SIO REFERENCE SERIES

79-12

MODEL SKILL AND MODEL SIGNIFICANCE IN LINEAR REGRESSION HINDCASTS

Rudolph W. Preisendorfer

University of California Scripps Institution of Oceanography

July 1979

University of California, San Diego

Scripps Institution of Oceanography

La Jolla, California

MODEL SKILL AND MODEL SIGNIFICANCE IN LINEAR REGRESSION HINDCASTS

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 Series #79-12

July 1979

Abstract

There are three main problems encountered when applying linear regression models to geophysical time series, namely the problems of: model significance, model hindcast skill and model forecast skill. In this note we solve the first two problems by the systematic introduction of various hindcast performance indexes of the linear regression model, such as canonic skill Q, classic skill S, and ineptness I, and by deriving their probability density functions on the assumption of gaussian noise governing the residual vectors. The notion of signal to noise ratio λ is introduced into the analyses of the problems of significance and skill, and it is shown how λ, as a parameter in the probability density function for Q, S, and I, can be used to generate confidence intervals for its estimation. As a result, by means of λ, it is possible to unify the problems of model significance and model hindcast skill in a way that suggests various basic strategies to maximize model hindcast skill subject to the constraint that a model be significant. In this way a framework for linear regression hindcast theory is provided on which the solution for the third main problem may eventually be based.

TABLE OF CONTENTS

MODEL SKILL AND MODEL SIGNIFICANCE

IN LINEAR REGRESSION HINDCASTS

by

Rudolph W. Preisendorfer

1. Introduction

From the point of view of a physical oceanographer or a meteorologist, the concept of linear regression provides an interesting mixture of dynamics and statistics in the sense that the usual form of a linear regression equation, namely

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon} \quad , \tag{1.1}$$

holds simultaneously within it the algebraic essence of a dynamical law: $\underline{y} = \underline{X}\,\underline{\beta}$, and a random perturbation $\underline{\varepsilon}$ of that law. Thus, as we shall briefly illustrate below, we may envision the matrix $\underline{X}$ as embodying a generalized force and $\underline{\beta}$ as the transfer function that converts $\underline{X}$ into an observable field $\underline{y}$ as seen through an intermediate haze of noise $\underline{\varepsilon}$. In such a dynamical context, $\underline{X}$ and $\underline{\beta}$ may rigorously take on a great variety of forms, ranging from simple ohm's law quantities in linear electric circuits, to the appropriate parts of solutions of linear wave equations arising in oceanography and meteorology.

In the present note we shall prepare a framework for the general solutions of two of the three main problems arising when (1.1) is directed toward the description of linear dynamical processes in random settings. In practice these three problems arise in ways which we shall now briefly describe.

A. Estimating The Model Parameter $\underline{\beta}$

The n x 1 vector $\underline{y}$ in (1.1) is imagined to be a set of n observations of a field which arises through the action of a set of driving forces situated at p locations in space, at each of which n observations of the force are made. Let '$\underline{x}_j$' denote the n x 1 vector summarizing n observations of the forcing field made at the jth point. Then write $\underline{X} \equiv [\underline{x}_1 \ \underline{x}_2 \ \dots \ \underline{x}_p]$, so that $\underline{X}$ is an n x p matrix. For example the $\underline{x}_j$'s can be $\underline{p}$ time series of sea level atmospheric pressure, and $\underline{y}$ can be the corresponding time series of sea surface temperatures at a point. By means of a least squares procedure, to be reviewed below, we can estimate the components of the vector $\underline{\beta}$, using the observed driving field $\underline{X}$ and observed resultant field $\underline{y}$; thus if $\hat{\underline{\beta}}$, is the desired estimate of $\underline{\beta}$, we find:

$$\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y} \ . \tag{1.2}$$

Here 'T' denotes matrix transpose. If there is no noise, i.e., if in (1.1), $\underline{\varepsilon} = 0$, then on substitution of $\underline{y} = \underline{X}\,\underline{\beta}$ into (1.2), we would find $\hat{\underline{\beta}} = \underline{\beta}$. In this case, the least square estimation technique allows us to determine exactly the essential physical parameter $\underline{\beta}$ of the linear regression model (1.1) in the absence of noise.

When noise is present in (1.1), then the solution (1.2) for $\hat{\underline{\beta}}$, on substitution of (1.1) for $\underline{y}$, becomes

$$\hat{\underline{\beta}} = \underline{\beta} + (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon} \ . \tag{1.3}$$

Now the physical parameter vector $\underline{\beta}$ is masked by the noise vector $(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}$. One no longer is certain that $\underline{\beta}$ really exists as a nonzero vector. Indeed, setting $\underline{\beta} = 0$ in (1.1) and (1.3) suggests that what we could observe is simply pure noise; and for any finite sample of size n, no statistical test can absolutely assure us that the observation $\underline{y}$ is not pure noise.

B. Problem of Model Significance

This brings us to the first main problem arising in the use of (1.1) to study physical systems in nature: *how does one decide, from the measurements* $\underline{y}$*,* $\underline{X}$ *and knowledge of the statistics of* $\underline{\varepsilon}$*, that* $\underline{\beta} \neq 0$? This is the *problem of model significance*. The term 'significance' is used to indicate that we cannot decide with certainty that $\underline{\beta} \neq 0$, but only to indicate with some stated measure of confidence (e.g., on the 95% level) that $\underline{\beta} \neq 0$. If we find that $\underline{\beta} \neq 0$, then we can view $\underline{y} = \underline{X}\ \underline{\beta}$, with some measure of confidence, as a non trivial (i.e., a not completely noisy) indicator of a law of nature worthy of closer scrutiny. For this is our principal attitude toward (1.1): namely that (1.1) is merely a preliminary indicator of a possibly significant mode of dynamic behavior of a portion of (say) the atmosphere/hydrosphere fluid system. This attitude does not rule out the possibility that the relevant law itself contains random structure; nor perhaps that the most we could ever know about the system would be certain simple refinements of (1.1) itself.*

C. Model Skills

It is quite possible that an estimated model $\hat{\underline{y}} = \underline{X}\ \hat{\underline{\beta}}$ of the law $\underline{y} = \underline{X}\ \underline{\beta}$ is significant in the above sense, but that (because of an overly-dominant $\underline{\varepsilon}$ term, e.g.) it may be of little value in describing the temporal variations of the field $\underline{y}$, i.e., that $\hat{\underline{y}} = \underline{X}\ \hat{\underline{\beta}}$ is not very skillful† in approximating $\underline{y}$ for the given field $\underline{X}$. A quantitative measure of such skill is the ratio

---

* The bulk formulas for the thermodynamic processes at the air/sea surface are currently of this kind; and some of the various parameterizations of physical processes incorporated in the currently most advanced general circulation models of the air and sea are also of this kind.

† This will be illustrated in §13B, C.

$$Q = \frac{||X\hat{\beta}||^2}{||\underline{y}-X\hat{\beta}||^2} = \frac{||\hat{\underline{y}}||^2}{||\underline{y}-\hat{\underline{y}}||^2} \qquad (1.4)$$

where $||\underline{x}||^2 = x_1{}^2 + \ldots + x_n{}^2$, for any n dimensional vector $\underline{x} = [x_1, x_2, \ldots, x_n]^T$ ('T' denotes transpose; all vectors are written as single columns of scalars). Thus Q is the ratio of the square of the length of $\hat{\underline{y}}$ (i.e., $\hat{y}_1{}^2 + \ldots + \hat{y}_n{}^2$) to the square of the length of the residual vector $\underline{y}-\hat{\underline{y}}$, the vector representing the error of the model in its attempt to describe $\underline{y}$. Clearly, the greater Q the better the model. Q is the *canonic skill* of the model.

Another measure of model fit is given by

$$S = \frac{||X\hat{\beta}||^2}{||\underline{y}||^2} = \frac{||\hat{\underline{y}}||^2}{||\underline{y}||^2} \qquad (1.5)$$

which is the ratio of the estimator's square to the estimand's square. Clearly, the greater S, the better the model. S is the *classic skill* of the model.

Still another index of the performance of the model $\hat{\underline{y}} = X\hat{\beta}$ in describing $\underline{y} = X\beta$ is the ratio

$$R = \frac{||\underline{y}-X\hat{\beta}||^2}{||\underline{y}||^2} = \frac{||\underline{y}-\hat{\underline{y}}||^2}{||\underline{y}||^2} \qquad (1.6)$$

The smaller R, the better the model. R is the *residual unskill* index. As we shall see below, R and S are simply related by:

$$R + S = 1 \; , \qquad (1.7)$$

using an n dimensional form of Pythagoras' theorem. From this we see that either R or S is sufficient to characterize the performance of the model. Further, one can readily see that:

$$Q = S/R = S/(1-S) = (1-R)/R \tag{1.8}$$

D. Problem of Model Hindcast Skill

All three indexes are closely tied together in their abilities to rate the performance of $\hat{\underline{y}} = X\hat{\underline{\beta}}$ in describing $\underline{y} = X\underline{\beta}$. For a chosen sample size n, we can watch how that performance is affected by varying the single remaining parameter in (1.1) available to us, namely the number p of time series used to describe $\underline{y}$. Thus the jth reading of y, namely y_j is given by the jth component of (1.1):

$$y_j = \sum_{k=1}^{p} x_{jk}\beta_k + \varepsilon_j \;, \quad j = 1, \ldots, n \tag{1.9}$$

Our options are limited by observing that: the driving forces x_{jk} are given by nature; the observations y_j are measured *in situ*; the noise ε_j is inevitable. With these as given, to improve our skill (to make Q, S greater or R smaller) it is left to us only to decide on which time series $\underline{x}_j$ to measure and how many there will be included in (1.1). It has been the experience of many practitioners of linear regression modeling over the years that an unrestrained growth in the number p of predictors $\underline{x}_j$ (holding n momentarily fixed) results in successively higher skill values Q, S (or lower residual unskill R) while simultaneously there results a decreasing model significance (i.e., one must drop the level of confidence in order to continue to assert model significance). It has taken the last several years of work by climate researchers studying the air/sea interaction problem using linear regression theory to allow this insight about skill/significance dependence on p to be so succinctly stated. (cf. Barnett and Hasselmann (1979), Davis (1978)). In this way we come to the statement of the second main problem of linear regression: *how does one choose the location and number of the predictor time series in $\underline{X}$ so as to maximize a given skill index subject to the constraint that the associated model be significant?* This is the problem of model hindcast skill.

E. Problem of Model Forecast Skill

The word 'hindcast' in 'the problem of model hindcast skill' emphasizes that we are momentarily concerned only about how well the model may be cast on the past; i.e., how well past observations $\underline{y}$ are fitted by $\underline{X}\hat{\underline{\beta}}$. There is no automatic guarantee that a significant, skillful hindcast of (1.1) over a particular data stretch $\underline{X}$ will continue to be skillful when the estimated $\hat{\underline{\beta}}$ is used on a fresh stretch of time series beyond that of $\underline{X}$. In this way we come to the third and final main problem of linear regression studies of physical processes: *how does one choose the location and number of the predictor time series so as to maximize a given forecast skill index, subject to the constraint that the associated model be significant in the hindcast mode?*

F. The Problems Studied in this Note and a Summary of Results

We shall lay the groundwork for the full statistical solution of the model significance and model hindcast skill problems defined above. In this way the advances of Lorenz, Davis, Barnett and Hasselmann can be consolidated and possibly extended. The third problem, that of model forecast skill, will not be considered here. In our studies below, we shall be motivated in particular to clarify the pioneering work in this area by Lorenz (1956), and shall be guided by the recent advances on the two problems by Barnett and Hasselmann (1979), and by Davis (1978). The work of Barnett and Hasselmann, in particular, has shown the importance of including the probability density function of the difference $\underline{\beta}-\hat{\underline{\beta}}$ in their analysis of the model significance problem. Inspired by their example, the work below turns to those parts of the work of Davis and Lorenz wherein the introduction of the probability density function (pdf) of the classic skill index S would correspondingly clarify their discussions of model hindcast skill. In the setting of homogeneous noise, i.e., where $<\underline{\varepsilon}\underline{\varepsilon}^T> = \sigma^2\underline{I}$, it will turn out that, by introducing the notion of the *signal to noise ratio* $\lambda \equiv ||\underline{X}\underline{\beta}||^2/\sigma^2$ into the settings of the skill and significance problems, we shall be able to unify the various approaches of Davis, Barnett and Hasselmann to these problems, so that the solution of each problem may cast light on the solution of the other. Specifically, the signal to noise ratio λ will be incorporated into the probability density functions for Q, S, R (and their three relatives) along with the sample size n and predictor number p. In this way we will be able to watch the simultaneous, coupled effects on model significance and model hindcast skill as p, n, and λ are varied. Some further corollaries of the presence of λ in the probability density functions for Q, S, R in the linear regression theory are: a unified geometric formulation of the hindcast performance indexes (the three skills Q, S, C, and the three unskills R, I, U); 'skeleton' Monte Carlo representations of the six performance indexes as random variables which, with the above geometric formulation, considerably clarify the p, n, λ behavior of these indexes; the derivation of an unbiased estimator of λ; a small-sample theory of the confidence limits of λ, based on the pdf of any of the six performance indexes; a large-sample theory of the confidence limits of λ, based on a form of the central limit theorem; and exact knowledge of the population means and variances of the performance indexes. The work concludes with two appendixes, the first giving a self-contained derivation of the general forms of the pdfs for the performance indexes, and the second appendix which gives finite-term integrals of the pdfs, yielding efficient numerical procedures to find the $\frac{1}{2}\alpha$, $1-\frac{1}{2}\alpha$ significance levels for each performance index. Also appended are figures and tables describing in a preliminary way some of the n, p, λ-behaviors of the performance indexes, thereby yielding information by which a user of linear regression representations of physical processes can deepen his understanding of those representations.

2. Dynamical Aspects of Regression Equations

Our introductory remarks referred to the dynamical laws inherent in the form (1.1). It is of considerable help when visualizing the physical applications of (1.1), particularly in geophysical settings, to see the $\underline{\beta}$ vector as a transfer function of some sort, and the $\underline{X}$ matrix as time series of variously located driver forces giving rise to the observed field $\underline{y}$. Some insight into the origins of $\underline{\varepsilon}$ are also forthcoming. In this section we will sketch the main stages of a derivation leading to (1.1) starting from a two-dimensional linear partial differential equation. The reader may imagine it describing damped long-wave motion in a fluid basin or equivalently, linearized atmospheric waves over oceanic or land regions. The essential ideas of the reduction to linear regression form are, of course, independent of the specific physical interpretation. The equation (2.1) below merely serves to draw our attention to certain general dynamical aspects inherent in the form and application of (1.1).

A. Wave Equation

We start with the two dimensional wave equation governing the field $\eta(\underline{z},t)$ where $\underline{z} = (x,y)$, over some region R,

$$\eta_{tt} + a\eta_t + b\eta - c^2(\eta_{xx} + \eta_{yy}) = f_* \quad . \tag{2.1}$$

Here a, b are constants, describing dissipative mechanisms in the fluid (or general medium) of interest. c is the speed of propagation of undamped waves. f_* is the driving force. For example, if $\eta(\underline{z},t)$ is wave elevation at point $\underline{z}$ at time t, $f_*(\underline{z},t)$ may be the sea level pressure at the same space time point.

B. Solution of Wave Equation

We are interested in a solution of (2.1) subject to the initial conditions

$$\eta(\underline{z},0) = f(\underline{z})$$

and

$$\eta_t(\underline{z},0) = g(\underline{z})$$

and boundary conditions

$$\alpha_1\eta_n(\underline{b},t) + \beta_1\eta(\underline{b},t) = 0$$

where η_n is a derivative normal to the fluid boundary at point $\underline{b} = (x_b, y_b)$, at each $\underline{b}$ of the boundary B of the region R over which (2.1) is to be solved.

It can be shown that under the preceding conditions there exist two Greens' functions G, H such that for every $\underline{z}$ in R, and $t \geq 0$,

$$\eta(\underline{z},t) = \int_R \int_0^t f_*(\underline{z}',t')G(\underline{z}',\underline{z},t-t')dt'dA(\underline{z}') + \int_R [\eta(\underline{z}',0)H(\underline{z}',\underline{z},t) + \eta_t(z',0)G(\underline{z}',\underline{z},t)]dA(\underline{z}') \quad (2.2)$$

where

$$G(\underline{z}',\underline{z},t) = e^{-\alpha t}\sum_{k=1}^{\infty} \frac{\sin\gamma_k t}{\gamma_k} u_k(\underline{z}')u_k(\underline{z})$$

$$H(\underline{z}',\underline{z},t) = e^{-\alpha t}\sum_{k=1}^{\infty} [\cos\gamma_k t + \frac{\alpha}{\gamma_k}\sin\gamma_k t]u_k(\underline{z}')u_k(\underline{z})$$

and where

$$\gamma_k = [\lambda_k{}^2 - \alpha^2] \; , \; \alpha = a/2, \; k = 1, \ldots, \infty$$

The λ_k are eigenvalues of the spatial Helmholtz equation associated with (2.1) and the given boundary conditions. Moreover, the functions $u_k(\underline{z})$ are the corresponding eigenfunctions of the spatial Helmholtz equation, and have the properties

$$\int_R u_k(\underline{z})u_\ell(\underline{z})dA(\underline{z}) = \delta_{k\ell}$$

and

$$\sum_{k=1}^{\infty} u_k(\underline{z})u_k(\underline{z}') = \delta(\underline{z}-\underline{z}') .$$

C. <u>Discretized Solution of the Wave Equation Diagnostic Mode</u>

We turn now to the simplification of (2.2) with an eye toward attaining the associated regression equation. The first term in (2.2) indicates the way the driving force $f_*(\underline{z}',t')$ makes itself felt at $\underline{z},t$ through the transfer function $G(\underline{z}',\underline{z},t-t')$, which communicates the cause at $\underline{z}',t'$, to the effect at $\underline{z},t$. It is the linearity of the process and the constancy of the coefficients a,b,c in (2.1) that allows G to depend only on t-t'. The second term in (2.2) shows how the initial state of the fluid system is felt at time t later. As time t grows, the exponential terms in G and H tend to make the system forget its original state, so that in the long run, i.e., for t greater than some τ_o, (2.2) can be shortened to

$$\eta(\underline{z},t) = \int_R \int_o^t f_*(\underline{z}',t')G(\underline{z}',\underline{z},t-t')dt'dA(\underline{z}') \tag{2.3}$$

In the diagram below we have partitioned the region R into r parts over each of which, at a given moment in time, we may approximate the spatial behavior of f_* by an appropriately chosen single number. Moreover, we can divide the time interval [o,t] into τ_o subintervals over each of which f_* can be represented by a single number. Thus by a mean value theorem of calculus we can write (2.3) as:

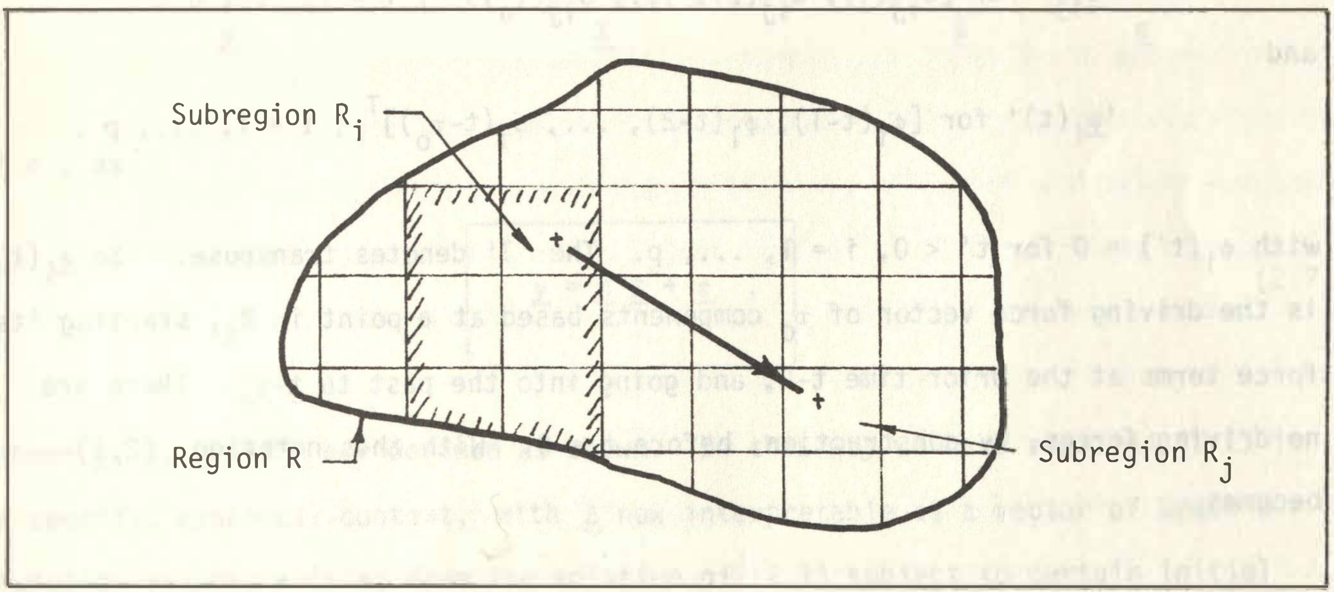
$$\eta(\underline{z},t) = \sum_{i=1}^{r} \sum_{k=1}^{\tau_o} \int_{R_i} \int_{t_{k-1}}^{t_k} f_*(\underline{z}',t')G(\underline{z}',\underline{z},t-t')dt'dA(\underline{z}')$$

or as

$$\eta(\underline{z},t) = \sum_{i=1}^{r} \sum_{\tau=o}^{\tau_o} \phi_i(t-\tau)\, G_{ij}(\tau) \tag{2.4}$$

where $\underline{z}$ is in R_j, and where $f_*(\underline{z}',k) \equiv \phi_i(k)$ for some $\underline{z}'$ in R_i and $k=t'$ in $[t_{k-1},t_k]$. Thus the time index has been discretized along with the space index, and τ_o is the integer such that $t_k > t_o$, when $k > \tau_o$. Moreover, we have set:

$$G_{ij}(\tau_o-k) \equiv \int_{R_i} \int_{t_{k-1}}^{t_k} G(\underline{z}',\underline{z},t-t')dt'dA(\underline{z}') .$$



We next decide that only p of the r subregions in R will contribute essential dynamical effects to $\eta(\underline{z},t)$ at $\underline{z}$ in R_j. Hence (2.4) can be written

$$\eta_j(t) = \sum_{i=1}^{p} \sum_{\tau=1}^{\tau_o} \phi_i(t-\tau)G_{ij}(\tau) + \sum_{i=p+1}^{r} \sum_{\tau=1}^{\tau_o} \phi_i(t-\tau)G_{ij}(\tau)$$

$$\equiv \sum_{i=1}^{p} \sum_{t=1}^{\tau_o} \phi_i(t-\tau)G_{ij}(\tau) + \varepsilon_j(t) \qquad (2.5)$$

In this way the second sum term in (2.5) becomes the noise $\varepsilon_j(t)$.

D. <u>The Linear Regression Equation</u>

It is now a simple pair of steps to the form (1.1). Let us write, for fixed j

'$\underline{G}_{ij}$' for $[G_{ij}(1), G_{ij}(2), \ldots, G_{ij}(\tau_o)]^T$, $i = 1, \ldots, p$.

and

'$\underline{\phi}_i(t)$' for $[\phi_i(t-1), \phi_1(t-2), \ldots, \phi_i(t-\tau_o)]^T$, $i = 1, \ldots, p$.

with $\phi_i(t') = 0$ for $t' < 0$, $i = 1, \ldots, p$. The 'T' denotes transpose. So $\underline{\phi}_i(t)$ is the driving force vector of τ_o components based at a point in R_i, starting its force terms at the prior time t-1, and going into the past to $t-\tau_o$. There are no driving forces, by construction, before t = 0. With this notation, (2.5) becomes

$$\eta_j(t) = [\underline{\phi}_1{}^T(t), \underline{\phi}_2(t), \ldots, \underline{\phi}_p{}^T(t)] \begin{bmatrix} \underline{G}_{1j} \\ \underline{G}_{2j} \\ \cdot \\ \cdot \\ \cdot \\ \underline{G}_{pj} \end{bmatrix} + \varepsilon_j(t) \qquad (2.6)$$

for all integer times $t \geq 0$.

We can write (2.6) out explicitly for times 1, ..., n, i.e., for any n times (not necessarily consecutive) representing n snapshots of the dynamical process in R. The resulting n copies of (2.6) can then be arranged in vector form:

$$\underbrace{\begin{bmatrix} \eta_j(1) \\ \eta_j(2) \\ \cdot \\ \cdot \\ \cdot \\ \eta_j(n) \end{bmatrix}}_{\underline{y}} = \underbrace{\begin{bmatrix} \underline{\phi}_1^T(1) & \underline{\phi}_2^T(1) & \dots & \underline{\phi}_p^T(1) \\ \underline{\phi}_1^T(2) & \underline{\phi}_2^T(2) & \dots & \underline{\phi}_p^T(2) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \underline{\phi}_1^T(n) & \underline{\phi}_2^T(n) & \dots & \underline{\phi}_p^T(n) \end{bmatrix}}_{\underline{X}} \underbrace{\begin{bmatrix} \underline{G}_{1j} \\ \underline{G}_{2j} \\ \cdot \\ \cdot \\ \cdot \\ \underline{G}_{pj} \end{bmatrix}}_{\underline{\beta}} + \underbrace{\begin{bmatrix} \varepsilon_j(1) \\ \varepsilon_j(n) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_j(n) \end{bmatrix}}_{\underline{\varepsilon}}$$

i.e., as

$$\boxed{\underline{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}\ .} \tag{2.7}$$

where $\underline{y}$, $\underline{X}$, $\underline{\beta}$ and $\underline{\varepsilon}$ are defined as shown. In this way we have realized (1.1) in a specific dynamical context, with $\underline{\beta}$ now interpretable as a vector of Green's function values, arising from the solution of (2.1) subject to certain initial and boundary conditions. The noise vector $\underline{\varepsilon}$ is seen to be the linear superposition of (in practice usually very many) perfectly legitimate pieces of information about the dynamical system in R. But by definition, unwanted information is 'noise'. By the grace of the central limit theorem, the successive realizations of $\underline{\varepsilon}$ arising from more or less independent successive n-samples of the η field in R can usefully be considered as drawn from an infinite ensemble of gaussianly distributed n-dimensional vectors.

E. Discretized Solution of the Wave Equation - Predictive Mode

We return to the discretized solution (2.5) and examine it for the possibility of yielding up a predictive equation. How must (2.5) be modified so as to have a prediction of $\eta_j(t)$ from knowledge of the driving forces $\phi_i(t-\tau)$? Clearly, to achieve this, the summation over τ must not begin at $\tau = 1$, but at some integer $\ell > 1$. For in order to predict $\eta_j(t)$ we must restrict use of driving terms to some finite time in the past of t. Thus we can write (2.5) in the predictive mode as:

$$y_j(t) = \sum_{i=1}^{p} \sum_{\tau=\ell}^{\tau_o} \phi_i(t-\tau)G_{ij}(\tau) + \sum_{i=p+1}^{r} \sum_{\tau=1}^{\tau_o} \phi_i(t-\tau)G_{ij}(\tau) + \sum_{i=1}^{p} \sum_{\tau=1}^{\ell-1} \phi_i(t-\tau)G_{ij}(\tau) \qquad (2.8)$$

where now the noise term $\varepsilon_j(t)$ contains information — all inaccessible by fiat — about effects at other places up to the present and effects at the same place in the immediate past. A reduction of (2.8) to (2.7) now can be made, with no major changes in the steps: The time lags in $\underline{G}_{ij}$ now being at $\ell > 0$ and continue to τ_o; the time arguments in $\underline{\phi}_i(t)$ now begin at $t-\ell$ and continue to $\tau-\tau_o$. The final form of the regression equation (2.7) is unchanged.

F. Discretized Solution of the Wave Equation - General Mode

The preceding modification (2.8) of (2.4) suggests still another. It is possible in principal to have information about the drivers $\phi_i(t-\tau)$ for $\tau = 1, \ldots, \ell$, then a gap of knowledge from $\ell+1, \ldots,$ to m, and then knowledge of $\phi_i(t-\tau)$ for $\tau = m+1, \ldots \tau_o$. The resultant form of (2.4) can be written in general as

$$y_j(t) = \sum_{i=1}^{p} \sum_{\tau\epsilon T} \phi_i(t-\tau)G_{ij}(\tau) + \epsilon_j(t) \quad (2.9)$$

where now T is a set of integers where the information about $\phi_i(t-\tau)$ is known for each τ in T. Clearly (2.9) covers both (2.8) and (2.5), and even (2.4). Once again the general regression form (2.7) results.

The form (2.9) is sufficiently general to allow even negative integers. The interpretation in this case is that of a postdiction of the observed field $y_j(t)$, i.e., a characterization of the past behavior in terms of its future behavior. This is not as absurd as it may first appear.

G. Postdiction vs Prediction

As we shall see in the next section, the determination of the $\underline{\beta}$ vector via least squares fit of $\underline{X}\ \beta$ to $\underline{y}$ is unconcerned about the specific information contained in $\underline{X}$ and $\underline{y}$. From an algebraic point of view, the normal equations will work on any $\underline{X}$ and any $\underline{y}$ to produce an estimate of $\underline{\beta}$. Yet there is something in our intuition that says (2.7) in the real world will be more successful in the predictive than the postdictive mode. Intuition is correct, but for reasons which are not easily stated in everyday terms. A partial explanation follows.

If we return to the wave equation (2.1) and set the dissipative term a to zero, the exponential terms in the Green's functions of (2.2) become unit-valued. In this case it can be shown that the predictive and postdictive modes of (2.7) are equally powerful with respect to any measure of hindcast skill and any measure of forecast skill we can reasonably devise. When a > 0, however, the predictive mode requires t > 0 and the e^{-at} terms tend to dampen the effects of $\epsilon_j(t)$ in (2.9), but the postdictive mode tends to magnify the effects of $\epsilon_j(t)$ since t < 0 and the e^{-at} terms can become enormous for reasonably-sized negative integers in T.

This situation is closely analogous to the numerical problem of trying to solve a partial differential equation, such as (2.1), *backwards* into time, starting from given initial conditions and boundary conditions as in par B above. As the numerical procedure is followed for a case in which $a > 0$, it is found that numerical instabilities arise and as one progresses into the past the solution literally blows up by producing enormous, unrealistic $\eta(\underline{z},t)$ values for $t < 0$. By the same token, solving (2.1) *forwards* into time, any slight numerical glitches (e.g., round off errors) arising in the machine's performance (which in the previous case were disastrous) are dampened by the presence of the e^{-at} effect, errors are forgotten, so to speak, and information about $\phi_i(t-\tau)$ for $t > \tau_o$, for some integer τ_o, does not contribute materially to $y_j(t)$, for large $t-\tau$.

H. Interim Conclusions

The net result of these observations about (2.7) vis a vis (2.1) indicates that we should expect our predictive uses of (2.7) to be generally more effective than the postdictive uses. For once in this real world of real frustrations besetting the forecaster of geophysical time series, something seems to be working in his favor: if he keeps good records, the forecaster doesn't have to worry about postdiction, and he can turn to overcome the evils of the lesser of the two tasks: prediction.

Yet the damping mechanism in (2.1) eventually catches up to the forecaster here, too. His records, no matter how well gathered and kept, will be relevant only for limited predictions into the future; in attempting a given prediction, damping makes irrelevant the use of information beyond (say) τ_o into the past; damping and unforseen wanderings of ϕ_i in the future and elsewhere make irrelevant his predictions beyond τ_o into the future. If he turns to predict the predictors ϕ_i, he could, if not careful, become enmeshed on the threshold of an infinite regress.

With these reflections, we turn to the exposition below with an overriding feeling (despite the aspect of precision and power it at first conveys) that it is merely an exercise in algebra and geometry bordering on the brink of futility.

3. Least Squares Estimate of $\underline{\beta}$

Having examined the dynamical basis of (1.1), we now turn to the practical matter of estimating the model parameter $\underline{\beta}$ and also the model noise $\underline{\varepsilon}$ in (1.1). To begin, we have the unknown, $\underline{\beta}$ and two knowns $\underline{y}$, $\underline{X}$ from which we attempt to find the best approximation $\hat{\underline{\beta}}$ to $\underline{\beta}$ in the least squares sense.

Let $\underline{X}$ represent an n x p matrix of p columns, each of which comprises n measurements of a driver force field. Thus if $\underline{X} = (\underline{x}_1\ \underline{x}_2\ \dots\ \underline{x}_p)$, then $\underline{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$ are the n measurements at the jth point in space. The corresponding n values of the observed field $\underline{y}$ are given by $\underline{y} = (y_1, \dots, y_n)^T$. Our discussions in §2 show that (1.1) may be taken in its general mode, so that what we are now to do holds equally well — in an algebraic sense — for both predictive and postdictive activities with (1.1).

We wish to represent the vector $\underline{y}$ as a linear combination of the vectors $\underline{x}_j$, j = 1, ..., p. Thus let us write

$$'\underline{\delta}' \text{ for } \underline{y} - \sum_{k=1}^{p} \underline{x}_k\, \alpha_k \quad . \tag{3.1}$$

With $\underline{y}$ and the $\underline{x}_j$ given, we search through the set of all p dimensional vectors $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)^T$ for that which minimizes $||\underline{\delta}||^2 = \sum_{j=1}^{n} \delta_j{}^2$. Clearly, for a useful and unique solution to this problem, we must postulate that $n \geq p$ at this stage.

Now from (3.1), the jth component of $\underline{\delta}$ is

$$\delta_j = y_j - \sum_{k=1}^{p} x_{jk}\alpha_k \quad , \quad j = 1, \ldots, n. \tag{3.2}$$

Thus we wish to find the α_j which minimize

$$r(\alpha_1,\alpha_2,\ldots,\alpha_p) \equiv ||\underline{\delta}||^2 = \sum_{j=1}^{n} \delta_j^2 = \sum_{j=1}^{n} [y_j - \sum_{k=1}^{p} x_{jk}\alpha_k]^2 \ . \tag{3.3}$$

A necessary condition for the minimum of the function r is the set of p conditions:

$$\frac{\partial r}{\partial \alpha_k} = 0 \ , \ k = 1, \ldots, p \tag{3.4}$$

Thus in (3.3) we require

$$\frac{\partial r}{\partial \alpha_\ell} = -2 \sum_{j=1}^{n} [y_j - \sum_{k=1}^{p} x_{jk}\alpha_k]\, x_{j\ell} = 0 \ , \quad \ell = 1, \ldots, p$$

whence

$$\boxed{\sum_{k=1}^{p} [\sum_{j=1}^{n} x_{jk}x_{j\ell}]\,\alpha_k = \sum_{j=1}^{n} y_j x_{j\ell} \ , \quad \ell = 1, \ldots, p \ .} \tag{3.5}$$

The set (3.5) is the desired collection of p linear equations in the unknowns α_k, k = 1, ..., p. Knowing the x_{jk} and the y_j, we can thus find the solutions of (3.5). We can put (3.5) into matrix form to simplify subsequent work with it and its solution vector. Towards this end we note that the right side of (3.5) is the inner product of $\underline{y}$ and $\underline{x}_\ell$, i.e., $\underline{y}^T\underline{x}_\ell = \underline{x}_\ell^T\underline{y}$. The quantity in square brackets on the left in (3.5) is the $k\ell$ element of the symmetric matrix $\underline{X}^T\underline{X} = \underline{Z}$, i.e.,

$z_{k\ell} = z_{\ell k}$. Hence (3.5) may be written

$$\sum_{k=1}^{p} z_{\ell k}\alpha_k = \underline{x}_\ell^T \underline{y} . \tag{3.6}$$

If we denote the ℓth row of $\underline{Z}$ by '$\underline{z}^\ell$', then (3.6) can be written

$$\underline{z}^\ell \underline{\alpha} = \underline{x}_\ell^T \underline{y} \quad , \ell = 1, \ldots, p .$$

Collecting these p equations together on a vertical stack:

$$\begin{bmatrix} \underline{z}^1 \\ \underline{z}^2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{z}^p \end{bmatrix} \underline{\alpha} = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_p^T \end{bmatrix} \underline{y} \tag{3.7}$$

which is

$$\underline{X}^T \underline{X}\, \underline{\alpha} = \underline{X}^T \underline{y} .$$

Solving for $\underline{\alpha}$ and henceforth denoting the solution by '$\hat{\underline{\beta}}$', we find

$$\boxed{\hat{\underline{\beta}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}} \tag{3.8}$$

This is the desired least squares estimate of the model parameter $\underline{\beta}$, using the known time series information in $\underline{X}$ and $\underline{y}$. In order for the inverse in (3.8) to exist, the rank of $\underline{X}$ must equal p, i.e., the p vectors $\underline{x}_j$, j=1, ..., p must be linearly independent. This we assume henceforth.

4. Analysis of the Residual Noise

We now inquire as to how well the approximation of the observed field $\underline{y}$ by linear combinations of the $\underline{x}_j$ went. There are two separate aspects of this approximation. Firstly, we write

$$'\underline{\varepsilon}_{n-p}' \text{ for } \underline{y} - \underline{X}\,\hat{\underline{\beta}} \tag{4.1}$$

Here $\underline{\varepsilon}_{n-p}$ is an n dimensional vector which summarizes the fit that we have made to $\underline{y}$. $||\underline{\varepsilon}_{n-p}||^2$ is the minimum value of $||\underline{\delta}||^2$ sought in §3. We can write (4.1) in the tautological form:

$$\underline{y} = \underline{X}\,\hat{\underline{\beta}} + \underline{\varepsilon}_{n-p} . \tag{4.2}$$

Next we inquire as to how well we have approximated the *signal* $\underline{X}\,\underline{\beta}$ by $\underline{X}\,\hat{\underline{\beta}}$. Thus, secondly we write,

$$'\underline{\varepsilon}_p' \text{ for } \underline{X}\,\hat{\underline{\beta}} - \underline{X}\,\underline{\beta} \tag{4.3}$$

Here $\underline{\varepsilon}_p$ is an n dimensional vector. We now can write another tautology:

$$\underline{X}\,\hat{\underline{\beta}} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}_p \tag{4.4}$$

Combining (4.2), (4.4), we find the general form of (1.1):

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon} \tag{4.5}$$

where we have written

$$'\underline{\varepsilon}' \text{ for } \underline{\varepsilon}_p + \underline{\varepsilon}_{n-p} \quad . \tag{4.6}$$

It should be noted that $\underline{\varepsilon}$ is introduced into the theory in a way which anticipates its determination in practice: (4.1) obtains by direct computation the portion $\underline{\varepsilon}_{n-p}$; and (4.3) obtains its orthogonal complement $\underline{\varepsilon}_p$. In practice $\underline{\varepsilon}_p$ can be partially estimated only after several samples of size n - i.e., several fits of (1.1) to fixed data sets $\underline{X}$, have been made, and provided the sampling has been done from the <u>same</u> noise population. In general, however, $\underline{\varepsilon}_p$ is not exactly estimable. It is simply not observable without some inkling of $\underline{\beta}$, our main unknown! This is the reason why $\underline{\varepsilon}$ is then given a uniform variance for each component. In our ignorance, it's the best we can do (see, however, §6D, E below - also note §10B).

A. The Data-Space Projector

In order to understand the physical and geometric implications of the above definitions of $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$, $\underline{X}\,\hat{\underline{\beta}}$, $\underline{X}\,\underline{\beta}$, and their interrelations, we digress here to introduce an important matrix $\underline{P}$, the *data-space projector*, and develop some of its consequences useful for linear regression theory.

When we form $\underline{X}\,\hat{\underline{\beta}}$, using the representation for $\hat{\underline{\beta}}$ in (3.8), we find

$$\boxed{\underline{X}\,\hat{\underline{\beta}} = \underline{P}\,\underline{y}} \tag{4.7}$$

where we have written

$$'\underline{P}' \text{ for } \underline{X}\,(\underline{X}^T\underline{X})^{-1}\underline{X}^T \tag{4.8}$$

By direct computation we find that the n x n matrix $\underline{P}$ has the following properties

$$\underline{P}\,\underline{X} = \underline{X} \tag{4.9a}$$

$$\underline{P}^T = \underline{P} \tag{4.9b}$$

$$\underline{P}\,\underline{P} = \underline{P} \tag{4.9c}$$

$$(\underline{I}-\underline{P})\underline{P} = \underline{P}(\underline{I}-\underline{P}) = 0 \quad (\underline{0}\text{: nxn zero matrix}) \tag{4.9d}$$

$$(\underline{I}-\underline{P})\,(\underline{I}-\underline{P}) = (\underline{I}-\underline{P}) \quad (\underline{I}\text{: nxn identity matrix}) \tag{4.9e}$$

Property (4.9a) states that $\underline{P}$ acting on $\underline{X}$ leaves $\underline{X}$ unchanged. Actually, $\underline{P}$ acting on each column vector $\underline{x}_j$ of $\underline{X}$ leaves $\underline{x}_j$ unchanged; for the meaning of $\underline{P}\underline{X}$ is $\underline{P}\,[\underline{x}_1\ \underline{x}_2\ \dots\ \underline{x}_p] = [\underline{P}\underline{x}_1\ \underline{P}\underline{x}_2\ \dots\ \underline{P}\underline{x}_p]$ as an application of the definition of matrix multiplication will show. Hence by the meaning of matrix equality, we conclude that for each j=1, ..., p, $\underline{P}\underline{x}_j = \underline{x}_j$.

Property (4.9b) says $\underline{P}$ is symmetric, while (4.9c) results from two applications of $\underline{P}$ when $\underline{P}$ is written on the form (4.8). Property (4.9c) and (4.9a) are equivalent when $\underline{X}$ has rank p.

Property (4.9d) follows immediately from (4.9c), and will be crucial below in our further analysis of noise and linear regression: it says that the operator $\underline{I}$-$\underline{P}$ is orthogonal to $\underline{P}$. The practical import of this orthogonality is that it carries over to vectors which are images, under $\underline{P}$ or ($\underline{I}$-$\underline{P}$), of other vectors. Thus if $\underline{b} = \underline{P}\underline{y}$ and $\underline{a} = (\underline{I}-\underline{P})\underline{x}$, then necessarily $\underline{a}$ and $\underline{b}$ are orthogonal. Indeed $\underline{a}^T\underline{b} = [\underline{x}^T(\underline{I}-\underline{P})^T](\underline{P}\underline{y}) = \underline{x}^T[(\underline{I}-\underline{P})^T\underline{P}]\underline{y} = \underline{x}^T[(\underline{I}-\underline{P})\underline{P}]\underline{y} = \underline{x}^T\underline{0}\ \underline{y} = 0$. In this deduction we used (4.9b), (4.9d) and the fact that $(\underline{A}\underline{B})^T = \underline{B}^T\underline{A}^T$ and $(\underline{A} + \underline{B})^T = \underline{A}^T + \underline{B}^T$, for any two commensurate matrices $\underline{A}$, $\underline{B}$.

Another useful and far-reaching consequence of the properties (4.9) is that: *any element of* E_n *can be uniquely decomposed into a sum of two vectors, one lying in the space* E_p *spanned by the columns of* $\underline{X}$ *and the other in the orthogonal complement* E_{n-p} *to this space.* To see this, let $R(\underline{P}) \equiv \{\underline{z}: \underline{P}\underline{z} = \underline{z}\}$ and $R(\underline{I}-\underline{P}) \equiv \{\underline{z}: (\underline{I}-\underline{P})\underline{z} = \underline{z}\}$. It is easy to see that both $R(\underline{P})$ and $R(\underline{I}-\underline{P})$ are subspaces of E_n. Then if $\underline{z}$ is any vector in E_n, $\underline{z} = \underline{P}\underline{z} + (\underline{I}-\underline{P})\underline{z}$ is the desired decomposition. To see this, let $M(\underline{X}) = \{\underline{x}: \text{for some } \underline{\gamma} = (\gamma, \ldots, \gamma_p)^T, \underline{x} = \underline{X}\underline{\gamma}\}$. $M(\underline{X})$ is the p dimensional vector space spanned by the columns of $\underline{X}$. We now show that $R(\underline{P}) = M(\underline{X})$. If $\underline{z} \in R(\underline{P})$, then $\underline{z} = \underline{P}\,\underline{z} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{z} = \underline{X}\underline{\alpha}$, where $\underline{\alpha} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{z}$. Hence $\underline{z} \in M(\underline{X})$; so $R(\underline{P}) < M(\underline{X})$. On the other hand, if $\underline{x} \in M(\underline{X})$, then for some $\underline{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$, $\underline{x} = \underline{X}\underline{\alpha}$, and $\underline{P}\underline{x} = \underline{P}\underline{X}\underline{\alpha} = \underline{X}\underline{\alpha} = \underline{x}$, and $\underline{x} \in R(\underline{P})$; so $M(\underline{X}) < R(\underline{P})$. Hence $M(\underline{X}) = R(\underline{P})$ and $R(\underline{P})$ has dimension p. Let $\underline{Y} = (\underline{y}_1, \ldots, \underline{y}_q)$ be a basis for $R(\underline{I}-\underline{P})$. Since $R(\underline{I}-\underline{P})$ is a subspace of E_n, we know at least that $q \leq n$. If $\underline{z} \in E_n$, then we can write any $\underline{z}$ in E_n, as shown above, as a linear combination of a vector in $R(\underline{P})$ and a vector in $R(\underline{I}-\underline{P})$, i.e., as a linear combination of the p vectors $\underline{x}_j$ and the q vectors $\underline{y}_j$. Therefore $\underline{X}$, $\underline{Y}$ together consist of a set of linearly independent vectors that span E_n. Hence we must have $p + q = n$, i.e., $q = n-p$. Clearly each element of $R(\underline{I}-\underline{P})$ is orthogonal to each $R(P)$ so $R(\underline{I}-\underline{P})$ is the orthogonal complement to $M(\underline{X})$ in E_n. Finally, there is only one way to write $\underline{z}$ as a sum of a vector in $R(\underline{P})$ and one in $R(\underline{I}-\underline{P})$. Suppose, e.g., that $\underline{z} = \underline{x}+\underline{y} = \underline{x}' + \underline{y}'$, with $\underline{x}, \underline{x}'$ in $R(P)$ and $\underline{y}, \underline{y}'$ in $R(\underline{I}-\underline{P})$. Then since $(\underline{x}-\underline{x}') + (\underline{y}-\underline{y}') = \underline{0}$, we can apply $\underline{P}$ to each side and find $\underline{P}(\underline{x}-\underline{x}') + \underline{P}(\underline{y}-\underline{y}') = \underline{P}(\underline{x}-\underline{x}') = \underline{0}$, whence $\underline{P}\underline{x} = \underline{P}\underline{x}'$, and by definition of $R(\underline{P})$, $\underline{x} = \underline{x}'$. On the other hand, applying $(\underline{I}-\underline{P})$ to $(\underline{x}-\underline{x}') + (\underline{y}-\underline{y}') = \underline{0}$ yields $\underline{y} = \underline{y}'$, in a similar manner. Thus the main assertion above is proved. Henceforth we will simply write 'E_p' for $R(\underline{P})$ and 'E_{n-p}' for $R(\underline{I}-\underline{P})$.

Since $\underline{P}$ maps E_n onto E_p, $\underline{P}$ has rank p; and since $(\underline{I}-\underline{P})$ maps E_n onto E_{n-p}, $(\underline{I}-\underline{P})$ has rank n-p. A further study of $\underline{P}$ and $(\underline{I}-\underline{P})$ is made in §2 of Appendix A.

B. Analysis of $\underline{\varepsilon}$

Returning now to the definitions of $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ in par A, we see that from (4.1), (4.7)

$$\underline{\varepsilon}_{n-p} = \underline{y}-\underline{X}\hat{\underline{\beta}} = \underline{y}-\underline{P}\underline{y} = (\underline{I}-\underline{P})\underline{y} \tag{4.10}$$

Hence $\underline{\varepsilon}_{n-p}$ is in E_{n-p}. By construction of $\underline{\varepsilon}_p$ (as a linear combination of the columns of $\underline{X}$ in (4.3)) we find $\underline{\varepsilon}_p$ is in E_p. Hence by our observation in par A, the decomposition (4.6) of $\underline{\varepsilon}$ into $\underline{\varepsilon}_p$ and $\underline{\varepsilon}_{n-p}$ is unique.

Alternately, we can arrive at the decomposition of $\underline{\varepsilon}$ by, applying $\underline{P}$ to each side of (4.5), using (4.7), and (4.3) for $\underline{\varepsilon}_p$, along with (4.9a); we arrive at:

$$\underline{\varepsilon}_p = \underline{P}\underline{\varepsilon}. \tag{4.11}$$

Using (4.5) for $\underline{y}$ in the right equality of (4.10), and (4.9a), we have

$$\underline{\varepsilon}_{n-p} = (\underline{I}-\underline{P})\underline{\varepsilon} . \tag{4.12}$$

Equation (4.10) gives us the constructive definition of $\underline{\varepsilon}_{n-p}$ in terms of $\underline{y}$ alone (as a projection onto E_{n-p}), while (4.11), (4.12) let us see $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ as projections onto E_p, E_{n-p} of the noise vector $\underline{\varepsilon}$. Also, Eq. (4.7) says $\underline{X}\hat{\underline{\beta}}$ is the projection of $\underline{y}$ onto E_p.

C. Analysis of $\underline{y}$, $\hat{\underline{\beta}}$, and $\underline{X}\,\hat{\underline{\beta}}$

Returning to (4.2) we can by (4.12) write that as

$$\underline{y} = \underline{X}\,\hat{\underline{\beta}} + (\underline{I}-\underline{P})\underline{\varepsilon} \quad (=\underline{P}\underline{y} + (\underline{I}-\underline{P})\underline{y}) \tag{4.13}$$

and (4.4) by (4.11) as

$$\underline{X}\,\underline{\hat{\beta}} = X\,\underline{\beta} + \underline{P}\,\underline{\varepsilon} \quad . \tag{4.14}$$

Moreover, from (3.8), with (4.5), and the orthogonal decomposition of $\underline{\varepsilon}$,

$$\underline{\hat{\beta}} = \beta + (\underline{X}^T X)^{-1} \underline{X}^T \varepsilon = \beta + (\underline{X}^T X)^{-1} \underline{X}^T \varepsilon_p \quad . \tag{4.15}$$

Here, very clearly, we see the roles in describing $\underline{y}$, $\underline{X}\,\hat{\underline{\beta}}$ of the two error-vectors $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ in (4.13) and (4.14) and their relative orthogonality. In (4.15) we see $\hat{\underline{\beta}}$ as a random perturbation of $\underline{\beta}$ either via the full $\underline{\varepsilon}$ or via its projection $\underline{\varepsilon}_p$ on E_p.

5. Standard Form of the Regression Equation

We will show that the regression equation (1.1), i.e.,

$$\underline{y} = X\beta + \underline{\varepsilon} \ , \tag{5.1}$$

if we know $\underline{X}$ and the statistics of $\underline{\varepsilon}$, can always be reduced to the form where

$$\underline{X}^T \underline{X} = \underline{I}_p \tag{5.2}$$

and

$$\langle\underline{\varepsilon}\rangle = \underline{0}, \ \langle\underline{\varepsilon}\,\underline{\varepsilon}^T\rangle = \sigma^2 \underline{I}_n \ . \tag{5.3}$$

Here $\underline{I}_p$, $\underline{I}_n$ are identity matrices of dimension p, n respectively. In other words, the nxp data matrix $\underline{X}$ can, without loss of generality, be considered as a set of p column vectors, each column a time series, such that the ith column $\underline{x}_i$ and the j column $\underline{x}_j$ of $\underline{X}$ are uncorrelated and of unit length:

$$\underline{x}_i^T \underline{x}_j = \delta_{ij} \ , \quad i, j = 1, \ldots, p \ .$$

Moreover (5.3) states that without loss of generality the noise simultaneously with (5.2) can be of zero mean and uncorrelated with uniform variance σ^2. That is, by (3.3)

$$\langle \varepsilon_j \rangle = 0 , \quad \langle \varepsilon_i \varepsilon_j \rangle = \sigma^2 \delta_{ij} , \quad i,j = 1, \ldots, n .$$

The ensemble average operation < > is over some specified set of random variables, e.g., the set of normally distributed n dimensional vectors alluded to in the closing remarks of §2D.

A. Singular Decompositions of Matrices

To facilitate the proof of assertions (5.1)-(5.3) we pause to gather the essential elements needed in that proof. The material here is general and of potential use in studies of linear regression of dynamical systems.

If $\underline{C}$ is any pxp symmetric matrix, then a fundamental theorem of linear algebra states that there exist p orthonormal px1 vectors $\underline{e}_1, \ldots, \underline{e}_p$, which we can gather together in a pxp matrix $\underline{E} = (\underline{e}_1 \underline{e}_2 \ldots \underline{e}_p)$, and there exist p eigenvalues $\ell_1, \ldots, \ell_p$ which we can put in pxp diagonal matrix form $\underline{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_p)$, with the property that

$$\underline{C}\,\underline{E} = \underline{E}\,\underline{L} \tag{5.4}$$

where

$$\underline{E}^T \underline{E} = \underline{E}\,\underline{E}^T = \underline{I}_p .$$

Hence we can express $\underline{C}$ as

$$\underline{C} = \underline{E}\,\underline{L}\,\underline{E}^T . \tag{5.5}$$

If we write

$$\text{'}\underline{L}^{\frac{1}{2}}\text{'} \text{ for diag } (\ell_1^{\frac{1}{2}}, \ldots, \ell_p^{\frac{1}{2}}) \tag{5.6}$$

then (5.5) can be written

$$\underline{C} = (\underline{E}\,\underline{L}^{\frac{1}{2}})\,(\underline{L}^{\frac{1}{2}}\underline{E}^T) = (\underline{E}\,\underline{L}^{\frac{1}{2}})\,(\underline{E}\,\underline{L}^{\frac{1}{2}})^T \tag{5.7}$$

Hence if we write

$$\text{'}\underline{S}\text{'} \text{ for } \underline{E}\,\underline{L}^{\frac{1}{2}} \qquad (\text{pxp}) \tag{5.8}$$

we have found the square root of $\underline{C}$, in the sense that :

$$\boxed{\underline{C} = \underline{S}\,\underline{S}^T} \; . \tag{5.9}$$

Next, suppose that we have any nxp matrix $\underline{Y}$. Let $\underline{C} \equiv \underline{Y}^T\underline{Y}$. Hence $\underline{C}$ is a pxp symmetric matrix and by the preceding analysis it has an associated pxp eigenvector matrix $\underline{E}$ and pxp eigenvalue matrix $\underline{L}$ with the properties stated below (5.4). Thus we can write

$$\underline{Y} = \underline{Y}(\underline{E}\,\underline{E}^T) = (\underline{Y}\,\underline{E})\underline{E}^T \tag{5.10}$$

and

$$\text{'}\underline{A}\text{'} \text{ for } \underline{Y}\,\underline{E}, \qquad (\text{nxp})$$

and observe that, on using (5.4),

$$\underline{A}^T\underline{A} = (\underline{YE})^T(\underline{YE}) = \underline{E}^T(\underline{Y}^T\underline{Y})\underline{E} = \underline{E}^T(\underline{CE})$$

$$= \underline{E}^T(\underline{EL}) = \underline{L} \; .$$

Thus if we write

$$'\underline{X}' \text{ for } \underline{A}\,\underline{L}^{-\frac{1}{2}} \qquad (n\text{x}p)$$

(assuming $\underline{C}$ is positive definite, i.e., all ℓ_j are positive) then

$$\underline{A} = \underline{X}\,\underline{L}^{\frac{1}{2}}$$

and (5.10) becomes

$$\boxed{\underline{Y} = \underline{X}\,\underline{L}^{\frac{1}{2}}\,\underline{E}^T} \tag{5.11}$$

where

$$\underline{X}^T\underline{X} = (\underline{A}\,\underline{L}^{-\frac{1}{2}})^T(\underline{A}\,\underline{L}^{-\frac{1}{2}}) = \underline{L}^{-\frac{1}{2}}(\underline{A}^T\underline{A})\underline{L}^{-\frac{1}{2}} = \underline{I}_p$$

This factoring of $\underline{Y}$ is its *singular decomposition*, with the n x p matrix $\underline{A}$ comprising in its columns the *principal components* of $\underline{Y}$, i.e., (5.10) in the form

$$\boxed{\underline{Y} = \underline{A}\,\underline{E}^T} \tag{5.12}$$

is the *principal component* (or *empirical orthogonal function*) *decomposition* of $\underline{Y}$, with the orthonormal vectors of $\underline{E}$ the *empirical orthogonal functions* or *principal vectors* of $\underline{Y}$.

B. Uncorrelating the Noise $\underline{\varepsilon}$

To demonstrate that (5.1) can be written with (5.3), we proceed as follows. Suppose we have a linear regression equation in the form:

$$\underline{x} = \underline{W}\,\underline{\alpha} + \underline{\delta} \tag{5.13}$$

Where $\underline{W}$ is n x p, and $\underline{\delta}$ is n x 1 with the assumed known property

$$\langle \underline{\delta}\,\underline{\delta}^T \rangle = \sigma^2 \underline{V} \tag{5.14}$$

We observe first that by subtracting $\langle \underline{\delta} \rangle$ from each side of (5.13), we can, after relabeling, satisfy the left condition in (5.3). Now, clearly the n x n matrix $\underline{V}$ is symmetric. Then by (5.9) we can find its n x n square root $\underline{S}$ such that

$$\underline{V} = \underline{S}\,\underline{S}^T.$$

Assuming $\underline{V}$ is positive definite,* we multiply each side of (5.13) by $\underline{S}^{-1}$:

$$\underline{S}^{-1}\underline{x} = \underline{S}^{-1}\underline{W}\,\underline{\alpha} + S^{-1}\underline{\delta} \tag{5.15}$$

and observe that

---

* If this is not the case, we can also handle the slight complications arising therefrom. To do so here will cause too much of a digression from the main line of the development. The main point to note in this paragraph is that, in order to reach (5.18) below in practice, we must have in hand the matrix $\underline{V}$ in (5.14) in some form.

$$<(\underline{S}^{-1}\underline{\delta})(\underline{S}^{-1}\underline{\delta})^T> = <\underline{S}^{-1}(\underline{\delta}\ \underline{\delta}^T)(\underline{S}^{-1})^T>$$

$$= \underline{S}^{-1}<\underline{\delta}\ \underline{\delta}^T>(\underline{S}^T)^{-1}$$

$$= \underline{S}^{-1}(\sigma^2\underline{V})(\underline{S}^T)^{-1}$$

$$= \sigma^2\underline{S}^{-1}\ \underline{S}\ \underline{S}^T(\underline{S}^T)^{-1}$$

$$= \sigma^2\underline{I}_n ,$$

as was to be shown in (5.3). Thus writing

$$'\underline{y}' \text{ for } \underline{S}^{-1}\underline{x}$$

$$'\underline{Y}' \text{ for } \underline{S}^{-1}\underline{W}$$

and

$$'\underline{\varepsilon}' \text{ for } \underline{S}^{-1}\underline{\delta} ,$$

(5.15) becomes

$$\underline{y} = \underline{Y}\ \underline{\alpha} + \underline{\varepsilon} \tag{5.18}$$

where $\underline{\varepsilon}$ has the property (5.3). Moreover, $\underline{\alpha}$ may be estimated via

$$\hat{\underline{\alpha}} = [\underline{Y}^T\underline{Y}]^{-1}\underline{Y}^T\underline{y} = [\underline{W}^T\underline{V}^{-1}\underline{W}]^{-1}(\underline{W}^T\underline{V}^{-1})\underline{x} \tag{518a}$$

Observe that $\underline{\alpha}$ in the noise-free case is in principle unaffected by pre multiplying

(5.13) by $\underline{S}^{-1}$. Hence in the case of no noise, (5.18a) should recover $\underline{\alpha}$ exactly.

C. Orthonormalizing the Data Matrix

Using the decomposition of $\underline{Y}$, given by (5.11), in (5.18), we can transform (5.18) to:

$$\underline{y} = \underline{X}\ \underline{\beta} + \underline{\varepsilon} \tag{5.19}$$

where we have

$$'\underline{\beta}' \text{ for } \underline{L}^{\frac{1}{2}}\underline{E}^{T}\underline{\alpha} \tag{5.20}$$

and where $\underline{X}$, $\underline{L}$ and $\underline{E}$ are as given in the preceding discussion of the singular decomposition of $\underline{Y}$ in par. A. Hence in (5.19)

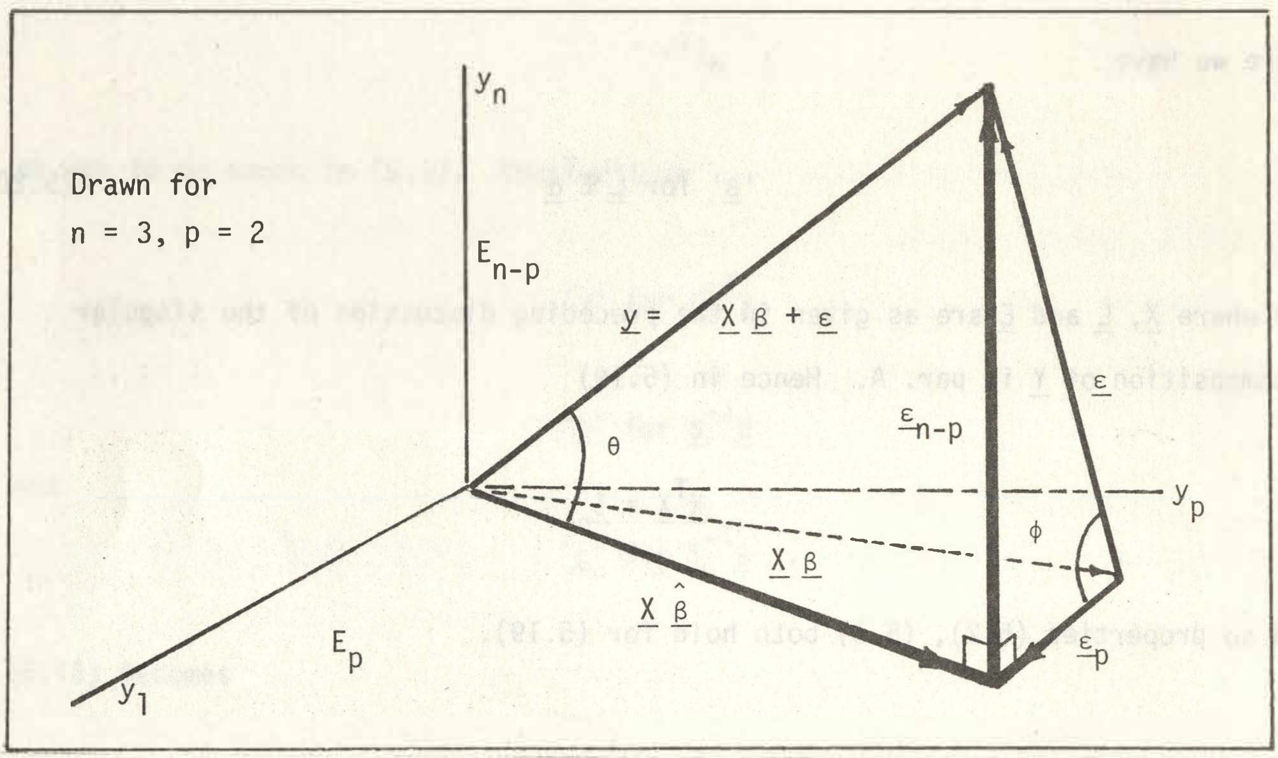
$$\underline{X}^{T}\underline{X} = \underline{I}_{p}$$

and so properties (5.2), (5.3) both hold for (5.19).

6. Geometry of Linear Regression

The analysis of the residual noise in §4 led to the introduction of a projection operator $\underline{P}$ whose geometrical interpretation suggests the following imagery in connection with linear regression studies.

The diagram below is drawn for the case of n = 3, p = 2. However, it contains all the essential elements of the general case and is labeled to suggest the general case.



A. Euclidean Geometry of the Diagram

Every formula in §4 and derivation there may be interpreted in the light of this diagram; and other formulas and definitions may be read directly from it prior to formal proofs or definitions. For example, from Pythagoras' theorem and the orthogonality of the pair $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$, and the orthogonality of the pair $\underline{X}\,\hat{\underline{\beta}}$, $\underline{\varepsilon}_{n-p}$, we find that

$$||\underline{y}||^2 = ||\underline{X}\ \hat{\underline{\beta}}||^2 + ||\underline{\varepsilon}_{n-p}||^2 \qquad (\theta \text{ triangle}) \qquad (6.1)$$

$$||\underline{\varepsilon}||^2 = ||\underline{\varepsilon}_p||^2 + ||\underline{\varepsilon}_{n-p}||^2 \qquad (\phi \text{ triangle}) \qquad (6.2)$$

These relations are read directly from the two right triangles in the figure (labeled via angles θ, ϕ). They are proved in general by means of the representation of the vectors on the right sides as appropriate projections via $\underline{P}$ or $(\underline{I}-\underline{P})$ of the vectors on the left sides; and then using (4.9), Another relation, based on the ϕ triangle, and (6.2), is:

$$||\underline{y}-\underline{X}\ \underline{\beta}||^2 = ||\underline{X}(\hat{\underline{\beta}}-\underline{\beta})||^2 + ||\underline{y}-\underline{X}\ \hat{\underline{\beta}}||^2 \quad (\phi \text{ triangle}) \qquad (6.3)$$

This relation shows that $||\underline{y}-\underline{X}\ \underline{\beta}||^2$ attains a minimum when $\underline{\beta}=\hat{\underline{\beta}}$ for a given $\underline{y}$, $\underline{X}$, n and p.

B. Kinematics of the Diagram

The kinematic aspects and random aspects of linear regression stand out in the diagram. Thus $\underline{X}\ \underline{\beta}$ is the underlying fixed signal which is perturbed by random additions of $\underline{\varepsilon}$, so that we may watch the random variable $\underline{y}$ twitter about as successive realizations of $\underline{\varepsilon}$ are added to the fixed vector $\underline{X}\ \underline{\beta}$. Our estimate $\underline{X}\ \hat{\underline{\beta}}$ of the underlying signal is also a random variable, its wanderings over the space E_p being propelled by the random vector $\underline{\varepsilon}_p$ in E_p. As we saw in (4.11), $\underline{\varepsilon}_p$ is the projection of $\underline{\varepsilon}$ onto E_p. These images suggest that the pair $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ and the pair $\underline{X}\ \hat{\underline{\beta}}$, $\underline{\varepsilon}_{n-p}$ are each independent pairs of random variables. These facts are borne out in our statistical studies in §§2, 5 of Appendix A and form the basis of the probability density derivations occupying the main portion of the study below.

C. Fixed-$X\beta$ Interpretation of the Diagram

In the interpretation of the diagram above it should be kept in mind that the diagram is for a random noise $\underline{\varepsilon}$ associated with a *fixed* $\underline{X}\ \underline{\beta}$ vector — a *fixed* signal associated with a specific set $\underline{y} = (y_1, \ldots, y_n)^T$ of observations and set $\underline{X} = (\underline{x}_1, \ldots, \underline{x}_p)$ of forcing field data. If we go on to a new set $\underline{y}$ and $\underline{X}$ down the time stream (say) it is possible that the pdf governing the residual noise vector $\underline{\varepsilon}$ (as discussed in §4) will be different. If that is the case, the successive realizations of $\underline{\varepsilon}$ in the diagram may be distributed quite differently relative to the first diagram. Thus it is generally not possible to associate the same diagram above with two successive (n-sample, p-predictor) experiments.

D. Definition of a Stationary Setting for the Diagram

We may put the preceding observation in perspective by stating it in a positive rather than negative way. If we have two or more successive (n-sample, p-predictor) experiments, and conducted in a milieu where the pdf of $\underline{\varepsilon}$ is the same for all experiments and so that the ratio $||\underline{X\beta}||^2/\sigma^2$ is the same in each experiment, then the same diagram holds for all the experiments. In this sense we may say that the random noise vector (or its pdf) is *stationary*, and that the experiments of the type (n-sample, p-predictor), occur in a *stationary setting*. This situation could arise in practice, and its earmark would be a definitive spread of $\underline{\varepsilon}$ vectors (as found in (4.6)) which, via a successful statistical test, are all judged to belong to the same population.

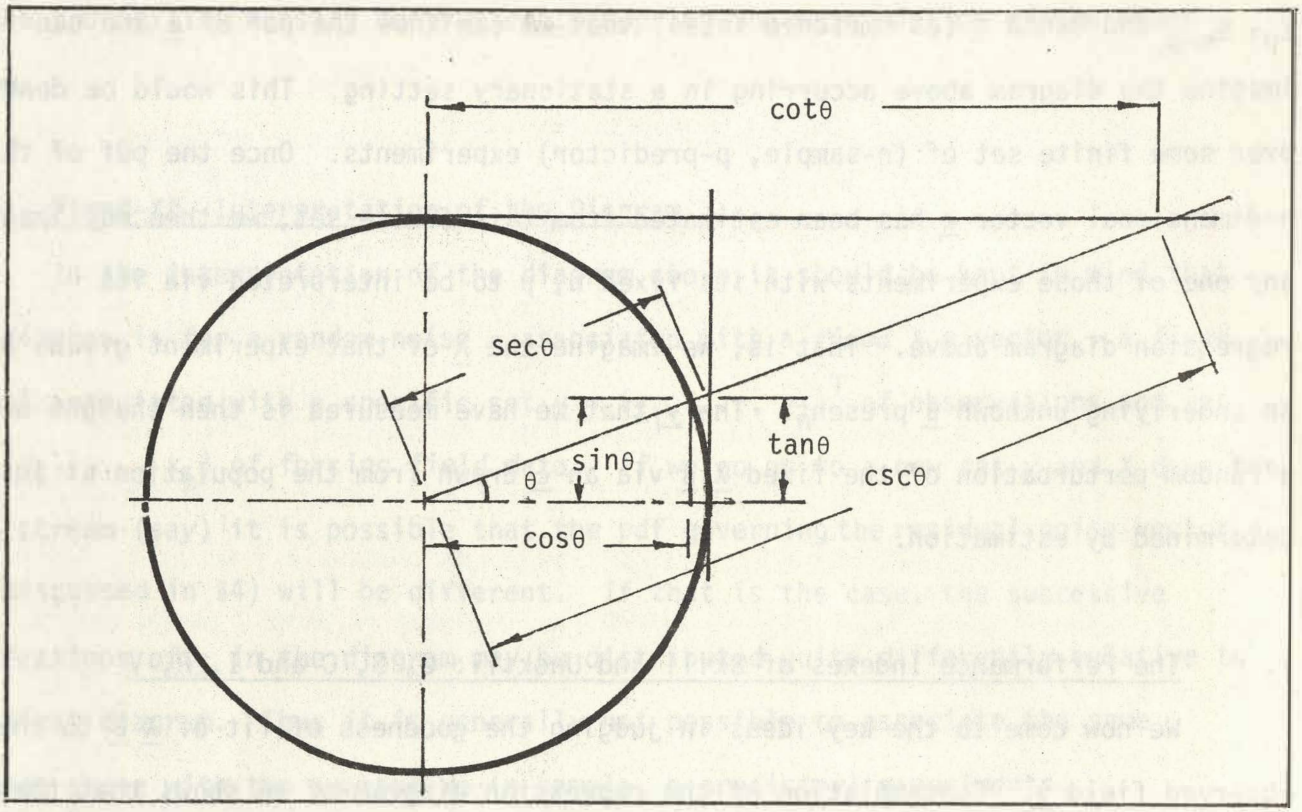
E. Determining Stationarity of a Setting-- The Associated Fixed-$X\beta$ Interpretation of the Diagram

It is, in the last analysis, only by direct experimental determination of

$\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ and hence $\underline{\varepsilon}$ (as sketched in §4) that we can know the pdf of $\underline{\varepsilon}$ and can imagine the diagram above occurring in a stationary setting. This would be done over some finite set of (n-sample, p-predictor) experiments. Once the pdf of the n-dimensional vector $\underline{\varepsilon}$ has been estimated from this finite set, we then may imagine any one of those experiments with its fixed n, p to be interpreted via its regression diagram above. That is, we imagine the $\underline{X}$ of that experiment given, and an underlying unknown $\underline{\beta}$ present. The $\underline{y}$ that we have measured is then thought of as a random perturbation of the fixed $\underline{X}\,\underline{\beta}$ via an $\underline{\varepsilon}$ drawn from the population as just determined by estimation.

7. The Performance Indexes of Skill and Unskill: Q, S, C and I, R, V

We now come to the key ideas in judging the goodness of fit of $\underline{X}\,\hat{\underline{\beta}}$ to the observed field $\underline{y}$. Contemplation of the regression diagram of §6 shows that the smaller $||\underline{\varepsilon}_{n-p}||$ is, all other things (n, p) the same, the better is the regression fit of $\underline{X}\,\hat{\underline{\beta}}$ to $\underline{y}$. In other words, the *smaller* θ *is, the better is the fit*. An intuitively desirable skill index would then increase as θ decreases. In order for the skill index to be free of units and scale sizes when describing goodness of fit we can adopt ratios of the lengths of various portions of the diagram to reflect the skill of the fit. The most natural candidates for such skill ratios are the trigonometric functions associated with the θ triangle. There are six trigonometric functions associated with θ (see the mnemonic diagram below): three of them decrease as θ decreases; namely, $\sin\theta$, $\tan\theta$, $\sec\theta$; and three increase as θ decreases, namely $\cos\theta$, $\cot\theta$, $\csc\theta$. It is this behavior of the latter three that suggests adopting them as skill indexes, and



hence assigning the former three as unskill indexes. The table below summarizes these definitions and the names and symbols we attach to them in order to facilitate discussion of their statistical properties and conventions later in this study. We use the *squares* of the trig functions because of the relatively simple algebraic and occasionally linear connections between them.

HINDCAST PERFORMANCE INDEXES

	Symbol	Name	Trig Analog	Basic Definition	Connections	pdf Ref.
Skills	Q	canonic skill	$\cot^2\theta$	$\lVert\underline{X}\,\hat{\underline{\beta}}\rVert^2/\lVert\underline{\varepsilon}_{n-p}\rVert^2$	$Q = \frac{S}{1-S} = C-1$	(8.1), (A48)
	S	classic skill	$\cos^2\theta$	$\lVert\underline{X}\,\hat{\underline{\beta}}\rVert^2/\lVert\underline{y}\rVert^2$	$S = 1-R = \frac{Q}{1+Q}$	(8.7), (A51)
	C	coskill	$\csc^2\theta$	$\lVert\underline{y}\rVert^2/\lVert\underline{\varepsilon}_{n-p}\rVert^2$	$C = \frac{1}{1-S} = 1+Q$	(A48)
Unskills	I	ineptness	$\tan^2\theta$	$\lVert\underline{\varepsilon}_{n-p}\rVert^2/\lVert\underline{X}\,\hat{\underline{\beta}}\rVert^2$	$I = \frac{1-S}{S} = U-1$	(8.4), (A49)
	R	residual unskill	$\sin^2\theta$	$\lVert\underline{\varepsilon}_{n-p}\rVert^2/\lVert\underline{y}\rVert^2$	$R = 1-S = \frac{I}{1+I}$	(A51)
	U	unskill	$\sec^2\theta$	$\lVert\underline{y}\rVert^2/\lVert\underline{X}\,\hat{\underline{\beta}}\rVert^2$	$U = \frac{1}{S} = 1+I$	(A49)

By using the various connections between the θ triangle and the noise components $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$, the basic definitions above can be given numerically equivalent forms. For example, we can also write Q as $\lVert\underline{X}\,\hat{\underline{\beta}}\rVert^2/\lVert\underline{y}-\underline{X}\,\hat{\underline{\beta}}\rVert^2$ using (4.2). In this way Q becomes directly computable from the observed field $\underline{y}$ and the data field $\underline{X}$, where $\hat{\underline{\beta}}$ is of course given by (3.8). From Q the remaining two skills follow by the indicated connections. Similarly, the ineptness I is simply the reciprocal of canonic skill Q,hence directly computable, and so the remaining unskills are readily forthcoming from I (and hence ultimately Q).

From a statistical aspect, the most basic of skill indexes is the canonic skill. Its probability density function (pdf), as we shall see in Appendix A, follows most simply from that of the residual noise vector $\underline{\varepsilon}$, the fountainhead of all the pdfs in linear regression theory. Moeover Q's mean and variance alone have simple closed expressions. All other five pdfs could follow (if one chose) from Q's alone by simple geometric and analytic considerations. There are three natural pairs among the six indexes: (Q, C), (S, R), (I, U): Since Q and C are simply related by a *linear* relation we need only study Q. Moreover, since S, R, and I, U are also *linearly* related pairs, we need only study (say) S, and I. We are particularly interested in Q and its arithmetic inverse I; their relation is not as simple as the linear relations among the three natural pairs.

The presence of S in the basic triplet Q, I, S and in the connecting relations was singled out (from the various other possibilities) because S is the classic skill index initiated by Lorenz, and later studied by Davis, and Barnett and Hasselmann.

8. Probability Density Functions for Q, I, and S: Their interpretation and their behavior

The probability density functions of the performance indexes allow us to see at a glance where the indexes mostly dwell in their respective ranges; they allow us to easily and exactly compute the means and variances of the indexes; they allow us to construct confidence regions for estimates and they generally allow us to theorize about statistical questions arising in regression studies of physical fields. Once the probability density of the noise vector $\underline{\varepsilon}$ is determined, the density of each index is fixed. In this study we have chosen the normal law governing $\underline{\varepsilon}$ because of its relatively frequent occurrence in natural phenomena and because of its mathematical tractability.* The details of the derivations of the six indexes based on the normal law for $\underline{\varepsilon}$, are given in Appendix A. The treatment there is rigorous, and essentially complete. In this section we single out for discussion three of the indexes, namely Q, I and S. The reasons for these choices were explained in §7. Throughout the discussions below, $\lambda = ||\underline{X}\ \underline{\beta}||^2/\sigma^2$ (signal to noise ratio), n = sample size of an experiment, p = number of predictors in an experiment.

---

* The reason for this choice is given just below (2.7).

A. Pdf and Moments of Canonic Skill Q (cf. (A48), (A55), (A58))

$$P_Q(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{x^{r+\frac{1}{2}p-1}}{(1+x)^{r+\frac{1}{2}n}}, \quad n>p\geq 1 \cdot \tag{8.1}$$

$(o \leq x < \infty)$

$$\mu_Q = \frac{\lambda+p}{n-p-2}, \quad n - p > 2 \tag{8.2}$$

$$\sigma^2_Q = \frac{2[\lambda^2+(n-2)(2\lambda+p)]}{[n-p-2]^2[n-p-4]}, \quad n - p > 4 \tag{8.3}$$

B. Pdf and Moments of Ineptness I (cf (A49), (A59), (A60))

$$P_I(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda)^s}{s!} \cdot \frac{\Gamma(s+\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n-p))\Gamma(s+\frac{1}{2}p)} \cdot \frac{x^{\frac{1}{2}(n-p)-1}}{(1+x)^{s+\frac{1}{2}n}}, \quad n>p\geq 1 \tag{8.4}$$

$(o \leq x < \infty)$

$$\mu_I = (n-p)e^{-\frac{1}{2}\lambda} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda)^s}{s!} \cdot \frac{1}{2s+p-2}, \quad p > 2 \ \text{(1st raw moment, } \mu'_1) \tag{8.5}$$

$$\mu'_2 = (n-p)[n-p+2]e^{-\frac{1}{2}\lambda} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda)^s}{s!} \cdot \frac{1}{[2s-p-2][2s+p-4]}, \quad p>4 \ \text{(2nd raw moment)} \tag{8.6}$$

The variance doesn't appear to have a simple closed form, and so $\sigma_I^2 = \mu_2 = \mu'_2 - \mu'^2_1$ may be determined numerically.

C. <u>pdf and Moments of Classic Skills (cf (A51), (A64), (A65))</u>

$$P_S(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot x^{r+\frac{1}{2}p-1}(1-x)^{\frac{1}{2}(n-p)-1}, \quad n>p\geq 1 \quad (0\leq x\leq 1) \tag{8.7}$$

$$\mu_S = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \left[\frac{2r+p}{2r+n}\right] \quad \text{(1st raw moment, } \mu'_1\text{)} \tag{8.8}$$

$$\mu'_2 = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{(2r+2+p)(2r+p)}{(2r+2+n)(2r+n)} \quad \text{(2nd raw moment)} \tag{8.9}$$

The variance is computed via $\sigma_S^2 = \mu'_2 - \mu'^2_1$.

For small signal to noise ratio λ :

$$\mu_S \cong (1-\frac{\lambda}{2})\frac{p}{n} + \frac{\lambda}{2}\left(\frac{p+2}{n+2}\right) \quad \text{(to first order in } \lambda\text{)} \tag{8.10}$$

$$= S_o + \tfrac{1}{2}\lambda(1-S_o), \quad S_o \equiv p/n.$$

Hence S_o is the mean value of S for the case $\lambda = 0$.

For small signal to noise ratio λ:

$$\sigma_S^2 \cong \frac{2(n-p)}{n(n+2)}\left[\frac{p}{n} - 2\lambda \cdot \frac{(n+np+p)}{n(n+4)}\right] \quad \text{(to first order in } \lambda\text{)} \tag{8.11}$$

$$= \frac{2(1-S_o)}{n+2}\left[S_o - 2\lambda\frac{(1+p+S_o)}{n+4}\right], \quad S_o = p/n .$$

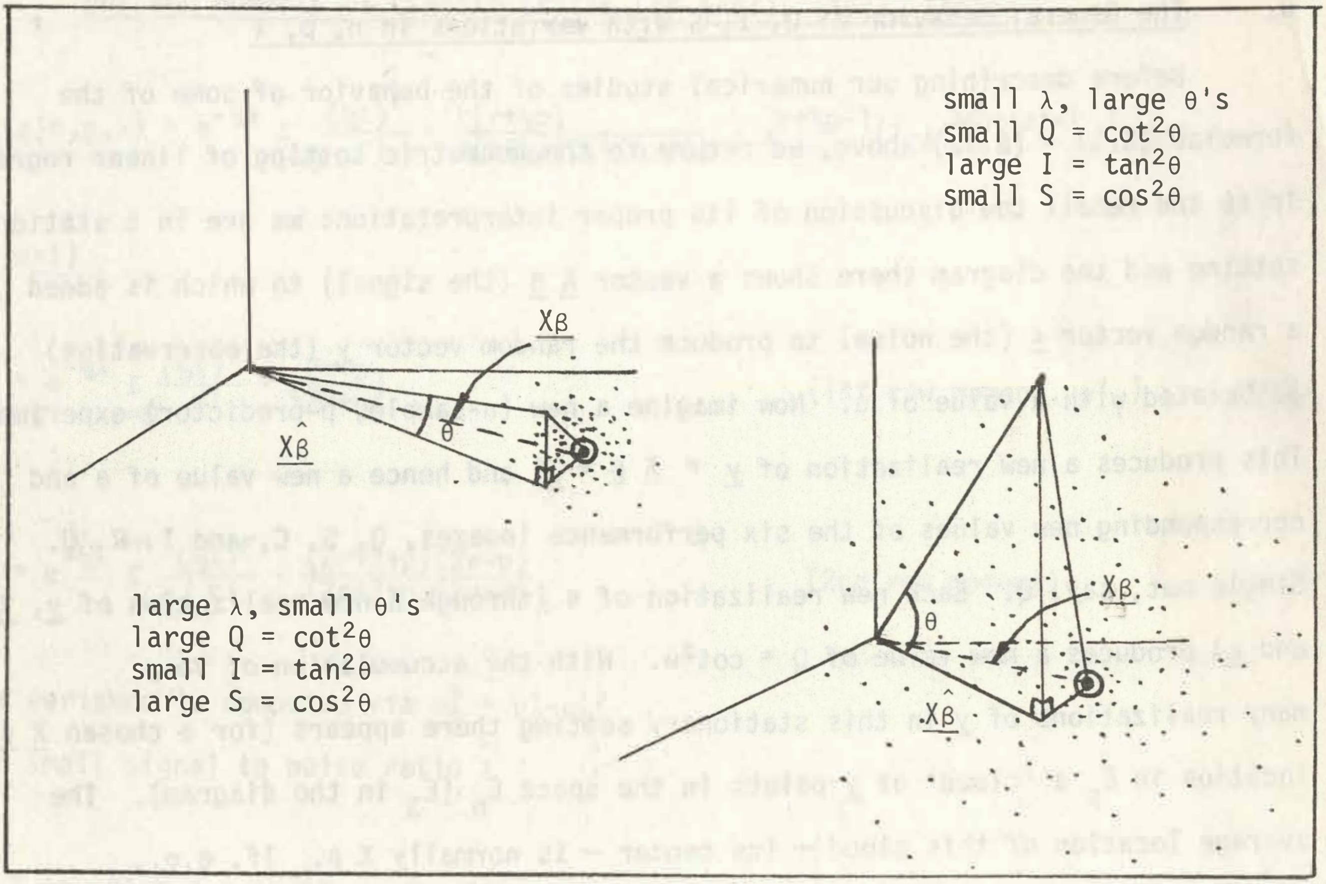
For $\lambda = 0$, the exact result holds:

$$\sigma_S^2 = \frac{2(1-S_o)S_o}{n+2} = \frac{2(1-p/n)(p/n)}{n+2} \tag{8.12}$$

D. The General Behavior of Q, I, S with variations in n, p, λ

Before describing our numerical studies of the behavior of some of the formulas (8.1) - (8.12) above, we return to the geometric setting of linear regression in §6 and recall the discussion of its proper interpretation: we are in a stationary setting and the diagram there shows a vector $\underline{X}\ \underline{\beta}$ (the signal) to which is added a random vector $\underline{\varepsilon}$ (the noise) to produce the random vector $\underline{y}$ (the observation) associated with a value of θ. Now imagine a new (n-sample, p-predictor) experiment. This produces a new realization of $\underline{y} = \underline{X}\ \underline{\beta} + \underline{\varepsilon}$) and hence a new value of θ and corresponding new values of the six performance indexes, Q, S, C, and I, R, U. Single out, say, Q. Each new realization of θ (through a new realization of $\underline{y}$, $\underline{X}$ and $\underline{\varepsilon}$) produces a new value of $Q = \cot^2\theta$. With the accumulation of very many realizations of $\underline{y}$ in this stationary setting there appears (for a chosen $\underline{X}\ \underline{\beta}$ location in E_p a 'cloud' of $\underline{y}$ points in the space E_n (E_3 in the diagram). The average location of this cloud — its center — is normally $\underline{X}\ \underline{\beta}$. If, e.g., $\underline{\varepsilon} \sim N_n\ (\underline{0},\sigma^2 \underline{I}_n)$, then the center is $\underline{X}\ \underline{\beta}$ and its size is governed by the size of σ^2. Thus the cloud will hover very near the plane E_p (E_2 in the diagram) if σ^2 is much smaller than $||\underline{X\beta}||^2$, i.e., if $\lambda = ||\underline{X\beta}||^2/\sigma^2$ is large. The value of θ for such a cloud will always be near 0 and so the associated sprinkling of the values $\cot^2\theta$ on the real line will be located a large distance from the origin. That is, for a large signal to noise ratio, canonic skill Q will tend to be large.

Returning to (8.2) we see that it corroborates our preceding conclusion that the average value of Q increases with λ for given fixed n, p. The diagrams below sketch the two clouds of $\underline{y}$ points for cases of small and large λ.



From these diagrams we see that ineptness I, for given n, p, decreases with increasing λ while both canonic skill Q and classic skill S increase. When $\lambda = 0$, the cloud engulfs the origin of the diagram and θ often is in the vicinity of 90°. The average values of Q, I, S in this case are easy to reckon:

$$\mu_Q = \frac{p}{n-p-2} \qquad n-p > 2 \tag{8.13}$$

$$\mu_I = \frac{n-p}{p-2} \qquad \lambda = 0\ ,\ n-p > 0\ ,\ p > 2 \tag{8.14}$$

$$\mu_S = \frac{p}{n} \qquad n-p > 0 \tag{8.15}$$

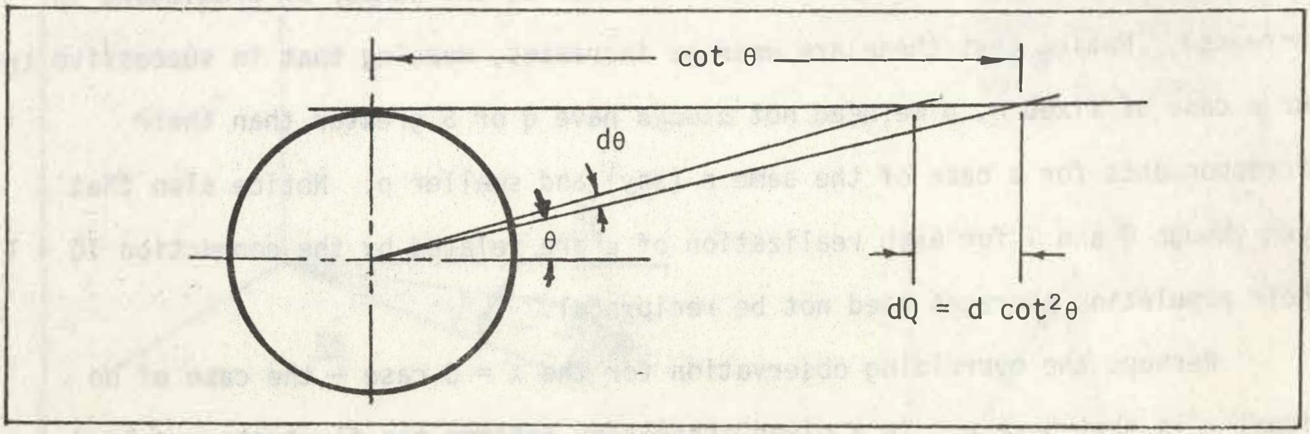
Observe that μ_Q, μ_S increase as p increases for fixed n showing that, in a stationary setting, hindcast skill on the average increases as the number of predictors is increased. Notice that these are *average* increases, meaning that in successive trials for a case of fixed n, p we need not *always* have Q or S greater than their correspondents for a case of the same n (say) and smaller p. Notice also that even though Q and I for each realization of $\underline{\varepsilon}$ are related by the connection IQ = 1, their population averages need not be reciprocal.

Perhaps the overriding observation for the λ = 0 case — the case of no signal — is that S, e.g., in a given stationary setting can fluctuate and land anywhere in its domain (0, 1) as we perform hindcast experiments in that setting. That is, just because there is no signal, the value of S need not always be 0. As (8.15) states, the average value of S is p/n. Similarly, Q need not always be zero, and the closer p is to n (within the stated condition n-p>2) the higher is the average value of Q. Even ineptness when there is no signal, can be brought quite low on the average over a set of successive experiments in a stationary setting by making p sufficiently near n.

E. Study of Some Specific Examples of the pdf's of Q, I, and S

1) We consider first the properties of the canonic skill Q. Figure Q-0 shows plots of (8.1) for the case of n = 10, p = 5 as λ takes on the five values λ = 0, 1, 2, 5, 10. The horizontal axis from 0 to ∞ is the range of Q (=x in (8.1)). The vertical axis is the probability (density) of Q. The area under each curve is of course unity. By (8.2) the area of each curve is balanced around $\mu_Q = (\lambda+p)/(n-p-2)$. Thus for the λ = 0 curve, the mean of Q is at 5/(10-5-2) = 5/3 = 1.67. We see that, as λ increases, the main mass of a distribution moves to larger Q values until at λ = 10, μ_Q is at (10+5)/(10-5-2) = 15/3 = 5. At the same time it is clear that the variance, or spread of the mass about μ_Q increases as λ increases. From (8.3) we find that $\sigma_Q^2 = 80/9$ for λ = 0, and $\sigma_Q^2 = 600/9$ for λ = 10.

This enormously accelerated spread of Q as λ increases is understandable from the unit circle diagram for $\cot^2\theta$ (cf §7).



As we saw above in paragraph D, small θ means large $\cot^2\theta = Q$. At the same time, small random changes in small θ can result in enormous changes in $\cot^2\theta$. Hence as λ increases, and σ^2 is fixed, the vector $\underline{y}$ is drawn down to E_p and held there at small θ on the average. But now the random perturbations of $\underline{X}\,\underline{\beta}$ by $\underline{\varepsilon}$ produce relatively great changes in Q from one realization to the next, i.e., $dQ = d\cot^2\theta = -2\cot\theta\csc^2\theta d\theta$. This sensitivity of Q at small θ (high Q) to changes in θ could be used to test effects of changes in p on a hindcast. Figs. Q-1 to Q-5 show the rapid shift in probability mass as p increases from 1 to 7 for fixed $n = 10$, for all five cases of λ from 0 to 10 shown. The graphs warn us at the same time about the relatively great spreads of Q readings possible when n and p are relatively close. Notice in particular as in Fig. Q-5 the spread in Q when $n = 10$, $p = 7$. This is anticipated from (8.3) by the presence of n-p in *both* factors in the denominator. This spread increases with increasing λ as seen in both sets of Figures Q-1 to Q-5, and Q-6 to Q-10. This spread is dramatically smaller when $\lambda = 0$, say. *Hence a tight cluster of Q readings around 0 indicates poor hindcast fits in a low signal to noise setting. The larger the λ the larger will be the spread of Q, and the better the fittings on the average.*

2) Consider next the properties of ineptness. Fig. I-0 should be compared with Fig. Q-0. The curves present clearly inverse characters. Now ineptness quickly decreases in Fig. I-0 as λ increases from 0 to 10 in five cases. The spread of I decreases as λ increases. *A tight set of I readings around 0 indicates good hindcast fits in a high signal to noise setting. The smaller the λ, the larger will be the spread of I, and the less good the fittings on the average.* The sharp rise of the pdf for I in Fig. I-1 for the case n = 4, p = 3 is indicative of a singularity at I = 0, as may be seen from (8.4). For in this case we have $\frac{1}{2}(n-p)-1 = -\frac{1}{2}$, so $P_I(x|4, 3, 0) \to \infty$ as $x \to 0$, but in an integrable way so that the area under $P_I(x|4, 3, 0)$ is still 1. Observe also that for n = 5, p = 3 we have $\frac{1}{2}(n-p)-1 = 0$, and so $P_I(x|5, 3, 0) \to a \neq 0$, i.e., its limit is a finite nonzero quantity. (The high-rise curve in Fig. Q-1 is an example of Q's singularity for p = 1. This is P_Q's only singularity, while P_I has one whenever n-p = 1).

3) Consider finally the classic skill S.

Fig. S-0 contains curves of $P_I(10, 5, \lambda)$ for five choices of λ = 0, 1, 2, 5, 10. The curves were drawn from numerical values based on (8.7). The range of S is (0, 1). The curve for $\lambda = 0$ is symmetric whenever n = 2p and of the general form:

$$P_S(x|n,p,0) = \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} x^{\frac{1}{2}p-1}(1-x)^{\frac{1}{2}(n-p)-1} \quad (8.7 \text{ with } \lambda=0)$$

As $n \to \infty$, and we fix $p/n = S_o$, the mean $\mu_S = S_o$ stays fixed and curve becomes more peaked (cf (8.12)) and can be shown to approach gauss' curve. In general, for any λ as $n \to \infty$ and we fix $p/n = S_o$, the curves will approach the gaussian bell shaped curve. This follows from an examination of the higher moments and the central limit theorem. In general, for fixed n, p, as λ increases, the mass of the S readings shifts toward 1, as expected. In the sets of curves shown in Figs. S-1 to S-5, we see the effect of increasing λ on moving the originally disparate curves in Fig. S-1 to near conformity in high skill in Fig. S-5. In Fig. S-1, incidentally, observe how for $p = 1$, $n = 10$ the mass of S is very close to 0. As p goes up through the ranks through 2, 3, 5 and 7, the curves' maxima move steadily toward 1. In the set of Figures S-6 to S-10 we watch the effect of increasing λ on various choices of n for fixed $p = 3$. The curve for $n = 4$, $p = 3$ in Fig. S-6 has an integrable singularity at $x = 1$, as may be seen from (8.7). There is also a singularity of P_S when $p = 1$ (see Fig. S-1 and the interesting case of $p = 1$ in Fig. S-4). The set of curves, S-6 to S-10, as well as S-1 to S-5, are particularly instructive in showing how, for fixed p, the classic skill deteriorates as n becomes larger, regardless of the size of λ. The latter, to be sure, for large λ, holds back this deterioration as n increases, but only by varying amounts does it stay the inevitable decrease of the average S to zero. Equation (8.10) expresses this phenomenon succinctly, but only approximately and for λ not too large.

9. <u>The Mean Signal to Noise Ratio $\bar{\lambda}$</u>

A. <u>Introduction:</u> The signal to noise ratio $\lambda = ||\underline{X}\underline{\beta}||^2/\sigma^2$ ostensibly depends on the data matrix $\underline{X}$ and the underlying physical process'Greens' functions (cf §2). It also depends on the dimensions n, p of $\underline{X}$ and $\underline{\beta}$. We shall now show that under normal working conditions we cannot let λ and p vary independently of each other without incurring problems of interpretation and application of the theory of the performance index pdfs studied in §8. It will be recalled that in §8 we allowed all three parameters n, p, λ to vary independently as we explored the geometry of the regression setting. This was permissible in that more or less abstract setting. But now we consider λ, as defined, and the implications of its connections to $\underline{X}$ and $\underline{\beta}$. This will lead to the introduction of $\bar{\lambda} = \lambda/p$.

B. Principal Representation of λ: Using the theory of §5, let ℓ_j, j = 1, ..., p, and $\underline{e}_j$, j = 1, ..., p be the eigenvalues and eigenvectors of the symmetric matrix $\underline{X}^T\underline{X}$. Define the nxp amplitude matrix $\underline{A} \equiv \underline{X}\,\underline{E}$, where $\underline{E} = [\underline{e}_1, \ldots, \underline{e}_p]$, and then the nxp basis matrix $\underline{B} \equiv \underline{A}\,\underline{L}^{-\frac{1}{2}}$, where $\underline{L}$ = diag $(\ell_1, \ldots, \ell_p)$. If $\underline{A} = [\underline{a}_1, \ldots, \underline{a}_p]$ and $\underline{B} = [\underline{b}_1, \ldots, \underline{b}_p]$, then we have respectively the principal component and singular decomposition representations of $\underline{X}$:

$$\underline{X} = \underline{A}\underline{E}^T = \underline{B}\underline{L}^{\frac{1}{2}}\underline{E}^T \tag{9.1}$$

which in vector form become

$$\underline{X} = \sum_{j=1}^{p} \underline{a}_j\underline{e}_j^T = \sum_{j=1}^{p} \ell_j^{\frac{1}{2}}\underline{b}_j\underline{e}_j^T \tag{9.2}$$

where $\underline{B}$, $\underline{E}$ are orthogonal matrices, i.e.,

$$\underline{e}_i^T\underline{e}_j = \delta_{ij} \;, \quad \underline{b}_i^T\underline{b}_j = \delta_{ij} \;, \quad i, j = 1, \ldots, p \tag{9.3}$$

The vectors $\underline{B} = [\underline{b}_1, \ldots, \underline{b}_p]$ are an orthonormal basis of E_p. We use them in §2 of appendix A. (For simplicity we drop the subscript p from $\underline{B}_p$).

We may go on to use this representation to write

$$\underline{X\beta} = \sum_{j=1}^{p} \ell_j^{\frac{1}{2}}\underline{b}_j(\underline{e}_j^T\underline{\beta}) = \sum_{j=1}^{p} \ell_j^{\frac{1}{2}}\underline{b}_j\beta_j$$

where $\beta_j \equiv \underline{e}_j^T\,\underline{\beta}$ is the jth component of $\underline{\beta}$ relative to the basis $\underline{E}$, the one used to give EOF representations of the spatial extent of the data matrix. The quantity $||\underline{X\beta}||^2$ used in the signal to noise definition can now be written (using (9.3)) as:

$$||\underline{X\beta}||^2 = (\underline{X\beta})^T(\underline{X\beta}) = (\sum_{j=1}^{p} \ell_j^{\frac{1}{2}}\beta_j\underline{b}_j)^T(\sum_{k=1}^{p} \ell_k^{\frac{1}{2}}\beta_k\underline{b}_k)$$

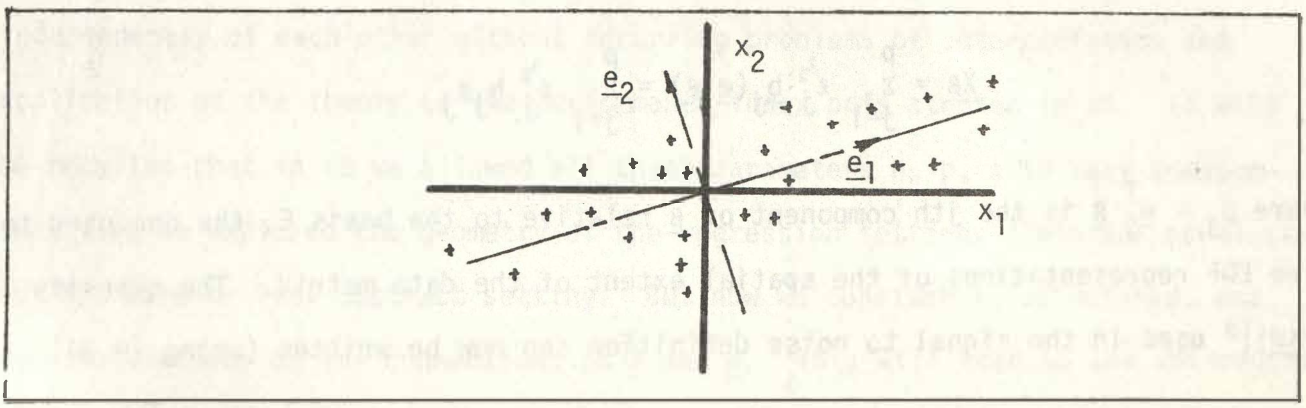
$$= \sum_{j=1}^{p} \ell_j\beta_j^2 \qquad (9.4)$$

Hence we derive at the *principal representation* of λ:

$$\boxed{\lambda = ||\underline{X\beta}||^2/\sigma^2 = \sum_{j=1}^{p} (\ell_j/\sigma^2)\beta_j^2} \qquad (9.5)$$

C. Geometric interpretation of the principal representation of λ.

The representation in (9.5) has the following geometric interpretation, relative to the linear regression diagram in §6. On the one hand the n dimensionality of the diagram in §6 arises from the sample size n taken in gathering up the n components y_j of $\underline{y}$. On the other, the p points in space (over the ocean, atmosphere, etc.) where those $\underline{n}$ samples are taken have, at any moment, associated with them p values x_{ij}, j = 1, ..., p (a row of $\underline{X}$) which we could plot as a point in a p dimensional space. There would be n of those p dimensional points (or vectors), and we schematically show them in the diagram below for the case p = 2.



We show in particular the two basis vectors $\underline{e}_1$, $\underline{e}_2$ which resolve the n row-vectors of the data nxp matrix $\underline{X}$ into their principal components. The ℓ_1 and ℓ_2 are the variances of the data set along these orthogonal principal axes $\underline{e}_1$, $\underline{e}_2$. Thus the dimensionless ratios ℓ_j/σ^2 are ultimately where the signal to noise ratio resides, namely in the comparison of the principal variances of the p time series in $\underline{X}$ with the variance σ^2 of the noise $\underline{\varepsilon}$. The values β_j^2 are intrinsic properties of the physical system and are presumably independent of $\underline{X}$ and $\underline{\varepsilon}$ (cf (2.7)).

D. Introduction of $\overline{\lambda}$: We now recast (9.5) as

$$\lambda = p\overline{\lambda} \tag{9.6}$$

where we have written

$$\text{'}\overline{\lambda}\text{'} \text{ for } \frac{1}{p}\sum_{j=1}^{p} (\ell_j/\sigma^2)\beta_j^2 \tag{9.7}$$

thereby defining the *mean signal to noise ratio*. In any given physical setting from which we can draw p time series out of a large reservoir of time series, of fixed sample size n we know intuitively that $\overline{\lambda}$ (despite the various fluctuations encountered as we draw from that reserve and increase p and continue to reckon the resulting $\overline{\lambda}$'s, i.e., we know that $\overline{\lambda}$) will remain generally in some relatively small interval of values. The β_j, being Greens' functions, essentially of the kind in (2.2), also present a more or less spatially homogeneous variation with j. There are fluctuations of the β_j with h, of course, but the *mean* or average of these values together with those of ℓ_j are expected to be relatively steady as p increases. In this way we argue that the signal to noise ratio λ should be given an explicit linear dependence on p, particularly for the purpose of exploring changes of the performance indexes under changes with p or λ.

E. Some Immediate Consequences of the Definition of $\bar{\lambda}$

Let us return to the closed forms for μ_Q, $\sigma^2{}_Q$ in (8.2), (8.3) and use in them the representation $\lambda = p\bar{\lambda}$ for λ. We note first of all that

$$\mu_Q = \frac{\lambda+p}{n-p-2} = \frac{p(1+\bar{\lambda})}{n-p-2} = \frac{p}{n}\frac{(1+\bar{\lambda})}{(1-p/n-2/n)}$$

For n large compared with 2, we can write this approximately as:

$$\mu_Q = \frac{S_o(1+\bar{\lambda})}{1-S_o-2/n} \cong \frac{S_o(1+\bar{\lambda})}{1-S_o} = Q_o(1+\bar{\lambda}) \tag{9.8}$$

Here $$S_o = p/n, \text{ and } Q_o \equiv S_o/(1-S_o) = p/(n-p) \tag{9.9}$$

The definition of Q_o is suggested by the general connection between S and Q in the Table of §7. That (9.8) arises so neatly this way, with its connections to the case of $\bar{\lambda} = 0$ (i.e. S_o, Q_o), is a good sign that (9.7) is a natural definition in the linear regression hindcast context. Except for the condition on n, (9.8) is exact. Equation (9.8) states that μ_Q grows linearly with the mean signal to noise ratio $\bar{\lambda}$.

We consider next (8.3) in which we substitute $p\bar{\lambda}$ for λ and find

$$\sigma_Q^2 = \frac{2[\lambda^2+(n-2)(2\lambda+p)]}{[n-p-2]^2[n-p-4]} = \frac{2[S_o^2\overline{\lambda}^2+(1-\frac{2}{n})S_o(2\overline{\lambda}+S_o)]}{n[1-S_o-2/n]^2[1-S_o-4/n]} \tag{9.10}$$

If n is large compared to 4, then we can write this as

$$\sigma_Q^2 \cong \frac{2}{n} \cdot \frac{[S_o^2(1+\overline{\lambda}^2) + 2S_o\overline{\lambda}]}{[1-S_o]^3} \tag{9.11}$$

$$= \frac{2}{n} \cdot Q_o(1+Q_o)\ [Q_o(1+\overline{\lambda}^2)+2\overline{\lambda}(1+Q_o)] \tag{9.12}$$

From this we see how, holding n, p fixed, σ_Q^2 grows parabolically with $\overline{\lambda}$, or alternately σ_Q^2 decreases as 1/n with increasing n for fixed S_o or Q_o. This latter fact is in accordance with large-sample theory. The p-dependence of σ_Q^2 is now essentially in the Q_o (cf (9.9)), and we can see rapid growth of σ_Q^2 with p holding n, $\overline{\lambda}$ fixed.

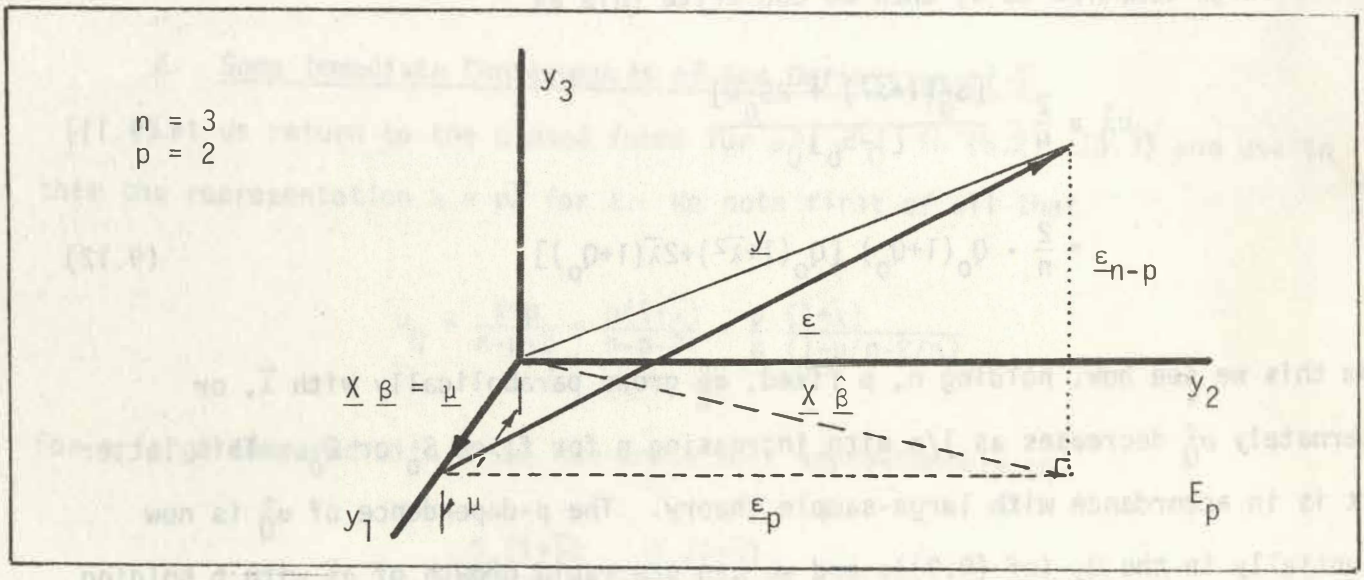
10. The Monte Carlo Skeleton of Linear Regression

It is possible to explore the linear regression problem by means of a Monte Carlo simulation of the noise vector $\underline{\varepsilon}$ added to a fixed signal vector $\underline{\mu} = \underline{X}\ \underline{\beta}$. No restrictions need then be imposed on the distribution of $\underline{\varepsilon}$ in order to gain an insight into the corresponding behavior of $\underline{y}$, $\underline{X}\ \hat{\underline{\beta}}$, and any of the performance indexes associated with these vectors. We outline the proof of this possibility in three stages.

A. The Standard Gaussian Case

To see how the simulation goes, first of all in the standard gaussian case (§5B), recall the regression diagram in §6. The vector $\underline{X}\ \underline{\beta}$ is fixed in E_n. To $\underline{X}\ \underline{\beta}$

is added the random n dimensional vector $\underline{\varepsilon}$ to yield the observation vector $\underline{y}$. The representation of $\underline{X}\,\underline{\beta}$ as a vector in E_n can be simplified by a rotational change of basis of the kind adopted in the derivation of the χ^2 distribution in §3 of Appendix A (Stage 3 there). Thus the diagram of §6 becomes:



That is, the vector $\underline{\mu} = \underline{X}\,\underline{\beta}$ is now aligned along the first axis of E_n. If we adopt the coordinate frame $\underline{B} = [\underline{B}_p\ \underline{B}_{n-p}]$ used in §2 of Appendix A, we can use the independent gaussian variates δ_j, j=1, ..., n, defined there to simulate the random activity in the diagram. Let $\underline{\delta} = (\delta_1, \ldots, \delta_n)^T$ be the vector of uncorrelated zero-mean unit-variance gaussian variates. Then $||\underline{X\beta}||^2/\sigma^2 \equiv \mu^2/\sigma^2 = \mu^2 = \lambda$ the signal to noise ratio for the present set up.

The Monte Carlo simulation of $\underline{X}\,\hat{\underline{\beta}}$ in this frame is then

$$\underline{X}\,\hat{\underline{\beta}} = (\mu+\delta_1)\underline{b}_1 + \delta_2\underline{b}_2 + \ldots + \delta_p\underline{b}_p \tag{10.0}$$

and that of $\underline{y}$ is

$$\underline{y} = (\mu+\delta_1)\underline{b}_1 + \delta_2\,\underline{b}_2 + \ldots + \delta_n\,\underline{b}_n \tag{10.1}$$

Further we have the simulations

$$\underline{\varepsilon}_p = \delta_1 \underline{b}_1 + \dots + \delta_p \underline{b}_p \tag{10.2}$$

$$\underline{\varepsilon}_{n-p} = \delta_{p+1} \underline{b}_{p+1} + \dots + \delta_n \underline{b}_n \tag{10.3}$$

$$\underline{\varepsilon} = \underline{\varepsilon}_p + \underline{\varepsilon}_{n-p} \tag{10.4}$$

Thus we can write $\underline{y}$ as

$$\underline{y} = (\mu+\delta_1)\underline{b}_1 + \dots + \delta_p \underline{b}_p + [\delta_{p+1}\underline{b}_p + \dots + \delta_n \underline{b}_n]$$

$$= \underline{X}\,\underline{\hat{\beta}} + \underline{\varepsilon}_{n-p} \tag{10.5}$$

as usual (cf. (4.2)). It is easy to check that squares of lengths of the above vectors use only the squares of the appropriate δ_j occurring in their representations. Thus, e.g.,

$$||\underline{X\hat{\beta}}||^2 = (\underline{X\hat{\beta}})^T(\underline{X\hat{\beta}}) = (\mu+\delta_1)^2 + \delta_2^2 + \dots + \delta_p^2$$

$$||\underline{y}||^2 = \underline{y}^T\underline{y} = (\mu+\delta_1)^2 + \delta_2^2 + \dots + \delta_n^2$$

The Monte Carlo representations of the performance indexes in §7 are then given as:

$$Q(n,p,\lambda) = ||\underline{X\hat{\beta}}||^2/||\underline{\varepsilon}_{n-p}||^2 = \frac{(\mu+\delta_1)^2 + \delta_2^2 + ,,, + \delta_p^2}{\delta_{p+1}^2 + \dots + \delta_n^2} \tag{10.6}$$

$$S(n,p,\lambda) = ||X\hat{\beta}||^2/||y||^2 = \frac{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2}{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2+(\delta_{p+1}^2+\ldots+\delta_n^2)} \qquad (10.7)$$

$$C(n,p,\lambda) = ||y||^2/||\varepsilon_{n-p}||^2 = \frac{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2+(\delta_{p+1}^2+\ldots+\delta_n^2)}{\delta_{p+1}^2+\ldots+\delta_n^2} \qquad (10.8)$$

$$I(n,p,\lambda) = ||\varepsilon_{n-p}||^2/||X\hat{\beta}||^2 = \frac{\delta_{p+1}^2+\ldots+\delta_n^2}{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2} \qquad (10.9)$$

$$R(n,p,\lambda) = ||\varepsilon_{n-p}||^2/||y||^2 = \frac{\delta_{p+1}^2+\ldots+\delta_n^2}{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2+(\delta_{p+1}^2+\ldots+\delta_n^2)} \qquad (10.10)$$

$$U(n,p,\lambda) = ||y||^2/||X\hat{\beta}||^2 = \frac{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2+(\delta_{p+1}^2+\ldots+\delta_n^2)}{(\mu+\delta_1)^2+\delta_2^2+\ldots+\delta_p^2} \qquad (10.11)$$

To operate these simulators: For each realization of say (10.6), generate the n realizations δ_j, $j=1, \ldots, n$. Then perform the remaining indicated operations in the numerator and denominator of (10.6). Repeat as often as desired. Collect the results and statistics as required. Observe carefully how the n realizations $\delta_1, \ldots, \delta_n$ are used in the fractions. After many such realizations the values of Q (say) will spread out on the positive real line $(0,\infty)$ with a density that approximates that given by (8.1), and the averages of Q in the simulations will approach that given by (8.2), etc. In fact, our analytic and algebraic derivations of the formulas in §8 were checked using (10.6), (10.7), (10.9) in thousands of realizations for each formula. This check also served to show how relatively cheaply the Monte Carlo simulations of regression settings can be carried out. Many interesting experiments are suggested by the formulas (10.0) - (10.5) and those in (10.6) - (10.11).

B. Correlated Gaussian Noise Simulation

A moment's reflection on (10.0) - (10.11) will suggest that their formulations are applicable, as they stand, to more general probability settings than the standard one. To see this, consider the rotational realignment of the axes of E_p to place $\underline{X}\ \underline{\beta}$ along the first axis in E_p (and hence E_n). This realignment does not change the correlation properties of the population of vectors $\underline{\varepsilon}$, provided we rotate the $\underline{\varepsilon}$ vectors along with the frame as we make the desired alignment. Thus if $\underline{M}$ is the orthogonal matrix used in going from (A22) to (A23), and the present version of (A22) is

$$\frac{|\underline{C}|^{-\frac{1}{2}}}{(2\pi)^{p/2}} \exp\{-\tfrac{1}{2}(\underline{x}-\underline{\mu})^T \underline{C}^{-1}(\underline{x}-\underline{\mu})\} \ ,$$

where $\underline{C}$ is the population covariance matrix of the noise vector $\underline{\varepsilon}$ of current interest, then clearly since,

$$(x-\underline{\mu})^T \underline{C}^{-1}(x-\underline{\mu}) = [M^T(\underline{x}-\underline{\mu})]^T[\underline{M}^T \underline{C}\underline{M}]^{-1}[\underline{M}^T(x-\underline{\mu})] \ ,$$

we would use the covariance matrix $\underline{M^T CM}$ in devising the simulation calculations in any of formulas (10.0) to (10.11). The generation of gaussian variates with a given covariance matrix $\underline{M^T CM}$ is easily effected. In this way we could generate several thousand trial values of Q, say, and get an impression of their mean values μ_Q and their spread σ_Q^2 and so on, when the noise is correlated.

There is an alternate Monte Carlo approach to finding the pdf of any performance index in the case of correlated gaussian noise. This is based on knowledge of the covariance matrix $\underline{C}$ of the noise and particularly on its square root matrix $\underline{S}$, where $\underline{S}\ \underline{S}^T = \underline{C}$. We use the canonic skill Q and the developments in §5 to explain the method. Suppose the data matrix comes to us as $\underline{W}$ and the residual noise vector is $\underline{\delta}$. Then $\langle\underline{\delta\delta}^T\rangle = \underline{C}$. Moreover, $\underline{Y} = \underline{S}^{-1}\underline{W}$; $\underline{\varepsilon} = \underline{S}^{-1}\underline{\delta}$ are respectively the new data matrix and uncorrelated noise vectors, with the latter having zero mean and unit variance. The canonic skill in this uncorrelated setting is $Q = ||\underline{Y}\hat{\underline{\alpha}}||^2/||\underline{\varepsilon}_{n-p}||^2$, by definition. Since $\sigma^2 = 1$, the signal to noise ratio λ is simply $|\underline{|Y\alpha|}|^2 \equiv \mu^2$. The Monte Carlo simulation then proceeds as in par A above. Thus, the vectors $\underline{Y}\hat{\underline{\alpha}} = [(\mu+\delta_1), \delta_2, \ldots, \delta_p]^T$ and $\underline{\varepsilon}_{n-p} = [\delta_{p+1}, \ldots, \delta_n]^T$ are formed. Then apply $\underline{S}$ to the vectors $\underline{Y}\hat{\underline{\alpha}}, \underline{\varepsilon}_{n-p}$ to form $\underline{S}(\underline{Y}\hat{\underline{\alpha}})$ and $\underline{S}(\underline{\varepsilon}_{n-p})$ and thus the quotient $Q' = ||\underline{S}(\underline{Y}\hat{\underline{\alpha}})||^2/||\underline{S}(\underline{\varepsilon}_{n-p})||^2$, *which is the canonic skill in the original correlated setting* - since $\underline{S}\underline{Y} = \underline{W}$ and $\underline{S}\,\underline{\varepsilon}_{n-p} = \underline{\delta}_{n-p}$. In this way each realization of the uncorrelated unit-variance variables $\delta_1, \ldots, \delta_p$ yields a realization of Q'. Many such realizations can be used to build a histogram, i.e., a finite approximant to the pdf of Q'. Observe that this procedure could assign a meaning to λ where it would not, *prima facie*, exist.

C. The General Case

The foregoing observations suggest that the Monte Carlo representations (10.0) - (10.11) can be used for any random noise population provided the pdf for the population is known in sufficient detail so as to allow a simulated sampling via the usual Monte Carlo techniques. Moreover the pdf should allow a rotation of itself into the preferred alignment of $\underline{X}\,\underline{\beta}$ along a particular (say, the first) axis of the coordinate system for E_n. Even the latter rotation is no longer needed if it becomes too arduous to perform the rotation. What would be needed in this event is the set of the n components of $\underline{X}\,\underline{\beta}$ in the $\underline{B}$-frame of E_n. If $\underline{X}\,\underline{\beta} = (\mu_1, \ldots, \mu_p, 0, \ldots, 0)^T$ are these components, then (10.0) would be replaced by

$$\underline{X}\,\hat{\underline{\beta}} = \sum_{j=1}^{p} (\mu_j+\delta_j)\,\underline{b}_j \qquad (10.12)$$

and (10.1) by

$$\underline{y} = \sum_{j=1}^{n} (\mu_j+\delta_j)\,\underline{b}_j \qquad (10.12)$$

The forms of (10.2) - (10.4) are unchanged. However, the original, simple notion of a signal-to-noise ratio λ no longer exists and we drop it from the notation. The simulation of Q, for example, would then be accomplished by the following generalization of (10.6):

$$Q(n,p) = ||X\hat{\underline{\beta}}||^2/||\underline{\varepsilon}_{n-p}||^2 = \frac{\sum_{j=1}^{p}(\mu_j+\delta_j)^2}{\sum_{j=p+1}^{n}\delta_j^2} \quad (= 1/I(n,p)) \tag{10.13}$$

The $\delta_1, \ldots, \delta_n$ would now be randomly drawn repeatedly from the n-variate population with the given pdf. As another example, (10.7) would become:

$$S(n,p) = ||X\hat{\underline{\beta}}||^2/||\underline{y}||^2 = \frac{\sum_{j=1}^{p}(\mu_j+\delta_j)^2}{\sum_{j=1}^{p}(\mu_j+\delta_j)^2 + \sum_{j=p+1}^{n}\delta_j^2} \tag{10.14}$$

11. Estimating the Signal to Noise Ratio λ

We have seen throughout the studies above the central role played by the signal to noise ratio λ. It is therefore of some importance to determine λ from hindcasts of real data. We shall now consider two methods leading to the determination of confidence limits for λ. The small-sample method is covered in pars A, B. The large-sample method is described in par C.

A. Confidence interval for λ via canonic skill— small-sample theory

Let us return to the pdf for canonic skill in (8.1). Select a value for n and p. Choose a value for the mean signal to noise parameter $\bar{\lambda}$. This then fixes $\lambda = p\bar{\lambda}$ (cf. (9.6)). Choose a confidence level $(1-\alpha)$ 100%. One can then find the $\sigma(\frac{1}{2}\alpha)$ and $\sigma(1-\frac{1}{2}\alpha)$ values of Q such that*

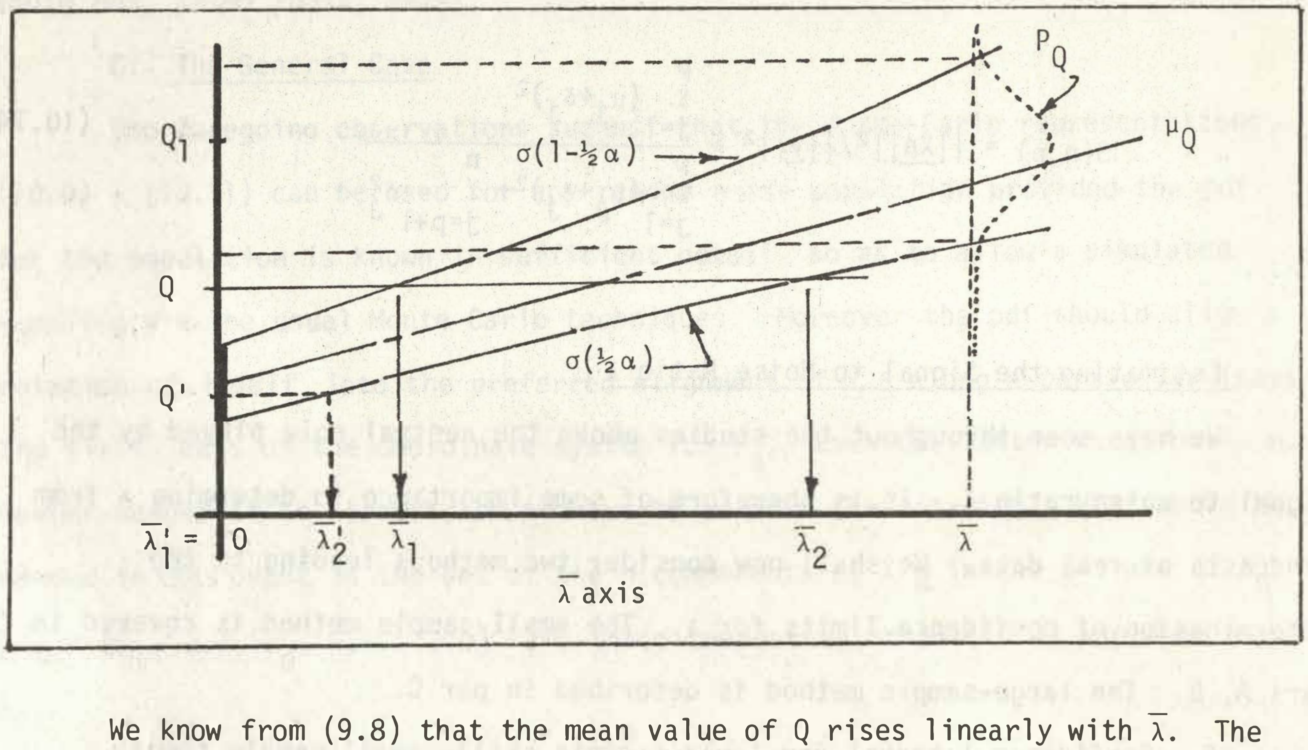
---

* Formulas for the determination of these integrals are given in Appendix B.

$$\int_0^{\sigma(\frac{1}{2}\alpha)} P_Q(x|n,p,\lambda)dx = \tfrac{1}{2}\alpha \tag{11.1}$$

$$\int_0^{\sigma(1-\frac{1}{2}\alpha)} P_Q(x|n,p,\lambda)dx = 1-\tfrac{1}{2}\alpha \tag{11.2}$$

If we repeat this determination of $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ for a selected set of $\overline{\lambda}$ values (for fixed n,p) then we can rough-in curves (as accurately as we wish) of $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ as functions of $\overline{\lambda}$. Let the results be as sketched below:



We know from (9.8) that the mean value of Q rises linearly with $\overline{\lambda}$. The curves for $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$, as suggested by (9.10), will diverge parabolically from the straight line for μ_Q. Again by (9.10), it is clear that this departure from the μ_Q line can be made arbitrarily small for n chosen sufficiently large, for a given fixed ratio $S_o = p/n$ or equivalently $Q_o = p/(n-p)$.

Suppose now that we have a value Q from a hindcast with the given n,p of the diagram. Draw a horizontal line through this value of Q and determine the $\bar{\lambda}$-values of the intersections of the horizontal line with the $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ curves. The resultant values $\bar{\lambda}_1$, $\bar{\lambda}_2$ determine the confidence interval for $\bar{\lambda}$. That is, with confidence* $(1-\alpha)100\%$, $\bar{\lambda}$ is in $[\bar{\lambda}_1,\bar{\lambda}_2]$. By our observations above, $[\bar{\lambda}_1,\bar{\lambda}_2]$ can be made arbitrarily small for n sufficiently large for a given ratio $p/n = S_o$. Hence the method in principle can pinpoint $\bar{\lambda}$ if there is enough of a data stretch over which we have a stationary setting.

B. The use of any performance index to find the confidence interval for λ

The observations in par A may obviously be extended to the use of P_S or P_I in §8 to find $[\bar{\lambda}_1,\bar{\lambda}_2]$. The relative capabilities of P_Q, P_I, P_S in this regard will not be studied here.

C. Large-sample estimates of λ

The large-sample method is derived from the following facts. For a given n, p, λ, the canonic skill Q of a hindcast model $\underline{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$ is distributed in a known way, such that the population mean of Q is

$$\mu_Q = \frac{\lambda+p}{n-p-2} \qquad (11.3)$$

and the population variance of Q is

---

* Proof: In the diagram, if the true value is $\bar{\lambda}$, then $(1-\alpha)100\%$ of all the horizontal lines randomly drawn through the axis of Q values will fall between the dashed lines formed by the $\frac{1}{2}\alpha$, $1-\frac{1}{2}\alpha$ points of the pdf at $\bar{\lambda}$. Therefore, if $\bar{\lambda}$ is the true value, then horizontal lines drawn through realized Q values will produce intervals $[\bar{\lambda}_1, \bar{\lambda}_2]$ such that $\bar{\lambda}$ will be in $[\bar{\lambda}_1,\bar{\lambda}_2]$, $(1-\alpha)100\%$ of the time. This is the correct interpretation of $[\bar{\lambda}_1,\bar{\lambda}_2]$.

$$\sigma_Q^2 = \frac{2\{\lambda^2+(n-2)(2\lambda+p)\}}{[n-p-2]^2[n-p-4]} \quad , \tag{11.4}$$

as we see in (A55), (A58).

If we apply the model $\underline{y} = \underline{X}\ \underline{\beta} + \underline{\varepsilon}$ repeatedly to independent data and observation sets $\underline{X}$, $\underline{y}$, and compute in each case $Q = ||\underline{X\hat{\beta}}||^2/||\underline{\varepsilon}_{n-p}||^2$, we obtain a set of (say) m Q-values which, in the limit of an infinite number of such independent trials (i.e., $m \to \infty$), are distributed with mean μ_Q and variance σ_Q^2. Therefore, the statistic Z determined by any finite sample of size m:

$$Z = (\overline{Q}-\mu_Q)/[\sigma_Q/m^{\frac{1}{2}}] \tag{11.5}$$

where,

$$\overline{Q} = m^{-1}(Q_1+\ldots+Q_m)$$

is distributed approximately *normally* with zero mean and unit variance. The larger the m, the closer the approximation. This fact follows from an application of a simple form of the Central Limit Theorem (Hoel, 1954, p107).

To apply the foregoing observation, decide on a level $1-\alpha$ of confidence. Let $Z_{\frac{1}{2}\alpha}$ be the two-sided normal pdf limit associated with $1-\alpha$. Then for a sample of size m, $\overline{Q} = m^{-1}(Q_1+\ldots+Q_m)$ is known. μ_Q and σ_Q are determined by p, n, but with λ unknown. Hence we have the bound-condition on λ given by

$$-Z_{\frac{1}{2}\alpha} < (\overline{Q}-\mu_Q)/[\sigma_Q m^{-\frac{1}{2}}] < Z_{\frac{1}{2}\alpha} \tag{11.6}$$

In principle we may now vary λ in (11.6) until those two values of λ are found that make the statistic Z take on the two extreme limit values $\pm Z_{\frac{1}{2}\alpha}$. These two values of λ will form the desired ends of the confidence interval for the true λ.

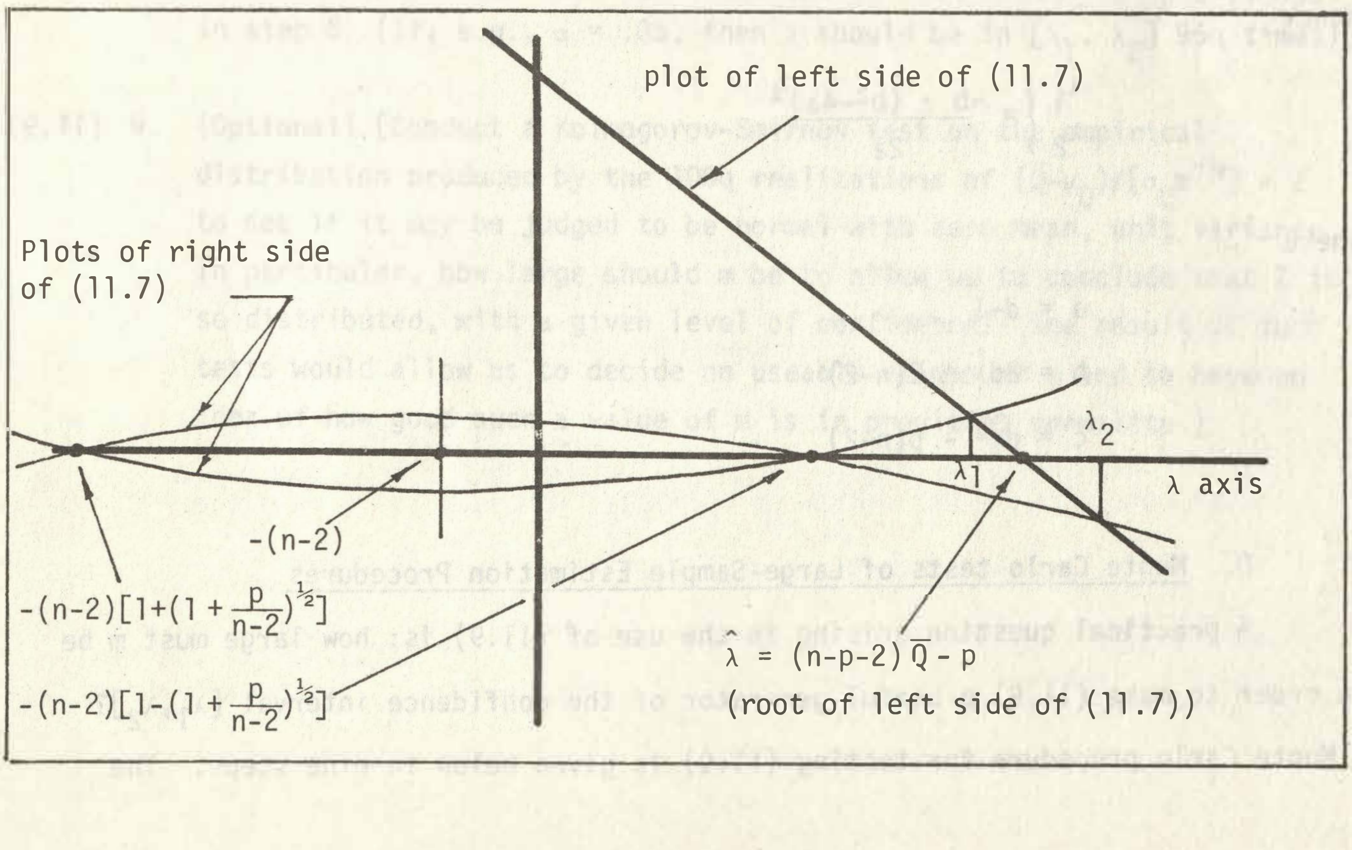
We can solve for these λ values by setting

$$\frac{\overline{Q}-\mu_Q}{\sigma_Q m^{-\frac{1}{2}}} = \pm Z_{\frac{1}{2}\alpha}$$

so that

$$\overline{Q} - \frac{\lambda+p}{n-p-2} = \pm \frac{Z_{\frac{1}{2}\alpha}}{m^{\frac{1}{2}}} \left[\frac{2\{\lambda^2+(n-2)(2\lambda+p)\}}{[n-p-2]^2[n-p-4]} \right]^{\frac{1}{2}} \tag{11.7}$$

It is easy to see, at least in principle that, for sufficiently large m, two roots λ_1, λ_2 of (11.7) will exist. Thus in the diagram below is a sketch of the straight line generated by varying λ in the left side of (11.7). Letting λ vary in the right side of (11.7) produces two parabolas, one for each sign. These are sketched as the two curved lines in the figure below.



The parabolas meet the straight line at abscissas λ_1, λ_2, the desired confidence limits of λ. The estimate $\hat{\lambda} = (n-p-2)\overline{Q}-p$ of λ always lies in the interval $[\lambda_1, \lambda_2]$. Clearly by (11.3), $\langle\hat{\lambda}\rangle = \langle(n-p-2)\overline{Q}-p\rangle = \lambda$, and so $\hat{\lambda}$ is an unbiased estimate of λ.

We now can see that the intersections at λ_1, λ_2 will always exist for a given p, n, since $m^{\frac{1}{2}}$ in (11.7) can be made arbitrarily large, thereby producing parabolas that are arbitrarily shallow, and hence, by their intersections with the straight line, produce an interval $[\lambda_1, \lambda_2]$ about $\hat{\lambda}$ that is arbitrarily small.

A mechanical procedure for determining λ_1, λ_2 is given as follows. Rearrange (11.7) into the form of a quadratic equation:

$$(d-1)\lambda^2 - 2(d\hat{\lambda}+(n-2))\lambda + [d\hat{\lambda}^2-p(n-2)] = 0 \qquad (11.8)$$

where

$$d = \frac{m}{2Z^2_{\frac{1}{2}\alpha}} \cdot \frac{[n-p-2]^2}{[n-p-4]}, \quad \hat{\lambda} = (n-p-2)\overline{Q}-p$$

Hence

$$\left.\begin{matrix}\lambda_1\\ \lambda_2\end{matrix}\right\} = \frac{-b \pm (b^2-4a)^{\frac{1}{2}}}{2a} \qquad (11.9)$$

where

$$a = d-1$$

$$b = 2d\hat{\lambda} + 2(n-2)$$

$$c = d\hat{\lambda}^2 - p(n-2)$$

D. <u>Monte Carlo tests of Large-Sample Estimation Procedures</u>

A practical question arising in the use of (11.9) is: how large must m be in order to make (11.9) a useful generator of the confidence interval $[\lambda_1,\lambda_2]$? A Monte Carlo procedure for testing (11.9) is given below in nine steps. The method uses the representation (10.6) of Q.

1. Fix n, p, λ, choose m, q (defined below).

2. Fix the confidence level $1-\alpha$ and hence $\frac{1}{2}\alpha$, $Z_{\frac{1}{2}\alpha}$.

3. Compute m realizations of $Q = [(\mu+\varepsilon_1)^2 + \varepsilon_2^2 +\ldots+\varepsilon_p^2]/[\varepsilon_{p+1}^2+\ldots+\varepsilon_n^2]$ (A fresh, randomly chosen batch of variates $\varepsilon_1,\ldots, \varepsilon_n$ is used for each realization).

4. Compute $\overline{Q} = m^{-1}(Q_1+\ldots+Q_m)$ from the result in 3.

5. Compute λ_1, λ_2 from (11.9).

6. Check to see if λ is in $[\lambda_1, \lambda_2]$.

7. Repeat 3-6 a large number, say 100q times where q = 1, 2, 3, ...

8. Make a tally of the number of times out of 100q that λ is in $[\lambda_1, \lambda_2]$, in step 6 (If, e.g., $\alpha = .05$, then λ should be in $[\lambda_1, \lambda_2]$ 95q times).

9. (Optional) [Conduct a Kolmogorov-Smirnov test on the empirical distribution produced by the 100q realizations of $(\overline{Q}-\mu_Q)/[\sigma_Q m^{-\frac{1}{2}}] = Z$ to see if it may be judged to be normal with zero mean, unit variance. In particular, how large should m be to allow us to conclude that Z is so distributed, with a given level of confidence? The result of such tests would allow us to decide on useable values of m and to have an idea of how good such a value of m is in providing normality.]

12. Model Significance

A. Solution of the problem

The problem of model significance, defined in §1B, can be solved by the technique described in §11A. It is clear from the diagram in §11A that if we have a value Q' from a hindcast which is such that in $[\bar{\lambda}_1', \bar{\lambda}_2'], \bar{\lambda}_1' = 0$, then with confidence $(1-\alpha)$ 100% the value $\bar{\lambda} = 0$ is in $[0, \bar{\lambda}_2']$. In other words, if Q' falls in the heavy interval (in the figure) associated with $\bar{\lambda} = 0$ (and hence $\lambda = 0$) the model is *not significant*, and this judgment is reached with confidence $(1-\alpha)$ 100%. This procedure can be effected by programming (11.1), (11.2) using (A48) in which $\lambda = 0$.

B. Equivalence with Barnett and Hasselmann

The preceding criterion of model insignificance, namely $\bar{\lambda}_1 = 0$, is equivalent to Barnett and Hasselmann's criterion that $\underline{\beta} = 0$. For if $||\underline{X\beta}||^2/\sigma^2 = \lambda = 0$ and $\underline{X}$ is of rank p (as it usually is taken to be in hindcasts) then it follows* that $\underline{\beta} = 0$. Conversely a zero $\underline{\beta}$ vector implies $\lambda = 0$. The procedure of Barnett and Hasselmann is based on (A44): The quantity $\hat{\underline{\beta}}$ is found; $\underline{\beta}$ is assumed zero by hypothesis. Then, if $||\hat{\underline{\beta}}||^2/\sigma^2$ does not exceed the (say) .95 significance level of the $\chi^2(p)$ distribution the model is judged insignificant.

C. Generalized Barnett and Hasselmann procedure to find confidence intervals

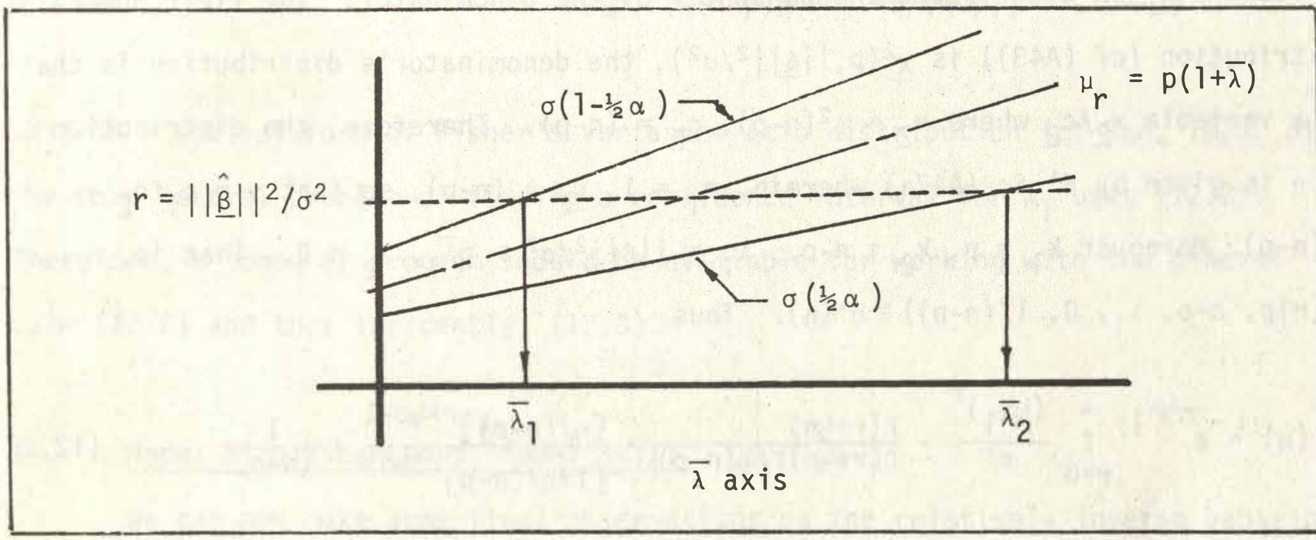
The procedure of Barnett and Hasselmann can be generalized as follows. Let $r = ||\hat{\underline{\beta}}||^2/\sigma^2$. We use (A25) to construct $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ curves via a selected set of $\bar{\lambda}$ values and given n, p. Now $\bar{\lambda} = \lambda_o/p$, λ_o as in (A43). From (A43), (A30) we have

---

* One may see this also from inspection of (9.5). All the terms in the sum are non-negative. Hence if the sum is zero, and the p values ℓ_j are not zero (this is the rank condition in another form) then necessarily the β_j are 0.

$$\mu_r = \lambda + p = p(1+\bar{\lambda})$$

so that the mean value of r rises linearly with $\bar{\lambda}$ for fixed p. Moreover, from (A33) we expect the $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ curves to diverge approximately linearly from μ_r (since $\sigma_r^2 = 2p(1+2\bar{\lambda})$). A sketch of the curves is given below.



In a hindcast $||\hat{\underline{\beta}}||^2/\sigma^2$ is determined. A horizontal line through this value fixes the $(1-\alpha)100\%$ confidence interval for $\bar{\lambda}$, namely $[\bar{\lambda}_1, \bar{\lambda}_2]$. If $\bar{\lambda}_1 = 0$, the model is judged insignificant. Otherwise, *we can then estimate* $[\bar{\lambda}_1, \bar{\lambda}_2]$ *of the significant model.*

D. <u>Further generalizations</u>

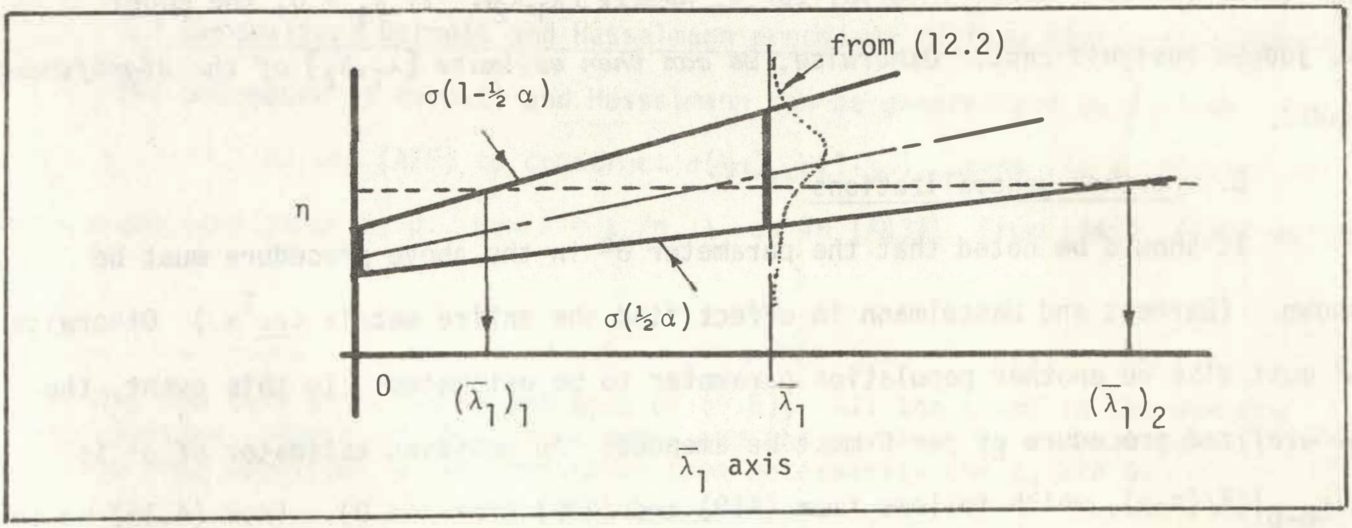
It should be noted that the parameter σ^2 in the above procedure must be known. (Barnett and Hasselmann in effect find the entire matrix $\langle \underline{\varepsilon}\underline{\varepsilon}^T \rangle$.) Otherwise σ^2 must also be another population parameter to be estimated. In this event, the generalized procedure of par C must be amended: An unbiased estimator of σ^2 is $||\underline{\varepsilon}_{n-p}||^2/(n-p)$, which follows from (A18) and (A30) (for $\lambda = 0$). From (4.15) we see that $\hat{\underline{\beta}} = \underline{\beta} + (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}_p$. If we adopt the orthonormalized data matrix $\underline{X}$, i.e., $\underline{X}^T\underline{X} = \underline{I}_p$, then $\hat{\underline{\beta}} = \underline{\beta} + \underline{X}^T\underline{\varepsilon}_p$. This coordinate frame was used in §2 of Appendix A to show the independence of $\underline{\varepsilon}_p$ and $\underline{\varepsilon}_{n-p}$. Hence we can form the statistic η from hindcast information:

$$\eta = [||\hat{\underline{\beta}}||^2/\sigma^2]/[||\underline{\varepsilon}_{n-p}||^2/(n-p)\sigma^2] = ||\hat{\underline{\beta}}||^2/[||\underline{y}-X\hat{\underline{\beta}}||^2/(n-p)] \qquad (12.1)$$

The numerator is distributed independently of the denominator. The first numerator's distribution (cf (A43)) is $\chi^2(p,||\underline{\beta}||^2/\sigma^2)$, the denominator's distribution is that of a variable x_2/c_2 where $x_2 \sim \chi^2(n-p)$, $c_2 = (n-p)$. Therefore, the distribution of η is given by H' in (A47a) wherein, $c_1 = 1$, $c_2 = (n-p)$, so that $\gamma = c_1/c_2 = 1/(n-p)$. Moreover $k_1 = p$, $k_2 = n-p$. $\lambda_1 = ||\underline{\beta}||^2/\sigma^2 \equiv p\bar{\lambda}_1$, $\lambda_2 = 0$. That is, $H'(\eta|p, n-p, \lambda_1, 0, 1/(n-p)) \equiv H'(\eta)$. Thus

$$H'(\eta) = e^{-\frac{1}{2}\lambda_1} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \cdot \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{[\eta/(n-p)]^{r+\frac{1}{2}p-1}}{[1+\eta/(n-p]^{r+\frac{1}{2}n}} \cdot \frac{1}{(n-p)} \qquad (12.2)$$

We may now compute $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ from (12.2) for various choices of $\bar{\lambda}_1$, thereby forming confidence curves as before, and making a diagram of the kind shown below:



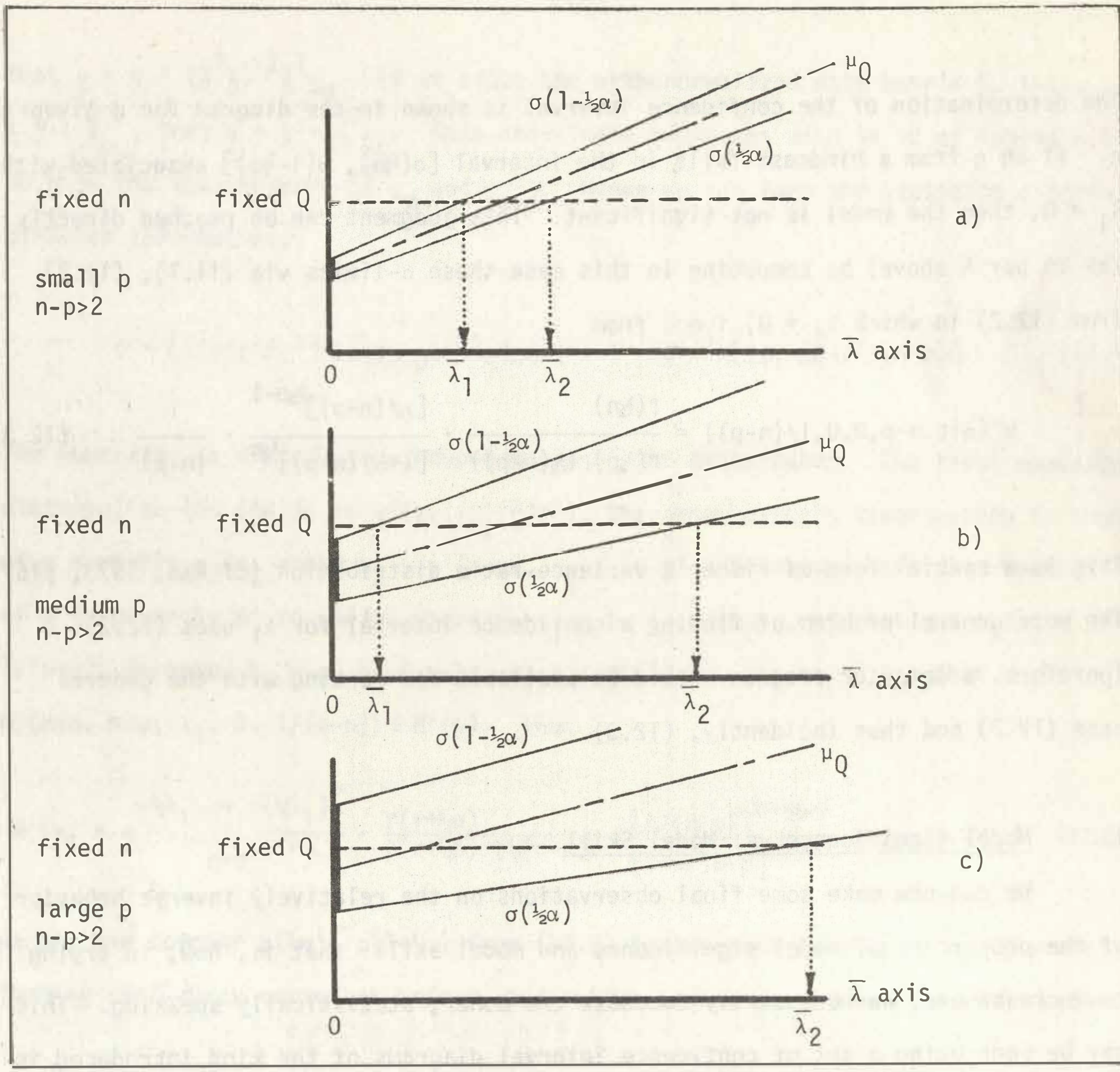
The determination of the confidence interval is shown in the diagram for a given value η. If an η from a hindcast falls in the interval $[\sigma(\frac{1}{2}\alpha), \sigma(1-\frac{1}{2}\alpha)]$ associated with $\overline{\lambda}_1 = 0$, then the model is not significant. This judgment can be reached directly (as in par A above) by computing in this case these σ-limits via (11.1), (11.2) from (12.2) in which $\lambda_1 = 0$, i.e., from

$$H'(\eta|p,n-p,0,0,1/(n-p)) = \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{[\eta/(n-p)]^{\frac{1}{2}p-1}}{[1+\eta/(n-p)]^{\frac{1}{2}n}} \cdot \frac{1}{(n-p)} \qquad (12.3)$$

This is a special form of Fisher's variance-ratio distribution (cf Rao, 1973, p167). The more general problem of finding a confidence interval for λ_1 uses (12.2). Therefore, a computer program should be available for working with the general case (12.2) and thus incidently, (12.3).

13. Model Significance vs. Model Skill

We can now make some final observations on the relatively inverse behavior of the properties of model significance and model skill: that is, how, in trying to increase one, we necessarily decrease the other, statistically speaking. This may be seen using a set of confidence interval diagrams of the kind introduced in §11. The changes in the diagrams below are the result of increasing the number p of predictors, holding the number n of samples fixed. The changes are observable for a continuum of mean signal to noise ratios $\overline{\lambda}$, and are based on the suggestions in (8.1), (8.2) as to how the mean μ_Q behaves and on how the $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ curves behave with changes in p.

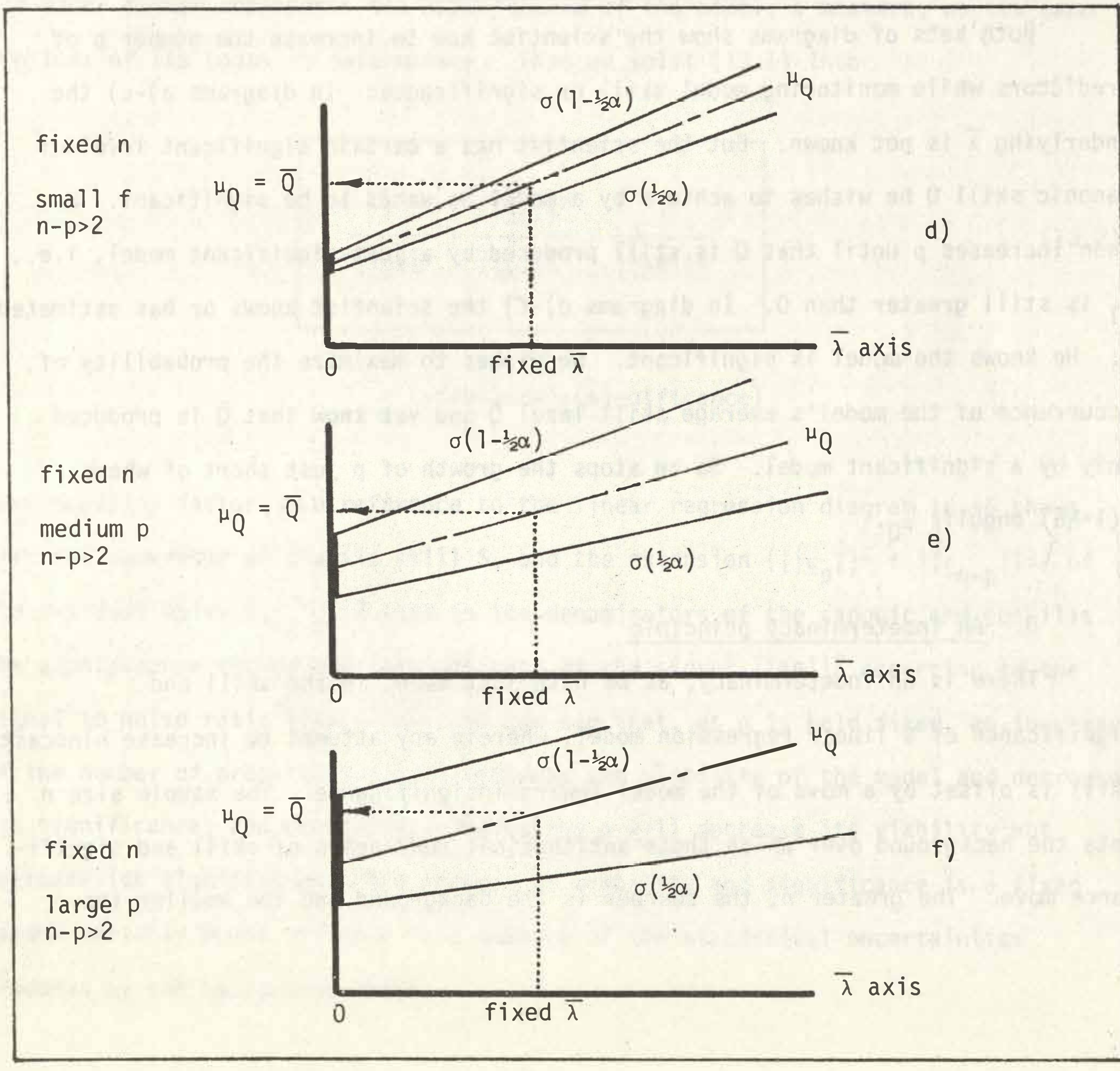


A. Significant-model strategy

In diagram a), p is small relative to n. An observed high value of canonic skill Q produces a pair of $\overline{\lambda}$ values well away from 0 on the $\overline{\lambda}$ axis and we have a highly significant model. Holding n fixed but increasing the number p of predictors generally raises the average Q at each $\overline{\lambda}$, as in diagram b). The increase in p also spreads the $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ curves away from the straight μ_Q line. Hence the same high Q in a) is now less spectacular probability-wise and still very good: but the confidence interval for $\overline{\lambda}$ has moved toward 0. Finally in c) p has been increased to a relatively large value just under n. The higher-spread Q values for this model now are very probable and engulf the same Q of the preceding two cases. Thus the Q is as good as before, on an absolute scale, but it is probabilistically mediocre. Moreover, it can be produced by a non significant model, since the confidence interval now includes 0 on the $\overline{\lambda}$ axis.

B. Significant-skill strategy

There is a complementary way of seeing the above phenomenon through the same general diagrams. Now they are sliced vertically by a fixed mean signal to noise ratio $\overline{\lambda}$.



In diagram d), the small p/n ratio with the given $\overline{\lambda}$ produces a $\mu_Q(=\overline{Q})$ as shown. Compared to models with $\overline{\lambda} = 0$, this is a very high Q score. It is highly significant. In e), p has increased so that we have an even higher μ_Q than before, but relative to the $\overline{\lambda} = 0$ model's Q scores, it is not as impressive (yet still good). This is because μ_Q is just outside the significance interval for $\overline{\lambda} = 0$. For the choice of p in f), where p is quite near n, μ_Q is considerably greater than the two previous μ_Q's, but now it is quite indistinguishable from run-of-the-mill Q scores produced by a model with $\overline{\lambda} = 0$.

C. The complementary model, skill strategies

Both sets of diagrams show the scientist how to increase the number p of predictors while monitoring model skill or significance: In diagrams a)-c) the underlying $\overline{\lambda}$ is not known. But the scientist has a certain significant level of canonic skill Q he wishes to achieve by a model he wants to be significant. He then increases p until that Q is still produced by a just-significant model, i.e., $\overline{\lambda}_1$ is still greater than 0. In diagrams d)-f) the scientist knows or has estimated $\overline{\lambda}$. He knows the model is significant. He wishes to maximize the probability of occurrence of the model's average skill level $\overline{Q}$ and yet know that $\overline{Q}$ is produced only by a significant model. So he stops the growth of p just short of where $\sigma(1-\frac{1}{2}\alpha)$ engulfs μ_Q.

D. An indeterminacy principle

There is an indeterminacy, as we have just seen, in the skill and significance of a linear regression model, wherein any attempt to increase hindcast skill is offset by a move of the model toward insignificance. The sample size n sets the background over which these antithetical tendencies of skill and significance move. The greater n, the sharper is the background and the smaller the uncertainties induced by changing the predictor count p (recall σ_Q^2 in (9.12)). Let us measure this background uncertainty by the reciprocal of the norm of the n-vector $\underline{\varepsilon}$:

$$\frac{1}{||\underline{\varepsilon}||^2} \tag{13.1}$$

Out of this background chaos we split apart two opposing factors: one factor represents the *viability* of the model, a meld of all the skill measures of §7; the other factor represents the *significance* of the model, a measure, as its name implies, of its roots in determinacy. Thus we split (13.1) into

$$\boxed{\frac{1}{||\underline{\varepsilon}||^2} = \frac{||\underline{X\hat{\beta}}||^2}{||\underline{\varepsilon}||^2} \cdot \frac{1}{||\underline{X\hat{\beta}}||^2}} \tag{13.2}$$

$$= (\text{viability})\cdot(\text{significance})$$

The *viability* factor, as a reference to the linear regression diagram in §6 shows, uses the numerator of classic skill S, and the extension $(||\underline{\varepsilon}_p||^2 + ||\underline{\varepsilon}_{n-p}||^2)$ of the residual noise $||\underline{\varepsilon}_{n-p}||^2$ used in the denominators of the canonic and coskills. The *significance* factor uses the estimate of the signal $||\underline{X\beta}||^2$ occurring in the signal to noise ratio $||\underline{X\beta}||^2/\sigma^2$. We now see that, as n is held fixed, an increase of the number of predictors p will increase the viability of the model and decrease its significance; and conversely, decreasing p will decrease its viability but increase its significance. The product of viability and significance is a fixed random variable whose variance is a measure of the statistical uncertainties produced by the background noise.

The split in (13.2) is not unique. But any way one cares to split $1/||\underline{\varepsilon}||^2$, using p-dependent pieces, one comes up with something like a viability and a significance, to wit:

$$\boxed{\frac{1}{||\underline{\varepsilon}||^2} = \frac{||\hat{\underline{\beta}}||^2}{||\underline{\varepsilon}||^2} \cdot \frac{1}{||\hat{\underline{\beta}}||^2}} \tag{13.3}$$

$$= \text{(viability)} \cdot \text{(significance)}$$

E. The roots of indeterminacy

The preceding examples of indeterminacy are somewhat forced and artificial. Nevertheless they and their immediate variants cannot be formulated without the statistical tendency for various quantities in E_p to spread as p increases. For example, the most fundamental manifestation of this spread is evident in the series of graphs of pdfs for Q (the series of Figures Q-0 to Q-10). In the sub-series that shows how (for fixed n, λ) the pdfs spread their Q-mass on the interval (0, ∞) with increasing p, we see the indeterminacy at work in its most basic way: in order, as p increases relative to n, to let Q reach the higher values, the sharp Q-distribution peaks for small p must be replaced by the broad shallow Q-humps for large p (recall (8.1), (8.3)). At the same time and for the same fundamental reason, the random quantity $||\hat{\underline{\beta}}||^2$ on the average grows as p increases (recall (A33), (A43)) simulating a random walk in spaces E_p of ever larger dimensions, making the location of $\underline{\beta}$, relative to $\hat{\underline{\beta}}$, harder to pin down.

14. Description of the Tables for Q, I, S Significance Levels

There are three sets of tables: one each for Q, I, and S, the canonic skill, ineptness, and classic skill, respectively (cf §7). For each of Q, I, and S we list $\sigma(05)$, $\sigma(95)$ and its mean on separate tables, for a variety of p, n, and λ values. The λ values are 0.0, 0.2, 0.3, 0.5, 0.7, 1.0, 1.5, and 2.0.

For example, consider the tables for canonic skill Q. Let $\lambda = 0.0$. Then there are three tables given for this value of λ: one for $\sigma(05)$, one for $\overline{Q}$, the mean of Q, and one for $\sigma(95)$. For example, the table for $\sigma(05)$ of Q lists p values across the top and n values down the left side. The tables were made using (B2) with (11.1), (11,2), and setting $\alpha = 0.10$. For instance, still with Q, we find $\sigma(05) = 0.288$ for n=6, p=4, λ=0.0, while $\sigma(95) = 38.494$ for the same triple of parameters. Note that the mean $\overline{Q}$ does not exist for this triple (because we must have $n-p>2$; recall (8.2)). However, $\overline{Q}$ exists for n=8, p=4, λ=0.0 and is $\overline{Q} = 2.000$.

The tables are included to show in a preliminary way the ranges of the 5% and 95% significance levels for the random variables Q, I, S under the assumption of zero-centered homogeneous-variance, gaussian noise (cf (A1)). The tables are not exhaustive, and perhaps not in their best form for practice, which would use $\overline{\lambda}$ rather than λ (cf §9). Probably the best way for a user of the present theory to retain knowledge of $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ and the mean of these performance indexes, would be not in tabular form but in the form of a computer program that would fire up the confidence limits $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ at will for any triple p, n, $\overline{\lambda}$, within reason. The formulas in appendix B have been tested for n up to 50 and λ up to 2.0.

15. Acknowledgments

The author's research done in this work was supported by Director John Apel of the Pacific Marine Environmental Laboratory, E.R.L., of NOAA, Seattle, Washington. The programming was done by Peter Crill who was supported under Grant NOAA04-8-MOI-188 of the NOAA Climate Program Office, Washington, D.C. Grace Johnston provided editorial assistance. Conversations over the past year with Tim Barnett continued to renew my enthusiasm for carrying through the work. The recent works of Barnett and Hasselmann, and of Davis were constant guides in my attempts to set up the foundations of the subject.

16 References

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

Barnett, T.P. and K. Hasselmann, 'Techniques of linear prediction, with application to oceanic and atmospheric fields in the tropical Pacific', Reviews of Geophysics and Space Physics, 17(5) 949-968. (1979).

Davis, R. E., 'Predictability of Sea Level Pressure Anomalies over the North Pacific Ocean', J. Phys. Oceanogr., 8, 233 (1978).

Hoel, P. G., Introduction to Mathematical Statistics, John Wiley & Sons, N.Y., (1954), 2nd Ed.

Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol 2 (3rd Ed), Hafner Publ. Co., N.Y. (1973).

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction, Rep. 1, Statist. Forecasting Proj., Mass. Inst. Tech., Cambridge.

Price, R., 'Some non-Central F-distributions expressed in closed form', Biometrika, 51, 107 (1964).

Rao, C. R., Linear Statistical Inference and its Applications, John Wiley & Sons, N.Y. (1973), 2nd Ed.

Tang, P. C., 'The Power Function and the Analysis of Variance Tests with Tables and Illustrations of their Use', Stat. Res. Mem., 2, 126 (1938).

Appendix A, Derivations of the Fundamental Probability Density Functions

The derivations below are of the basic probability densities needed in the study of model significance and skill in the present Linear Regression theory. Our observations in §§4,5 showed that we may base all our formulas on the uncorrelated, zero-mean, uniform-variance case. In the present work we shall therefore assume that the noise vector $\underline{\varepsilon}$ is an n dimensional random variate such that

$$\langle\underline{\varepsilon}\rangle = \underline{0}, \quad \langle \underline{\varepsilon\varepsilon}^T \rangle = \sigma^2 \underline{I}_n$$

i.e.,

$$\langle\varepsilon_i\rangle = 0 \qquad i = 1, \ldots, n$$

$$\langle\varepsilon_i\varepsilon_j\rangle = \sigma^2\delta_{ij} \qquad i, j=1, \ldots, n ,$$

and in particular that:

$$P(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)d\varepsilon_1 d\varepsilon_2 \ldots d\varepsilon_n = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}(\varepsilon_1{}^2+\varepsilon_2{}^2+\ldots+\varepsilon_n{}^2)\}d\varepsilon_1 d\varepsilon_2 \ldots d\varepsilon_n$$

i.e., we assume

$$\underline{\varepsilon} \sim N_n(\underline{0},\sigma^2\underline{I}_n) . \qquad \text{(A1)}$$

The coordinate system and units in which we work are originally defined by the physical setting from which the data are drawn.

1. χ^2 Distribution and Gamma Distribution for $||\varepsilon||^2/\sigma^2$

The error vector $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$, obeys (A1) and we wish to find the distribution of $||\underline{\varepsilon}|^2 = \varepsilon_1{}^2+\varepsilon_2{}^2+\ldots+\varepsilon_n{}^2$. Since $||\underline{\varepsilon}|^2$ depends only on the length of $\underline{\varepsilon}$ and not its orientation in E_n, we introduce polar coordinates in E_n:

$$\varepsilon_1 = r \cos\phi_1$$

$$\varepsilon_2 = r \sin\phi_1 \cos\phi_2$$

$$\vdots$$

$$\varepsilon_{n-1} = r \sin\phi_1 \sin\phi_2 \ldots \sin\phi_{n-2} \cos\phi_{n-1}$$

$$\varepsilon_n = r \sin\phi_1 \sin\phi_2 \ldots \sin\phi_{n-2} \sin\phi_{n-1}$$

From this,

$$r^2 = \varepsilon_1{}^2+\varepsilon_2{}^2+\ldots+\varepsilon_n{}^2.$$

This is the generalization of the familiar case for n = 3:

$$\varepsilon_1 = r \cos\phi_1$$

$$\varepsilon_2 = r \sin\phi_1 \cos\phi_2$$

$$\varepsilon_3 = r \sin\phi_1 \sin\phi_2$$

In making the change of variables, the differentials of volume are related by

$$d\varepsilon_1 d\varepsilon_2 \ldots d\varepsilon_n = \frac{\partial(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)}{\partial(r_1, \phi_1, \ldots, \phi_{n-1})} dr\, d\phi_1 \ldots d\phi_{n-1}$$

$$= r^{n-1}\, dr\, d\Omega_{n-1}$$

where $d\Omega_{n-1}$ is the differential of area of the unit sphere in E_n. For n = 3, $d\Omega_2 = \sin\phi_1\, d\phi_1 d\phi_2$, and in E_3 this quantity is usually called an 'element of solid angle'. Hence (A1) can be written

$$P(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)d\varepsilon_1 d\varepsilon_2\ldots d\varepsilon_n = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}n}} \exp\{\frac{-r^2}{2\sigma^2}\}\, r^{n-1} dr d\Omega_{n-1} \qquad (A2)$$

From n dimensional geometry*

$$\int_S d\Omega_{n-1} = \frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \qquad (A3)$$

where S is the rectangle in n-1 dimensional ϕ space such that $0 \le \phi_i \le \pi$, i = 1, ..., n-2, and $0 \le \phi_{n-1} \le 2\pi$. Integrating $d\Omega_{n-1}$ over this rectangle is equivalent to integrating the ε_i over the unit sphere in E_n. Thus using (A3) in (A2) we find the probability element for $r^2 = ||\underline{\varepsilon}||^2$:

$$P(r^2)d(r^2) \equiv \frac{(\frac{1}{2\sigma^2})^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \exp\{-(\frac{1}{2\sigma^2})r^2\}\, (r^2)^{\frac{1}{2}n-1}\, d(r^2) \qquad (A4a)$$

or

$$Q((r/\sigma)^2)d(r/\sigma)^2 \equiv \frac{(\frac{1}{2})^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \exp\{-\tfrac{1}{2}(r/\sigma)^2\}\, [(r/\sigma)^2]^{\frac{1}{2}n-1} d(r/\sigma)^2 \qquad (A4b)$$

This shows that the pdf of r^2 goes most naturally into the form for $(r/\sigma)^2$, a dimensionless variable. So (A4b) can be written, without $d(r/\sigma)^2$, as:

---

* See, e.g., (Anderson, 1958, p176).

$$\chi^2(x|n) \equiv \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} \exp\{-\tfrac{1}{2}x\}\, x^{\frac{1}{2}n-1} \qquad \text{(A5)}$$

where we set

$$\text{'x' for } (r/\sigma)^2 \quad (=||\underline{\varepsilon}||^2/\sigma^2)$$

Equation (A5) has the familiar form of the χ^2-distribution. Both (A4a) and (A4b), along with (A5) fall under the general form of the gamma distribution:

$$G(x|\alpha,p) \equiv \frac{\alpha^p}{\Gamma(p)} \exp\{-\alpha x\} x^{p-1}\ ,\ 0<x<\infty \qquad \text{(A6)}$$

The transition from one form of (A4) to another is facilitated by the general property

$$G(kz|\alpha,p) = \frac{1}{k}\, G(z|k\alpha,p) \qquad \text{(A7)}$$

where $k = 1/\sigma^2$. Another useful property of (A6) is readily verified by direct calculation:

$$G(x|\alpha,p+q) = \int_0^x G(y|\alpha,p)G(x-y|\alpha,q)dy \qquad \text{(A8)}$$

The verification requires the use of the beta function.
The connection between the χ^2 and G notation is:

$$\chi^2(x|n) = G(x|\tfrac{1}{2},\tfrac{1}{2}n) \qquad \text{(A9)}$$

or in function form:

$$\chi^2(n) = G(\tfrac{1}{2},\tfrac{1}{2}n)$$

Thus the main result of this section may be stated as

$$||\underline{\varepsilon}||^2/\sigma^2 \sim \chi^2(n)$$

or $$||\underline{\varepsilon}||^2/\sigma^2 \sim G(\tfrac{1}{2},\tfrac{1}{2}n) \qquad \text{(A10)}$$

or $$||\underline{\varepsilon}||^2 \sim G(\frac{1}{2\sigma^2}, \tfrac{1}{2}n)$$

2. χ^2 Distributions for $||\underline{\varepsilon}_p||^2/\sigma^2$ and $||\underline{\varepsilon}_{n-p}||^2/\sigma^2$

We now derive the pdf's of $||\underline{\varepsilon}_p||^2/\sigma^2$ and $||\underline{\varepsilon}_{n-p}||^2/\sigma^2$. The noise vectors $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$, as defined in (4.11), (4.12), are n dimensional. They are formed by projecting the n dimensional noise vector $\underline{\varepsilon}$ onto the subspaces E_p, E_{n-p} of E_n. The vectors $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ are in E_p and E_{n-p}, respectively, and as $\underline{\varepsilon}$ twitters about in E_n, these vectors $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$ are confined to their respectively smaller dimensioned spaces. This almost by itself is enough to assure that e.g., $\underline{\varepsilon}_p$ is a p dimensional gaussian variate, but its n dimensionality must be stripped down to p dimensionality to be perfectly sure about this, and the uncorrelatedness of the components of $\underline{\varepsilon}_p$ and $\underline{\varepsilon}_{n-p}$ in their respective spaces must be checked out before we can apply the result (A10) of §1, above.

Consider first the matrix $\underline{P}$. The nxn projection matrix $\underline{P}$ is symmetric (cf 4.9b) and hence by (5.4) has a set of n orthonormal eigenvectors and associated eigenvalues. Since $\underline{P}$ has rank p, only p of those eigenvalues are not zero.

Those that are not zero are all of unit value. This may be seen by operating on an eigenvector $\underline{b}$ of $\underline{P}$. By definition of $\underline{b}$ and λ, $\underline{Pb} = \lambda\underline{b}$. Operating on each side of this equation, with $\underline{P}$ and using (4.9c), $\underline{PPb} = \underline{P}\lambda\underline{b}$, so $\underline{Pb} = \lambda^2\underline{b}$. Therefore $\lambda^2 = \lambda$, i.e., $\lambda(\lambda-1) = 0$, so that the eigenvalues of $\underline{P}$ are either 0 or 1. Let $\underline{b}_1, \ldots, \underline{b}_p$ be any set of eigenvectors associated with the unit eigenvalues. This set is not unique, but can be fixed in any of several ways* (the remaining eigenvectors also arise in an infinite number of ways* — they lie in E_{n-p}). Note that these $\underline{b}_j$ are in general distinct from the $\underline{x}_j$ in §4A, for the latter are generally not orthonormal. Thus $\underline{Pb}_j = \underline{b}_j$ for $j = 1, \ldots, p$. Let $\underline{B}_p = [\underline{b}_1\underline{b}_2\ldots\underline{b}_p]$ be the nxp matrix of these eigenvectors. Then $\underline{B}_p^T \underline{B}_p = \underline{I}_p$, which states compactly that $\underline{b}_i^T \underline{b}_j = \delta_{ij}$, $i, j = 1, \ldots, p$. Moreover, we find, $\underline{PB}_p = \underline{B}_p$ and $\underline{B}_p^T\underline{P} = \underline{B}_p^T$.

Consider next the matrix I-P. This, too, is a projection matrix, symmetric of rank (n-p). Hence it has n-p eigenvectors $\underline{b}_{p+1}, \ldots, \underline{b}_n$ with unit eigenvalues, such that $(\underline{I}-\underline{P})\underline{b}_j = \underline{b}_j$, $j = p+1, \ldots, n$. Let $\underline{B}_{n-p} = [\underline{b}_{p+1}, \ldots, \underline{b}_n]$ be the n x (n-p) matrix of these eigenvectors. Then $\underline{B}_{n-p}^T \underline{B}_{n-p} = \underline{I}_{n-p}$. Moreover, $(\underline{I}-\underline{P})\underline{B}_{n-p} = \underline{B}_{n-p}$, and $\underline{B}_{n-p}^T(\underline{I}-\underline{P}) = \underline{B}_{n-p}^T$.

By our observations in §4A, since every $\underline{b}_j$ in $\underline{B}_p$ is of the form $\underline{Pb}_j$, and every $\underline{b}_j$ in $\underline{B}_{n-p}$ is of the form $(\underline{I}-\underline{P})\underline{b}_j$, it follows that $\underline{B}_p^T\underline{B}_{n-p} = \underline{0}_{px(n-p)}$, the p x (n-p) zero matrix; and also that $\underline{B}_{n-p}^T\underline{B}_p = \underline{0}_{(n-p)xp}$, the (n-p)xp zero matrix. In like manner, $\underline{PB}_{n-p} = \underline{0}_{nx(n-p)}$, $(\underline{I}-\underline{P})\underline{B}_p = \underline{0}_{nxp}$. Companion relations follow on taking transposes of each side of these equations and using $\underline{P} = \underline{P}^T$.

---

* To fix the $\underline{b}_j$, j=1, ..., p, we observe that the numerical construction of the $\underline{b}_j$, j=1, ..., p, can arise automatically in the singular decomposition of the data matrix $\underline{X} = \sum_{j=1}^{p} \underline{a}_j\underline{e}_j^T = \sum_{j=1}^{p} \ell_j^{1/2}\underline{b}_j\underline{e}_j^T$, (cf §5A, §9). The construction of $\underline{B}_{n-p}$, however, is not uniquely guided by the data, and may be done in any of several ways.

We next construct the nxn matrix $\underline{B} = [\underline{B}_p \ \underline{B}_{n-p}]$, and observe that

$$\underline{B}^T\underline{B} = \begin{bmatrix} \underline{B}_p^T \\ \\ \underline{B}_{n-p}^T \end{bmatrix} [\underline{B}_p \ \underline{B}_{n-p}] = \left[\begin{array}{c:c} \underline{I}_p & \underline{0}_{px(n-p)} \\ \hdashline \underline{0}_{(n-p)xp} & \underline{I}_{n-p} \end{array}\right] = \underline{I}_n$$

Hence the n column vectors comprising $\underline{B}$ form an orthonormal basis of E_n. This also means that B^T and $\underline{B}$ are mutual inverses. In particular $\underline{BB}^T = \underline{I}_n$ also. This can be verified alternately by noting that $\underline{BB}^T = \sum_{j=1}^{n} \underline{b}_j\underline{b}_j^T$, which acts like $\underline{I}_n$ for every $\underline{y}$ in E_n. The operation $B^T\underline{\varepsilon}$ finds the components of the noise vector in the new coordinate frame. Using the composite form of $\underline{B}$, we find

$$\underline{B}^T\underline{\varepsilon} = \begin{bmatrix} \underline{B}_p^T \\ \underline{B}_{n-p}^T \end{bmatrix} \underline{\varepsilon} = \begin{bmatrix} \underline{B}_p^T\underline{\varepsilon} \\ \underline{B}_{n-p}^T\underline{\varepsilon} \end{bmatrix} \equiv \begin{bmatrix} \underline{\delta}_p \\ \underline{\delta}_{n-p} \end{bmatrix} \equiv \underline{\delta}$$

Here $\underline{\delta}_p = (\delta_1, \ldots, \delta_p)^T$ is a p component vector and $\underline{\delta}_{n-p} = (\delta_{p+1}, \ldots, \delta n)^T$ an (n-p) component vector. From the orthonormality of $\underline{B}$, we find

$$\delta_1^{\,2} + \ldots + \delta_p^{\,2} + \delta_{p+1}^{\,2} + \ldots + \delta_n^{\,2} = \underline{\delta}^T\underline{\delta} = (\underline{B}^T\underline{\varepsilon})^T(\underline{B}^T\underline{\varepsilon})$$

$$= \underline{\varepsilon}^T\underline{BB}^T\ \underline{\varepsilon}$$

$$= \underline{\varepsilon}^T\underline{\varepsilon} = \varepsilon_1^{\,2} + \ldots + \varepsilon_n^{\,2}$$

Now the transformation $\underline{B}^T$ from $\underline{\varepsilon}$ to $\underline{\delta}$ is volume-preserving in E_n (the determinant of $\underline{B}$ is unity — since $|\underline{B}^T\underline{B}| = |\underline{B}^T||\underline{B}| = |\underline{B}|^2 = |\underline{I}_n| = 1$). Hence the pdf (A1) of $\underline{\varepsilon}$ is identical in form for $\underline{\delta}$. Thus the δ_j, j=1,...,n are pairwise uncorrelated, zero mean gaussian variates of uniform variance σ^2. That is

$$\underline{\delta} \sim N_n(\underline{0}, \sigma^2 \underline{I}_n) \tag{A11}$$

and in particular $\delta_j \sim N(0, \sigma^2) \quad i = 1, \ldots, n$ (A12)

and also that $\underline{\delta}_p \sim N_p(\underline{0}, \sigma^2 \underline{I}_p)$ (A13)

$$\underline{\delta}_{n-p} \sim N_{n-p}(\underline{0}, \sigma^2 \underline{I}_{n-p}) \tag{A14}$$

all of which may be read off from (A1) now with δ_j replacing ε_j, $j = 1, \ldots, n$. Moreover, $\underline{\delta}_p$ and $\underline{\delta}_{n-p}$ are independent.

It follows from §1, in particular (A10), that

$$||\underline{\delta}_p||_p^2/\sigma^2 \sim \chi^2(p) \tag{A15}$$

$$||\underline{\delta}_{n-p}||_{n-p}^2 /\sigma^2 \sim \chi^2(n-p) \tag{A16}$$

where the subscripts on the norm bars remind us that the sums they represent run over p, and n-p terms, respectively.

The final step observes that, from the definition of $\underline{\varepsilon}_p$ in (4.11),

$$\underline{B}^T \underline{\varepsilon}_p = \underline{B}^T \underline{P}\, \underline{\varepsilon} = \begin{bmatrix} \underline{B}_p^T \\ \underline{B}_{n-p}^T \end{bmatrix} \underline{P}\, \underline{\varepsilon} = \begin{bmatrix} \underline{B}_p^T \underline{P} \\ \underline{B}_{n-p}^T \underline{P} \end{bmatrix} \underline{\varepsilon}$$

$$= \begin{bmatrix} \underline{B}_p^T \\ \underline{0}_{(n-p)\times n} \end{bmatrix} \underline{\varepsilon} = \begin{bmatrix} \underline{B}_p^T \underline{\varepsilon} \\ \underline{0} \end{bmatrix} = \begin{bmatrix} \underline{\delta}_p \\ \underline{0} \end{bmatrix}$$

Hence

$$||\underline{\varepsilon}_p||^2 = ||\underline{B}^T\ \underline{\varepsilon}_p||^2 = ||\underline{\delta}_p||^2_p,$$

and so $||\varepsilon_p||^2/\sigma^2$ is indentically distributed as $||\underline{\delta}_p||^2{}_p/\sigma^2$. Therefore, by (A15), and a closely analogous argument* for $\underline{\varepsilon}_{n-p}$, we find

$$||\underline{\varepsilon}_p||^2/\sigma^2 \sim \chi^2(p) \tag{A17}$$

$$||\underline{\varepsilon}_{n-p}||^2/\sigma^2 \sim \chi^2(n-p) \tag{A18}$$

which was to be shown.

* i.e., replace $\underline{P}$ by $\underline{I}$-$\underline{P}$ in the preceding argument, i.e., use $\underline{\varepsilon}_{n-p} = (\underline{I}-\underline{P})\underline{\varepsilon}$ from (4.12).

3. Theory of the non central χ^2 distribution

We now pause to develop the pdf for variates of the form $||\underline{x}||^2/\sigma^2$ where $\underline{x} \sim N_p(\underline{\mu}, \sigma^2 \underline{I}_p)$, i.e., $\underline{x}$ is normally distributed such that its p components are uncorrelated but not of zero mean. While accounts of the theory of such $\underline{x}$ exist in the literature,* there is not readily available a single, simply-connected derivation to my liking; and since the non central χ^2 distribution is crucial to our further derivations we keep the arguments of this appendix essentially self-contained by the observations summarized in this section. The work will proceed in four stages: the first stage sets up the one dimensional case; the second stage reduces the general p dimensional case to the one dimension and the central χ^2 cases; the third stage combines these special cases into the general; and in the fourth stage we develop formulas for all the moments of the non central χ^2 distribution.

Stage 1: *Let* $x \sim N(\mu,\sigma^2)$, *i.e., let the scalar-valued random variable* x *be distributed normally with mean* μ *and variance* σ^2. *We are interested in the pdf of* $y = x^2$.

Thus, by hypothesis:

$$P(x)\,dx = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}\,dx$$

From the change of variable $y = x^2$, we have $dy = 2\,x\,dx = \pm 2y^{\frac{1}{2}}\,dx$. As x varies over $(-\infty,\infty)$ it causes the positive square root $y^{\frac{1}{2}}$ to vary only over $(0,\infty)$ unless we also select $-y^{\frac{1}{2}}$ to cover the case when x is in $(-\infty,0)$. Hence the pdf of y is

---

* See, e.g., (Rao, 1973, p181). This reference suggested the main line of derivation below. However, our treatment, in stage 4 below, of the problem of the moments of χ^2, seems new.

$$p(x)dx = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{(y^{\frac{1}{2}}-\mu)^2}{2\sigma^2}\right\} \frac{dy}{2y^{\frac{1}{2}}}$$

$$+ \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{(-y^{\frac{1}{2}}-\mu)^2}{2\sigma^2}\right\} \frac{dy}{2y^{\frac{1}{2}}}$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma y^{\frac{1}{2}}} e^{-(y+\mu^2)/2\sigma^2} \left[\frac{e^{y^{\frac{1}{2}}\mu/\sigma^2} + e^{-y^{\frac{1}{2}}\mu/\sigma^2}}{2}\right] dy$$

i.e.,

$$p(x)dx = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma y^{\frac{1}{2}}} e^{-(y+\mu^2)/2\sigma^2} \cosh\left(\frac{y^{\frac{1}{2}}\mu}{\sigma^2}\right) dy \qquad (A19)$$

This is one form of the required pdf. However, we may place it into a form that uses the gamma distribution, something which will facilitate later manipulations. Thus, we expand the cosh term into an infinite series:

$$\cosh\left(\frac{y^{\frac{1}{2}}\mu}{\sigma^2}\right) = \sum_{r=o}^{\infty} \frac{\left(\frac{y^{\frac{1}{2}}\mu}{\sigma^2}\right)^{2r}}{(2r)!} = \sum_{r=o}^{\infty} \frac{\left(\frac{y}{2\sigma^2}\right)^r \cdot 2^{2r} \cdot \left(\frac{\mu^2}{2\sigma^2}\right)^r}{\Gamma(2r+1)} \qquad (a)$$

and write

$$\frac{1}{(2\pi)^{\frac{1}{2}}\sigma y^{\frac{1}{2}}} = \frac{1}{2\pi^{\frac{1}{2}}\left(\frac{y}{2\sigma^2}\right)^{\frac{1}{2}}\sigma^2} \qquad (b)$$

Moreover, using the duplication formula for gamma functions:

$$\Gamma(2j) = \frac{2^{2j-1}}{\pi^{\frac{1}{2}}} \cdot \Gamma(j)\Gamma(j+\tfrac{1}{2}) \qquad (c)$$

we can write

$$\Gamma(2r+1) = \Gamma(2[r+\tfrac{1}{2}])$$

$$= \frac{2^{2r}}{\pi^{\frac{1}{2}}}\,\Gamma(r+\tfrac{1}{2})\Gamma(r+1) \qquad \text{(d)}$$

Using (a)-(d), (A19) can be written as

$$p(x)dx = e^{-\mu^2/2\sigma^2} \sum_{r=o}^{\infty} \frac{(\frac{\mu^2}{2\sigma^2})^r}{r!} \cdot \left[\frac{e^{-y/2\sigma^2}\,(y/\sigma^2)^{r-\frac{1}{2}}}{2^{r+\frac{1}{2}}\,\Gamma(r+\frac{1}{2})}\right] \cdot d(y/\sigma^2)$$

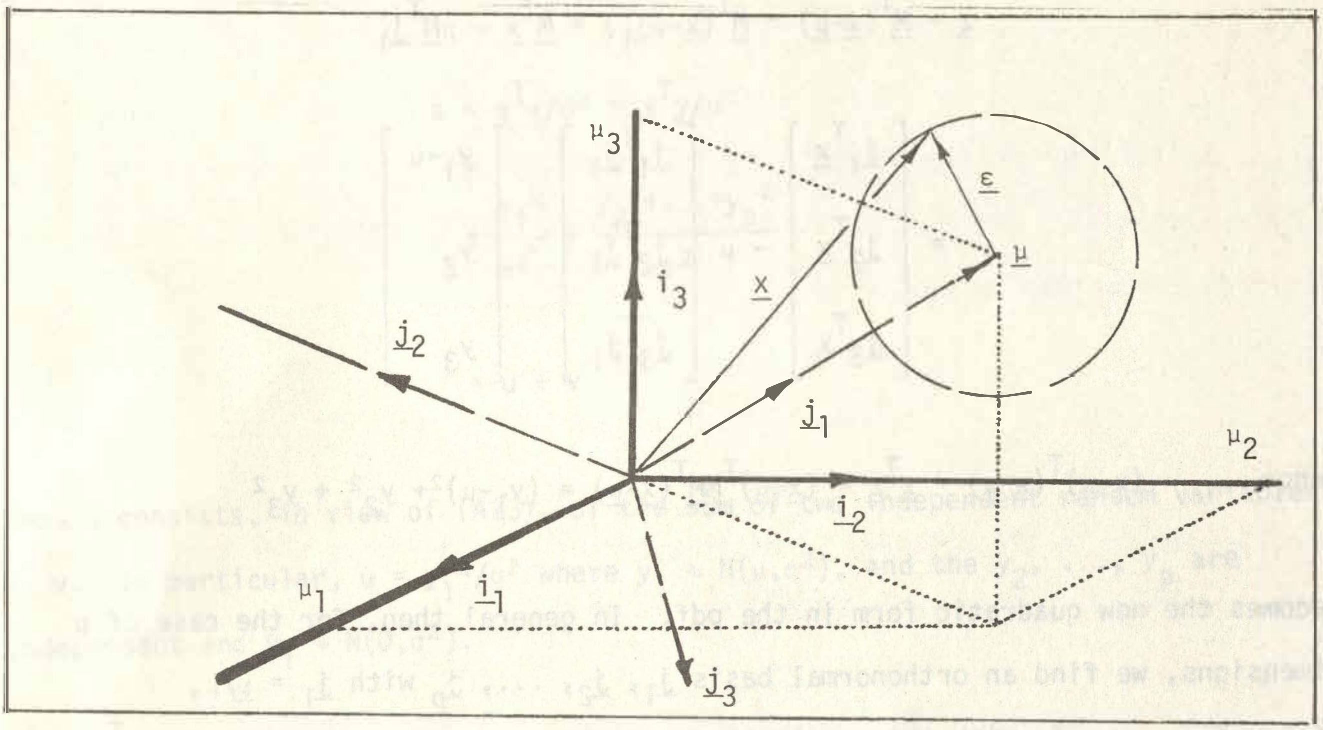
$$= e^{-\mu^2/2\sigma^2} \sum_{r=o}^{\infty} \frac{(\frac{\mu^2}{2\sigma^2})^r}{r!} \cdot G(y/\sigma^2 \mid \tfrac{1}{2},\, r+\tfrac{1}{2}) \cdot d(y/\sigma^2) \qquad \text{(A20)}$$

$$= e^{-\mu^2/2\sigma^2} \sum_{r=o}^{\infty} \frac{(\frac{\mu^2}{2\sigma^2})^r}{r!}\, G(y|1/(2\sigma^2),\, r+\tfrac{1}{2})\, dy\ , \quad 0<y<\infty \qquad \text{(A21)}$$

the last step by (A7). The quantity μ^2/σ^2 will become the signal to noise ratio in a later stage.

Stage 2: Let $x_1, \ldots, x_p$ *be* p *independently distributed gaussian variates each of variance* σ^2 *and mean* $\mu_1, \ldots, \mu_p$, *respectively. That is,* $x_i \sim N(\mu_i,\sigma^2)$, $i = 1, \ldots, p$. *We wish to find the pdf of* $s = \sum_{i=1}^{p} x_i^2/\sigma^2$.

The approach we use is suggested by the following three dimensional case.



The vector $\underline{\mu}$ in the diagram represents the mean position of the random vector $\underline{x} = (x_1, x_2, x_3)^T$. We establish a new orthogonal coordinate frame of unit vectors so that the unit vector $\underline{\mu}/||\underline{\mu}|| \equiv \underline{j}_1$ becomes the first basis vector of that frame. We construct the remaining two $\underline{j}_2$, $\underline{j}_3$, so that $\underline{M} = (\underline{j}_1, \underline{j}_2, \underline{j}_3)$ forms a basis of E_3 in matrix form. Then make the change of variables: $\underline{y} = \underline{M}^T\underline{x}$. The components of $\underline{y}$ are the projections of $\underline{x}$ on $\underline{j}_1$, $\underline{j}_2$, $\underline{j}_3$; hence they are the coordinates of $\underline{x}$ in the new frame. With this change, the vector $\underline{x}-\underline{\mu}$ in the quadratic form $(\underline{x}-\underline{\mu})^T(\underline{x}-\underline{\mu})$ occurring in the pdf:

$$\frac{1}{(2\pi\sigma^2)^{p/2}} \exp\{-\frac{1}{2\sigma^2}(\underline{x}-\underline{\mu})^T(\underline{x}-\underline{\mu})\} \quad \text{(A22)}$$

becomes (with $\mu \equiv ||\underline{\mu}||$, $\underline{y} = \underline{M}^T\underline{x} = (y_1, y_2, y_3)^T$):

$$\underline{z} = \underline{M}^T(\underline{x}-\underline{\mu}) = \underline{M}^T(\underline{x}-\mu\underline{j}_1) = \underline{M}^T\underline{x} - \mu\underline{M}^T\underline{j}_1$$

$$= \begin{bmatrix} \underline{j}_1^T\underline{x} \\ \underline{j}_2^T\underline{x} \\ \underline{j}_3^T\underline{x} \end{bmatrix} - \mu \begin{bmatrix} \underline{j}_1^T\underline{j}_1 \\ \underline{j}_2^T\underline{j}_1 \\ \underline{j}_3^T\underline{j}_1 \end{bmatrix} = \begin{bmatrix} y_1-\mu \\ y_2 \\ y_3 \end{bmatrix}$$

Hence $(\underline{x}-\underline{\mu})^T(\underline{x}-\underline{\mu}) = \underline{z}^T\underline{z} = (\underline{x}-\underline{\mu})^T\underline{MM}^T(\underline{x}-\underline{\mu}) = (y_1-\mu)^2 + y_2^2 + y_3^2$

becomes the new quadratic form in the pdf. In general then, for the case of p dimensions, we find an orthonormal basis $\underline{j}_1, \underline{j}_2, \ldots, \underline{j}_p$ with $\underline{j}_1 = \underline{\mu}/\mu$, $\mu^2 = \mu_1^2 + \ldots + \mu_p^2$, with the result that, on making the transformation $\underline{y} = \underline{M}^T\underline{x}$, $\underline{z} = \underline{M}^T(\underline{x}-\underline{\mu})$,

$$(\underline{x}-\underline{\mu})^T(\underline{x}-\underline{\mu}) = \underline{z}^T\underline{z} = (y_1-\mu)^2 + y_2^2 + \ldots + y_p^2$$

In this way (A22) is transformed to

$$\frac{1}{(2\pi\sigma^2)^{p/2}} \exp\{-\frac{1}{2\sigma^2}[(y_1-\mu)^2 + y_2^2 + \ldots + y_p^2]\}$$

(A23)

$$= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\{-\frac{1}{2\sigma^2}(y_1-\mu)^2\} \cdot \frac{1}{(2\pi\sigma^2)^{(p-1)/2}} \exp\{-\frac{1}{2\sigma^2}(y_2^2 + \ldots + y_p^2)\}$$

Hence, since $\underline{M}$ is orthogonal,

$$s = \underline{x}^T\underline{x}/\sigma^2 = \underline{y}^T\underline{y}/\sigma^2$$

$$= \frac{y_1^2}{\sigma^2} + \frac{y_2^2+\ldots+y_p^2}{\sigma^2}$$

$$= u + v$$

Thus s consists, in view of (A23), of the sum of two independent random variables u, v. In particular, $u = y_1^2/\sigma^2$ where $y_1 \sim N(\mu,\sigma^2)$, and the $y_2, \ldots, y_p$ are independent and $y_i \sim N(0,\sigma^2)$.

Stage 3: Synthesis of Results

From our work in stage 1, the pdf of $u = y_1/\sigma^2$ is given by (A20). From (A10), v is distributed as $\chi^2(p-1)$. Our conclusion in stage 2 was that u and v are independent. Thus the pdf of s is found by convolving the pdfs of u and v, i.e., the pdf of s is, using the pdfs in (A10), (A20):

$$e^{-\mu^2/2\sigma^2} \sum_{r=o}^{\infty} \frac{\left(\frac{\mu^2}{2\sigma^2}\right)^r}{r!} \int_o^s G(u|\tfrac{1}{2},r+\tfrac{1}{2})G(s-u|\tfrac{1}{2},\tfrac{1}{2}(p-1))du \qquad \text{(A24)}$$

The integral may be reduced via (A8). In this way we arrive at the pdf of $s = \sum_{r=1}^{p} x_i^2/\sigma^2$, being of the form

$$\boxed{\chi^2(s|p,\lambda) \equiv e^{-\frac{1}{2}\lambda} \sum_{r=}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} G(s|\tfrac{1}{2}, r+\tfrac{1}{2}p)} \qquad \text{(A25)}$$

where $\lambda = \mu^2/\sigma^2$, $\mu^2 = \mu_1{}^2 + \ldots + \mu_p{}^2$, *and the* x_i *are distributed independently as* $N(\mu_i, \sigma^2)$.

The notation on the left in (A25) is standard for a non central χ^2 distribution with p degrees of freedom and non centrality parameter λ. If the latter is zero, then by (A9), and (A25)

$$\chi^2(s|p,0) \equiv \chi^2(s|p) = G(s|\tfrac{1}{2}, \tfrac{1}{2}p) \tag{A26}$$

i.e., we return to the ordinary χ^2 distribution for a variate s with p degrees of freedom. Written out in full, (A25) is:

$$\chi^2(s|p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \left[\frac{1}{2^{r+\frac{1}{2}p}} \frac{1}{\Gamma(r+\frac{1}{2}p)} e^{-\frac{1}{2}s} s^{r+\frac{1}{2}p-1} \right] \tag{A27}$$

Stage 4: Moments of $\chi^2(x|p,\lambda)$.

We shall need some of the lower moments of a non centrally distributed χ^2 variate. Write

$$'\mu'_m{}' \text{ for } \int_o^{\infty} x^m \chi^2(x|p,\lambda)dx \tag{A28}$$

Hence

$$\mu'_m = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \int_o^{\infty} x^m \, G(x|\tfrac{1}{2}, r+\tfrac{1}{2}p)dx$$

$$= e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{(\frac{1}{2})^{r+\frac{1}{2}p}}{\Gamma(r+\frac{1}{2}p)} \cdot \int_o^{\infty} e^{-\frac{1}{2}x} x^{r+m+\frac{1}{2}-1} dx$$

$$= e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{(\frac{1}{2})^{r+\frac{1}{2}p}}{\Gamma(r+\frac{1}{2}p)} \cdot \frac{\Gamma(r+m+\frac{1}{2}p)}{(\frac{1}{2})^{r+m+\frac{1}{2}p}}$$

Thus the mth moment of $x \sim \chi^2(p,\lambda)$ is:

$$\mu_m' = 2^m e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{\Gamma(r+m+\frac{1}{2}p)}{\Gamma(r+\frac{1}{2}p)} \tag{A29}$$

In particular, we find

$$\mu_1' = 2\, e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} (r+\tfrac{1}{2}p)$$

$$= 2\, e^{-\frac{1}{2}\lambda} \left[\sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot r + \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \tfrac{1}{2}p\right]$$

$$= 2\, e^{-\frac{1}{2}\lambda} [\tfrac{1}{2}\lambda + \tfrac{1}{2}p]\, e^{\frac{1}{2}\lambda}$$

Thus the mean of $x \sim \chi^2(p,\lambda)$ is:

$$\mu_1' = \lambda+p \tag{A30}$$

Moreover, from (A29):

$$\mu_2' = 2^2\, e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} (r+1+\tfrac{1}{2}p)(r+\tfrac{1}{2}p)$$

$$= 2^2\, e^{\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} [r^2 + r(p+1) + \tfrac{1}{4}p^2+\tfrac{1}{2}p]$$

This requires us to sum series of the form:

$$f_n(x) = \sum_{r=o}^{\infty} \frac{x^r}{r!} r^n$$

This may be done as follows:

Now $$\frac{df_n(x)}{dx} = \sum_{r=o}^{\infty} \frac{r\, x^{r-1}\, r^n}{r!} = \sum_{r=1}^{\infty} \frac{x^{r-1}}{(r-1)!} r^n = \sum_{r=o}^{\infty} \frac{x^r}{r!}(r+1)^n$$

$$= \sum_{r=o}^{\infty} \frac{x^r}{r!} \sum_{j=o}^{n} {}^nC_j\, r^j \quad , \quad {}^nC_j = \frac{n,(n-1)\ldots(n-j+1)}{1,2,\ldots j}$$

$$= \sum_{j=o}^{n} {}^nC_j\, f_j(x)$$

i.e.,

$$\frac{df_n(x)}{dx} = \sum_{j=1}^{n-1} {}^nC_j\, f_j(x) + f_o(x) + f_n(x) \quad , \quad n = 2, 3, \ldots \qquad \text{(A31)}$$

This provides a differential equation for $f_n(x)$ in terms of the lower order functions $f_o(x)$, $f_1(x)$, ..., $f_{n-1}(x)$. The chain of equations (A31) can be solved for n = 2, 3, ... since we know that

$$f_o(x) = \sum_{r=o}^{\infty} \frac{x^r}{r!} = e^x$$

and that

$$f_1(x) = \sum_{r=o}^{\infty} \frac{x^r}{r!} r = \sum_{r=1}^{\infty} \frac{x^r}{r!} = \sum_{r=1}^{\infty} \frac{x^r}{(r-1)!} = \sum_{r=o}^{\infty} \frac{x^{r+1}}{r!}$$

$$= x\, e^x$$

Thus we can solve (A31) for the case n = 2,

$$\frac{df_2(x)}{dx} = {}^2C_1\ f_1(x) + f_o(x) + f_2(x) = f_2(x) + (2x+1)e^x$$

subject to the initial condition

$$f_2(o) = 0\ .$$

We see that the general solution is:

$$f_2(x) = f_2(o)e^x + \int_o^x\ [(2t+1)e^t]e^{x-t}\ dt$$

and so

$$f_2(x) = (x+x^2)e^x$$

Returning to μ_2' we find:

$$\mu_2' = 2^2e^{-\frac{1}{2}\lambda}\ [f_2(\tfrac{1}{2}\lambda) + (p+1)f_1(\tfrac{1}{2}\lambda) + (\tfrac{1}{4}p^2+\tfrac{1}{2}p)\ f_o(\tfrac{1}{2}\lambda)]$$

so

$$\boxed{\mu_2' = \lambda^2 + (p+2)(2\lambda+p)} \qquad \text{(A32)}$$

The variance of $x \sim \chi^2(p,\lambda)$ is

$$\mu_2 = \mu_2'-\mu_1'^2$$

which comes out to be

$$\boxed{\mu_2 = 4\lambda + 2p} \tag{A33}$$

Higher moments of $x \sim \chi^2(p,\lambda)$ can be found similarly, using (A31) with (A29).

4. <u>Non Central χ^2 Distributions for $||\underline{y}||^2/\sigma^2$ and $||X\hat{\underline{\beta}}||^2/\sigma^2$.</u>

We consider first the simpler case, namely that of $\underline{y}$. From (4.5), and (A1) it follows at once that $\underline{y} \sim N_n(X\underline{\beta}, \sigma^2 \underline{I}_n)$. Thus each component y_j of $\underline{y}$ has the property $y_j \sim N(\mu_j, \sigma^2)$ where μ_j is the jth component of $X\underline{\beta}$, i.e., $\mu_j = \sum_{k=1}^{p} x_{jk}\beta_k$, $j = 1, \ldots, n$. These y_j are independently distributed, and so by (A25),

$$||\underline{y}||^2/\sigma^2 \sim \chi^2(n,\lambda) \tag{A34}$$

with $\lambda = \mu^2/\sigma^2$, $\mu^2 = \mu_1^2 + \ldots + \mu_n^2$. Here λ is the *signal to noise* ratio.

We will next show that

$$||X\hat{\underline{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda) \tag{A35}$$

This result is plausible because, in view of the diagram in §6, $\underline{X}\,\hat{\underline{\beta}}$, even though it is a vector of n components, moves only in the subspace E_p spanned by the p columns of $\underline{X}$. The main goal of the following argument will be to find a p-component vector which is known to always have the same length as $\underline{X}\,\hat{\underline{\beta}}$ and whose components are independent gaussian variates with mean μ_i and variance σ^2.

We may use for this purpose the basis $\underline{B}$ of E_n constructed in §2 of this appendix. The components of $\underline{X}\,\hat{\underline{\beta}}$ in this frame are reckoned by (recalling (4.7)):

$$\underline{B}^T(\underline{X\hat{\beta}}) = \begin{bmatrix} \underline{B}_p^T \\ \underline{B}_{n-p}^T \end{bmatrix} \underline{X}\,\hat{\underline{\beta}} = \begin{bmatrix} \underline{B}_p^T\underline{X}\,\hat{\underline{\beta}} \\ \underline{B}_{n-p}^T\underline{X}\,\hat{\underline{\beta}} \end{bmatrix} = \begin{bmatrix} \underline{B}_p^T\underline{X}\,\hat{\underline{\beta}} \\ \underline{0}_{(n-p)xn} \end{bmatrix}$$

Hence

$$||\underline{X\hat{\beta}}||^2 = ||\underline{B}^T(\underline{X\hat{\beta}})||^2 = ||\underline{B}_p^T\underline{X\hat{\beta}}||_p^2 \tag{A36}$$

and similarly we can show:

$$||\underline{X\beta}||^2 = ||\underline{B}^T(\underline{X\beta})||^2 = ||\underline{B}_p^T\underline{X\beta}||_p^2 \tag{A37}$$

Now, from (4.4) we find

$$\underline{B}_p^T(\underline{X\hat{\beta}}) = \underline{B}_p^T\,(\underline{X\beta}) + \underline{B}_p^T\,\underline{P}\,\underline{\varepsilon}$$

$$= \underline{B}_p^T\,(\underline{X\beta}) + \underline{B}_p^T\,\underline{\varepsilon}$$

$$= \underline{B}_p^T\,(\underline{X\beta}) + \underline{\delta}_p$$

From (A13), we know that

$$\underline{\delta}_p \sim N_p(\underline{0}, \sigma^2\underline{I}_p)$$

Hence

$$\underline{B}_p^T(\underline{X\hat{\beta}}) \sim N_p(\underline{B}_p^T(\underline{X\beta}), \sigma^2\underline{I}_p) \tag{A38}$$

By this and (A25)

$$||\underline{B}_p^T(\underline{X\hat{\beta}})||_p^2 \sim \chi^2(p,\lambda)$$

where

$$\lambda = ||\underline{B}_p^T(\underline{X\beta})||^2/\sigma^2 = ||\underline{X\beta}||^2/\sigma^2 ,$$

the last step, from (A37).

By (A36) we know that $||\underline{X\hat{\beta}}||^2$ and $||\underline{B}_p^T\underline{X\hat{\beta}}||_p^2$ are distributed identically. Hence,

$$||\underline{X\hat{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda) \tag{A39}$$

with $\lambda = ||\underline{X\beta}||^2/\sigma^2$, as was to be shown.

5. Independence of $||\underline{X\hat{\beta}}||^2/\sigma$ and $||\underline{\varepsilon}_{n-p}||^2$

We now make the observation that $\underline{X}\ \underline{\hat{\beta}}$ and $\underline{\varepsilon}_{n-p}$ are independent variates. This is fairly clear from the linear regression diagram in §6. Since $\underline{\varepsilon}$ is resolved into the independent variates $\underline{\varepsilon}_p$, $\underline{\varepsilon}_{n-p}$, the twitter of $\underline{X\hat{\beta}} = \underline{X\beta} + \underline{\varepsilon}_p$ is due to $\underline{\varepsilon}_p$ only. However, this may also be established formally by using the basis $\underline{B}$ of E_n introduced in §2. Starting with the representation (4.2) of $\underline{y}$; and recalling the definitions of $\underline{\delta}_p$, $\underline{\delta}_{n-p}$ in §2,

$$\begin{bmatrix} \underline{B}_p^T \\ \underline{B}_{n-p}^T \end{bmatrix} \underline{y} = \begin{bmatrix} \underline{B}_p^T \\ \underline{B}_{n-p}^T \end{bmatrix} (\underline{X\hat{\beta}} + \underline{\varepsilon}_{n-p})$$

$$= \begin{bmatrix} \underline{B}_p^T\ \underline{X}\ \underline{\hat{\beta}} \\ \underline{B}_{n-p}^T\ \underline{\varepsilon}_{n-p} \end{bmatrix} = \begin{bmatrix} \underline{B}_p^T\ \underline{X}\ \underline{\beta} + \underline{B}_p^T\ \underline{\varepsilon}_p \\ \underline{B}_{n-p}^T\ \underline{\varepsilon}_{n-p} \end{bmatrix} \quad \text{(via (4.4))}$$

$$= \begin{bmatrix} \underline{B}_p^T\ \underline{X}\ \underline{\beta} + \underline{\delta}_p \\ \underline{\delta}_{n-p} \end{bmatrix}$$

From this we read:

$$\underline{B}_p^T(\underline{X\hat{\beta}}) \sim N_p(\underline{B}_p^T \underline{X\beta}, \sigma^2 \underline{I}_p) \tag{A40}$$

$$B_{n-p}^T(\underline{\varepsilon}_{n-p}) \sim N_{n-p}(\underline{0}, \sigma^2 \underline{I}_{n-p}) \tag{A41}$$

and since, as seen in (A11), (A13), (A14), $\underline{\delta}_p$, $\underline{\delta}_{n-p}$ are independent, the result follows. An immediate corollary of this is that $||\underline{X\hat{\beta}}||^2/\sigma^2$ and $||\underline{\varepsilon}_{n-p}||^2/\sigma^2$ are independent (functions of independent variates are themselves independent).

6. χ^2 Distributions for $||\underline{\hat{\beta}}||^2/\sigma^2$, $||\underline{\hat{\beta}}-\underline{\beta}||^2/\sigma^2$

From (4.15), we can think of $\underline{\hat{\beta}}$ as $\underline{\beta}$ that has been linearly perturbed by $\underline{\varepsilon}$:

$$\underline{\hat{\beta}} = \underline{\beta} + (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}$$

and so we suspect that $\underline{\hat{\beta}}$ will be normally distributed with mean $\underline{\beta}$. To find its covariance matrix we use the

Theorem.* Let $\underline{u} \sim N_n(\underline{\mu},\underline{\Sigma})$, i.e., let $\underline{u}$ be an n dimensional gaussian variate with mean $\underline{\mu}$ and covariance $\underline{\Sigma}$. Define a p dimensional variate $\underline{v} = \underline{C}\,\underline{u}$ by means of a pxn matrix transformation $\underline{C}$. Then $\underline{v} \sim N_p(\underline{C}\underline{\mu}, \underline{C}\,\underline{\Sigma}\,\underline{C}^T)$.

To apply this theorem we return to (4.5) and (A1) and note that $y \sim N_n(\underline{X\beta}, \sigma^2\underline{I}_n)$. Thus $\underline{\mu} = \underline{X}\,\underline{\beta}$ and $\underline{\Sigma} = \sigma^2\underline{I}_n$, for $\underline{y} = u$. Then from (3.8) we have the requisite form of $\underline{C} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T$. By the theorem, $\underline{v} = \underline{\hat{\beta}}$ has mean $\underline{C\mu} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T(\underline{X\beta}) = \underline{\beta}$, and covariance $\underline{C}\,\underline{\Sigma}\,\underline{C}^T = (\underline{X}^T\underline{X})^{-1}X^T(\sigma^2\underline{I}_n)\,\underline{X}\,(\underline{X}^T\underline{X})^{-1} = \sigma^2(\underline{X}^T\underline{X})^{-1}$. Hence

---

* see, e.g. (Rao, 1973, pp 522).

$$\hat{\underline{\beta}} \sim N_p\,(\underline{\beta},\ \sigma^2(\underline{X}^T\underline{X})^{-1}) \tag{A42}$$

From this we see that for a given data matrix $\underline{X}$, the components of $\hat{\underline{\beta}}$ and $\hat{\underline{\beta}}-\underline{\beta}$ are generally correlated. In order to apply χ^2 statistics, e.g., we would adapt X so that $X^T X = I_p$ (cf (5.2)). Then* by (A25),

$$\left.\begin{array}{ll} ||\hat{\underline{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda_o)\ , \quad \lambda_o \equiv ||\underline{\beta}||^2/\sigma^2 & \text{(A43)} \\ \text{or} & \\ ||\hat{\underline{\beta}}-\underline{\beta}||^2/\sigma^2 \sim \chi^2(p) & \text{(A44)} \end{array}\right\} \quad \underline{X}^T\underline{X} = I_p$$

7. The H-Distribution

We now consider the derivation of the pdf underlying the canonic skill Q and ineptness I (cf §7 of the main text). We will pose at the outset a slightly more general problem and then reduce it to the Q and I cases: *Let* x_1, x_2 *be two independent variates such that* $x_1 \sim \chi^2(k_1, \lambda_1)$ *and* $x_2 \sim \chi^2(k_2, \lambda_2)$. *It is required to find the pdf of* x_1/x_2.

The derivation requires the following preliminary observations on transformations of random variables. Suppose x_1, x_2 are two random variables with joint pdf $p(x_1, x_2)$. We wish to make a change of variables from x_1, x_2 to y_1, y_2, where

$$x_1 = f(y_1, y_2)$$

$$x_2 = g(y_1, y_2)$$

---

* The λ_o in (A43) is simply defined to be $||\underline{\beta}||^2/\sigma^2$ for the present application of (A25). However, see the discussion of the quantity $\lambda = ||\underline{X}\underline{\beta}||^2/\sigma^2$ when $\underline{X}^T\underline{X} = I_p$ (cf (9.5)). In that setting, $\lambda_o = \lambda$. Recall also (5.20).

To see how the differential dx_1dx_2 transforms, we compute the differentials

$$dx_1 = f_1\, dy_1 + f_2\, dy_2$$

$$dx_2 = g_1\, dy_1 + g_2\, dy_2$$

where f_i and g_i are derivatives of f with respect to y_i. Then by the calculus of exterior differential forms (or equivalently, Jacobian theory of change of variables):

$$dx_1dx_2 = (f_1dy_1 + f_2dy_2)\ (g_1dy_1 + g_2dy_2)$$

and this is reduced using $(dy_j)^2 = 0$, $(dy_idy_j) = -(dy_jdy_i)$, $i \neq j$.

The element of area dx_1dx_2 thus transforms as

$$dx_1dx_2 = (f_1g_2 - f_2g_1)dy_1dy_2$$

The quantity in parentheses is the Jacobian of the transformation. Hence the related probability elements are

$$p(x_1,x_2)dx_1dx_2 = p(f(y_1,y_2),\ g(y_1,y_2))\ (f_1g_2 - f_2f_1)dy_1dy_2$$

$$\equiv q(y_1,y_2)dy_1dy_2$$

Where q is defined in context, i.e.,

$$q(y_1,y_2) = p(f(y_1,y_2)\ ,\ g(y_1,y_2))\ (f_1g_2 - f_2g_1) \qquad \text{(A45)}$$

Returning to the problem of the distribution of $x_1/x_2 = y_1$, we make the change of variables

$$x_1 = f(y_1, y_2) = y_1 y_2$$

$$x_2 = g(y_1, y_2) = y_2$$

so that the Jacobian is $f_1 g_2 - f_2 g_1 = y_2$

and from (A45):

$$q(y_1, y_2) = p(y_1 y_2,\ y_2) y_2 \,.$$

Since x_1, x_2 are independent, $p(x_1, x_2) = p_1(x_1) p_2(x_2)$, and so

$$q(y_1, y_2) = p_1(y_1 y_2) p_2(y_2) y_2$$

We can now drop the subscript on y_1 and revert from y_2 to x_2. The joint pdf $q(y,\ x_2)$ for $x_1/x_2 = y$ and x_2 is then

$$p_1(y x_2) p_2(x_2) x_2 \qquad \text{(A46)}$$

Now $p_1 = \chi^2(k_1, \lambda_1)$, $p_2 = \chi^2(k_2, \lambda_2)$, by hypothesis. The pdf for y is obtained by integrating (A46) over the range of x_2, namely $(0, \infty)$. Hence from (A46) and (A25):

$$H(y|k_1,k_2,\lambda_1,\lambda_2) \equiv \int_0^\infty p_1(yx_2)p_2(x_2)x_2\,dx_2$$

$$= \int_0^\infty \chi^2(yx_2|k_1,\lambda_1)\chi^2(x_2|k_2,\lambda_2)x_2\,dx_2$$

$$= e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty}\sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!}\frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \underbrace{\int_0^\infty G(yx_2|\tfrac{1}{2},r+\tfrac{1}{2}k_1)G(x_2|\tfrac{1}{2},s+\tfrac{1}{2}k_2)x_2\,dx_2}_{J}$$

Here, using (A6):

$$J = \int_0^\infty \frac{(\frac{1}{2})^{r+\frac{1}{2}k_1}}{\Gamma(r+\frac{1}{2}k_1)} e^{-\frac{1}{2}yx_2}(yx_2)^{r+\frac{1}{2}k_1-1} \cdot \frac{(\frac{1}{2})^{s+\frac{1}{2}k_2}}{\Gamma(s+\frac{1}{2}k_2)} e^{-\frac{1}{2}x_2} x_2^{s+\frac{1}{2}} \,dx_2$$

$$= \frac{(\frac{1}{2})^{r+s+\frac{1}{2}(k_1+k_2)}\, y^{r+\frac{1}{2}k_1-1}}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} \cdot \int_0^\infty e^{-\frac{1}{2}(1+y)x_2}\, x_2^{r+s+\frac{1}{2}(k_1+k_2)-1}\,dx_2$$

Using the known gamma function integral

$$\int_0^\infty e^{-ax}x^n dx = \frac{\Gamma(n+1)}{a^{n+1}}$$

with $a = \frac{1}{2}(1+y)$, $n = r + s + \frac{1}{2}(k_1+k_2)-1$, we find

$$J = \frac{(\tfrac{1}{2})^{r+s+\frac{1}{2}(k_1+k_2)}\, y^{r+\frac{1}{2}k_1-1}}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} \cdot \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{(\tfrac{1}{2})^{r+s+\frac{1}{2}(k_1+k_2)}\,(1+y)^{r+s+\frac{1}{2}(k_1+k_2)}}$$

$$= \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} \cdot \frac{y^{r+\frac{1}{2}k_1-1}}{(1+y)^{r+s+\frac{1}{2}(k_1+k_2)}}$$

In this way we come to the pdf for $x_1/x_2 = y$, where x_1 x_2 are independent non central χ^2 variates, $x_1 \sim \chi^2(k_1,\lambda_1)$, $x_2 \sim \chi^2(k_2,\lambda_2)$:

$$H(y|k_1,k_2,\lambda_1,\lambda_2) =$$

$$= e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} \cdot \frac{y^{r+\frac{1}{2}k_1-1}}{(1+y)^{r+s+\frac{1}{2}(k_1+k_2)}}, \quad \text{(A47)}$$

$$0<y<\infty$$

A. Generalization of H

We can generalize (A47) to account for ratios of the form $y = x_1/x_2 = c_1\xi_1/c_2\xi_2$, i.e., where numerator and denominator are independent variates ξ_1, ξ_2, multiplied by constants c_1, c_2, and where $x_i \sim \chi^2(k_i,\lambda_i)$, i = 1,2. Thus let $y = (c_1/c_2)\ (\xi_1/\xi_2) \equiv \gamma\eta$.

Then in (A47)

$$H(y|k_1,k_2,\lambda_1,\lambda_2)dy = H(\gamma\eta|k_1,k_2,\lambda_1,\lambda_2)\gamma d\eta$$

write

$$\text{'}H'(\eta|k_1,k_2,\lambda_1,\lambda_2,\gamma)\text{'} \text{ for } H(\gamma\eta|k_1,k_2,\lambda_1,\lambda_2)\gamma \qquad \text{(A47a)}$$

This is the required new pdf for the ratio $\eta = \xi_1/\xi_2$, where $\xi_1 = x_1/c_1$, $\xi_2 = x_2/c_2$ and $x_1 \sim \chi^2(k_1,\lambda_1)$, $x_2 \sim \chi^2(k_2,\lambda_2)$. That is, $H'(\eta|k_1,k_2,\lambda_1,\lambda_2,\gamma)$ gives the pdf of ξ_1/ξ_2 where the numerator and denominator each differ by a fixed factor from a pure χ^2 variate. This new pdf H' is found from (A47) by performing on H the indicated operations on the right in (A47a). An example of the use of (A47a) is given in (12.2).

B. <u>The pdf for $Q = ||\underline{X\hat{\beta}}||^2/||\underline{\varepsilon}_{n-p}||^2$</u>

As a special case of (A47), we have from (A35) and (A18), and the fact that $||\underline{X\hat{\beta}}||^2$ and $||\underline{\varepsilon}_{n-p}||^2$ are independent (cf §5, Appendix A), i.e., since

$$x_1 = ||\underline{X\hat{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda)$$

$$x_2 = ||\underline{\varepsilon}_{n-p}||^2/\sigma^2 \sim \chi^2(n-p,0)$$

we can set

$$k_1 = p,\ k_2 = n-p,\ \lambda_1 = \lambda = ||\underline{X\beta}||^2/\sigma^2\ ,\ \lambda_2 = 0$$

and find:

$$P_Q(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{x^{r+\frac{1}{2}p-1}}{(1+x)^{r+\frac{1}{2}n}} \qquad \text{(A48)}$$

$$o \leq x < \infty, \; n > p \geq 1$$

By virtue of the connection between Q and C in §7, i.e., Q = C-1, the pdf for coskill C follows at once from A(48) by replacing 'x' by 'x-1' on the right side and 'P_Q' by 'P_C' on the left. The range of C is (1,∞). P_Q is also known as the 'non central f' distribution (Rao, 1973, p216). The signal to noise ratio λ (as in (A39)) is also known as the 'noncentrality parameter' in advanced statistical theory when no specific physical imagery is available.

C. The pdf for $I = ||\underline{\varepsilon}_{n-p}||^2/||X\hat{\underline{\beta}}||^2$

As a special case of (A47), we have from (A35) and (A18) and the fact that $||X\hat{\underline{\beta}}||^2$ and $||\underline{\varepsilon}_{n-p}||^2$ are independent (cf §5, Appendix A), i.e., since

$$x_1 = ||\underline{\varepsilon}_{n-p}||^2 \sim \chi^2(n-p,o)$$

$$x_2 = ||X\hat{\underline{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda)$$

we can set

$$k_1 = n-p,\; k_2 = p,\; \lambda_1 = 0,\; \lambda_2 = \lambda = ||X\underline{\beta}||^2/\sigma^2 ,$$

and find

$$P_I(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda)^s}{s!} \cdot \frac{\Gamma(s+\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n-p))\Gamma(s+\frac{1}{2}p)} \cdot \frac{x^{\frac{1}{2}(n-p)-1}}{(1+x)^{s+\frac{1}{2}n}} \qquad \text{(A49)}$$

$$o \leq x < \infty, \; n > p \geq 1$$

By virtue of the connection between I and U in §7, i.e., I = U-1, the pdf for unskill U follows at once from (A49) by replacing 'x' by 'x-1' on the right side and 'p_I' by 'P_U' on the left. The range of U is $(1,\infty)$.

The essential difference in the distributions for Q and I is in the exponent of x: there is no summation dummy s in the exponent of x in (A49). The H-distribution appears to have been first studied in (Tang, 1938) and (Price, 1964); cf also (Kendall and Stuart, vol 2, 1973, p262).

8. The J-Distribution

The pdf for classic skill S may be obtained from those of $||\underline{X}\hat{\beta}||^2$ and $||\underline{\varepsilon}_{n-p}||^2$ by observing that $||\underline{y}||^2 = ||\underline{X}\hat{\beta}||^2 + ||\underline{\varepsilon}_{n-p}||^2$ (cf (6.1)). In §5 of this Appendix the independence of the summands was established and we know that $x_1 = ||\underline{X}\hat{\beta}||^2/\sigma^2 \sim \chi^2(p,\lambda)$ and $x_2 = ||\underline{\varepsilon}_{n-p}||^2/\sigma^2 \sim \chi^2(p,0)$. It remains then to deduce the pdf for $y = x_1/(x_1+x_2)$.

We will derive the general pdf for $y = x_1/(x_1+x_2)$ where the independent variates x_1,x_2 are such that $x_1 \sim \chi^2(k_1,\lambda_1)$, $x_2 = \chi^2(k_2,\lambda_2)$. Following the transformation procedure in §7 above, let

$$x_1 = f(y_1,y_2) = \frac{y_1y_2}{1-y_1}$$

$$x_2 = g(y_1,y_2) = y_2$$

The first transformation is motivated by the defining relation $y = x_1/(x_1+x_2)$ solved for x_1 and relabeling x_2 as 'y_2'. The Jacobian of the transformation is

$$f_1g_2-f_2g_1 = \frac{y_2}{(1-y_1)^2}$$

Then

$$p(x_1,x_2)dx_1dx_2 = p(\frac{y_1y_2}{1-y_1},y_2)\,\frac{y_2}{(1-y_1)^2}\,dy_1dy_2$$

$$= p_1(\frac{yx_2}{1-y})p_2(x_2)\,\frac{x_2}{(1-y)^2}\,dy\,dx_2$$

where we have used the independence of x_1, x_2, and reset $y = y_1$ and $x_2 = y_2$. The required pdf for y is then found using $p_1 = \chi^2(k_1,\lambda_1)$, $p_2 = \chi^2(k_2,\lambda_2)$ with (A25):

$$J(y|k_1,k_2,\lambda_1,\lambda_2) = \int_o^\infty p_1(\frac{yx_2}{1-y})\,p_2(x_2)\,\frac{x_2}{(1-y)^2}\,dx_2$$

$$= e^{-\frac{1}{2}(\lambda_1+\lambda_2)}\sum_{r=o}^{\infty}\sum_{s=o}^{\infty}\frac{(\frac{1}{2}\lambda_1)^r(\frac{1}{2}\lambda_2)^s}{r!\,s!}\cdot\underbrace{\left[\int_o^\infty G(\frac{yx_2}{1-y}|\tfrac{1}{2},r+\tfrac{1}{2}k_1)G(x_2|\tfrac{1}{2},s+\tfrac{1}{2}k_2)x_2dx_2\right]}_{K}\cdot\frac{1}{(1-y)^2}$$

Here, using (A6):

$$K = \int_o^\infty \frac{(\frac{1}{2})^{r+\frac{1}{2}k_1}}{\Gamma(r+\frac{1}{2}k_1)}\cdot e^{-\frac{1}{2}[\frac{yx_2}{1-y}]}\cdot[\frac{yx_2}{1-y}]^{r+\frac{1}{2}k_1-1}\cdot\frac{(\frac{1}{2})^{s+\frac{1}{2}k_2}}{\Gamma(s+\frac{1}{2}k_2)}\cdot e^{-\frac{1}{2}x_2}\,x_2^{s+\frac{1}{2}k_2-1}\,x_2dx_2$$

$$= \frac{(\frac{1}{2})^{r+s+\frac{1}{2}(k_1+k_2)}}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)}\cdot[\frac{y}{1-y}]^{r+\frac{1}{2}k_1-1}\cdot\int_o^\infty e^{-\frac{1}{2}x_2[\frac{1}{1-y}]}\,x_2^{r+s+\frac{1}{2}(k_1+k_2)-1}\,dx_2$$

The gamma function integral in §7 of the Appendix can be used here with $a = 1/2(1-y)$, $m = r+s+\frac{1}{2}(k_1+k_2)-1$. Thus

$$K = \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} y^{r+\frac{1}{2}k_1-1} (1-y)^{s+\frac{1}{2}k_2+1}$$

Hence we end up with

$$J(y|k_1,k_2,\lambda_1,\lambda_2) =$$

$$= e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s-o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{\Gamma(r+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)} y^{r+\frac{1}{2}k_1-1} (1-y)^{s+\frac{1}{2}k_2-1} \quad \text{(A50)}$$

$$o \le y \le 1$$

which is the pdf for $y = x_1/(x_1+x_2)$ where $x_1 \sim \chi^2(k_1,\lambda_1)$, $x_2 \sim \chi^2(k_2,\lambda_2)$ and x_1,x_2 are independent.

A. The pdf for $S = ||X\hat{\underline{\beta}}||^2/||\underline{y}||^2$

As a special case of (A50) we have

$$x_1 \sim ||X\hat{\underline{\beta}}||^2/\sigma^2 \sim \chi^2(p,\lambda)$$

$$x_2 \sim ||\underline{\varepsilon}_{n-p}||^2/\sigma^2 \sim \chi^2(n-p,0)$$

and can set

$$k_1 = p,\ k_2 = n-p,\ \lambda_1 = \lambda = ||X\underline{\beta}||^2/\sigma^2\ ,\ \lambda_2 = 0$$

and find

$$P_S(x|n,p,\lambda) = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} x^{r+\frac{1}{2}p-1}(1-x)^{\frac{1}{2}(n-p)-1} \tag{A51}$$

$$0 \le x \le 1, \quad n > p \ge 1$$

By virtue of the connection between R and S in §7 of the main text, i.e., S = 1-R, the pdf for residual unskill R follows at once from (A51) by replacing 'x' by '1-x' on the right side and 'P_S' by 'P_R' on the left. The range of R is (0,1). P_S is also known as the 'non-central beta' distribution (Rao, 1973, p217).

9. Calculation of the Moments of the H and J Distributions

The mth raw moment of $y \sim H(k_1,k_2,\lambda_1,\lambda_2)$ is found from (A47) via

$$\mu'_m = \int_o^{\infty} y^m \, H(y|k_1,k_2,\lambda_1,\lambda_2)dy$$

This requires the evaluation of

$$\int_o^{\infty} \frac{y^{r+m+\frac{1}{2}k_1-1}}{(1+y)^{r+s+\frac{1}{2}(k_1+k_2)}} dy = \frac{\Gamma(r+\frac{1}{2}k_1+m)\Gamma(s+\frac{1}{2}k_2-m)}{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}$$

using a variation of the beta function integrand. Hence in general

$$\mu'_m = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{\Gamma(r+\frac{1}{2}k_1+m)\Gamma(s+\frac{1}{2}k_2-m)}{\Gamma(r+\frac{1}{2}k_1)\ \Gamma(s+\frac{1}{2}k_2)}, \quad m = 0,1,2,\ldots \tag{A52}$$

As special cases of this,

$$\mu_1' = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{(2r+k_1)}{(2s+k_2-2)} \tag{A53}$$

$$\mu_2' = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{(2r+k_1+2)}{(2s+k_2-1)} \frac{(2r+k_1)}{(2s+k_2-2)} \tag{A54}$$

As further special cases, we have

A. First Moment of Q

In (A53) for $Q = ||X\hat{\beta}||^2/||\underline{\varepsilon}_{n-p}||^2$, we set $k_1=p$, $k_2=n-p$, $\lambda_1=\lambda= ||X\beta||^2/\sigma^2$, $\lambda_2=0$, and find

$$\boxed{\mu_1' = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \frac{[2r+p]}{[n-p-2]}}$$

$$= \frac{1}{n-p-2} e^{-\frac{1}{2}\lambda} [2f_1(\tfrac{1}{2}\lambda) + p\, f_o(\tfrac{1}{2}\lambda)] \quad \text{(cf. (A31)}$$

$$= \frac{1}{n-p-2} e^{-\frac{1}{2}\lambda} [2(\tfrac{1}{2}\lambda)e^{\frac{1}{2}\lambda} + p\, e^{\frac{1}{2}\lambda}]$$

whence

$$\boxed{\mu_1' = \frac{\lambda+p}{n-p-2}} \quad (\equiv\mu_Q) \tag{A55}$$

This exists when n-p>2.

B. Second Moment of Q

In (A54) we make the same substitutions leading to (A55), and find

$$\mu_2' = \frac{e^{-\frac{1}{2}\lambda}}{(n-p-2)(n-p-4)} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} (2r+p+2)(2r+p)$$

since

$$(2r+p+2)(2r+p) = 4r^2 + 4r(p+1) + (p^2+2p)$$

we can write

$$\mu_2' = \frac{e^{-\frac{1}{2}\lambda}}{(n-p-2)(n-p-4)} [4f_2(\tfrac{1}{2}\lambda) + 4(p+1)f_1(\tfrac{1}{2}\lambda) + (p^2+2p)f_o(\tfrac{1}{2}\lambda)]$$

using the functions $f_n(x)$ defined in (A31). This may be reduced to

$$\boxed{\mu_2' = \frac{\lambda^2 + (p+2)(2\lambda+p)}{[n-p-2]\,[n-p-4]}} \qquad \text{(A56)}$$

This exists when n-p>4.

C. Variance of Q

In general the variance is given by

$$\mu_2 = \mu_2' - \mu_1^2 \qquad \text{(A57)}$$

Using (A55), (A56) in this we have, on reduction,

$$\mu_2 = \frac{2[\lambda^2 + (n-2)(2\lambda+p)]}{[n-p-2]^2[n-p-4]} \qquad (\equiv\sigma_Q^2) \qquad \text{(A58)}$$

This exists when n-p>4.

D. First and Second Moments of I

From (A53) for $I = ||\underline{\varepsilon}_{n-p}||^2/||\underline{X\hat{\beta}}||^2$, we set

$$k_1 = n-p,\ k_2 = p,\ \lambda_1 = 0,\ \lambda_2 = \lambda = ||\underline{X\beta}||^2/\sigma^2$$

and find

$$(\mu_I =)\ \mu_1' = (n-p)e^{-\frac{1}{2}\lambda}\sum_{s=o}^{\infty}\frac{(\frac{1}{2}\lambda)^s}{s!}\cdot\frac{1}{2s+p-2} \qquad \text{(A59)}$$

and from (A54):

$$\mu_2' = (n-p)[n-p+2]e^{-\frac{1}{2}\lambda}\sum_{s=o}^{\infty}\frac{(\frac{1}{2}\lambda)^s}{s!}\cdot\frac{1}{[2s+p-2][2s+p-4]} \qquad \text{(A60)}$$

μ_2 is best found numerically in this case, using (A57), (A59), (A60). The moments μ_1', μ_2' exist if p>2, p>4, respectively.

The mth raw moment of $y \sim J(k_1, k_2, \lambda_1, \lambda_2)$ is found from (A50) via

$$\mu_m' = \int_o^1 y^m\, J(y|k_1,k_2,\lambda_1,\lambda_2)\, dy$$

This requires the evaluation of

$$\int_o^1 y^{r+m+\frac{1}{2}k_1-1}(1-y)^{s+\frac{1}{2}k_2-1}\, dy = \frac{\Gamma(r+m+\frac{1}{2}k_1)\Gamma(s+\frac{1}{2}k_2)}{\Gamma(r+s+m+\frac{1}{2}(k_1+k_2))}$$

using the beta function. Hence in general

$$\mu_m' = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{\Gamma(r+m+\frac{1}{2}k_1)}{\Gamma(r+\frac{1}{2}k_1)} \frac{\Gamma(r+s+\frac{1}{2}(k_1+k_2))}{\Gamma(r+s+m+\frac{1}{2}(k_1+k_2))}, \quad m=0,1,2\ldots \qquad \text{(A61)}$$

As special cases of this

$$\mu_1' = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s=o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{(2r+k_1)}{(2r+2s+k_1+k_2)} \qquad \text{(A62)}$$

$$\mu_2' = e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{r=o}^{\infty} \sum_{s-o}^{\infty} \frac{(\frac{1}{2}\lambda_1)^r}{r!} \frac{(\frac{1}{2}\lambda_2)^s}{s!} \cdot \frac{(2r+2+k_1)(2r+k_1)}{(2r+2s+2+k_1+k_2)(2r+2s+k_1+k_2)} \qquad \text{(A63)}$$

E. First Moment of S

In (A62) for $S = ||\underline{X\hat{\beta}}||^2/||\underline{y}||^2$, we set

$$k_1 = p,\ k_2 = n-p,\ \lambda_1 = \lambda = ||\underline{X\hat{\beta}}||^2/\sigma^2,\ \lambda_2 = 0$$

and find

$$\mu_1' = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \left[\frac{2r+p}{2r+n}\right] \qquad (= \mu_S) \qquad \text{(A64)}$$

F. Second Moment of S

In (A63) we make the same substitutions leading to (A64), and find

$$\mu_2' = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{(2r+2+p)(2r+p)}{(2r+2+n)(2r+n)} \tag{A65}$$

The variance of S is best found numerically from (A57), (A64), (A65) for any given set of n, p, λ values. Some approximations may be possible, as we show below.

G. First Moment of S for Small Signal to Noise Ratio λ

Expanding the exponential series in (A64) and retaining only first powers of λ, we find

$$(\mu_s =)\ (\overline{S} =)\ \mu_1' \cong (1-\frac{\lambda}{2})\ \frac{p}{n} + \frac{\lambda}{2}\ (\frac{p+2}{n+2}) \tag{A64a}$$

As $\lambda \to 0$, $\mu_1' \to p/n$, as may also be seen from (A64). If we write 'S_o' for p/n, and n is large compared to 2, then (A64a) becomes

$$(\mu_s =)\ (\overline{S} =)\ \mu_1' \cong S_o + \tfrac{1}{2}\lambda(1-S_o) \tag{A64b}$$

This reduces to the exact classic skill's mean for the case of zero signal to noise:

$$\mu_s = S_o = p/n\ . \tag{A64c}$$

H. The Second Moment of S for Small Signal to Noise Ratio λ

From (A65), expanding the exponential, and retaining only the first power of λ,

$$\mu_2' \cong (\frac{p+2}{n+2})\ [\frac{p}{n} + \lambda\ \frac{2(n-p)}{n(n+4)}] \tag{A65a}$$

I. Variance of S for Small Signal to Noise Ratio λ

Since

$$\sigma_S^2 = \mu_2' - \mu_1'^2 \ ,$$

We have from (A64a) and (A65a), on retaining only the first power of λ,

$$\sigma_S^2 \cong \frac{2(n-p)}{n(n+2)} \left[\frac{p}{n} - 2\lambda \ \frac{(n+np+p)}{n(n+4)}\right]$$

$$= \frac{2(1-S_o)}{n+2} \left[S_o - 2\lambda \ \frac{(1+p+S_o)}{n+4}\right] \qquad \text{(A65b)}$$

This reduces to exactly to the classic skill's variance for the case of zero signal to noise:

$$\sigma_S^2 = \frac{2(1-S_o)S_o}{n+2} = \frac{2(1-p/n)(p/n)}{n+2} \qquad \text{(A65c)}$$

10. Classic Hindcast Skill and the Multiple Correlation Coefficient

There is a general intuitive connection between the ideas of linear regression and multiple correlation that would lead one to suspect a correspondingly general formal connection between all the salient parameters in each of these two domains. In this section we shall show the exact formal correspondence between classic skill S and the square R^2 of the multiple correlation coefficient, and also the explicit connection between the signal to noise ratio λ and the population correlation coefficient $\underline{R}^2$.

T prepare for the demonstration we rewrite (A51) as follows:

$$P_S(x|n,p,\lambda) =$$

$$= \frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2}(n-p))\Gamma(\frac{1}{2}(p-1))} x^{\frac{1}{2}p-1} (1-x)^{\frac{1}{2}(n-p)-1} \cdot e^{-\frac{1}{2}\lambda} \cdot$$

$$\cdot \sum_{r=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+2r))\ \Gamma(\frac{1}{2}(p-1))}{\Gamma(\frac{1}{2}(p+2r))\ \Gamma(\frac{1}{2}(n-1))} \cdot \frac{(\frac{1}{2}\lambda x)^r}{r!} \tag{A66}$$

In (Kendall and Stuart, vol 2, 1972, p358), given as an exercise, is the following form for the pdf of the square of the multiple correlation coefficient (using their notation and taking the liberty to make some rearrangements and to open up their beta function, so as to facilitate the comparison):

$$dF =$$

$$= \frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2}(n-p)\Gamma(\frac{1}{2}(p-1))} (R^2)^{\frac{1}{2}(p-3)} (1-R^2)^{\frac{1}{2}(n-p-2)} e^{-\frac{1}{2}(n-p)\underline{R}^2} \cdot$$

$$\cdot \sum_{r=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n-1+2r))}{\Gamma(p-1+2r)} \frac{\Gamma(\frac{1}{2}(p-1))}{\Gamma(\frac{1}{2}(n-1))} \cdot \frac{(\frac{1}{2}(n-p)\underline{R}^2R^2)^r}{r!} dR^2 \tag{A67}$$

Ovserve that certain terms can be cancelled in (A66), such as $\Gamma(\frac{1}{2}(n-1))$ and $\Gamma(\frac{1}{2}(p-1))$. These may also be cancelled in (A67). They were put in by Kendall and Stuart ('K and S') to 'pretty up' the results, and we followed suit. When a comparison between (A66), (A67) is made in their simplified forms, the following correspondences are evident:

Multiple Correlation and Classic Skill

K and S	Here
$n-1$	n
$p-1$	p
R^2	x
$(n-p)\underline{R}^2$	λ

In this way we discover the connection between our signal to noise ratio λ and the population correlation coefficient $\underline{R}^2$:

$$||X\underline{\hat{\beta}}||^2/\sigma^2 = \lambda = (n-p)\underline{R}^2 \ . \tag{A68}$$

There is an important proviso regarding (A68), namely that $\underline{R}^2$ by construction is always bound by $0 \leq R^2 \leq 1$, whereas λ clearly can exceed 1, as a perusal of the linear regression diagram in §6 of the main text shows. We can fix $||X\underline{\beta}||$ and imagine the vector $\underline{\varepsilon}$ to have any σ^2, large or small. In terms of our dynamical studies in §2 (particularly recall (2.5)), the signal $||X\underline{\beta}||^2$ of the retained drivers and the noise σ^2 (of the discarded drivers) may be independently chosen. It is particularly this fact and to a somewhat lesser extent the specialized cast of multiple correlation theory in the domain of statistics that suggested retaining our independent development of the theory of λ. Still another correspondence can be set up using a result in (Rao, 1973, p600).

Appendix B, Finite-term Formulas for Cumulative Probabilities

The numerical determination of the $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ significance levels for a given performance skill Q, S, or I, (as in §11) is facilitated by the formulas presented below. The formulas are based on the fact that when n-p or p (as the case may be) are even integers, the indefinite integrals of the densities P_Q, P_S, P_I presented in §8 can be expressed as the results of a finite number of elementary operations. A computer program based on these finite-term formulas is much faster than one that integrates the densities using, say, Simpson's rule. The tables below are based on these finite-term integrals. It is found that tabulations of $\sigma(05)$, $\sigma(95)$ for n up to about 50 can be handled this way before numerical problems of accuracy arise. Beyond n = 50, the determination of $\sigma(\frac{1}{2}\alpha)$, $\sigma(1-\frac{1}{2}\alpha)$ for $\alpha = .10$ (say) must be done with Simpson's rule and double precision, or some other integration procedure with controllable accuracy, such as Runge-Kutta schemes.

1. Formulas for Canonic Skill Q

Starting with (8.1) we integrate $P_Q(x|n,p,\lambda)$ from x = 0 to some arbitrary value y. This requires the evaluation of the x-dependent part of P_Q in the form:

$$H_r(y|p,\lambda) = \int_0^y \frac{x^{r+\frac{1}{2}p-1}}{(1+x)^{r+\frac{1}{2}n}}\,dx$$

Make the substitution of variables: $1 + x = u^2$, then $dx = 2u\,du$, and so $x^{r+\frac{1}{2}p-1} = (u^2-1)^{r+\frac{1}{2}p-1}$. When x = 0, u = 1. So

$$H_r(y|p,n) = 2\int_1^y \frac{(u^2-1)^{\frac{1}{2}p+r-1}}{u^{n+2r-1}}\,du ,$$

where we leave the upper limit u arbitrary, say of value y.

Let p/2 be an integer. Then with

$$\binom{n}{j} \equiv {}^{n}C_{j} = \frac{n.(n-1)\ldots(n-j+1)}{1.2\ldots j},$$

$$(u^2-1)^{\frac{1}{2}p+r-1} = \sum_{j=o}^{\frac{1}{2}p+r-1} \binom{\frac{1}{2}p+r-1}{j} (u^2)^j (-1)^{\frac{1}{2}p+r-1-j}$$

So

$$H_r(y|p,n) = \sum_{j=o}^{\frac{1}{2}p+r-1} \binom{\frac{1}{2}p+r-1}{j} (-1)^{\frac{1}{2}p+r-j-1} \left[\frac{(1+y)^{j+1-(\frac{1}{2}n+r)}-1}{j+1-(\frac{1}{2}n+r)}\right] \quad \text{(B1)}$$

Therefore

$$\int_o^y P_Q(x|n,p,\lambda)dx = e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot H_r(y|p,n) \quad \text{(B2)}$$

$\frac{1}{2}p$ an integer, $n-p>1$

$$\overline{Q} = \frac{\lambda+p}{n-p-2}, \quad n-p>2 \quad \text{(B3)}$$

In applications of (B2), one should keep in mind the important option of using the representation $\lambda = p\overline{\lambda}$ for the signal to noise ratio (cf §9). In our preliminary study of (B2), summarized in the tables below, p, n, λ were treated as independent variables. In practical applications, it is suggested that the representation $\lambda = p\overline{\lambda}$ be used since, as explained in §9, $\overline{\lambda}$ is then more or less independent of p, and so n, p, $\overline{\lambda}$ are independent parameters. These comments of course hold for the formulas below.

2. Formulas for Ineptness I

Starting with (8.4), we integrate $P_I(x|n,p,\lambda)$ from x = 0 to some arbitrary value y. This requires the evaluation of the x-dependent part of P_I in the form:

$$J_r(y|p,n) = \int_o^y \frac{x^{\frac{1}{2}(n-p)-1}}{(1+x)^{r+\frac{1}{2}n}}\,dx$$

$$= \sum_{j=o}^{\frac{1}{2}(n-p)-1} \binom{\frac{1}{2}(n-p)-1}{j} (-1)^{\frac{1}{2}(n-p)-j-1} \left[\frac{(1+y)^{j+1-(\frac{1}{2}p+r)}-1}{j+1-(\frac{1}{2}n+r)}\right] \qquad (B4)$$

It is seen that this differs from $H_r(y|p,\lambda)$ only by the interchange of (n-p), p, and the absences of certain r presences in (B4). To evaluate (B4) we used the assumption that n-p is an integer. Hence

$$\int_o^y P_I(x|n,p,\lambda)dx = e^{-\frac{1}{2}} \sum_{r=o}^{\infty} \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot J_r(y|p,n) \qquad (B5)$$

$\frac{1}{2}(n-p)$ a positive integer,

$$\overline{I} = (n-p)e^{-\frac{1}{2}\lambda} \sum_{r=o}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \frac{1}{(p+2r-2)} \quad , \; p > 2 \qquad (B6)$$

3. Formulas for Classic Skill S

Starting with (8.7), we integrate $P_S\,(x|n,p,\lambda)$ from $x = 0$ to some arbitrary value y. This requires the evaluation of the x-dependent part of P_S in the form (with the assumption that n-p is even):

$$L_r(y|p,n) = \int_0^y x^{r+\frac{1}{2}p-1}\,(1-x)^{\frac{1}{2}(n-p)-1}\,dx$$

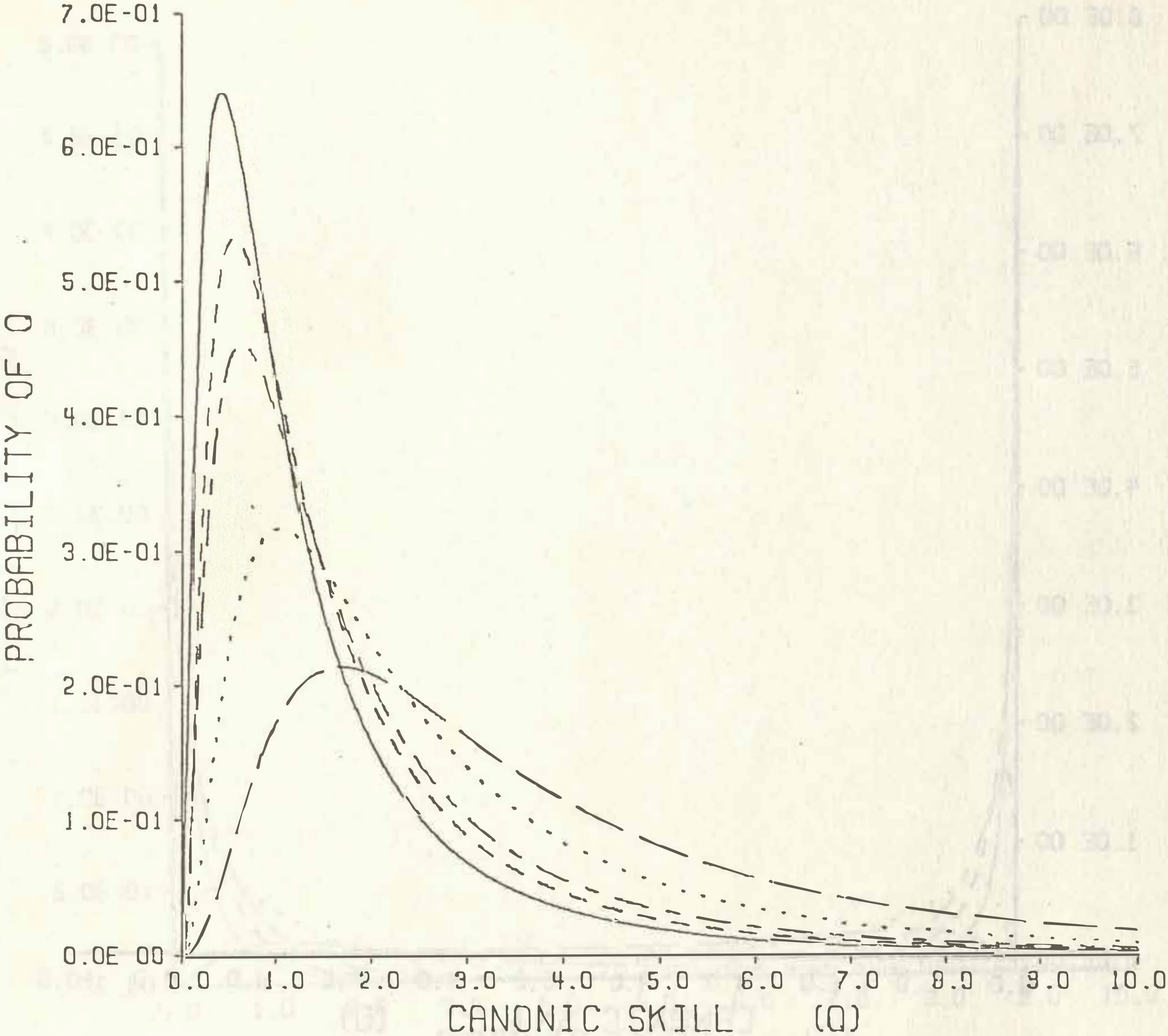
$$= (-1)^{n-p-2} \sum_{j=0}^{\frac{1}{2}(n-p)-1} \binom{\frac{1}{2}(n-p)-1}{j} (-1)^j \frac{x^{\frac{1}{2}p+r+j}}{\frac{1}{2}p+r+j} \qquad \text{(B7)}$$

Hence

$$\int_0^y P_S(x|n,p,\lambda)dx = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{\Gamma(r+\frac{1}{2}n)}{\Gamma(r+\frac{1}{2}p)\Gamma(\frac{1}{2}(n-p))} \cdot \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot L_r(y|n,p) \qquad \text{(B8)}$$

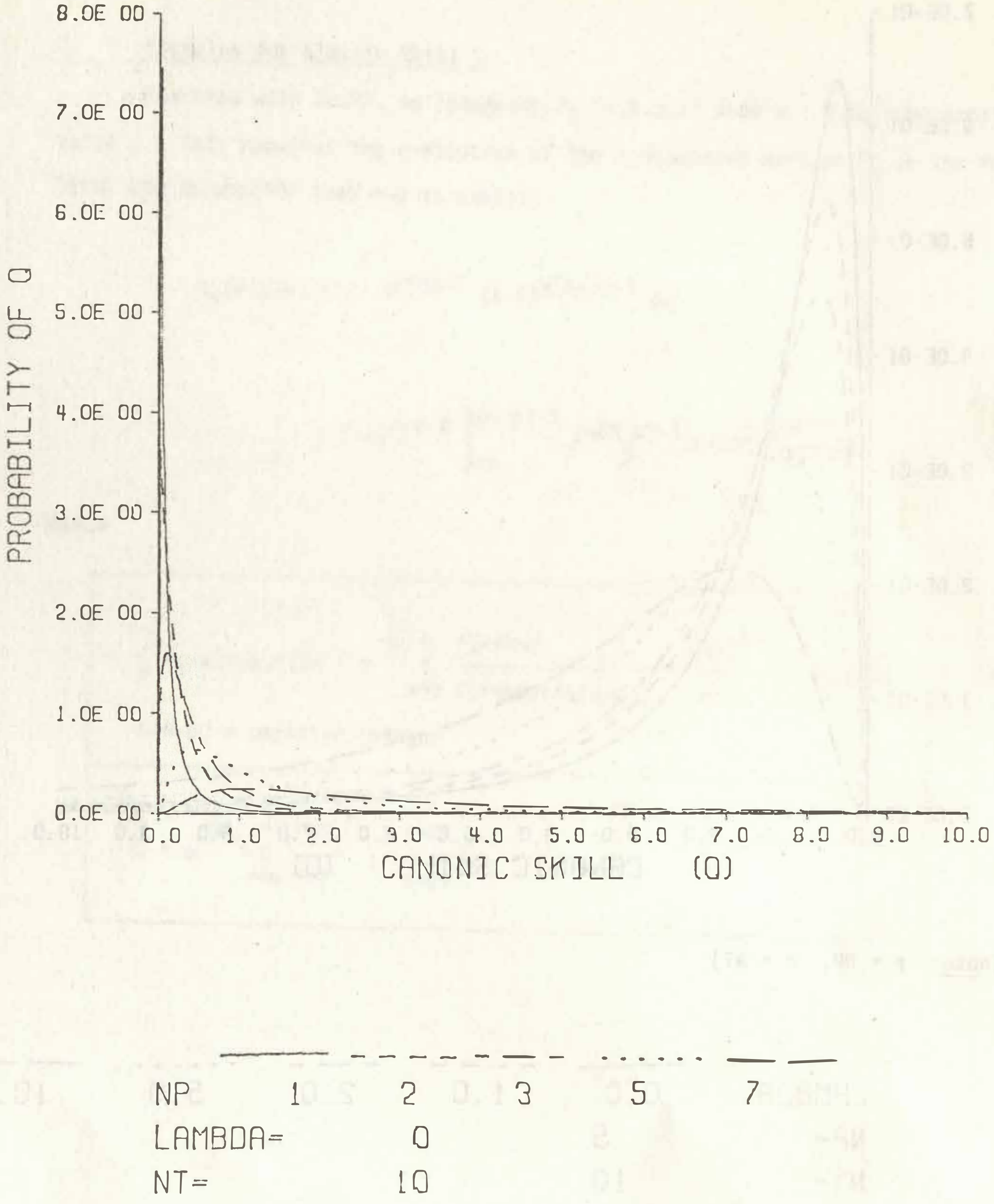
$\frac{1}{2}(n-p)$ a positive integer

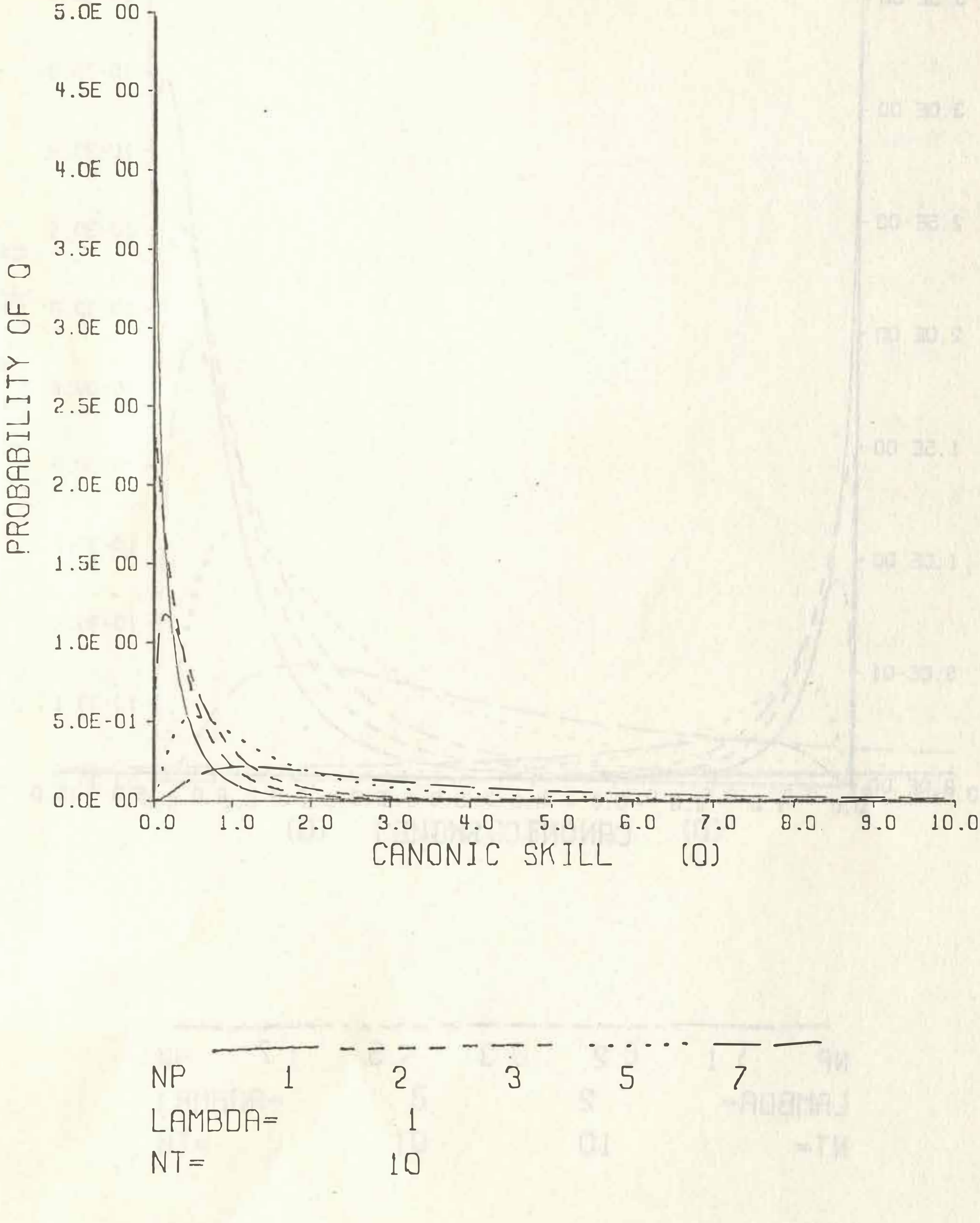
$$\overline{S} = e^{-\frac{1}{2}\lambda} \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \cdot \left[\frac{p+2r}{n+2r}\right] \qquad \text{(B9)}$$



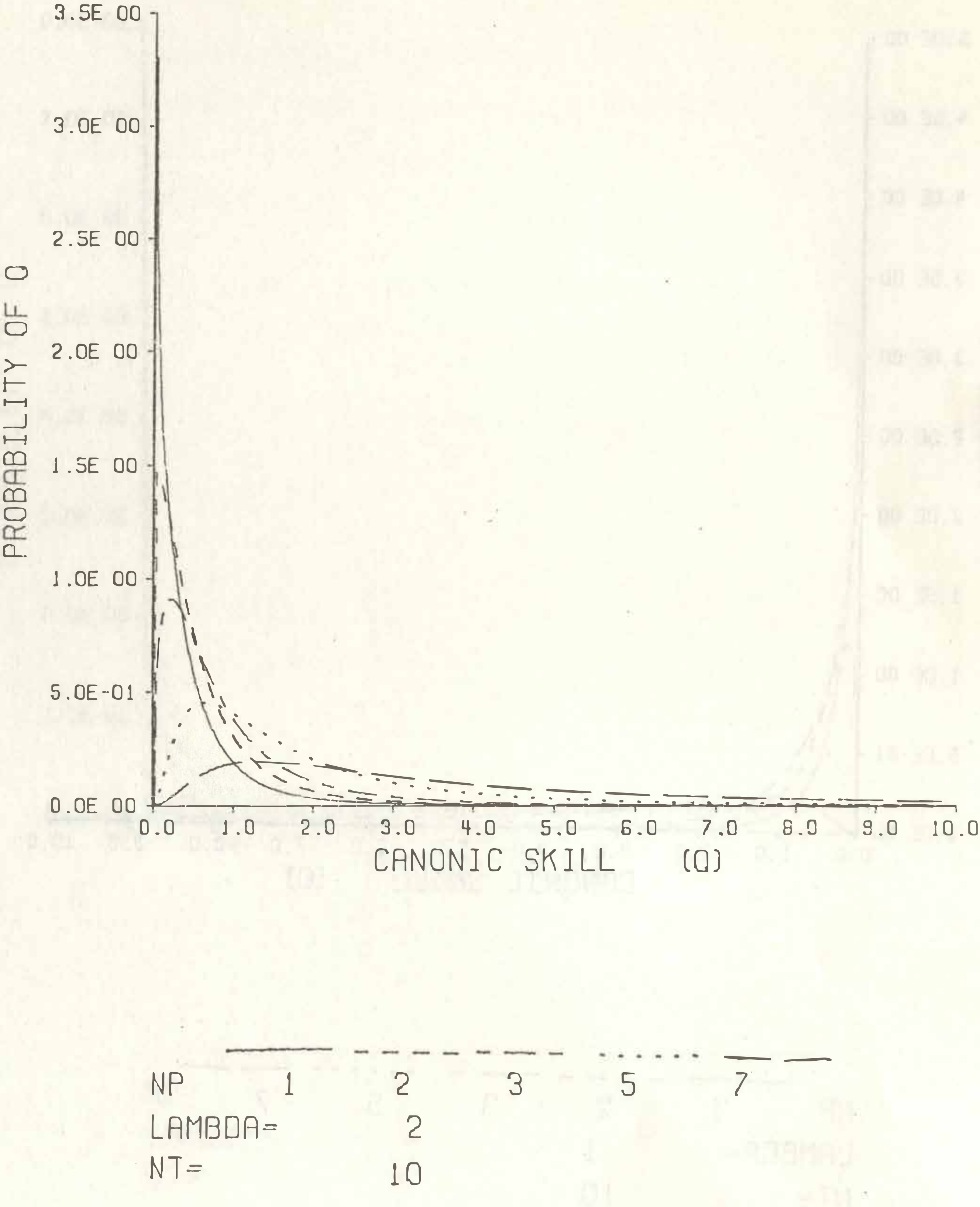
(Note: p = NP, n = NT)

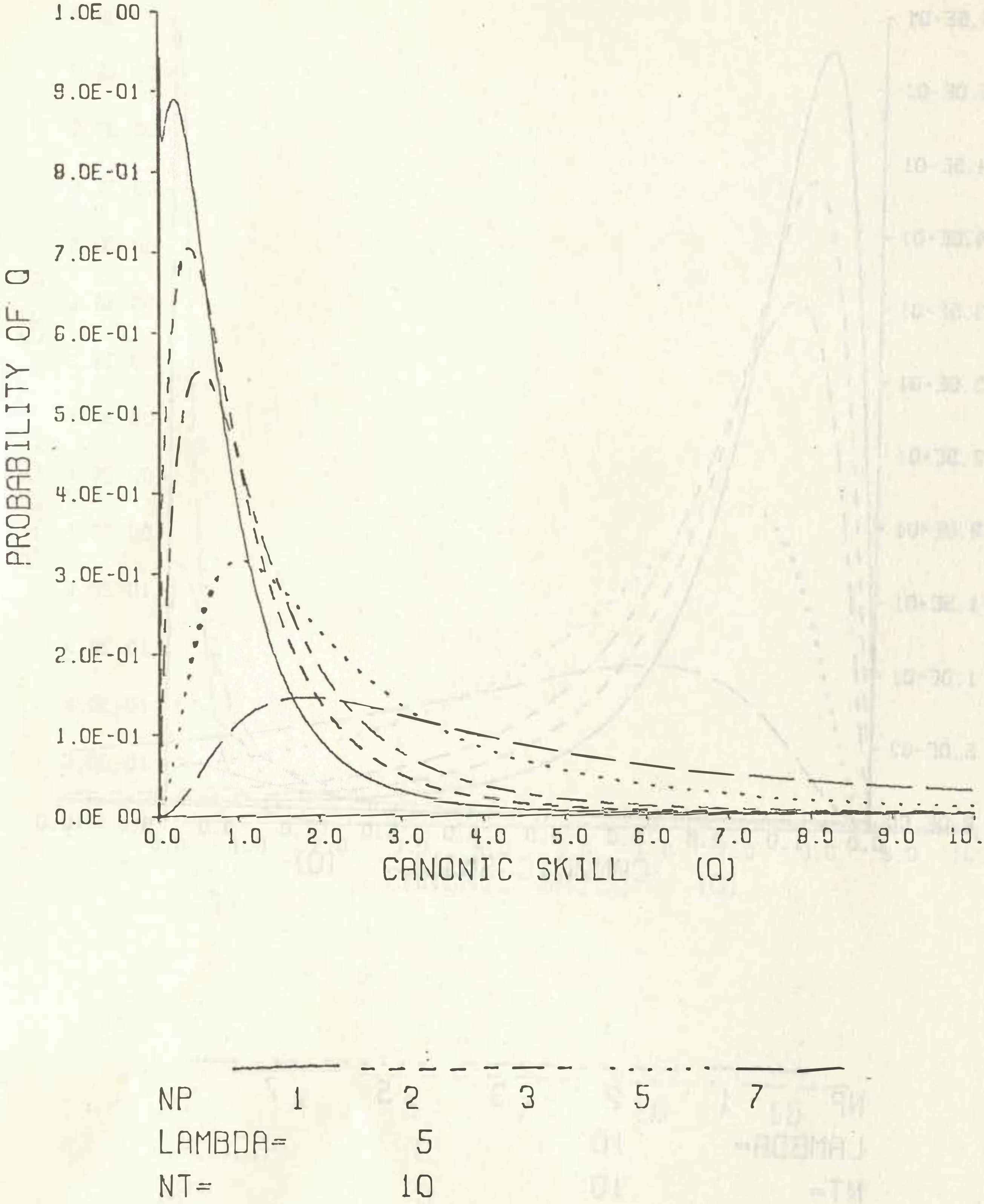
LAMBDA	0.0	1.0	2.0	5.0	10.0
NP=	5				
NT=	10				

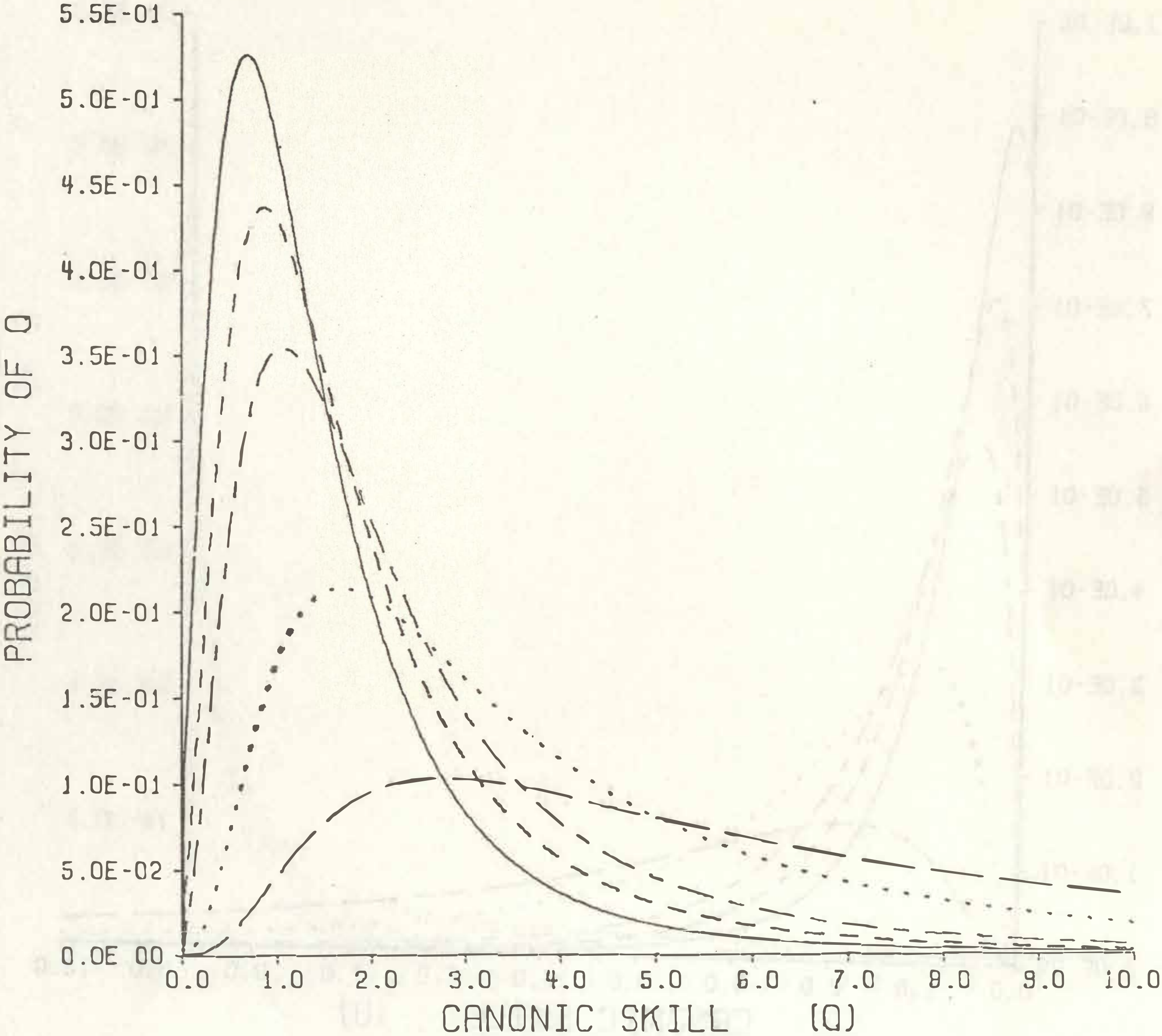

8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF Q
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
CANONIC SKILL (Q)
NP 1 2 3 5 7
LAMBDA= 0
NT= 10



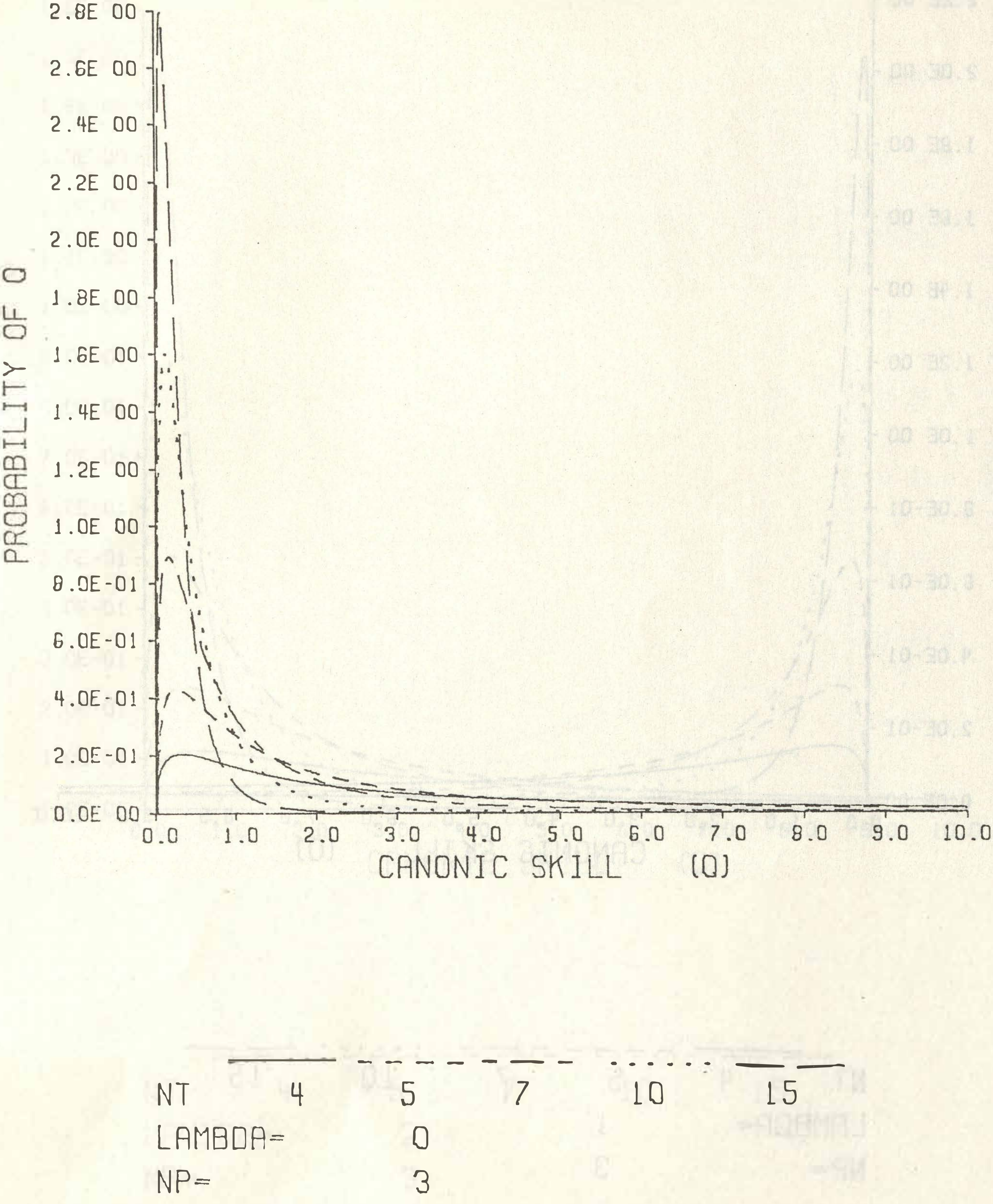
5.0E 00
4.5E 00
4.0E 00
3.5E 00
3.0E 00
2.5E 00
2.0E 00
1.5E 00
1.0E 00
5.0E-01
0.0E 00
PROBABILITY OF Q
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
CANONIC SKILL (Q)
NP 1 2 3 5 7
LAMBDA= 1
NT= 10


3.5E 00
3.0E 00
2.5E 00
2.0E 00
1.5E 00
1.0E 00
5.0E-01
0.0E 00
PROBABILITY OF Q
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
CANONIC SKILL (Q)
NP 1 2 3 5 7
LAMBDA= 2
NT= 10

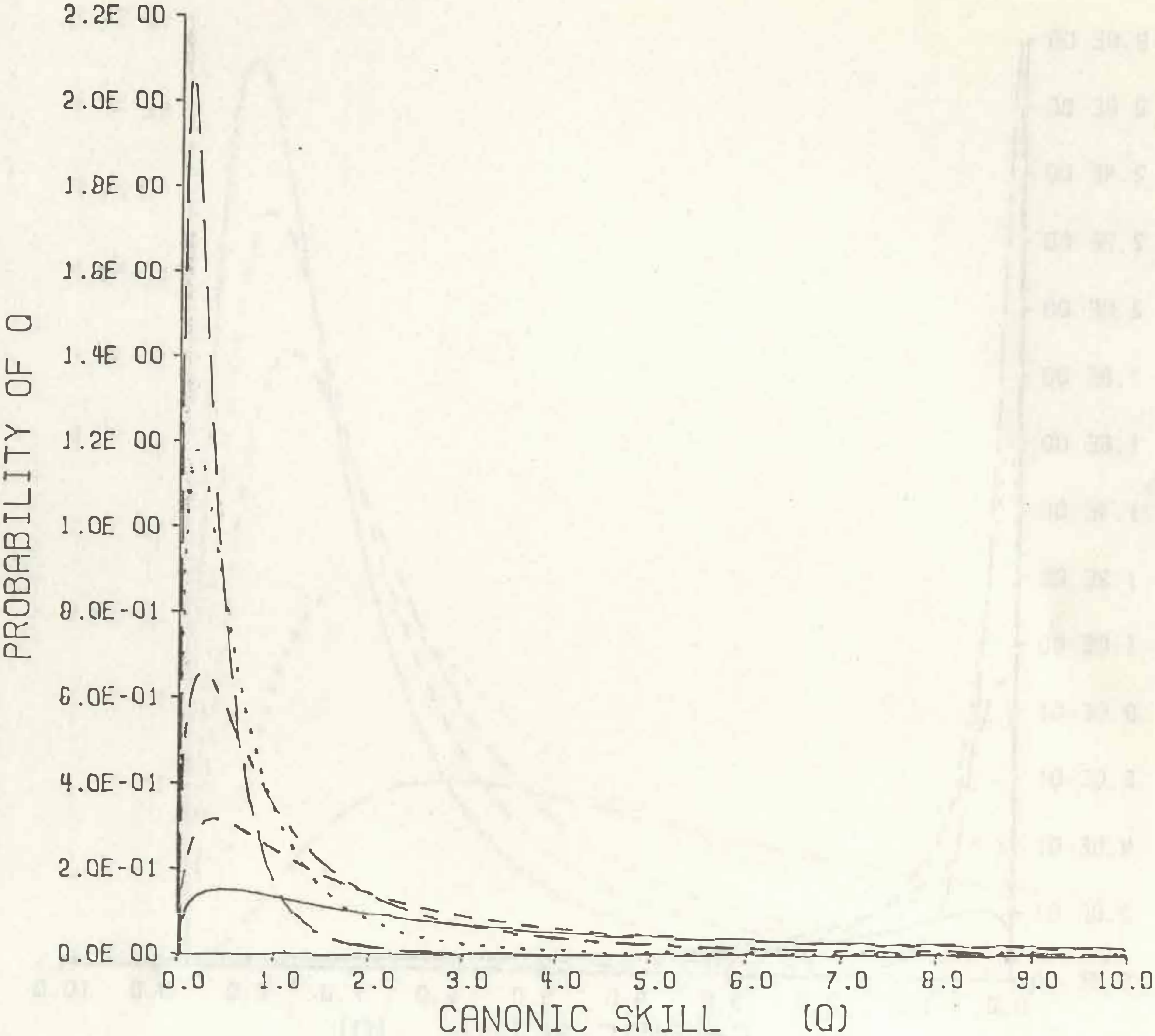

1.0E 00
9.0E-01
8.0E-01
7.0E-01
6.0E-01
5.0E-01
4.0E-01
3.0E-01
2.0E-01
1.0E-01
0.0E 00
PROBABILITY OF Q
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
CANONIC SKILL (Q)
NP 1 2 3 5 7
LAMBDA= 5
NT= 10



NP	1	2	3	5	7
LAMBDA=		10			
NT=		10			



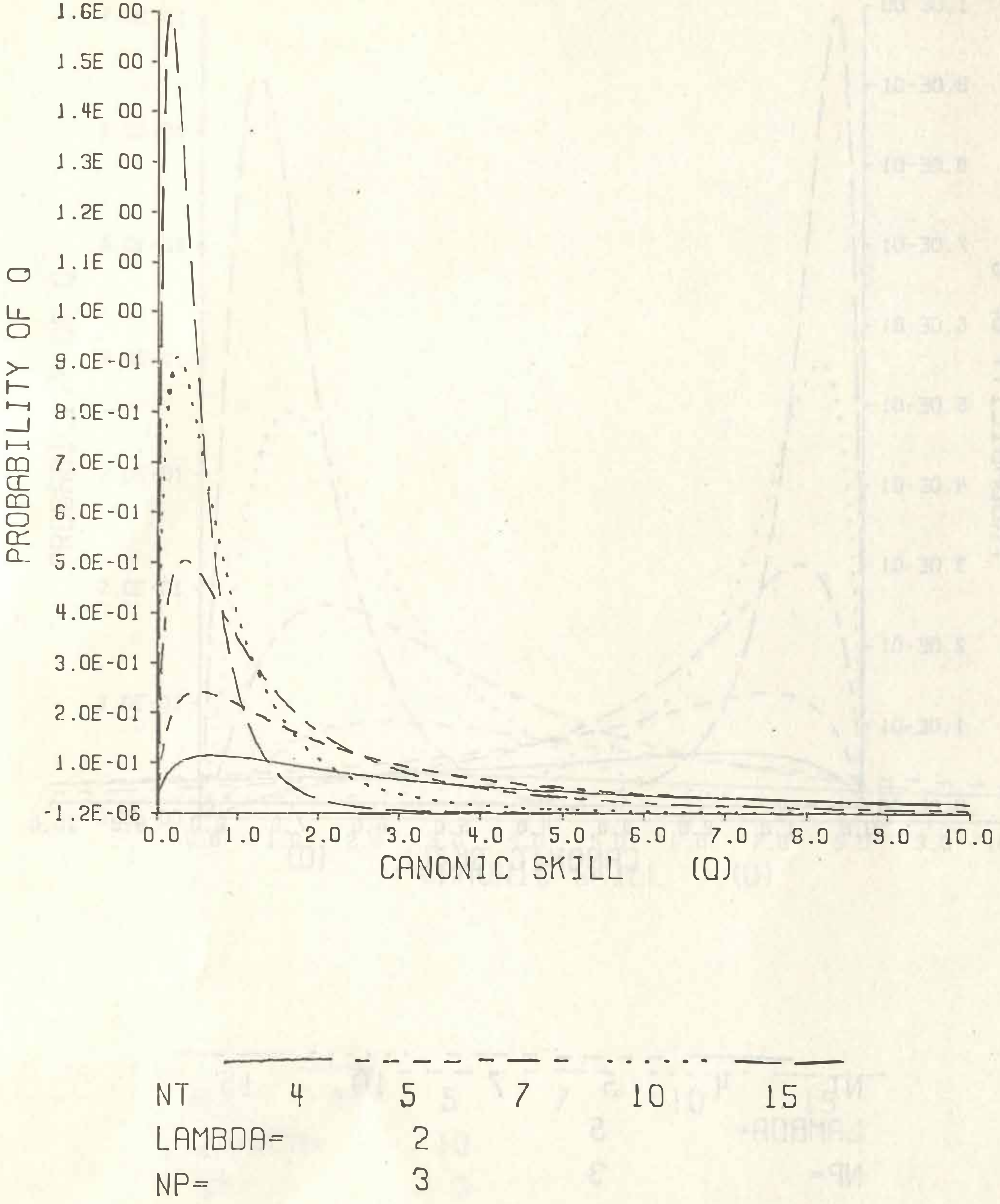
PROBABILITY OF Q
2.8E 00
2.6E 00
2.4E 00
2.2E 00
2.0E 00
1.8E 00
1.6E 00
1.4E 00
1.2E 00
1.0E 00
8.0E-01
6.0E-01
4.0E-01
2.0E-01
0.0E 00
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
CANONIC SKILL (Q)
NT 4 5 7 10 15
LAMBDA= 0
NP= 3



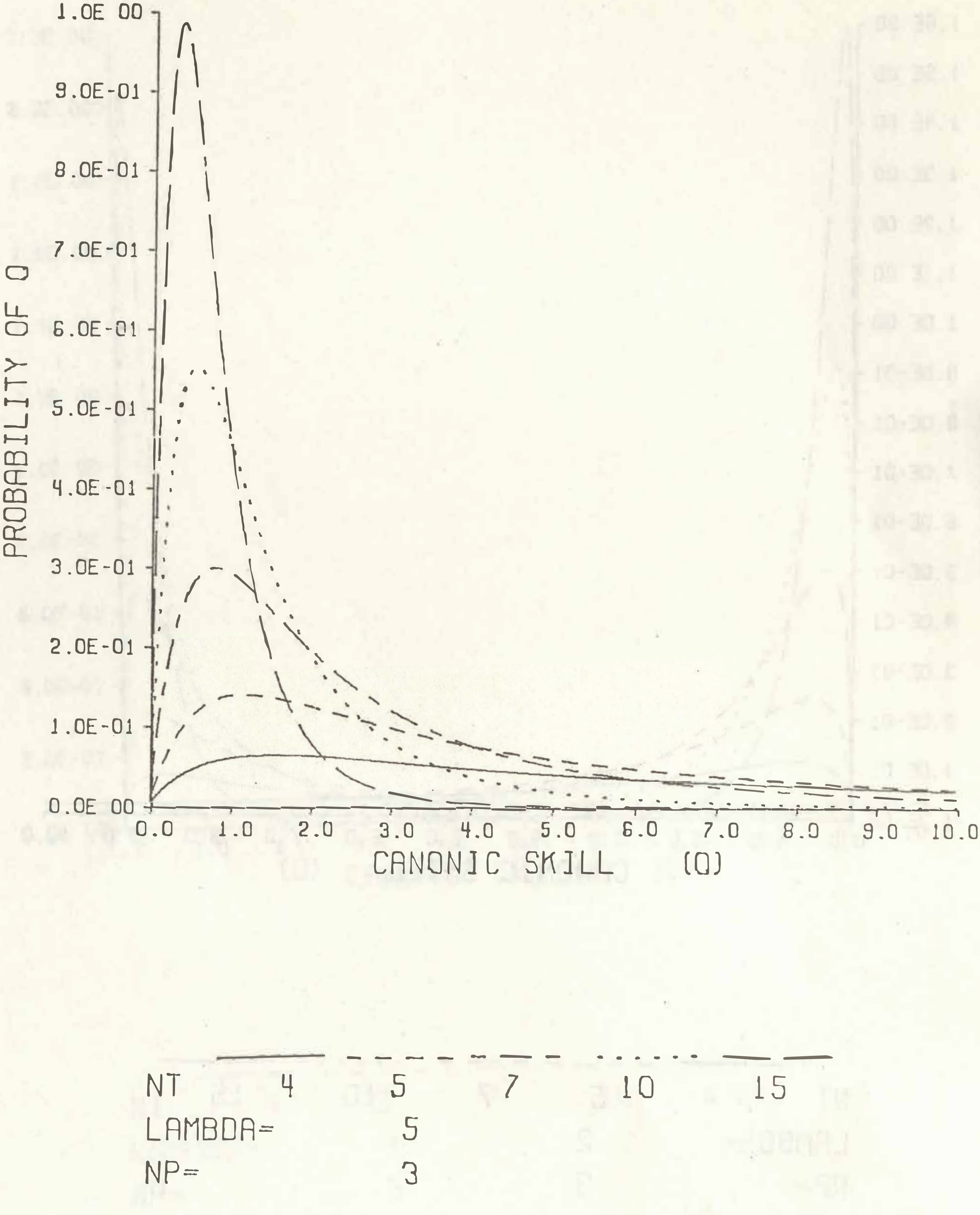
NT 4 5 7 10 15

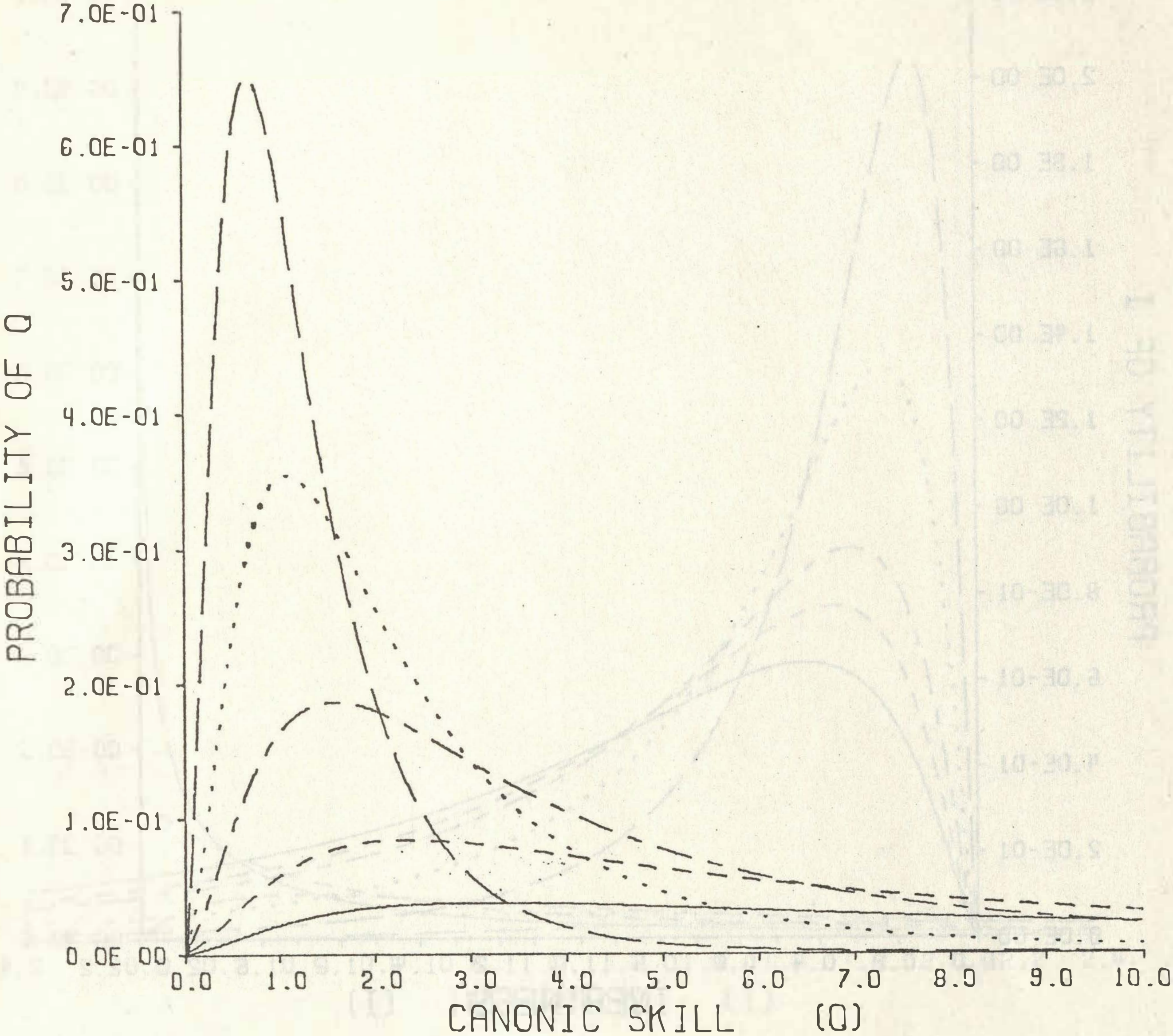
LAMBDA= 1

NP= 3

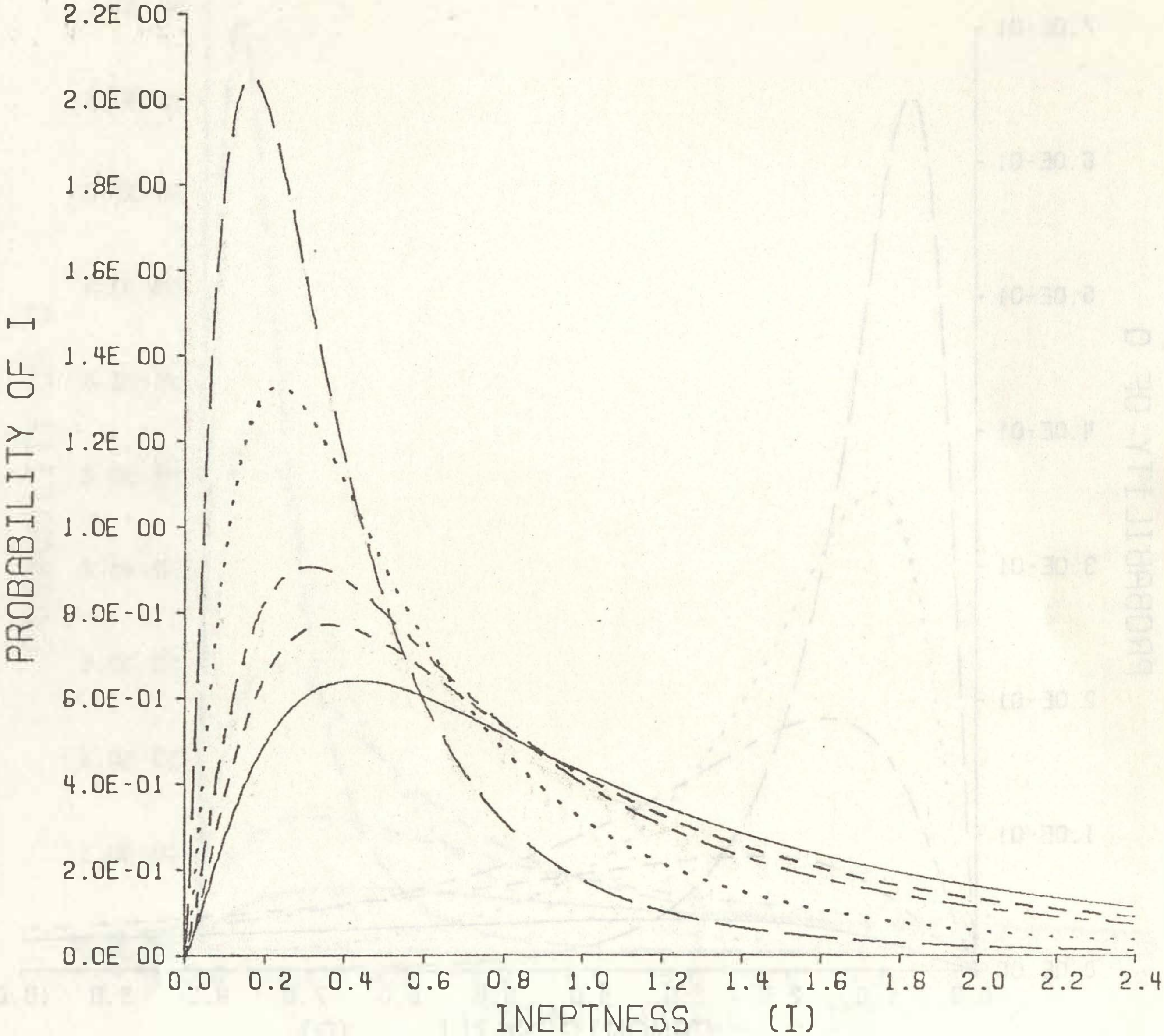


1.6E 00
1.5E 00
1.4E 00
1.3E 00
1.2E 00
1.1E 00
1.0E 00
9.0E-01
8.0E-01
7.0E-01
6.0E-01
5.0E-01
4.0E-01
3.0E-01
2.0E-01
1.0E-01
-1.2E-06
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
PROBABILITY OF Q
CANONIC SKILL (Q)
NT 4 5 7 10 15
LAMBDA= 2
NP= 3


1.0E 00
9.0E-01
8.0E-01
7.0E-01
6.0E-01
5.0E-01
4.0E-01
3.0E-01
2.0E-01
1.0E-01
0.0E 00
PROBABILITY OF Q
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0
CANONIC SKILL (Q)
NT 4 5 7 10 15
LAMBDA= 5
NP= 3

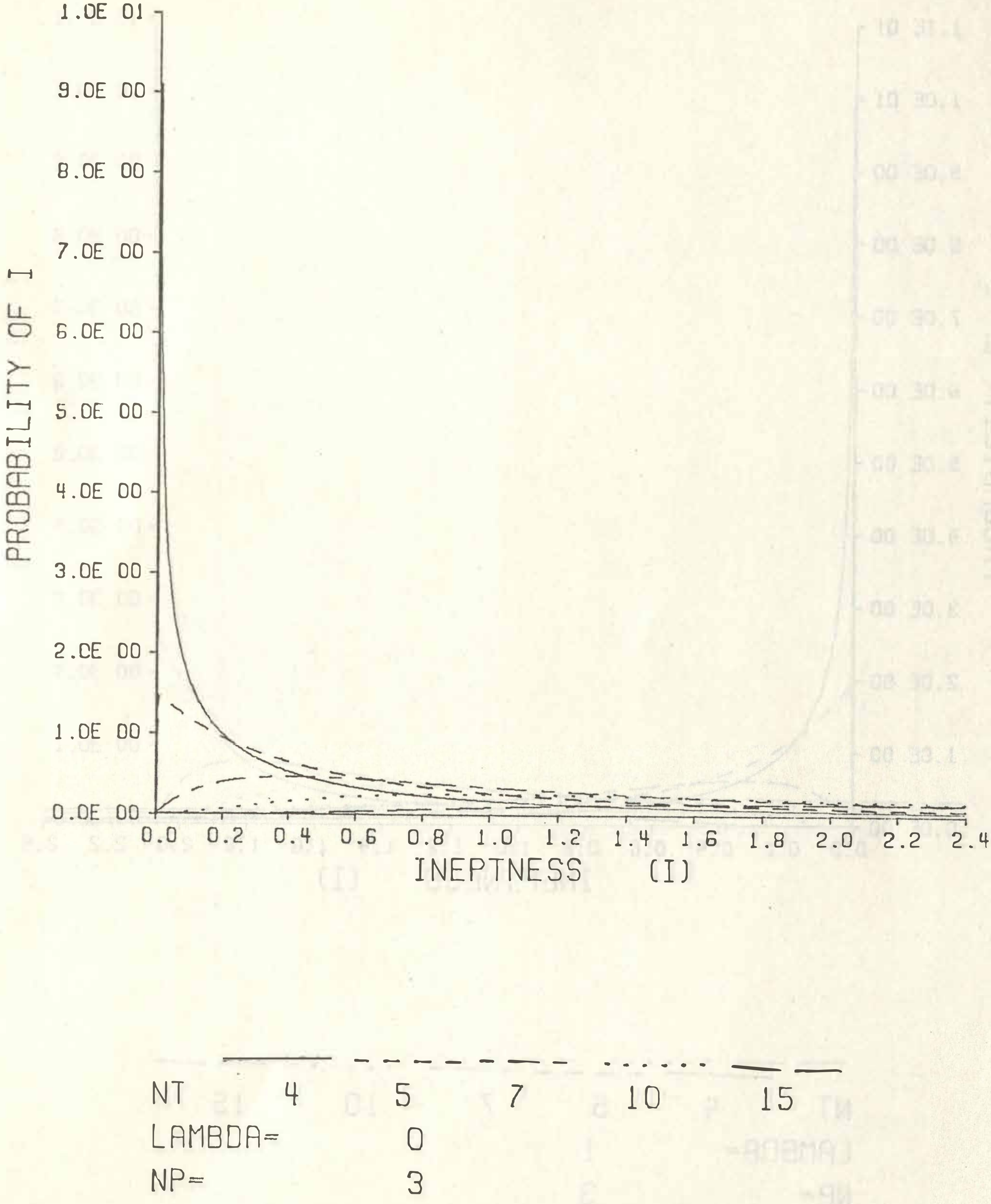


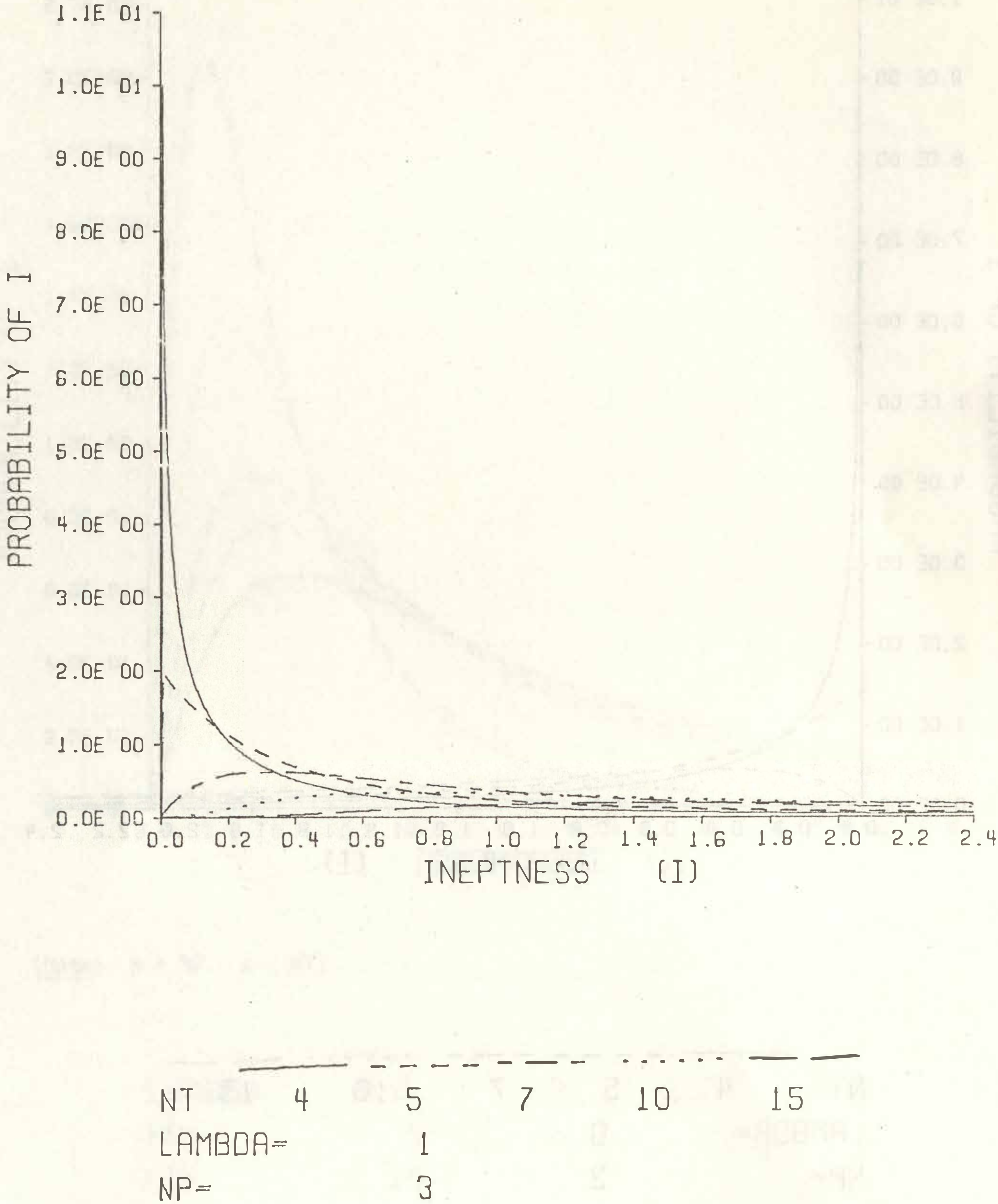
NT 4 5 7 10 15
LAMBDA= 10
NP= 3

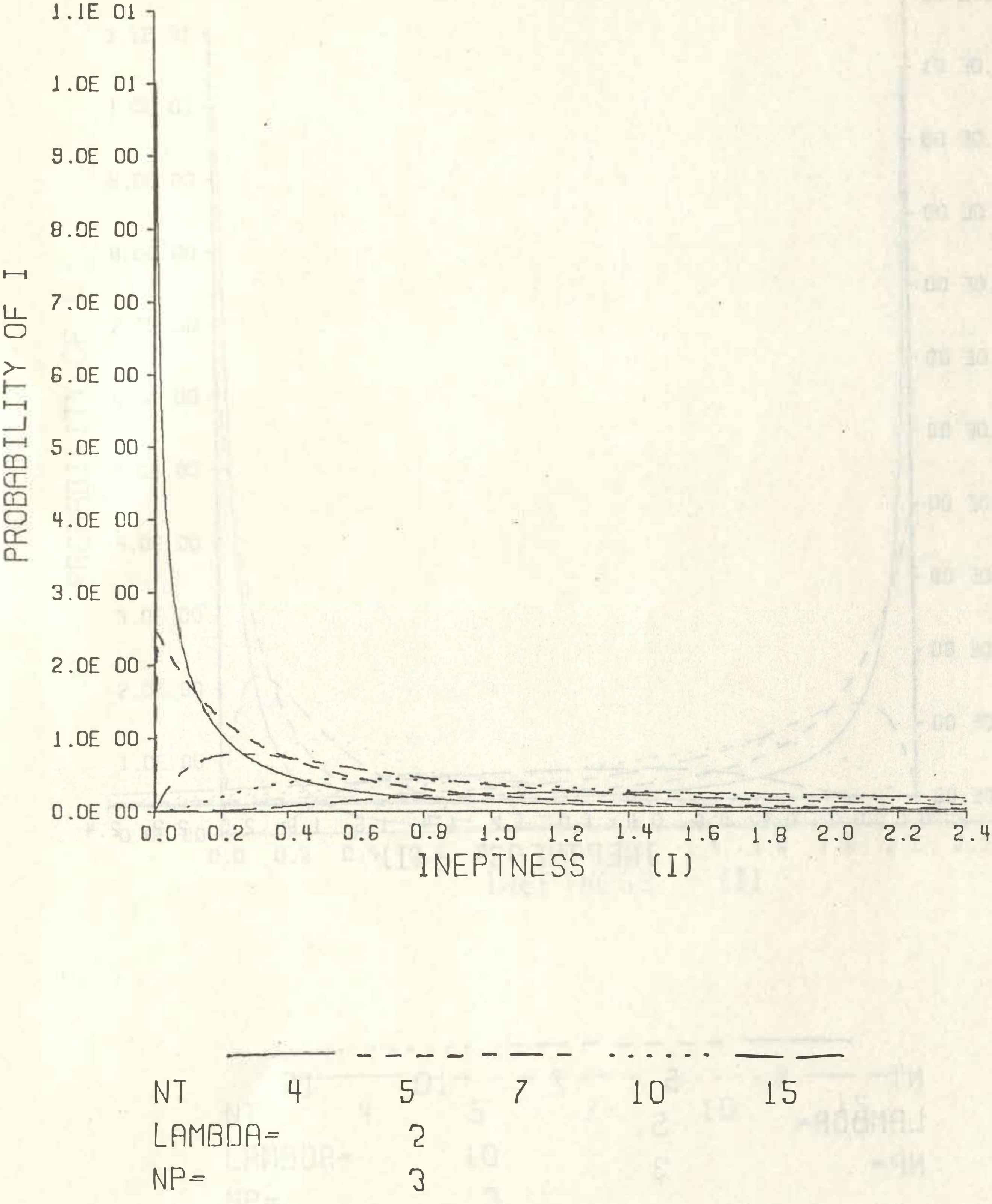


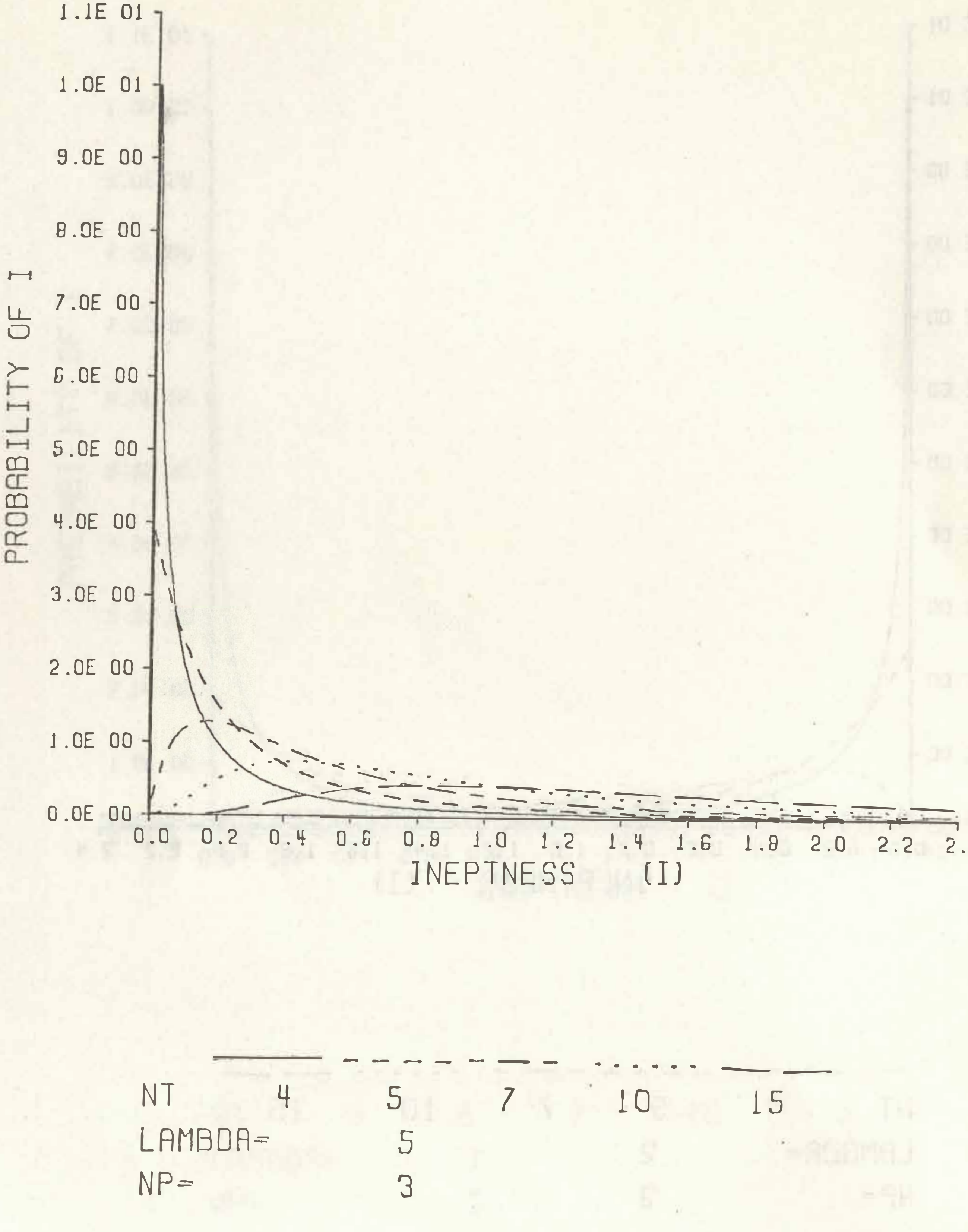
(<u>Note:</u> p = NP, n = NT)

LAMBDA	0 0	1.0	2.0	5.0	10.0
NP=	5				
NT=	10				

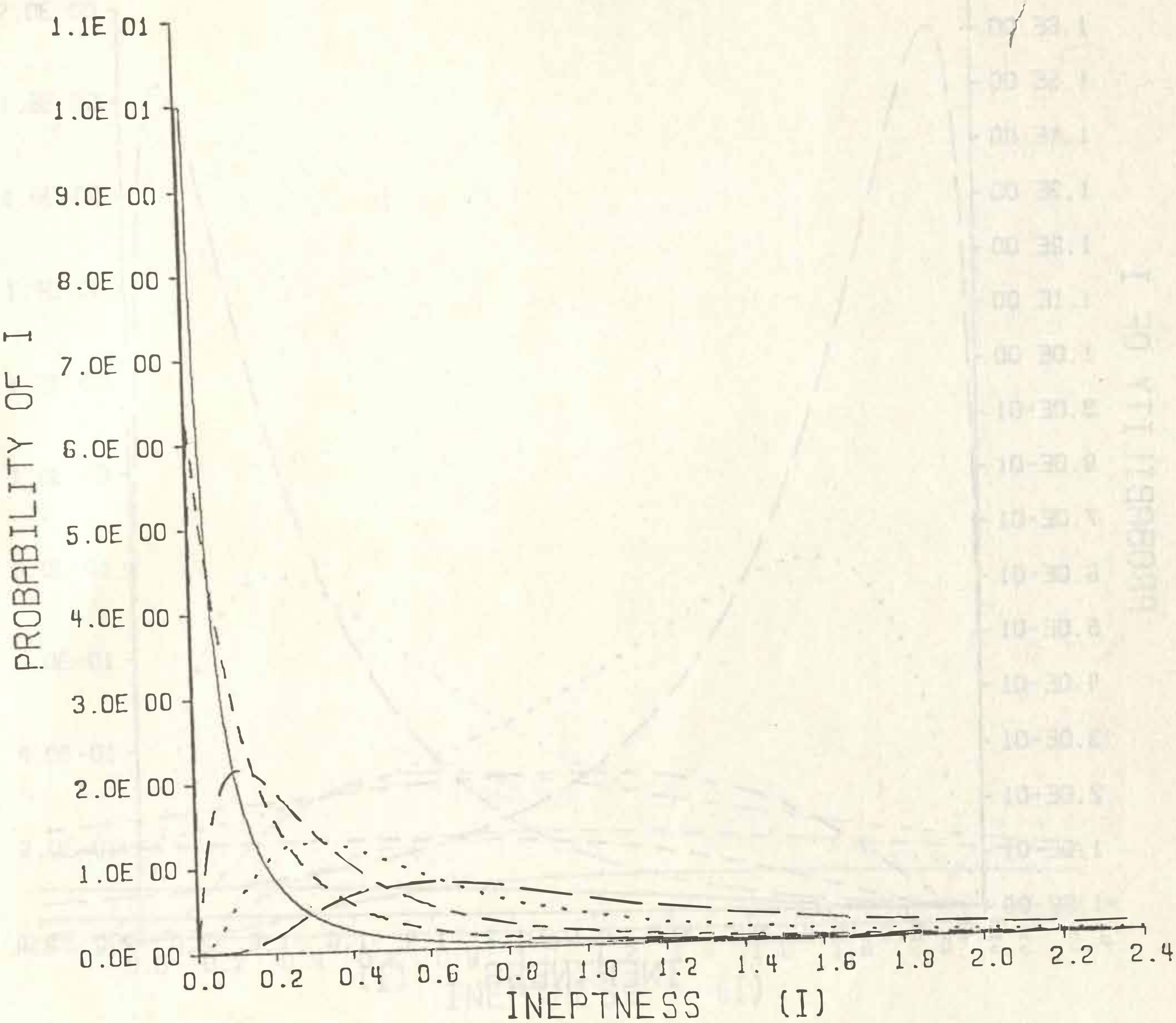

1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NT 4 5 7 10 15
LAMBDA= 0
NP= 3


1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NT 4 5 7 10 15
LAMBDA= 1
NP= 3


1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NT 4 5 7 10 15
LAMBDA= 2
NP= 3



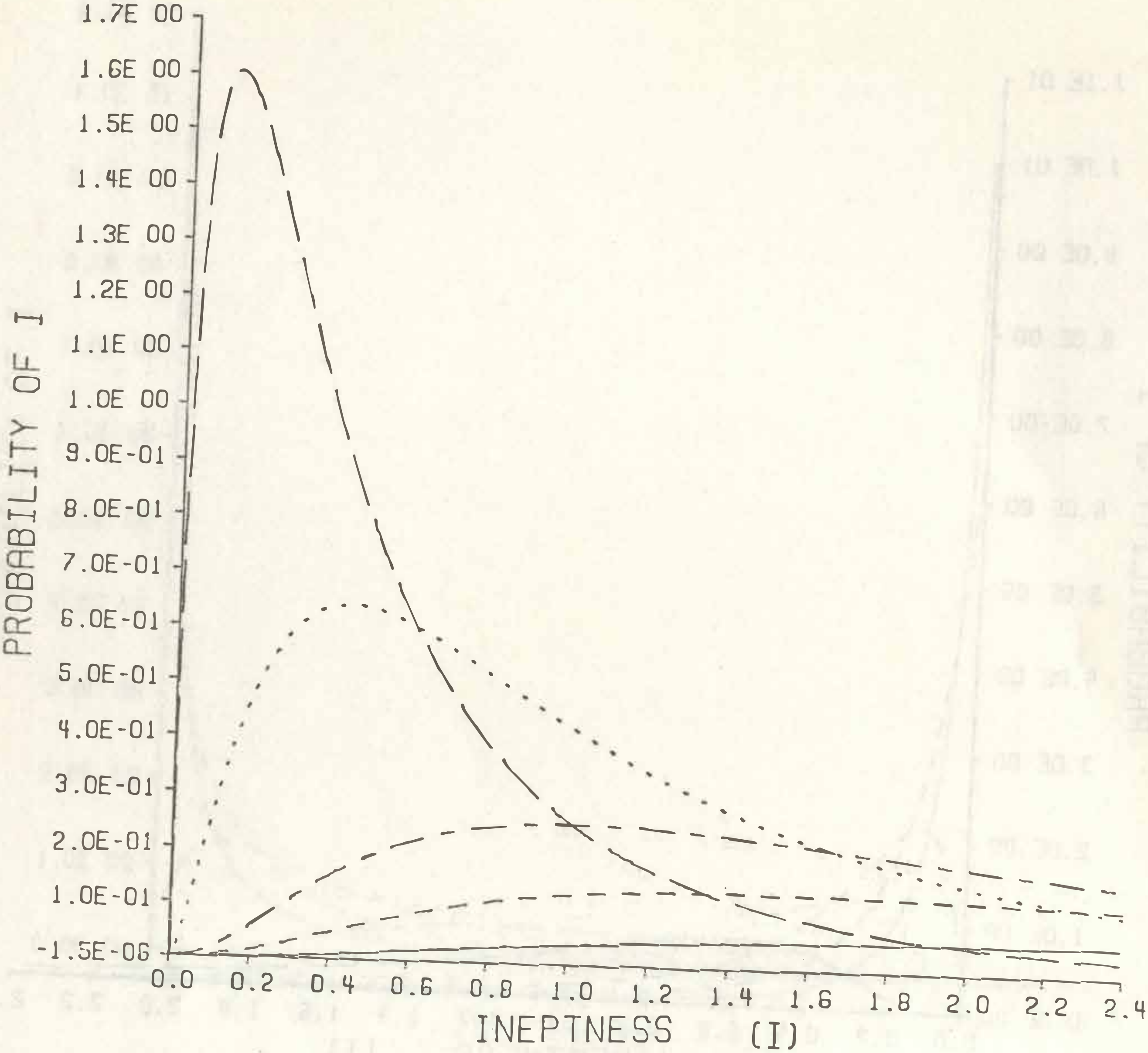
1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NT 4 5 7 10 15
LAMBDA= 5
NP= 3



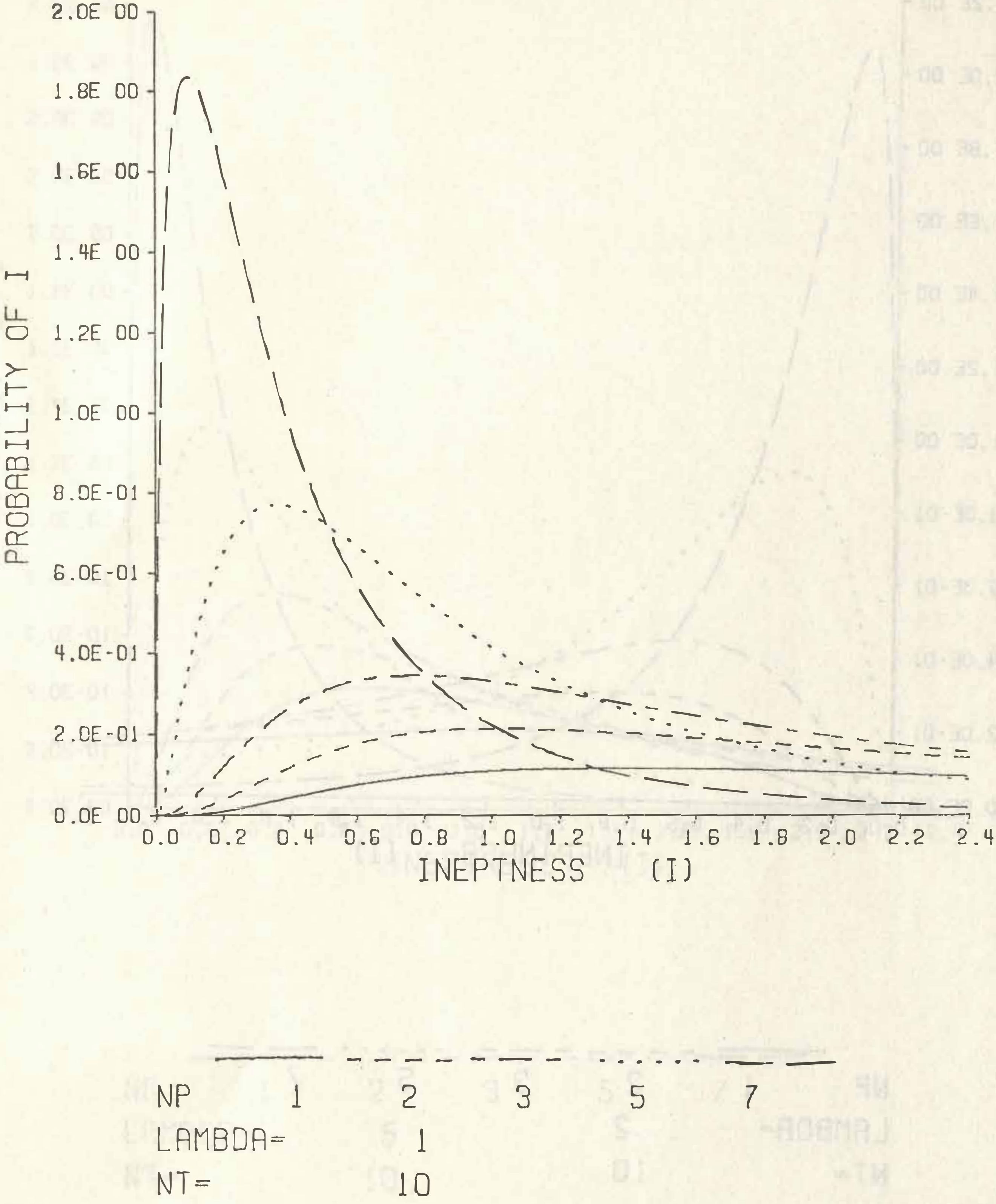
NT 4 5 7 10 15

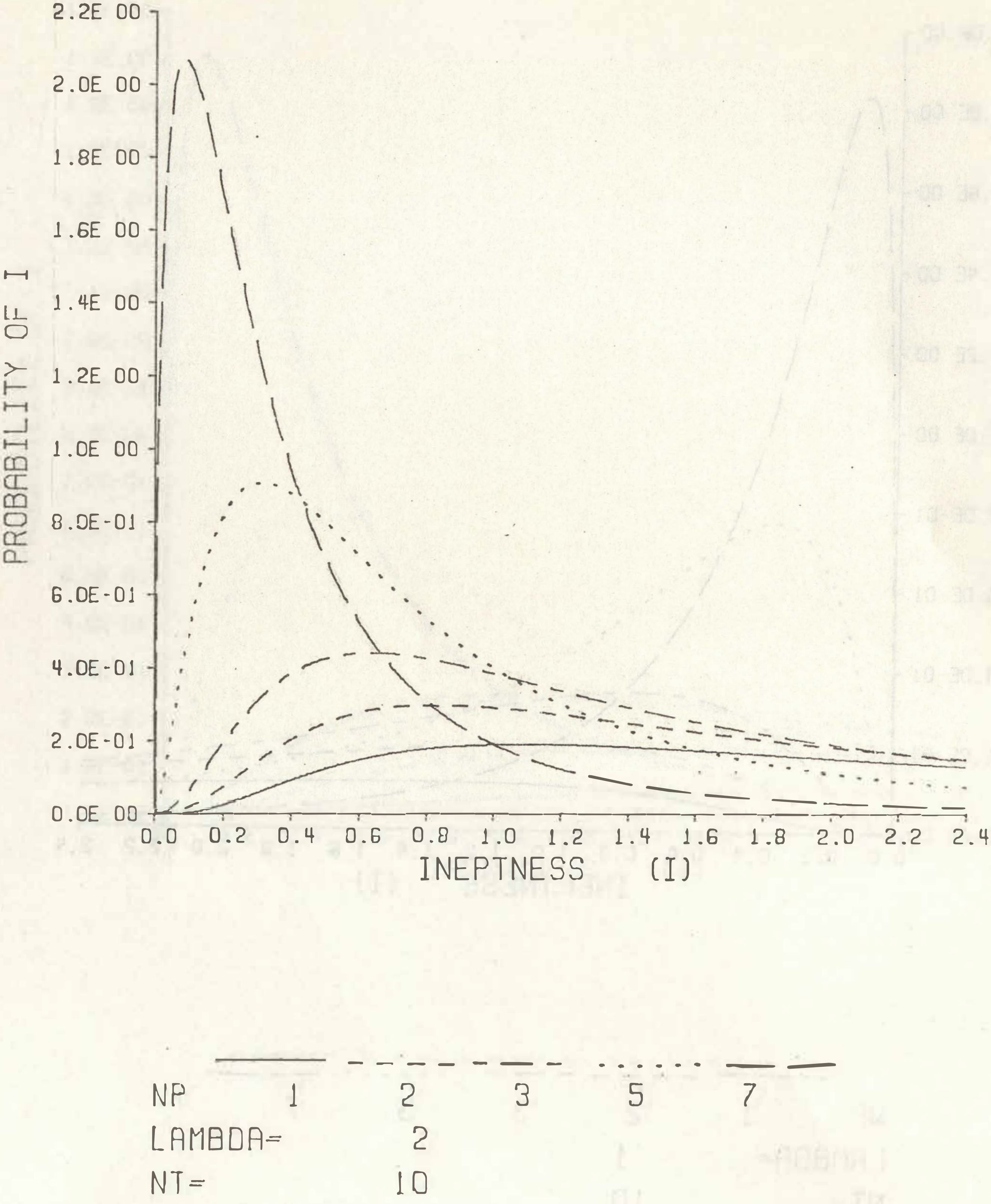
LAMBDA= 10

NP= 3

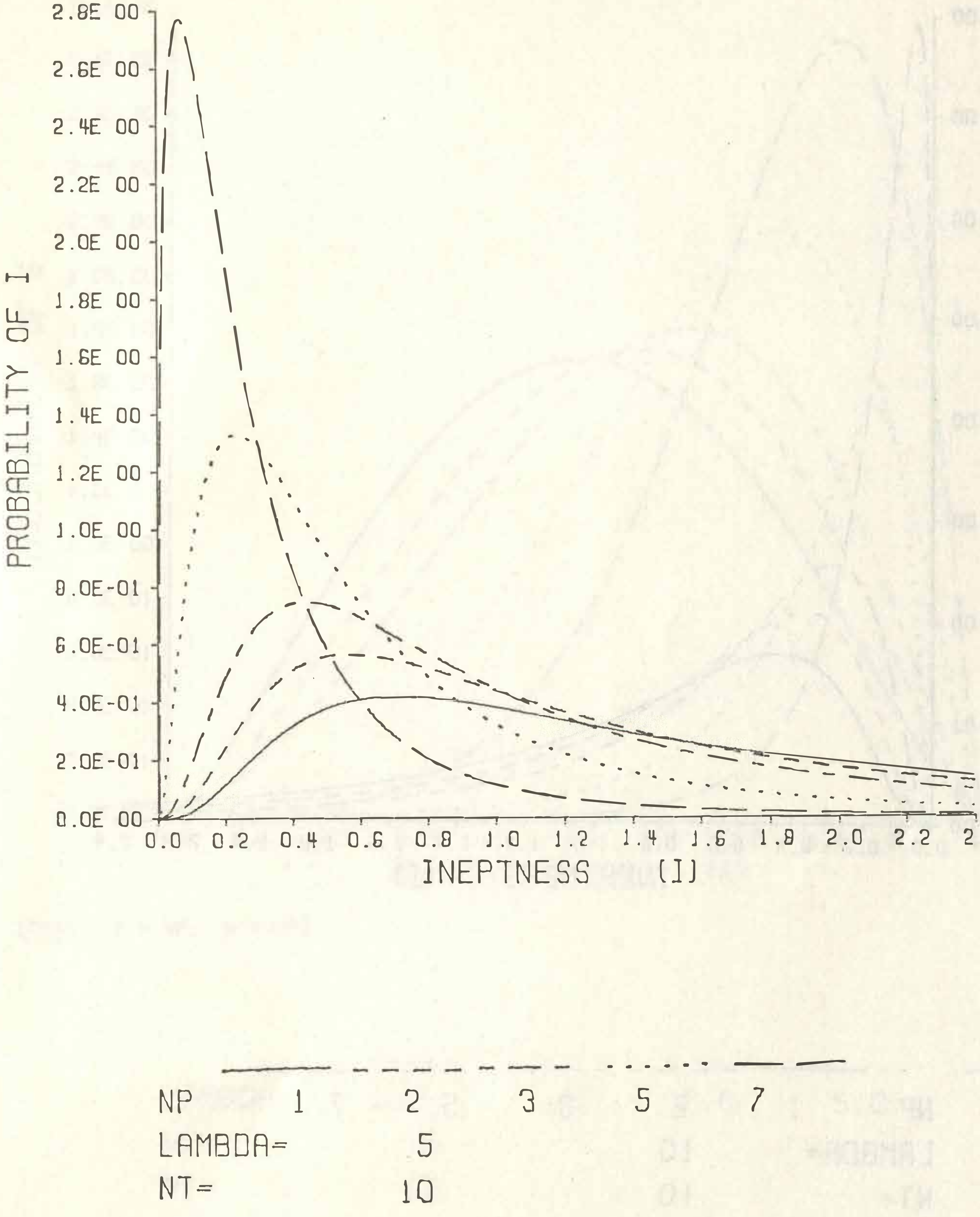


NP	1	2	3	5	7
LAMBDA=		0			
NT=		10			

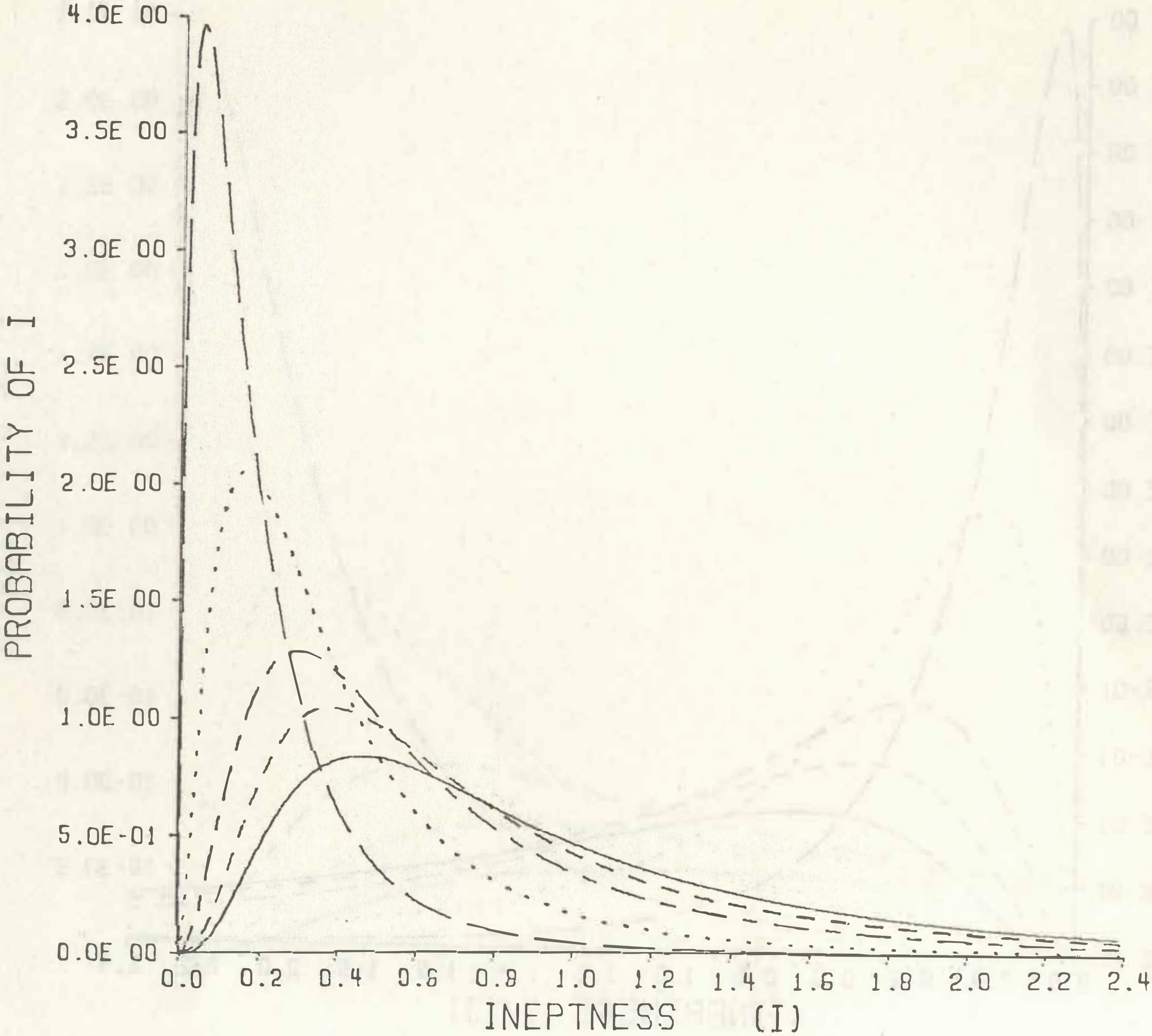

2.0E 00
1.8E 00
1.6E 00
1.4E 00
1.2E 00
1.0E 00
8.0E-01
6.0E-01
4.0E-01
2.0E-01
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NP 1 2 3 5 7
LAMBDA= 1
NT= 10



2.2E 00
2.0E 00
1.8E 00
1.6E 00
1.4E 00
1.2E 00
1.0E 00
8.0E-01
6.0E-01
4.0E-01
2.0E-01
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NP 1 2 3 5 7
LAMBDA= 2
NT= 10



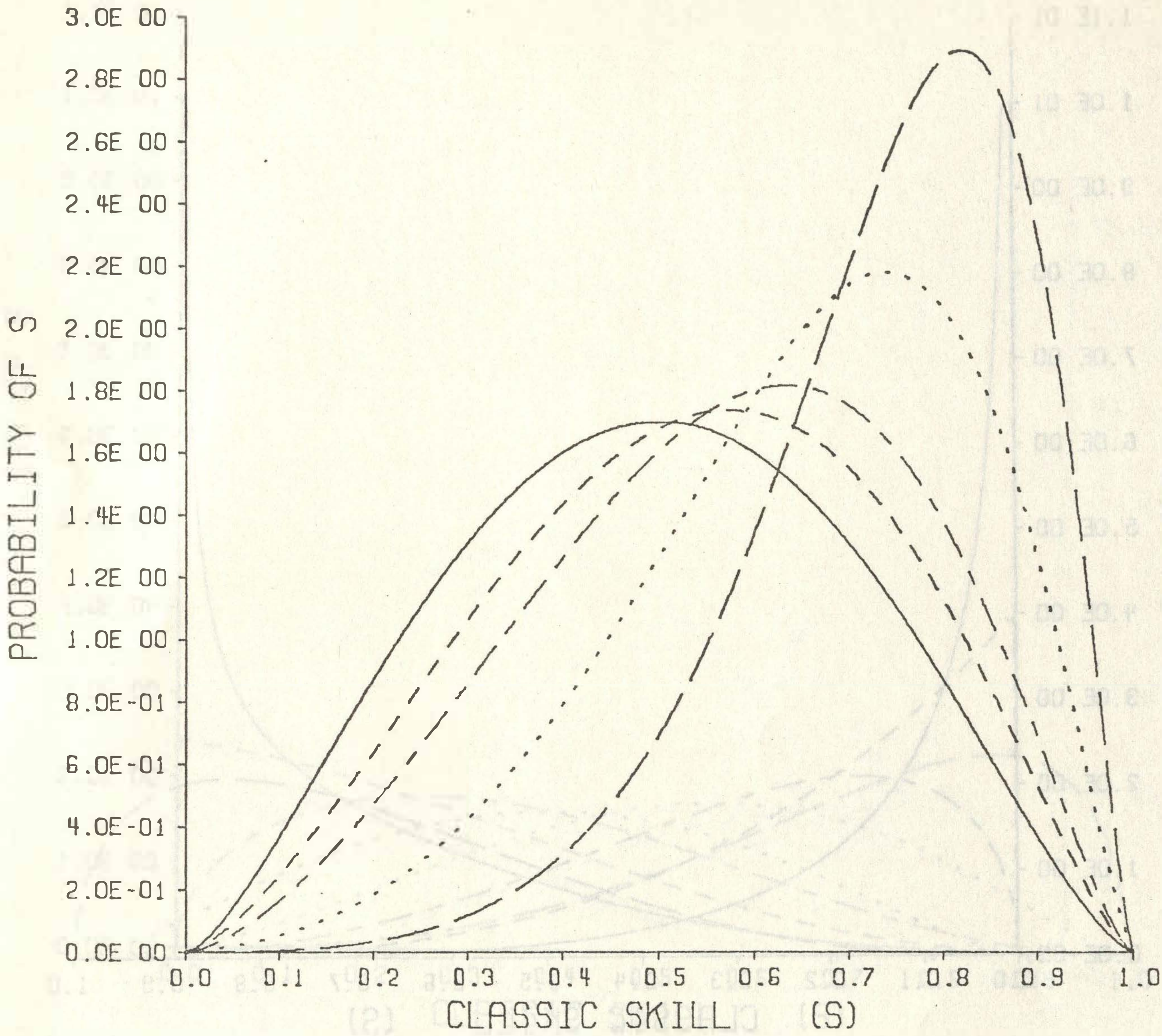
2.8E 00
2.6E 00
2.4E 00
2.2E 00
2.0E 00
1.8E 00
1.6E 00
1.4E 00
1.2E 00
1.0E 00
8.0E-01
6.0E-01
4.0E-01
2.0E-01
0.0E 00
PROBABILITY OF I
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
INEPTNESS (I)
NP 1 2 3 5 7
LAMBDA= 5
NT= 10



NP 1 2 3 5 7

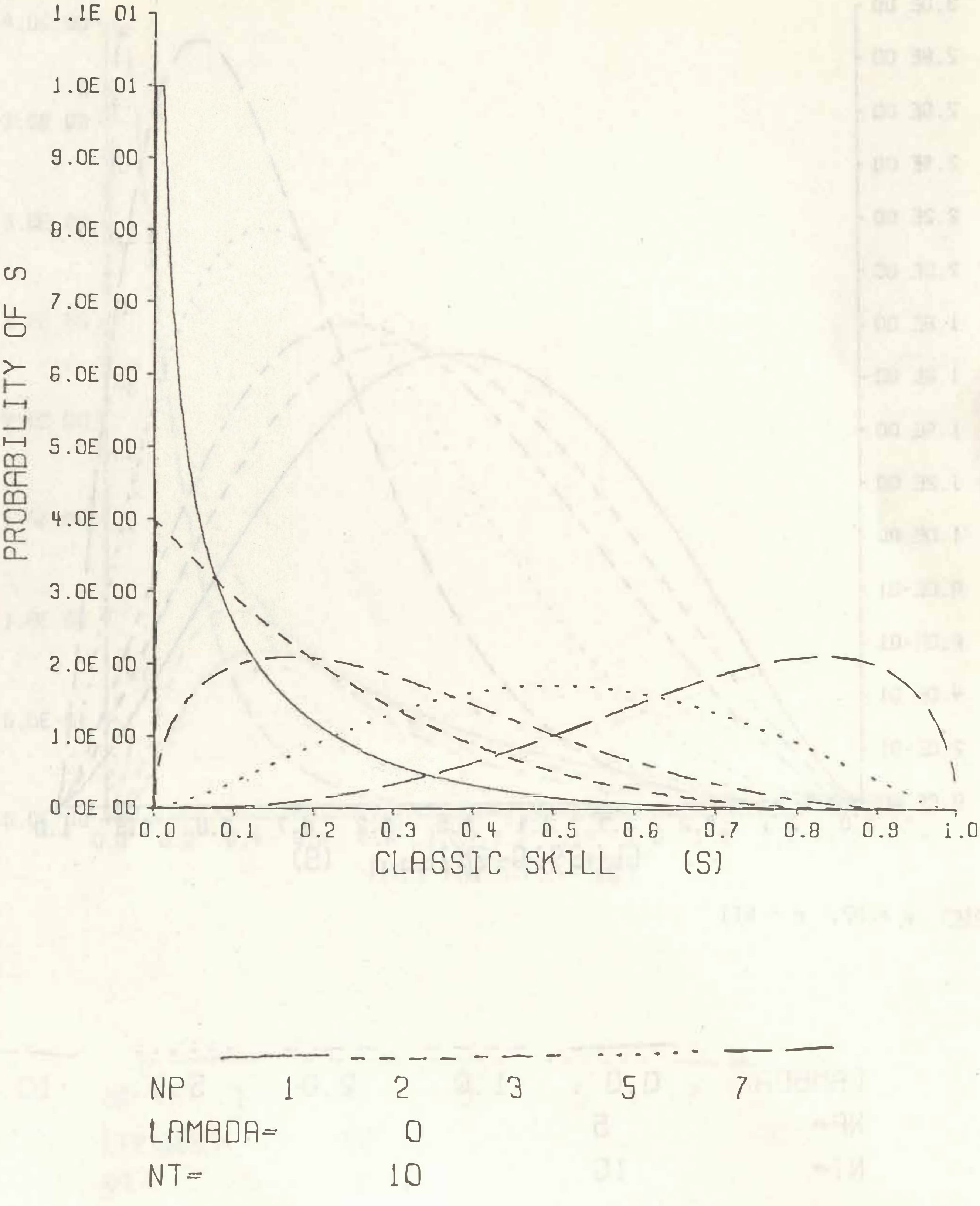
LAMBDA= 10

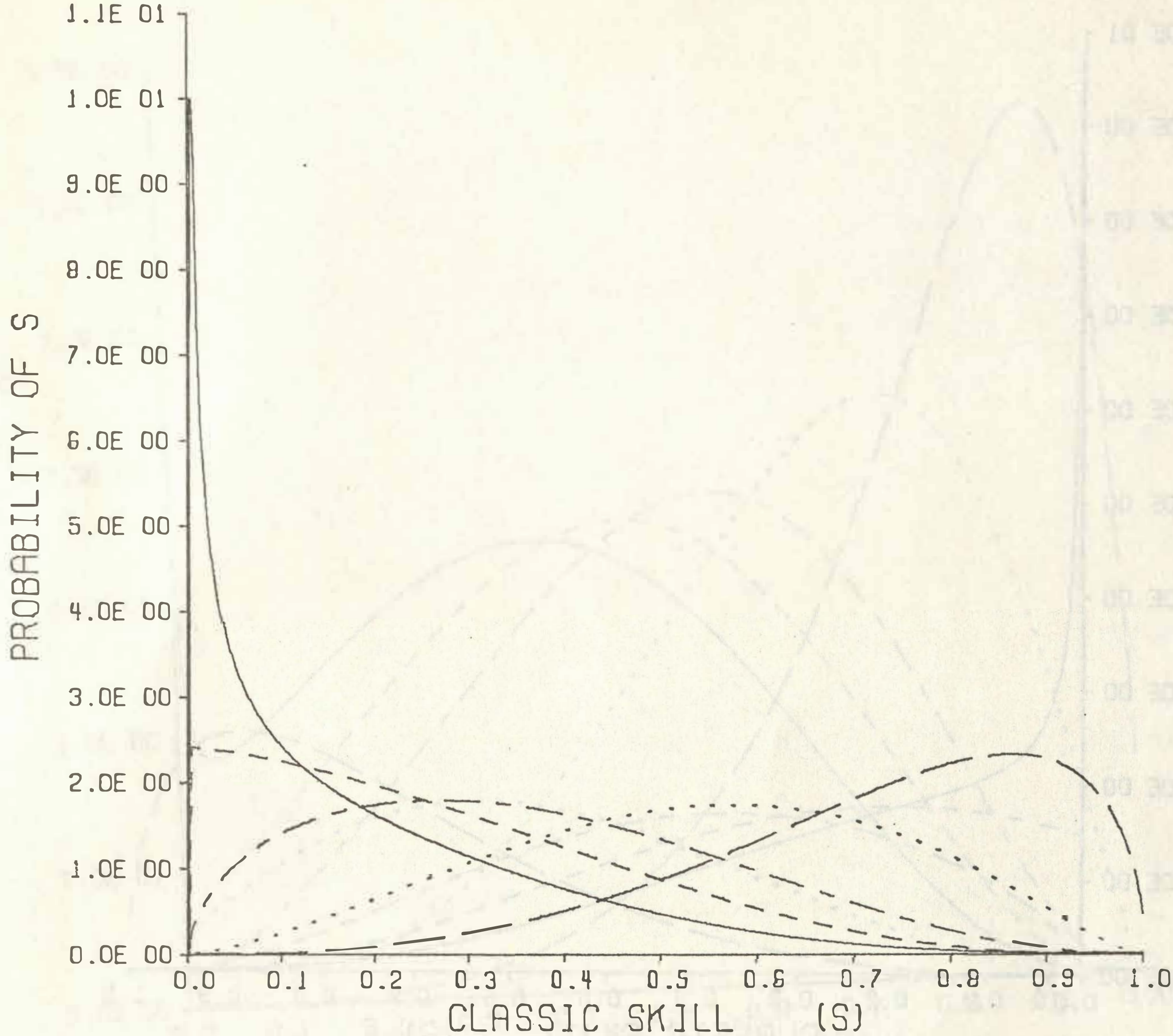
NT= 10



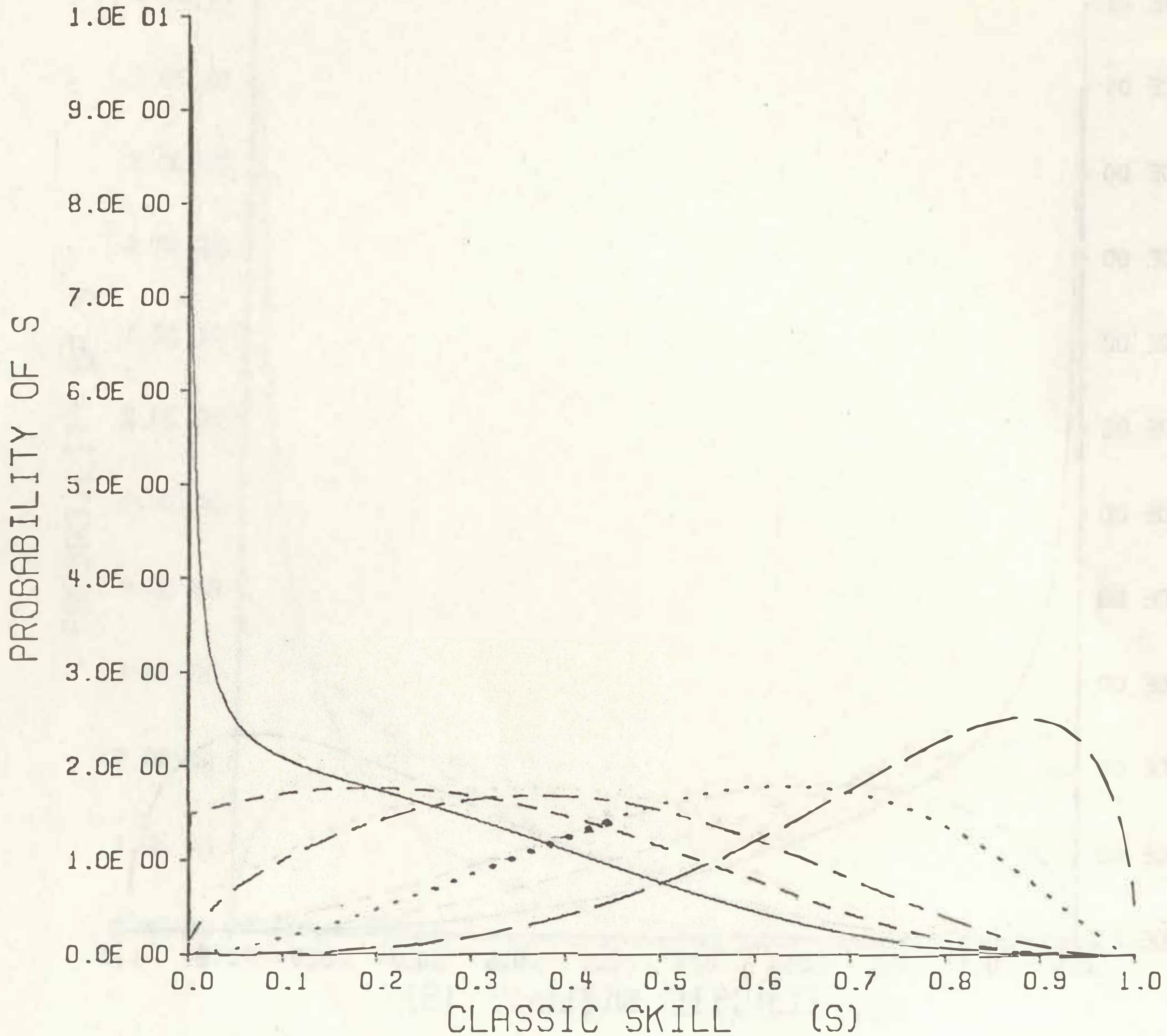
(Note: p = NP, n = NT)

LAMBDA	0.0	1.0	2.0	5.0	10.0
NP=	5				
NT=	10				


1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF S
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
CLASSIC SKILL (S)
NP 1 2 3 5 7
LAMBDA= 0
NT= 10



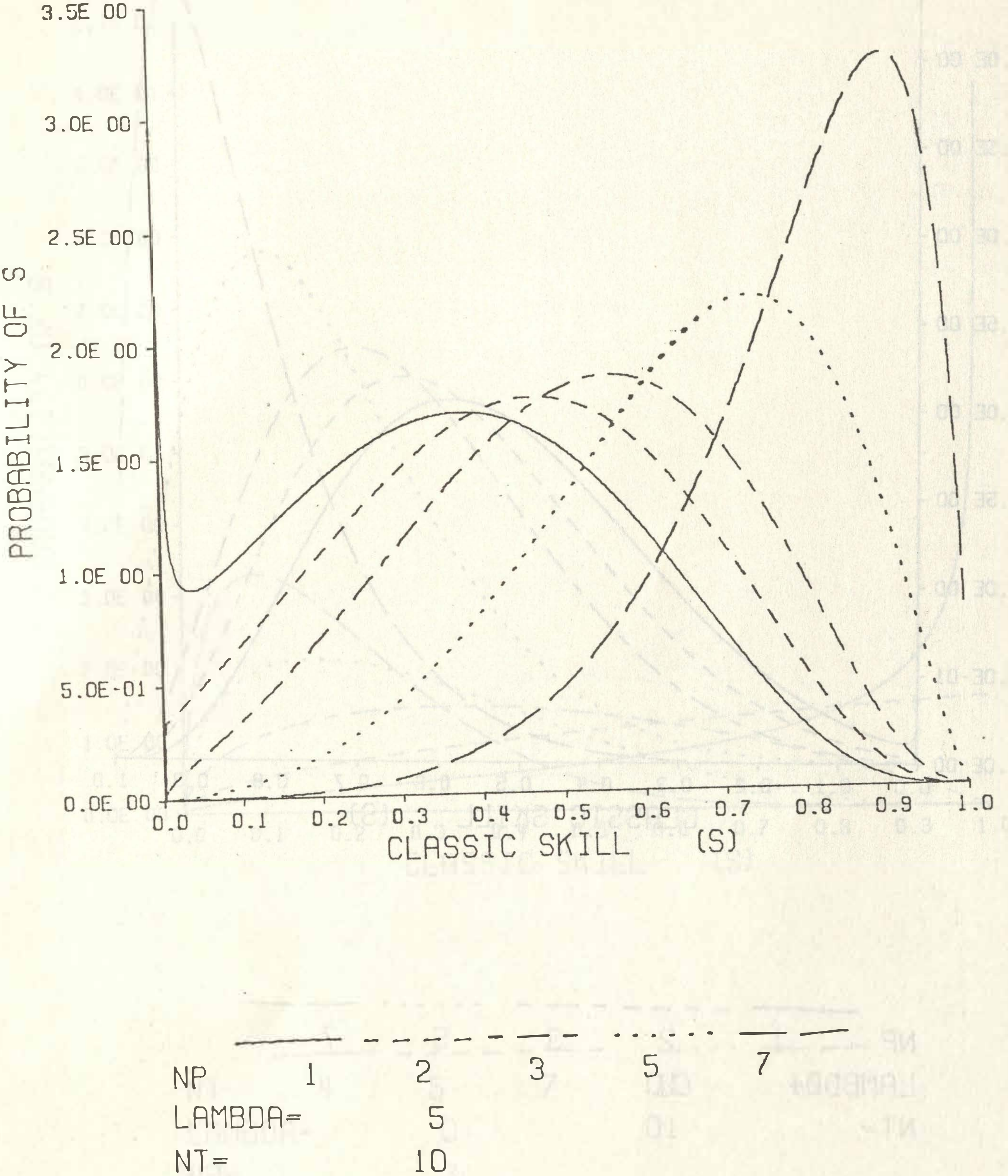
NP	1	2	3	5	7
LAMBDA=		1			
NT=		10			

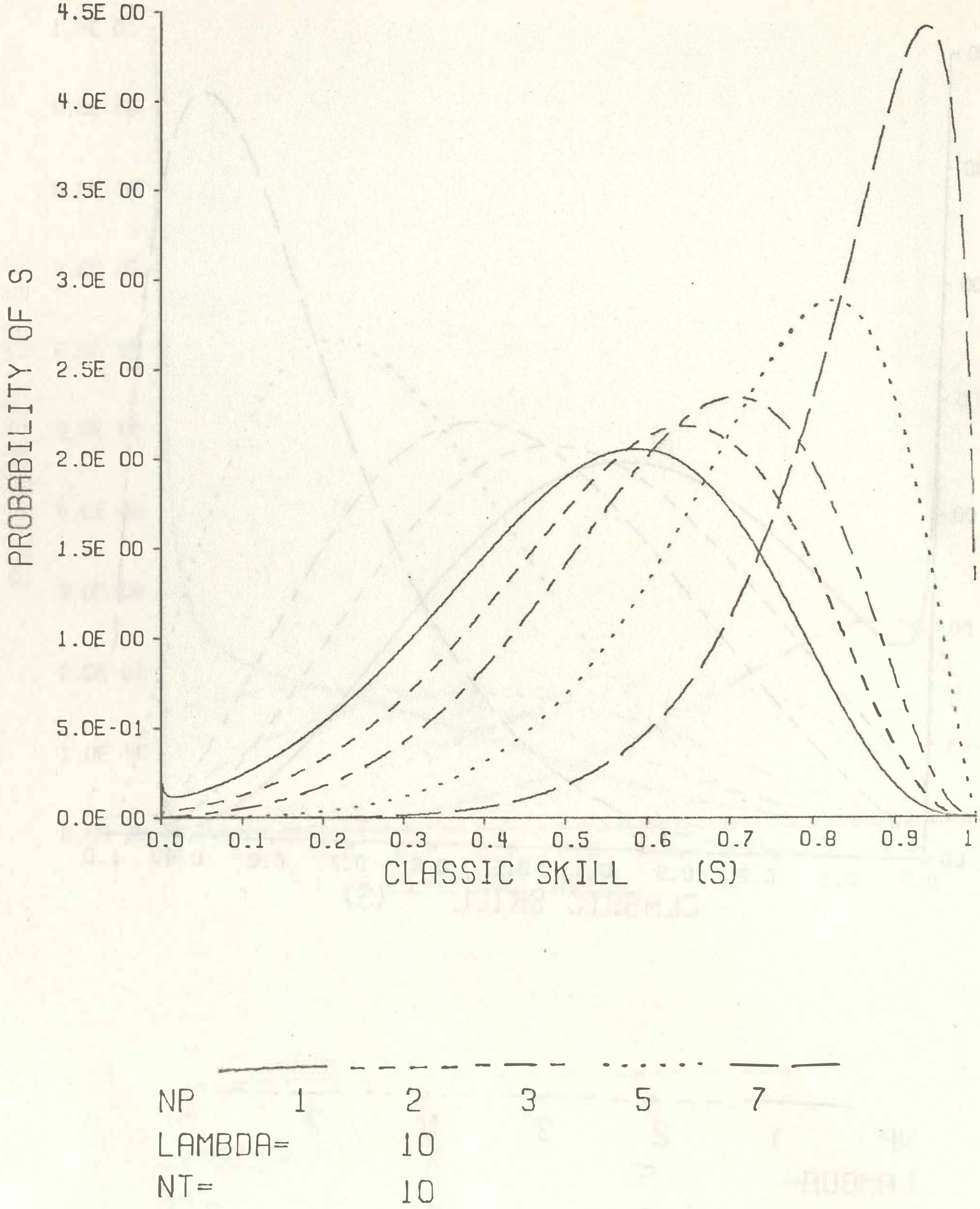


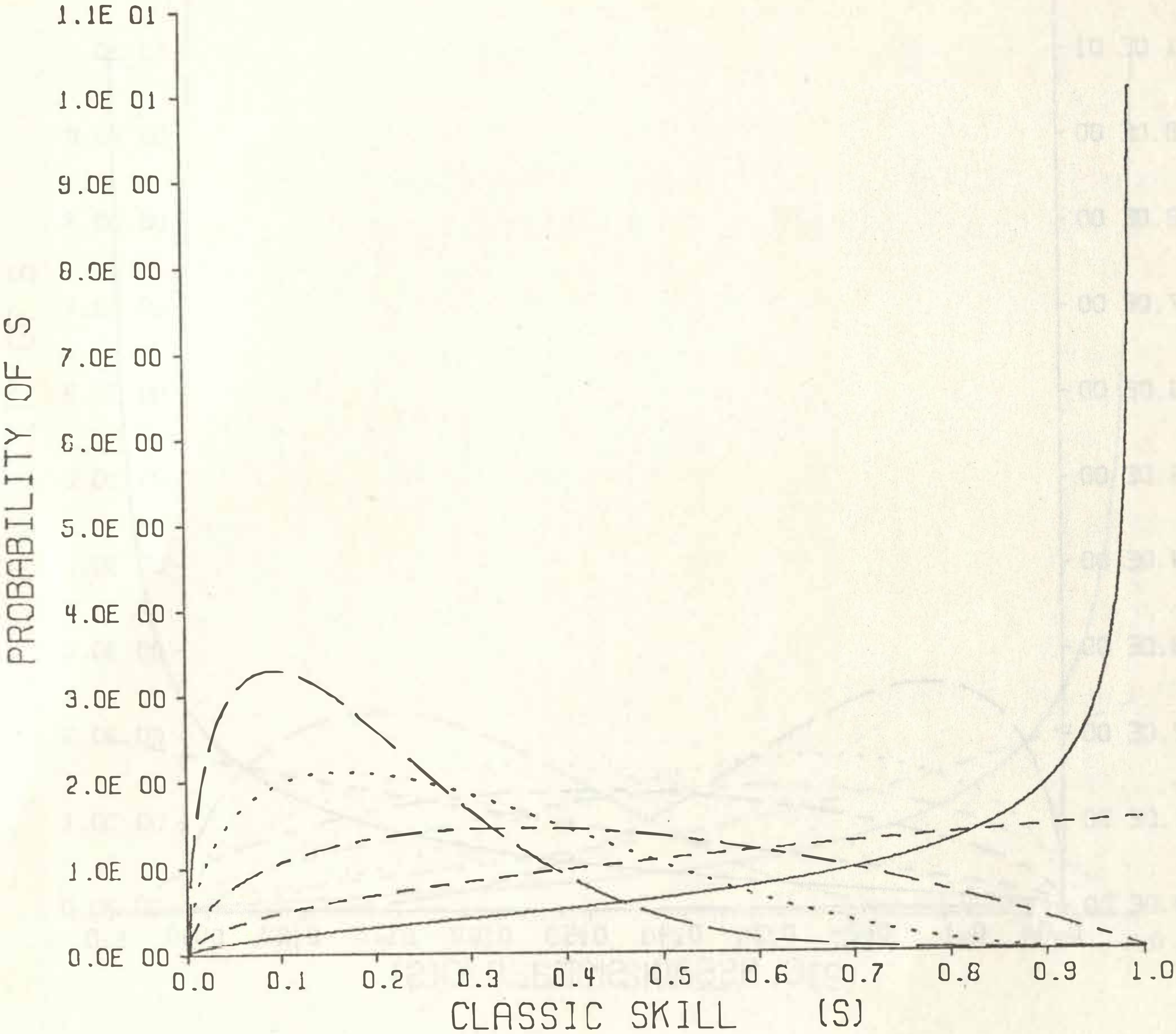
NP 1 2 3 5 7

LAMBDA= 2

NT= 10


3.5E 00
3.0E 00
2.5E 00
2.0E 00
1.5E 00
1.0E 00
5.0E-01
0.0E 00
PROBABILITY OF S
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
CLASSIC SKILL (S)
NP 1 2 3 5 7
LAMBDA= 5
NT= 10

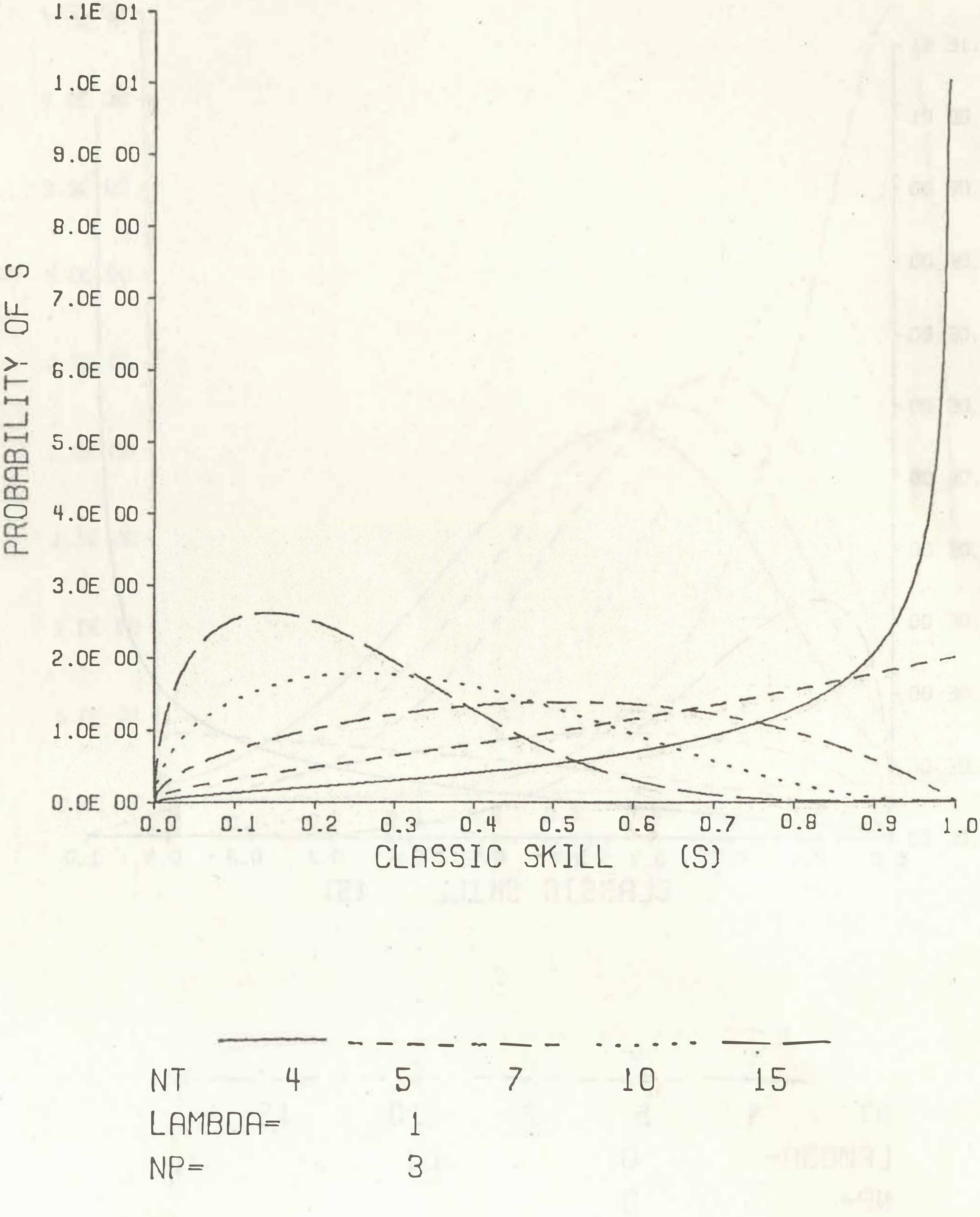

4.5E 00
4.0E 00
3.5E 00
3.0E 00
2.5E 00
2.0E 00
1.5E 00
1.0E 00
5.0E-01
0.0E 00
PROBABILITY OF S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CLASSIC SKILL (S)
NP 1 2 3 5 7
LAMBDA= 10
NT= 10



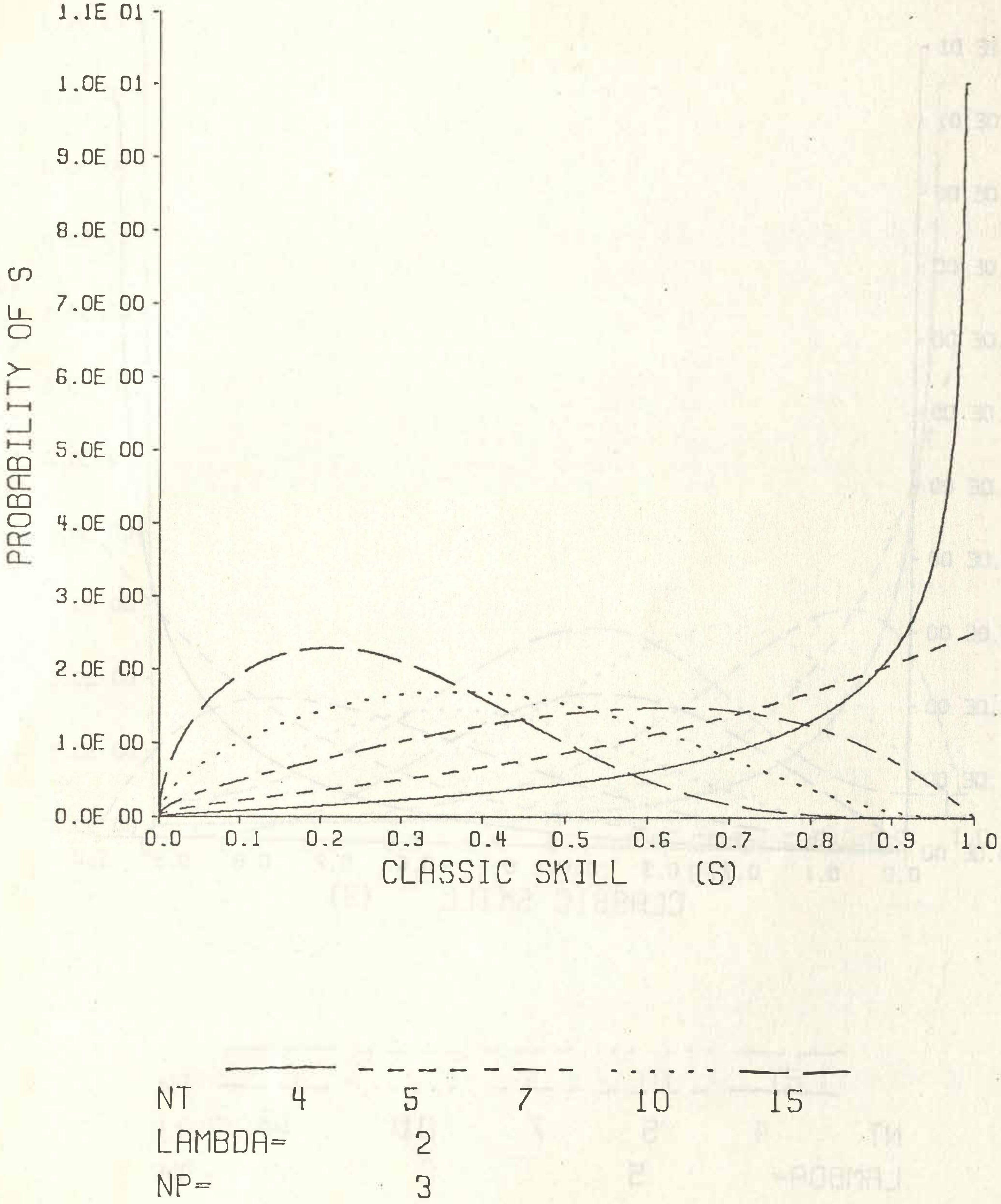
NT 4 5 7 10 15

LAMBDA= 0

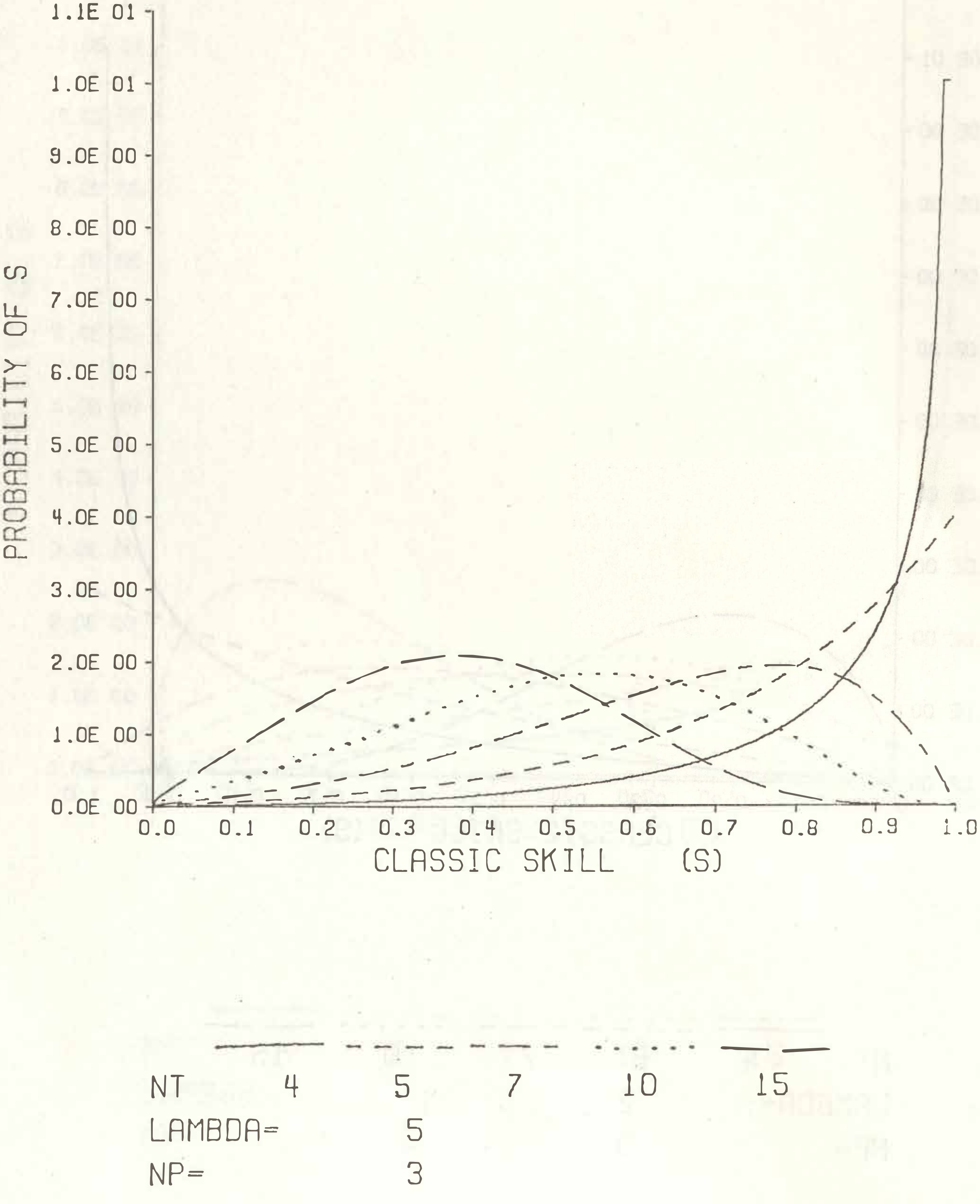
NP= 3



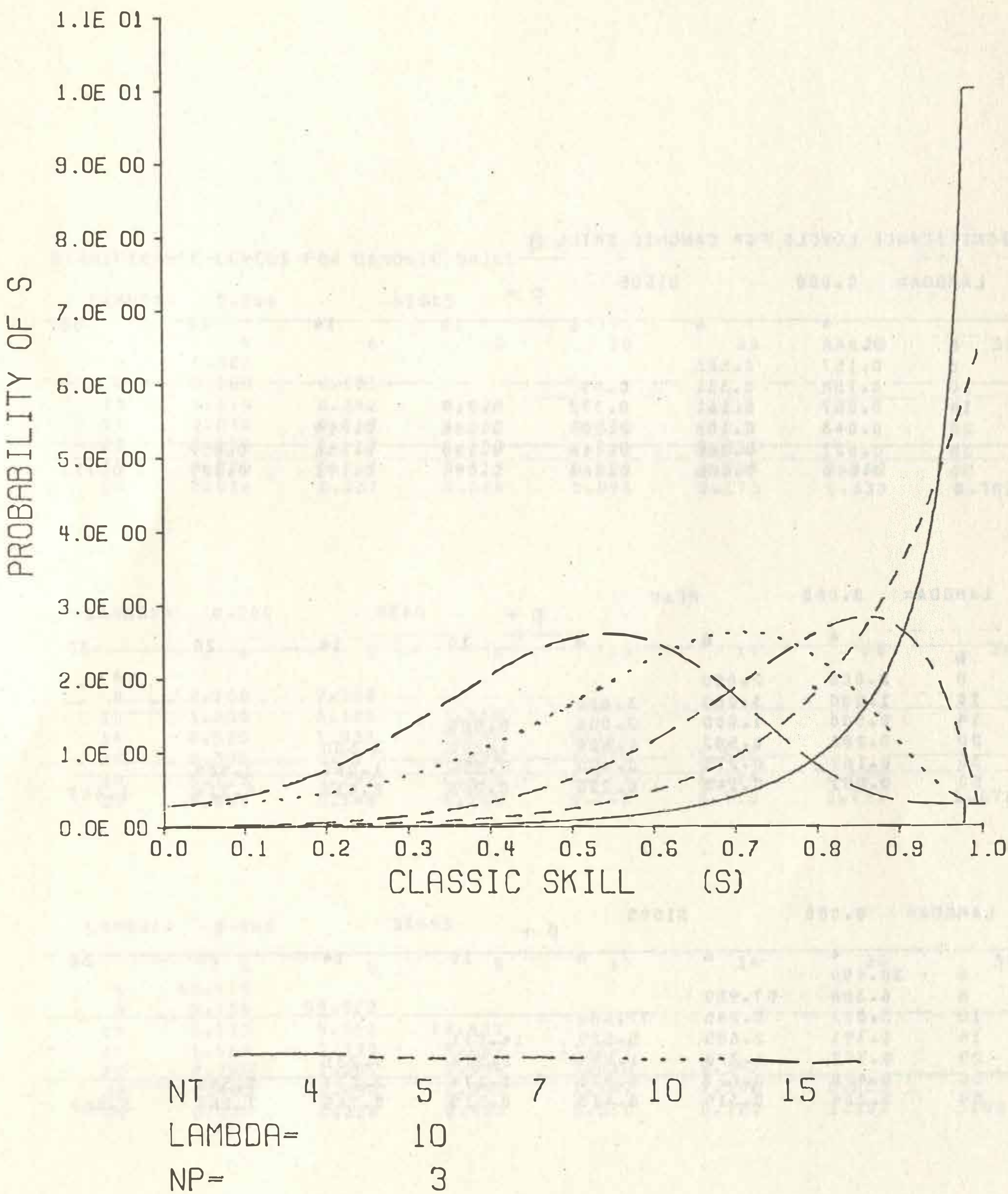
1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CLASSIC SKILL (S)
NT 4 5 7 10 15
LAMBDA= 1
NP= 3



1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CLASSIC SKILL (S)
NT 4 5 7 10 15
LAMBDA= 2
NP= 3



1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CLASSIC SKILL (S)
NT 4 5 7 10 15
LAMBDA= 5
NP= 3


1.1E 01
1.0E 01
9.0E 00
8.0E 00
7.0E 00
6.0E 00
5.0E 00
4.0E 00
3.0E 00
2.0E 00
1.0E 00
0.0E 00
PROBABILITY OF S
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
CLASSIC SKILL (S)
NT 4 5 7 10 15
LAMBDA= 10
NP= 3

SIGNIFICANCE LEVELS FOR CANONIC SKILL Q

LAMBDA= 0.000 SIG05

n ↓ \ p →	4	6	8	10	14	20	30
6	0.288						
8	0.157	0.583					
10	0.108	0.331	0.897				
14	0.067	0.161	0.372	0.719			
20	0.043	0.108.	0.203	0.336	0.819		
30	0.027	0.065	0.116.	0.180	0.358	0.852	
50	0.015.	0.036	0.063	0.094.	0.171	0.327	0.777.

LAMBDA= 0.000 MEAN

n ↓ \ p →	4	6	8	10	14	20	30
6							
8	2.000.	2.000					
10	1.000.	3.000.	3.000				
14	0.500	1.000.	2.000	5.000			
20	0.286	0.500.	0.800	1.250	3.500		
30	0.167	0.273	0.400	0.556	1.000	2.500	
50	0.091.	0.143	0.200.	0.263.	0.412.	0.714	1.667

LAMBDA= 0.000 SIG95

n ↓ \ p →	4	6	8	10	14	20	30
6	38.494						
8	6.388	57.989					
10	3.022	9.245	77.484.				
14	1.391	2.685	5.529	14.911			
20	0.752	1.220	1.899	2.978	9.231		
30	0.422	0.627	0.871	1.174	2.077	5.548.	
50	0.224	0.315	0.413	0.519	0.769	1.288	3.060

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 0.200 SIG05

	4	6	8	10	14	20	30
6	0.303						
8	0.165	0.603					
10	0.114	0.342	0.920				
14	0.070	0.187	0.382	0.733			
20	0.045	0.112	0.208	0.343	0.831		
30	0.028	0.067	0.119	0.184	0.363	0.860	
50	0.016	0.037	0.064	0.096	0.173	0.330	0.782

LAMBDA= 0.200 MEAN

	4	6	8	10	14	20	30
6							
8	2.100	2.100					
10	1.050	3.100	3.100				
14	0.525	1.033	2.050	5.100			
20	0.300	0.517	0.820	1.275	3.550		
30	0.175	0.282	0.410	0.567	1.014	2.529	
50	0.095	0.148	0.205	0.268	0.418	0.721	1.678

LAMBDA= 0.200 SIG95

	4	6	8	10	14	20	30
6	40.419						
8	6.706	59.922					
10	3.173	9.552	79.421				
14	1.460	2.775	5.667	15.209			
20	0.789	1.261	1.946	3.038	9.362		
30	0.443	0.648	0.893	1.197	2.106	5.603	
50	0.235	0.326	0.423	0.530	0.780	1.301	3.081

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 0.300 SIG05

	4	6	8.	10	14	20	30
6	0.310						
8	0.169	0.613.					
10	0.117.	0.348.	0.931				
14	0.072	0.190.	0.387.	0.741			
20	0.046	0.114	0.211	0.346	0.837		
30	0.029.	0.068	0.121	0.186	0.366	0.865	
50	0.016	0.038	0.065	0.097	0.174	0.332	0.785

LAMBDA= 0.300 MEAN

	4	6	8	10	14	20	30
6							
8	2.150	2.150					
10.	1.075.	3.150	3.150				
14	0.537.	1.050.	2.075.	5.150.			
20	0.307	0.525	0.830	1.287	3.575		
30	0.179	0.286	0.415	0.572	1.021	2.537	
50	0.098	0.150	0.207	0.271	0.421	0.725	1.683

LAMBDA= 0.300 SIG95

	4	6	8.	10.	14	20	30
6	41.383						
8	6.865	60.889					
10	3.247	9.706	80.390				
14	1.494	2.819	5.736	15.358			
20	0.807	1.281	1.970	3.067	9.428		
30	0.453	0.658	0.904	1.209	2.121	5.631	
50	0.240	0.331	0.428	0.535	0.785.	1.307	3.092

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 0.500 SIG05

	4	6	8	10	14	20	30
6	0.326						
8	0.177	0.633					
10	0.122	0.359	0.954				
14	0.076	0.196	0.396	0.755			
20	0.048	0.118	0.216	0.353	0.849		
30	0.030	0.071	0.124	0.189	0.371	0.873	
50	0.017	0.039	0.067	0.099	0.177	0.335	0.790

LAMBDA= 0.500 MEAN

	4	6	8	10	14	20	30
6							
8	2.250	2.250					
10	1.125	3.250	3.250				
14	0.563	1.083	2.125	5.250			
20	0.321	0.542	0.850	1.313	3.625		
30	0.188	0.295	0.425	0.583	1.036	2.563	
50	0.102	0.155	0.212	0.276	0.426	0.732	1.694

LAMBDA= 0.500 SIG95

	4	6	8	10	14	20	30
6	43.312						
8	7.180	62.824					
10	3.395	10.011	82.329				
14	1.561	2.907	5.873	15.655			
20	0.843	1.321	2.017	3.126	9.559		
30	0.473	0.678	0.925	1.232	2.150	5.686	
50	0.251	0.341	0.438	0.545	0.796	1.320	3.113

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 0.700 SIG05

	4	6	8	10	14	20	30
6	0.342						
8	0.186	0.654					
10	0.129	0.371	0.977				
14	0.080	0.203	0.406	0.770			
20	0.051	0.122	0.221	0.360	0.861		
30	0.032	0.073	0.127	0.193	0.376	0.882	
50	0.018	0.041	0.068	0.101	0.179	0.338	0.795

LAMBDA= 0.700 MEAN

	4	6	8	10	14	20	30
6							
8	2.350	2.350					
10	1.175	3.350	3.350				
14	0.567	1.117	2.175	5.350			
20	0.336	0.558	0.870	1.337	3.675		
30	0.196	0.305	0.435	0.594	1.050	2.587	
50	0.107	0.160	0.217	0.282	0.432	0.739	1.706

LAMBDA= 0.700 SIG95

	4	6	8	10	14	20	30
6	45.243						
8	7.492	64.760					
10	3.541	10.316	84.267				
14	1.628	2.994	6.009	15.951			
20	0.879	1.360	2.063	3.185	9.691		
30	0.493	0.698	0.946	1.255	2.180	5.742	
50	0.261	0.351	0.448	0.555	0.807	1.333	3.131

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 1.000 SIG05

	4	6	8	10	14	20	30
6	0.367						
8	0.200	0.685					
10	0.138	0.389	1.013				
14	0.086	0.213	0.421	0.792			
20	0.055	0.128	0.229	0.370	0.879		
30	0.034	0.077	0.131	0.199	0.384	0.895	
50	0.019	0.043	0.071	0.104	0.183	0.343	0.804

LAMBDA= 1.000 MEAN

	4	6	8	10	14	20	30
6							
8	2.500	2.500					
10	1.250	3.500	3.500				
14	0.625	1.167	2.250	5.500			
20	0.357	0.583	0.900	1.375	3.750		
30	0.208	0.318	0.450	0.611	1.071	2.625	
50	0.114	0.167	0.225	0.289	0.441	0.750	1.722

LAMBDA= 1.000 SIG95

	4	6	8	10	14	20	30
6	48.142						
8	7.958	67.665					
10	3.757	10.771	87.176				
14	1.725	3.124	6.213	16.396			
20	0.931	1.418	2.132	3.273	9.887		
30	0.522	0.728	0.978	1.289	2.224	5.824	
50	0.277	0.366	0.463	0.570	0.823	1.352	3.167

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 1.500 SIG05

	4	6	8	10	14	20	30
6	0.412						
8	0.224	0.740					
10	0.155	0.420	1.073				
14	0.096	0.230	0.446	0.831			
20	0.062	0.138	0.243	0.388	0.909		
30	0.038	0.083	0.139	0.208	0.397	0.917	
50	0.022	0.046	0.075	0.109	0.189	0.352	0.817

LAMBDA= 1.500 MEAN

	4	6	8	10	14	20	30
6							
8	2.750	2.7500					
10	1.375	3.750	3.750				
14	0.6880	1.250	2.375	5.750			
20	0.3930	0.625	0.950	1.438	3.875		
30	0.2290	0.341	0.475	0.639	1.107	2.688	
50	0.1250	0.179	0.238	0.303	0.456	0.7680	1.750

LAMBDA= 1.5000 SIG95

	4	6	8	10	14	20	30
6	52.979						
8	8.727	72.511					
10	4.112	11.524	92.027				
14	1.884	3.337	6.550	17.134			
20	1.015	1.514	2.246	3.418	10.213		
30	0.568	0.776	1.030	1.346	2.296	5.962	
50	0.301	0.390	0.488	0.595	0.850	1.384	3.220

SIGNIFICANCE LEVELS FOR CANONIC SKILL

LAMBDA= 2.000 SIG05

	4	6	8	10	14	20	30
6	0.460						
8	0.251	0.797					
10	0.174	0.453	1.135				
14	0.108	0.248	0.472	0.869			
20	0.069	0.149	0.258	0.407	0.940		
30	0.043	0.090	0.148	0.218	0.411	0.939	
50	0.025	0.050	0.080	0.114	0.196	0.360	0.832

LAMBDA= 2.000 MEAN

	4	6	8	10	14	20	30
6							
8	3.000	3.000					
10	1.500	4.000	4.000				
14	0.750	1.333	2.500	6.000			
20	0.429	0.667	1.000	1.500	4.000		
30	0.250	0.364	0.500	0.667	1.143	2.750	
50	0.136	0.190	0.250	0.316	0.471	0.786	1.778

LAMBDA= 2.000 SIG95

	4	6	8	10	14	20	30
6	57.823						
8	9.487	77.359					
10	4.460	12.273	96.880				
14	2.039	3.548	6.885	17.869			
20	1.097	1.607	2.359	3.561	10.538		
30	0.613	0.824	1.080	1.402	2.368	6.099	
50	0.325	0.414	0.511	0.619	0.876	1.416	3.286

SIGNIFICANCE LEVELS FOR INEPTNESS I

LAMBDA= 0.000 X05

n ↓ \ p →	4	6	8	10	14	20	30
6	0.026						
8	0.157	0.017					
10	0.331	0.108	0.013				
14	0.719	0.372	0.181	0.067			
20	1.330	0.819	0.527	0.336	0.108		
30	2.370	1.595	1.147	0.852	0.482	0.180	
50	4.468	3.170	2.421	1.926	1.301	0.777	0.327

LAMBDA= 0.000 MEAN

n ↓ \ p →	4	6	8	10	14	20	30
6	1.000						
8	2.000	0.500					
10	3.000	1.000	0.333				
14	5.000	2.000	1.000	0.500			
20	8.000	3.500	2.000	1.250	0.500		
30	13.000	6.000	3.667	2.500	1.333	0.556	
50	25.000	11.000	7.000	5.000	3.000	1.667	0.714

LAMBDA= 0.000 X95

n ↓ \ p →	4	6	8	10	14	20	30
6	3.472						
8	6.388	1.714					
10	9.245	3.022	1.115				
14	14.911	5.529	2.685	1.391			
20	23.376	9.230	4.926	2.978	1.220		
30	37.463	15.365	8.611	5.548	2.794	1.174	
50	65.614	27.612	15.945	10.644	5.869	3.060	1.288

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 0.200 X05

	4	6	8	10	14	20	30
6	0.025						
8	0.149	0.017					
10	0.315	0.105	0.013				
14	0.685	0.360	0.176	0.066			
20	1.268	0.793	0.514	0.329	0.107		
30	2.258	1.544	1.120	0.835	0.475	0.178	
50	4.258	3.069	2.363	1.888	1.283	0.769	0.325

LAMBDA= 0.200 MEAN

	4	6	8	10	14	20	30
6	0.952						
8	1.903	0.484					
10	2.855	0.967	0.325				
14	4.758	1.935	0.975	0.490			
20	7.613	3.386	1.951	1.225	0.493		
30	12.371	5.805	3.577	2.451	1.315	0.550	
50	21.887	10.642	6.826	4.902	2.958	1.650	0.710

LAMBDA= 0.200 X95

	4	6	8	10	14	20	30
6	3.304						
8	6.078	1.659					
10	8.796	2.924	1.087				
14	14.187	5.349	2.620	1.364			
20	22.241	8.929	4.805	2.920	1.203		
30	35.543	14.664	8.399	5.439	2.754	1.152	
50	62.427	26.713	15.552	10.435	5.787	3.030	1.279

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 0.300 X05

	4	6	8	10	14	20	30
6	0.024.						
8	0.146.	0.016.					
10	0.308.	0.103.	0.012				
14	0.669	0.355	0.174.	0.065			
20	1.239.	0.781.	0.508.	0.326.	0.106		
30	2.207	1.520.	1.106.	0.827.	0.471.	0.178	
50	4.162.	3.021	2.334.	1.870	1.274.	0.756	0.324

LAMBDA= 0.300. MEAN

	4	6	8	10	14	20	30
6	0.929						
8	1.857.	0.476					
10	2.786	0.952	0.321				
14	4.643	1.904.	0.964.	0.485.			
20	7.429	3.331	1.927.	1.213.	0.489		
30	12.072	5.711	3.533.	2.427	1.305	0.547	
50	21.358	10.470	6.745.	4.854	2.937	1.642	0.707.

LAMBDA= 0.300 X95

	4	6	8	10	14	20	30
6	3.223						
8	5.930	1.632					
10	8.581	2.876	1.074.				
14.	13.840	5.261	2.587.	1.350.			
20.	21.696	8.783	4.746.	2.891.	1.195		
30.	34.769	14.621	8.296.	5.365	2.735.	1.157	
50.	60.896	26.276	15.360.	10.332	5.746	3.015	1.275

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 0.500 X051

	4	6	8	10	14	20	30
6	0.023						
8	0.139	0.0161					
10	0.295	0.100	0.012				
14	0.640	0.3441	0.170	0.0541			
20	1.186	0.757	0.4961	0.3201	0.1051		
30	2.113	1.474	1.0811	0.812	0.465	0.176	
50	3.985	2.931	2.2811	1.835	1.257	0.758	0.3221

LAMBDA= 0.500 MEAN

	4	6	8	10	14	20	30
6	0.885						
8	1.770	0.4611					
10	2.6541	0.922	0.313				
14	4.424	1.643	0.940	0.4761			
20	7.078	3.226	1.881	1.1901	0.483		
30	11.5021	5.530	3.448	2.3801	1.287	0.5421	
50	20.3501	10.138	6.583	4.7501	2.8961	1.626	0.7021

LAMBDA= 0.500 X95

	4	6	8	10	14	20	30
6	3.070						
8	5.6461	1.580					
10	8.170	2.784	1.0481				
14	13.176	5.092	2.525	1.324			
20	20.655	8.550	4.631	2.835	1.178		
30	33.099	14.150	8.0951	5.280	2.697	1.145	
50	57.969	25.432	14.985	10.131	5.668	2.986	1.267

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 0.700 X05

	4	6	8	10	14	20	30
6.	0.022						
8	0.133	0.015					
10	0.282	0.097	0.012.				
14	0.614	0.334	0.166	0.063			
20	1.138	0.735.	0.485	0.314	0.103.		
30	2.029	1.432.	1.057	0.797	0.459.	0.174	
50	3.826	2.848	2.230	1.802	1.240	0.751	0.320

LAMBDA= 0.700 MEAN

	4	6	8	10	14	20	30
6	0.844						
8	1.687	0.446					
10	2.531.	0.893	0.306				
14	4.219	1.786.	0.918.	0.467			
20	6.750.	3.125	1.837	1.167.	0.476.		
30	10.969	5.357	3.367	2.335	1.269.	0.537.	
50	19.466	9.821	6.428	4.669	2.856.	1.610	0.698

LAMBDA= 0.700. X95

	4	6	8	10	14	20	30
6	2.925						
8	5.379.	1.530.					
10	7.782	2.696	1.023.				
14.	12.548	4.930	2.464	1.299			
20.	19.669	8.278.	4.519	2.780	1.162		
30	31.517	13.696.	7.899	5.178	2.659	1.134.	
50	55.197	24.621.	14.621	9.935	5.590	2.957.	1.258

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 1.000 X05

	4	6	8	10	14	20	30
6	0.021						
8	0.126	0.015					
10	0.266	0.093	0.011				
14	0.580	0.320	0.161	0.061			
20	1.074	0.705	0.469	0.306	0.101		
30	1.916	1.373	1.023	0.776	0.450	0.172	
50	3.616	2.733	2.158	1.755	1.217	0.740	0.316

LAMBDA= 1.000 MEAN

	4	6	8	10	14	20	30
6	0.787						
8	1.574	0.426					
10	2.361	0.852	0.296				
14	3.935	1.704	0.887	0.454			
20	6.295	2.983	1.773	1.135	0.466		
30	10.230	5.113.	3.251	2.269	1.244	0.529	
50	18.100	9.375	6.206	4.539	2.798	1.587	0.691

LAMBDA= 1.000 X95

	4	6	8	10	14	20	30
6	2.725						
8	5.006	1.459					
10	7.241	2.570	0.987				
14	11.673	4.699	2.377	1.262			
20	18.294	7.842	4.359	2.701	1.138		
30	29.311	13.051	7.618	5.030	2.605	1.118	
50	51.329	23.470	14.096	9.652	5.477	2.914	1.246

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 1.500 X05

	4	6	8	10	14	20	30
6	0.019						
8	0.115	0.014					
10	0.243	0.087	0.011				
14	0.531	0.300	0.153	0.058			
20	0.985	0.661	0.445	0.293	0.098		
30	1.759	1.288	0.971	0.743	0.435	0.168	
50	3.323	2.567	2.050	1.682	1.180	0.724	0.311

LAMBDA= 1.500 MEAN

	4	6	8	10	14	20	30
6	0.704						
8	1.407	0.395					
10	2.111	0.791	0.279				
14	3.518	1.561	0.837	0.433			
20	5.628	2.767	1.675	1.084	0.451		
30	9.146	4.744	3.071	2.167	1.203	0.517	
50	16.181	8.697	5.862	4.335	2.706	1.550	0.680

LAMBDA= 1.500 X95

	4	6	8	10	14	20	30
6	2.429						
8	4.456	1.351					
10	6.440	2.378	0.932				
14	10.374	4.345	2.242	1.204			
20	16.252	7.249	4.109	2.576	1.100		
30	26.032	12.061	7.180	4.796	2.517	1.091	
50	45.580	21.718	13.273	9.207	5.299	2.845	1.226

SIGNIFICANCE LEVELS FOR INEPTNESS

LAMBDA= 2.000 X05

	4	6	8	10	14	20	30
6	0.017						
8	0.105	0.013					
10	0.224	0.081	0.010				
14	0.490	0.282	0.145	0.056			
20	0.912	0.622	0.424	0.281	0.095		
30	1.630	1.214	0.926	0.713	0.422	0.164	
50	3.083	2.426	1.953	1.617	1.146	0.798	0.307

LAMBDA= 2.000 MEAN

	4	6	8	10	14	20	30
6	0.532						
8	1.264	0.368					
10	1.896	0.736	0.264				
14	3.161	1.471	0.793	0.415			
20	5.057	2.575	1.585	1.036	0.437		
30	8.217	4.414	2.907	2.073	1.164	0.505	
50	14.539	8.093	5.549	4.145	2.619	1.514	0.669

LAMBDA= 2.000 X95

	4	6	8	10	14	20	30
6	2.175						
8	3.982	1.255					
10	5.749	2.206	0.981				
14	9.253	4.027	2.117	1.150			
20	14.488	6.715	3.879	2.459	1.064		
30	23.196	11.169	6.777	4.579	2.434	1.066	
50	40.611	20.166	12.510	8.794	5.135	2.779	1.207

SIGNIFICANCE LEVELS FOR HINDCAST SKILL S 0

LAMBDA= 0.000 X05

p →

n ↓	2	30	5	7	10	15	20	30	50	70	100
3	0.10										
5	0.03	0.140									
7	0.02	0.080	0.30								
10	0.01	0.050	0.16	0.350							
15	0.01	0.03	0.10	0.19	0.38						
20	0.01	0.02	0.07	0.13	0.25	0.51					
30	0.00	0.01	0.04	0.08	0.15	0.29	0.460				
50	0.000	0.010	0.020	0.05	0.09	0.16	0.25	0.44			
70	0.00	0.00	0.02	0.03	0.06	0.11	0.17	0.29	0.58		
100	0.000	0.00	0.01	0.02	0.04	0.08	0.11	0.20	0.38	0.59	
150	0.00	0.00	0.01	0.01	0.03	0.05	0.07	0.13	0.25	0.37	0.57

SIGNIFICANCE LEVELS FOR HINDCAST SKILL S

LAMBDA= 0.0000 SBAR

p →

n ↓	2	3	5	7	10	15	20	30	50	70	100
3	0.67										
5	0.46	0.60									
7	0.29	0.430	0.71								
10	0.20	0.30	0.500	0.70							
15	0.13	0.200	0.33	0.47	0.67						
20	0.10	0.15	0.25	0.35	0.500	0.75					
30	0.07	0.10	0.170	0.23	0.330	0.500	0.67				
50	0.04	0.06	0.10	0.14	0.20	0.30	0.400	0.60			
70	0.03	0.04	0.07	0.10	0.14	0.21	0.29	0.43	0.710		
100	0.02	0.030	0.05	0.07	0.10	0.15	0.20	0.300	0.500	0.70	
150	0.010	0.020	0.03	0.05	0.07	0.10	0.13	0.20	0.33	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL S 0

LAMBDA= 0.000 X95

p →

n ↓	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.85	0.97									
7	0.70	0.83	0.98								
10	0.53	0.65	0.83	0.95							
15	0.37	0.47	0.62	0.75	0.90						
20	0.290	0.360	0.49	0.60	0.75	0.930					
30	0.20	0.25	0.34	0.43	0.54	0.71	0.85				
50	0.12	0.15	0.21	0.27	0.34	0.46	0.56	0.75			
70	0.09	0.11	0.15	0.19	0.25	0.34	0.42	0.57	0.83		
100	0.060	0.080	0.11	0.140	0.18	0.24	0.30	0.410	0.610	0.800	
150	0.04	0.05	0.07	0.09	0.12	0.16	0.20	0.28	0.42	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.200 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.11										
5	0.04	0.14									
7	0.02	0.08	0.31								
10	0.01	0.05	0.17	0.36							
15	0.01	0.03	0.10	0.19	0.38						
20	0.01	0.02	0.07	0.13	0.25	0.51					
30	0.00	0.01	0.04	0.08	0.15	0.30	0.46				
50	0.00	0.01	0.02	0.05	0.09	0.16	0.25	0.44			
70	0.00	0.01	0.02	0.03	0.06	0.11	0.17	0.30	0.58		
100	0.00	0.00	0.01	0.02	0.04	0.08	0.12	0.20	0.39	0.59	
150	0.00	0.00	0.01	0.01	0.03	0.05	0.08	0.13	0.25	0.37	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.200 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.68										
5	0.42	0.61									
7	0.30	0.44	0.72								
10	0.21	0.31	0.51	0.70							
15	0.14	0.21	0.34	0.47	0.67						
20	0.11	0.16	0.26	0.36	0.50	0.75					
30	0.07	0.11	0.17	0.24	0.34	0.50	0.67				
50	0.04	0.06	0.10	0.14	0.20	0.30	0.40	0.60			
70	0.03	0.05	0.07	0.10	0.15	0.22	0.29	0.43	0.72		
100	0.02	0.03	0.05	0.07	0.10	0.15	0.20	0.30	0.50	0.70	
150	0.01	0.02	0.03	0.05	0.07	0.10	0.13	0.20	0.33	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.200 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.88	0.97									
7	0.72	0.84	0.98								
10	0.55	0.66	0.84	0.95							
15	0.39	0.48	0.63	0.76	0.91						
20	0.30	0.38	0.50	0.61	0.75	0.93					
30	0.21	0.26	0.35	0.43	0.54	0.71	0.85				
50	0.13	0.16	0.22	0.27	0.35	0.46	0.57	0.75			
70	0.09	0.12	0.16	0.20	0.25	0.34	0.42	0.57	0.83		
100	0.07	0.08	0.11	0.14	0.18	0.24	0.30	0.41	0.62	0.80	
150	0.05	0.05	0.08	0.09	0.12	0.16	0.20	0.28	0.43	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.300 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.11										
5	0.04	0.152									
7	0.02	0.08	0.31								
10	0.012	0.05	0.17	0.36							
15	0.01	0.03	0.102	0.20	0.38						
20	0.01	0.022	0.07	0.14	0.262	0.512					
30	0.00	0.012	0.04	0.08	0.162	0.302	0.46				
50	0.00	0.012	0.032	0.05	0.09	0.16	0.25	0.44			
70	0.00	0.01	0.022	0.03	0.06	0.11	0.17	0.30	0.58		
100	0.00	0.00	0.012	0.02	0.04	0.082	0.12	0.20	0.392	0.59	
150	0.00	0.00	0.01	0.022	0.032	0.05	0.08	0.132	0.25	0.37	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.300 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.69										
5	0.422	0.622									
7	0.312	0.452	0.72								
10	0.222	0.322	0.51	0.71							
15	0.15	0.212	0.34	0.48	0.67						
20	0.11	0.162	0.26	0.36	0.51	0.75					
30	0.08	0.112	0.17	0.24	0.34	0.50	0.67				
50	0.05	0.07	0.11	0.14	0.20	0.30	0.40	0.60			
70	0.032	0.05	0.08	0.10	0.15	0.22	0.29	0.432	0.72		
100	0.02	0.032	0.05	0.07	0.10	0.15	0.20	0.302	0.50	0.70	
150	0.02	0.02	0.04	0.05	0.07	0.10	0.142	0.20	0.33	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.300 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.88	0.97									
7	0.73	0.842	0.98								
10	0.56	0.67	0.84	0.96							
15	0.402	0.492	0.64	0.76	0.942						
20	0.31	0.38	0.51	0.612	0.752	0.932					
30	0.22	0.27	0.36	0.442	0.552	0.712	0.85				
50	0.13	0.16	0.22	0.272	0.352	0.462	0.57	0.76			
70	0.102	0.12	0.16	0.202	0.25	0.342	0.42	0.57	0.83		
100	0.07	0.08	0.11	0.142	0.182	0.242	0.302	0.41	0.62	0.802	
150	0.05	0.06	0.08	0.102	0.12	0.16	0.20	0.28	0.43	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.500 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.120										
5	0.04	0.16									
7	0.03	0.09	0.320								
10	0.020	0.050	0.180	0.37							
150	0.01	0.03	0.100	0.20	0.39						
200	0.01	0.02	0.070	0.140	0.26	0.520					
30	0.000	0.010	0.050	0.090	0.16	0.300	0.47				
50	0.00	0.01	0.03	0.05	0.09	0.17	0.25	0.440			
70	0.00	0.01	0.02	0.03	0.06	0.11	0.17	0.30	0.59		
100	0.00	0.000	0.010	0.02	0.04	0.080	0.12	0.200	0.39	0.59	
150	0.00	0.000	0.010	0.02	0.03	0.050	0.08	0.13	0.25	0.37	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.5000 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.70										
5	0.44	0.63									
7	0.32	0.46	0.730								
10	0.23	0.330	0.520	0.71							
15	0.16	0.220	0.35	0.48	0.68						
20	0.120	0.17	0.27	0.36	0.510	0.760					
30	0.08	0.11	0.18	0.250	0.34	0.510	0.67				
50	0.05	0.07	0.11	0.150	0.21	0.310	0.410	0.60			
70	0.04	0.05	0.080	0.11	0.15	0.220	0.290	0.430	0.72		
100	0.02	0.03	0.050	0.07	0.10	0.15	0.20	0.300	0.500	0.70	
150	0.02	0.02	0.040	0.05	0.07	0.10	0.14	0.200	0.340	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.500 X95

	2	3	5	7	100	15	20	30	50	70	100
3	1.00										
5	0.890	0.97									
7	0.74	0.85	0.98								
10	0.58	0.68	0.85	0.96							
150	0.42	0.50	0.65	0.770	0.91						
20	0.330	0.400	0.51	0.62	0.76	0.930					
30	0.230	0.280	0.360	0.44	0.55	0.710	0.85				
50	0.140	0.17	0.230	0.28	0.35	0.46	0.57	0.76			
70	0.10	0.12	0.17	0.20	0.26	0.340	0.42	0.57	0.83		
100	0.07	0.09	0.12	0.14	0.18	0.250	0.300	0.41	0.62	0.80	
150	0.05	0.06	0.08	0.10	0.12	0.17	0.21	0.28	0.430	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.700 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.13										
5	0.050	0.16									
7	0.030	0.09	0.33								
10	0.020	0.06	0.18	0.37							
15	0.01	0.030	0.11	0.21	0.390						
20	0.010	0.02	0.080	0.14	0.260	0.52					
30	0.01	0.02	0.050	0.09	0.160	0.30	0.47				
50	0.00	0.01	0.03	0.05	0.09	0.17	0.25	0.44			
70	0.00	0.01	0.02	0.04	0.06	0.12	0.17	0.30	0.59		
100	0.00	0.00	0.01	0.02	0.040	0.08	0.12	0.20	0.39	0.59	
150	0.00	0.00	0.01	0.020	0.03	0.05	0.08	0.13	0.25	0.37	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.700 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.71										
5	0.46	0.64									
7	0.340	0.470	0.74								
10	0.240	0.340	0.53	0.72							
15	0.17	0.23	0.36	0.49	0.68						
20	0.13	0.18	0.270	0.37	0.52	0.760					
30	0.09	0.120	0.180	0.25	0.35	0.510	0.67				
50	0.05	0.07	0.11	0.15	0.21	0.31	0.41	0.61			
70	0.040	0.050	0.08	0.11	0.150	0.22	0.29	0.430	0.72		
100	0.030	0.040	0.06	0.080	0.110	0.16	0.21	0.300	0.500	0.70	
150	0.02	0.02	0.04	0.05	0.07	0.10	0.14	0.20	0.34	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 0.700 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.000										
5	0.90	0.97									
7	0.76	0.86	0.98								
10	0.60	0.70	0.850	0.96							
15	0.44	0.52	0.65	0.77	0.91						
20	0.350	0.410	0.52	0.630	0.760	0.94					
30	0.240	0.290	0.37	0.450	0.560	0.72	0.850				
50	0.15	0.18	0.230	0.28	0.36	0.47	0.57	0.76			
70	0.11	0.13	0.17	0.21	0.26	0.35	0.42	0.57	0.830		
100	0.08	0.090	0.120	0.15	0.19	0.250	0.31	0.42	0.62	0.80	
150	0.050	0.06	0.08	0.190	0.13	0.17	0.21	0.28	0.43	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 1.000 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.15										
5	0.05	0.18									
7	0.03	0.10	0.34								
10	0.02	0.06	0.19	0.38							
15	0.01	0.04	0.11	0.21	0.40						
20	0.01	0.03	0.08	0.15	0.27	0.52					
30	0.01	0.02	0.05	0.09	0.17	0.31	0.47				
50	0.00	0.01	0.03	0.05	0.09	0.17	0.26	0.44			
70	0.00	0.01	0.02	0.04	0.07	0.12	0.18	0.30	0.55		
100	0.00	0.00	0.01	0.03	0.04	0.08	0.12	0.20	0.39	0.59	
150	0.00	0.00	0.01	0.02	0.03	0.05	0.08	0.13	0.25	0.38	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 1.000 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.72										
5	0.48	0.65									
7	0.36	0.49	0.74								
10	0.26	0.35	0.54	0.72							
15	0.18	0.24	0.37	0.50	0.69						
20	0.14	0.19	0.28	0.38	0.52	0.76					
30	0.09	0.13	0.19	0.26	0.35	0.52	0.68				
50	0.06	0.08	0.12	0.16	0.22	0.31	0.41	0.61			
70	0.04	0.06	0.08	0.11	0.15	0.23	0.30	0.44	0.72		
100	0.03	0.04	0.06	0.08	0.11	0.16	0.21	0.31	0.50	0.70	
150	0.02	0.03	0.04	0.05	0.07	0.11	0.14	0.21	0.34	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 1.000 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.90	0.97									
7	0.77	0.87	0.98								
10	0.62	0.71	0.86	0.96							
15	0.46	0.53	0.66	0.78	0.91						
20	0.37	0.43	0.54	0.63	0.77	0.94					
30	0.26	0.30	0.38	0.46	0.56	0.72	0.85				
50	0.16	0.19	0.24	0.29	0.36	0.47	0.57	0.76			
70	0.12	0.14	0.18	0.21	0.27	0.35	0.43	0.57	0.83		
100	0.09	0.10	0.13	0.15	0.19	0.25	0.31	0.42	0.62	0.80	
150	0.06	0.07	0.09	0.10	0.13	0.17	0.21	0.29	0.43	0.56	0.75

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 1.500 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.18										
5	0.07	0.20									
7	0.04	0.12	0.36								
10	0.03	0.07	0.21	0.40							
15	0.02	0.04	0.12	0.22	0.41						
20	0.01	0.03	0.09	0.16	0.28	0.53					
30	0.01	0.02	0.06	0.10	0.17	0.31	0.48				
50	0.00	0.01	0.03	0.06	0.10	0.17	0.26	0.45			
70	0.00	0.01	0.02	0.04	0.07	0.12	0.18	0.31	0.59		
100	0.00	0.01	0.02	0.03	0.05	0.08	0.12	0.21	0.39	0.59	
150	0.00	0.00	0.01	0.02	0.03	0.05	0.08	0.13	0.25	0.38	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMPDA= 1.500 SBAR

	2	3	5	7	10	15	20	30	50	70	100
3	0.75										
5	0.51	0.67									
7	0.39	0.51	0.76								
10	0.29	0.38	0.56	0.73							
15	0.20	0.27	0.39	0.51	0.69						
20	0.16	0.20	0.30	0.39	0.53	0.77					
30	0.11	0.14	0.20	0.27	0.36	0.52	0.68				
50	0.07	0.09	0.13	0.16	0.22	0.32	0.42	0.61			
70	0.05	0.06	0.09	0.12	0.16	0.23	0.30	0.44	0.72		
100	0.03	0.04	0.06	0.08	0.11	0.16	0.21	0.31	0.51	0.70	
150	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.21	0.34	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 1.500 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.92	0.58									
7	0.80	0.88	0.98								
10	0.65	0.73	0.87	0.96							
15	0.49	0.56	0.68	0.79	0.92						
20	0.40	0.45	0.55	0.65	0.77	0.94					
30	0.28	0.32	0.40	0.47	0.57	0.73	0.86				
50	0.18	0.21	0.26	0.30	0.37	0.48	0.58	0.76			
70	0.13	0.15	0.19	0.22	0.28	0.36	0.43	0.58	0.83		
100	0.10	0.11	0.13	0.16	0.20	0.26	0.31	0.42	0.62	0.80	
150	0.07	0.07	0.09	0.11	0.13	0.17	0.21	0.29	0.43	0.57	0.76

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 2.000 X05

	2	3	5	7	10	15	20	30	50	70	100
3	0.22										
5	0.08	0.23									
7	0.05	0.13	0.39								
10	0.03	0.08	0.22	0.41							
15	0.02	0.05	0.13	0.24	0.42	0					
20	0.01	0.04	0.10	0.170	0.29	0.54					
30	0.01	0.02	0.06	0.10	0.18	0.32	0.480				
50	0.01	0.01	0.03	0.060	0.10	0.18	0.26	0.45			
70	0.00	0.010	0.02	0.04	0.07	0.120	0.18	0.31	0.59		
100	0.00	0.010	0.02	0.03	0.05	0.090	0.13	0.21	0.390	0.60	
150	0.00	0.00	0.01	0.02	0.03	0.06	0.08	0.14	0.25	0.38	0.58

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 2.000 SBAR

	2	3	50	7	10	15	20	30	50	70	100
3	0.77										
5	0.54	0.69									
7	0.42	0.54	0.77								
10	0.32	0.40	0.570	0.74							
150	0.23	0.29	0.40	0.52	0.70						
20	0.18	0.22	0.31	0.400	0.54	0.77					
300	0.12	0.15	0.22	0.28	0.37	0.53	0.69				
50	0.08	0.09	0.13	0.17	0.23	0.33	0.42	0.61			
70	0.050	0.07	0.100	0.12	0.17	0.24	0.31	0.440	0.72		
100	0.04	0.050	0.07	0.090	0.12	0.17	0.22	0.310	0.51	0.71	
150	0.03	0.03	0.05	0.06	0.08	0.11	0.14	0.21	0.34	0.47	0.67

SIGNIFICANCE LEVELS FOR HINDCAST SKILL

LAMBDA= 2.0000 X95

	2	3	5	7	10	15	20	30	50	70	100
3	1.00										
5	0.93	0.98									
7	0.82	0.89	0.99								
10	0.68	0.75	0.87	0.96							
15	0.52	0.58	0.70	0.80	0.920						
20	0.42	0.47	0.57	0.66	0.78	0.94					
30	0.30	0.34	0.42	0.49	0.58	0.73	0.86				
50	0.20	0.22	0.27	0.32	0.38	0.49	0.59	0.77			
70	0.14	0.16	0.20	0.23	0.28	0.36	0.44	0.58	0.84		
100	0.10	0.12	0.14	0.17	0.20	0.26	0.32	0.43	0.62	0.81	
150	0.07	0.08	0.10	0.11	0.14	0.180	0.22	0.29	0.43	0.57	0.76