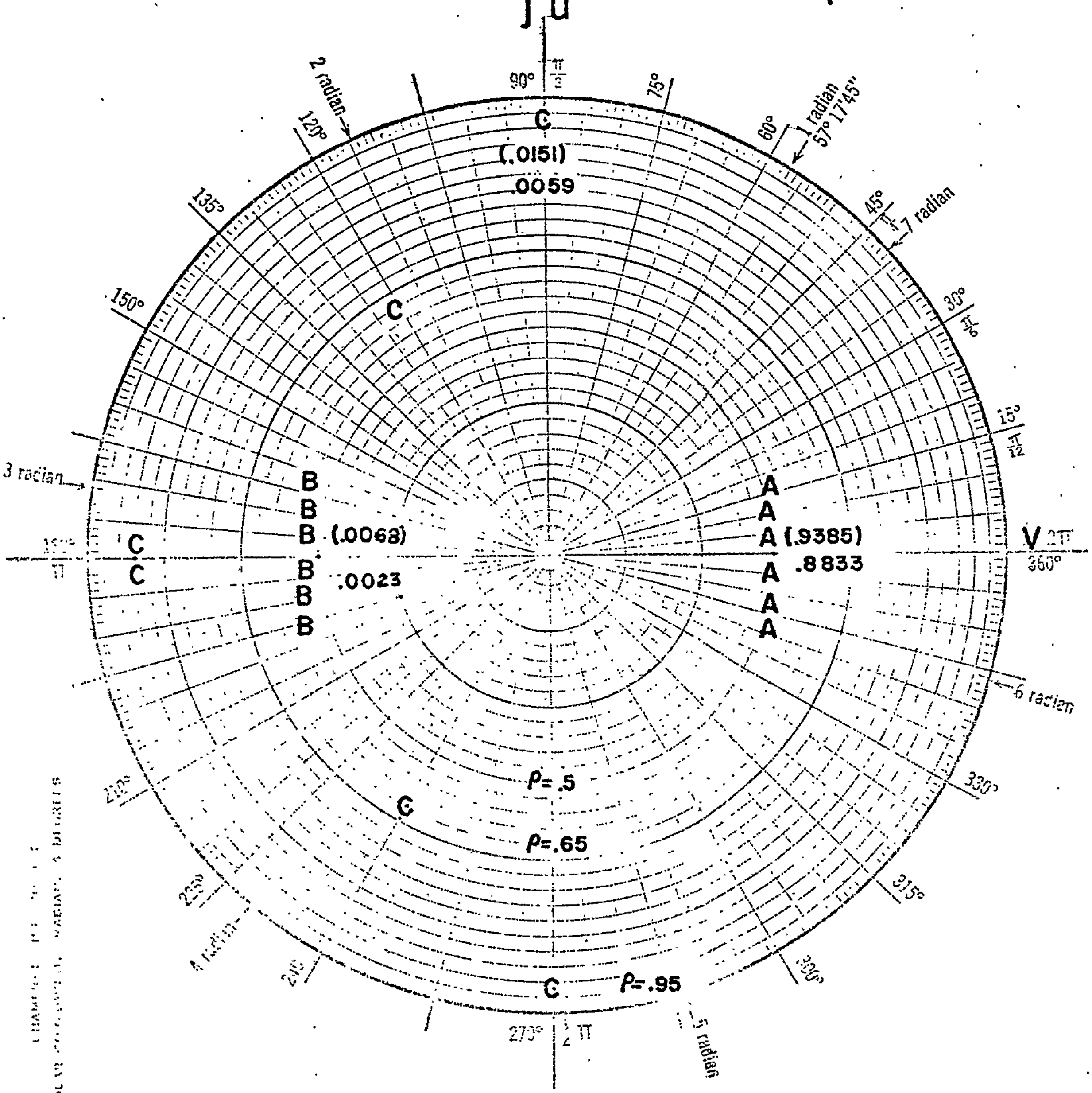

FIG. 3