SIAM REVIEW
Vol. 31, No. 1, pp. 103–109, March 1989

004

CLASSROOM NOTES

EDITED BY MURRAY S. KLAMKIN

This section contains brief notes which are essentially self-contained applications of mathematics that can be used in the classroom. New applications are preferred, but exemplary applications not well known or readily available are accepted.

Both "modern" and "classical" applications are welcome, especially modern applications to current real world problems.

Notes should be submitted to M. S. Klamkin, Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1.

SOME RESULTS ON MAXIMUM ENTROPY DISTRIBUTIONS FOR PARAMETERS KNOWN TO LIE IN FINITE INTERVALS*

FIRDAUS E. UDWADIA†

Abstract. This paper deals with maximum entropy distributions for uncertain parameters that lie between two finite values. Such parameter uncertainties often arise in the modeling of physical systems. The paper shows that these maximally unpresumptive distributions depend on the nature of the a priori information available about the uncertain parameter. In particular, three commonly occurring situations met with in engineering systems are considered: (1) only the interval in which the uncertain parameter lies is known a priori; (2) the interval as well as the mean value of the parameter is known; (3) the interval, the mean value of the parameter, and the parameter's variance are all known. The nature of the probability distributions is determined and closed form solutions for these three situations are provided.

Key words. uncertain parameters, finite intervals, maximum entropy, a priori information, mean, variance, closed-form probability distributions

AMS(MOS) subject classification. 62C

1. Introduction. We often model a class of engineering systems through the use of generic types of mathematical models. These models often contain one or more parameter constants. When these constants are set to their proper values, the mathematical model may be thought of as representing a particular, specific system out of the class to which the mathematical model is applicable. For example, the response $x(t)$ of a one-story building subjected to a force $f(t)$ is often described by the generic model

$$mx'' + cx' + kx = f(t),$$

where m, k, and c are parameter constants. To model a particular structure from the class we must provide the value of the parameters appropriate to that specific structure.

Although it is often possible to provide the range of values in which the various parameters may lie, it is usually difficult to obtain the exact parameter values. This therefore generally leads to the use of certain nominal values and eventually causes much ad hoc hedging around the nominal analysis—a concept that has become increasingly common in most fields of engineering design and analysis.

To circumvent this difficulty it is necessary not only to specify the nominal values of the parameters, but also to admit our prior ignorance by considering the possible

* Received by the editors June 25, 1987; accepted for publication (in revised form) April 8, 1988.

† Department of Civil Engineering, Decision Systems, and Mechanical Engineering, University of Southern California, Los Angeles, California 90089-1114.

deviations of these parameters from their ascribed nominal values. This can be done by assigning a probability distribution to the parameters. However, we are seldom provided with sufficient empirical data to adduce such a probability model. We must rely on limited statistical information about the parameters and induce a probability model which is consistent with prior knowledge and which admits the greatest ignorance in matters where prior knowledge is unavailable. The following method of obtaining such a probability model maximizes our ignorance while including the available statistical database: We first define a suitable measure of information, the entropy, and then determine the probability distribution that maximizes this entropy subject to the constraints imposed by the available data. In this paper, we shall consider an uncertain parameter k which is known to lie between two **finite** values, say a and b with $b>a$. Numerous examples of such uncertain parameters are encountered in engineering and science. We provide explicit expressions for the maximally unpresumptive probability distributions for three commonly occurring cases in engineering practice. Each reflects a different amount of a priori information about the uncertain parameter. Several of these results are new and, to the best of the author's knowledge, have not appeared anywhere in the literature.

2. Probability density of *k* given information about moments of its distribution. Denoting by $p_k(k)$, the probability distribution of k, the a priori ignorance is described by the Shannon measure [1], [2]:

$$J=-\int_a^b p_k(k)\ln\{p_k(k)\}\,dk. \tag{1}$$

Often additional information about k is available in the form of moments of its distribution (e.g., the mean, variance, etc.). Thus we have $n+1$ constraints of the form

$$\int_a^b k^i p_k(k)\,dk=d_i, \qquad i=0,1,2,\cdots,n, \tag{2}$$

where the d_i, $i=1,2,\cdots,n$ are given, and $d_0=1$. Using the method of Lagrange multipliers, we consider the functional

$$\bar{J}=J+\sum_{i=0}^{i=n}\lambda_i\left[\int_a^b k^i p_k(k)\,dk-d_i\right], \tag{3}$$

where λ_i are the Lagrange multipliers, and set its variation to zero so that

$$\delta\bar{J}=\int_a^b\left[-\ln p_k(k)-1+\sum_{i=0}^{i=n}\lambda_i k^i\right]\delta p_k(k)\,dk=0. \tag{4}$$

Equation (4) now yields the density of k as

$$p_k(k)=\exp\left[-1+\sum_{i=0}^{i=n}\lambda_i k^i\right]. \tag{5}$$

The multipliers λ_i are determined from the $n+1$ equations of (2). It appears that Boekee [3] was the first to obtain this result.

We now consider three situations that commonly occur in engineering practice: (i) k is known to lie between two finite values a and b, where we shall assume that $b>a$; (ii) k is known to lie between two finite values a and b and its mean is known to be m; and (iii) k is known to lie in the finite range a to b, its mean is $(a+b)/2$, and its variance is known to be v. Rather than use the parameter k, it is convenient to use the normalized variable x, which lies in the range -1 to $+1$. Let $r=(a+b)/2$ and $s=(b-a)/2$. Then we can go from the variable x to the variable k through the usual linear transformation $k=r+sx$, and $p_k(k)=(1/s)p_x[(k-r)/s]$. We now take up the counterparts of the above-mentioned three cases in terms of our normalized variable x.

Case 1 (i). The variable x is known to lie in the range -1 to $+1$. Using (5) and noting that our a priori information regarding x corresponds to the situation where $n=0$ in (2), we find that the maximally unpresumptive density of x is simply a constant. Noting that the area under the density curve is unity we have

$$p_x(x)=\begin{cases}\frac{1}{2} & \text{for } -1<x<1,\\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

Thus the distribution of our original variable k is uniform between a and b. The mean of the distribution is zero and its variance is $(b-a)^2/12$.

Case 2 (ii). The variable x is known to lie between -1 and 1, and its mean is μ. If we use (5) with $n=1$, the density of x becomes

$$p_x(x)=\begin{cases}C\exp(\lambda x) & \text{for } -1<x<1,\\ 0 & \text{otherwise,}\end{cases} \tag{7}$$

where C is a positive constant. Using the relations

$$\int_{-1}^{1} p_x(x)\,dx=1 \quad \text{and} \quad \int_{-1}^{1} xp_x(x)\,dx=\mu, \tag{8}$$

we obtain

$$C=\frac{\lambda}{2\sinh\lambda}, \tag{9a}$$

$$\tanh\lambda=\frac{\lambda}{(\mu\lambda+1)}. \tag{9b}$$

From relation (9b) we see that when $\mu\to 0$, $\lambda\to 0$ and we obtain a uniform distribution identical to that given by (6). Also, for $\mu\to\pm 1$, $\lambda\to\pm\infty$. Furthermore, λ is an odd function of μ. For other values of λ, the corresponding values of μ can be found by inverting (9b) to read

$$\mu=\frac{\lambda-\tanh\lambda}{\lambda\tanh\lambda}. \tag{9c}$$

Figure 1(a) shows λ as a function of μ numerically calculated on a pocket calculator. The corresponding probability densities of x for positive values of μ are shown in Fig. 1(b). The densities for negative μ values are obtained by reflecting the densities for the corresponding positive μ values in the y-axis. As $\mu\to\pm 1$, the densities tend toward delta distributions.

Case 3 (iii). The variable x is known to lie between -1 and $+1$, its mean is zero, and its variance is σ^2. Using (5) with $n=2$, we have

$$p_x(x)=\begin{cases}D\exp[\lambda_1 x+\lambda_2 x^2], & -1<x<1,\\ 0 & \text{otherwise,}\end{cases} \tag{10}$$

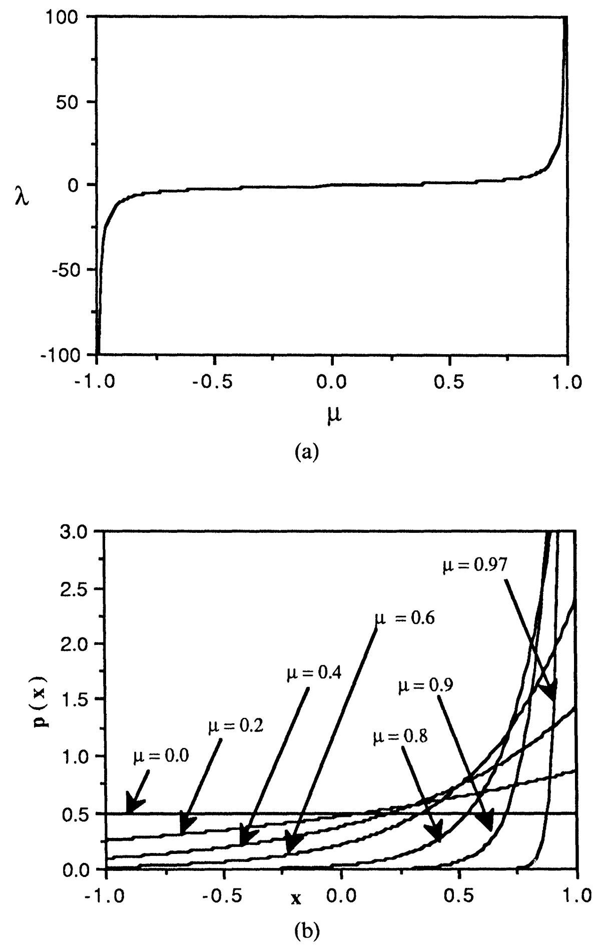

FIG. 1

where D is a positive constant. Consider the distribution given by

$$p_x(x) = \begin{cases} D \exp[\lambda x^2], & -1 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Since it is an even function of x, its mean is zero. Furthermore, the relations

$$\int_{-1}^{1} p_x(x)\,dx = 1 \quad \text{and} \quad \int_{-1}^{1} x^2 p_x(x)\,dx = \sigma^2 \tag{12}$$

require

$$I := D^{-1} = \int_{-1}^{1} \exp[\lambda x^2]\,dx, \tag{13}$$

$$\sigma^2 = \int_{-1}^{1} x^2 D \exp\{\lambda x^2\}\,dx. \tag{14}$$

Equation (14) can now be expressed using (13) as follows:

$$\frac{dI}{d\lambda} = \sigma^2 I, \qquad I(0) = 2. \tag{15}$$

Solving (15) we get

$$I := \int_{-1}^{1} \exp[\lambda x^2]\, dx = 2 \exp[\sigma^2 \lambda], \tag{16}$$

from which we get

$$\sigma^2 = \left\{ \ln\left[\frac{1}{2}\int_{-1}^{1} \exp[\lambda x^2]\, dx\right]\right\} \Big/ \lambda. \tag{17}$$

We now have two results regarding the relation between σ and λ, which will be used in assessing the nature of the distribution given by (11).

RESULT 1. *The parameters* λ *and* σ *are related such that*

(a) *When* $\lambda \to 0$, $\sigma^2 \to \frac{1}{3}$;

(b) *When* $\lambda \to -\infty$, $\sigma^2 \to 0$.

Proof. Expanding $\exp[\lambda x^2]$ on the right-hand side of (17) for small values of λ, integrating, and again expanding the logarithm, Result 1(a) follows.

To prove Result 1(b), write $\lambda = -1/(2\alpha^2)$, and note that

$$\sigma^2 = \frac{(1/\alpha\sqrt{2\pi}) \int_{-1}^{1} x^2 \exp[-x^2/(2\alpha^2)]\, dx}{(1/\alpha\sqrt{2\pi}) \int_{-1}^{1} \exp[-x^2/(2\alpha^2)]}. \tag{18}$$

For $\lambda \to -\infty$, which implies $\alpha = \varepsilon \to 0$, we can approximate the integrals using the properties of Gaussian distributions so that

$$\frac{1}{\alpha\sqrt{2\pi}} \int_{-1}^{1} x^2 \exp\left[-\frac{x^2}{2\alpha^2}\right] dx \approx \frac{1}{\alpha\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left[-\frac{x^2}{2\alpha^2}\right] dx = \varepsilon^2, \tag{19}$$

$$\frac{1}{\alpha\sqrt{2\pi}} \int_{-1}^{1} \exp\left[-\frac{x^2}{2\alpha^2}\right] dx \approx \frac{1}{\alpha\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2\alpha^2}\right] dx = 1. \tag{20}$$

When we use (19) and (20) in (18), the result follows. □

RESULT 2. σ^2 *is an increasing function of* λ.

Proof. Differentiating (18) after replacing $-1/(2\alpha^2)$ by λ, we get

$$\frac{d\sigma^2}{d\lambda} = \frac{\int_{-1}^{1} \exp[\lambda x^2]\, dx \int_{-1}^{1} x^4 \exp[\lambda x^2]\, dx - \int_{-1}^{1} x^2 \exp[\lambda x^2]\, dx \int_{-1}^{1} x^2 \exp[\lambda x^2]\, dx}{[\int_{-1}^{1} \exp[\lambda x^2]\, dx]^2}. \tag{21}$$

Invoking the Schwartz–Buniakowsky inequality for the numerator, we get

$$\frac{d\sigma^2}{d\lambda} \geqq 0. \tag{22}$$

Noting that for finite λ the equality cannot occur, the result follows. □

From (17), the numerically determined plot for $\sigma = f(\lambda)$ is shown in Fig. 2(a). We note that for a given σ, once λ is determined, the constant D is obtained from the relation

$$D = I^{-1} = \frac{1}{2} \exp[-\sigma^2 \lambda] \tag{23}$$

so that we can obtain the two parameters D and λ that characterize the distribution given by (11).

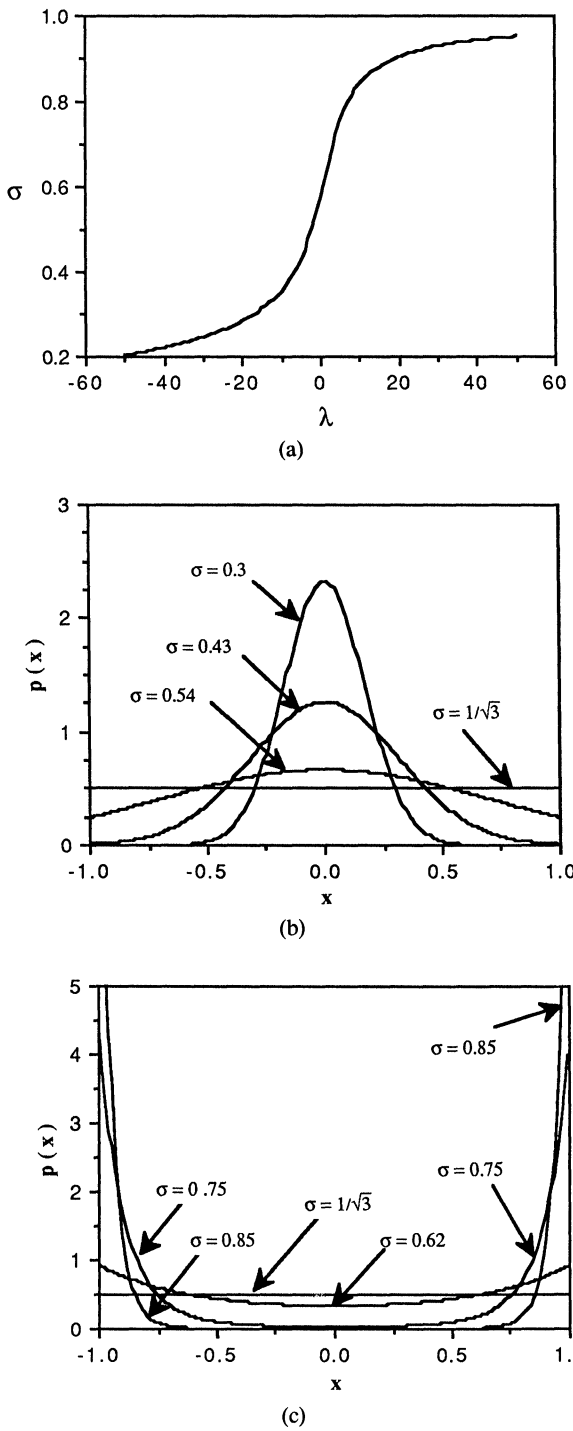

FIG. 2

Since σ^2 is an increasing function of λ, using Result 1 we find that for $0 < \sigma < 1/\sqrt{3}$ the value of λ is always negative and the resulting probability density given by (11) is a truncated Gaussian distribution. Figure 2(b) shows the probability distributions of x for various values of σ in this range. These distributions are determined using the plot of Fig. 2(a) and (23). At $\sigma = 1/\sqrt{3}$, the value of λ is zero, and the distribution becomes a uniform distribution over the range -1 to 1 (see Fig. 2(b)). For $1/\sqrt{3} < \sigma < 1$, λ is positive and the distribution given by (11) has a positive exponent. Distributions for various σ values in this range are shown in Fig. 2(c). We note that the variance σ cannot exceed unity (see Fig. 2(a)). This situation arises when we have two probability masses (delta distributions) located at -1 and $+1$, each of magnitude (strength) $\frac{1}{2}$. For values of σ close to unity, Fig. 2(c) shows that the probability distributions move toward this limit.

3. Conclusions. (1) When an uncertain parameter k is known to lie within a finite interval (a, b) the maximally unpresumptive distribution consistent with the data is a uniform distribution over the interval.

(2) When an uncertain parameter is known to lie within a finite interval (a, b) and its mean m is known, the maximally unpresumptive distribution consistent with the data is an exponential distribution. In particular, when $m = (a + b)/2$, the distribution reverts to a uniform distribution over the interval.

(3) When an uncertain parameter is known to lie within a finite interval (a, b), its mean m is known to equal $(a + b)/2$, and its variance v is given, the maximally unpresumptive distribution is symmetric about the mean value. As long as the prescribed variance v is less than that of a uniform distribution over the same interval, i.e., when

$$v < v_0,$$

where

$$v_0 = \frac{1}{3}\left[\frac{b-a}{2}\right]^2, \tag{24}$$

the maximum entropy distribution of k is a truncated Gaussian distribution. When $v = v_0$, the maximum entropy distribution reverts to a uniform distribution over the interval. For values of v greater than v_0, the distribution is of the form $\exp[\lambda x^2]$, where λ is a positive number. As v increases beyond v_0, the probability area gets increasingly concentrated away from the mean value and toward the ends of the interval. This culminates in two delta distributions, each centered at the ends of the interval with a corresponding maximum variance of $3v_0$.

REFERENCES

[1] S. Watanabe, *Knowing and Guessing*, John Wiley, New York, 1969.

[2] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949.

[3] D. E. Boekee, *Maximum information in continuous systems with constraints*, in Eighth International Congress on Cybernetics, AIC, Namur, Belgium, pp. 243–259.