REMARKS ON UNCERTAINTY ASSESSMENT AND MANAGEMENT IN MODELING AND COMPUTATION *

H.T. Banks
Center for Research in Scientific Computation
Box 8205
North Carolina State University
Raleigh, NC 27695-8205
Fax: 919 515-1636
email: htbanks@eos.ncsu.edu

ABSTRACT

We discuss questions related to uncertainty in scientific computations for mathematical models. A computationally tractable probabilistic framework to treat uncertainty in the estimation of parameters or inverse problems is given. The theory is illustrated by a simple computational example for the estimation of constant parameters in differential equations by treating the parameters as random variables.

Keywords: Computational uncertainty, inverse problems, probability distributions

*Invited lecture, Conference on Computation and Control V, Montana State University, Bozeman, MT, August 4-7, 1998. Research supported in part by the Air Force Office of Scientific Research under grants AFOSR F49620-95-1-0447 and F49620-98-1-0180.

1 Introduction

The remarks on uncertainty offered here were stimulated by a special session at the SIAM National Meeting in Toronto in July, 1998. In particular, presentations by Mac Hyman and Linda Petzold prompted the simple but hopefully useful observations we make. The general topic of "quantifiying uncertainty" and then minimizing or controlling it was the focus of a SIAM/NSF workshop in Washington in June 1998 and discussions with participants of that workshop also influenced us to think about the topic.

Uncertainty (recognized or not) is present in much of what we do as applied mathematicians. It arises in our approximations to physical and biological systems when we write dynamical or static equations and constraints. Our models may be compartmental or distributed and often involve reduction in order and subsequent discretization either at the finite or infinite dimensional level. Once we have models for the processes of interest, we often introduce computational uncertainty through forward simulations, inverse or parameter estimation procedures, and sensitivity and robustness analyses (as in control design).

Most often our efforts do not include discussions of these uncertainties. Indeed there is a general absence of accepted scientific frameworks for quantifying and controlling uncertainties from underlying sources of error in estimates, computations and analysis.

Seldom are we able to treat with care the relationship between levels of uncertainty and the nondeterministic nature of currently popular "softer" sciences involving biological, sociological, economic and demographic (especially population behavior) studies. There are an overwhelming number of related questions. What do we mean by a "good model"? How do we know when we have one? How do we assess accuracy in fitting of data to competing models? How much weight do we give to predictive capability of models versus their preservation of certain qualitative, quantitative or asymptotic behaviors? In regard to forward simulations or inverse problem estimates, how do we attach "goodness" or validity to numbers that we produce from computations? More specifically, how do we introduce nondeterministic features in our models and calcualtions such as numerical simulations or parameter estimation algorithms?

Immediate possibilities come to mind in response to such philosophical cogitations. Among the more traditional approaches are the following:

1) Use of *stochastic* ordinary or partial differential equation models. These models may involve simple additive "noise" terms and random initial values or they may include a much more comprehensive treatment of stochasticity in coefficients (parameterizations), nonlinearities, boundary formulations, constraints, etc.

2) Use of "error bars" as uncertainty bounds in simulation results. These error bars may be in recognition of modeling error including parameterization error, initial data error, computational (formula and/or machine)

error, etc. This approach could encompass the reporting of computational results in a manner similar to that used by experimentalists with data, requiring scientists to discuss the variability and uncertainty associated with a particular set of numbers generated by simulations.

3) Use of uncertainty or error bars in parameter estimates, optimal controls, sensitivity and robustness bounds. This is essentially the same approach as in 2) extended to a much larger range of computational questions. For example, consider a parameter estimation problem involving a best estimate q^* to be chosen from a set Q of admissable values using data $\hat{x}_i$ for $x(t_i)$ where $\dot{x}(t) = f(t, x(t), q)$. As an alternative, one might calculate not q^*, but a range of values in $[\tilde{q} - \delta, \tilde{q} + \delta]$ where δ embodies probablistic information and $\tilde{q}$ approximates q^*.

To continue with the rather concrete items mentioned above, we note that while 1)–3) above each include stochasticity or uncertainty, 1) versus 2),3) entail very different paradigms for inclusion of uncertainty. The stochastic differential equation approach which leads to a stochastic calculus, for example, for equations of the form $dx(t) = f(t, x(t), q)dt + dw(t)$, has resulted over the past decades in a tremendous amount of mathematical literature and numerous theoretical tools. However, there are a number of limitations in this approach. In particular, one generally requires very specific types of noise (additive, initial data, white noise, etc.) to obtain a rigorous theory. Stochastic parameters (such as rate constants, delays, nonlinearities) not in a special class are generally not amenable to a theoretical treatment. Overall the theoretical frameworks developed have not had the desired impact on practical aspects of uncertainty in modeling and computation. The research literature (applied as well as theoretical) also contains many specialized stochastic differential equation simulation methods, but ensembles of solutions are often computationally expensive for systems of interest in applications.

With respect to 2) and 3), the nearest semblance to an "error bar" theory presently available are the *a priori* error bounds that can be derived for many numerical approximation schemes. However, these *a priori* bounds are often not available, not used properly, or simply not very helpful in addressing questions related to uncertainty in numerical computations. As we shall see in the next section, there are some significant mathematical foundations in probability theory that can be used in inverse problem methods. But to date we have not made very good use of them to develop practical uncertainty assessment tools.

Happily, there is a *"bad news, good news"* scenario in all of this. To date, there is a dearth of good practical tools available for the assessment and control of uncertainty in the computational sciences. The good news resides in the potential for significant contributions in such a wide open field — there are reputations (and maybe money too!) to be made by those who can contribute even simple, sound mathematical theories resulting in practical, user friendly tools.

In the next section we will outline an approach to inverse or parameter estimation problems that relies heavily on known results in probability theory. As we shall see, this approach allows one to attach quantitative measures of variability to estimated parameters by viewing them as random variables.

2 Estimation of Parameters in Dynamical Systems

To illustrate our ideas, we consider here the estimation of constant parameters in a system of ordinary differential equations. The ideas are readily extended to partial differential equations with unknown functional (e.g., spatially and/or time dependent coefficients) parameters as well as to many other systems of interest in applications.

A typical estimation problem consists of using observations $\{\hat{x}\}_{i=1}^{n}$ for $x(t_i)$, $i = 1, 2, \ldots, n$, to estimate parameters $q \in R^m$ in the vector dynamical system

$$\dot{x}(t) = f(t, x(t), q). \tag{1}$$

This is often done using a least squares formulation to find a best parameter value q^* in some admissible parameter set $Q \subset R^m$. Thus, we seek to find $q^* \in Q$ that provides a minimum for

$$J(q) = \sum_{i=1}^{n} |x(t_i; q) - \hat{x}_i|^2 \tag{2}$$

where $x(t; q)$ is the solution of (1) for a given $q \in Q$.

To introduce uncertainty into this deterministic process, we treat the parameters as realizations for a random variable and use the data to estimate the *probability distribution function* (PDF) for this random variable. More precisely, let $\mathcal{P}(Q)$ denote the linear space of probablilty distributions on Q and treat the data $\{\hat{x}_i\}$ as observations for the expected value

$$\mathcal{E}\left[x(t_i; q)|P\right] = \int_Q x(t_i; q) dP(q) \tag{3}$$

for a given PDF $P \in \mathcal{P}(Q)$. If P is a discrete PDF with atoms $\{q_j\} \subset Q$ and probabilities $\{p_j\}, p_j \geq 0, \sum p_j = 1$, then (3) can be written

$$\int_Q x(t_i; q) dP(q) = \sum_j x(t_i; q_j) p_j. \tag{4}$$

In any case, the least squares estimation problem becomes: Find $P^* \in \mathcal{P}(Q)$ to minimize

$$J(P) = \sum_{i=1}^{n} |\mathcal{E}\left[x(t_i;q)|P\right] - \hat{x}_i|^2 \tag{5}$$

over $P \in \mathcal{P}(Q)$. For this problem to be tractable, it would be most helpful to have a topology on $\mathcal{P}(Q)$, continuity of the map $P \to J(P)$ in this topology, compatible compactness results, and some approximation results leading to implementable computational algorithms. Fortunately, probability theory offers a great start toward a possible complete and tractable computational methodology! We summarize useful results from Billingsly [1].

We begin with a metric, the *Prohorov metric*, for $\mathcal{P}(Q)$, the set of probability measures on the Borel subsets of Q, where Q can be any complete metric space with metric d. For any closed subset $F \subset Q$ and $\varepsilon > 0$, we define an $\varepsilon-$ neighborhood of F by

$$F^{\varepsilon} \equiv \{q \in Q \,|d(\tilde{q}, q) < \varepsilon, \tilde{q} \in F\}\,.$$

We then define $\rho : \mathcal{P}(Q) \times \mathcal{P}(Q) \to R_+$ by

$$\rho(P_1, P_2) \equiv \inf\{\varepsilon > 0 \,|P_1[F] \leq P_2[F^{\varepsilon}] + \varepsilon,\ F \text{ closed} \subset Q\}\,.$$

The following results are well-known:

(i) ρ is a metric (called the Prohorov metric) on $\mathcal{P}(Q)$

(ii) $(\mathcal{P}(Q), \rho)$ is a complete metric space;

(iii) if Q is compact, then $(\mathcal{P}(Q), \rho)$ is a compact metric space.

To develop any approximation ideas, we need to understand the meaning of the convergence $P_k \to P$ in the ρ metric. The definition of the Prohorov metric is neither intuitive nor easily used directly. Again, we are fortunate to have the following theorem on equivalent formulations.

Theorem 1 *Suppose (Q, d) is a complete metric space and $(\mathcal{P}(Q), \rho)$ is defined as above. Then for P_k, P in $\mathcal{P}(Q)$, the following are equivalent:*

(i.) $\rho(P_k, P) \to 0$*;*

(ii.) $\int_Q f dP_k(q) \to \int_Q f dP(q)$ *for all bounded, uniformly continuous $f : Q \to R^1$;*

(iii.) $P_k[A] \to P[A]$ *for all Borel sets $A \subset Q$ with $P[\partial A] = 0$, where ∂A denotes the boundary of A.*

Thus we see that convergence in the ρ metric is equivalent to convergence in distribution. Or, if we consider $\mathcal{P}(Q) \subset C_B(Q)^*$, where $C_B(Q)$ denotes the space of bounded, continuous functions on Q with supremum norm, then convergence in the ρ topology is equivalent to weak* convergence in $\mathcal{P}(Q)$. More importantly to us for our discussions here, $\rho(P_k, P) \to 0$ is equivalent to

$$\int_Q x(t;q) dP_k(q) \to \int_Q x(t;q) dP(q),$$

or, using standard notation from probability theory,

$$\mathcal{E}[x(t;q) \,|P_k] \to \mathcal{E}[x(t;q)| \, P].$$

This is precisely the convergence one needs to establish continuity in the ρ topology of the map

$$P \to J(P) = \sum_{i=1}^{n} |\mathcal{E}[x(t_i;q) \,|P] - \hat{x}_i \,|^2 .$$

Continuity of this map along with the compactness of Q, which in turn guarantees compactness of $\mathcal{P}(Q)$ in the ρ metric, is then sufficient to establish existence of solutions to the problem of minimizing J of (5) over $\mathcal{P}(Q)$.

Assuming that existence questions are dealt with, we note that $\mathcal{P}(Q)$ with the ρ metric is generally an infinite dimensional space. Thus, to treat computational issues one must consider approximations ideas. Once again, probability theory provides the needed results.

Theorem 2 *Suppose that Q is a complete separable metric space and let $Q_r = \{q_j\}_{j=1}^{\infty}$ be a countable dense subset of Q. Then the set of $P \in \mathcal{P}(Q)$ such that P has finite support in Q_r is dense in $\mathcal{P}(Q)$ in the ρ metric. That is,*

$$\mathcal{P}_r(Q) \equiv \{P = \textstyle\sum_{j=1}^{l} p_j \delta_{q_j} \,\Big|\, p_j \geq 0, \sum_{j=1}^{l} p_j = 1, l \in N^+, q_j \in Q_r \}$$

is dense in $\mathcal{P}(Q)$. Here δ_{q_j} is the Dirac measure with atom at q_j and N^+ are the positive integers.

For a given positive integer M and Q_r as in the theorem, define

$$\mathcal{P}^M = \{P \in \mathcal{P}_r(Q) \,\Big|\, P = \textstyle\sum_{j=1}^{M} p_j \delta_{q_j} \}. \tag{6}$$

Then $\mathcal{P}^M$ is a compact subset of $\mathcal{P}(Q)$ with $\mathcal{P}^M \subset \mathcal{P}^{M+1}$ and $\cup_{M=1}^{\infty} \mathcal{P}^M = \mathcal{P}_r(Q)$. Moreover, the denseness of $\mathcal{P}_r(Q)$ in $\mathcal{P}(Q)$ allows us to approximate any element $P \in \mathcal{P}(Q)$ by a sequence $\{P_M\}$, $P_M \in \mathcal{P}^M$, such that $\rho(P_M, P) \to 0$ as $M \to \infty$.

To define a family of approximate minimization problems, we fix $\{q_1, q_2, \ldots, q_M\}$ in Q_r with associated $\mathcal{P}^M$ defined by (6). Then for $P_M = \sum_{j=1}^{M} p_j \delta_{q_j} \in \mathcal{P}^M$, the minimization criterion (5) reduces to

$$J(P_M) = \sum_{i=1}^{n} \left| \sum_{j=1}^{M} x(t_i; q_j) p_j - \hat{x}_i \right|^2 \tag{7}$$

and our approximate estimation problem becomes one of minimizing $J(P_M)$ of (7) over $P_M \in \mathcal{P}^M$. We observe that this problem is computationally most attractive. Indeed it is a *constrained quadratic programming problem.* Letting $\bar{p} = (p_1, \ldots, p_M) \in R^M$ with $p_j \geq 0, \sum_1^M p_j = 1$, we see that minimizing (7) is equivalent to minimizing

$$\begin{aligned} \mathcal{J}(\bar{p}) &= \bar{p} \cdot A\bar{p} + 2\bar{p} \cdot \bar{b} + c \qquad (8) \\ \text{where} & \\ A_{kj} &= \sum_{i=1}^{n} x(t_i; q_k) x(t_i; q_j) \quad k, j = 1, 2, \ldots, M \\ b_j &= -\sum_{i=1}^{n} \hat{x}_i x(t_i; q_j) \quad j = 1, 2, \ldots, M \\ c &= \sum_{i=1}^{n} \hat{x}_i^2. \end{aligned}$$

That is, we are reduced to seeking a minimizer $\bar{p}^*$ for $\mathcal{J}(\bar{p})$ subject to $\bar{p} \cdot \bar{1}_j \geq 0, j = 1, 2, \ldots, M$, and $\bar{p} \cdot \bar{1} = 1$. Here $\bar{1}_j$ is the M vector with zero components except for a 1 in the jth component and $\bar{1} = (1, 1, \ldots, 1) \in R^M$.

For this constrained optimization problem there are many suitable techniques, e.g., Lagrange multiplier methods, etc. Indeed, MATLAB has routines that render this problem quickly solvable and we shall illustrate this by example below.

We remark that any solution $\bar{p}^*$ must satisfy

$$A\bar{p}^* = -\bar{b}$$

and if A is nonsingular, $\bar{p}^*$ is given uniquely and depends continuously on $-\bar{b}$ and hence continuously on $\{\hat{x}_i\}_{i=1}^n$. This is precisely the requirements for the finite dimensional problems to each be *well-posed* or *stable* in *inverse problem* terminology (i.e., existence, uniqueness, and continuous dependence on data for $\bar{p}^*$).

3 Example

We present a very simple example to illustrate a computational algorithm resulting from theoretical discussions of the previous section. The particular system

we use is motivated by a problem in the assessment of the efficiency of a vaccination program [2, 3] by using data $\{\hat{x}_i\}$ for the aggregate population $x(t_i)$ of vaccinated but not infected individuals at time t_i. The evolution of population is given by

$$\dot{x}(t) = -qG(t)x(t), \quad x(0) = x_0, \tag{9}$$

where G(t) represents the known rate of exposure to infection, q is the susceptibility to "environmental exposure" subsequent to vaccination of the population at time $t = 0$, and x_0 is the known number of individuals initially vaccinated. The best susceptibility parameter q is to be chosen from $Q = [0, 1]$ and in the spirit of the presentation of the previous section, we seek to identify a probability distribution $P(q) \in \mathcal{P}(Q)$ using the data $\{\hat{x}_i\}$ as the expected number of vaccinated but not yet infected individuals for a given distribution.

All calculations were carried out using MATLAB routines. "Data" $\{\hat{x}_i\}$ was prepared by solving (9) with a specific value of $q = q^*$ (= .5 in the example presented here) and $G(t)$ given by

$$G(t) = \begin{cases} 0 & 0 \le t \le .195 \\ 100(t - .195) & .195 \le t \le .205 \\ 1 & .205 \le t. \end{cases}$$

Relative random noise was added to these solutions so that the "data" was given by $\hat{x}_i = x(t_i; q^*)[1+\varepsilon_i]$ where the ε_i are independent Gaussian random variables with mean zero and variance σ^2 ($\sigma = .005$ was used in the computations detailed below).

To discretize Q in formulating the approximation class given in (6), for a given M we used a uniform partition of Q defined by $\{q_1, \ldots, q_M\}$ where $q_1 = 0, q_M = 1$ and $q_{j+1} - q_j = \frac{1}{M-1}$. The resulting constrained problem using (8) was solved using the `qp` (constrained quadratic programming) algorithm in MATLAB. Since the system (9) was particularly simple in this case, the values $x(t_i; q^*)$ were obtained analytically.

We present in the figures below the estimated discrete probability *densities* represented by $\bar{p} = (p_1, \ldots, p_M)$ and the corresponding discontinuous *distributions* $P_M = \sum_{j=1}^{M} p_j \delta q_j$ for several values of M. We note that in general we observe convergence of the probablity distributions in the Prohorov metric (as guaranteed by the theory of the previous section), while the discrete densities do not in general converge in any meaningful sense on $Q = [0, 1]$. In this example the true probability distribution P^* is given by $P^*[A] = 0$ if $.5 \notin A, P^*[A] = 1$ if $.5 \in A$. The computational algorithm performs similarly in cases where the underlying distribution is a continuous distribution with a corresponding probability density function (e.g., see [2]).

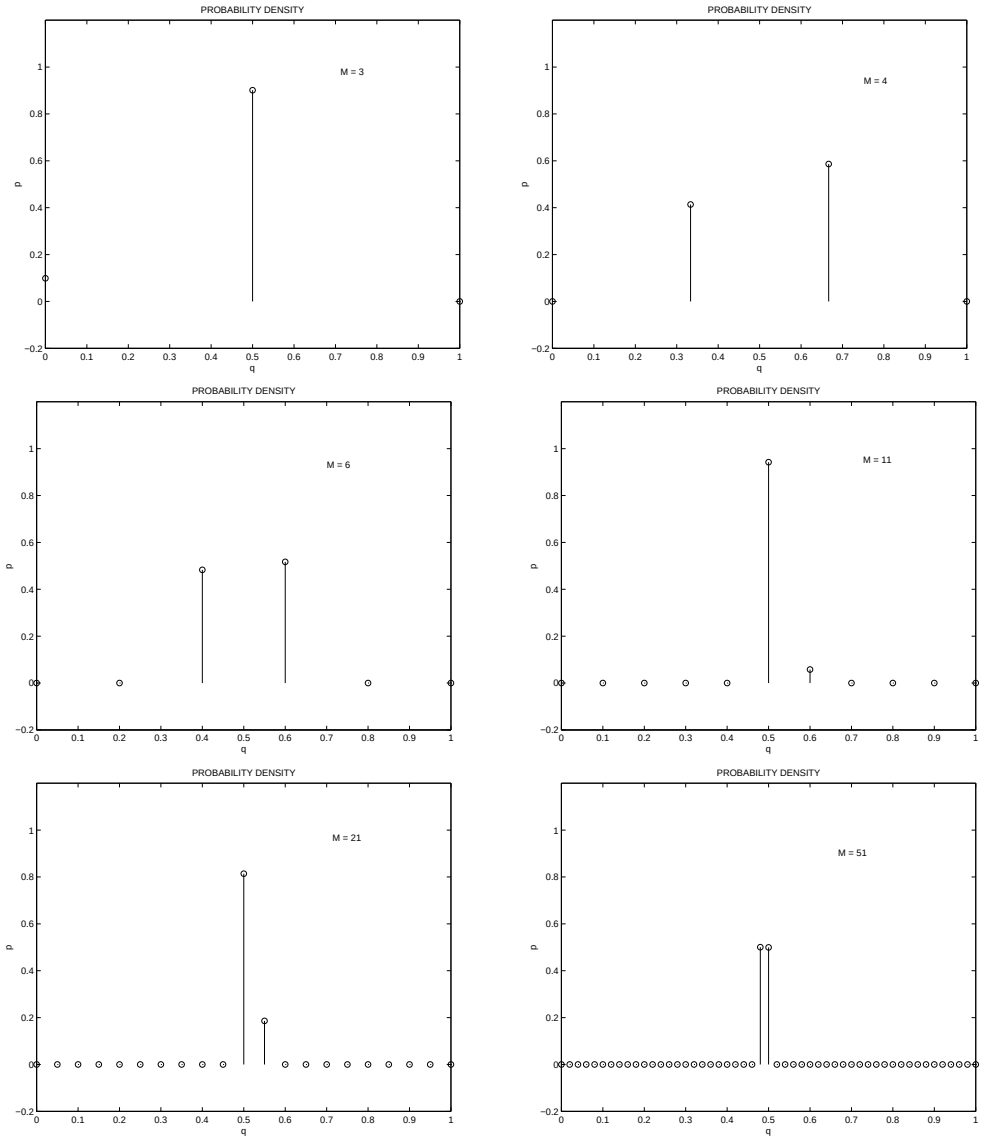

Figure 1. Approximate probability densities for a sequence of M values.

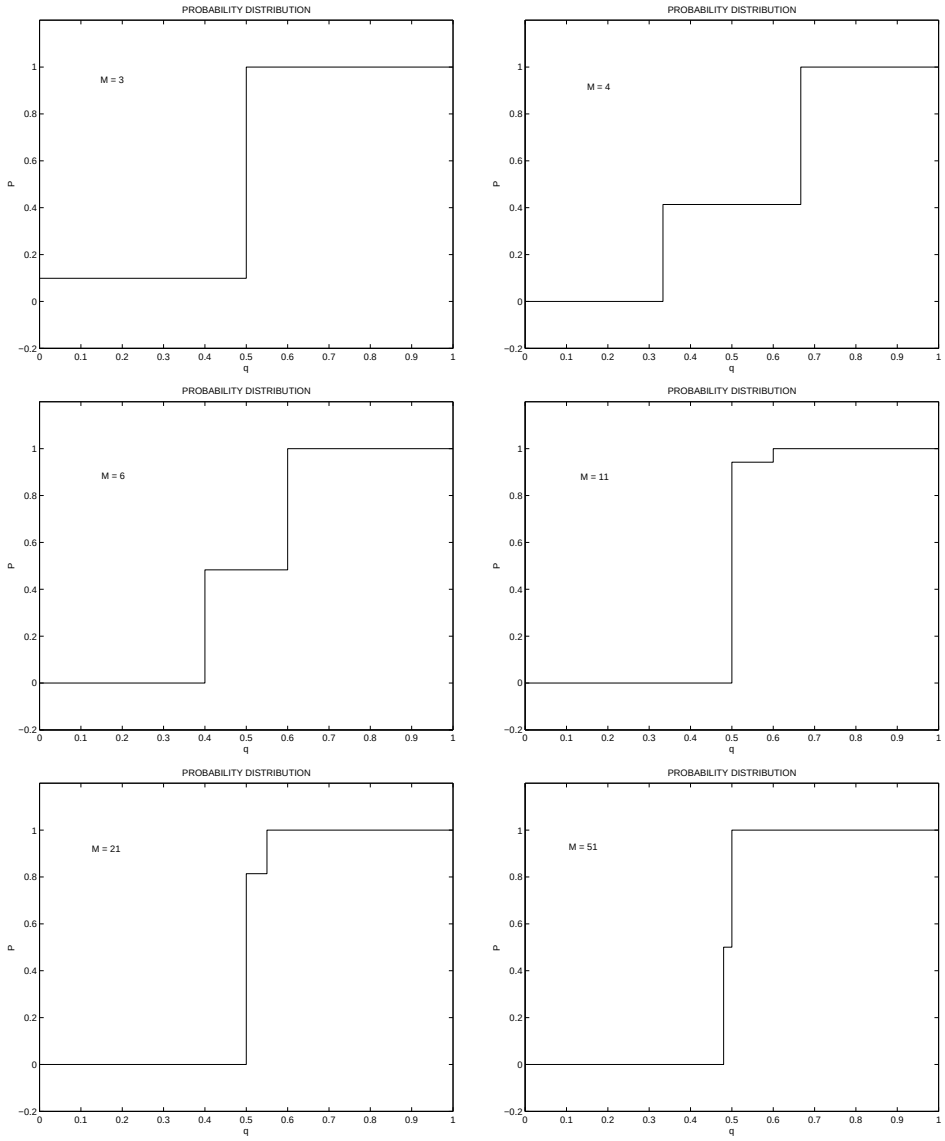

Figure 2. Approximate probability distributions for a sequence of M values.

4 Concluding Remarks

The probablistic ideas presented in Section 2 above have been used in a fundamental way in a diverse number of applications. They are the foundation of approximation ideas developed in the 1950's–1960's in control theory (e.g., chattering controls, Young's measures, sliding regimes, relaxed curves, etc. in the classical works of L.C. Young, E.J. McShane, A.F. Filippov and J. Warga between 1937 and 1970 — see [4] for discussions and references) and in the theory of effective material moduli in composite materials in the 1970's – 1980's. More recently they have been used in estimation of distributed growth rates from aggregate population data [5, 6, 7] and in an application to the Preisach theory of hysteresis in input operators for shape memory alloys [8, 9].

The considerations in the previous sections (which are simple initial remarks) suggest the possiblity of a useful *reformulation* of many of the concepts and approaches in inverse problem methodology, both theory and computation.

There are a number of other relevant important topics that, due to lack of time, we will not mention in detail here. Included are model approximation and order reduction and the question of "goodness" of the reduced order model. The popular Karhunen – Loève or proper orthogonal decomposition (POD) techniques offer great promise here. These model reduction techniques have been shown to be quite useful in both open loop and feedback or closed loop control computational methodologies (see [10, 11] and the references therein) and are currently being tested in inverse problem methodologies. The POD formulation itself is of interest since it inherently contains techniques for computing the efficiency of the reduced order model in capturing the dynamic energy present in the data or large scale model used to generate "snapshots". It thus has potential for use in a "goodness of model" framework.

In inverse problem methodology, there are least squares probablistic based statistical tests for comparing the fits of different models to data wherever the models are special cases of a larger model (see [12] and in particular Chapter V.5-7 of [13]). The underlying ideas can be exploited to develop a type of uncertainty measure in inverse problem techniques.

Acknowledgements

The author is grateful to Mac Hyman, Linda Petzold and Jim Crowley among others for initial discussions on uncertainty in modeling and computation. He also expresses thanks to Laura Potter for teaching him some MATLAB plotting tricks.

References

[1] P. Billingsly, *Convergence of Probability Measures*, Wiley, New York, 1968.

[2] R.C. Brunet, C.J. Struchiner and M.E. Halloran, "On the Distribution of Vaccine Protection Under Heterogeneous Response", *Math. Biosc.*, 116 (1993), pp. 111-125.

[3] H.T. Banks, B.G. Fitzpatrick, and Y. Zhang, "Estimation of Distributed Individual Rates from Aggregate Population Data", CRSC-TR94-13, Sept. 1994; in *Differential Equations and Applications to Biology and to Industry*, ed. by M. Martelli et al, World Scientific Press, 1996, pp. 13-22.

[4] J. Warga, "Optimal Control of Differential and Functional Equations", Academic Press, New York, 1972.

[5] H.T. Banks and B.G. Fitzpatrick, " Estimation of Growth Rate Distributions in Size-Structured Population Models", CAMS Tech. Rep. 90-2, January, 1990, University of Southern California, *Quart. Appl. Math.*, 49 (1991), pp. 215-235.

[6] H.T. Banks, " Some Remarks on Estimation Techniques for Size-Structured Population Models", CRSC-TR92-11, October, 1992; *Frontiers of Theoretical Biology* (Lecture Notes in Biomath 100), S. Levin, Ed. (Springer-Verlag, Berlin) 100 (1994), pp. 609-623.

[7] H.T. Banks, B.G. Fitzpatrick, L.K. Potter, and Y. Zhang, " Estimation of Probability Distributions for Individual Parameters Using Aggregate Population Observations", CRSC-TR98-6, January, 1998; *Stochastic Analysis, Control, Optimization and Applications* (W. McEneaney, G. Yin, and Q. Zhang, eds.) Birkhäuser Verlag, Basel, to appear.

[8] H.T. Banks, A.J. Kurdila, and G. Webb, " Identification of Hysteretic Control Influence Operators Representing Smart Actuators, Part I: Formulation", CRSC-TR96-14, April, 1996; *Mathematical Problems in Engineering*, 3 (1997), pp. 287-328.

[9] H.T. Banks, A.J. Kurdila, and G. Webb, "Identification of Hysteretic Control Influence Operators Representing Smart Actuators, Part II: Convergent Approximations", CRSC-TR97-7, April, 1997; *J. of Intelligent Material Systems and Structures*, 8 (1997) pp. 536-550.

[10] H.V. Ly and H.T. Tran, " Proper Orthogonal Decomposition for Flow Calculations and Optimal Control in a Horizontal CVD Reactor", CRSC-TR93-13, March, 1998; *Quarterly of Applied Math*, to appear.

[11] H.T. Banks, R.C.H. del Rosario, and R.C. Smith, " Reduced Order Model Feedback Control Design: Numerical Implementation in a Thin Shell Model", CRSC-TR98-27, July, 1998; *IEEE Trans. Auto. Control*, submitted.

[12] H.T. Banks and B.G. Fitzpatrick, "Statistical Methods for Model Comparison in Parameter Estimation Problems for Distributed Systems", CAMS Tech. Rep. 89-4, September, 1989, University of Southern California; *J. Math. Biol.*, 28 (1990), pp. 501-527.

[13] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhaüser Boston, 1989, pp. 320.