SANDIA REPORT

SAND2009-6072
Unlimited Release
Printed September 2009

Practical Reliability and Uncertainty Quantification in Complex Systems: Final Report

Paul T. Boggs, Matthew D. Grace, Youssef M. Marzouk, Philippe P. Pébay, John Red-Horse, James T. Ringland, Kathleen Diegert, and Rena Zurn

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/ordering.htm

2009-6072
Unlimited Release
Printed Sept 2009

Practical Reliability and Uncertainty Quantification in Complex Systems: Final Report

Paul T. Boggs
Sandia National Laboratories
M.S. 9159, P.O. Box 969
Livermore, CA 94551, U.S.A.
ptboggs@sandia.gov

Matthew D. Grace
Sandia National Laboratories
M.S. 9159, P.O. Box 969
Livermore, CA 94551, U.S.A.
mgrace@sandia.gov

Youssef M. Marzouk
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
ymarz@mit.edu

Philippe P. Pébay
Sandia National Laboratories
M.S. 9051, P.O. Box 969
Livermore, CA 94551, U.S.A.
pppebay@sandia.gov

John Red-Horse
Sandia National Laboratories
M.S. 0828, P.O. Box 5800
Albuquerque, NM 87185, U.S.A.
jrredho@sandia.gov

James T. Ringland
Sandia National Laboratories
M.S. 9007, P.O. Box 969
Livermore, CA 94551, U.S.A.
jtringl@sandia.gov

Kathleen Diegert
Sandia National Laboratories
M.S. 0830, P.O. Box 5800
Albuquerque, NM 87185, U.S.A.
kvdiege@sandia.gov

Rena Zurn
Sandia National Laboratories
M.S. 9007, P.O. Box 969
Livermore, CA 94551, U.S.A.
rmzurn@sandia.gov

Abstract

The purpose of this project was to investigate the use of Bayesian methods for the estimation of the reliability of complex systems. The goals were to find methods for dealing with continuous data, rather than simple pass/fail data; to avoid assumptions of specific probability distributions, especially Gaussian, or normal, distributions; to compute not only an estimate of the reliability of the system, but also a measure of the confidence in that estimate; to develop procedures to address time-dependent or aging aspects in such systems, and to use these models and results to derive optimal testing strategies. The system is assumed to be a system of systems, i.e., a system with discrete components that are themselves systems. Furthermore, the system is "engineered" in the sense that each node is designed to do something and that we have a mathematical description of that process. In the time-dependent case, the assumption is that we have a general, nonlinear, time-dependent function describing the process.

The major results of the project are described in this report. In summary, we developed a sophisticated mathematical framework based on modern probability theory and Bayesian analysis. This framework encompasses all aspects of epistemic uncertainty and easily incorporates steady-state and time-dependent systems. Based on Markov chain, Monte Carlo methods, we devised a computational strategy for general probability density estimation in the steady-state case. This enabled us to compute a distribution of the reliability from which many questions, including confidence, could be addressed. We then extended this to the time domain and implemented procedures to estimate the reliability over time, including the use of the method to predict the reliability at a future time. Finally, we used certain aspects of Bayesian decision analysis to create a novel method for determining an optimal testing strategy, e.g., we can estimate the "best" location to take the next test to minimize the risk of making a wrong decision about the fitness of a system. We conclude this report by proposing additional fruitful areas of research.

Keywords: Complex systems, system-of systems, modern probability, polynomial chaos, Markov Chain Monte Carlo, Bayesian analysis, Karhunan-Loève expansion

Contents

Acronyms

CDF cumulative distribution function

KL Karhunen-Loève

KS Kolmogorov-Smirnov

MCMC Markov chain Monte Carlo

NW nuclear weapon

PCE polynomial chaos expansion

PDF probability density function

RV random variable

Nomenclature

a_i	Action to be taken `\action{}`
$\mathscr{A}$	Space of all possible actions `\Aspace`
$F^{\boldsymbol{\pi}}(\boldsymbol{\theta})$	Cumulative distribution function of $\boldsymbol{\pi}(\boldsymbol{\theta})$ `\cdf{}`
$\mathscr{D}_i$	Data collected at the ith node (no subscript denotes the collection of all data) `\data{}`
$\delta_i(X)$	Decision rule given data X `\dec{}`
$\mathscr{X}^i$	Space of all possible data (superscript denotes the ith subspace `\Dspace{}`
η	Error node of the composite system `\error`
$E^{\boldsymbol{\pi}}_{\boldsymbol{\theta}}[\cdot]$	Expectation value of some function or random variable given a state of nature $\boldsymbol{\theta}$ and a probability distribution $\boldsymbol{\pi}$ `\expect{}{}`
Φ_i	Function mapping outputs from parent nodes of the jth node to output at the ith node `\fn{}`
H_i	The ith Hermite polynomial `\herm{}`
$f(x,\boldsymbol{\theta})$	Likelihood function for data x given the state of nature $\boldsymbol{\theta}$ `\llh`
$L(\boldsymbol{\theta},a)$	Loss function quantifying the loss of taking action a given the state of nature $\boldsymbol{\theta}$
n_ℓ	Total number of leaf nodes of the composite system `\nl`
$n_{\mathrm{p},i}$	Total number of parent nodes for a particular node `\np{}`
ξ_i	Gaussian or normal random variable from a polynomial chaos expansion from the ith node, i.e, $\xi_j \sim \mathscr{N}(0,1)$ `\nrv{}`
q_i	Order of the polynomial chaos expansion for the ith node `\ord{}`
$\underline{\Omega}$	Lower limit of root output (Ω_1) for reliability quantification `\outliml`
$\overline{\Omega}$	Upper limit of root output (Ω_1) for reliability quantification `\outlimu`

Ω_i Output at the ith node, calculated by a polynomial chaos expansion `\out{}`

p_i Set of parent indices for the ith node `\parents{}`

$\pi(\boldsymbol{\theta})$ Probability density function for the random variable $\boldsymbol{\theta}$ `\pdf`

$\Pr^{\pi}(\cdot)$ Probability of some "thing" given a probability distribution π `\prob{}`

$\mathscr{R}(\boldsymbol{\theta})$ Reliability for the composite system given the state $\boldsymbol{\theta}$ `\rel`

$\mathscr{R}_{\text{req}}$ Required reliability for the composite system `\relreq`

X^i Random variable associated with data collection (superscript indicates the ith node) `\rvd{}`

Θ_i Space of all possible states of nature (subscript denotes the ith subspace) `\Sspace{}`

$\boldsymbol{\theta}$ State vector corresponding to coefficients in all of the polynomial chaos expansions `\state`

θ_i^j State vector/coefficients of the polynomial chaos expansion for the ith node and jth component or power (superscript is context-dependent) `\sve{}{}`

$m(x)$ Unconditional marginal density of x `\umd`

$\mathscr{U}$ Notation for a uniform probability density function `\updf`

v_i Designation for the ith node where data was not collected `\vv{}`

Chapter 1

Introduction and Executive Summary

1.1 Problem, Goals, and Approach

One of the major goals of this research was to provide a comprehensive framework to modernize reliability estimation for nuclear weapon (NW) systems. This was motivated by a strong desire to reduce or eliminate the assumptions of Gaussian (normal), or other specific distributions on the random variables associated with the system. Engineers wanted to be able to use continuous, as opposed to discrete or pass/fail data and they wanted a way to model and assess the effects of aging on the systems. Figure 1.1 gives a diagram of current NW models and our version with our contributions.

Our approach was to create a framework for including general forms of epistemic uncertainty based on modern probability methods and Bayesian analysis. We use a network model of hierarchical systems, including system-of-systems models, from which certain factorizations and computational efficiencies can be derived; we employ a general representation of all random variables that avoids any assumption of specific distribution; we allow testing at arbitrary nodes in the system; and we use Bayesian techniques to propagate the effects of such tests through the system. This approach allows us to estimate the reliability of the system *and* a distribution of the reliability, so that many questions about the reliability, including confidence in any estimate, can be addressed. Our framework provides the ability to perform testing over time and therefore can be used to predict the reliability distribution at a future time. Finally, our analysis lends itself to the techniques of Bayesian decision theory, which, in turn, provides us with a systematic mechanism to specify optimal testing strategies that improve both our estimates and our confidence in them. In the next section, we present our basic model. We then discuss the key technical constituents in our approach, the difficulties that were addressed to create computationally efficient algorithms, and finally, our major accomplishments and their significance. In the process, we give an outline of the subsequent chapters in this report.

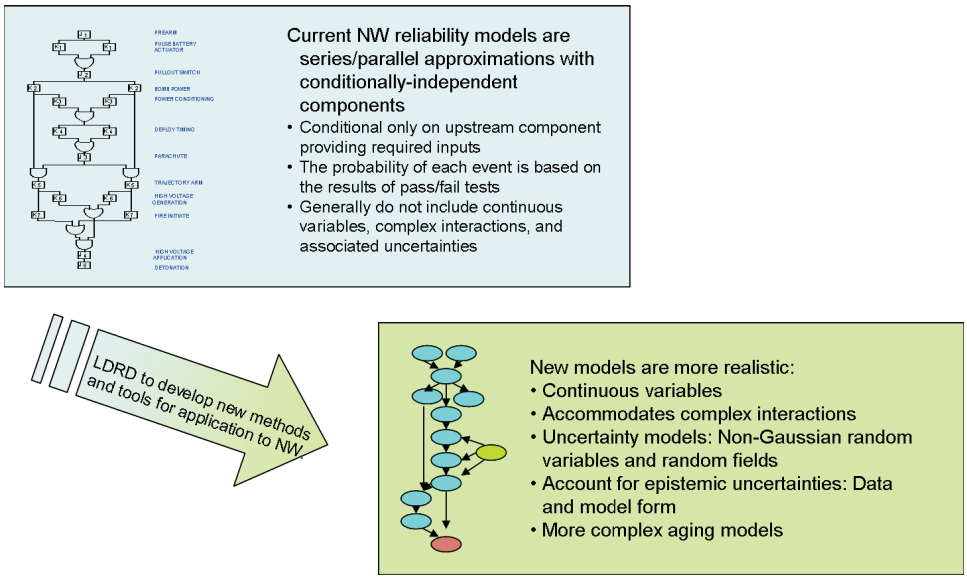


Figure 1.1. Current NW reliability models and their capabilities compared to the new models that are addressed in this project.

1.2 Model Description

We are assuming a system-of-systems model where a system consists of subsystems, each of which is a system. We can continue this to any level. Thus, we have a classic hierarchy tree as our basic model for the systems of interest. We can extend our results to other structures, a point that we will consider in our discussion of time-dependent systems in 4. Figure 1.2 shows a simple model of the type of system that we consider.

To set terminology, we refer to nodes 2 and 6-12 as leaf nodes and to the other nodes as system nodes. The output at any node i is denoted θ_i. We assume a function relationship between the output of any system node and its parents, i.e.,

$$\theta_i = \Phi_i(\theta_j,\ j \in p_i), \tag{1.2.1}$$

where p_i is the parent set of node θ_i and Φ_i maps output from parent set p_i to ith system node. For many of our early examples, we used a simple, linear model, i.e.,

$$\theta_i = \sum_{j \in p_i} \alpha_j \theta_j,$$

where the α_j are specified constants, but we emphasize that general nonlinear functions are allowed here, giving great flexibility to our models. For example, we could consider a discontinuous function that "turns on" at a certain threshold or a function that selects one of two possible parents to allow for redundancy.

At the leaf nodes, we assume that the output is a random variable (RV). We avoid making any assumptions about the distribution of this RV by representing it as a polynomial

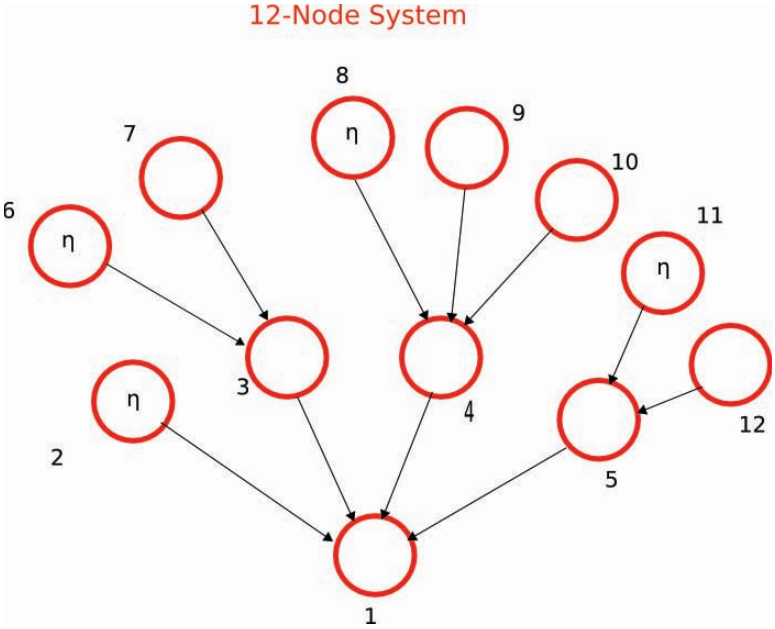


Figure 1.2. Example of a model of our system of systems. The nodes marked η are error nodes.

chaos expansion (PCE). The mathematical background and theory for PCEs are discussed in chapter 2. The PCE at leaf node i is

$$\theta_i = \sum_{j=0}^{q_i} \theta_{ij} H_j(\xi_i), \tag{1.2.2}$$

where q_i is the order of the PCE, H_j is the j^{th} Hermite polynomial, and ξ_i is a normal random variable with mean 0 and variance 1, i.e., $\xi_i \sim \mathcal{N}(0,1)$. A basic result given in chapter 2 is that any RV with finite variance can be expressed by a PCE (in fact, an even more general result is presented). Our approach is based on Bayesian methods and a fundamental tool is Bayes' Theorem given by

$$\pi(\boldsymbol{\theta} \mid \mathscr{D}) = \frac{f(\mathscr{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_\Theta f(\mathscr{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}. \tag{1.2.3}$$

The interpretation of (1.2.3) is as follows:

- The term $\pi(\boldsymbol{\theta} \mid \mathscr{D})$ is the posterior distribution of the state variable $\boldsymbol{\theta}$ given the data $\mathscr{D}$.
- The term $f(\mathscr{D} \mid \boldsymbol{\theta})$ is the probability of seeing the data $\mathscr{D}$ given a fixed value of $\boldsymbol{\theta}$. This is also referred to as the "likelihood function".
- The term $\pi(\boldsymbol{\theta})$ is the prior distribution of the state variable $\boldsymbol{\theta}$. This is set by the user.

- The denominator merely scales the posterior distribution so that it integrates to 1.

The evaluation of the likelihood function can be computationally expensive, so we pay particular attention to calculating this efficiently. Some aspects of this are in chapter 3 and more details are in chapter 4. It turns out that a particularly effective way to compute these likelihoods is to use an exact method mentioned in chapter 2. This requires an ability to find the real roots of polynomials. An extremely effective method has recently been developed and we have successfully applied it to our problems. A disadvantage with using this polynomial solving evaluation technique is the fact that it works most effectively with a stochastic dimension of one. To deal to this situation, we have adapted a projection method to reduce the dimension to one. This work is also reported in chapter 2.

To put our work into perspective and to give an indication of how the above tools are employed, we state the overall procedure for estimating the reliability distribution. For simplicity, we state it in terms of the steady-state problem; the extension to time-dependent problems is readily accomplished in this framework. (See chapter 4.)

First, we address the question of computing the reliability of the system given a specific value of the state variables $\boldsymbol{\theta}$. Computationally, at leaf node i, we sample the independent RV ξ_i and compute a sample output at node i by using (1.2.2). We then propagate these values through the system using the function relations at the system nodes given by (1.2.1), ultimately yielding a sample at node 1. Here, we assume that the system is deemed "reliable" if the output is between specified upper and lower bounds. Thus, we can simply compute the percentage of points in the sample that are contained in these bounds as the reliability of the system, given the state of nature $\boldsymbol{\theta}$.

Recall that the interpretation of Bayes theorem (1.2.3) is that we obtain a posterior distribution of the state $\boldsymbol{\theta}$ given test data $\mathscr{D}$. Thus, we want to sample the posterior distribution and for each value of the sample, we compute the reliability as just described. In this way we obtain a distribution of the reliability, as desired.

Because we do not have analytic formulas to describe the posterior distributions, it must be numerically sampled and constructed. A standard method for doing this is Markov chain, Monte Carlo (MCMC) [17]. The advantage of using MCMC is that it can be adapted (by using so-called "reversible-jump" techniques) to estimate the proper order q_i as well to provide samples of the state $\boldsymbol{\theta}$. This is very important since it relieves the user of the need to specify these orders in the majority of cases where there will be little to guide such a choice. Our version of reversible-jump MCMC is described in detail in chapter 4.

1.3 Key Accomplishments and Their Significance

This project has resulted in the novel exploitation of various analytical and probabilistic tools to provide general density estimation based on testing at arbitrary nodes in the system.

We first developed our approach in the context of steady-state systems, which avoided some complicating issues. Nevertheless, our approach here was, by design, general enough to include time-dependent systems. This approach provides, for the first time, a framework for dealing with all forms of epistemic uncertainty. Finally, we have been able to exploit this framework in the context of Bayesian decision analysis to provide optimal strategies for subsequent testing. Although this work was accomplished in the context of the network described in figure 1.2, our approach is much more general and can be applied to many other network problems. Some of these are mentioned in section 1.4.

We summarize the key accomplishments of this project under four headings, each with its own chapter. Each of these chapters is meant to be a stand-alone paper; as such, each will be submitted for publication.

Basic Theory There has been some confusion in the literature about the basic properties of PCEs. In Chapter 2 we discuss these issues, including the senses of equality and convergence of the associated approximations and how it affects our use of the approximations in a given analysis scenario. We also give one method for generating a given class of PCEs with observed data, which we assume are manifested as a finite collection of observations. Finally, we give a research path to connect the more general function analytic probabilistic approach with the field of statistical decision theory.

Steady-State Systems Our first work on this model considered only a steady-state system and only a simple version of the underlying network. This avoids some computational difficulties and allows us to concentrate on the basic procedures and algorithms. Thus, in chapter 3 we pose the problem as that of general density estimation based on the acquisition of data. Here we adapt the MCMC method to our system and investigate the use of reversible-jump techniques. We show that the chains produced by MCMC display appropriate mixing and that the reversible-jump procedures examine various orders before settling on the best choice. One computational piece that is required, even in this case, is the evaluation of the likelihood function. We use a result in chapter 2 that gives an exact value provided that the roots of a polymomial can be calculated quickly. A newly developed algorithm for this was used and shown to be approximately ten times faster than standard kernel density estimation methods. In addition, an enhancement of the polynomial-solving method was developed that produced an even faster evaluation. These techniques form the basis for the extension to larger models and the dime-dependent case, discussed next.

Time-Dependent Systems In chapter 4, we address two important issues. First, we consider the general time-dependent model and show that, in the first version of the problem, it is readily accommodated in our framework. Second, we consider the computational issues of a multi-node system that were avoided in chapter 3.

We formulate our model based on using parameterized, time-dependent coefficients in the PCE expansion given in (1.2.2) and we show that Bayes' Theorem continues to be the basis for propagating the effects of test data through a time-dependent system

to obtain a joint distribution for all of the parameters. As noted above, we use MCMC to sample this distribution. In fact, we use a reversible-jump technique and we give the details of the method, including the procedures for taking so-called "up-order" and "down-order" moves. The use of MCMC requires the evaluation of the likelihood function and we show that the network structure here allows for a factorization of the likelihood function that results in a much more computationally efficient method. In fact, this is one of the main bottlenecks in the computation and will impose limits on the size of the problem that can be addressed, but we note that our factorization routine allows the use of parallel computing and we show a simple example of how we obtain nearly linear speedup in certain cases. A more general parallel approach is still being explored.

The algorithm is tested on several representative systems with both a simple linear, time-dependent model, i.e., the coefficients of the PCE are linear functions of t and with a more realistic negative exponential mode that allows controlled drift from one state to another. In both cases, our algorithm was able to estimate the reliability of the system at the times where data were collected. More interestingly, our algorithm was able to accurately predict the reliability at a future time. We conclude this paper with some ideas for future research; these ideas are also briefly mentioned below.

Optimization Questions In chapter 5, we present an optimal sampling strategy based on Bayesian decision analysis [3]. After an initial estimate of reliability is calculated for a complex system-of-systems structure (e.g., figure 1.2) using the available test data and prior information, improving the accuracy of the reliability estimate requires the collection of additional data. However, testing all possible sub-systems may not be cost-effective, feasible, or even *necessary* to achieve an improvement. To address this sampling issue, we formulate a Bayesian methodology that systematically determines the optimal sampling strategy that will maximally improve the reliability estimate of the composite system, e.g., by reducing the variance of the reliability distribution.

This methodology involves calculating the "Bayes risk of a decision rule" for each available sampling strategy. We accomplish this by defining a suitable loss function, which, given a particular action (e.g., "accept" or "reject") and state of nature $\boldsymbol{\theta}$, quantifies the cost or loss associated with that action, e.g., the cost of accepting a device although it is not performing within specifications. Next, we introduce a decision rule that, given additional data, selects the action that minimizes the expected loss. The loss function and the decision rule are combined in the Bayes risk of a decision rule, which then allows us to calculate relative risk for each possible sampling strategy. In this scenario, risk quantifies the relative effect that each sampling strategy could have on the reliability estimate.

1.4 Future Research

As noted above, we believe that our framework is extensible to many other application areas where systems can be represented by networks, including nuclear weapons, electric power grids, certain biological systems, some aspects of cyber security, water distribution, and complex interacting systems.

In particular, some characteristics of networks that we intend to study include networks with feedback and networks with redundant components. A long-term goal is to extend the results to adaptive networks, i.e., networks where connections between nodes changes over time as a function of adaptive pressures.

Feedback induces a new kind of complexity into the network and is often associated with complicated, nonlinear behavior, including emergent behavior. Here, we think that the components of a feedback loop can be tested as above to assess the reliability of the loop itself. The output of the loop can then be forward-propagated to determine its impact on the full system. In the case of redundant components, assume for simplicity that a component is duplicated and that only one of the two instances has to function properly. The child node of these two parents then has to use either of its two parents, whichever is operating correctly. Another possibility is to have several redundant parents with the child node using a voting strategy on its inputs. This could be handled by writing an appropriate output function at that system node. Additionally, the Bayesian methodology for identifying optimal sampling strategies (see chapter 5) may be adapted to also identify components of a system where redundancy could significantly improve reliability, e.g., components to which operation of the composite system is most sensitive. More general network topologies imply a different set of modeling issues, but the Bayesian framework can still be applied. The feature that is most important will be its connectedness. If the network is appropriately structured so that it can be exploited in the computation of the likelihood function, then we believe that our procedures can be applied and that we can achieve scalability. Finally, in some networks, including cyber networks and power grids, a node may have been compromised and is now reporting fake data. Our models have a natural way of approaching this situation since we compute the likelihood of seeing the data given the state of the system. That is, we can assess the likelihood of the data in the context of the data that is being received from the other nodes. The challenge is to distinguish between good data that implies a fault from the node and fake data that would induce an operator to take action that would cause a disaster.

In the case of time-dependent systems, we have identified several mathematical and algorithmic questions that merit further study. For example, most stochastic processes cannot be described well with only one stochastic dimension. We propose, therefore, using a general Karhunan-Loève (KL) expansion (see chapter 2) [12] to model the process at node i:

$$\theta_j = \sum_{i=0}^{\infty} \rho_i(t)\gamma_i(\omega) \tag{1.4.1}$$

where the γ_i are RVs on some probability sample space Ω with $\omega \in \Omega$, and the $\rho_i(t)$ are the

eigenfunctions of the covariance operator of $\boldsymbol{\theta}_i$, i.e., for

$$C(t,t') = E^{\Omega}[\boldsymbol{\theta}_j(t,\omega)\boldsymbol{\theta}_j(t',\omega)]$$

the functions ρ_i satisfy

$$\int_T C(t,t')\rho_i(t')dt' = \lambda_i\rho_i(t).$$

By using several terms of this expansion and approximating the RVs γ_i by PCEs, we obtain a model with a higher stochastic dimension, but one that still fits into our framework. The number of variables, however, grows rapidly as we increase the stochastic dimension, so both practical and algorithmic considerations must be addressed to make this a tractable option.

A related extension is the case in which the nodes of the system have more than one correlated output. (If the outputs are independent, then we could model them by separate nodes and we would be back to our original model.) In the case of several outputs, the PCE at a leaf node would have several stochastic dimensions, but we believe that our procedure for rapid evaluation of the likelihood functions can be extended. In this case, we would have to solve systems of polynomial equations, but the polynomial solver described in Chapter 4 could still be used.

Chapter 2

Theory

2.1 Introduction

Probabilistic approaches have long been the standard bearers in scientific analyses involving uncertainty quantification (UQ). In the field of probability, there are two primary methods for effecting this UQ. The first of these is what we will refer to as the traditional probabilistic approach, in which one is concerned with properties of certain probabilistic entities, such as cumulative distribution functions (CDFs), probability density functions (PDFs), or statistical moments of various random functions, and their behavior under transformation or limit operations.

There is, however, an alternative analysis path, which we will refer to as the function analytic approach to probability where we look at features of the random functions themselves under, say, transformation.

The latter approach is based on recognizing that random variables (RVs), vector random variables (VRVs) and random fields (RFs) are, under appropriate conditions, (measurable) functional mappings with at least a subset of their domains consisting of a sample space, Ω, of elementary events that is well-defined in the context of a probability space. A probability space consists of a triple, $(\Omega, \mathscr{S}, P)$, which includes, in addition to the sample space, a σ-algebra $\mathscr{S}$ of subsets of Ω called events, and a probability measure P. Each of these entities has well-established and precise mathematical properties [40].

Using this structure, one sees immediately that a random function is, in fact, not random at all, and that whatever randomness that exists in the framework is entirely associated with the occurrence of events in the probability space. Thus, it seems reasonable: (1) To cast these random functions in a function analytic/function space setting, and, within this setting to approximate them in ways that are identical to deterministic approximation schemes; and, (2) To analyze the behavior of these approximations under various mathematical operations in ways that are similar to those employed in common deterministic computational

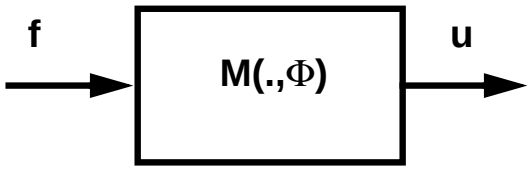


Figure 2.1. Constituents Of Our Model Problem

methods

In fact, given the common features present in, say, a finite element-formulated computational method and one that accommodates UQ using the latter approach described above, it shouldn't be surprising that it is becoming a popular way to perform computational UQ in engineering applications. However, this new popularity has given rise to some misunderstandings concerning some of the practical probabilistic aspects one should consider when using it. In this document, we seek to describe some of these, and to discuss ways to avoid misinterpretation of UQ results using methods based on the functional approach.

2.2 Functional Analytic Approach to Probability

Much of the function analytic approach to UQ can be traced to the early work of N. Wiener [48], who developed the polynomial chaos expansion (PCE) as a means to integrate operators possessing differential Brownian motion, at the time viewed as chaotic, as an external forcing influence. While the time-indexed RF, often called a stochastic process (SP), he described is quite general, we will build to this using a simpler route: After giving a brief description of the general model problem in terms of RFs, we will consider the RV case, then describe the process of transition to the more general cases of vectors containing RVs as components and the more general RFs/SPs.

Consider a model, **M**, which we will assume to define a mapping from input to output for a given physical process. We emphasize that **M** can depict mappings that are experimentally defined, analysis based, or a combination of the two, such as one might encounter in a model validation, or experiment design exercise.

We depict the broad constituents of this input-output relationship graphically in Figure 2.1. Many engineering analysis processes are modeled by ordinary or partial differential equations; ODEs and PDEs, respectively. For that case, we can think of **M** formally as an inverse operator for a given differential equation, and as taking into account any computa-

tional artifacts, such as discretization and floating point arithmetic.

In mechanics, we generally think of the inputs, $\mathbf{f}$, as providing external energy, or loadings, into the system; we often referred to these loadings as excitations. Naturally, our model $\mathbf{f}$ can be used to describe uncertainty, both in the inherent and reducible sense. In addition, a given model may contain any number of passive parametric entities, or descriptors. Examples of these parameters, again in a structural mechanics context, include values associated with material properties, such as the elastic modulus, E, and Poisson ratio, ν, for an linear isotropic elastic material. Often these parameters have physical bases like the material properties, but these parameters can have purely phenomenological origins, such as in parameters in a contact model along a structural interface; they can also be present entirely due to computational considerations, such as element size, etc. Any subset of these parameters that are to be modeled as uncertain entities, whether as discrete or continuous variables, or as spatially dependent field quantities, will be collected into the vector, Φ, which is also shown in Figure 2.1.

We consider, then, a given analytical system to be well described, with particular focus on uncertainty, by the following equation:

$$\mathbf{u} = \mathbf{M}(\mathbf{f}, \Phi) \tag{2.2.1}$$

where $\mathbf{f}$ and $\mathbf{u}$ are the (possibly) vector-valued system excitation and output, or response, respectively.

Let's briefly consider the case where $\mathbf{f}$ is deterministic and Φ is composed solely of a scalar RF, $\alpha = \alpha(x, \omega)$, where the arguments are points in the deterministic and probability constituent spaces, respectively, a point that we will make clearer shortly. We note that for this situation, u will also be an RF, and, in a computational setting our task is to find an approximation for the RF u based on approximation for RF α. Often, this response is used further downstream in a given analysis to calculate additional quantities of interest; in this context, we need to understand just what our solution process means to ensure that all of our results are taken in the proper context.

The most common way to build our approximations is to construct truncated Polynomial Chaos Expansions (PCEs) in the style of Wiener [48].

The PCE method for random fields is based on the discovery of a more basic collection of RVs on which to build orthogonal expansions; specifically, a countably infinite collection of independent, zero-mean Gaussian, or standard normal, RVs, which are the underlying variables in multi-dimensional Hermite polynomials [10]. In anticipation of our need to restrict this collection to finite-dimension on which we will base our approximations, we will attempt to provide a description from a more approachable perspective.

2.2.1 Scalar RVs

Consider two real-valued scalar RVs, X and $\bar{Y}$, where X is defined on a probability space, or triple, $(\Omega, \mathscr{S}, P)$, and $\bar{Y}$ is defined on a (generally different) probability triple, $(\bar{\Omega}, \bar{\mathscr{S}}, \bar{P})$. That is, $X : \Omega \longrightarrow \mathbf{R}$ and $\bar{Y} : \bar{\Omega} \longrightarrow \mathbf{R}$, with X and $\bar{Y}$ assumed to be measurable mappings to the real line, $\mathbf{R}$.

For a variety of reasons, it is often of interest to push forward from these domains to entities that are defined on $\mathbf{R}$. This is done by considering the so-called distribution of the RV, μ, and various quantities defined using it. For example, for X, we have distribution μ_X, which is defined using the measurability of X by

$$\mu_X(B) = P(X^{-1}(B)), \forall B \in \mathscr{B} \tag{2.2.2}$$

where $X^{-1}(B)$ is guaranteed to be in $\mathscr{S}$ by the definition of measurability and $\mathscr{B}$ is the Borel σ-algebra on $\mathbf{R}$ [41]. To complete the context for our push forward, we can see that we have a new probability triple, defined on the real line, $(\mathbf{R}, \mathscr{B}, \mu_X)$.

We will need to consider the existence of a functional transformation between X and $\bar{Y}$, $\bar{Y} = T(X)$, where we are able to specify the cumulative distribution function (CDF), $F(x) = \mu_X\{X \leq x\}$, on the input X. We begin with the following theorem, which we state, verbatim, from [40]:

Theorem 2.2.1. *Let U be a uniform RV over $[0,1]$, and let $F : \mathbf{R} \longrightarrow [0,1]$ be any CDF. Define Ψ as*

$$\Psi(u) = \inf_{x \in \mathbf{R}} \{x : F(x) \geq u\} \quad \forall u \in (0,1).$$

Then $Z = \Psi(U)$ is a RV whose CDF is F.

Proof. Since the CDF, F, is non-decreasing by definition, if $F(z) \leq u$, then $\Psi(u) \geq z$. On the other hand, since F is right continuous, we have that $\inf\{x \in \mathbf{R} : F(x) \geq u\} = \inf\{x \in \mathbf{R} : F(x) \geq u\}$; that is, the infimum is obtained. If follows that if $F(z) \geq u$, then $\Psi(u) \leq z$. Hence, $\Psi(u) \leq z$ if and only if $u \leq F(z)$. Since $0 \leq F(z) \leq 1$, we obtain $F(z) = \mu_U\{\Psi(U) \leq z\} = \mu_U\{U \leq F(z)\}$. □

In the proof, we understand the term $\mu_U\{\,\cdot\,\}$ as relative to a uniform-distributed RV on the probability space $([0,1], \mathscr{B}[0,1], \mu_U)$.

Per this theorem, we are guaranteed the existence of a transformation from an arbitrarily distributed RV, X, with CDF, F, to a uniform RV, U and that the function $\Psi(u)$ defined in the theorem satisfies the functional relationship $\Psi(u) = F(x)$. We can certainly repeat this process for a second RV, $\bar{Y}$, with CDF $\hat{F}$ such that $\hat{\Psi}(u) = \hat{F}(x)$; and, using the uniform RV as a pivot and inverting the relationships for $\hat{F}$ and Ψ, we can transform from a RV X to a RV $\bar{Y}$. That is, we have a process for determining T explicitly when the constituent CDFs are known explicitly.

Since our subsequent algorithms all rely implicitly on the existence of T, it is extremely important to observe from Theorem 2.2.1 that the sense of equality is "in distribution" only. That is, we've related two RVs through a transformation based entirely on their CDFs. We emphasize that this sense of equality, while weaker than others in probability theory, is still useful if used properly, such as in the case where one is computing probabilities of events. An artifact of this weaker sense of equality, is that the RVs X and $\bar{Y}$ do not even have to be defined on the same probability spaces. To phrase this differently, they don't necessarily need to be driven by the same underlying phenomena; they are related purely through their CDFs.

2.2.1.1 Generalized Fourier Expansion of $T(\,\cdot\,)$

We constructed T as a function based on criteria related to the distributions of the input and output RVs in Section 2.2.1. As for the RVs themselves, we relied solely on their measurability.

We now place an additional constraint on these functions; namely, that they be members of the function space of RVs with finite variance. We also now create a new RV, $Y = T(X)$, which we distinguish from $\bar{Y}$ by requiring that the equality holds "almost everywhere." The chief motive for defining this alternative RV to $\bar{Y}$ is to ensure that both RVs involved, X and Y, are defined on the same probability space, in this case, $(\Omega, \mathscr{S}, P)$. Almost everywhere equality means that the probability that these two mappings are equal is one, $P(Y = T(X)) = 1$.

Since both RVs in $Y = T(X)$ are finite variance by assumption, they are members of the class of square-integrable functions on their domain, the sample space; thus $Y \in L_2(\Omega)$. It is also known that members of L_2 constitute a Hilbert space [41] of functions. Specifically, $L_2(\Omega)$ is a space of functions that is a complete, normed, inner-product space with

$$E[\,\cdot\,] = \int_{\Omega} \cdot \, dP(\omega), \tag{2.2.3}$$

the operator of mathematical expectation, the inner product operator. One conceptual motive for effecting push forward is the transformation of integration operations, such as the one in Eq 2.2.3, from abstract sample spaces of events to the real line. We present the following theorem without proof to justify this [40], pg 67–68.

Theorem 2.2.2 (Change of Variable Theorem)**.** *Let X be an RV on a given probability space, $(\Omega, \mathscr{S}, P)$, with distribution μ_X. Then for any Borel measurable function, $f : \mathbf{R} \longrightarrow \mathbf{R}$, we have*

$$\int_{\Omega} f(X(\omega)) P(d\omega) = \int_{\infty}^{\infty} f(t) \mu_X(dt) \tag{2.2.4}$$

provided that the integral on either side of the equality exists. We note that the integrals are understood to be interpreted in the Lebesgue sense [41], which coincides with the Riemann integral when it exists.

As one might expect, there are a number of mathematical properties that membership in the function space L_2 entails. Most important to our purposes is that we know that it is possible to construct a generalized Fourier expansion [21] to approximate T, and that this can be done to arbitrary accuracy.

For the case of PCEs, and for reasons we will discuss later, we choose to construct this generalized Fourier expansion in terms of a set of orthogonal polynomials in a standard normal random variable, $X = \xi$, that we will denote as $\{\Gamma_i(\xi)\}$.

The mathematical construction we have described above ensures that we can construct approximations, $Y^{(n)}$, such that the norm of the difference $\|Y - Y^{(n)}\|_{L_2}$ can be made arbitrarily small when taking a large enough n. In other words,

$$\lim_{n\to\infty} \|Y - Y^{(n)}\|_2 = 0,$$

and the approximations are said to converge in the mean.

The $\{\Gamma_i\}$ are referred to as Hermite polynomials, and their properties can be found in any of a number of references documenting orthogonal polynomials such as [1]. In one dimension, these polynomials are specified by the following formula,

$$\Gamma_i(\xi) = \sum_{j=0}^{[i/2]} (-1)^j \frac{i!}{(i-2j)!\,j!\,2^j} \xi^{i-2j} \tag{2.2.5}$$

where the expression $[r]$ evaluates to the largest integer less than or equal to r and $0! = 1$. The first four of these are

$$\begin{aligned} \Gamma_0(\xi) &= 1 \\ \Gamma_1(\xi) &= \xi \\ \Gamma_2(\xi) &= \xi^2 - 1 \\ \Gamma_3(\xi) &= \xi^3 - 3\xi. \end{aligned} \tag{2.2.6}$$

Many orthogonal polynomials derive from particular ordinary differential equations defined on Hilbert spaces [50] known as Sturm-Liouville systems. These systems induce an associated inner product weighting function that defines their orthogonality properties. For the Hermite polynomials this weighting function is, to within a constant, readily recognized to be identical to the PDF of a standard normal RV, which explains our choice above.

With these specifics in hand, we can now exploit the orthogonality and properties of the inner product to build our PCE-based approximations:

$$Y^{(n)} = T^{(n)}(\xi) = \sum_{i=0}^{n} g_i \Gamma_i(\xi), \tag{2.2.7}$$

where the generalized Fourier coefficients, g_i, are given by

$$g_i = \frac{EX\Gamma_i(\xi)}{E\Gamma_i^2(\xi)}. \tag{2.2.8}$$

Note that for the discrete case it is possible to generalize Eqs 2.2.5 and 2.2.7 to higher dimensions in ξ, say $\xi = (\xi_1, \ldots, \xi_m)$, where m is referred to as the stochastic dimension. We caution, though, that the process is rather complicated; we refer the reader to [12] for the necessary details. The passage to the infinite-dimensional case, under suitable constraints, is a RF and is the subject of [48].

We also note that there are a number of other known probability density functions that also can be identified, simply by inspection, as weighting functions for inner products from different Sturm-Liouville systems. For example, an exponentially-distributed RV defined on the interval $[0, \infty)$, can be seen to be affiliated with the Laguerre polynomials. Note that for the case of finite dimensional collections, none of the underlying theory, nor the operations necessary for constructing a generalized Fourier expansion rely on a particular RV/orthogonal polynomial pairing. Thus, theoretically, it is possible that PCEs can be generalized to any appropriate pairing. These generalizations, often termed Askey Expansions in the literature, are the subject of active research.

2.2.1.2 On Convergence

For the case of computation, we are forced to pick an n in Eq 2.2.7, and work with the now-approximated RV.

But first we pause to investigate exactly what constraints our use of these approximations imposes on us. Recall that we began with our process by establishing the existence of a transformation from a new RV to our original one. We observed that the equality was guaranteed only in the sense that the probability distributions would be identical. We then considered an almost everywhere replacement RV to force everything onto identical probability spaces. We pushed forward to $(\mathbf{R}, \mathscr{B}, \mu_X)$ and exploited the properties of $L_2(\mathbf{R})$ to justify a generalized Fourier expansion representation of the replacement RV. Finally, we truncated this infinite series to arrive at an approximation and observed that this approximation is convergent in the mean. This means that the approximation is also in L_2, thus has a second moment that converges to the almost everywhere convergent surrogate.

But all of this begs the question of what happens when the original transformation yields an output RV with higher moments that are known to exist? Do the approximations also have this property? The answer, sadly, is no. Thankfully, our approximations are still useful despite this.

It is well-known that convergence in the mean implies convergence in distribution, but unfortunately the converse does not hold [29]. Convergence in distribution is also commonly known as weak convergence, not only because it is "weaker" in some heuristic sense, but

because it is consistent with weak convergence in a rigorous mathematical context [4]. Defining this explicitly is beyond our purpose here, but we state the following theorem to help the reader make sense of things [40]:

Theorem 2.2.3. *The following two conditions are equivalent:*

- $Y^{(n)} \longrightarrow Y$ *in distribution;*
- $\int_{\mathbf{R}} f d\mu_Y^{(n)} \longrightarrow \int_{\mathbf{R}} f d\mu_Y$ *for every bounded, Borel measurable real function,* f*, such that* $\mu(D_f) = 0$ *where* D_f *is the set of points at which* f *is discontinuous;*

Since moments involve unbounded functions, we cannot guarantee that they will converge, even if all of the integrals exist. However, it is important to realize that the probability of many interesting events, such as of intervals, is convergent according to theorem, so it is always safe to use such probabilities in analyses involving our series-based approximations. That this is true can be seen from the fact that in this case

$$\mu_Y(B) = \int_B d\mu_Y = \int_{\mathbf{R}} I_B(t) dmu_Y(t). \tag{2.2.9}$$

where $f = I_B$ is the indicator function of the event, $B \in \mathscr{B}$ defined as

$$I_B(x) = \begin{cases} 1, x \in B \\ \\ 0, \text{Otherwise.} \end{cases} \tag{2.2.10}$$

Some examples where these issues are illustrated and discussed can be found in [37].

In general, one has to be very careful when dealing with measures and quantities derived from them. Additional criteria are necessary to relax the boundedness constraint on f in Theorem 2.2.3 (viz. [47]). We can show that we get convergence if the quantities $f(Y^{(n)})$ are also asymptotically uniformly integrable; that is if they satisfy

$$\lim_{M \to \infty} \limsup_{n \to \infty} E|f(Y^{(n)})| I|f(Y^{(n)})| > M = 0. \tag{2.2.11}$$

We end this section with a final cautionary note concerning the issue of the relevance of moments. Even if we have a set of moments that we know converge to the true moments of the system, we still have one more hurdle to clear to say whether or not we could use them to determine a unique distribution: We need to establish whether or not the true system distribution can be determined uniquely by its moments; that is, is the distribution know to be what is termed Moment Determinant. In the general case it cannot. This is known as the Moment Problem, and it is an area of fairly current research. We recommend that the interested reader see [45] for further information on this.

2.2.2 Vector-Valued RVs (VRVs)

When we assemble RVs into vectors, or VRVs, $\mathbf{X}$ and $\bar{\mathbf{Y}}$, it is possible to generalize the procedure above to guarantee a transformation, $\bar{\mathbf{Y}} = \mathbf{T}(\mathbf{X})$. This is done by replacing F for each individual RV described above with an appropriate conditional CDF function, exploiting the pivoting uniform RV, and combining the results. When the VRV $\mathbf{X}$ is composed entirely of absolutely continuous RVs, X_j, we write this as follows

$$\begin{aligned} U_1 &= PX_1 \leq x_1 = F_1(x_1) \\ U_2 &= PX_2 \leq x_2 | X_1 = x_1 = F_2(x_2|x_1) \\ &\vdots \; = \\ U_k &= PX_k \leq x_k | X_{k-1} = x_{k-1}, \ldots, X_1 = x_1 = F_k(x_k|x_{k-1}, \ldots, x_1) \end{aligned} \tag{2.2.12}$$

where the $F(\,\cdot\,|\,|\,)$ are conditional CDFs [40] and the VRV, $\mathbf{U}$, can be shown to have components, U_j, that are independent, uniform RVs on the interval $[0,1]$ of $\mathbf{R}$. We refer the interested reader to [39] for further information on this process, but we note that for the case where the X_j have discontinuous conditional CDFs, we build the left hand side of Eq 2.2.12 in the same way we did in Theorem 2.2.1.

2.2.3 Back to Random Fields (RF/SPs)

Now we reconnect to the original story line we started in Section 2.2: Approximating RFs, using the material in the preceding sections as building blocks.

Under well-known assumptions on a given real-valued, scalar RF, such as measurability, membership in L_2 and the topological property of separability [23], our baseline RF, the measurable mapping $\alpha : \mathbf{D} \times \Omega \longrightarrow \mathbf{R}$ such that

$$\alpha(d,\omega) \in L_2(\mathbf{D} \times \Omega), \;\; d \in \mathbf{D}; \omega \in \Omega. \tag{2.2.13}$$

Here we've referred to a generic deterministic constituent domain, $\mathbf{D}$, where the reader should keep in mind this can be sets such as $\mathbf{R}^2$ for a two-dimensional problem, or $\mathbf{R}^3 \times [0,\infty)$ for a three-dimensional time-dependent problem, etc.

Under suitable assumptions, a zero-mean RF α can be decomposed using a so-called Karhunen-Loève (KL) expansion [44]

$$\alpha(d,\omega) = \sum_{i=0}^{\infty} \sqrt{\lambda_i}\phi_i(d)\eta_i(\omega) \tag{2.2.14}$$

where the $\{\lambda_i\}$ and $\phi_i(d)$ come from the KL eigenproblem

$$\int_{\mathbf{D}_\alpha} r(x,y)\phi_i(y)dy = \lambda_i \phi_i(x). \tag{2.2.15}$$

In Eq 2.2.15, $\mathbf{D}_\alpha(\subset \mathbf{D})$ is deterministic problem subdomain occupied by the RF, α, $r(x,y) = E[\alpha(x,\omega)\alpha(y,\omega)]$ is the correlation function of the RF. Finally, we note that Eqs 2.2.14 and 2.2.15 provide information about the deterministic properties of the RF based on its correlation function. The only information we get on $\eta_i(\omega)$, which derives from the canonical nature of the KL relationships, is that $E[\eta_i(\omega)\eta_j(\omega)] = 0 \forall i \neq j$. This means that all we can say at this point is that the $\{[\eta_i(\omega)\}$ are uncorrelated RVs. We emphasize that one should not take this to mean that they have any particular independence characteristics whatsoever.

One final note regarding features of α: If $\mathbf{D}_\alpha$ is an uncountable set as is the case, for example, if it's an interval on the real line, then the RF can be thought of as an uncountable collection of RVs. The assumption of separability allows us to reduce our considerations to a countably infinite number, and from this countable set, we use some criteria to restrict this set to a finite number. We do this by truncating the sum in Eq 2.2.14 to arrive at a finite number. We will often refer to this number as the stochastic dimension of the RF.

Once we've arrived at our approximation for α, for which the stochastic constituent is expressed in terms of a finite-dimensional vector, $\{\eta_i\}$, we can extend our strategy for one-dimensional PCEs to a multi-dimensional one by applying the same conceptual path to this VRV that we applied in Section 2.2.2 to arrive at a functional relationship to VRV of independent Gaussian RVs, $\{\xi_i\}$. As for the sense of convergence, it is as we saw in Section 2.2.1, we have a two stage path: First we equate the vectors $\{\eta_i\}$ and $\{\xi_i\}$ in probability, followed by an almost sure equality to a VRV that we will approximate in an n_{kl}-dimensional generalize Fourier expansion in multidimensional Hermite polynomials [12].

The matter of what criteria are appropriate for selecting the stochastic dimension for α is one of active research. Often one will use estimates for a dimension n_{kl}. One example of this is to require that $\sum_{i=1}^{\mu} \lambda_i / \sum_{i=1}^{N} \lambda_i$ is sufficiently close to one for some ad-hoc large number, N.

Certainly there are open questions regarding how to pick the number N, but there are other issues as well. These include how to account properly for the effects of any operator transforming our RF, or for properties of the underlying function spaces in the domain and range of a given operator, for example, when the deterministic problem is restricted to as space such as H^1 as opposed to L_2.

Often one encounters a dimension reduction in the output quantity of interest, and the question is, can one effect a dimension reduction that accounts for this. Of course this is a possibility, and it underscores the fact that this issue is not found solely in the realm of RFs, it can affect any situation where one is analyzing transformations of VRVs. We discuss this briefly in the next Section.

2.2.3.1 Stochastic Dimension Reduction

We can readily demonstrate some of the aspects of dimension reduction with an RV and a low-order VRV. Suppose that have an input VRV of dimension two, $\mathbf{S} = \{S_1, S_2\}^T$, where T is the symbol indicating a matrix transposition operation, and we assume that this VRV gets transformed to a one-dimensional output via an algebraic mathematical operation. That is we have $\hat{W} = g(S_1, S_2)$. Now, we can see that $\hat{W}$ has stochastic dimension two, but it is also a scalar entity, an RV, so it would make sense to try to compute its CDF, $F_{\hat{W}}(w)$; and, if we were looking to develop and approximation in the vein of Section 2.2.1.1, specify an appropriate RV-measure pair and use this to find the suitable secondary transformation to set up the affiliated expansion.

Let's say that we were going to develop a one-dimensional PCE expansion for $\hat{W}$. Then, assuming that $g(\,\cdot\,,\,\cdot\,)$ has the properties necessary for $\hat{W} \in L_2(\mathbf{R})$ and following our strategy, we'd use an intermediate RF, $\bar{W}$, and determine the transformation such that

$$\bar{W} = G(\xi) \tag{2.2.16}$$

where $\bar{W}$ now has stochastic dimension one. Note that in Eq 2.2.16 ξ is a new, standard normal RV, and that this is true even if the original VRV, $\mathbf{S}$, was itself composed of standard normal RVs.

Since we know that our original RV, $\hat{W}$, exists, we can conceptually guarantee that $G(\,\cdot\,)$ in Eq 2.2.16 exists and that $\hat{W}$ and $\bar{W}$ meet the criteria for equality in distribution.

Finally, as before, we define yet another RV $W = G(\xi)$ where we require that the equality be understood to be in the almost everywhere sense so that we have them defined on the same probability space. Then, we invoke our L_2 assumption and build a PCE

$$W(\xi) = \sum_{i=1}^{\infty} w_i \Gamma_i(\xi) \tag{2.2.17}$$

and carry on with our usual approximation process.

Now, all of the above discussion demonstrates implicit reasoning; there is still the matter of how to determine $G(\,\cdot\,)$ in Eq 2.2.16 when we don't have an analytical expression of the CDF of $\hat{W}$.

There is also the issue of uniqueness of representation for $\hat{W}$, which we can readily demonstrate by assuming that it is a Chi-Square distributed RV of degree of freedom two. That is,

$$\hat{W} = \xi_1 + \xi_2 \ . \tag{2.2.18}$$

Thus $\mathbf{S}$ consists of two standard normal RVs by definition, and we would have an analytical expression for its CDF allowing us to determine $G(\,\cdot\,)$ analytically. This would result in an infinite-order one-dimensional expansion for W as in Eq 2.2.17.

Contrast this with the known expansion of (yet another) RV $\tilde{W}$ with a two-dimensional expansion in a pair of standard normal RVs where the expansion, as we see from Eq 2.2.18,

is known to be of order two. We know that $\hat{W} = \tilde{W}$ everywhere by definition, and that $\hat{W} = W$ in probability and must be used thereafter in recognition of this fact, but the key issue here is that the one-dimensional expansion has an infinite number of non-zero terms and the two-dimensional expansion has only two.

All of this serves as a stark reminder that there is often no set formula for determining how best to approach the issue of selecting an optimal stochastic dimension, and, while we've demonstrated this fact by this simply low-order example, the fact can only be more complicated for the case of RFs.

2.2.4 Epistemic Uncertainty Formulation

No where in the preceding discussion do we address the specifics associated with how to arrive at a means to connect our random functions to observed data. There's a good reason for this: It's a difficult task not readily amenable to doing so. In this section, we discuss some of the possibilities using the functional approach.

2.2.4.1 Modeling Data-Related Uncertainties

The fact is, we are always limited in our ability to acquire sufficient data to develop a complete probabilistic picture of a given random function. For RVs, we can illustrate this fact by the existence of such classical probability limit theorems as the weak and strong laws of large numbers. The story doesn't get better for more general cases of RFs, or even of data that we acquire amidst a variety of errors, such as measurement error or though fitting of raw data to intermediate test-derived models.

One viewpoint is to approach these issues by modeling them as so-called epistemic uncertainties; that is, uncertainties in our knowledge. While there is much more to accomplish, we have had success doing this with the functional approach to modeling RVs, VRVs and RFs.

2.2.4.2 Augmenting Representations: Basic Product Space Approach

Here we describe our epistemic uncertainty approach, but first we note that the functional approach that we have been describing in this document is predicated on a simple mathematical concept: Product space extensions of original function domains [23] and associated function product spaces [46]. For example, in Eq 2.2.13, the $\mathbf{D} \times \Omega$ is a product space extension of the original deterministic domain, $\mathbf{D}$, to include a probabilistic sample space as a constituent. Our approach to modeling epistemic uncertainty will follow a similar path.

Our key assumption is that each measured quantity, j, can be represented with an $L_2(\bar{\Omega})$ random variable, Y_j, defined on a mathematical product space, $\bar{\Omega} = \Omega \times \Omega_1$ where Ω is the

original RV domain and Ω_1 is the sample space that represents the model for the reducible uncertainty, such as due to the restriction to finite observations for the j^{th} RV. We can write the PCE for Y_j as

$$Y_j(\xi_j, y_j) = \sum_{i=1}^{\infty} y_j(\omega_1)\Psi_i(\xi_j(\omega)) \; . \tag{2.2.19}$$

In Eq 2.2.19, note that we are abstracting the explicit connection for both sample spaces, Ω and Ω_1, into the real line, $\mathbf{R}^1$, and considering measure triples defined for each as dictated through the associated distributions, μ_{y_j} and μ_{ξ_j}. Thus, for Ω, for example, we have shifted our description from the original probability triple, $(\Omega, \mathscr{S}, P)$, into the the measure triple $(\mathbf{R}^1, \mathscr{B}, \mu_\xi)$ where $\mathscr{B}$ is the Borel σ-algebra of subsets of $\mathbf{R}^1$ generated by the open intervals and $\mu_\xi(B) = P(\xi^{-1}(B))$ for all sets $B \in \mathscr{B}$. We note, as we did in Section 2.2.1.2, that this abstraction constrains us to an understanding of equality and, for finite-order series approximations, convergence in distribution, but allows us the freedom to neglect mundane issues relating to the physical mechanisms that comprise the true source of the uncertainties. The case for y_j is handled in a similar manner.

Once we have this framework, we can proceed to associate the measurements with the series descriptions. One method for doing this, albeit a crude one, is to use the data to develop estimates for certain statistical quantities, such as means, correlations or higher-order moments, and perhaps other probabilistic quantities, such as exceedance probabilities, and account for the uncertainties due to finite samples in these estimates by specifying them as intervals. We then use these intervals as constraints for determining data-compliant sets of RVs. We can develop a family of possible RVs that can't be eliminated based on the available information by considering a number of instances of Eq 2.2.19 by taking a variety of series lengths by selecting numbers, n_{y_j}, as the upper limit of the summation in Eq 2.2.19, and sampling the y_j according to some assumed probability law, such as a uniform distribution. For each member of this family, one computes the statistical or probabilistic quantity and tests them all versus the set of constraints derived from the data. If the constraint set is satisfied, we keep it in the subset of the family that we cannot reject based on the data, and any auxiliary assumptions we make to set the bounds on the intervals. Once we've acquired a "large" number of members in this non-reject collection, we consider the data "characterized."

One perspective taken by some researchers [14] is to consider the initial distribution assumed for y_j in Eq 2.2.19 to be Bayesian priors, and to perform Bayesian updating on these distributions as new data are acquired. This connection shows promise in making the determination of members of the non-reject set less computationally burdensome, and could facilitate the calculation of indicators for sensitivity to the acquisition of additional data.

Obviously, there is more work to do to make all of this more rigorous, but the early successes based on the above approach have been promising [34, 36]. The connection to the field of theoretical statistical experiments [8] is obvious, and, specifically, to a function analytic approach to statistical experiments [7] may provide such a path.

2.3 Concluding Remarks

The recent resurgence in popularity of functional approaches to probability has given rise to some misunderstandings concerning some of the practical aspects one should consider when using them. Motivated by the general case of building and using approximations for random fields, we addressed several of these as they pertain to the various key building blocks for doing so.

Among these were establishing the existence of and describing methods for transforming between various random variables and vectors of random variables. We used these new random variables to develop expansions, using generalized Fourier series, which are based on function space criteria such as presence in the Hilbert space, L_2. We truncated these series expansions to arrive at our approximations.

Then we addressed issues relating to different senses of equality and convergence for random variables and approximations to them. The key point there was to understand that our series approximations converged "in distribution" and described ways to use these approximations that avoid known shortcomings. In particular we warned of the danger of using statistical moments inappropriately, since additional criteria beyond the existence of an approximation in a given sense must be satisfied to guarantee that the distribution of the associated random variable can be determined by its moments.

Finally, we described one approach to connecting to limited data via an epistemic formulation built on series approximations, and discussed the connection that this approach has to the field of statistical expeiments.

Chapter 3

Steady-State Density Estimation

3.1 Introduction

Density estimation is a standard and ubiquitous problem in statistical inference, wherein one seeks to characterize a random quantity from a limited number of observations. Observations may themselves reflect additional measurement error.

General outline:

- Introduce and review the density estimation problem in statistical inference. Why is it important and where does it arise. Mention our system reliability problem, among others.
- Recall various methods for semi-parametric or non-parametric density estimation: kernel density estimation, orthogonal series (Bernstein polynomials), what else?
- *Bayesian* density estimation. In particular, Bayesian nonparametrics. Key modern methods: Dirichlet process mixtures, Polya trees.
- Alternatively, we seek to represent random variables directly, rather than their probability densities. Why would one want to do this; what are the advantages? Functional representations of random variables or processes.
- Leads naturally to polynomial chaos expansions. Review this notion, its applications to engineering and physical systems, and its range of validity. (Cite Janson for a precise statement on stochastic dimension.)

  Polynomial chaos (PC) expansions represent random variables and processes in terms of orthogonal polynomial functionals of i.i.d. random variables. In principle, any finite-variance random quantity can be written in this form. Projection onto PC bases underlies efficient and flexible methods for propagating uncertainty through ordinary differential equation and partial differential equation models.

- Set up the need, and describe what we'll do: using polynomial chaos expansions as an infinitely parametric representation of a random quantity. Estimating these with a hierarchical Bayesian construction. Exploit the connection between polynomial roots and the resulting density. Varying the polynomial degree provides a natural space of nested models, which we traverse with reversible-jump MCMC. Univariate and multivariate cases. Examples in systems reliability.

3.2 Formulation

We begin by introducing essential building blocks of our formulation: blah, blah, and blah.

3.2.1 Polynomial chaos expansions

Let $(\Omega, \mathscr{U}, P)$ be a probability space, where Ω is a sample space, $\mathscr{U}$ is a σ-algebra over Ω, and P is a probability measure on $\mathscr{U}$. Let G be the Gaussian Hilbert space spanned by $\{\xi_i\}$, a set (finite or infinite) of independent standard normal random variables defined on Ω [**?**]. Then any square-integrable random variable $X : \Omega \to \mathbb{R}$ that is measurable with respect to $\mathscr{U}(G)$, the σ-algebra generated by G, has the following representation:

$$\begin{aligned} X(\omega) &= a_0\Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\xi_{i_1},\xi_{i_2}) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\xi_{i_1},\xi_{i_2},\xi_{i_3}) + \cdots \end{aligned} \tag{3.2.1}$$

In other words, $L^2(\Omega, \mathscr{U}(G), P)$ has the Wiener chaos decomposition

To be general, introduce the overall Askey scheme, not just Gauss-Hermite.

Polynomial chaos expansions underlie many efficient and flexible methods for propagating uncertainty through ordinary differential equation and partial differential equations.

Here, we will focus on the univariate case.

3.2.2 Bayesian estimation of polynomial chaos representations

Describe the Bayesian estimation of PC coefficients, i.e., the joint posterior density of the PC coeffs $p(a_k|\mathscr{D})$.

Next, extend to noisy observations; this is our hierarchical Bayesian scheme.

Discuss prior modeling: Gaussian prior on the mean, and inverse-Gamma prior on the variance. Show how inverse-Gamma prior on the variance leads to a joint prior on the PC coeffs $\{a_1, \ldots, a_p\}$. Include proper normalization, as we'll need it later for model comparison.

3.2.3 Likelihood evaluations

Show how the density $p(\mathscr{D}|a_k$ can be obtained by polynomial solving.

Discuss the Stürm sequence polynomial solver, with appropriate references.

This is faster and more accurate than KDE, which others have attempted before.

Include the formula for the coefficients of the modified Hermite polynomials.

Initial present univariate case; generalize to multivariate (polynomial systems).

3.2.4 Model selection with reversible-jump MCMC

Describe how we apply RJ-MCMC to this problem. Detail the construction of up-moves (currenty stochastic) and the corresponding down-moves (currently deterministic). Show the acceptance probability.

Explicitly outline the steps of our overall sampler, perhaps in some kind of pseudo-code. This should encompass the hierarchical scheme for errors-in-observations.

3.3 Results

Maybe call this section "Demonstrations" or something like that?

Present a series of examples of increasing complexity.

3.3.1 Simplification to the normal-conjugate case

Restrict to $p = 1$ and compare the numerical results to analytical expressions for the posterior distribution of the mean and variance of a normal distribution (assuming conjugate priors). They should match. *Is this even worth doing? It's so simple!*

3.3.2 Univariate PC

(Better title later.)

Generate data from a PC expansion; estimate the density using the full variable-order formulation. Try the following:

- Vary the amount of data (small/large).
- Generate data from a high-degree PCe versus a low-degree PCe, and see what happens with large/small amounts of data.
- Pick a PCe and a certain amount of data, then compare the case of *no* observational error to two cases of successively larger observational errors.
- What is the impact of the prior hyperparameters (e.g., α and β of the inverse-Gamma prior on the variance)?

In all these cases, plot the following:

- Posterior probability over models (i.e., bars to represent probability masses).
- Mean density, maybe median density, other sampled densities.
- Distribution of probability mass between certain fixed bounds (some expected to be in the center, some in the tails). These are more general than just a single definition of reliability.
- Alternatively, plot the posterior distribution of a certain quantile (e.g., where does the 20% quantile lie, where does the median lie, etc).

3.3.3 Univariate non-PC

Now generate data from something *other* than a PC expansion— for instance, a Gaussian mixture or a mixture of t-distributions. Try something not-very bimodal; should work ok. Something strongly bimodal; won't work very well at all, even at high order. Discuss limitations. Mention how piecewise polynomials and domain partitioning should be used here, and can be done with reversible jump.

Should we compare directly with other Bayesian methods (e.g., DPMs?) No; too much of a distraction.

3.3.4 Multivariate

Try a 2-D example with the poly system solver

3.3.5 Density estimation in hierarchical engineered systems

Formulate the large system, and show how arbitrary nonlinear functions can be accommodated.

3.4 Discussion

Strengths and limitations; broader context [cover the statistical and systems-reliability literatures thoroughly before writing this].

What to do with all this information, in a predictive simulation context? For one, it can be collapsed into a single PC expansion of larger stochastic dimension. Must we generate PC expansions for the coefficients themselves via the inverse Rosenblatt transformation? [Any references?]

What about different polynomial chaos bases (e.g., Jacobi-Beta rather than Gauss-Hermite)? What about wavelets?

What's involved in moving to multivariate density estimation? Point to Maurice's results on polynomial system solving.

Chapter 4

The Time-Dependent Problems

4.1 Introduction

In previous work [26] the authors introduced a method to propagate the effects of testing through a system in the steady-state case to assess the reliability of that system. More generally, we developed a technique to perform density estimation and thus we obtained a distribution of the reliability rather than just an estimate or some other statistic. This allowed us to state a confidence in the reliability, or answer more general statistical questions. In this paper, we extend these results to the time-dependent case.

As in [26] we want to allow general continuous data to be collected, but now we want this data to be collected at specified times, t_i. Given that we can obtain such data, we seek to understand how the system reliability is changing over time and to predict the reliability of the system at some future time. This would provide a powerful tool for understanding the effects of aging in systems or in the components of systems. To accomplish this, we must expand the model used in [26] to allow time dependent components and we must expand the algorithm to deal correctly with this model. In section 4.2 we describe the basic model and notation for the steady-state case. In section 4.3 we describe how the test data are collected and provide the notation for general testing. Then is section 4.4 we develop our extension to handle general nonlinear time-dependent models. These models, however, are in only one stochastic dimension; in 4.7 we discuss extensions that allow higher stochastic dimensions. In section 4.5 we describe in some detail the algorithm that use. In particular, we note that the computation of the likelihood function is the most expensive part, and we derive a means of factoring this function so that it can be rapidly computed in parallel. We also consider other details of the algorithm. In section 4.6 we show the results of our algorithm on a representative examples of time-dependent systems and show that our procedures and algorithms are effective. Finally, in section 4.7 we consider several extensions.

The significance of these results is that we have provided a powerful tool to understand and quantify the uncertainty in the reliability (or other function) of time-dependent systems

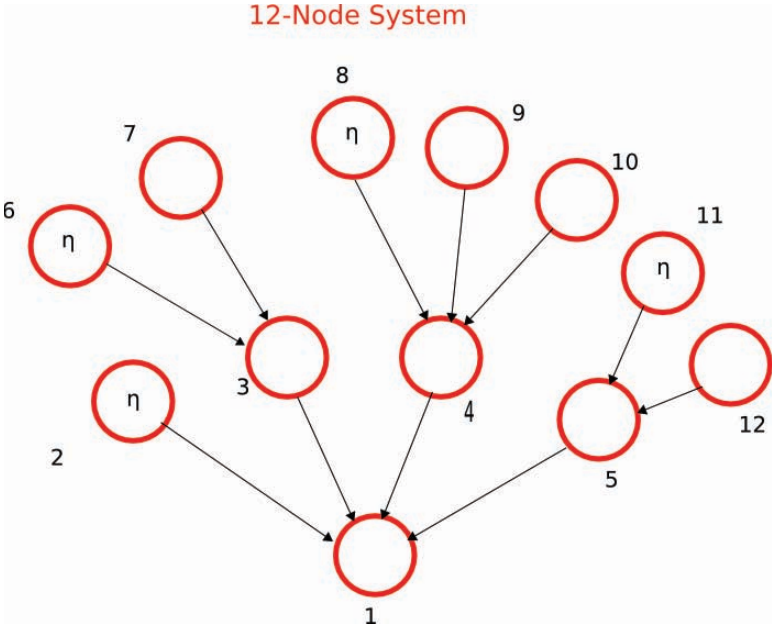


Figure 4.1. 12-Node system with leaf nodes 2 and 6–12. The nodes labeled η are the error nodes.

based on test data. Our framework encompasses all forms of epistemic uncertainty, makes no assumptions on the distributions of the random variables (RV) involved, and allows general nonlinear models. The algorithm is based on Bayesian analysis and uses Markov chain Monte Carlo (MCMC) to sample the posterior distribution. This work has been extended in [38] by using Bayesian decision analysis to develop optimal testing strategies.

4.2 The System Model

Our basic model is a "system of systems" model represented by a tree structure where the root node is the entire system and its parents represent the major components of the system. Each component is thought of as a system; this can be taken to an arbitrary level. Figure 4.1 shows a 12-node example of such a system. We assume that each leaf node, nodes 2 and 6-12, produces an output, represented by a random variable (RV), that is transmitted to its internal, or system, node. Each system node combines the information from its parents to produce an output that is then transmitted down. Eventually, node 1 produces the ultimate output of the system. The output is, of course, a RV and we deem it acceptable if it is between specified upper and lower bounds. The percentage of times that the output is acceptable is defined as the reliability of the system.

We take a probabilistic approach to these systems and assume that the output of each leaf node is a random variable (RV). As noted above, we wish to avoid distributional assump-

tions. Thus we use a polynomial chaos expansion (PCE) to represent the RVs at the leaf nodes. To fix notation, let the output of leaf node j be given by

$$r_j = \sum_{i=0}^{\infty} \theta_{ij} \Gamma_i(\xi_j) \tag{4.2.1}$$

where the θ_{ij} are real coefficients, Γ are (orthogonal) Hermite polynomials, and $\xi \sim \mathcal{N}(0,1)$, i.e., normally distributed with mean 0 and variance 1. It can be shown (see [35]) that any RV with finite variance can be represented in such a series. This representation can be extended; see [35] for relevant details. Obviously, however, we cannot do computations with an infinite series, so we must truncate the series. Let q_j be the order that we use in r_j, yielding in place of (4.2.1)

$$r_j = \sum_{i=0}^{q_j} \theta_{ij} \Gamma_i(\xi_j). \tag{4.2.2}$$

Each system node combines the information from its parents. Let p_j be the set of parents of system node j. Then the output of system node k is given by

$$r_j = \Phi_j(p_j),$$

where Φ_j is an arbitrary function.

4.3 The Test Data

We assume that tests can be taken at specified subsets of the nodes. A major assumption is that the nodes labeled with η are special "error" nodes that cannot be directly observed. Instead, these nodes allow for the possibility that the system nodes have an certain error associated with them. The testing that we do aims to characterize these errors as well as to estimate the reliability

The notation that we use for the tests is as follows: Let

$$N_i = \{n_{i1}, \ldots n_{ik}\}$$

be a set of k nodes at which simultaneous measurements are taken. The index i goes over all allowable tests. This enables us to specify which tests can be done simultaneously. For example, doing a "full-system" test, i.e., testing at node 1, may not permit obtaining a measurement at a higher level component, e.g., node 12 in figure 4.1. The data we obtain is then of the form

$$d_j(t, \rho_j)$$

where t is the time at which the test was taken and ρ_j is the number of replications. A set of the d_j, call it $\mathscr{D}$, is the test suite that we will use to update the state of our system.

4.4 The Time-Dependent Case

We briefly review the steady-state formulation and then show how to extend it to the time-dependent case. Let $\boldsymbol{\theta}$ be the collection of all of the θ_{ij} from all of the PCEs at the leaf nodes as given by (4.2.2). Then, what we want to compute is the joint distribution of $\boldsymbol{\theta}$ given the test data $\mathscr{D}$. We use Bayes' Theorem to obtain

$$\pi(\boldsymbol{\theta} \mid \mathscr{D}) \propto \pi(\mathscr{D} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}), \tag{4.4.1}$$

where $\pi(\boldsymbol{\theta} \mid \mathscr{D})$ is called the posterior, $\pi(\mathscr{D} \mid \boldsymbol{\theta})$ is the probability of the data given the parameters $\boldsymbol{\theta}$ (also known as the likelihood), and $\pi(\boldsymbol{\theta})$ is the prior. We need to discuss the computation of each of these in the context of updating our assessment of the reliability. First, however, we want to provide a high-level understanding of what are proposing.

We desire to compute a distribution of the reliability of the system. Given an instance of $\boldsymbol{\theta}$, we can sample each of the ξ_j and, using the appropriate part of $\boldsymbol{\theta}$, compute a sample of the output at each of the leaf nodes. We can then propagate this through the system to obtain a sample output at node 1. Finally, we can then easily compute the percentage of these values that fall between the specified upper and lower bounds to obtain an estimate of the reliability. To compute the distribution of the reliability, we propose sampling the posterior distribution given in (4.4.1) to obtain a list of $\boldsymbol{\theta}$ values. For each of these samples, we can compute an estimate of the reliability as just described. This has been completed for the steady-state case and results are reported in [26].

For the time-dependent case, there are some changes that are necessary. The general framework for the time-dependent problem is given in [35] where a general Karhunan-Loève (KL) expansion is used to represent a stochastic process. That is, at node j we have a stochastic process represented by

$$\boldsymbol{\theta}_j = \sum_{i=0}^{\infty} \rho_i(t)\gamma_i(\omega) \tag{4.4.2}$$

where the γ_i are RVs on some probability sample space Ω with $\omega \in \Omega$, and the $\rho_i(t)$ are the eigenfunctions of the covariance operator of $\boldsymbol{\theta}_j$, i.e., for

$$C(t,t') = E(\boldsymbol{\theta}_j(t,\omega)\boldsymbol{\theta}_j(t',\omega))$$

the functions ρ_i satisfy

$$\int_T C(t,t')\rho_i(t')dt' = \lambda_i\rho_i(t).$$

Here we use just one term of the expansion in (4.4.2) and can thus represent a simple time-dependent model that has the form of (4.2.2) where each of the coefficients is now a time-dependent function. We obtain

$$r_j = \sum_{i=0}^{q_j} \theta_{ij}(t)\Gamma_i(\xi_j). \tag{4.4.3}$$

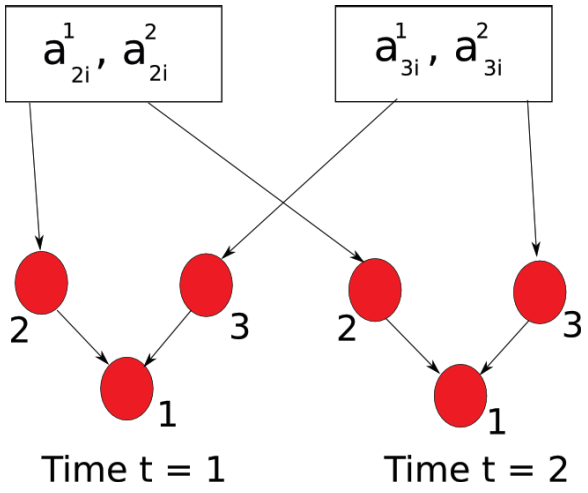


Figure 4.2. Time-dependent system showing two times. The coefficients are the same for both; this is indicated by the arrows that go from the coefficients to each instance of the system.

Here we assume that the functions $\theta_{ij}(t)$ depend on parameters that we will be trying to estimate. For example, the function may be linear:

$$\theta_{ij}(t) = \theta^1_{ij} + \theta^2_{ij} t,$$

where we now wish to create a distribution for θ^1_{ij} and θ^2_{ij}. We illustrate the general situation in figure 4.2 which shows that tests can be taken at time $t = 1$ and/or time $t = 2$ and that we are obtaining distributions of the coefficients θ^1_{ij} and θ^2_{ij} for $i = 2,\ 3$.

If, as for the steady-state case, we let $\boldsymbol{\theta}$ be the collection of all of the coefficients, now including the time-dependent ones, we have the same situation as above. Bayes' Theorem, (4.4.1) still holds and samples taken from the posterior can be used in the same way. But here we can propagate the results through the system at both $t = 1$ and $t = 2$ to obtain the reliability distributions at both times. Furthermore, we can predict the reliability distribution at a future time.

We are now ready to discuss the algorithm for sampling the posterior. A naive implementation will, in large systems, be computationally intractable, so care must be taken.

4.5 The Algorithm

The basic problem in sampling the posterior is that there is no closed-form expression for it. Thus the sampling procedure that we use is a slight modification of the Markov chain Monte Carlo (MCMC) technique used in [26]. A powerful feature of that technique is the

fact that it uses a "reversible-jump" strategy that allows it to simultaneously sample the orders, q_j as well as the coefficients $\boldsymbol{\theta}$. Thus the test data are used to determine the order and the coefficients and frees the user of having to make what would be hard-to-justify decisions on the order. Of course, the data have to support the order used; there is no point in attempting to estimate a high-order approximation with very little data.

4.5.1 The MCMC Procedure

The basic MCMC algorithm that we use is the Metropolis-Hastings procedure, but with the reversible-jump capability so that we can examine various orders. The MCMC algorithm is only aware of the basic structure of the application. In particular, it knows is the number of leaf nodes, an initial specification of the order q_j of the PCE at each leaf node, an order range, $[q_j^L, q_j^U]$, for each leaf node, and an initial vector $\boldsymbol{\theta}_0$ of the coefficients at each leaf node. It is initialized by computing the values of the likelihood and the prior at the initial conditions.

Each iteration of the MCMC process is as follows:

1. for each leaf node
   (a) Decide whether to attempt up-order move, down-order move or within-order move
   (b) If up-order move
      i. Increase the order by 1
      ii. Generate new, random coefficients and appropriate scaling parameters
      iii. Evaluate the likelihood function and the prior
      iv. Compute the jump acceptance probability (call it η)
      v. Choose $\mu \sim \mathscr{U}(0,1)$
      vi. Accept if $\eta \geq \mu$
   (c) If down-order move
      i. Decrease the order by 1
      ii. Evaluate appropriate scaling parameters
      iii. Evaluate likelihood function and the prior
      iv. Compute the jump acceptance probability, η
      v. Choose $\mu \sim \mathscr{U}(0,1)$
      vi. Accept if $\eta \geq \mu$
   (d) If within-order move
      i. For each coefficient in leaf-node PCE
         A. Propose a move
         B. Evaluate the likelihood and prior

C. Compute acceptance probability, η
D. Choose $\mu \sim \mathscr{U}(0,1)$
E. Accept if $\eta \geq \mu$

Some discussion of these steps is in order.

4.5.2 Within-Order Move

The within-order move is standard. For each coefficient of the PCE, (this includes all of the coefficients of the time-dependent model) we generate a proposed new value as follows: Let ω be a parameter specifying the proposal width. Let

$$\alpha \sim \mathscr{N}(0,1)$$

Then set

$$\text{aProposed} = \text{aProposed} + \alpha * \omega.$$

We then compute the product of the likelihood and the prior and compare it to the existing value in the standard manner. To be specific, let R be the value of the product of the likelihood and the prior before a proposed change in a coefficient and let R^+ be the value after the change. Choose $\mu \sim \mathscr{U}[0,1]$. Then accept the proposed move if

$$\frac{R^+}{R} \geq \mu.$$

4.5.3 Up-Order and Down-Order Moves

We use a straightforward procedure for these moves. As noted above, we provide a range of possible orders for each leaf node. If the algorithm is currently not at either end of its range, then the probability of going up, going down, and staying put is 1/3. If it is at either end point, then the probability of moving into the range is 1/2 and the probability of staying put is 1/2. We generate a uniform random number in [0,1] to determine what to do.

Let t_j be the number of coefficients in the time-dependent model at leaf node j. Then to perform an up-order move, we do the following. Set the new order to be $q_j = q_j + 1$ For each of the new t_j coefficients we choose a random number from $\mathscr{N}(0,1)$ and scale it by $1/(q_j!)$. I am looking for a reference. Youssef: what did you use? Is this the right scaling for adding more that one coefficient?

Once an up-order move has been calculated, we evaluate the likelihood and the prior, but we have a more complicated version of the test for acceptance that takes into account the fact that the dimension of the space was increased. This is done as follows: Again, I need a reference.

4.5.4 Evaluation of the Likelihood

We now consider the evaluation of the likelihood function $\pi(\mathscr{D} \mid \boldsymbol{\theta})$. By taking advantage of the structure of the system that is being tested, we can factor the likelihood function into smaller pieces that can be calculated more quickly. In fact, the pieces can be evaluated in parallel, providing the possibility of even further increases in efficiency.

The best way to see the procedure is from an example. Suppose that for the system in figure 4.1 we have data at nodes $[1,3,5,7,12]$. Recall that Φ_1 is the function that combines the output of the parents of node 1. Then we obtain

$$\delta_1 = \Phi_1(v_2, \delta_3, v_4, \delta_5),$$

where δ_j is the data at node j and v_k denotes the general value at node k. We continue to expand the general values (the v_k) until we reach leaf nodes where the $\boldsymbol{\theta}$ are defined. Thus we obtain

$$\delta_1 = \Phi_1(v_2, \delta_3, \Phi_4(v_8, v_9, v_{10}), \delta_5),$$

at which point we have expressed δ_1 as a function of the other data that has been collected or of the leaf node coefficients. Now, we can see that one factor of $\pi(\mathscr{D} \mid \boldsymbol{\theta})$ is given by

$$\pi(\delta_1, \delta_3, \delta_5 \mid \theta_2, \theta_8, \theta_9, \theta_{10}). \tag{4.5.1}$$

We discuss the numerical evaluation of this below. First, we note that the likelihood must also take into account that the data at node 3 allows us to compute the likelihood in terms of δ_7 and θ_6 as follows:

$$\pi(d_3, \delta_7 \mid \theta_6). \tag{4.5.2}$$

Finally, with data at the leaf nodes δ_7 and δ_{12} we can compute

$$\pi(\delta_7 \mid \theta_7) \tag{4.5.3}$$

and

$$\pi(\delta_{12} \mid \theta_{12}). \tag{4.5.4}$$

Thus to compute $\pi(\mathscr{D} \mid \boldsymbol{\theta})$, we need to compute the product of the likelihoods in (4.5.1 – 4.5.4).

This example, though simple, illustrates the general procedure of factoring the likelihood function. We conclude this section with the specifics of how we evaluate each of the factors. We first consider the evaluation of (4.5.3) (the evaluation of (4.5.4) is the same).

4.5.5 Evaluation of the Priors

From the above, it is clear that priors are only specified at the leaf nodes; we allow the prior at each leaf node to be set independently. Thus we can write

$$\pi(\boldsymbol{\theta}) = \pi(\theta_1)\pi(\theta_2)\cdots\pi(a[k])$$

where k is the number of leaf nodes. In all of our examples we used a prior based on the distribution of the mean and the distribution of the variance of the RV at the leaf node. In particular, we assume that the mean of node i is distributed normally with a given mean μ_i and standard deviation σ_i and that the variance has a gamma distribution, (G), with shape parameters α and β. Now the variance of a PCE is given by

$$var = \sum_{k=1}^{q_i} \theta_k / k!.$$

Thus we can write the prior as

$$\pi(\theta_i) = \mathcal{N}(\theta_i; \mu_i, \sigma_i) * G(var; \alpha, \beta). \tag{4.5.5}$$

This is commonly used in Bayesian analysis, but if other priors are available, they can be easily accommodated.

4.6 Results

We now present some results obtained from our algorithm on some example problems. We do not aim to test the procedure exhaustively, but rather we desire to show how the procedure works on a reasonable variety of test cases that illustrate some of the features. We begin with a simple linear model with a small number of nodes.

4.6.1 Example 1

Consider the system as illustrated in figure 4.2. First, we take node 2 to be the η-node of node 1, i.e., the node that represents the error introduced by node 1. Recall that this node is conceptual, not real, and that we cannot observe this node in testing; we can only infer its properties by testing at nodes 1 and 3. For these tests, we assume that the PCE representing the output at node 2 is of order 1 (i.e., it is a normal random variable). We specify the "true" system distribution to be $\mathcal{N}(0, .01)$ at time $t = 0$. We allow this distribution to drift linearly over time. Thus we have

$$r_2 = \sum_{i=0}^{1} (\theta_{ij}^1 + \theta_{ij}^2 t) \Gamma_i(\xi_j).$$

We take $\theta_{0j}^1 = 0$ and $\theta_{1j}^1 = .01$ so that the initial distribution is, indeed, $\mathcal{N}(0, .01)$. We specify the time drift by taking $\theta_{0j}^2 = .001$ and $\theta_{1j}^2 = .001$. This system is degrading slowly in this component.

For node 3, we specify the true output as a PCE of order 3. Using a linear model for drift over time as we did for node 2, we specify the constant terms as a vector of length 4. Specifically, we set

$$\begin{aligned}\theta^1 &= (1.0, 1.0, .2, .1)\\ \theta^1 &= (1.002, 0.001, 0.002, 0.001),\end{aligned}$$

so that there is quite a bit of drift in this component.

Finally, the output of nodes 2 and 3 are combined at node 1 to yield the output of the system. Here, we take the functional form of this combination to be linear; in particular,

$$r_1 = r_2 + 2.1 r_3. \tag{4.6.1}$$

To assess the reliability of this system, we specify upper and lower bounds on the output r_1. We arbitrarily choose the desired reliability to be 80%. We then sample the normal RVs at nodes 2 and 3 to create a sample of the output at these nodes. These are passed to node 1 and combined as specified by (4.6.1). We then pick upper and lower bounds so that the middle 80% of the output of node 1 is within these bounds.

Next, we have to specify the model that we will use. At node 2, we use a PCE of fixed order equal to 1. At node 3, however, we use a PCE of variable order from 2 through 4. The MCMC routine needs to have a starting value of the coefficients. We use

$$\begin{aligned}\theta^1_{(2\ \cdot)} &= (0.0, .02)\\ \theta^2_{(2\ \cdot)} &= (.0, .0)\end{aligned}$$

for node 2 and

$$\begin{aligned}\theta^1_{(3\ \cdot)} &= (0.00, 1..2.1.01)\\ \theta^2_{(3\ \cdot)} &= (0, 0, 0, 0, 0)\end{aligned}$$

To complete the model specification, we need to provide a prior. As noted above, our priors are on the mean and standard deviations of the $\boldsymbol{\theta}$ associated with each node. We use $N(0,1)$ for the prior on the mean of node 2 and $N(1,1)$ for the prior on the mean of node 3. We use inverse gamma distribution with parameters $(1,1)$ for the distributions of the variance at both nodes 2 and 3.

We now specify the test data that were used to estimate the reliability. We chose to obtain

test data at three points in time, $t = 0, 1, 2$. Using the notation above, the data we use are

$$\begin{aligned}
\delta &= [0, [1], 150] \\
\delta &= [0, [3], 150] \\
\delta &= [1, [1], 150] \\
\delta &= [1, [3], 150] \\
\delta &= [2, [1], 150] \\
\delta &= [2, [3], 150].
\end{aligned}$$

(Recall that we could have taken the data simultaneously at nodes 1 and 3, but this would have required a much longer computation time. As a test, we ran this form with little difference in the final result.) The data are generated by using the true description of the system.

The results of this test are in figure 4.3. where we show the reliability distributions at times $t = 1, 2$, and 3 where the data were collected. The true reliabilities are 0.8010, 0.7060, and 0.2170. One can observe that our distributions capture the true reliability. Jim: What's the right thing to say? Since we are also interested in predicting the reliability at a later time, we calculate the distribution of the reliability at $t = 4$. The true reliability at $t = 4$ is 0.0160. This is plot 4.3(d) in figure 4.3.

We then ran the test again with 300 replications of the data rather than 150. These results are in figure 4.4.

Note that the extra data improve the distributions.

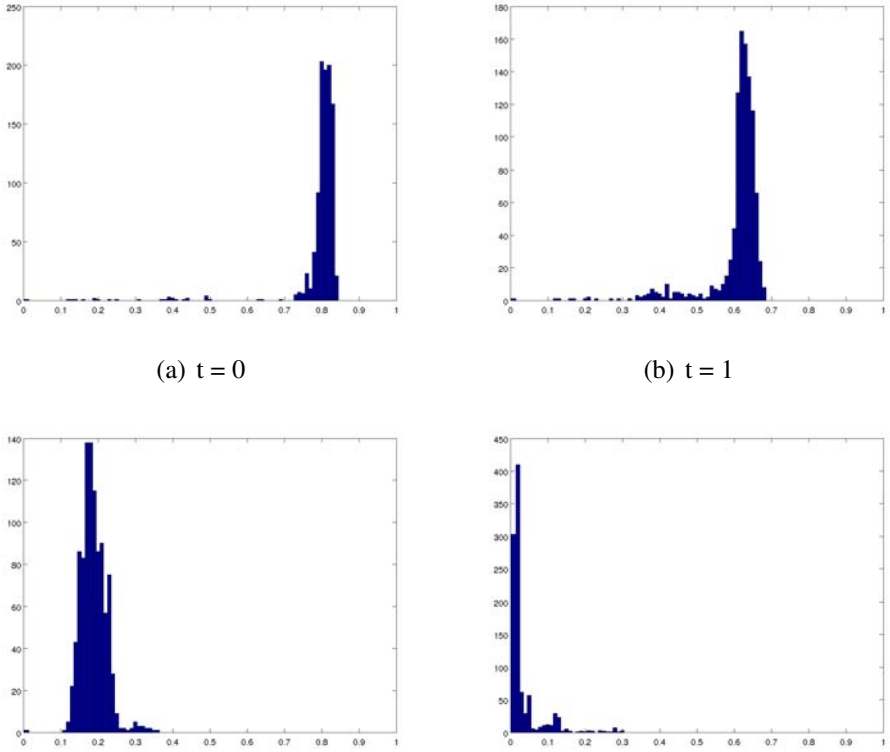
4.6.2 Example 2

In this example, we use a nonlinear model to illustrate the fact that we can solve problems with significant nonlinearities. For the coefficients in (4.4.3) we use

$$\theta_{ij}(t) = \theta_{ij}^1 + \theta_{ij}^2 e^{-\theta_{ij}^3 t}.$$

At $t = 0$, $\theta_{ij}(0) = \theta_{ij}^1 + \theta_{ij}^2$ and as $t \to \infty$, $\theta_{ij}(t) \to \theta_{ij}^1$ with decay rate θ_{ij}^3. Do we have an example?

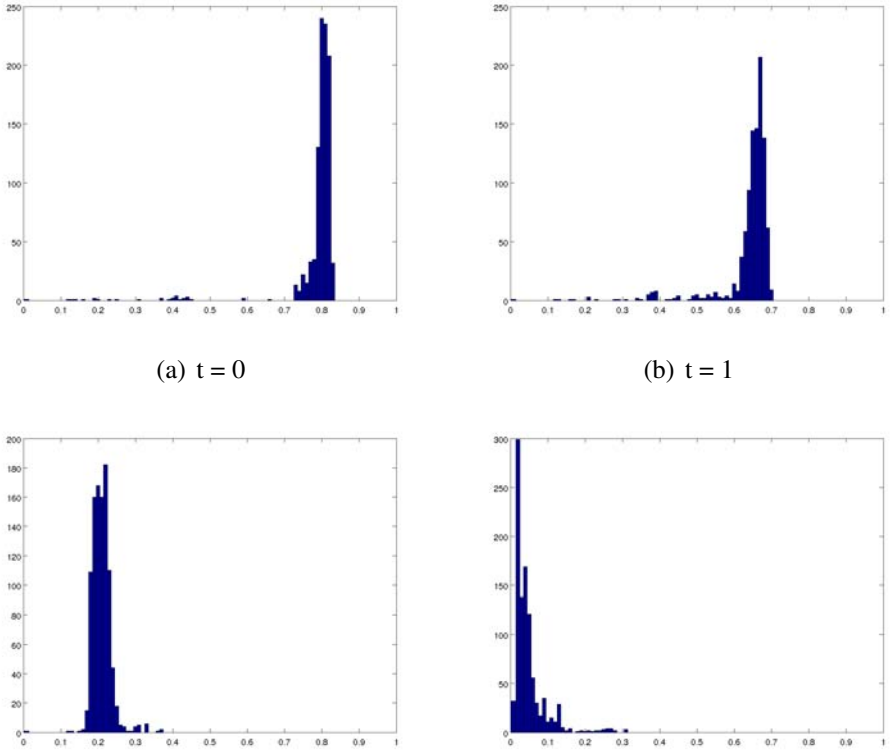
The prior is chosen as follows: At each time step we compute $\theta_{ij}(t)$. We then specify a distribution on the mean and the variance as above and compute the product of these for all of the time steps. If the component is rapidly changing, i.e., if ${\theta_{ij}}^3$ is large, then the prior that is appropriate at $t = 0$ is necessarily poorer as t increases. The usual way to overcome a bad prior is with more data, and so the runs here compare 150 samples at each data node with 300 samples. The data are taken as in Example 1.



(a) t = 0 (b) t = 1

(c) t = 2 (d) t = 3

Figure 4.3. Reliability results for linear time-dependent model with 150 replications of the data.



(a) t = 0 (b) t = 1

(c) t = 2 (d) t = 3

Figure 4.4. Reliability results for linear time-dependent model with 300 replications of the data.

For our tests, we have chosen the true values of the ${\theta_{ij}}^k$ as follows:

$$\begin{aligned}\theta^1_{0,2} &= .02,\\ \theta^2_{0,2} &= -.02,\\ \theta^3_{0,2} &= .7,\end{aligned}$$

and

$$\begin{aligned}\theta^1_{1,2} &= -.02,\\ \theta^2_{1,2} &= -.021,\\ \theta^3_{1,2} &= .8.\end{aligned}$$

As above, these values are used to generate the data. Our starting value also assumed a PCE order of 1 and we used

$$\begin{aligned}\theta^1_{0,2} &= .015,\\ \theta^2_{0,2} &= -.015,\\ \theta^3_{0,2} &= .6,\end{aligned}$$

and

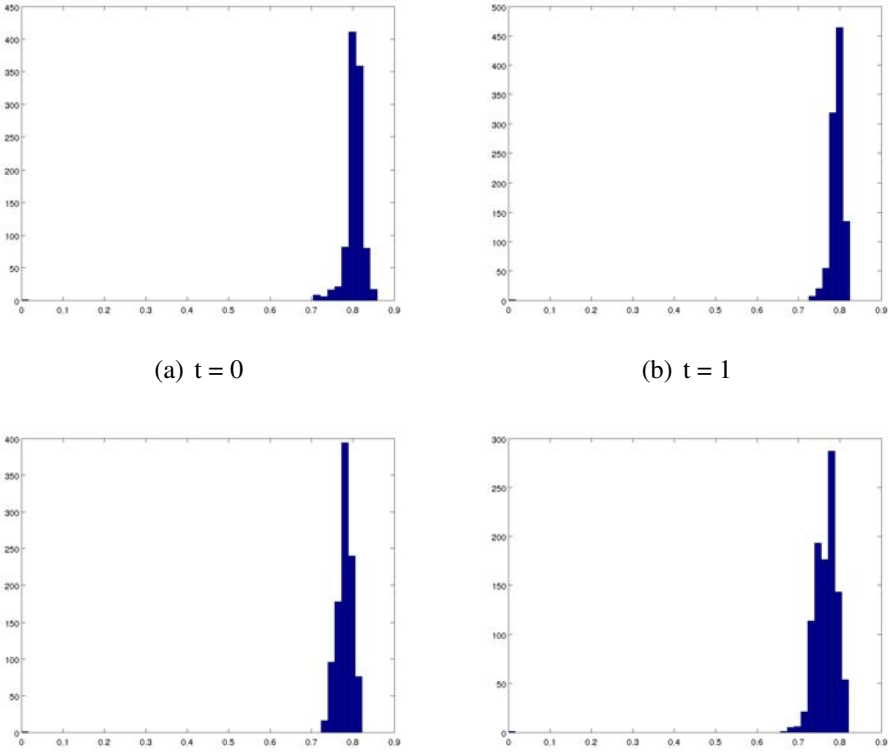
$$\begin{aligned}\theta^1_{1,2} &= -.025,\\ \theta^2_{1,2} &= -.015,\\ \theta^3_{1,2} &= .9.\end{aligned}$$

For node 3, the true PCE order is 3 and the values are ...

The true values of the reliability are 0.8010, 0.7750, 0.7570, and 0.7370 for $t = 0,1,2,3$, respectively. This represents a gradual degradation in reliability over time; the results for 300 data points again appear to be better.

4.7 Conclusions and Future Work

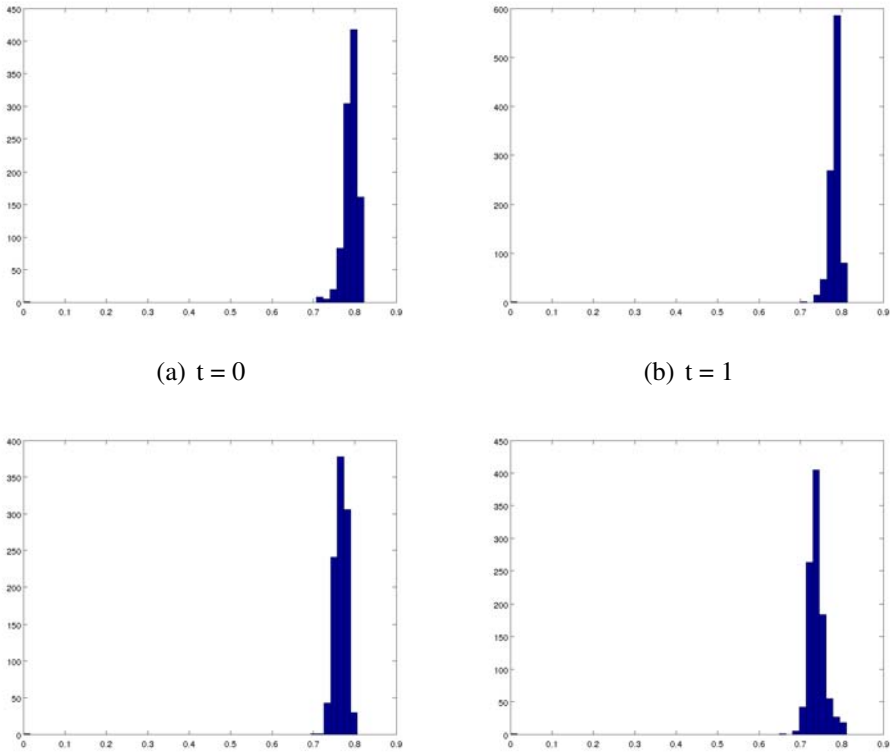
In this paper, we have shown that the ideas for density estimation developed in [26] can be extended in a computationally reasonable way to more general network systems with time-dependent components. We were able to construct distributions of the reliability over



(a) t = 0 (b) t = 1

(c) t = 2 (d) t = 3

Figure 4.5. Reliability results for nonlinear time-dependent model with 150 replications of the data.



(a) t = 0 (b) t = 1

(c) t = 2 (d) t = 3

Figure 4.6. Reliability results for nonlinear time-dependent model with 300 replications of the data.

time and use these to accurately predict the reliability at the times at which data were collected and at future times as well. We developed a version of MCMC that worked well in our tests and we developed computational strategies that allow the algorithm to perform in reasonable times. Nevertheless, there are several areas that need further enhancements.

More terms in KL: again, our framework easily includes this, gives extra stochastic dimensions that are often necessary in stochastic systems. implies more parameters and thus more computationally expensive. Our factorization procedure for the likelihood function can be parallelized. We did some simple examples based on the system in Example 2 above and, for some testing strategies, achieved almost linear speedup by using simple parallel extensions. The issues surrounding parallel computing need to be thought through carefully, but we think that much could be done to enable this approach to apply to much larger systems.

Errors in the test data as Youssef did.

Other structures that are not trees: We have relied on the tree structure of the systems we consider, but this is not strictly necessary. In fact the time dependent model figure 4.2 is not a tree, although at each time step it is. Nevertheless, the important thing is that the system has a network structure with the connection between nodes structured and limited. Thus, even for other networks, we believe that many of our techniques can be adapted. Examples include power grids, water dist networks, and certain biological systems.

Chapter 5

Optimal Testing Strategies

5.1 Introduction

In the previous chapters we have shown how to compute the reliability distribution given a set of data for steady-state and time-dependent cases. In this chapter, we consider the optimization questions that we wish to answer, For example, what is the best choice of data collection tests, given a fixed budget, that will maximally reduce the uncertainty in the reliability estimate? We begin by making some observations:

- The simplest scenario is to assume that we only have a prior distribution and determine which sampling scheme reduces the variance of the reliability distribution the most.
- We first consider only one data collection test. Later we will consider the problem of choosing the best suite of tests. Recall that our MCMC procedure (see chapter 4) produces a set of possible $\boldsymbol{\theta}$ values, and that for each of these values we can compute one value of the reliability.
- Even if we know the values of $\boldsymbol{\theta}$ corresponding to a particular node extremely well, i.e., there is very little variation in the $\boldsymbol{\theta}$ values, it still may be that the component is intrinsically quite variable. In general, the intrinsic variability of a component contributes to the variability of the overall output of the system; large variabilities will generally mean that the composite system is less reliable. This contribution will need to be considered relative to the contribution of the component to the overall system. A large variability in the $\boldsymbol{\theta}$ values of a component will, by contrast, be reflected in a large spread in the reliability distribution.
- The implication of this last point suggests that re-testing a component whose $\boldsymbol{\theta}$ values are already well determined, despite any inherent variability, will not decrease the spread of the reliability distribution. Instead, we will want to concentrate on those components whose coefficients are still widely variable.

- Because a test can take on a wide range of values, we clearly cannot test all of them. Thus, we need to compute some form of an expected effect of a given test on the computed reliability distribution instead.

Our approach is first to develop the ideal optimization problem that we would like to solve, even though it may be computationally intractable. We can then simplify the problem as necessary to obtain a computationally tractable approach. In the next section, we present a formal analysis and derivation based on Bayesian decision analysis.

5.2 Foundational Concepts and Notational Conventions

The following notation and formulation follows James Berger's classic text, *Statistical Decision Theory and Bayesian Analysis* [3]. Formal development of statistical procedures, be they estimation, hypothesis testing, design of experiments, etc., is typically structured as a decision analysis problem. By developing this formalism for our problem, this chapter develops a rigorous approach to the optimization problem. As appropriate we follow Berger's notation, clarifying our use of it and making adjustments to meet the needs of our problem.

This chapter is organized as follows. Section 5.2 works through the basic formalism of Bayesian decision analysis, while Section 5.3 addresses the optimization/design-of-experiments question specifically. Berger calls the latter "pre-posterior" analysis. Section 5.4 formulates the risk function minimization to identify the optimal data collection scheme.

The following notational conventions apply generally to the decision theoretic foundations of Bayesian analysis as used by Berger [3]. This list also introduces the basic formalism of Bayesian decision analysis. Let

- $\boldsymbol{\theta}$ (a vector) denote an unknown state of nature, with elements θ_{ij}, that captures *epistemic* uncertainties (i.e., the variation in the θ_{ij}'s is reduced as more data are collected). In our system, output at each of the n_ℓ "leaf" node (e.g., nodes 7, 9, 10, and 12 in Fig. 5.1; nodes marked with an η are unmeasurable error nodes) is given by a random variable (RV) Ω_j expressed as a polynomial chaos expansion (PCE) [48, 9, 28, 11]:

$$\Omega_j = \sum_{i=0}^{q_j} \theta_{ij} H_i(\xi), \tag{5.2.1}$$

  where q_j is the order of the PCE, H_i represents the ith Hermite polynomial, and ξ_j is a Gaussian RV (also referred to as a normal RV): $\xi_j \sim \mathcal{N}(0,1)$. Subscripts i correspond to the indices for the PCE expansion of the jth node, which begin with 0 for consistency with standard Hermite polynomial notation. Between the leaf and/or "parent" nodes and an intermediate system node k, output is given by a deterministic

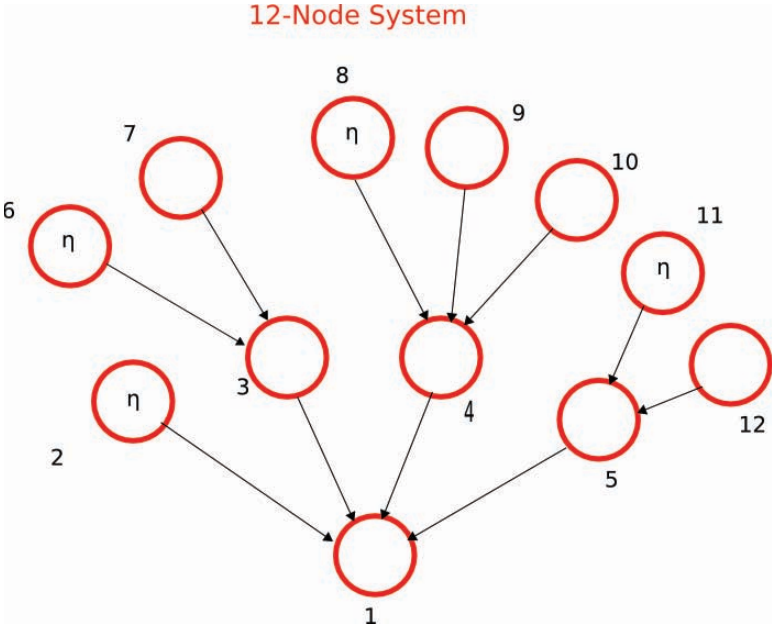


Figure 5.1. Graphical representation of a 12-node example model of our system of systems. Nodes marked η are (unmeasurable) error nodes.

function Φ_k:

$$\Omega_k = \Phi_k(\Omega_\alpha, \Omega_\beta, \ldots, \Omega_\omega), \tag{5.2.2}$$

where $\Omega_\alpha, \Omega_\beta, \ldots, \Omega_\omega$ denote the corresponding output of the $n_{p,k}$ parent nodes ($k < \alpha, \beta, \ldots, \omega$) and $\Phi_k : \mathbb{R}^{n_k} \to \mathbb{R}$, e.g., $\Omega_4 = \Phi_4(\Omega_9, \Omega_{10})$ in Fig. 5.1. The mapping Φ_1 is part of an intermediate calculation, producing output at the "root" node (node 1 in Fig. 5.1). Using either a projection or the Kolmogorov-Smirnov distance-based method [5], output at the root node can be then reduced from a stochastic dimension of n_ℓ to a PCE with a stochastic dimension of one [27]:

$$\Omega_1 = \sum_{i,j} \theta_{ij} H_i(\xi_j). \tag{5.2.3}$$

- Θ denote the space of possible states of nature. In general, Θ is not a proper vector space (it may not be closed under vector addition and scalar multiplication operations), but a subspace of a real vector space, e.g., $\Theta \subset \mathbb{R}^d$, where $d = \sum_j q_j + n_\ell$.

- $\mathscr{R}(\boldsymbol{\theta})$ denote system reliability, here, $\mathscr{R}(\boldsymbol{\theta}) = \Pr(\Omega_1 \in \tilde{\Omega}_1 \,|\, \boldsymbol{\theta})$ for some set $\tilde{\Omega}$ of acceptable values. This probability comes from the PCE evaluated at a given value of $\boldsymbol{\theta}$ and pre-defined upper and lower limits on the output, $\underline{\Omega}$ and $\overline{\Omega}$, respectively. In other words,

$$\mathscr{R}(\boldsymbol{\theta}) = \int_{\underline{\Omega}}^{\overline{\Omega}} D(\Omega_1|\boldsymbol{\theta}) \mathrm{d}\Omega_1, \tag{5.2.4}$$

where $D(\Omega_1|\boldsymbol{\theta})$ is the conditional distribution of Ω_1 given $\boldsymbol{\theta}$.

- $\mathscr{R}_{\text{req}}$ be the required reliability of the composite system of interest. The optimization methodology will use this as a reference value.

- $\pi(\boldsymbol{\theta})$ denote a prior density or probability function capturing the epistemic uncertainty about $\boldsymbol{\theta}$ *at the time the optimization problem is performed*. This may correspond to the posterior after some original prior has been updated with available data.

- $F^{\boldsymbol{\pi}}(\boldsymbol{\theta})$ denote the cumulative distribution function capturing the epistemic uncertainty about $\boldsymbol{\theta}$. For continuous univariate distributions. $\pi(\boldsymbol{\theta}) = \mathrm{d}F^{\boldsymbol{\pi}}(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}$. There is a notation convention here: superscripts give additional information about the source of the free variable in a function and subscripts provide information about conditioning variables.

- a denote an "action" to be taken. In this formulation, the action refers to the ultimate conclusion drawn from whatever data is collected, not the intermediate decisions about what data to collect. For estimation problems, the action is typically the reported vector of estimated values for $\boldsymbol{\theta}$. For hypothesis testing, where the choice is whether $\boldsymbol{\theta} \in \Theta_0$ or $\boldsymbol{\theta} \in \Theta_1$, where $\Theta = \Theta_0 \cup \Theta_1$, action typically assumes values: a_0 (assert $\boldsymbol{\theta} \in \Theta_0$ is correct) or a_1 (assert $\boldsymbol{\theta} \in \Theta_1$ is correct). For the development of our problem we follow the hypothesis testing format and choose among two actions:

  1. Conclude that the system does meet its reliability requirements
  2. Conclude that the system does not meet its reliability requirements

- $\mathscr{A}$ denote the space of all possible actions

- $L(\boldsymbol{\theta}, a)$ denote the "loss" (some measure of infidelity or regret) if one takes action a when $\boldsymbol{\theta}$ is the true state of nature (i.e., if all epistemic uncertainty was removed). Note that this does not depend in any way on the data or the distribution of $\boldsymbol{\theta}$. In addition, since we don't actually know the true value of $\boldsymbol{\theta}$, we can never compute a "true" value for $L(\boldsymbol{\theta}, a)$. What is important though is that the loss function is computable for all $\boldsymbol{\theta} \in \Theta$ and $a \in \mathscr{A}$. For estimation, a favorite choice for $L(\boldsymbol{\theta}, a)$ is the L^2 distance between $\boldsymbol{\theta}$ and a. For hypothesis testing the typical formulation is

$$L(\boldsymbol{\theta}, a_0) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in \Theta_0 \\ 1 & \text{if } \boldsymbol{\theta} \in \Theta_1 \end{cases} \tag{5.2.5a}$$

and

$$L(\boldsymbol{\theta}, a_1) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \Theta_0 \\ 0 & \text{if } \boldsymbol{\theta} \in \Theta_1 \end{cases} . \tag{5.2.5b}$$

An alternative formulation might replace the 1's above with some function of the distance between $\boldsymbol{\theta}$ and the nearest point on the boundary of Θ_0 or Θ_1, depending on the case.

- $\rho(\pi, a)$ denote the expected loss or risk functional for action a:

$$\rho(\pi, a) = E^{\pi}[L(\boldsymbol{\theta}, a)] = \int_{\Theta} L(\boldsymbol{\theta}, a) \mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{5.2.6}$$

  The Bayesian would seek to choose the action a that minimizes $\rho(\pi, a)$. Note the superscript on the expectation, indicating the distribution over which the integration is performed. For example, the π here is an argument to the risk functional representing the prior distribution on $\boldsymbol{\theta}$, but this formulation can also be applied to other distributions on $\boldsymbol{\theta}$, such as a posterior after some data is observed.

- X denote the RV associated with the data that we might collect.

- $\mathscr{X}$ denote the sample space for all possible data values, where X maps from the event space to $\mathscr{X}$.

- x denote an element of $\mathscr{X}$.

- $f(x|\boldsymbol{\theta})$ denote the likelihood function for x given the state of nature $\boldsymbol{\theta}$.

- $F^X(x|\boldsymbol{\theta})$ denote the cumulative distribution of x given the state of nature $\boldsymbol{\theta}$:

$$f(x|\boldsymbol{\theta}) = \mathrm{d}F^X(x|\boldsymbol{\theta})/\mathrm{d}x. \tag{5.2.7}$$

- $m(x)$ denote the unconditional marginal density of x:

$$m(x) = \int_{\boldsymbol{\theta}} f(x|\boldsymbol{\theta}) \mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{5.2.8}$$

- $F^m(x)$ denote the unconditional cumulative distribution of x:

$$m(x) = \mathrm{d}F^m(x)/\mathrm{d}x. \tag{5.2.9}$$

- $\pi(\boldsymbol{\theta}|x)$ denote the posterior density for $\boldsymbol{\theta}$; by Bayes theorem,

$$\pi(\boldsymbol{\theta}|x) = \frac{f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(x)}. \tag{5.2.10}$$

- $F^{\boldsymbol{\theta}|X}(\boldsymbol{\theta})$ denote the cumulative distribution of $\boldsymbol{\theta}$ given the data x:

$$\pi(\boldsymbol{\theta}|x) = \mathrm{d}F^{\boldsymbol{\theta}|X}(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}. \tag{5.2.11}$$

- $E_{\boldsymbol{\theta}}[h(X)]$ denote the expected value of some RV $h(X)$ given the state of nature $\boldsymbol{\theta}$.

$$E_{\boldsymbol{\theta}}[h(X)] = \int_{\mathscr{X}} h(x)\mathrm{d}F^{X}(x|\boldsymbol{\theta}). \tag{5.2.12}$$

  Note the *subscript* convention: $\boldsymbol{\theta}$ is the conditioning variable. If the context requires more clarity, we use the notation $E^{X}_{\boldsymbol{\theta}}[h(X)]$ to emphasize the integration is with respect to X conditional on $\boldsymbol{\theta}$.

- $\delta(X)$ denote the decision rule giving the action taken when X is observed. Thus, $\delta : \mathscr{X} \to \mathscr{A}$.

- $r(\pi, \delta)$ denote the "Bayes risk of a decision rule" with respect to a prior:

$$r(\pi,\delta) = E^{\pi}E^{X}_{\boldsymbol{\theta}}\left\{L[\boldsymbol{\theta},\delta(X)]\right\} = \int_{\Theta}\int_{\mathscr{X}} L[\boldsymbol{\theta},\delta(x)]\mathrm{d}F^{X}(x|\boldsymbol{\theta})\mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{5.2.13}$$

  When there are probability densities, this becomes

$$r(\pi,\delta) = \int_{\Theta}\int_{\mathscr{X}} L[\boldsymbol{\theta},\delta(x)]f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}x\mathrm{d}\boldsymbol{\theta}. \tag{5.2.14}$$

  An application of Fubini's theorem and some algebraic manipulation yields

$$r(\pi,\delta) = E^{m}\left\{\rho[\pi,\delta(X)]\right\} = \int_{\mathscr{X}}\int_{\Theta} L[\boldsymbol{\theta},\delta(x)]\mathrm{d}F^{\boldsymbol{\theta}|X}(\boldsymbol{\theta})\mathrm{d}F^{m}(x), \tag{5.2.15a}$$

$$= \int_{\mathscr{X}}\int_{\Theta} L[\boldsymbol{\theta},\delta(x)]\pi(\boldsymbol{\theta}|x)m(x)\mathrm{d}\boldsymbol{\theta}\mathrm{d}x. \tag{5.2.15b}$$

  This form is often easier to use. The Bayes risk of a decision rule is the metric that combines prior information, data collection, and subsequent conclusions.

5.3 Pre-posterior Analysis

There are several elements that one must combine: (a) the possible ways one might collect data, (b) the possible outcomes, (c) the conclusions that one might draw once the data are collected, and (d) the resulting risk of doing so. The Bayesian analyst would then run through this for all the ways one might collect data and then select the scheme with lowest risk.

5.3.1 Actions, loss, and decision rules for reliability problems

First, we need to augment the basic concepts introduced in the previous section to make them specific to our problem. Let

- a_{a} and a_{r} be the actions. For our problem, we propose

  a_{a}: conclude that the system does meet its reliability requirements (i.e., accept)

  a_{r}: conclude that the system does *not* meet its reliability requirements (i.e., reject)

- s be the index of possible data collection schemes. For the initial analysis, the strategy might be to collect one more observation at any one of the leaf nodes in the system. The notation, however, allows for any collection of pre-specified data collection options. For example, a simple extension would be to consider an ensemble of data collection schemes $s(i,n)$ where one collects n new observations at node i [1].

- $L(\boldsymbol{\theta}, a, s)$ denote an expanded loss function where $\boldsymbol{\theta}$ is the true state of nature, a is the ultimate action taken, and s is the sampling strategy for collecting data. This expansion allows for measuring the cost of collecting data. For our purposes, We propose that we start with a very simple right/wrong loss function that *does not* depend on s, i.e.,

$$L(\boldsymbol{\theta}, a_{\mathrm{a}}, s) = \begin{cases} 1 \text{ if } \mathscr{R}(\boldsymbol{\theta}) < \mathscr{R}_{\mathrm{req}} & \text{system is not OK and we disagree} \\ 0 \text{ if } \mathscr{R}(\boldsymbol{\theta}) \geq \mathscr{R}_{\mathrm{req}} & \text{system is OK and we agree} \end{cases} \tag{5.3.1a}$$

  and

$$L(\boldsymbol{\theta}, a_{\mathrm{r}}, s) = \begin{cases} 0 \text{ if } \mathscr{R}(\boldsymbol{\theta}) < \mathscr{R}_{\mathrm{req}} & \text{system is not OK and agree} \\ 1 \text{ if } \mathscr{R}(\boldsymbol{\theta}) \geq \mathscr{R}_{\mathrm{req}} & \text{system is OK and we disagree} \end{cases} \tag{5.3.1b}$$

  This formulation only works when the data collection alternatives are of essentially equal difficulty and expense. We address the issue of incorporating data collection costs in Section 5.5. For notational simplicity, let Θ_{a} denote the set of $\boldsymbol{\theta} \in \Theta$ such that $\mathscr{R}(\boldsymbol{\theta}) \geq \mathscr{R}_{\mathrm{req}}$ and $\Theta_{\mathrm{r}} = \Theta / \Theta_{\mathrm{a}}$, so $\Theta_{\mathrm{a}} \cup \Theta_{\mathrm{r}} = \Theta$.

- X^s denote a possible outcome that could be observed under data collection scheme s, i.e., X^s is the RV associated with the data that we might collect with scheme s.

- $\mathscr{X}^s$ denote the outcome space of data collection under scheme s. When it is clear from the context, x will be used to refer to an element of $\mathscr{X}^s$, otherwise x^s will be used.

- $\delta_s(X^s)$ denote the decision rule that states when we chose data collection scheme s and observe data X^s, then we choose action $\delta_s(X^s)$, which for our problem, is either a_{a} or a_{r}. Once data has been collected, a natural thing to do is to let $\delta_s(X^s)$ be the

[1] There are many potential realizations for $s(i,n)$, e.g., mortal, venial, etc.

action that minimizes the risk ρ, i.e., the action whose posterior probability of being correct corresponds to a maximum:

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \rho[\pi(\boldsymbol{\theta}|X^s), a_\mathrm{a}] < \rho[\pi(\boldsymbol{\theta}|X^s), a_\mathrm{r}] \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{5.3.2}$$

Applying the definition of expected loss [Eq. (5.2.6)] yields

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \int_\Theta L(\boldsymbol{\theta}, a_\mathrm{a}, s) \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) < \int_\Theta L(\boldsymbol{\theta}, a_\mathrm{r}, s) \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \\ a_\mathrm{r} & \text{otherwise} \end{cases} , \tag{5.3.3}$$

where $\boldsymbol{\theta}|X^s$ represents the conditional distribution of $\boldsymbol{\theta}$ given the result of the mapping X^s. Because $L(\boldsymbol{\theta}, a, s)$ is simple [Eq. (5.3.1)], this reduces to

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \int_{\Theta_\mathrm{r}} \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) < \int_{\Theta_\mathrm{a}} \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{5.3.4}$$

However, these integrands are just the posterior probabilities:

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{r}] < \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}] \\ a_\mathrm{r} & \text{otherwise} \end{cases} , \tag{5.3.5}$$

or equivalently,

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}] < 0.5 \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{5.3.6}$$

5.3.2 Bayes risk for reliability problems

This structure basically describes what we would do in practice for the military, although we don't perform the formal Bayesian calculation. We would tell them a system no longer meets its requirements if our best guess of its reliability (which we often interpret as our 50% confidence number), rightly or wrongly, fell below the requirement value.

Moreover, our often imprecise usage of "confidence", a word from the *frequentist*, not the Bayesian world, links well to Bayes risk. When applying the decision rule above, the value

of the Bayes risk of the decision is:

$$\rho[\pi(\boldsymbol{\theta}|X^s), \delta_s(X^s)] = \begin{cases} \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_r] & \text{if } \delta_s(X^s) = a_a \\ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_a] & \text{if } \delta_s(X^s) = a_r \end{cases} \tag{5.3.7a}$$

$$= \min\left\{\Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_a],\ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_r]\right\} \tag{5.3.7b}$$

$$= \min\left\{\Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_a],\ 1 - \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_a]\right\} \tag{5.3.7c}$$

In the context of our pre-defined decision rule δ_s, this is the posterior probability of making the wrong conclusion.

5.4 The Pre-posterior Problem

With the notation and formalism from Sections 5.2 and 5.3, we have the mathematical tools to address the pre-posterior problem. For the various sampling options indexed by s, one seeks to minimize the overall decision risk functional

$$r(\pi, \delta_s, s) = E^{\pi} E^{X^s}_{\boldsymbol{\theta}} \{L[\boldsymbol{\theta}, \delta_s(X^s), s]\} = \int_{\Theta} \int_{\mathscr{X}^s} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{X^s|\boldsymbol{\theta}}(x) \mathrm{d}F^{\pi}(\boldsymbol{\theta}), \tag{5.4.1}$$

just as one does in any Bayesian decision problem. Note the addition of the argument "s" to the risk function to index the data collection scheme. This looks straightforward, but there is an *inner* calculation:

$$\int_{\mathscr{X}^s} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{X^s|\boldsymbol{\theta}}(x), \tag{5.4.2}$$

which is evaluated at *all* values of $\boldsymbol{\theta}$ in the outer integral (integrating over the prior on $\boldsymbol{\theta}$), but the evaluation of the decision rule $\delta_s(X^s)$ depends on probabilities on the posterior distribution of $\boldsymbol{\theta}$ given X^s. That posterior draws on all of the prior $\pi(\boldsymbol{\theta})$, not just on the one $\boldsymbol{\theta}$ from the outer integral. We are writing the rule to determine what we will do with the data when we get there, and then, in the outer integral, work through how we could have gotten there. We have not written rules about how to use data based on an unknown true single value of $\boldsymbol{\theta}$.

Rather than develop the calculus with the original form of $r(\pi, \delta_s, s)$ from Eq. (5.4.1), it is more straightforward to utilize the alternate form

$$r(\pi, \delta_s, s) = E^m \{\rho[\pi(\boldsymbol{\theta}|X^s), \delta_s(X^s)]\} = \int_{\mathscr{X}^s} \int_{\Theta} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \mathrm{d}F^m(x). \tag{5.4.3}$$

This form emphasizes the fact that we first choose X^s according to the *unconditional* marginal distribution for possible data and then base the risk confidence on the posterior

development given that data. The key difference from our earlier formulation is that the valuation function (the loss) is inside of the integrals and is done separately and possibly differently for each possible data value. It is not done *outside* the integral for some averaged outcome. Combining everything yields [cf. Eq. (5.3.7)]

$$r(\pi, \delta_s, s) = \int_{\mathscr{X}^s} \min \left\{ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}],\ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{r}] \right\} m(x)\mathrm{d}x, \tag{5.4.4a}$$

where

$$\Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}] = \int_{\Theta_\mathrm{a}} \frac{f(X^s|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(X^s)}\mathrm{d}\boldsymbol{\theta} = \int_{\Theta_\mathrm{a}} \pi(\boldsymbol{\theta}|X^s)\mathrm{d}\boldsymbol{\theta}. \tag{5.4.4b}$$

Note that Eq. (5.4.4b) depends on X^s, not $\boldsymbol{\theta}$. The fairly simple equation above (at least analytically) is what we need to compare sampling schemes. Equation (5.4.4a) may also be expressed as

$$r(\pi, \delta_s, s) = \int_{\mathscr{X}^s} \min_{\Theta'} \left\{ \int_{\Theta'} f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \right\} \mathrm{d}x, \tag{5.4.5}$$

where Θ' is either Θ_a or Θ_r. Equation (5.4.4a) can be incorporated as part of an objective functional:

$$J_1 = r(\pi, \delta_s, s) + \lambda C(s) \tag{5.4.6a}$$

or

$$J_2 = r(\pi, \delta_s, s) \times [1 + C(s)], \tag{5.4.6b}$$

where $C(s)$ quantifies the cost of the sampling scheme s and $\lambda \geq 0$ weights $C(s)$ relative to $r(\pi, \delta_s, s)$. Now one can minimize J_1 or J_2 to locate the optimal scheme $\hat{s}$:

$$\hat{s} = \arg\min_s J, \tag{5.4.7}$$

where J can be either J_1 or $J = J_2$.

5.5 Example: Numerical Optimization

5.5.1 Model description

Consider a simple three-node system consisting of two leaf/parent nodes and one root/system node, illustrated in Fig. 5.2.

Output Ω_2 and Ω_3 at the leaf nodes 2 and 3, are given by respective PCEs [Eq. (5.2.1)] and output Ω_1 at the root node "1" is given by $\Omega_1 = \Phi_1(\Omega_2, \Omega_3)$. In this example, we consider two sampling schemes, corresponding to collecting additional data at either node 2 or 3.

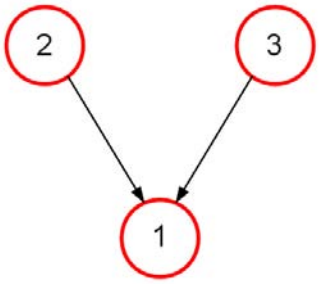


Figure 5.2. Graphical representation of a simple 3-node example model.

5.5.2 Limiting cases

For illustrative purposes, we further partition the subspaces Θ_a and Θ_r by node: $\Theta_{a,2}$, $\Theta_{r,2}$, $\Theta_{a,3}$ and $\Theta_{r,3}$, and posit that

$$\Theta_a = \Theta_{a,2} \otimes \Theta_{a,3} \tag{5.5.1a}$$

and

$$\Theta_r = \{\Theta_{a,2} \otimes \Theta_{r,3}, \Theta_{r,2} \otimes \Theta_{a,3}, \Theta_{r,2} \otimes \Theta_{r,3} = \Theta_r\}. \tag{5.5.1b}$$

Using this decomposition of Θ, consider $r(\pi, \delta_2, 2)$ from Eq. (5.4.5):

$$r(\pi, \delta_2, 2) = \int_{\mathscr{X}^2} \min_{\Theta'} \left\{ \int_{\Theta'} f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \right\} \mathrm{d}x, \tag{5.5.2a}$$

$$= \int_{\mathscr{X}^2} \min \left\{ \int_{\Theta_{a,2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \times \int_{\Theta_{a,3}} \pi(\theta_3)\mathrm{d}\theta_3, \right.$$
$$\left. \sum_{(i,j)} \int_{\Theta_{i,2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \times \int_{\Theta_{j,3}} \pi(\theta_3)\mathrm{d}\theta_3 \right\} \mathrm{d}x, \tag{5.5.2b}$$

where the indices i and j specify pairs $\{(a,r),(r,a),(r,r)\}$ corresponding to integration over the "reject" subspace Θ_r. By exchanging the node indices of Eq. (5.5.2), the corresponding analytical form for $r(\pi, \delta_3, 3)$ results. With Bayes risk expressed this way, we can test some limiting cases against intuition.

5.5.2.1 Informative priors for both leaf nodes

Suppose both nodes represent unique sub-systems which are each extremely well characterized. In other words, their respective priors are essentially multivariate delta functions,

e.g., for node 2,

$$\pi(\theta_2) \approx \delta\left(\theta_2^{\mathrm{t}} - \theta_2\right), \tag{5.5.3}$$

where $\theta_2^{\mathrm{t}} \in \Theta_{\mathrm{a},2}$ is the true value for θ_2 (analogously for node 3, where $\theta_3^{\mathrm{t}} \in \Theta_{\mathrm{a},3}$). Based on this information, $r(\pi, \delta_s, s) \approx 0$ for all s. For $s = 2$, all (i, j)-pairs in the second term of the minimization [Eq. (5.5.2)] are zero:

$$\int_{\Theta_{i,2}} f(x|\theta_2)\delta\left(\theta_2^{\mathrm{t}} - \theta_2\right) \mathrm{d}\theta_2 \times \int_{\theta_{j,3}} \delta\left(\theta_3^{\mathrm{t}} - \theta_3\right) \mathrm{d}\theta_3 = 0. \tag{5.5.4}$$

The same result holds for $s = 3$. Since both nodes are well characterized, there is no relative risk in choosing to sample either node 2 or 3, which is in agreement with intuition.

5.5.2.2 Uninformative priors for both leaf nodes

Now, suppose that both nodes are very poorly characterized, i.e., their priors are approximately multivariate uniform distributions. In this case, for $s = 2$, the Bayes risk becomes

$$\begin{aligned} r(\pi, \delta_2, 2) \approx \int_{\mathscr{X}^2} \min \Big\{ & \Pi_{\mathrm{a},3} \int_{\Theta_{\mathrm{a},2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2, \\ & \Pi_{\mathrm{r},3} \int_{\Theta_2} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 + \pi_{\mathrm{a},3} \int_{\Theta_{\mathrm{r},2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \Big\} \mathrm{d}x, \end{aligned} \tag{5.5.5a}$$

where

$$\Pi_{j,3} = \int_{\Theta_{j,3}} \pi(\theta_3)\mathrm{d}\theta_3, \tag{5.5.5b}$$

and analogously for $s = 3$. Given this intermediate result, the Bayes risk for each sampling scheme is *likely* to be

$$r(\pi, \delta_2, 2) \approx \int_{\mathscr{X}^2} \Pi_{\mathrm{a},3} \int_{\Theta_{\mathrm{a},2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2\mathrm{d}x \text{ for } s = 2 \tag{5.5.6a}$$

and

$$r(\pi, \delta_3, 3) \approx \int_{\mathscr{X}^3} \Pi_{\mathrm{a},2} \int_{\Theta_{\mathrm{a},3}} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3\mathrm{d}x \text{ for } s = 3. \tag{5.5.6b}$$

Without further information on the sampling spaces $\mathscr{X}^s$, the state spaces $\Theta_{\mathrm{a},s}$, and the likelihood functions $f(x|\theta_s)$, no conclusion can be drawn supporting the intuitive result of sampling at either node 2 or 3 with equal risk, unless, e.g., it is assumed that the systems represented by these nodes are identical, then the integrals in Eq. (5.5.6) are identical.

5.5.2.3 Informative prior for one leaf node

As a final example, suppose that node 2 is extremely well characterized and node 3 is very poorly characterized. Based on the previous examples, this corresponds to an approximate delta function for the prior on node 2 [Eq. (5.5.3)] and a uniform distribution for the prior on node 3. With these assumptions, the Bayes risk for each sampling scheme is

$$r(\pi,\delta_2,2)\min\left\{\Pi_{\mathrm{a},3},\Pi_{\mathrm{r},3}\right\}\times\approx\int_{\mathscr{X}^2} f\left(x|\theta_2^{\mathrm{t}}\right)\mathrm{d}x \text{ for } s=2 \tag{5.5.7a}$$

and

$$r(\pi,\delta_3,3)\approx\int_{\mathscr{X}^3}\min_{\Theta_3'}\int_{\Theta_3'} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3\mathrm{d}x \text{ for } s=3, \tag{5.5.7b}$$

where Θ_3' is either $\Theta_{\mathrm{a},3}$ or $\Theta_{\mathrm{r},3}$. We argue that $r(\pi,\delta_2,2)\approx 0.5$ because …, which is an approximate maximum of $r(\pi,\delta_s,s)$, supporting a data collection scheme at node 3.

5.5.3 Model initialization

To set $\underline{\Omega}$ and $\overline{\Omega}$, which collectively determine system reliability, we use the "truth", denoted as $\boldsymbol{\theta}^{\mathrm{t}}$ for the leaf nodes. For our tests, the set of θ_{ij} has either linear or exponential time-dependent forms:

$$\theta_{ij}=\begin{cases}\theta_{ij}=\theta_{ij}^1+\theta_{ij}^2 t & \text{linear}\\ \theta_{ij}=\theta_{ij}^1+\theta_{ij}^2\exp\left(-\theta_{ij}^3 t\right) & \text{exponential}\end{cases}. \tag{5.5.8}$$

With a sampling of ξ_j for each leaf node and the resulting output Ω_1 [determined from Eqs. (5.2.1) and (5.2.2)], we then set $\underline{\Omega}$ and $\overline{\Omega}$ such that $\mathscr{R}(\boldsymbol{\theta}^{\mathrm{t}})=z$ (e.g., a desired reliability of $z=0.8$):

$$\int_{\underline{\Omega}}^{\overline{\Omega}} D(\Omega_1|\boldsymbol{\theta}^{\mathrm{t}})\mathrm{d}\Omega_1=z. \tag{5.5.9}$$

This calculation allows for elements of Θ to be categorized as in either Θ_{a} or Θ_{r}, which is necessary to evaluate Eq. (5.4.5). For our initial optimizations, time will be fixed (e.g., $t=0$ or $t=1$) to simplify calculations.

5.5.4 Integration limits

5.5.4.1 State of nature priors

For the jth leaf node, where $\theta_j = \left\{\theta_{ij}\right\}_{i=0}^{q_j}$, the prior $\pi(\theta_j)$ is a product of Gaussian and gamma probability density functions (PDFs):

$$\pi(\theta_j) = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left[-\frac{(\theta_{0j}-\mu_j)}{2\sigma_j^2}\right] \times \frac{1}{\Gamma(\alpha)}\beta^{-\alpha}\nu_j^{\alpha-1}\exp\left(-\frac{\nu_j}{\beta}\right), \tag{5.5.10}$$

where μ_j and σ_j are the mean and standard deviation of the Gaussian PDF, $\Gamma(\alpha)$ is the standard gamma function, $\nu_j = \sum_{i=1}^{q_j}\theta_{ij}^2/i!$ (the variance of the jth PCE), and α and β are the shape and scale parameters for the gamma PDF, respectively. Typically, μ_j is known for a given node/sub-component, $\sigma_j = 0.5$, and $\alpha = \beta = 1$ for all j. The prior can be used to estimate reasonable integration limits for Eq. (5.4.5):

$$\mu_j - 4\sigma_j \leq \theta_{0j} \leq \mu_j + 4\sigma_j \tag{5.5.11a}$$

and

$$0 \leq \sum_{i=1}^{q_j}\theta_{ij}^2/i! \leq 6, \tag{5.5.11b}$$

where $\sum_i \theta_{ij}^2/i! \leq 6$ corresponds to a hyperellipsoid. For simplicity, the hyperellipsoid is approximated as q_j-orthotope (i.e., a generalized q_j-dimensional hypercube), with a hypervolume given by $\prod_{i=1}^{q_j} 2\sqrt{6i!}$.

5.5.4.2 Limits of the sampling spaces

Integration limits for the sampling spaces either can be set arbitrarily (which is probably not very efficient numerically) or can use the available data or truth model to get an estimate on $\mathscr{X}^s$.

5.5.5 Evaluation of the Bayes risk of a decision rule

For convenience, Eq. (5.4.5) is repeated here:

$$r(\pi, \delta_s, s) = \int_{\mathscr{X}^s} \min_{\Theta'}\left\{\int_{\Theta'} f(x|\theta)\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\right\}\mathrm{d}x,$$

which, for sampling schemes at leaf nodes, corresponds to a (q_s+1)-dimensional integral in Θ-space, with q_s (the order of the s-leaf-node PCE) corresponding to the sampling

scheme. Monte Carlo integration with quasi-random sequences [17, 33] is an appropriate method for numerical integration of Bayes risk:

$$r(\pi, \delta_s, s) \approx \frac{|\mathscr{X}^s|}{N_x} \sum_{i=1}^{N_x} \min_{\Theta'} \left\{ \frac{|\Theta'|}{N_{\theta'}} \sum_{j=1}^{N_{\theta'}} f\left(x^{(i)}|\boldsymbol{\theta}^{(j)}\right) \pi\left(\boldsymbol{\theta}^{(j)}\right) \right\}, \tag{5.5.12}$$

where $|\mathscr{X}^s|$ and $|\Theta'|$ denote the volumes of the respective spaces. For each of the N_x sample integration points of $\mathscr{X}^s$, Eq. (5.5.12) requires *two* integrations of $f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ because of the minimization over Θ_a and Θ_r. The $N_{\theta'}$ sample integration points will be generated according to the following routine:

1. Based on the limits specified by Eq. (5.5.11), generate a set of N_θ sample points, each denoted as $\boldsymbol{\theta}^{(i)}$, via quasi-random sequences (e.g., the Faure sequence [43]) [17, 33].
2. For every point, determine whether $\boldsymbol{\theta}^{(i)}$ is an element of Θ_a or Θ_r. This calculation is similar as to what is described in Section 5.5.3: If $\mathscr{R}(\boldsymbol{\theta}(i)) \geq \mathscr{R}_\mathrm{req}$, then $\boldsymbol{\theta}^{(i)} \in \Theta_\mathrm{a}$, otherwise $\boldsymbol{\theta}^{(i)} \in \Theta_\mathrm{r}$.
3. Repeat steps 1 and 2 for all N_θ points.

For a given data point x and a state $\boldsymbol{\theta}$, the likelihood function $f(x|\boldsymbol{\theta})$ in Eq. (5.5.12) is evaluated in the following manner. Let

$$\Omega(\xi|\boldsymbol{\theta}) = \sum_{i=1}^{q} \theta_i H_i(\xi), \tag{5.5.13}$$

i.e., leaf-node output Ω is represented by PCE, which is a function of the Gaussian RV ξ, conditional on $\boldsymbol{\theta}$. Then, the likelihood $f(x|\boldsymbol{\theta})$ is given by

$$f(x|\boldsymbol{\theta}) = \sum_{\xi^* \in \Xi^*} \frac{P_\xi(\xi^*)}{|\Omega'(\xi^*|\theta)|}, \tag{5.5.14}$$

where $P_\xi(\cdot)$ is the known probability distribution of ξ [in this work, $\xi \sim \mathscr{N}(0,1)$] and Ξ^* is the set of roots of $x - \Omega(\xi)$. The set Ξ^* is determined from an efficient polynomial solver [31, 32].

5.6 Conclusion and Future Work

In this chapter we present a methodology to select an optimal sampling strategy based on Bayesian decision theory, incorporating a combination of data and prior information via Bayes theorem. By defining an appropriate loss function [Eq. (5.3.1)]and decision rule [Eq. (5.3.2)], this methodology determines which component/node should be tested to

optimally improve the reliability estimate. A numerical implementation is described that scales linearly with the number of components available for testing, although evaluating integrals of Eq. (5.5.12) may be expensive in general (depending on the PCE order of the leaf nodes, the number of sample points required for accurate integration, etc.).

Future work involves a numerical demonstration of this sampling strategy for a multicomponent system, potentially adapting the formalism to determine which components might benefit from redundancy or other fault-tolerant techniques to further improve the reliability estimate. In addition, developing criteria that *a priori* limit the number of integral evaluations (e.g., by performing a preliminary analysis of the likelihood functions and prior distributions) may decrease the computational expense of this strategy and are also being considered.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. 9th Printing, Dover Publications, Inc., New York, 1970.

[2] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, NY, 1972.

[3] James O. Berger. *Statistical decision theory and Bayesian analysis*. Springer series in statistics. Springer, New York, NY, 2nd edition, 1985.

[4] Patrick Billingsley. *Probability and Measure*. J. Wiley and Sons, New York, 1995.

[5] Paul T. Boggs, Youssef M. Marzouk, Philippe P. Pébay, John Red-Horse, James T. Ringland, Kathleen Diegert, and Rena Zurn. Bayesian methods for estimating the reliability in complex hierarchical networks (interim report). Technical Report 2007-2758, Sandia National Laboratories, 2007.

[6] Paul T. Boggs, Youssef M. Marzouk, James T. Ringland, Matthew D. Grace, Philippe P. Pébay, John Red-Horse, James T. Ringland, Kathleen Diegert, and Rena Zurn. Determining the reliability distribution in a time-dependent hierarchical system. Technical Report 2009-xxxx, Sandia National Laboratories, 2009.

[7] Immanuel M. Bomze. *A functional analytic approach to statistical experiments*. Pitman Research Notes in Mathematics. Longman, London, 1990.

[8] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, Berlin, Heidelberg, 1986.

[9] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. Math.*, 48(2):385–392, 1947.

[10] R. H. Cameron and W. T. Martin. The orthogonal development of nonlinear functionals in series of fourier-hermite functionals. *Ann. Math*, pages 385–392, 1947.

[11] Thierry Crestaux, Olivier Le Maître, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.

[12] R. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Dover Publications, Inc, Mineola, NY, revised edition edition, 2003.

[13] R. G. Ghanem and A. Doostan. On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *Journal of Computational Physics*, 217:63–81, 2006.

[14] Roger G. Ghanem and Alireza Doostan. On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *Journal of Computational Physics*, 217(1):63–81, 2006.

[15] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC Press, New York, NY, 1996.

[16] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[17] F. James. Monte Carlo theory and practice. *Rep. Prog. Phys.*, 43(9):1145–1189, 1980.

[18] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Springer, 2001.

[19] M. I. Jordan. Graphical models. *Statistical Science*, 19:140–155, 2004.

[20] S. Kakutani. Spectral analysis of stationary Gaussian processes. In J. Neyman, editor, *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 239–247. University of California Press, 1961.

[21] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, Mineola, NY, 1957.

[22] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, Mineola, NY, 1957.

[23] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. R. A. Silverman (Translator). Dover Publications, Inc, 1975.

[24] H. F. Martz and R. A. Waller. Bayesian reliability analysis of complex series/parallel systems of binomial subsystems and components. *Technometrics*, 32:407–416, 1990.

[25] H. F. Martz, R. A. Waller, and E.T. Fickas. Bayesian reliability analysis of series systems of binomial subsystems and components. *Technometrics*, 30:143–154, 1988.

[26] Youssef M. Marzouk, Paul T. Boggs, Philippe P. Pébay, John Red-Horse, James T. Ringland, Kathleen Diegert, and Rena Zurn. Nonparametric bayesian density estimation using polynomial chaos ex pansions. Technical Report 2009-xxxx, Sandia National Laboratories, 2009.

[27] Youssef M. Marzouk and Habib N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *J. Comp. Phys.*, 228(6):1862–1902, 2009.

[28] Youssef M. Marzouk, Habib N. Najm, and Larry A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *J. Comp. Phys.*, 224(2):560–586, jun 2007.

[29] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1965.

[30] A. Papoulis and S. U. Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 4th edition, 2002.

[31] P. Pébay, M. J. Rojas, and D. Thompson. Optimization and np_r-completeness of certain fewnomials. In *Proc. 3rd International Workshop on Symbolic-Numeric Computation*, Kyoto, Japan, August 2009.

[32] P. Pébay, D. Thompson, and J. M. Rojas. Computational algebraic geometry for statistical modeling FY09Q2 progress. Sandia Report SAND2009-2011, Sandia National Laboratories, April 2009.

[33] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical recipes: The art of scientific computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

[34] J. R. Red-Horse and A. S. Benjamin. A probabilistic approach to uncertainty quantification with limited information. *Reliability Engineering and System Safety*, 85:183–190, 2004.

[35] John Red-Horse, Philippe P. Pébay, Paul T. Boggs, Youssef M. Marzouk, James T. Ringland, Kathleen Diegert, and Rena Zurn. Practical considerations in modeling random variables, vectors and fields. Technical Report 2009-xxxx, Sandia National Laboratories, 2009.

[36] John R. Red-Horse and Allan S. Benjamin. A probabilistic approach to uncertainty quantification with limited information. In *9thASCE EMD-SEI-GI-AD Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability*, Albuquerque, NM, July 24–26 2004.

[37] Jr. Richard V. Field and Mircea D. Grigoriu. Convergence properties of polynomial chaos approximations for l_2 random variables. Technical Report SAND2007-1262, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, March 2007.

[38] James T. Ringland, Matthew D. Grace, John Red-Horse, Paul T. Boggs, Youssef M. Marzouk, Philippe P. Pébay, James T. Ringland, Kathleen Diegert, and Rena Zurn. Determining the optimal sampling strategy in a hierarchical system. Technical Report 2009-xxxx, Sandia National Laboratories, 2009.

[39] Murray Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952.

[40] J. S. Rosenthal. *A First Look At Rigorous Probability Theory*. World Scientific, 2^{nd} edition, 2006.

[41] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, third edition, 1987.

[42] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 3rd edition, 1987.

[43] P. K. Sarkar and M. A. Prasad. A comparative study of pseudo and quasi random sequences for the solution of integral equations. *J. Comp. Phys.*, 68(1):66–88, 1987.

[44] Christoph Schwab and Radu Alexandru Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, Jan 2006.

[45] Jordan Stoyanov. Krein condition in probabilistic moment problems. *Bernoulli*, 6(5):939–949, 2000.

[46] Francois Treves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press, 1967.

[47] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[48] N. Wiener. The homogeneous chaos. *Amer. J. Math.*, 60:897–936, 1938.

[49] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comp. Phys.*, 187(1):137–167, 2003.

[50] N. Young. *An Introduction to Hilbert Space*. Cambridge University Press, Cambridge, UK, 1997.