Journal of Engineering Mathematics, Vol. 4, No. 3, July 1970
Wolters-Noordhoff Publishing–Groningen
Printed in the Netherlands

Bayesian Identification of System Parameters*

R. B. KERR

Duke University, Department of Electrical Engineering, Durham, N.C., U.S.A.

(Received June 12, 1969)

SUMMARY

A common problem is that of identifying system parameters from short-term signal observations. If the signals have random components, and the parameters of interest are treated as random variables, then a Bayesian point of view provides at least one general approach to the problem. Although not too practical from a computational standpoint, this approach can give valuable insight into the fundamental limitations on how rapidly and accurately system parameters can be estimated, in situations where speed and accuracy may be critically important, and the number of unknown parameters is not too large.

Two situations are discussed in this paper—that in which the input (driving) signals of the system are observable, and that in which they are not observable, but have known statistics. An example, involving the estimation of the natural frequency of a lightly damped second-order system in each of these situations is presented, in which the "threshold effect" in accuracy versus observation time is clearly illustrated.

1. Introduction

During recent years, one trend that has become quite evident in the areas of advanced control and communication theory is the tendency of these two disciplines to grow closer together—i.e., workers in the two fields have more and more found themselves using common mathematical methods which previously were considered to belong almost exclusively to one or the other of the two fields. This is particularly evident in the application of statistical methods, where estimation and decision theory, as applied to problems of signal detection and classification by communication engineers, have more recently been applied by control engineers to problems of state estimation and system parameter estimation. The use of a Bayesian or *a posteriori* probability approach to the problem of estimating the state variables of a discrete system, with the assumption of Gaussian statistics, results in the discrete Kalman filter formulation, as has been demonstrated by Ho [1]. Applying this approach to the problem of estimating system parameters is considerably more difficult, even for Gaussian statistics, since in this case the estimation problem is usually a non-linear one, and sufficient statistics for the parameters of interest usually do not exist [2]. However, the Bayesian viewpoint does have at least two merits in considering system parameter identification problems. First, it provides a unified point of view in attacking estimation problems. This has been particularly lacking in the area of system identification, which has generally been characterized by a large number of widely diverse and specialized methods, each designed for its own special identification problem, often without any optimization basis. Secondly, it is a viewpoint that allows a study of the basic statistical limitations on system identification to be made; i.e., with certain assumed statistical signal and noise properties, it allows one to investigate the limitations on the rapidity and accuracy with which one can estimate system parameters. Because of the non-linear character of system parameter estimates based upon *a posteriori* distributions, the computer speed and capacity requirements are such that it is not very practical even with large modern computers for real, on-line system identification involving large numbers of parameters. There are situations, however, where it is critically important to be able to rapidly estimate one or two dominant parameters from noisy signals—such as, for example, the frequency of the first body-bending mode of a large flexible vehicle. In applying optimum statistical methods to system

* This work was supported by National Aeronautics and Space Administration Sustaining Grant No. NGR-34-001-005, Duke University.

parameter estimation, threshold effects, analogous to those which occur in signal detection problems, are evident. For example, if the variances of the estimates are plotted against the record length of the observed signals processed, the curves show a well defined "knee", as illustrated by the example contained in this paper, and if speed of identification is critical, it is important to know where this knee occurs. Thus, while the computational difficulties limit the practical usefulness of this approach for on-line identification, "basic-limitation" type studies can yield useful results and a better understanding of system identification from a statistical point of view.

2. Two System Models

When considering system identification problems from a statistical viewpoint, one is usually confronted with one of two possible situations, as indicated in Figure 1. In 1(a), the input signal vector is not available for observation, but its statistics are assumed known. The system, for

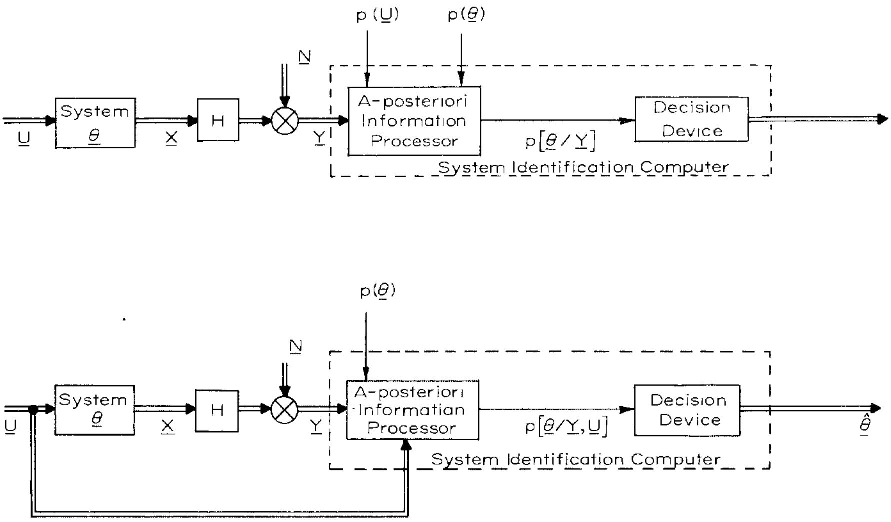


Figure 1 (a and b): Two basic system identification situations.

example, might be subject to random wind or ocean wave forces which cannot be directly measured, but about which statistical information may be available. In 1(b), it is assumed that the input signals are observable, such as would be the case, for example, if the inputs represented control surface deflections, valve openings, etc. In this case, of course, more information is available about the system, and a better job of identification can be done, both in terms of greater speed and accuracy, and in terms of less computer time required.

For the purpose of digital simulation, and also for the purpose of obtaining a statistical formulation which is based on a finite dimensional sample space, it is necessary to make use of a discrete formulation of the system time behavior. The most usual and conceptually simplest way of arriving at a discrete time model for the continuous system is to assume discrete-ized input signals as indicated in Figure 2. In a sampled data configuration, these could represent the true input signals, or in a continuous configuration they could represent approximations to the true continuous input signals. In Figure 1, the system is assumed to be describable by a set of parameters $\boldsymbol{\theta}$. The value of the state vector at the i'th sampling instant will be denoted by $\boldsymbol{x}_i$, and the $\boldsymbol{x}_i$ will be used to form a single vector $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} \triangleq \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \tag{1}$$

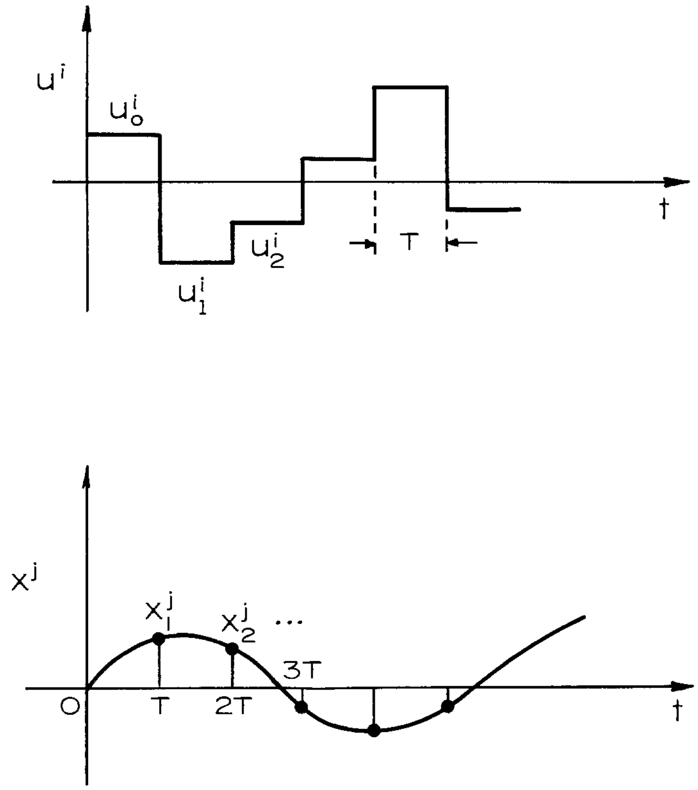


Figure 2. Typical input and sampled output signal components.

Similarly, the samples of the input, noise, and observable output vectors will be arranged as in equation (1) and will be denoted by $\boldsymbol{U}$, $\boldsymbol{N}$, and $\boldsymbol{Y}$ respectively. If the system is linear and time invariant, then in continuous time its behavior will be described by the state equations

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}\,, \qquad \boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n}\,. \tag{2}$$

Using the discrete signal model, these equations become

$$\boldsymbol{x}_{K+1} = C\boldsymbol{x}_K + D\boldsymbol{u}_K\,, \qquad \boldsymbol{y}_K = H\boldsymbol{x}_K + \boldsymbol{n}_K \tag{3}$$

where

$$C = \phi(T) = \varepsilon^{AT}\,, \qquad D = \left[\int_0^T \phi(\tau)d\tau\right] B\,. \tag{4}$$

Assuming for simplicity that the system starts with zero initial conditions at time $t=0$,

$$\boldsymbol{X} = G\boldsymbol{U}\,, \tag{5}$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ are as defined previously and G is the matrix

$$G \triangleq \begin{bmatrix} D & \phi & \phi & \phi & \dots \\ CD & D & \phi & \phi & \dots \\ C^2D & CD & D & \phi & \dots \\ \vdots & & & & \vdots \\ \vdots & \dots & \dots & \dots & D \end{bmatrix} \tag{6}$$

[ϕ denotes a null matrix].

Similarly, we may write the second equation of (3) as

$$Y = MX + N \tag{7}$$

where

$$M \triangleq \begin{bmatrix} H & \phi & \phi & \\ \phi & H & \phi & \dots \\ \phi & \phi & H & \dots \\ \vdots & & & \vdots \\ \vdots & \dots & \dots & H \end{bmatrix} \tag{8}$$

Equations (5) and (7) then provide a model upon which to base the calculation of *a posteriori* distribution densities.

3. The Identification Computer

The model of Figure 1(a) is first considered. In this model, $\boldsymbol{U}$ is not observable, but it will be assumed that the statistics of $\boldsymbol{U}$ are known. In particular, it will be assumed that the elements of both $\boldsymbol{U}$ and $\boldsymbol{N}$ have multivariate Gaussian distributions with zero means and variance-covariance matrices V_U and V_N respectively. This would correspond, in a continuous formulation, to the system being driven by a random signal vector with a Gaussian amplitude distribution and known spectral and cross-spectral densities. The variance-covariance matrix of $\boldsymbol{X}$ is then given by

$$V_X = G V_U G^T \tag{9}$$

where G^T denotes G-transpose, and if $\boldsymbol{U}$ and $\boldsymbol{N}$ are statistically independent, $\boldsymbol{Y}$ will have a multivariate Gaussian distribution with zero mean and variance-covariance matrix

$$V_Y = M V_X M^T + V_N = M G V_U G^T M^T + V_N\,. \tag{10}$$

The likelihood function is

$$p(Y|\theta) = \frac{1}{(2\pi)^{\frac{1}{2}m} |V_Y|^{\frac{1}{2}}} \exp\left\{-\tfrac{1}{2} Y^T V_Y^{-1} Y\right\} \tag{11}$$

where $|\ |$ denotes determinant and m is the order of the $\boldsymbol{Y}$ vector. Treating $\boldsymbol{\theta}$ as a random vector and applying Bayes rule yields the *a posteriori* distribution

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{p(Y|\theta)p(\theta)}{\int_\theta p(Y|\theta)p(\theta)d\theta}\,. \tag{12}$$

Two logical choices for an estimate $\hat{\theta}$ are the "most probable value", obtained by maximizing $p(\theta|Y)$ with respect to θ, or the "conditional mean value", obtained by calculating

$$E(\theta|Y) = \int_\theta \theta\, p(\theta|Y)d\theta\,. \tag{13}$$

In calculating the most probable estimate, $p(Y)$ can be treated as a constant and the distribution

$$p(\theta|Y) = (\text{const.})\, p(\theta) |V_Y|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2} Y^T V_Y^{-1} Y\right\} \tag{14}$$

may be maximized with respect to θ. Since $p(\theta|Y)$ is not Gaussian, the most probable and conditional mean estimates will in general not coincide. The most probable estimate reduces to the classical maximum-likelihood estimate if $p(\theta)$ is assumed uniform (i.e., if all values of θ are *a priori* equally likely), while the conditional mean estimate minimizes

$$E\{\|\theta - \theta\|^2 | Y\} = \int_\theta \|\theta - \theta\|^2 p(\theta|Y)d\theta\,. \tag{15}$$

Since in general the $\boldsymbol{\theta}$ parameters will appear in a highly non-linear and transcendental form in V_Y and $|V_Y|$, one must either conduct a computer search for the values which maximize the *a posteriori* distribution, or use numerical techniques to integrate equation (15). For more than a very small number of parameters, either of these procedures will be formidable, particularly if one is desirous of experimentally determining statistical properties of the estimates obtained. From a fundamental limitations point of view, however, this optimum formulation may sometimes prove valuable.

If the input signals to the system are observable, as in Figure 1(b), the expression for $p(\boldsymbol{\theta}|\boldsymbol{Y})$ is not so formidable. In this case,

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{1}{2}m}|V_N|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}(\boldsymbol{Y}-MG\boldsymbol{U})^T V_N^{-1}(\boldsymbol{Y}-MG\boldsymbol{U})\} \tag{16}$$

and

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = (\text{const.}) \exp.\{-\tfrac{1}{2}(\boldsymbol{Y}-MG\boldsymbol{U})^T V_N^{-1}(\boldsymbol{Y}-MG\boldsymbol{U})\}\,. \tag{17}$$

In this case the system parameters are contained only in G, and not in the output variance-covariance matrix, which is in this case simply V_N. This means that the amount of computation required to compute either of the estimates previously mentioned is vastly reduced, since the computer does not have to generate V_Y and $|V_Y|$ for each "trial" value of $\boldsymbol{\theta}$.

4. A Second Order Example

To illustrate a "basic limitations" study on speed versus accuracy of system identification, a second order system described by the state equations

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \omega_0^2 x_1 - 2\zeta\omega_0 x_2 + u(t) \end{aligned} \tag{18}$$

was simulated on the digital computer, and incorporated into the models of both Figure 1(a) and 1(b). It was assumed that a single parameter, namely the undamped natural frequency ω_0, was to be estimated. (For a lightly damped system, the results were found to not be very sensitive to the value chosen for the damping ratio parameter ζ.) H matrices were chosen so that either or both of the state variables x_1 and x_2 could be observed, with the same output signal-to-noise ratios chosen for both. The samples of $\boldsymbol{u}$ and $\boldsymbol{N}$* were generated from independent, "white," Gaussian random sequences with zero means. The observable output vector $\boldsymbol{Y}$ was stored after a given operation time of the system (record length), and the computer was programmed to then calculate the values of $p(\boldsymbol{\theta}|\boldsymbol{Y})$ for various assumed values of $\boldsymbol{\theta}$ (i.e., ω_0), and conduct a search for the value of ω_0 to maximize this function. All values of ω_0 were assumed *a priori* equally likely (i.e., $p(\omega_0)$ was assumed uniform within wide limits), so the estimates obtained correspond to maximum likelihood estimates. To evaluate the quality of these estimates, three hundred estimates were made for each record length and the sample mean, variance, and standard deviation computed.

The effect of sampling interval choice was first investigated by keeping the record length constant and varying the sampling interval within this record length. The curves, examplified by those of Figure 3, exhibit a sharp knee occurring when the number of samples is equal to two per natural cycle, as would be expected. For a sampling rate greater than this, the standard deviations of the estimates of ω_0 do not change appreciably, although there is a slight improvement as the sampling rate increases, due probably to the fact that the system does not have a sharply defined upper cut-off frequency. For convenience a sampling rate of 3.14 samples per natural cycle of ω_0 was chosen for the remaining tests. This is beyond the "knee" of the curves, and one is thus assured of obtaining results very close to those that would be obtained if the data could be processed in continuous rather than sampled form (with continuous input and noise signals).

* Figure 1.

With equal output signal-to-noise ratios on both of the state variables, it was found that no improvement (in fact, no appreciable difference) could be noticed in the estimates when both state variables were observed, as contrasted to when only x_1 was observed (with noise). This is reasonable, since in the absence of any noise, x_2 can presumably be determined exactly from x_1 (and known initial conditions), and contributes no additional information. In the presence of noise, presumably the noisy x_2 signal does contribute additional information along with the noisy x_1 signal (since it allows additional noise smoothing), but this is nullified by the additional noise which is presented to the processor. Hence, in subsequent tests, H was chosen

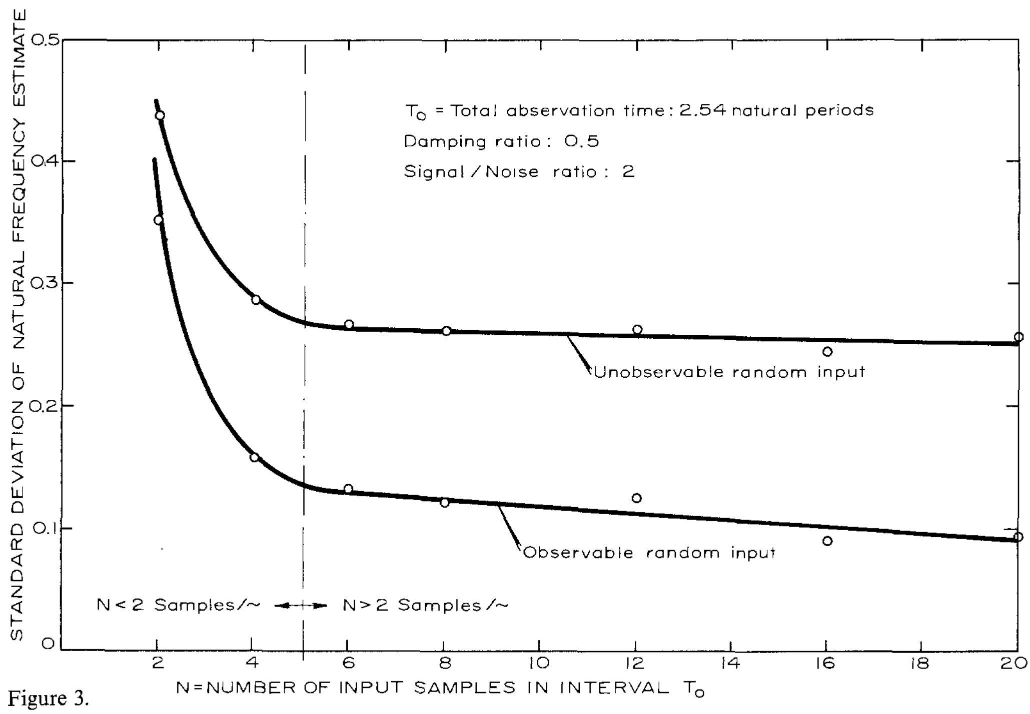
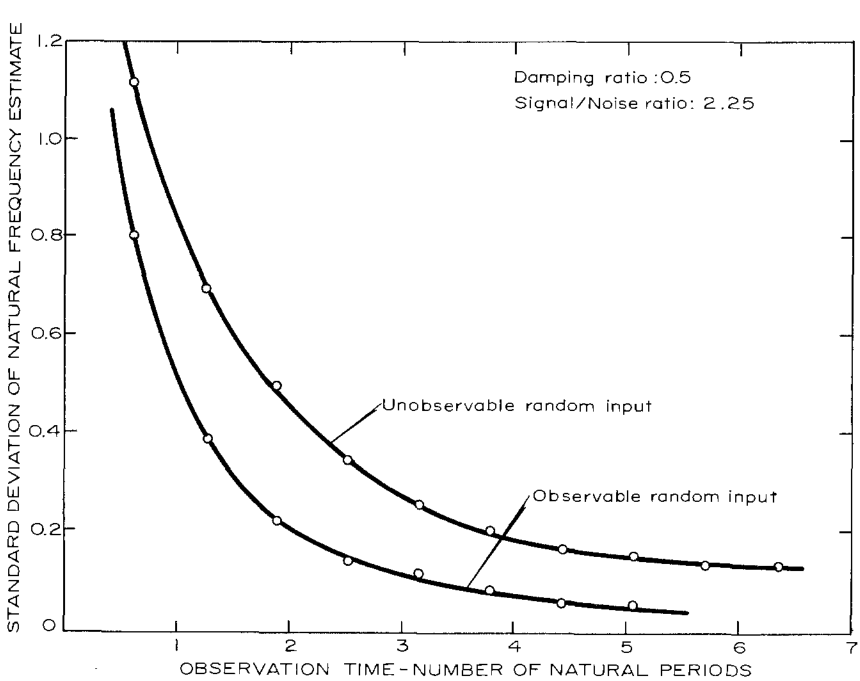
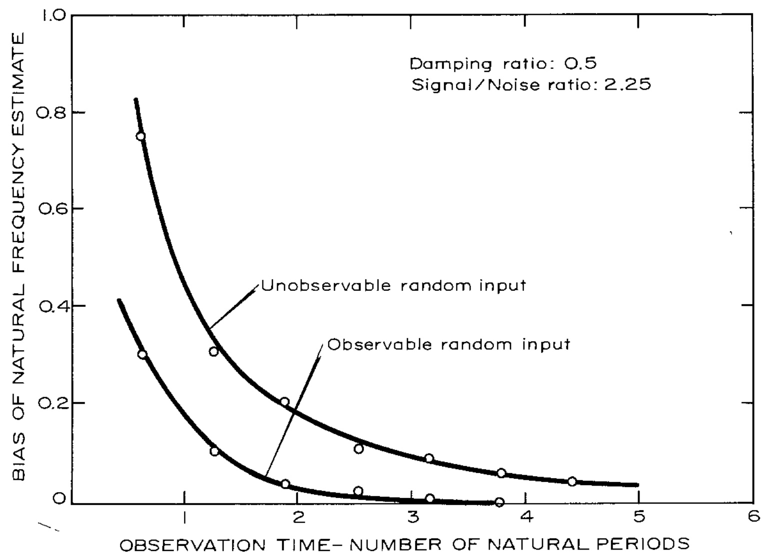
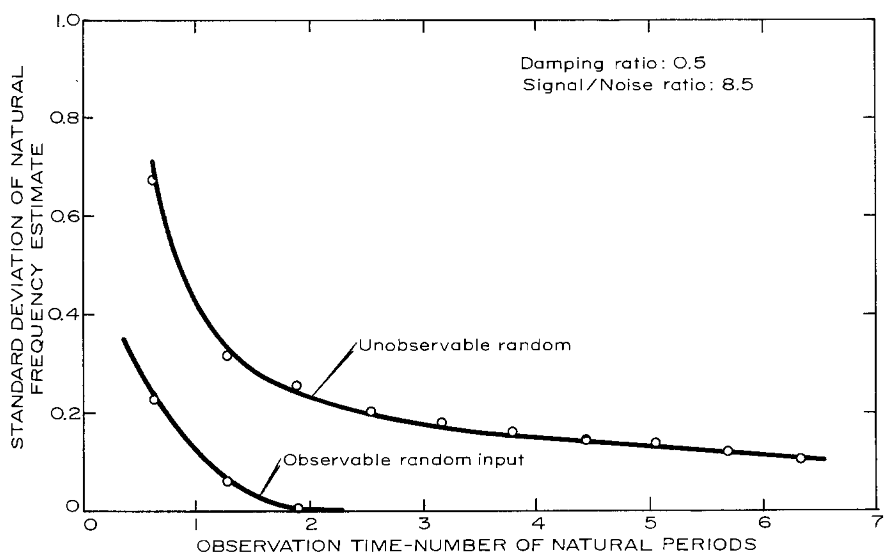


Figure 3.



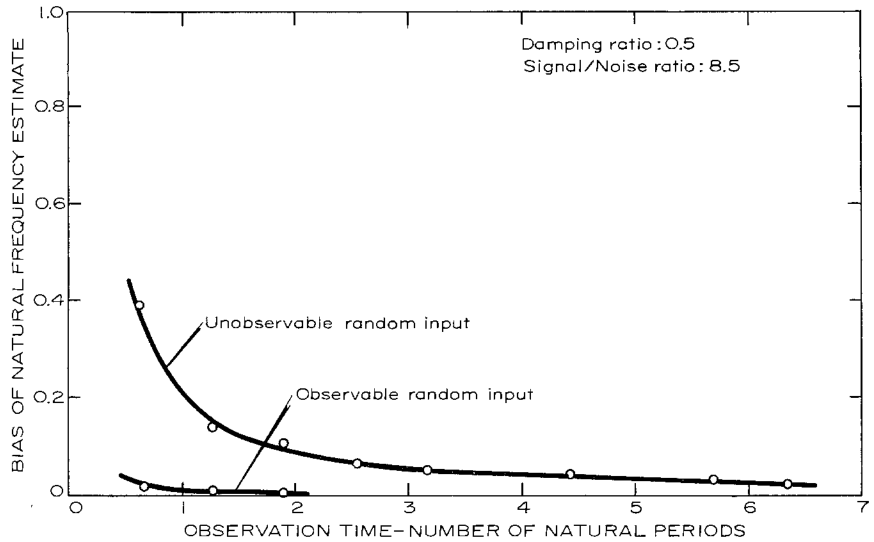


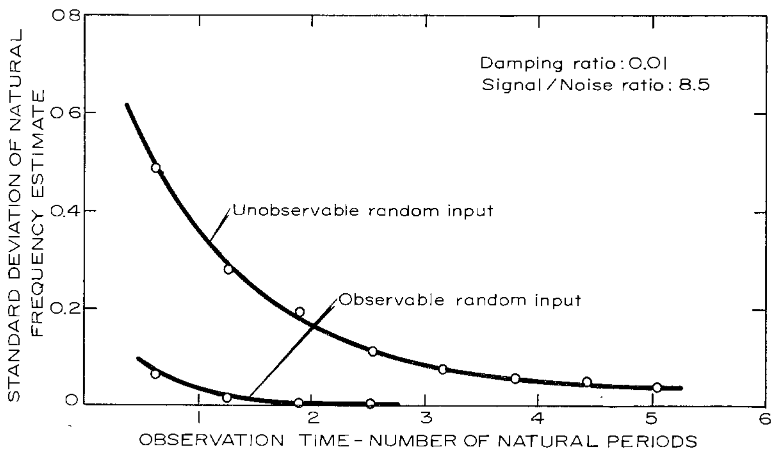


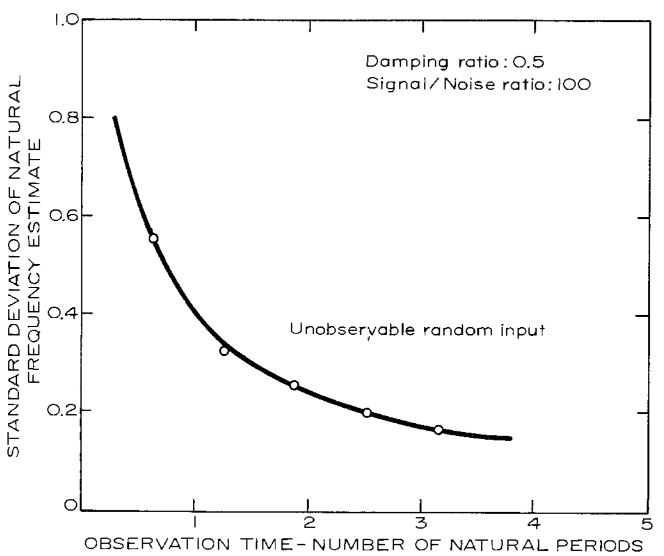
Figures 4–6. Illustration of natural frequency estimate properties for a second-order system.

so that only the noisy x_1 signal was "observed."

Figures 4–9 show typical results which were obtained for the bias and standard deviation of the natural frequency estimates, as functions of the observation time or record length (indicated by the number of natural periods). The signal-to-noise ratios indicated are power ratios in the sense that they are the ratios of output signal variance to noise variance. The "true" value of ω_0 was chosen for convenience to be unity. The curves are all of a characteristic shape, indicating that there is a fairly well defined critical record length in identifying the natural frequency. Beyond this critical record length, subsequent increasing of the observation time does not appreciably improve the quality of the estimates, unless, of course, extremely long record lengths are used.







Figures 7–9. Illustration of natural frequency estimate properties for a second-order system.

5. Summary

The Bayesian point of view provides at least one unified way of attacking system identification problems. Unlike the state estimation problem, however, the system parameter estimation problem is nearly always one of nonlinear estimation, for which computational constraints are severe. However, this approach should be considered when: (a) it is critically important to quickly and accurately determine a small number of parameters when faced with noisy signals, and (b) "basic limitation" type studies on speed and accuracy of identification are desired. With the ever-improving capabilities of digital computational equipment, such optimum statistical methods should become more and more practical in the area of system identification.

REFERENCES

[1] Yu-Chi and Lee Ho and C. K. Robert, A Bayesian Approach to Problems in Stochastic Estimation and Control, *Joint Automatic Control Conference Proceedings*, (1964) 382–387.
[2] R. B. Kerr, Statistical, Structural, and Regression Models in Linear Process Identification, *Joint Automatic Control Conference Proceedings*, (1965) 458–466.

Journal of Engineering Math., Vol. 4 (1970) 273–281