SANDIA REPORT

SAND2010-6661
Unlimited Release
Printed Sept 2010

Determining the Bayesian Optimal Sampling Strategy in a Hierarchical System

Paul T. Boggs, Matthew D. Grace, Philippe P. Pebay, and James T. Ringland

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

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

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.osti.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/help/ordermethods.asp?loc=7-4-0#online

SAND2010-6661
Unlimited Release
Printed Sept 2010

Determining the Bayesian Optimal Sampling Strategy in a Hierarchical System

Paul T. Boggs
Quantitative Modeling & Analysis (08954)
Matthew D. Grace
Scalable & Secure Systems Research & (08961)
Philippe P. Pebay
Scalable Modeling & Analysis (08953)
James T. Ringland
Reliability & Electrical Systems (08245)
Sandia National Laboratories
Livermore, CA 94550, U.S.A.
ptboggs@sandia.gov, mgrace@sandia.gov,
pppebay@sandia.gov, & jtringl@sandia.gov

Abstract

Consider a classic hierarchy tree as a basic model of a "system-of-systems" network, where each node represents a component system (which may itself consist of a set of sub-systems). For this general composite system, we present a technique for computing the optimal testing strategy, which is based on Bayesian decision analysis. In previous work, we developed a Bayesian approach for computing the distribution of the reliability of a system-of-systems structure that uses test data and prior information. This allows for the determination of both an estimate of the reliability and a quantification of confidence in the estimate. Improving the accuracy of the reliability estimate and increasing the corresponding confidence require the collection of additional data. However, testing all possible sub-systems may not be cost-effective, feasible, or even *necessary* to achieve an improvement in the reliability estimate. To address this sampling issue, we formulate a Bayesian methodology that systematically determines the optimal sampling strategy under specified constraints and costs 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, where risk quantifies the relative effect that each sampling strategy could have on the reliability estimate. A general numerical algorithm is developed and tested using an example multicomponent system. The results show that the procedure scales linearly with the number of components available for testing.

Contents

Figures

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`

$F^{\pi}(\boldsymbol{\theta})$ Cumulative distribution function of $\pi(\boldsymbol{\theta})$ `\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`

$E^{\pi}_{\boldsymbol{\theta}}[\cdot]$ Expectation value of some function or random variable given a state of nature $\boldsymbol{\theta}$ and a probability distribution π `\expect{}{}`

Ω_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{}`

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 [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. Aleatory versus epistemic issue, system model and mapping from node to node to root.
- 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 of the possible large number of components and corresponding testing schemes *and* the potential wide range of outcomes for a particular test, we clearly cannot explore all possibilities. 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.

2 Foundational Concepts and Notational Conventions

The following notation and formulation follows James Berger's classic text, *Statistical Decision Theory and Bayesian Analysis* [2]. 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 2 works through the basic formalism of Bayesian decision analysis, while Section 3 addresses the optimization/design-of-experiments question specifically. Berger calls the latter "pre-posterior" analysis. Section 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 [2]. 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}, the distribution of which 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. 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) [14, 5, 9, 6]:

$$\Omega_j = \sum_{i=0}^{q_j} \theta_{ij} H_i(\xi_j), \tag{2.1}$$

  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 function:

$$\Omega_k = \Omega_k(\Omega_\alpha, \Omega_\beta, \ldots, \Omega_\omega), \tag{2.2}$$

  where $\Omega_\alpha, \Omega_\beta, \ldots, \Omega_\omega$ denote the corresponding output of the parent nodes ($k < \alpha, \beta, \ldots, \omega$) and $\Omega_k : \mathbb{R}^{n_k} \to \mathbb{R}$, e.g., $\Omega_4 = \Omega_4(\Omega_9, \Omega_{10})$ in Fig. 1. The mapping Ω_1 is part of an intermediate calculation, producing output at the "root" node (node 1 in Fig. 1). At this point, the output at the root node has a stochastic dimension of n_ℓ. Using a projection method [3], output at the root node can be then approximated by a PCE with a stochastic dimension of one [8]:

$$\Omega_1 = \sum_{i,j} \theta_{ij} H_i(\xi_j). \tag{2.3}$$

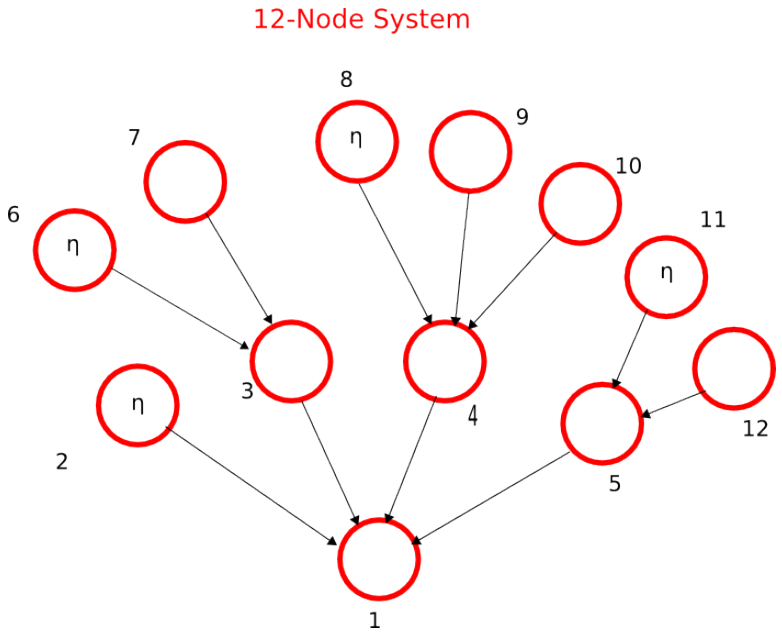

Figure 1: Graphical representation of a 12-node example model of our system of systems. Nodes marked η are (unmeasurable) error nodes.

- 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, $\mathscr{R}(\boldsymbol{\theta}) = \Pr(\Omega_1 \in [\underline{\Omega}, \overline{\Omega}] \,|\, \boldsymbol{\theta})$ for some set $[\underline{\Omega}, \overline{\Omega}]$ 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,

$$\mathscr{R}(\boldsymbol{\theta}) = \int_{\underline{\Omega}}^{\overline{\Omega}} D(\Omega_1|\boldsymbol{\theta}) \mathrm{d}\Omega_1, \tag{2.4}$$

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

$$\mathscr{R}(\boldsymbol{\theta}) = \int_{\Omega} D(\Omega_1|\boldsymbol{\theta}) \mathrm{d}\Omega_1, \tag{2.5}$$

- $\mathscr{R}_{\text{req}}$ 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.

- $F^{\pi}(\boldsymbol{\theta})$ denote the cumulative distribution function capturing the epistemic uncertainty about $\boldsymbol{\theta}$. For continuous univariate distributions. $\pi(\boldsymbol{\theta}) = \mathrm{d}F^{\pi}(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}$. 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 are 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 that $\boldsymbol{\theta} \in \Theta_0$ is correct) or a_1 (assert that $\boldsymbol{\theta} \in \Theta_1$ 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 hypothesis testing the typical formulation is

$$L(\boldsymbol{\theta}, a_0) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in \Theta_0 \\ 1 & \text{if } \boldsymbol{\theta} \in \Theta_1 \end{cases} \tag{2.6a}$$

  and

$$L(\boldsymbol{\theta}, a_1) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \Theta_0 \\ 0 & \text{if } \boldsymbol{\theta} \in \Theta_1 \end{cases} . \tag{2.6b}$$

  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:

$$\rho(\pi, a) = E^{\pi}[L(\boldsymbol{\theta}, a)] = \int_{\Theta} L(\boldsymbol{\theta}, a) \mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{2.7}$$

  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 are 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 $\boldsymbol{\theta}$:

$$f(x|\boldsymbol{\theta}) = \mathrm{d}F^X(x|\boldsymbol{\theta})/\mathrm{d}x. \tag{2.8}$$

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

$$m(x) = \int_{\boldsymbol{\theta}} f(x|\boldsymbol{\theta})\mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{2.9}$$

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

$$m(x) = \mathrm{d}F^m(x)/\mathrm{d}x. \tag{2.10}$$

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

$$\pi(\boldsymbol{\theta}|x) = \frac{f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(x)}. \tag{2.11}$$

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

$$\pi(\boldsymbol{\theta}|x) = \mathrm{d}F^{\boldsymbol{\theta}|X}(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}. \tag{2.12}$$

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

$$E_{\boldsymbol{\theta}}[h(X)] = \int_{\mathscr{X}} h(x)\mathrm{d}F^X(x|\boldsymbol{\theta}). \tag{2.13}$$

  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:

$$r(\pi,\delta) = E^{\pi}E^X_{\boldsymbol{\theta}}\left\{L[\boldsymbol{\theta},\delta(X)]\right\} = \int_{\Theta}\int_{\mathscr{X}} L[\boldsymbol{\theta},\delta(x)]\mathrm{d}F^X(x|\boldsymbol{\theta})\mathrm{d}F^{\pi}(\boldsymbol{\theta}). \tag{2.14}$$

When there are probability densities, this becomes

$$r(\pi, \delta) = \int_{\Theta} \int_{\mathscr{X}} L[\boldsymbol{\theta}, \delta(x)] f(x|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}x \mathrm{d}\boldsymbol{\theta}. \tag{2.15}$$

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

$$r(\pi, \delta) = E^m \left\{ \rho[\pi, \delta(X)] \right\} = \int_{\mathscr{X}} \int_{\Theta} L[\boldsymbol{\theta}, \delta(x)] \mathrm{d}F^{\boldsymbol{\theta}|X}(\boldsymbol{\theta}) \mathrm{d}F^m(x), \tag{2.16a}$$

$$= \int_{\mathscr{X}} \int_{\Theta} L[\boldsymbol{\theta}, \delta(x)] \pi(\boldsymbol{\theta}|x) m(x) \mathrm{d}\boldsymbol{\theta} \mathrm{d}x. \tag{2.16b}$$

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.

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.

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.,

$$L(\boldsymbol{\theta}, a_{\mathrm{a}}, s) = \begin{cases} 1 \text{ if } \mathscr{R}(\boldsymbol{\theta}) < \mathscr{R}_{\mathrm{req}} & \text{system is not OK and we disagree} \\ 0 \text{ if } \mathscr{R}(\boldsymbol{\theta}) \geq \mathscr{R}_{\mathrm{req}} & \text{system is OK and we agree} \end{cases} \tag{3.1a}$$

  and

$$L(\boldsymbol{\theta}, a_{\mathrm{r}}, s) = \begin{cases} 0 \text{ if } \mathscr{R}(\boldsymbol{\theta}) < \mathscr{R}_{\mathrm{req}} & \text{system is not OK and we agree} \\ 1 \text{ if } \mathscr{R}(\boldsymbol{\theta}) \geq \mathscr{R}_{\mathrm{req}} & \text{system is OK and we disagree} \end{cases} \tag{3.1b}$$

  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 6. 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.

[1] There are many potential realizations for $s(i, n)$, e.g., envy, gluttony, greed, lust, pride, sloth, wrath [1].

- $\mathscr{X}^s$ 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 action that minimizes the risk ρ, i.e., the action whose posterior probability of being correct corresponds to a maximum:

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \rho[\pi(\boldsymbol{\theta}|X^s), a_\mathrm{a}] < \rho[\pi(\boldsymbol{\theta}|X^s), a_\mathrm{r}] \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{3.2}$$

Applying the definition of expected loss [Eq. (2.7)] yields

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \int_\Theta L(\boldsymbol{\theta}, a_\mathrm{a}, s)\mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) < \int_\Theta L(\boldsymbol{\theta}, a_\mathrm{r}, s)\mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \\ a_\mathrm{r} & \text{otherwise} \end{cases} , \tag{3.3}$$

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. (3.1)], this reduces to

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \int_{\Theta_\mathrm{r}} \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) < \int_{\Theta_\mathrm{a}} \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{3.4}$$

However, these integrands are just the posterior probabilities:

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{r}] < \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}] \\ a_\mathrm{r} & \text{otherwise} \end{cases} , \tag{3.5}$$

or equivalently,

$$\delta_s(X^s) = \begin{cases} a_\mathrm{a} & \text{if } \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_\mathrm{a}] < 0.5 \\ a_\mathrm{r} & \text{otherwise} \end{cases} . \tag{3.6}$$

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:

$$\rho[\pi(\boldsymbol{\theta}|X^s), \delta_s(X^s)] = \begin{cases} \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{r}}] & \text{if } \delta_s(X^s) = a_{\mathrm{a}} \\ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}] & \text{if } \delta_s(X^s) = a_{\mathrm{r}} \end{cases} \tag{3.7a}$$

$$= \min\left\{\Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}],\ \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{r}}]\right\} \tag{3.7b}$$

$$= \min\left\{\Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}],\ 1 - \Pr^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}]\right\} \tag{3.7c}$$

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

4 The Pre-posterior Problem

With the notation and formalism from Sections 2 and 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

$$r(\pi, \delta_s, s) = E^{\pi} E_{\boldsymbol{\theta}}^{X^s} \{L[\boldsymbol{\theta}, \delta_s(X^s), s]\} = \int_{\Theta} \int_{\mathscr{X}^s} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{X^s|\boldsymbol{\theta}}(x) \mathrm{d}F^{\pi}(\boldsymbol{\theta}), \quad (4.1)$$

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:

$$\int_{\mathscr{X}^s} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{X^s|\boldsymbol{\theta}}(x), \quad (4.2)$$

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. (4.1), it is more straightforward to utilize the alternate form

$$r(\pi, \delta_s, s) = E^{m} \{\rho[\pi(\boldsymbol{\theta}|X^s), \delta_s(X^s)]\} = \int_{\mathscr{X}^s} \int_{\Theta} L[\boldsymbol{\theta}, \delta_s(X^s), s] \mathrm{d}F^{\boldsymbol{\theta}|X^s}(\boldsymbol{\theta}) \mathrm{d}F^{m}(x). \quad (4.3)$$

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. (3.7)]

$$r(\pi, \delta_s, s) = \int_{\mathscr{X}^s} \min \left\{ \mathrm{Pr}^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}],\ \mathrm{Pr}^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{r}}] \right\} m(x) \mathrm{d}x, \quad (4.4a)$$

where

$$\mathrm{Pr}^{\pi(\boldsymbol{\theta}|X^s)}[\Theta_{\mathrm{a}}] = \int_{\Theta_{\mathrm{a}}} \frac{f(X^s|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(X^s)} \mathrm{d}\boldsymbol{\theta} = \int_{\Theta_{\mathrm{a}}} \pi(\boldsymbol{\theta}|X^s) \mathrm{d}\boldsymbol{\theta}. \quad (4.4b)$$

Note that Eq. (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 (4.4a) may also be expressed as

$$r(\pi, \delta_s, s) = \int_{\mathscr{X}^s} \min_{\Theta'} \left\{ \int_{\Theta'} f(x|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right\} \mathrm{d}x, \quad (4.5)$$

where Θ' is either Θ_{a} or Θ_{r}. Equation (4.4a) can be incorporated as part of an objective functional:

$$J_1 = r(\pi, \delta_s, s) + \lambda C(s) \tag{4.6a}$$

or

$$J_2 = r(\pi, \delta_s, s) \times [1 + C(s)], \tag{4.6b}$$

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 $\hat{s}$:

$$\hat{s} = \arg\min_{s} J, \tag{4.7}$$

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

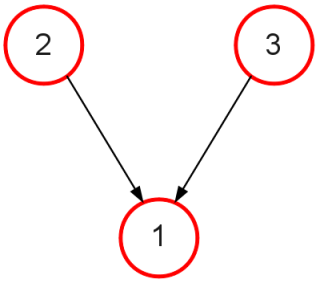

Figure 2: Graphical representation of a simple 3-node example model.

5 Examples: Limiting Cases

Here we work out in detail some small examples to show that our formula for risk makes sense and gives the correct result in some limiting cases. This also gives some insight into the process.

Consider a simple three-node system consisting of two leaf/parent nodes and one root/system node, illustrated in Fig. 2.

Output Ω_2 and Ω_3 at the leaf nodes 2 and 3, are given by respective PCEs [Eq. (2.1)] and output Ω_1 at the root node "1" is given by $\Omega_1 = \Omega_1(\Omega_2, \Omega_3)$. In this example, we consider two sampling schemes, corresponding to collecting additional data at either node 2 or 3.

For our purposes, we make the simplifying assumption that we can partition the subspaces Θ_a and Θ_r by node: $\Theta_{\mathrm{a},2}$, $\Theta_{\mathrm{r},2}$, $\Theta_{\mathrm{a},3}$ and $\Theta_{\mathrm{r},3}$, and posit that

$$\Theta_\mathrm{a} = \Theta_{\mathrm{a},2} \otimes \Theta_{\mathrm{a},3} \tag{5.1a}$$

and

$$\Theta_\mathrm{r} = \{\Theta_{\mathrm{a},2} \otimes \Theta_{\mathrm{r},3}, \Theta_{\mathrm{r},2} \otimes \Theta_{\mathrm{a},3}, \Theta_{\mathrm{r},2} \otimes \Theta_{\mathrm{r},3}\}\,. \tag{5.1b}$$

Using this decomposition of Θ, consider $r(\pi, \delta_2, 2)$ from Eq. (4.5):

$$r(\pi, \delta_2, 2) = \int_{\mathscr{X}^2} \min_{\Theta'} \left\{ \int_{\Theta'} f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \right\} \mathrm{d}x, \tag{5.2a}$$

$$\begin{aligned} &= \int_{\mathscr{X}^2} \min \left\{ \int_{\Theta_{\mathrm{a},2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \times \int_{\Theta_{\mathrm{a},3}} \pi(\theta_3)\mathrm{d}\theta_3, \right. \\ &\quad \left. \sum_{(i,j)} \int_{\Theta_{i,2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \times \int_{\Theta_{j,3}} \pi(\theta_3)\mathrm{d}\theta_3 \right\} \mathrm{d}x, \end{aligned} \tag{5.2b}$$

where the indices i and j specify pairs $\{(\mathrm{a},\mathrm{r}), (\mathrm{r},\mathrm{a}), (\mathrm{r},\mathrm{r})\}$ corresponding to integration over the "reject" subspace Θ_r. By exchanging the node indices of Eq. (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.

5.1 Example 1: Node 2 is completely characterized

Suppose node 2 represents a sub-system that is completely characterized. In other words, its prior is a multivariate delta function given by

$$\pi(\theta_2) = \delta\left(\theta_2^{\mathrm{t}} - \theta_2\right), \tag{5.3}$$

where, without loss of generality, we assume $\theta_2^{\mathrm{t}} \in \Theta_{\mathrm{a},2}$ is the true value for θ_2. We define

$$\Pi_{j,k} = \int_{\Theta_{j,k}} \pi(\theta_k)\mathrm{d}\theta_k, \tag{5.4}$$

where j is either a or r and k is 2 or 3. Thus from our assumptions, $\Pi_{a,2} = 1$ and $\Pi_{r,2} = 0$. Now, from [Eq. (5.2)], we compute

$$\begin{aligned}
r(\pi, \delta_2, 2) &= \int_{\mathscr{X}^2} \min\left\{ \Pi_{a,3} \int_{\Theta_{\mathrm{a},2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2, \right. && (5.5)\\
&\qquad \left. \Pi_{r,3} \int_{\Theta_{i,2}} f(x|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2 \right\}\mathrm{d}x && (5.6)\\
&= \int_{\mathscr{X}^2} \min\left\{ \Pi_{a,3}, \Pi_{r,3} \right\} f(x|\theta_2)\mathrm{d}x && (5.7)\\
&= \min\left\{ \Pi_{a,3}, \Pi_{r,3} \right\} \int_{\mathscr{X}^2} f(x|\theta_2)\mathrm{d}x && (5.8)\\
&= \min\left\{ \Pi_{a,3}, \Pi_{r,3} \right\}. && (5.9)
\end{aligned}$$

This says that the risk of testing a node that is already completely characterized, i.e., where all of the epistemic uncertainty has already been removed, is independent of that node. That is, no useful information will be gained. This is in accordance with our intuition. If node 3 is also completely characterized, then it follows that the risk of testing at either node 2 or node 3 is 0; testing at either node yields no new information, but there is no risk involved.

5.2 Example 2: Node 2 is completely characterized, but node 3 is not

As above and recalling that $\Pi_{a,2} = 1$ and $\Pi_{r,2} = 0$, we can write the risk at node 3 as

$$\begin{aligned} r(\pi, \delta_3, 3) &= \int_{\mathscr{X}^3} \min \left\{ \int_{\Theta_{\mathrm{a},3}} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3, \right. && (5.10) \\ &\qquad \left. \int_{\Theta_{\mathrm{r},3}} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3 \right\} \mathrm{d}x && (5.11) \\ &\leq \min \left\{ \int_{\mathscr{X}^3} \int_{\Theta_{\mathrm{a},3}} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3\mathrm{d}x, && (5.12) \\ &\qquad \left. \int_{\mathscr{X}^3} \int_{\Theta_{i,3}} f(x|\theta_3)\pi(\theta_3)\mathrm{d}\theta_3\mathrm{d}x \right\}. && (5.13) \end{aligned}$$

Now, by interchanging the order of integration using Fubini's theorem and recalling that

$$\int_{\mathscr{X}^3} f(x|\theta_3)\mathrm{d}x = 1,$$

for any value of θ_3, we have

$$\begin{aligned} r(\pi, \delta_3, 3) &\leq \min \left\{ \int_{\Theta_{\mathrm{a},3}} \pi(\theta_3)\mathrm{d}\theta_3, \int_{\Theta_{\mathrm{r},2}} \pi(\theta_3)\mathrm{d}\theta_3 \right\} && (5.14) \\ &\leq \min \left\{ \Pi_{a,3}, \Pi_{r,3} \right\} && (5.15) \\ &\leq r(\pi, \delta_2, 2). && (5.16) \end{aligned}$$

This shows that the risk of testing at node 3 when node 3 is not completely characterized is always less than testing at node 2, again in accordance with our intuition.

6 Numerical Implementation

6.1 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:

$$\theta_{ij} = \begin{cases} \theta_{ij} = \theta_{ij}^1 + \theta_{ij}^2 t & \text{linear time dependence} \\ \theta_{ij} = \theta_{ij}^1 + \theta_{ij}^2 \exp\left(-\theta_{ij}^3 t\right) & \text{exponential time dependence} \end{cases}. \tag{6.1}$$

With a sampling of ξ_j for each leaf node and the resulting output Ω_1 [determined from Eqs. (2.1) and (2.2)], we then set $\underline{\Omega}$ and $\overline{\Omega}$ such that $\mathscr{R}(\boldsymbol{\theta}^{\mathrm{t}}) = \mathscr{R}_{\mathrm{req}}$ (e.g., a desired reliability of $\mathscr{R}_{\mathrm{req}} = 0.8$):

$$\mathscr{R}(\boldsymbol{\theta}^{\mathrm{t}}) = \int_{\underline{\Omega}}^{\overline{\Omega}} D(\Omega_1|\boldsymbol{\theta}^{\mathrm{t}})\mathrm{d}\Omega_1 = \mathscr{R}_{\mathrm{req}}. \tag{6.2}$$

This calculation allows for elements of Θ to be categorized as in either Θ_{a} or Θ_{r}, which is necessary to evaluate Eq. (4.5). For our initial optimizations, time will be fixed (e.g., $t = 0$ or $t = 1$) to simplify calculations.

6.2 Integration limits

State of initial prior distributions

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):

$$\pi(\theta_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(\theta_{0j} - \mu_j)}{2\sigma_j^2}\right] \times \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} \nu_j^{\alpha-1} \exp\left(-\frac{\nu_j}{\beta}\right), \tag{6.3}$$

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 estimated or known for a given node/sub-component, e.g., $\sigma_j = 0.5$, and $\alpha = \beta = 1$ for all j. This prior can be used to estimate reasonable integration limits for Eq. (4.5):

$$\mu_j - b_0\sigma_j \leq \theta_{0j} \leq \mu_j + b_0\sigma_j \tag{6.4a}$$

and

$$0 \leq \sum_{i=1}^{q_j} \theta_{ij}^2/i! \leq b_1, \tag{6.4b}$$

where $\nu_j = \sum_i \theta_{ij}^2/i! \leq b_1$ 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 $\prod_{i=1}^{q_j} 2\sqrt{b_1 i!}$. In practice, $b_0 = 3$ and $b_1 = 5$ may be sufficient

Limits of the sampling spaces

Integration limits for the sampling spaces can be based on the available data or truth model to get an estimate on $\mathscr{X}^s$, e.g., within a few standard deviations of the mean.

6.3 Evaluation of the Bayes risk of a decision rule

As an illustrative example, we consider the simple three-node system consisting of two leaf nodes and one root node in Fig. 2, with only 2 possible sampling strategies, $s = 2$ or $s = 3$, corresponding to sampling at node 2 or node 3, respectively.

For convenience, Eq. (4.5) is repeated here:

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

where $\Theta' = \{\Theta_a, \Theta_r\}$. For sampling schemes at leaf nodes, this 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 [7, 12] is a suitable method for numerical integration of Bayes risk:

$$r(\pi, \delta_s, s) \approx \frac{|\mathscr{X}^s|}{N_x} \sum_{i=1}^{N_x} \min_{\Theta'} \left\{ \frac{|\Theta'|}{N_{\theta'}} \sum_{j=1}^{N_{\theta'}} f\left(x^{(i)}|\boldsymbol{\theta}^{(j)}\right) \pi\left(\boldsymbol{\theta}^{(j)}\right) \right\}, \tag{6.5}$$

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. (6.5) requires *two* integrations of $f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ because of the minimization over Θ' involves Θ_a and Θ_r. For a given leaf node sampling scheme s, the $N_{\theta'}$ sample integration points will be generated according to the following routine:

1. Based on the limits specified by Eq. (6.4), generate a set of N_θ sample points, each denoted as $\boldsymbol{\theta}^{(i)}$, via, e.g., quasi-random sequences such as the Faure sequence [13, 7, 12].

2. For every point, determine whether $\boldsymbol{\theta}^{(i)}$ is an element of Θ_a or Θ_r. This calculation is similar to what is described in Section 6.1: If $\mathscr{R}\left[\boldsymbol{\theta}(i)\right] \geq \mathscr{R}_\mathrm{req}$, 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. (6.5) is evaluated in the following manner. Let

$$\Omega(\xi|\boldsymbol{\theta}) = \sum_{i=1}^{q} \theta_i H_i(\xi), \tag{6.6}$$

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

$$f(x|\boldsymbol{\theta}) = \sum_{\xi^* \in \Xi^*} \frac{P_\xi(\xi^*)}{|\Omega'(\xi^*|\theta)|}, \tag{6.7}$$

where $P_\xi(\cdot)$ is the known probability distribution of ξ [in this work, $\xi \sim \mathcal{N}(0,1)$] and Ξ^* is the set of roots of $x - \Omega(\xi)$. The set Ξ^* is determined from an efficient polynomial solver [10, 11].

7 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. (3.1)] and decision rule [Eq. (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. (6.5) 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] Dante Alighieri. *La Divina Commedia*. 1555. Canticas 1.

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

[3] 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 SAND2007-2758, Sandia National Laboratories, 2007.

[4] 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 SAND2009-xxxx, Sandia National Laboratories, 2009.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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, Aug 2009.

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

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

[13] 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.

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

Sandia National Laboratories