ESAIM: M2AN 47 (2013) 635–662
DOI: 10.1051/m2an/2012038

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

DISTINGUISHING AND INTEGRATING ALEATORIC AND EPISTEMIC VARIATION IN UNCERTAINTY QUANTIFICATION*

Kamaljit Chowdhary[1] and Paul Dupuis[2]

Abstract. Much of uncertainty quantification to date has focused on determining the effect of variables modeled probabilistically, and with a known distribution, on some physical or engineering system. We develop methods to obtain information on the system when the distributions of some variables are known exactly, others are known only approximately, and perhaps others are not modeled as random variables at all. The main tool used is the duality between risk-sensitive integrals and relative entropy, and we obtain explicit bounds on standard performance measures (variances, exceedance probabilities) over families of distributions whose distance from a nominal distribution is measured by relative entropy. The evaluation of the risk-sensitive expectations is based on polynomial chaos expansions, which help keep the computational aspects tractable.

Mathematics Subject Classification. 41A10, 60H35, 65C30, 65C50.

Received March 22, 2011. Revised March 13, 2012.
Published online March 29, 2013.

1. Introduction

Uncertainty quantification refers to a broad set of techniques for understanding the impact of uncertainties in complicated mechanical, physical, and computational systems. In this context "uncertainty" can take on many meanings, but we follow the convention of dividing uncertainty into aleatoric and epistemic categories. In short, aleatoric uncertainty refers to inherent uncertainty due to stochastic or probabilistic variability. This type of uncertainty is *irreducible* in that there will always be positive variance since the underlying variables are truly random. Epistemic uncertainty refers to limited knowledge we may have about the model or system. This type of uncertainty is *reducible* in that if we have more information, *e.g.*, take more measurements, then this type of uncertainty can be reduced. However, for many problems where uncertainty quantification is important, the acquisition of data is difficult or expensive. The epistemic uncertainty cannot be removed entirely, and so one needs modeling and computational techniques which can also accommodate this form of uncertainty.

Keywords and phrases. Epistemic uncertainty, aleatoric uncertainty, generalized polynomial chaos, relative entropy, uncertainty quantification, spectral methods, stochastic differential equations, monte Carlo integration, stochastic collocation method, quadrature.

* *This research was supported in part by the Air Force Office of Scientific Research (FA9550-07-1-0425 and FA9550-09-1-0378), the National Science Foundation (DMS-1008331), and the Department of Energy (DE-SCOO02413).*

[1] Division of Applied Mathematics, Brown University, Providence, RI, 02912,USA. kchowdhary@brown.edu

[2] Lefschetz Center for Dynamical Systems, Division of Applied Mathematics,Brown University, Providence, RI, 02912, USA. dupuis@dam.brown.edu

Article published by EDP Sciences

© EDP Sciences, SMAI 2013

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Much of the work to date has focused on aleatoric uncertainty and what one might call the propagation of uncertainty. Here, one characterizes uncertainty concerning one or more elements of a physical system *via* some probability distribution, and attempts to quantify how that uncertainty will be propagated throughout the system under its constitutive equations and laws. The simplest and most straightforward method to characterize this output distribution would be to sample from the input probability distribution, solve the system equations, and thereby produce output samples (*i.e.*, standard Monte Carlo). Although this is very simple conceptually, it can be far from practical, especially when it is computationally intensive to construct the mapping that takes input variables to output. Straightforward schemes such as Monte Carlo would require that the system be solved for each sample of the random variable, a situation that is often not practical. In recent years efficient computational methods have been developed to calculate particular functionals of the induced distribution that may be required for a particular application (*e.g.*, covariances or error probabilities). Most recently, polynomial chaos has become popular as an efficient method for approximating the distribution of the output variable.

We note that random variables with a well-defined and given distribution are often used in this context even when there is no justification for their use, as in the case of modeling error. In other words, it is (perhaps implicitly) assumed that epistemic uncertainty can be modeled by aleatoric uncertainty. One reason is that, at least as they have been developed to date, most computational techniques (*e.g.*, polynomial chaos and Monte Carlo) are based on the assumption that the user can identify some "appropriate" distribution for each uncertain aspect of the system, regardless of the type of uncertainty, aleatoric or epistemic. If one is interested in just basic qualitative properties of the system then this may not be a central issue, since virtually any model of uncertainty will give information on the sensitivities of the system. However, when the intended use of uncertainty quantification is for regulatory assessment or some other application where performance measures are sensitive to distributional assumptions, the issue becomes much more important, and one should carefully distinguish how one accounts for the two types of uncertainty.

The aim of the present paper is to describe an approach that (i) logically distinguishes those aspects of uncertainty that are treated as stochastic variability from other forms of uncertainty, (ii) in cases where a stochastic model is theoretically valid but for which determination of the distribution is not practical, gives bounds for performance measures that are valid for explicitly identified families of distributions, and (iii) is computationally feasible if ordinary uncertainty propagation is feasible. The intended audience is broad, including numerical analysts interested in robust performance bounds as well as applied probabilists for whom certain computational aspects of uncertainty quantification may be novel. Since a typical reader might not be familiar with the terminology and methods from both fields, we have included background material for both the probabilistic and numerical analysis approaches used to make the paper broadly accessible.

The paper is organized as follows. In Section 2 we further discuss the distinction between aleatoric and epistemic uncertainty, and explain some of the limitations to achieving useful performance bounds in the presence of epistemic uncertainty. Section 3 introduces risk-sensitive performance measures and discusses their robust properties. Several hybrid forms are introduced that are more useful than the simplest form when both aleatoric and epistemic uncertainties are present. Various properties of the risk-sensitive measures that are useful in applications (monotonicity, optimization) are also discussed. Section 4 presents numerical examples, reviews the computational methods used to evaluate the risk-sensitive performance measures, and finishes with a Section of discussion and conclusions.

2. Aleatoric and epistemic uncertainties

As noted in the Introduction, *uncertainty* can be divided into two categories, aleatoric and epistemic. From the perspective of mathematical formulation and modeling, aleatoric uncertainty is in some sense simpler. In contrast, epistemic uncertainty can mean different things in different contexts. *Lack of knowledge* is an ambiguous term that encompasses many different scenarios.

To illustrate, we consider an elementary example. Consider a system described by some well-posed partial differential equation (PDE), but with an uncertain boundary condition. The boundary condition is expressed in

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

terms of a random variable X (*e.g.*, $u_x(0,t) = X$) with a particular fixed distribution, and the numerical analysis problem is to characterize the induced distribution of the solution to the PDE at some time and location (*e.g.*, $u(x_0, t_0; X)$). Suppose we know that the boundary condition is properly modeled as a random variable, but we do not know the correct value of some parameter in that distribution. For example, based on a central limit type argument, one might claim the distribution is known to be Gaussian, but still unknown are the "true" mean and/or variance. In this case, the *lack of knowledge* is this missing information about the parameters of the distribution. The lack of information regarding these parameters is a form of epistemic uncertainty. Hence in this example aleatoric and epistemic uncertainty are mingled. Assuming that samples are available, one could use the empirical mean and variance from data to approximate the true mean and variance, and thus reduce this uncertainty. However, in the common situation where sampling is necessarily limited, some epistemic uncertainty is inherent in the model due to practical limitations on data acquisition and modeling.

Another type of epistemic uncertainty is the omission of important aspects of the system model. This form is possibly the most difficult to quantify. For example, there might be a hidden random variable in the true physical system. In the PDE example we have assumed the boundary condition is modeled *via* a random variable, but there may be other coefficients in the true physical system that are random, but are modeled as constants. It is also frequently true that the mathematical model used for the system is only approximately true. For example, the model might simplify the geometry of the true system, nonlinearities may be approximated by linear relations, etc. In the context of the PDE example, boundary conditions of Dirichlet form were chosen, when in fact a more realistic model might use a mixed form (*e.g.*, Robin boundary conditions).

For almost all forms of epistemic uncertainty there is little justification for the use of random variables to model the uncertainty. Nonetheless, it is common practice to do just that, and in practice randomness and random variables are used in many situations to account for errors in the physical model or modeling ignorance. The reasons were mentioned previously–that basic distributional properties of the output (*e.g.*, variance) will still provide a reasonable sensitivity analysis for the problem, and existing computational methods are largely based on aleatoric uncertainty. However, with little justification for the use of any particular distribution (or even the use of randomness at all), in more critical applications one may insist on rigorous bounds on performance that are valid for a particular family of distributions. In the extreme case where no probabilistic model is considered acceptable, one may wish to only prescribe bounds on certain parameters, and then obtain tight bounds on performance over all values of the parameters that satisfy the bounds.

Thus there are many different types of uncertainty that one should account for in an analysis of the effects of uncertainty. These include: (i) aleatoric with known distribution; (ii) aleatoric with partly known distribution (mingled aleatoric and epistemic); (iii) epistemic for which one is willing to model by a family of aleatoric uncertainties, and (iv) epistemic where one is only willing to place bounds on the uncertainties. As remarked in the Introduction, this paper will introduce an approach that allows these uncertainties to be handled within a single framework that can exploit computational methods originally developed just for the treatment of aleatoric uncertainties.

3. Duality for exponential integrals

Our development of performance measures that distinguish forms of uncertainty, and which in particular allow for robustness with respect to epistemic uncertainties, depends on a duality relation between exponential integrals and relative entropy. Its first use along these lines but with regard to estimation appears to be in [1], and for optimization in [3]. We first state the basic duality result, and then define two functionals which will allow aleatoric and epistemic uncertainties to be analyzed simultaneously but at the same time distinguished.

The general duality is stated for random variables that take values in a Polish metric space (*i.e.*, a complete, separable metric space) $\mathcal{X}$. The associated σ-algebra is the Borel σ-algebra. A typical example of $\mathcal{X}$ for our purposes is a closed subset of some Euclidean space $\mathbb{R}^d$. Let $\mathcal{P}(\mathcal{X})$ denote the collection of probability measures on $\mathcal{X}$, and let $\mu \in \mathcal{P}(\mathcal{X})$. Given any $\nu \in \mathcal{P}(\mathcal{X})$ that is absolutely continuous with respect to μ, we define the

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

relative entropy of ν with respect to μ by

$$R(\nu\|\mu) = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\nu(\mathrm{d}x)$$

whenever $\log(\mathrm{d}\nu/\mathrm{d}\mu(x))$ is integrable with respect to ν. In all other cases $R(\nu\|\mu)$ is defined to be ∞.

Relative entropy defines a mapping $(\nu,\mu) \to R(\nu\|\mu)$ from $\mathcal{P}(\mathcal{X})^2$ to $\mathbb{R}\cup\infty$. This mapping has a number of very attractive properties. For example, relative entropy is non-negative, with $R(\nu\|\mu) = 0$ if and only if $\nu = \mu$. In addition, the mapping is jointly convex and lower semicontinuous [2], Lemma 1.4.3. It is important to note, however, that relative entropy is not a true metric since it is not symmetric, nor does it satisfy the triangle inequality. A property of particular interest for our purposes is the following variational formula for exponential integrals, which can be considered as an infinite-dimensional version of the Legendre transform. Given any bounded and continuous function $F : \mathcal{X} \to \mathbb{R}$ and any $c \in (0,\infty)$,

$$\Lambda_c \doteq \frac{1}{c}\log\int_{\mathcal{X}} \mathrm{e}^{cF(x)}\mu(\mathrm{d}x) = \sup_{\nu\in\mathcal{P}(\mathcal{X})}\left[-\frac{1}{c}R(\nu\|\mu) + \int_{\mathcal{X}} F(x)\nu(\mathrm{d}x)\right]. \tag{3.1}$$

For a proof see [2], Proposition 1.4.2. The duality continues to hold as stated if F is bounded from above, and also when bounded from below if one restricts the supremum to $\nu \in \mathcal{P}(\mathcal{X})$ for which $R(\nu\|\mu) < \infty$ [2], Proposition 4.5.1.

As an elementary example on how such a formula can be useful, suppose that F is a performance measure (*e.g.*, variance or an error probability), and that μ is a model for some random phenomena. We consider μ to be the *nominal model*, *e.g.*, our best guess[3]. However, we are uncertain if this is the correct model, and would like a measure of performance that also holds for alternative models, but with a penalty for deviation from the nominal model. Λ_c is just such a performance measure, since by (3.1), for any alternative model ν the bound

$$\int_{\mathcal{X}} F(x)\nu(\mathrm{d}x) \le \Lambda_c + \frac{1}{c}R(\nu\|\mu) \tag{3.2}$$

applies. Hence we obtain bounds on performance over a family of alternative models. The parameter c allows one to balance robustness with respect to possible model inaccuracies against tighter bounds. In particular, as c tends to 0, Λ_c converges to $\int_{\mathcal{X}} F(x)\mu(\mathrm{d}x)$, which is the performance measure under the nominal model, but in the limit the bound is meaningful only when $\nu = \mu$.

Suppose that a bound on performance over a specific family of distributions is needed. Let R^* denote the maximum of relative entropy with respect to the nominal model over this family. Then the tightest possible bound is obtained by minimizing

$$\Lambda_c + \frac{1}{c}R^*$$

over $c > 0$. We show in Proposition 3.3 below that this function has only one local minimum over $c \in (0,\infty]$, and thus the global minimum is easy to compute.

Functionals of the exponential form that appears in the definition of Λ_c are sometimes called *risk-sensitive* because the exponential amplifies the effect of any large values of the performance measure F. In the next two sections we consider two hybrid risk-sensitive functionals that will allow aleatoric and epistemic variables to be combined and yet distinguished.

3.1. Two hybrid forms

In this section and the next random variables with a *known* distribution will take values in a Polish space $\mathcal{X}$. Variables whose distribution is *not known* or are otherwise of the epistemic variety take values in the Polish metric space $\mathcal{Y}$.

[3]This is similar to the prior distribution in the Bayesian framework.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

The performance measure of interest for some given problem is assumed to be of the form

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x,y)\lambda(\mathrm{d}y)\nu(\mathrm{d}x), \tag{3.3}$$

where ν (resp., λ) is a probability measure on $\mathcal{X}$ (resp., $\mathcal{Y}$). If X and Y are independent random variables with distributions ν and λ, then $F(X,Y)$ represents both the performance measure (*e.g.*, a second moment) as well as the underlying physical or mechanical system that maps these aleatoric and epistemic inputs into outputs. Now, let μ and γ represent the nominal distribution of X and Y, resp., while ν and λ represent the alternative probability distributions of X and Y, resp. Applying (3.2) to the joint measures $\mu \times \gamma$ (nominal) and $\nu \times \lambda$ (alternative), we obtain bounds on the performance measure

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x,y)\lambda(\mathrm{d}y)\nu(\mathrm{d}x) \leq \frac{1}{c} R\left(\lambda \,\|\gamma\right) + \frac{1}{c} R\left(\nu \,\|\mu\right) + \Lambda_c \tag{3.4}$$

where the ordinary risk-sensitive performance measure is

$$\Lambda_c = \frac{1}{c} \log \int_{\mathcal{X}} \int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)} \gamma(\mathrm{d}y)\mu(\mathrm{d}x). \tag{3.5}$$

Since the distribution of X is known, we can take $\nu = \mu$ so that (3.4) now becomes

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x,y)\lambda(\mathrm{d}y)\nu(\mathrm{d}x) \leq \frac{1}{c} R\left(\lambda \,\|\gamma\right) + \Lambda_c. \tag{3.6}$$

While (3.6) seems to give upper bounds on the performance measure under uncertainty in the distribution of Y only, the measure (3.5) does not distinguish the variables according to type (aleatoric or epistemic). In fact, the risk-sensitive performance measure (3.5) is the same in both (3.4) and (3.6). Therefore, the use of a risk-sensitive form (3.5) is not well motivated for the aleatoric variables. Indeed, use of this measure will give bounds that are robust with respect to variations on a distribution that is known (as can be seen in (3.4)), and obviously such bounds are not as tight as possible.

The first form we consider for a hybrid measure is

$$\Lambda_c^1 = \frac{1}{c} \log \int_{\mathcal{Y}} \mathrm{e}^{\int_{\mathcal{X}} cF(x,y)\mu(\mathrm{d}x)} \gamma(\mathrm{d}y). \tag{3.7}$$

Note that by Jensen's inequality (applied to the convex function exp), this is smaller than (and in general strictly smaller than) Λ_c. Using the relative entropy representation for exponential integrals given in (3.1) [but with $\mathcal{Y}$ in place of $\mathcal{X}$ and letting $\int_{\mathcal{X}} cF(x,y)\mu(\mathrm{d}x)$ be the cost function], it follows that for any distribution $\theta(\mathrm{d}y)$

$$\int_{\mathcal{Y}} \int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x)\theta(\mathrm{d}y) \leq \frac{1}{c} R\left(\theta(\mathrm{d}y) \,\|\gamma(\mathrm{d}y)\right) + \Lambda_c^1. \tag{3.8}$$

This gives a bound on the performance measure for an arbitrary distribution of Y, but with the distribution of X equal to the known true distribution. The distributions thus play very different roles. In particular, we think of γ as a *nominal* distribution of Y, which should be distinguished from a possible *true* distribution. The risk sensitive functional Λ_c^1, whose numerical evaluation can be carried out by a variety of methods (including polynomial chaos expansion as discussed below), is based on the nominal distribution. Through the relative entropy duality, it will yield various bounds (depending on c) on a families of distributions, with the relative entropy distance the key metric. Ideally one would consider the smallest family that contains the "true" distribution (if it exists), and get the tightest bound by optimizing over c. With the hybrid risk-sensitive formulation, μ is allowed to be both the nominal (computational) distribution and also the true distribution.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

The second form one could consider for a hybrid measure is

$$\Lambda_c^2 = \frac{1}{c}\int_{\mathcal{X}} \left[\log \int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)}\gamma(\mathrm{d}y)\right]\mu(\mathrm{d}x). \tag{3.9}$$

Note that by Jensen's inequality (applied now to the concave function log), this is again in general strictly smaller than Λ_c. The relative entropy representation for exponential integrals gives

$$\int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y|x) \le \frac{1}{c} R\left(\theta(\mathrm{d}y|x)\,\|\gamma(\mathrm{d}y)\right) + \frac{1}{c}\log\int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)}\gamma(\mathrm{d}y).$$

Here $\theta(\mathrm{d}y|x)$ is any stochastic kernel on $\mathcal{Y}$ given $\mathcal{X}$, which is essentially a conditional distribution on $\mathcal{Y}$ given $X = x$ (for the precise definition see [2], p. 35). Integrating over $\mathcal{X}$ gives

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y|x)\mu(\mathrm{d}x) \le \frac{1}{c}\int_{\mathcal{X}} R\left(\theta(\mathrm{d}y|x)\,\|\gamma(\mathrm{d}y)\right)\mu(\mathrm{d}x) + \Lambda_c^2.$$

In the case where $\theta(\mathrm{d}y|x)$ is independent of x, we obtain

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y)\mu(\mathrm{d}x) \le \frac{1}{c} R\left(\theta(\mathrm{d}y)\,\|\gamma(\mathrm{d}y)\right) + \Lambda_c^2.$$

Note that the second hybrid cost gives a more general bound, in that it allows the alternative distribution on Y (*i.e.*, $\theta(\mathrm{d}y|x)$) to depend on the value taken by X. This additional flexibility comes at a cost, and indeed we will show in the next Section that the following inequalities hold (typically in a strict fashion):

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} F(x,y)\gamma(\mathrm{d}y)\mu(\mathrm{d}x) \le \Lambda_c^1 \le \Lambda_c^2 \le \Lambda_c. \tag{3.10}$$

Hence if one is concerned with performance bounds for distributions on Y that do not depend on the variable X, then the tightest bound is obtained using Λ_c^1.

Before studying properties of the risk-sensitive measures, we note some elementary generalizations that are possible. Suppose that the nominal distribution of X and Y is not of product form. In the setting of the first form, we could let γ be the marginal distribution of Y and let $\mu(\mathrm{d}x|y)$ be the conditional distribution of X given $Y = y$. Using the extended definition

$$\bar{\Lambda}_c^1 = \frac{1}{c}\log\int_{\mathcal{Y}} \mathrm{e}^{\int_{\mathcal{X}} cF(x,y)\mu(\mathrm{d}x|y)}\gamma(\mathrm{d}y),$$

we obtain the bound

$$\int_{\mathcal{Y}}\int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x|y)\theta(\mathrm{d}y) \le \frac{1}{c} R\left(\theta(\mathrm{d}y)\,\|\gamma(\mathrm{d}y)\right) + \bar{\Lambda}_c^1.$$

For the second form, we let μ be the marginal distribution of X and let $\gamma(\mathrm{d}y|x)$ be the conditional distribution of Y given $X = x$. With the definition

$$\bar{\Lambda}_c^2 = \frac{1}{c}\int_{\mathcal{X}} \left[\log\int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)}\gamma(\mathrm{d}y|x)\right]\mu(\mathrm{d}x),$$

we obtain

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y|x)\mu(\mathrm{d}x) \le \frac{1}{c}\int_{\mathcal{X}} R\left(\theta(\mathrm{d}y|x)\,\|\gamma(\mathrm{d}y|x)\right)\mu(\mathrm{d}x) + \bar{\Lambda}_c^2.$$

It is indeed sometimes useful to allow the distribution of one type of variable to depend on the value taken by the other type of variable. As an elementary example, consider the situation where an aleatoric variable appears in a system, and while it is known that this variable has a Gaussian distribution with mean zero, the variance of the variable is however not known. Then the variance itself might be modeled as an uncertainty whose distribution is not known precisely, *i.e.*, as an epistemic uncertainty. In this case $\bar{\Lambda}_c^1$ would be the relevant risk-sensitive formulation. An example of this sort is presented in Section 4.2.2.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Remark 3.1. In problems of regulatory assessment epistemic uncertainty can pose a unique set of challenges. The collection of data and model validation are often particularly difficult or expensive, while at the same time stringent bounds on performance might be demanded. In these circumstances, the framework presented here, wherein tight bounds are obtained for a known family of alternative models, would seem very natural. Prior to the collection of data or testing of samples to determine compliance, all critical elements of the model, including selection of the nominal model, size of the family of alternative models allowed by the relative entropy bounds, and the performance measures that would be required to hold uniformly for this family, would all be determined by negotiation among the interested parties. This would allow various competing needs (*e.g.*, expense of model validation versus adequate protection of consumers) to be balanced according to the interests of the participants.

3.2. Properties of the risk-sensitive forms

In this section we prove properties of the two hybrid risk-sensitive forms. In particular we derive an inequality relating the two, and study the limit as $c \to \infty$. For the results to hold as stated we often need F be bounded from below. This is a mild assumption. Indeed, standard measures of performance such as second moments and error probabilities satisfy this condition.

The first proposition shows the inequality between the two hybrid forms.

Proposition 3.2. *Assume that F is bounded from below and consider the functionals defined in* (3.3)–(3.9). *Then the inequalities* (3.10) *hold.*

Proof. The only inequality which does not follow from Jensen's inequality is $\Lambda_c^1 \le \Lambda_c^2$. The relative entropy duality as stated in (3.1) applied to Λ_c^1 gives

$$\Lambda_c^1 = \sup_{\theta \in \mathcal{P}(\mathcal{Y})} \left[\int_{\mathcal{Y}} \int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x)\theta(\mathrm{d}y) - \frac{1}{c} R\left(\theta(\mathrm{d}y) \,\|\gamma(\mathrm{d}y)\right) \right],$$

where $\mathcal{P}(\mathcal{Y})$ is the space of probability measures on $\mathcal{Y}$. The same representation applied to

$$\frac{1}{c} \log \int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)} \gamma(\mathrm{d}y)$$

gives

$$\frac{1}{c} \log \int_{\mathcal{Y}} \mathrm{e}^{cF(x,y)} \gamma(\mathrm{d}y) = \sup_{\theta \in \mathcal{P}(\mathcal{Y})} \left[\int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y) - \frac{1}{c} R\left(\theta(\mathrm{d}y) \,\|\gamma(\mathrm{d}y)\right) \right].$$

The supremum operation in the last display can be done in such a way that an optimizing (or near optimizing) θ is a Borel measurable function of x, *i.e.*, a stochastic kernel $\theta(\mathrm{d}y|x)$ on $\mathcal{Y}$ given $\mathcal{X}$ [2]. Let $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ denote the space of such stochastic kernels. Integrating the last display with respect to μ gives

$$\Lambda_c^2 = \sup_{\theta \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \left[\int_{\mathcal{X}} \int_{\mathcal{Y}} F(x,y)\theta(\mathrm{d}y|x)\mu(\mathrm{d}x) - \frac{1}{c} \int_{\mathcal{X}} R\left(\theta(\mathrm{d}y|x) \,\|\gamma(\mathrm{d}y)\right) \mu(\mathrm{d}x) \right].$$

Since $\mathcal{P}(\mathcal{Y}) \subset \mathcal{P}(\mathcal{Y}|\mathcal{X})$, $\Lambda_c^2 \ge \Lambda_c^1$ follows. □

The next proposition shows how to obtain bounds on the performance over a family of distributions defined in terms of a maximum relative entropy B.

Proposition 3.3. *Consider the functionals defined in* (3.5)–(3.9), *and let* $D = \{c : \Lambda_c < \infty\}$ (resp., $D^i = \{c : \Lambda_c^i < \infty\}, i = 1, 2$). *Assume that the interior of D* (resp., D^i) *is nonempty. Then Λ_c* (resp., Λ_c^i) *is differentiable on the interior of D* (resp., D^i). *Assume also that F is bounded from below by zero. Then $c \mapsto \Lambda_c$* (resp., $c \mapsto \Lambda_c^i$) *is nondecreasing for $c \ge 0$. Let $B > 0$ be given. Then there is a unique $c \in (0, \infty]$ at which*

$$c \mapsto \frac{1}{c} B + \Lambda_c \quad \left(\text{resp.}, \frac{1}{c} B + \Lambda_c^i \right)$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

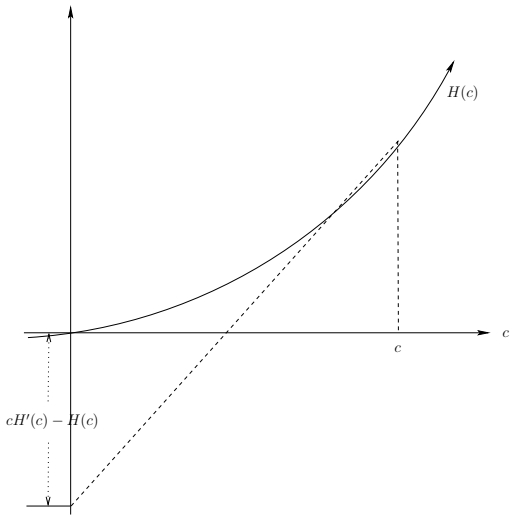


FIGURE 1. $f(c)$ is monotone increasing.

attains a local minimum, where the statement that the minimum occurs at $c = \infty$ means that $\Lambda_c + B/c > \Lambda_\infty$ for a well defined limit Λ_∞ and all $c < \infty$.

Proof. To simplify we give the proof only for the case of Λ_c and omit the y variable. Thus we consider $\Lambda_c = \frac{1}{c}\log\int_{\mathcal{X}} \mathrm{e}^{cF(x)}\mu(\mathrm{d}x)$. Proofs for the other functionals are analogous. It is well known that $H(c) = \log\int_{\mathcal{X}} \mathrm{e}^{cF(x)}\mu(\mathrm{d}x)$ is a convex function taking values in $\mathbb{R}\cup\{\infty\}$, and strictly convex and infinitely differentiable on the interior of $\{c : H(c) < \infty\}$ (see, *e.g.*, [6]).

Next we prove the monotonicity. Differentiating gives

$$H'(c) = \frac{\int_{\mathcal{X}} F(x)\mathrm{e}^{cF(x)}\mu(\mathrm{d}x)}{\int_{\mathcal{X}} \mathrm{e}^{cF(x)}\mu(\mathrm{d}x)}.$$

Since $F \geq 0$ the derivative is non-negative, and convexity implies it is non-decreasing. Using

$$\frac{1}{c}H(c) = \frac{1}{c}\left(H(c) - H(0)\right) = \frac{1}{c}\int_0^c H'(s)\mathrm{d}s,$$

it follows that $\Lambda_c = H(c)/c$ is nondecreasing for $c \geq 0$.

First assume that $M \doteq \sup D = \infty$, and observe that

$$\frac{\mathrm{d}}{\mathrm{d}c}\left[\frac{1}{c}B + \frac{1}{c}H(c)\right] = \frac{1}{c^2}\left[-B + cH'(c) - H(c)\right].$$

Since H is strictly convex and $H'(0) \geq 0$, the mapping $f(c) \mapsto cH'(c) - H(c)$ is monotone increasing and takes the value 0 at $c = 0$ (see Fig. 1).

Let K denote the limit of $f(c)$ as $c \to \infty$. If $K = \infty$ then there is a unique solution to

$$cH'(c) - H(c) = B,$$

and we are done. The same is true if $0 \leq B < K$. Hence the only case left is when $B \geq K$. In this case $(B + H(c))/c$ is monotone decreasing for all $c \in [0, \infty)$. Since $H(c)/c \geq H'(0) \geq 0$, there is a well-defined limit Λ_∞ that is necessarily the minimum.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Next assume that $M \doteq \sup D \in (0,\infty)$. We claim that in this case $cH^{'}(c) - H(c) \uparrow \infty$ as $c \uparrow M$, and therefore we can argue just as before. By monotone convergence it must be that $H(c) \uparrow \infty$ as $c \uparrow M$, which implies that necessarily $H^{'}(c) \uparrow \infty$ as $c \uparrow M$. To prove the claim, let $0 < \beta < c < M$. Then since $H^{'}(s)$ is increasing

$$H(c) = \int_0^c H^{'}(s)\mathrm{d}s \leq \beta H^{'}(\beta) + (c-\beta)H^{'}(c).$$

This implies

$$cH^{'}(c) - H(c) \geq \beta \left[H^{'}(c) - H^{'}(\beta)\right].$$

Letting $c \uparrow M$ and using that $\beta > 0$, the claim is proved, thus completing the proof of the theorem. □

Next, we consider the sense in which the infimum over c of right hand side of (3.2) provides a tight bound on the performance measure. To simplify we give the proof only for the case of Λ_c and omit the y variable. The proofs for the hybrid forms (*e.g.*, (3.8)) are analogous.

Theorem 3.4. *Consider the functional defined in* (3.1) *and assume the conditions of Proposition* 3.3. *Suppose that* $L \in \mathbb{R}$, $B \in [0,\infty)$ *are given, and that* $c^* < \infty$ *minimizes*

$$c \mapsto \frac{1}{c}B_c + \Lambda_c$$

so that $\frac{1}{c^*}B + \Lambda_{c^*} < \infty$. *Define* $\mathcal{F}_B \doteq \{\gamma \in \mathcal{P}(\mathcal{X}) : R(\gamma \,\|\mu) \leq B\}$, i.e., *the family of alternative distributions on* $\mathcal{X}$. *Then for all* $\nu \in \mathcal{F}_B$,

$$\int_{\mathcal{X}} F(x)\nu(\mathrm{d}x) \leq L \Leftrightarrow \frac{1}{c^*}B + \Lambda_{c^*} \leq L.$$

Proof. If $\frac{1}{c^*}B + \Lambda_{c^*} \leq L$ then, by (3.2), $\int_{\mathcal{X}} F(x)\nu(\mathrm{d}x) \leq L$ for all $\nu \in \mathcal{F}_B$. For the other direction, we will show that there exists a $\nu(\mathrm{d}x) \in \mathcal{F}_B$ such that

$$\int_{\mathcal{X}} F(x)\nu(\mathrm{d}x) = \frac{1}{c^*}B + \Lambda_{c^*}.$$

Hence, if $\frac{1}{c^*}B + \Lambda_{c^*} > L$, then, for at least one $\nu(\mathrm{d}x) \in \mathcal{F}_B$, $\int_{\mathcal{X}} F(x)\nu(\mathrm{d}x) > L$.

First, recall from (3.1) that for any given $c \in (0,\infty)$

$$\Lambda_c = \sup_{\nu \in \mathcal{P}(\mathcal{X})} \left[-\frac{1}{c}R(\nu \,\|\mu) + \int_{\mathcal{X}} F(x)\nu(\mathrm{d}x)\right].$$

The supremum is achieved at $\nu^c(\mathrm{d}x) = \mathrm{e}^{cF(x)}\mu(\mathrm{d}x)/Z(c)$, where $Z(c) = \int_{\mathcal{X}} \mathrm{e}^{cF(x)}\mu(\mathrm{d}x)$ [2], Proposition 1.4.2. Thus

$$\int_{\mathcal{X}} F(x)\nu^{c^*}(\mathrm{d}x) = \frac{1}{c^*}R(\nu^{c^*} \,\|\mu) + \Lambda_{c^*}.$$

By the previous proposition, $c^* < \infty$ must lie in the interior of D. Therefore, we can differentiate $\frac{1}{c}B + \Lambda_c$ with respect to c, and we get that c^* satisfies

$$-\frac{1}{(c^*)^2}\log\int_{\mathcal{X}} \mathrm{e}^{c^*F(x)}\mu(\mathrm{d}x) + \frac{1}{c^*}\frac{1}{Z(c^*)}\int_{\mathcal{X}} F(x)\mathrm{e}^{cF(x)}\mu(\mathrm{d}x) - \frac{B}{(c^*)^2} = 0. \tag{3.11}$$

Also,

$$\begin{aligned} R(\nu^{c^*} \,\|\mu) &= \int_{\mathcal{X}} \frac{[c^*F(x) - \log Z(c^*)]}{Z(c^*)}\mathrm{e}^{c^*F}\mu(\mathrm{d}x) \\ &= \frac{1}{Z(c^*)}\int_{\mathcal{X}} c^*F(x)\mathrm{e}^{c^*F(x)}\mu(\mathrm{d}x) - \log Z(c^*), \end{aligned}$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

and so

$$\frac{1}{(c^*)^2} R(\nu^{c^*} \| \mu) = \frac{1}{Z(c^*)} \int c^* F \mathrm{e}^{c^* F} \mathrm{d}\mu - \frac{1}{(c^*)^2} \log Z(c^*). \tag{3.12}$$

Comparing (3.11) and (3.12) we have $R(\nu^{c^*} \| \mu) = B$ and therefore $\nu^{c^*} \in \mathcal{F}_B$. □

In the context of regulatory assessment, it may be required that the performance measure of interest be no greater than a particular value, *e.g.*, L. What this theorem tells us is that the performance measure will pass the regulatory criteria for a given family of alternative distributions, if and only if the upper bound also passes the regulatory criteria. If the upper bound fails to pass the regulatory requirement, all we can say is that there exists at least one distribution in the given family of alternative distributions that fails this criteria. It may very well be that there are distributions within the family that actually pass the regulatory criteria, but, as a whole, the family fails to satisfy the requirement.

One of the situations of interest when aleatoric and epistemic variables appear simultaneously is the case where all that is known regarding the epistemic variables are bounds. It turns out that this problem is well posed (*i.e.*, the relevant performance criteria are finite) essentially in those cases where $\Lambda^1_\infty < \infty$. In fact, when this is the case the value of Λ^1_∞ can be used to establish optimal bounds subject to the constraint on the epistemic variables. For simplicity, we assume that the set A appearing in the statement of the following theorem is bounded and that γ, the nominal distribution for the epistemic variables, is the uniform distribution (in the limit $c \to \infty$ the precise form of the nominal distribution is not important, and in fact it is only the support of the distribution that matters in the limit (see the remark after the Theorem)). If A is not bounded one can use an appropriate distribution whose support is all of A (*e.g.*, the exponential distribution could be used for $A = [0, \infty)$. The choice of γ as uniform when A is bounded is simplest and also one that is convenient for most uses. A related statement holds for Λ^2_∞, but it is not as useful (at least in this setting), since $\Lambda^1_\infty \le \Lambda^2_\infty$.

Theorem 3.5. *Let $\mathcal{X}$ and $\mathcal{Y}$ be subsets of a finite dimensional Euclidean space. Suppose that $A \subset \mathcal{Y}$ is bounded and the closure of its interior and that γ is the uniform distribution on A. Assume that F is lower semicontinuous in y for each $x \in \mathcal{X}$ and bounded from below. Define the risk-sensitive functional Λ^1_c by (3.7) and let $\Lambda^1_\infty = \lim_{c\to\infty} \Lambda^1_c$. Then*

$$\sup_{y \in A} \int_{\mathcal{X}} F(x, y) \mu(\mathrm{d}x) = \Lambda^1_\infty.$$

Proof. Since $c\Lambda^1_c$ is convex the limit of Λ^1_c is well defined, though it might take the value ∞. We give the proof for the case $\Lambda^1_\infty < \infty$. The extension to the case $\Lambda^1_\infty = \infty$ is straightforward.

We first prove that the left side of the last display is bounded above by the right side. Fix any $\bar{y} \in A^\circ$, the interior of A. For small $\varepsilon > 0$ let B_ε be the ball in A about $\bar{y}$ of volume ε, and let M be the volume of A. Let $\theta(\mathrm{d}y)$ be the measure whose density with respect to Lebesgue measure is $1/\varepsilon$ on B_ε, and zero elsewhere. Then

$$\begin{aligned} R\left(\theta(\mathrm{d}y) \| \gamma(\mathrm{d}y)\right) &= \int_{B_\varepsilon} \log\left[(1/\varepsilon)/(1/M)\right] \theta(\mathrm{d}y) \\ &= (1/\varepsilon) \log\left[(1/\varepsilon)/(1/M)\right]. \end{aligned}$$

Then (3.8) gives

$$\int_{\mathcal{X}} \int_{B_\varepsilon} F(x, y) \theta(\mathrm{d}y) \mu(\mathrm{d}x) \le \frac{1}{c} (1/\varepsilon) \log\left[(1/\varepsilon)/(1/M)\right] + \Lambda^1_c. \tag{3.13}$$

Since $y \to F(x, y)$ is lower semicontinuous

$$\liminf_{\varepsilon \to 0} \int_{B_\varepsilon} F(x, y) \theta(\mathrm{d}y) \ge F(x, \bar{y}).$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Letting first $c \to \infty$ and then $\varepsilon \to 0$ in (3.13), Fatou's lemma gives $\int_{\mathcal{X}} F(x,\bar{y})\mu(\mathrm{d}x) \le \Lambda^1_\infty$. Since $\bar{y} \in A^\circ$ is arbitrary, the lower semicontinuity and another use of Fatou give the bound for all $y \in A$.

For the reverse inequality, we use the fact that the minimizing measure in the definition of Λ^1_c exists. In fact,

$$\int_{\mathcal{Y}} \int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x)\theta^*_c(\mathrm{d}y) = \frac{1}{c} R\left(\theta^*_c(\mathrm{d}y) \,\|\gamma(\mathrm{d}y)\right) + \Lambda^1_c$$

precisely at θ^* defined by

$$\frac{\mathrm{d}\theta^*_c(\cdot)}{\mathrm{d}\gamma(\cdot)}(y) = \mathrm{e}^{\int_{\mathcal{X}} cF(x,y)\mu(\mathrm{d}x)} \Big/ \int_A \mathrm{e}^{\int_{\mathcal{X}} cF(x,y)\mu(\mathrm{d}x)}\gamma(\mathrm{d}y)$$

(see [2], Prop. 1.4.2). By the non-negativity of relative entropy

$$\begin{aligned}
\Lambda^1_\infty &= \lim_{c\to\infty} \Lambda^1_c \\
&\le \limsup_{c\to\infty} \int_{\mathcal{Y}} \int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x)\theta^*_c(\mathrm{d}y) \\
&\le \sup_{y\in A} \int_{\mathcal{X}} F(x,y)\mu(\mathrm{d}x).
\end{aligned}$$

□

Remark 3.6. The set A need not be bounded, nor must the uniform distribution be used. Indeed, assume only that $\gamma(\mathrm{d}y)$ has a density $f(y)$ such that $f(y) \ge \alpha > 0$ for $y \in B_\varepsilon$. Then

$$\begin{aligned}
R\left(\theta(\mathrm{d}y) \,\|\gamma(\mathrm{d}y)\right) &\le \int_{B_\varepsilon} \log\left[(1/\varepsilon)/\alpha\right]\theta(\mathrm{d}y) \\
&= (1/\varepsilon)\log\left[1/(\varepsilon\alpha)\right],
\end{aligned}$$

and the result still holds.

4. Examples and computational methods

To indicate how the robust performance bounds might be used in practice, we present some numerical examples, which illustrate the relationship between the different risk sensitive integrals and the relevant techniques and computational issues. We first review the polynomial chaos techniques that are used to compute the risk-sensitive integrals. The reader familiar with this material can skip to the next section.

Remark 4.1. Although previously the random variables with known and unknown distributions were denoted by X and Y, in the rest of the paper these random variables will be denoted by Z_1 and Z_2. This is done to free up x and y to be spatial variables in various PDE and related equations that might be used to define the mapping F.

4.1. Review of polynomial chaos methods

This Section reviews the polynomial chaos method that we use to compute the risk-sensitive integrals. The reader familiar with this material can skip to Section 4.2, or refer to [5] for a more comprehensive treatment. It should also be noted that one need not use polynomial chaos methods for the calculation of the risk-sensitive integrals. Various Monte Carlo type techniques can certainly be applied. However, for the examples in this paper, we found that polynomial chaos methods worked better (higher accuracy with a smaller number of sample points) than straightforward Monte Carlo integration.

To calculate risk sensitive integrals, one could use Monte Carlo integration, which requires evaluating $F(Z_1, Z_2)$ for many replicas of (Z_1, Z_2). Because the order of convergence for Monte Carlo integration is

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

$\mathcal{O}(N^{-1/2})$, where N is the number of sample points, this method can take a long time to converge(though not necessarily for the given examples), and so we utilize generalized polynomial chaos (gPC) methods to approximate $F(z_1, z_2)$ and evaluate various statistical properties of $F(Z_1, Z_2)$ (see, for example [11]). We briefly recall the mathematical framework of polynomial chaos methods, along with a simple example in one dimension.

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space where Ω is the sample space, $\mathcal{A}$ the associated σ-algebra, and $\mathcal{P}$ the probability measure on $\mathcal{A}$. Define a random vector $\mathbf{Z}(\omega) \doteq (Z_1(\omega), \ldots, Z_N(\omega)) \in \mathbb{R}^N$ on this probability space. Consider a d-dimensional bounded domain $D \subset \mathbb{R}^d$ with boundary ∂D and let $t \in [0, T], T \in (0, \infty)$. We consider random fields $u(t, x; \mathbf{Z}(\omega)) : \bar{D} \times [0, T] \times \Omega \to \mathbb{R}$ that are defined by requiring that for a.e. $\omega \in \Omega$

$$\mathcal{L}(t, x, u; \mathbf{Z}(\omega)) = f(t, x; \mathbf{Z}(\omega)), \qquad (x, t, \omega) \in D \times [0, T] \times \Omega,$$

subject to the boundary and initial conditions

$$\begin{aligned} \mathcal{B}(t, x, u; \mathbf{Z}(\omega)) &= g(t, x; \mathbf{Z}(\omega)), \qquad (x, t, \omega) \in \partial D \times [0, T] \times \Omega, \\ u(t, x; \mathbf{Z}(\omega)) &= u_0(x; \mathbf{Z}(\omega)), \qquad (x, t, \omega) \in D \times \{0\} \times \Omega. \end{aligned}$$

Here $x = (x_1, ..., x_d) \in \mathbb{R}^d$, $\mathcal{L}$ is a linear or nonlinear differential operator, and $\mathcal{B}$ is a boundary operator. We assume for simplicity that a unique classical sense solution exists to the differential equation and/or boundary conditions. Note that for the first two examples of the last Section the problem involves only time dependence, and hence there is neither a spatial variable nor a boundary condition.

We assume that $\{Z_i\}_{i=1}^N$ are independent random variables, with support $\{\Gamma_i\}_{i=1}^N$ and with probability density functions $\{\rho_i(z_i)\}_{i=1}^N$, respectively. Hence the joint density is $\rho(\mathbf{z}) = \prod_{i=1}^N \rho_i(z_i)$. Let Ω be the canonical space $\prod_{i=1}^N \Gamma_i$, in which case $\mathbf{Z}(\omega) = \omega$ and we identify ω with $\mathbf{z}$. Thus the equations above become

$$\mathcal{L}(t, x, u; \mathbf{z}) = f(t, x; \mathbf{z}), \qquad (x, t, \mathbf{z}) \in D \times [0, T] \times \Omega, \tag{4.1}$$

subject to the boundary and initial conditions

$$\begin{aligned} \mathcal{B}(t, x, u; \mathbf{z}) &= g(t, x; \mathbf{z}), \qquad (x, t, \mathbf{z}) \in \partial D \times [0, T] \times \Omega, \\ u(t, x; \mathbf{z}) &= u_0(x; \mathbf{z}), \qquad (x, t, \mathbf{z}) \in D \times \{0\} \times \Omega. \end{aligned} \tag{4.2}$$

We consider approximating the mapping as a function of the stochastic variable, *i.e.*, $\mathbf{z} \to u(t, x; \mathbf{z})$ at some particular (t, x), *via* a finite sum of orthogonal basis functions.

We first define finite dimensional subspaces for $L^2(\Gamma_i)$ according to

$$W_i^{P_i} = \left\{ v : \Gamma_i \to \mathbb{R} : v \in \operatorname{span} \{\phi_{i,m}(z_i)\}_{m=0}^{P_i} \right\}, \qquad i = 1, ..., N.$$

Here P_i is the highest degree of the polynomial basis function, and $\{\phi_{i,m}(z_i)\}$ are a set of orthonormal polynomials with respect to the weight ρ_i, *i.e.*, for $m \neq n$

$$\int_{\Gamma_i} \phi_{i,m}(z_i)\phi_{i,n}(z_i)\rho_i(z_i)\mathrm{d}z_i = 0$$

and

$$\int_{\Gamma_i} \phi_{i,n}^2(z_i)\rho_i(z_i)\mathrm{d}z_i = 1.$$

The orthonormal basis representation is determined by the probability density function ρ_i. For example, if the density is uniform or Gaussian, then Legendre or Hermite orthogonal polynomials, respectively, are used. A table of polynomial basis functions and their respective distributions is listed at the end of this section (also see [8]). A finite dimensional subspace for $L^2(\Gamma)$, where $\Gamma \doteq \Gamma_1 \times \cdots \times \Gamma_N = \Omega$, can either be defined as

$$W_N^P = \bigotimes_{|\mathbf{P}| \leq P} W_i^{P_i}$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

where the tensor product is over all combinations of the multi-index $\mathbf{P} = (P_1, \ldots, P_N) \in \mathbb{N}_0^N$ with $|\mathbf{P}| = \sum_{i=1}^N P_i \leq P$, or

$$\tilde{W}_N^P = \bigotimes_{i=1}^N W_i^P.$$

Thus, W_N^P is the space of N-variate orthogonal polynomials of total degree at most P, whereas $\tilde{W}_N^P$ is the full tensor product of the one-dimensional polynomial spaces with each highest degree P. Note that $\dim(W_N^P) = \binom{N+P}{P}$ and $\dim(\tilde{W}_N^P) = (P+1)^N$, and that for large N, $\binom{N+P}{P} \ll (P+1)^N$. Since our examples only consider $N = 2$, we will use the full tensor product space, $\tilde{W}_N^P$. Let $\Phi_j(\mathbf{z}), j = 1, \ldots, M$ denote the elements of $\tilde{W}_N^P$, where $M = \dim(\tilde{W}_N^P)$.

There are two standard methods for constructing gPC approximations, referred to as the stochastic Galerkin method (see, *e.g.*, [8]) and the stochastic collocation method (see, *e.g.*, [9]). Since the calculations presented below use only the collocation method, we restrict discussion to this method. With stochastic collocation, we first consider an approximation to the solution $u(t, x; \mathbf{z})$ in terms of a Lagrange interpolant of the form

$$v(t, x; \mathbf{z}) = \sum_{k=1}^{N_p} v(t, x; \mathbf{z}_k) \mathbf{L}_k(\mathbf{z}), \tag{4.3}$$

where $\mathbf{L}_k(\mathbf{z})$ is a Lagrange polynomial of degree $\dim(W_N^P)$, satisfies $\mathbf{L}_k(\mathbf{z}_l) = \delta_{kl}$, and $\{\mathbf{z}_k\}_{k=1}^{N_p}$ are a set of prescribed nodes in the N-dimensional space Γ. We require that the residuals

$$R(t, x; \mathbf{z}) \doteq \mathcal{L}(t, x, v; \mathbf{z}) - f(t, x; \mathbf{z})$$

$$R^{BC}(t, x; \mathbf{z}) \doteq \mathcal{B}(t, x, v; \mathbf{z}) - g(t, x; \mathbf{z})$$

$$R^{IC}(x; \mathbf{z}) \doteq v(0, x; \mathbf{z}) - u_0(x; \mathbf{z})$$

vanish at the collocation points $\{\mathbf{z}_k\}_{k=1}^{N_p}$, *i.e.*,

$$\mathcal{L}(t, x, v(t, x; \mathbf{z}_k); \mathbf{z}_k) - f(t, x; \mathbf{z}_k) = 0, \text{ for } k = 1, \ldots, N_p$$

$$\mathcal{B}(t, x, v(t, x; \mathbf{z}_k); \mathbf{z}_k) - g(t, x; \mathbf{z}_k) = 0, \text{ for } k = 1, \ldots, N_p$$

$$v(0, x; \mathbf{z}_k) - u_0(x; \mathbf{z}_k) = 0, \text{ for } k = 1, \ldots, N_p.$$

Thus, the stochastic collocation method produces a set of N_p decoupled equations, where each value of $v(t, x; \mathbf{z}_k)$ coincides with the exact solution $u(t, x; \mathbf{z})$ for the given $\mathbf{z}_k$, since both satisfy the same equations.

Once the values of $\{v(t, x; \mathbf{z}_k)\}_{k=1}^{N_p}$ have been determined at the collocation points, one can construct a gPC approximation in the orthogonal basis representation in the form

$$v(t, x; \mathbf{z}) = \sum_{j=1}^{M} \hat{v}_j(t, x) \Phi_j(\mathbf{z}), \tag{4.4}$$

which, under certain conditions, will be equivalent to the Lagrange basis formulation. The coefficients $\{\hat{v}_j\}_{j=1}^M$ can be determined by inverting a Vandermonde matrix, where invertibility is dependent on the choice of collocation points. If one chooses quadrature or cubature[4] points for the collocation points, then invertibility

[4]cubature points just refers to quadratures in more than one dimension.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

is guaranteed. Furthermore, if the cubature rule is exact up to polynomials of degree $2M$, inversion of the Vandermonde matrix is not necessary[5], and the coefficients $\{\hat{v}_j\}_{j=1}^{M}$ can be computed by evaluating

$$\hat{v}_j(t,x) = \sum_{k=1}^{N_p} v(t,x;\mathbf{z}_k)\Phi_j(\mathbf{z}_k)w_k, \tag{4.5}$$

where $\{w_k\}_{k=1}^{N_p}$ are the respective weights according to the choice of quadrature or cubature points.

Statistical information is easy to obtain from the form (4.4). For example,

$$\mathbb{E}_{\mathbf{z}}[v(t,x;\mathbf{z})] = \int_{\Gamma} \left(\sum_{j=1}^{M} \hat{v}_j(t,x)\Phi_j(\mathbf{z}) \right) \rho(\mathbf{z})\mathrm{d}\mathbf{z} = \hat{v}_1(t,x)$$

and

$$\mathbb{E}_{\mathbf{z}}[(v(t,x;\mathbf{z}))^2] = \int_{\Gamma} \left(\sum_{j=1}^{M} \hat{v}_j(t,x)\Phi_j(\mathbf{z}) \right)^2 \rho(\mathbf{z})\mathrm{d}\mathbf{z} = \sum_{j=1}^{M} \hat{v}_j(t,x)^2$$

as a consequence of the orthonormality. Other statistical information, such as higher moments and sensitivity coefficients, are also easily calculated (see [8]). The ease with which one can calculate such moments explains why orthogonal representation (4.4) is preferred over the Lagrange form (4.3).

To illustrate the approach, we look at a simplified version of Example 4.2. Let $g(z_2) = u_0$, so that the only uncertainty is in the decay rate k, which depends on the random variable Z. The simplified problem becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -k(z)u(t), \qquad u(0) = u_0,$$

where $u(t;z)$ denotes the solution parameterized by z, with z a point in the range of Z. Requiring that the residuals vanish at the collocation points $\{z_n\}_{n=1}^{N_p}$ gives

$$R(t;z_n) = \frac{\mathrm{d}}{\mathrm{d}t}v(t;z_n) + k(z_n)v(t;z_n) = 0$$

and

$$R^{IC}(z_n) = v(0;z_n) - u_0 = 0.$$

Note that there are indeed N_p decoupled equations, which are solved independently for each n. Here the solutions are exact:

$$v(t;z_n) = u_0 \mathrm{e}^{-k(z_n)t}, \ \ for = 1,\ldots,N_p.$$

One can then obtain the approximation (4.4) with respect to the orthogonal basis, assuming that the appropriate collocation points and weights have been chosen.

We illustrate this example with $k(z) = z$ and $Z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$. Here the $\{\Phi_j\}_{j=1}^{M}$ are the Hermite polynomials, which are orthonormal with respect to the weighting function $\rho(z) = \mathrm{e}^{-(z-\mu)^2/2\sigma^2}/\sigma\sqrt{2\pi}$ with support $\Gamma = (-\infty, +\infty)$. Figure 2 shows the stochastic collocation approximation $v(1,z)$ at $t = 1$ as a function of z, with $N_p = 2,3,4,5$, versus the exact solution $u_{\text{exact}}(1,z) = \mathrm{e}^{-z}$. Note that the collocation approximation is an interpolation of the exact solution at N_p points.

Figure 3 shows the relative error between the exact and approximated mean and variance in a semilog plot. The linear decrease in relative error on the semilog plot implies an exponential decay in the error. Errors become a constant at the limits of the accuracy of the numerical scheme.

[5]This is especially useful if the Vandermonde matrix is ill-conditioned.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

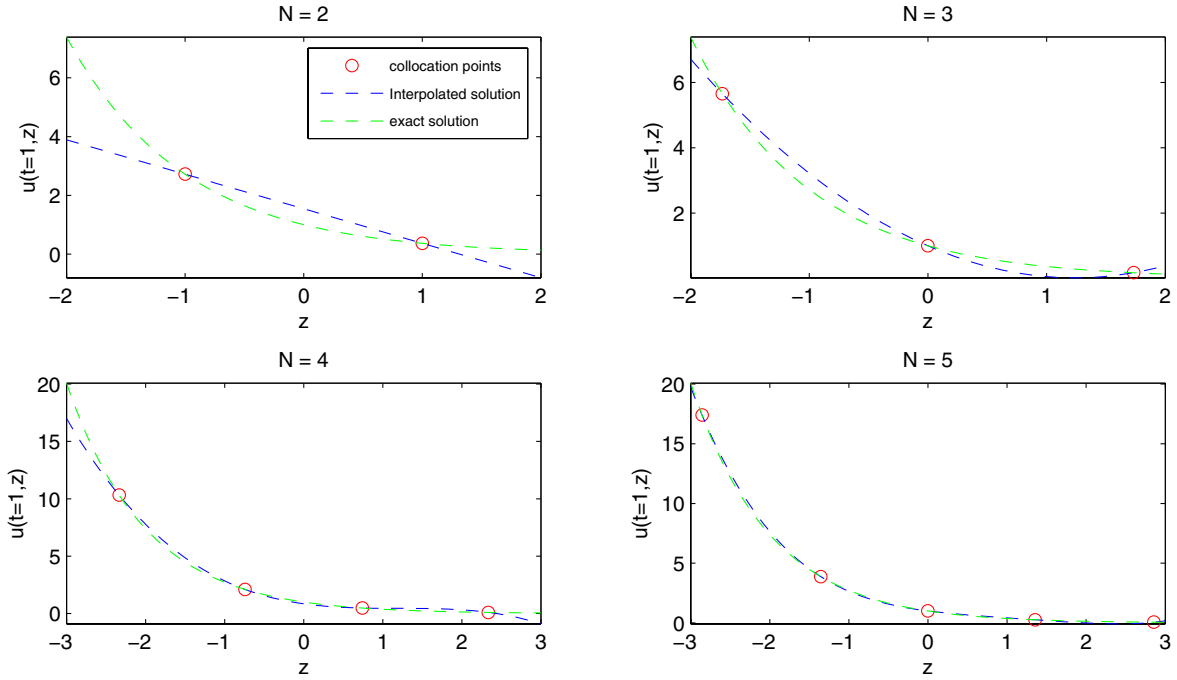


FIGURE 2. Interpolated solutions versus the exact solution at $t = 1$.

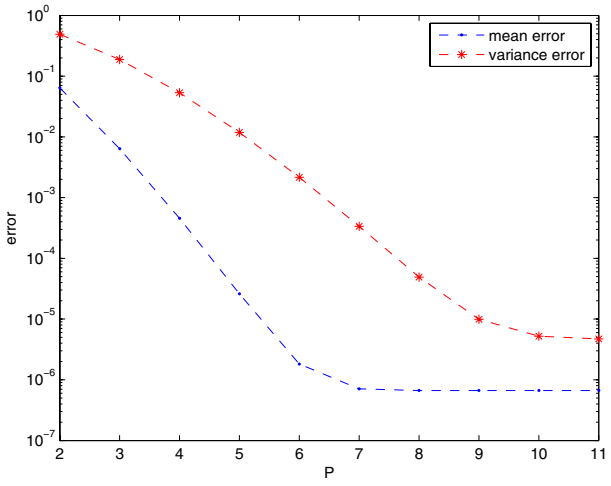


FIGURE 3. Relative errors.

4.2. Numerical examples and interpretation

We now discuss numerical calculation of the risk-sensitive integrals. For convenience, we recall the three types of risk-sensitive integrals introduced in Section 3.1, the ordinary risk sensitive cost Λ_c and the two hybrid

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

integrals Λ_c^1 and Λ_c^2:

$$\Lambda_c = \frac{1}{c}\log\int_{\mathcal{Z}_1}\int_{\mathcal{Z}_2} \mathrm{e}^{cF(z_1,z_2)}\gamma(\mathrm{d}z_2)\mu(\mathrm{d}z_1), \tag{4.6}$$
$$\Lambda_c^1 = \frac{1}{c}\log\int_{\mathcal{Z}_2} \mathrm{e}^{\int_{\mathcal{Z}_1} cF(z_1,z_2)\mu(\mathrm{d}z_1)}\gamma(\mathrm{d}z_2),$$
$$\Lambda_c^2 = \frac{1}{c}\int_{\mathcal{Z}_1}\left[\log\int_{\mathcal{Z}_2} \mathrm{e}^{cF(z_1,z_2)}\gamma(\mathrm{d}z_2)\right]\mu(\mathrm{d}z_1),$$

where $\mathcal{Z}_1, \mathcal{Z}_2$ are the range spaces for the random variables Z_1, Z_2, respectively. Recall the relationships $\Lambda_c^1 \leq \Lambda_c^2 \leq \Lambda_c$. It is assumed that the distribution of Z_1 is known, but this is not true for Z_2. Instead, we assume there is a nominal distribution, or best guess, for Z_2, i.e, $\gamma(\mathrm{d}z_2)$.

In the stochastic collocation approach, we replace $F(z_1, z_2)$ in (4.6) with a polynomial interpolant, $\tilde{F}(z_1, z_2)$, determined by the full tensor product of the one dimensional collocation points, $\{z_{1,i}\}_{i=1}^{N_{1,q}} \otimes \{z_{2,i}\}_{i=1}^{N_{2,q}}$ in $\mathcal{Z}_1 \times \mathcal{Z}_2$. Thus

$$\tilde{F}(z_1, z_2) = \sum_{j=1}^{M} \hat{F}_j(t, x)\Phi_j(z_1, z_2),$$

where $M = N_{1,q} \cdot N_{2,q}$, Φ_j is the full tensor product basis, and $\hat{F}_j(t, x)$ is computed using (4.5). From here, one can efficiently compute the risk sensitive integrals (4.6) *via* either Monte Carlo sampling or quadrature. For both methods, once the samples or quadrature points have been chosen, one can calculate the risk sensitive integrals for different values of c without resampling or choosing different quadrature points. Hence, it is computationally inexpensive to compute the risk sensitive integral for different values of c. Note that for higher dimensions, a sparse grid is preferred and, in most cases, necessary for these types of computations. It is typical to use the Smolyak algorithm to generate these sparse grids, which are based on a one dimensional quadrature rule (for example, see Sect. 4.1.2 in [7] for a more detailed explanation and further references).

In this paper, we opt for the numerical quadrature approach. Since $\tilde{F}(z_1, z_2)$ can be evaluated very quickly, we can compute the risk sensitive integrals with high accuracy and efficiency using a high order quadrature rule. The number of quadrature points chosen to compute the risk sensitive integrals need not be the same as the number of collocation points used in computing the coefficients of the gPC expansion, though the type of quadrature points is the same. For example, if we were to use Legendre-Gauss quadrature points to evaluate the gPC coefficients, we would again use Legendre-Gauss quadrature points to calculate the risk sensitive integrals. In the examples presented below a much larger number of quadrature points are used to compute the risk sensitive integrals. To avoid confusion, to the set of quadrature points used to evaluate the risk sensitive integrals is denoted by $\{\tilde{z}_{1,i}\}_{i=1}^{\tilde{N}_{1,q}} \otimes \{\tilde{z}_{2,i}\}_{i=1}^{\tilde{N}_{2,q}}$, *i.e.*, the full tensor product of one dimensional quadrature points. In the examples, $\tilde{N}_{1,q}$ and $\tilde{N}_{2,q}$ equal 2^8, while $N_{1,q}$ and $N_{2,q}$ are 8 for smooth F and 12 when F is an indicator functions.

We also implemented the Monte Carlo approach, but observed that numerical quadrature gave a much better approximation for the same computational effort, at least for our examples. It should also be noted that a variation on traditional Monte Carlo is needed for the hybrid forms of the risk sensitive integrals. For the original risk-sensitive cost function (3.5), the standard approach is to use many independent samples of the pair (Z_1, Z_2). This produces an estimate which has a mean square error of $\mathcal{O}(N^{-1/2})$, where N is the total number of Monte Carlo samples. However, for the first hybrid form (3.7), one must approximate the inner integral $\int_{\mathcal{X}} cF(z_1, z_2)\mu(\mathrm{d}z_1)$ for every fixed sample of Z_2. Thus one has to simulate more Z_1 samples than Z_2 samples. Similarly, to compute the second form of the hybrid (3.9), one must approximate the inner integral $\int_{\mathcal{Y}} \mathrm{e}^{cF(z_1,z_2)}\gamma(\mathrm{d}z_2)$ for every fixed Z_1. In particular, for both hybrid forms one does not simply choose independent samples of (Z_1, Z_2).

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

4.2.1. Independent aleatoric and epistemic uncertainties

When the aleatoric and nominal epistemic variables are independent, the numerical quadrature approach would evaluate the risk sensitive integrals via the formulas

$$\begin{aligned}
\Lambda_c &\approx \frac{1}{c}\log\left(\sum_{i=1}^{\tilde{N}_{1,q}}\sum_{j=1}^{\tilde{N}_{2,q}} \mathrm{e}^{cF(\tilde{z}_{1,i},\,\tilde{z}_{2,j})}\tilde{w}_{1,i}\tilde{w}_{2,j}\right) \\
\Lambda_c^1 &\approx \frac{1}{c}\log\left(\sum_{j=1}^{\tilde{N}_{2,q}} \mathrm{e}^{\left(\sum_{i=1}^{\tilde{N}_{1,q}} cF(\tilde{z}_{1,i},\,\tilde{z}_{2,j})\tilde{w}_{1,i}\right)}\tilde{w}_{2,j}\right) \\
\Lambda_c^2 &\approx \frac{1}{c}\sum_{i=1}^{\tilde{N}_{1,q}}\left(\log\sum_{j=1}^{\tilde{N}_{2,q}} \mathrm{e}^{cF(\tilde{z}_{1,i},\tilde{z}_{2,j})}\tilde{w}_{2,j}\right)\tilde{w}_{1,i}.
\end{aligned} \tag{4.7}$$

where $\{\tilde{z}_{1,i},\tilde{w}_{1,i}\}_{i=1}^{\tilde{N}_{1,q}}$, $\{\tilde{z}_{2,i},\tilde{w}_{2,i}\}_{i=1}^{\tilde{N}_{2,q}}$ are the pairs of quadrature points and weights chosen based on the underlying distribution.[6]

The first example is taken from [10] and involves a simple ODE. The example is convenient because analytic solutions to the risk-sensitive integrals can be obtained, and then compared with the numerical approximations described in the next section. The second example is a random oscillator, where the stochastic parameters are the spring and dampening coefficients. The third example involves a heat equation with random heat capacity and thermal conductivity.

Example 4.2. We consider $F(Z_1, Z_2)$ defined in terms of the solution of the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -k(z_1)u(t), \quad u(0) = g(z_2),$$

where k and g are known functions. Letting $u(t; z_1, z_2)$ denote the solution parameterized by (z_1, z_2), set $F(Z_1, Z_2) = h(u(t; Z_1, Z_2))$, where typical choices of h are $h(u) = u, h(u) = u^2, h(u) = \mathbf{1}\{a \le u \le b\}$.

Here we take $k(z_1) = z_1, g(z_2) = z_2$, let $Z_1 \sim U[0,1]$, $Z_2 \sim U[0,1]$, and compute risk sensitive integrals for $F(z_1, z_2) = [u(1; z_1, z_2)]^2$, and $F(z_1, z_2) = \mathbf{1}\{0.8 \le u(1; z_1, z_2) \le 1\}$. Hence the goal is to obtain robust bounds on the second moment and the probability that the solution to the stochastic ODE falls within the interval $[0.8, 1]$ at time $t = 1$, respectively. Note that the indicator function is not smooth, and thus more collocation points are required to obtain an accurate approximation. We choose a Legendre polynomial basis for both Z_1 and Z_2 and use the Legendre–Gauss quadrature points and weights for each dimension.

Figures 4 and 5 show the approximations as a function of c for $F(u) = u^2$ and $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$, respectively. As expected Λ_c^1 gives the best bounds, while the original form Λ_c gives the worst upper bounds of the three.

The plots also suggest what happens as $c \to \infty$. The measure Λ_c is expected to yield robust bounds if one is uncertain regarding the distributions of both random variables Z_1 and Z_2, and $\lim_{c\to\infty}\Lambda_c$ represents the tightest upper bound on performance if we know nothing about these random variables except for their support. In this case, the limit in Figure 5 appears to be 1, which means that for some choice of z_1 and z_2, $u(1; z_1, z_2) \in [0.8, 1]$. Hence without more information on the distributions of Z_1, Z_2 we do not obtain information other than this from this performance bound. On the other hand, $\lim_{c\to\infty}\Lambda_c^i$ for $i = 1, 2$ is strictly less than 1 and thus by Theorem 3.5 gives a meaningful performance bound when the only information available about Z_2 is its support. Note also that while Λ_c^1 and Λ_c^2 seem close for $F(u) = u^2$, the differ considerably for $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$.

[6] In some of the examples below, the function of interest, F, is more easy to evaluate than surrogate $\tilde{F}$. In these cases, the point is merely to illustrate the use and accuracy of gPC surrogate models.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

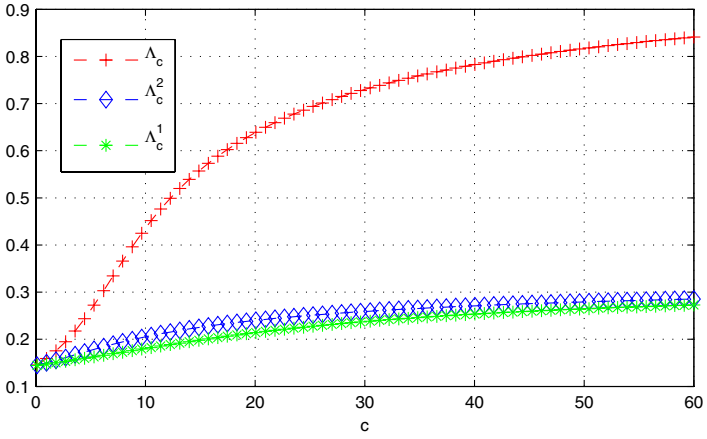


FIGURE 4. Example 4.2, Λ_c, $\{\Lambda_c^i\}_{i=1}^2$ for $F(u) = u^2$.

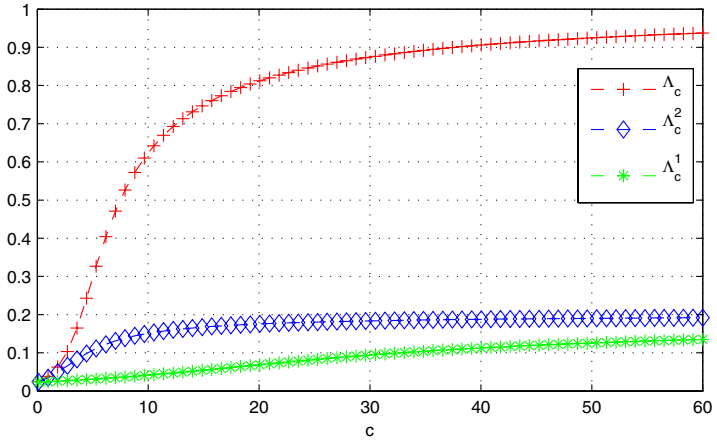


FIGURE 5. Example 4.2, Λ_c, $\{\Lambda_c^i\}_{i=1}^2$ for $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$.

In Figures 6 and 7 we display the relative error $|\Lambda_c^i - \Lambda_{c,\text{exact}}^i| / |\Lambda_{c,\text{exact}}^i|$ between the three risk sensitive integrals and the exact solution at $c = 30$ and $c = 60$ as a function of the number of quadrature points, N^q. The exact solutions were calculated by reducing the risk-sensitive integrals to one-dimensional integrals, and then using an adaptive quadrature method. We note that the rate of convergence is significantly slower for $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$ than $F(u) = u^2$. This is not surprising since the former is discontinuous. Also, the calculation of risk-sensitive integrals for $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$ is not as sensitive to large c values when compared to $F(u) = u^2$. In the former case, because $0 \le F(u) \le 1$, large values of c do not significantly affect the growth of the exponents in the risk-sensitive integrals which could cause large round-off errors. However, with $F(u) = u^2$, this is not the case, and Figure 7 shows that there is a reduction in accuracy between $c = 30$ and $c = 60$ for the same number of quadrature points. Finally, Figure 7 also reveals which risk-sensitive integrals are are the slowest to converge, for a fixed c. In order from slowest to fastest convergence, we have Λ_c, Λ_c^2, and Λ_c^1. The reason that Λ_c^1 converges more quickly than Λ_c or Λ_c^2 is due to the summation of the exponential terms

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

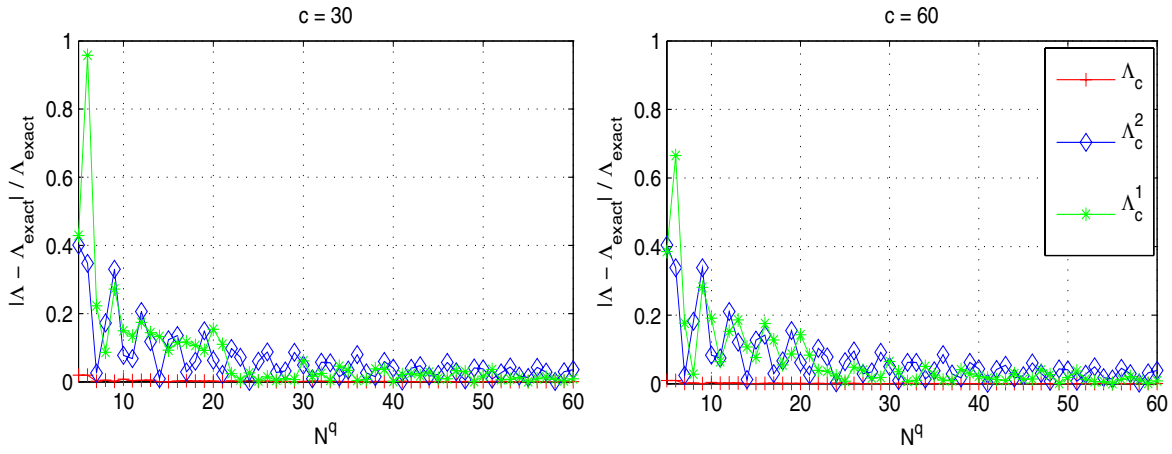


FIGURE 6. Example 4.2, relative error for Λ_c, $\{\Lambda_c^i\}_{i=1}^2$ with $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$ for $c = 30$ (left) *vs.* $c = 60$ (right) for different values of $N^q = 5, \ldots, 60$.

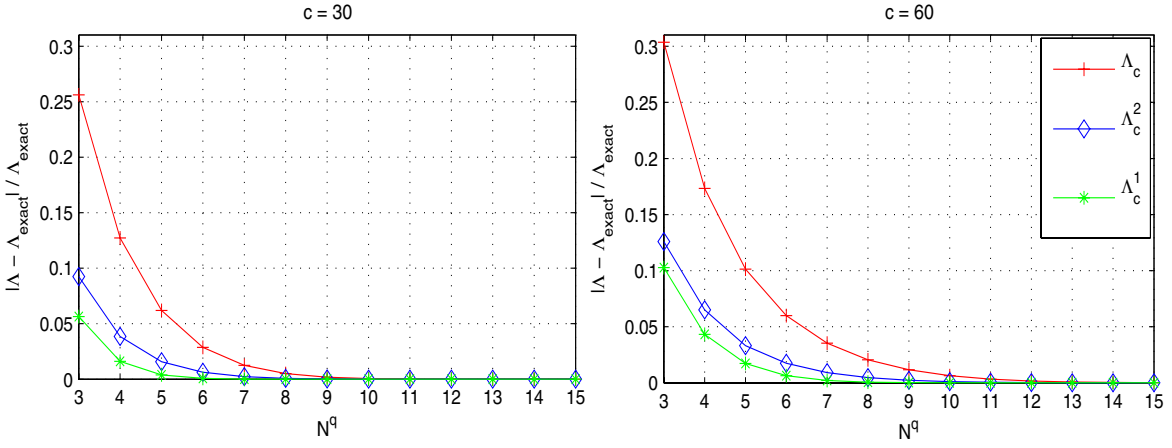


FIGURE 7. Example 4.2, relative error for Λ_c, $\{\Lambda_c^i\}_{i=1}^2$ with $F(u) = u^2$ for $c = 30$ (left) *vs.* $c = 60$ (right) for different values of $N^q = 3, \ldots, 15$.

in (4.7) for large values of c. The calculation of Λ_c^1 involves half as many computations of exponential terms (since the summation is in the exponent), which is the main source of the loss in accuracy for large values of c.

Next we consider the problem of computing the value of c which minimizes $c \mapsto \frac{1}{c}B + \Lambda_c^i$ for $i = 0, 1, 2$.[7] Here B is a maximum relative entropy distance that will be allowed between the "true" distribution, $\theta(\mathrm{d}z_2)$, and the nominal distribution, $\gamma(\mathrm{d}z_2)$, and the minimization produces the tightest possible bounds. For the example, suppose that the family of distributions for which bounds are to be valid includes $\theta(\mathrm{d}z_2) \sim \text{beta}(\alpha = 1.5, \beta = 1.5)$, which implies $B \approx 0.0484$. In this example $B \le f(c) \doteq c\frac{\mathrm{d}}{\mathrm{d}c}H^i(c) - H(c)$ for some c, where $H^i(c) \doteq c\Lambda_c^i$. As shown in Proposition 3.3, this implies there is a unique local minimum for $c \in (0, \infty)$. See Figures 8 and 9. Alternatively, if $B > f(c)$ for all c, then Proposition 3.3 implies the minimum is achieved in the limit as $c \to \infty$.

[7]Note that the risk sensitive integrals are infinite for $c = 0$. Thus in the numerical calculations, we actually compute c starting from 0.01.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

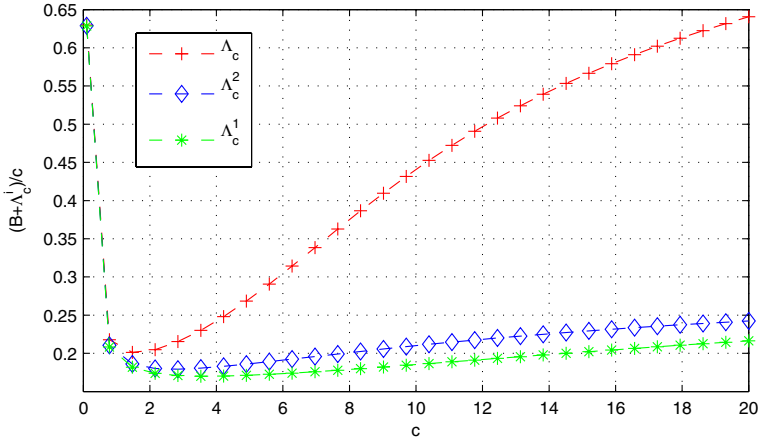


FIGURE 8. Example 4.2, $\frac{1}{c}(B + \Lambda_c), \{\frac{1}{c}(B + \Lambda_c^i)\}_{i=1}^2$ with $F(u) = u^2$.

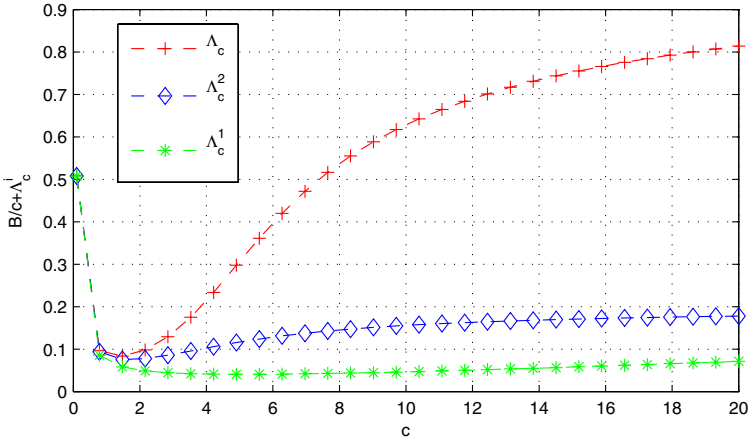


FIGURE 9. Example 4.2, $\frac{1}{c}(B + \Lambda_c), \{\frac{1}{c}(B + \Lambda_c^i)\}_{i=1}^2$ with $F(u) = \mathbf{1}\{0.8 \le u \le 1\}$.

From Figure 9 we find a minimum for $\frac{1}{c}B + \Lambda_c^1$ of approximately 0.04 at $c \approx 5.12$. Thus for all distributions on Z_2 whose relative entropy distance to $U[0,1]$ is less than 0.0484, the probability that the solution to the random ODE falls between 0.8 and 1 at time 1 is less than 0.04.

Note that the minimum is at different values of c for the three different integrals, but the minimum is always the smallest for the first form of the hybrid. In the case when the minimum occurs at some $c < \infty$, the minimum is easily calculated, since this is a one-dimensional unconstrained minimization problem (for example, one can use a golden method search algorithm, such as MATLAB's fminbnd function). In the case when the minimum occurs in the limit as $c \to \infty$, in order to find the minimum, one can iterate Λ_c^i until the successive iterations differ by less than a prescribed tolerance.

The second example is a random oscillator, where the stochastic parameters are the spring and dampening coefficients.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

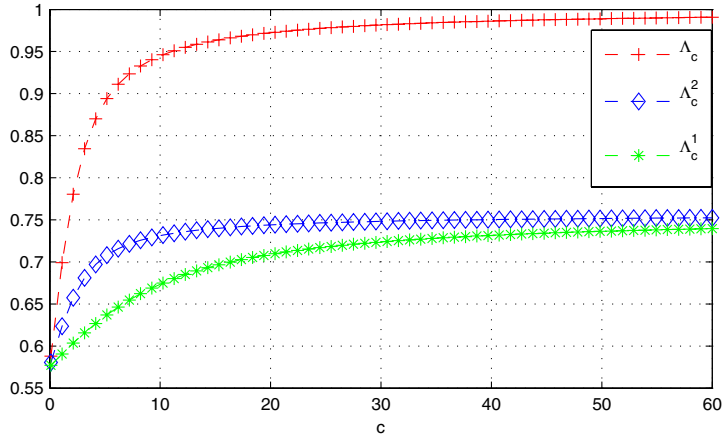


FIGURE 10. Example 4.3, Λ_c, $\{\Lambda_c^i\}_{i=1}^2$ for $F(u) = \mathbf{1}\{u \le 2\}$.

Example 4.3. We consider $F(Z_1, Z_2)$ defined in terms of the solution of the ODE

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}u(t) + \gamma(z_2)\frac{\mathrm{d}}{\mathrm{d}t}u(t) + k(z_1)u(t) = f(t), \quad u(0) = u_0, \quad u'(0) = u_0',$$

where k and γ are known functions, which represent the spring and dampening coefficients, respectively. The outcome of interest is whether or not the position of the random oscillator falls within a specified range, $[a, b]$, at a specified time, t_{critical}. Letting $u(t; z_1, z_2)$ denote the solution parameterized by (z_1, z_2), set $F(Z_1, Z_2) = \mathbf{1}\{a \le u(t_{\text{critical}}; Z_1, Z_2) \le b\}$.

Here we consider the random oscillator with $k(z_1) = 4z_1, \gamma(z_2) = z_2/5$, and $f(t) = 10\cos(10t) + 3$, and choose $Z_1 \sim \text{beta}(\alpha = 5, \beta = 5)$, $Z_2 \sim \text{beta}(\alpha = 5, \beta = 5)$, $t_{\text{critical}} = 4$, and $[a, b] = (-\infty, 2]$. Here we use the Gauss–Jacobi quadrature points and weights. In this example we are concerned with the position of the oscillator at a critical time. Figure 10 plots the risk sensitive integrals for Example 4.3 with $F(u) = \mathbf{1}\{u \le 2\}$. Figure 11 depicts $c \mapsto \frac{1}{c}(B + \Lambda_c^i)$, where $B = R\left(\theta(\mathrm{d}z_2) \,\|\, \gamma(\mathrm{d}z_2)\right) \approx .0587$, with $\theta(\mathrm{d}z_2) \sim \text{beta}(\alpha = 10, \beta = 10)$.

Finally, the third example involves a heat equation with random heat capacity and thermal conductivity.

Example 4.4. We consider $F(Z_1, Z_2)$ defined in terms of the solution of the PDE

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t, x) = \frac{k(u; z_1)}{m(z_2)}\frac{d^2}{\mathrm{d}x^2}u(t, x), \qquad u(0, x) = u_0(x)$$

on $x \in (0, L)$ with boundary conditions

$$-k(u; z_1)\frac{\mathrm{d}}{\mathrm{d}x}u(t, 0) = q, \quad \frac{\mathrm{d}}{\mathrm{d}x}u(t, L) = 0.$$

Here k and m are the thermal conductivity and heat capacity, respectively. Note that in this example, the randomness affects the diffusivity and the boundary conditions. We are interested in the probability that the material exceeds a particular temperature, T_{critical}, at the time t_{final} and at some point $x^\star \in [0, L]$. Letting $u(t; z_1, z_2)$ denote the solution parameterized by (z_1, z_2), set $F(Z_1, Z_2) = \mathbf{1}\{u(t_{\text{final}}, x^\star; Z_1, Z_2) \ge T_{\text{critical}}\}$.

Here we consider the random heat equation with (nonlinear) Neumann boundary conditions. We use $k(u, z_1) = z_1 + (1.5 \times 10^{-7})u$, $m(z_2) = (10^{-6})z_2$, $q = 0.35$, $L = 1.90$, $u_0(x) = 25$, and $T_{\text{critical}} = 980$, $x^\star = 0$, $t_{\text{final}} = 1000$,

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

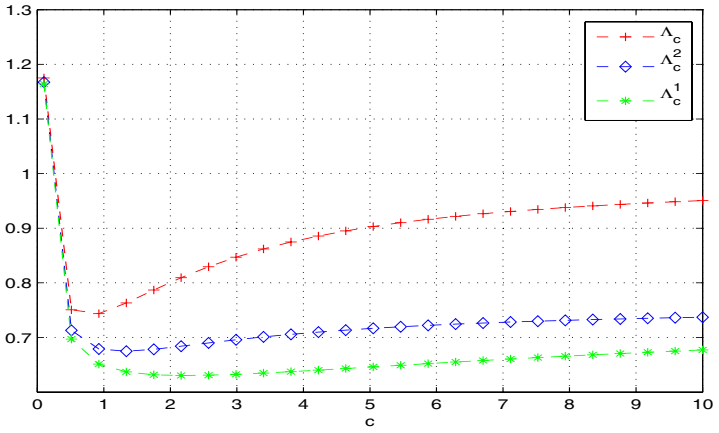


FIGURE 11. Example 4.3, $\frac{1}{c}(B + \Lambda_c), \left\{\frac{1}{c}(B + \Lambda_c^i)\right\}_{i=1}^2$ with $F(u) = \mathbf{1}\{u \le 2\}$.

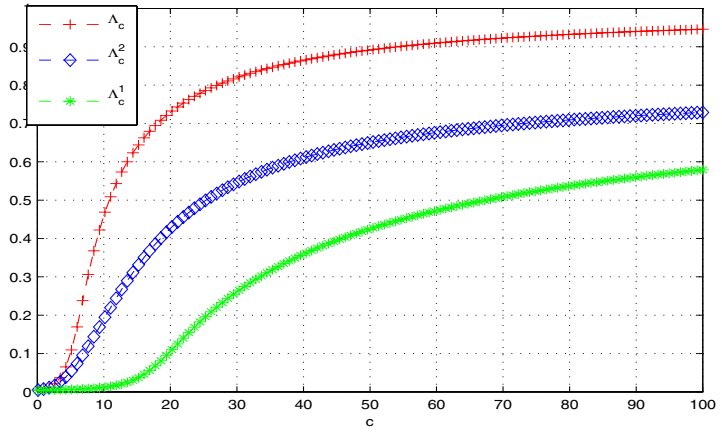


FIGURE 12. Example 4.4, $\frac{1}{c}(B + \Lambda_c), \left\{\frac{1}{c}(B + \Lambda_c^i)\right\}_{i=1}^2$ for $F(u) = \mathbf{1}\{u \ge T_{\text{critical}}\}$.

and set $F(Z_1, Z_2) = \mathbf{1}\{u(t_{\text{final}}, x^\star; Z_1, Z_2) \ge T_{\text{critical}}\}$. For the random variables, we introduce two independent random variables $\tilde{Z}_1 \sim \text{beta}(\alpha = 5, \beta = 5)$, $\tilde{Z}_2 \sim \text{beta}(\alpha = 5, \beta = 5)$ and let $Z_1 = (2 \times 10^{-3})\tilde{Z}_1 + 3 \times 10^{-3}$, $Z_2 = (0.11)\tilde{Z}_2 + 0.30$. Figures 12 and 13 plot the risk sensitive integrals as a function of c and illustrate $c \to \frac{1}{c}(B + \Lambda_c^i)$, where $B = R\left(\theta(\mathrm{d}z_2) \,\|\gamma(\mathrm{d}z_2)\right) \approx 0.0587$. and with $\theta(\mathrm{d}z_2) \sim \text{beta}(\alpha = 10, \beta = 10)$. Again, we use the Gauss–Jacobi quadrature points and weights.

4.2.2. Dependent aleatoric and epistemic uncertainties

In this section we consider problems where the distribution of Z_1 can depend on the value of Z_2, or vice versa. In the case when Z_1 depends on Z_2, we use the following inequality, which holds for the alternative of the first hybrid form (see the discussion in Sect. 3.1):

$$\int_{\mathcal{Z}_2} \int_{\mathcal{Z}_1} F(z_1, z_2)\mu(\mathrm{d}z_1|z_2)\theta(\mathrm{d}z_2) \le \frac{1}{c} R\left(\theta(\mathrm{d}z_2) \,\|\gamma(\mathrm{d}z_2)\right) + \bar{\Lambda}_c^1,$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

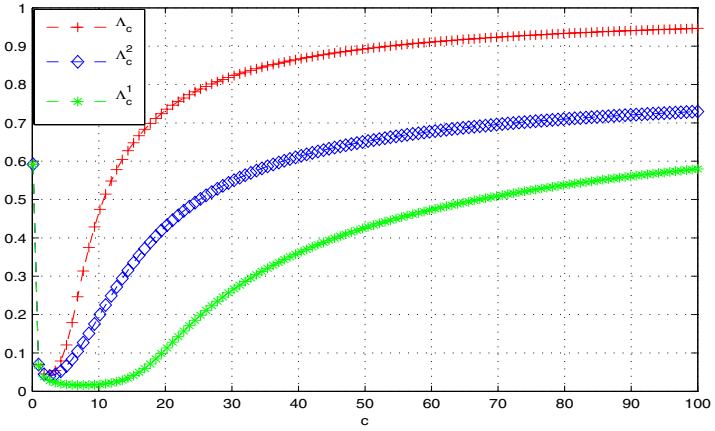


FIGURE 13. Example 4.4, $\frac{1}{c}(B+\Lambda_c)$, $\left\{\frac{1}{c}(B+\Lambda_c^i)\right\}_{i=1}^2$ with $F(u)=\mathbf{1}\{u\geq T_{\text{critical}}\}$.

where

$$\bar{\Lambda}_c^1=\frac{1}{c}\log\int_{\mathcal{Z}_2}\exp\left(\int_{\mathcal{Z}_1}cF(z_1,z_2)\mu(\mathrm{d}z_1|z_2)\right)\gamma(\mathrm{d}z_2).$$

Recall that the relative entropy term $\frac{1}{c}R\left(\theta(\mathrm{d}z_2)\,\|\gamma(\mathrm{d}z_2)\right)$ represents our lack of knowledge about the distribution of Z_2, *i.e.*, the epistemic uncertainty where we want robustness, whereas Z_1 represents aleatoric uncertainty. A question that now arises is whether or not it is still possible to use gPC methods to evaluate this integral. Consider the example with $Z_2\sim U[0,1]$ and $Z_1\sim N(Z_2,1)$, so that Z_1 is Gaussian with variance 1 and mean Z_2. Certainly Z_1 and Z_2 are not independent random variables. However, we can introduce an auxiliary normal random variable $Z\sim\mathcal{N}(0,1)$, and write $Z_1\doteq Z+Z_2$, where Z_2 and Z are independent random variables. Now we consider $G(Z,Z_2)$ instead of $F(Z_1,Z_2)$, and note that $G(Z,Z_2)=F(Z+Z_2,Z_2)$. Furthermore, $\bar{\Lambda}_c^1$ can be written as

$$\bar{\Lambda}_c^1=\frac{1}{c}\log\int_{\mathcal{Z}_2}\exp\left(\int_{\mathcal{Z}}cG(z,z_2)\mu(\mathrm{d}z)\right)\gamma(\mathrm{d}z_2).$$

Hence, $\bar{\Lambda}_c^1$ can be calculated just like Λ_c^1 via a gPC approximation. Similar results hold for the case when Z_2 depends on Z_1, but the epistemic uncertainty is in Z_2.

In general, as long as we can find a smooth, invertible mapping from (Z_1,Z_2) to an independent pair of random variables, one can evaluate the integrals as before. Figures 14 and 15 show the performance of risk sensitive integrals for dependent random variables as above for Example 4.2 and with $F(u)=\mathbf{1}\{1/2\leq u\leq 1\}$.

4.3. Discussion and Conclusion

There exist alternative methods for integrating aleatoric and epistemic variation in uncertainty quantification. The most straightforward techniques amount to treating the aleatoric and epistemic variables separately, and perform nested iterations, iterating over a finite collection of epistemic parameters on the outer loop, then sampling over aleatoric variables on the inner loop [4]. From this, one can plot the ensemble of cumulative distribution functions to visualize the family of distributions, sometimes called horsetail plots. For example, suppose the aleatoric random variable is Gaussian with an epistemic mean, where all we know about this epistemic variable is its support ($X\sim\mathcal{N}(Y,1)$, $Y\in[a,b]$). Then, for a set of N values contained within the known support (outer iteration), $\{Y_i=y_i:\ y_i\in[a,b]\}_{i=1}^N$, one can compute the performance (inner iteration) based on the corresponding set of input distributions, $\{X_i:X_i\sim\mathcal{N}(y_i,1)\}_{i=1}^N$.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

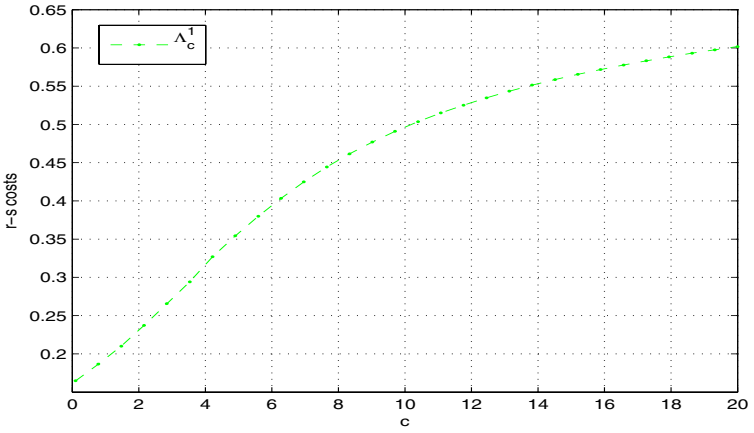


FIGURE 14. Example 4.2, $\bar{\Lambda}_c^1$ for $F(u) = \mathbf{1}\{1/2 \leq u \leq 1\}$.

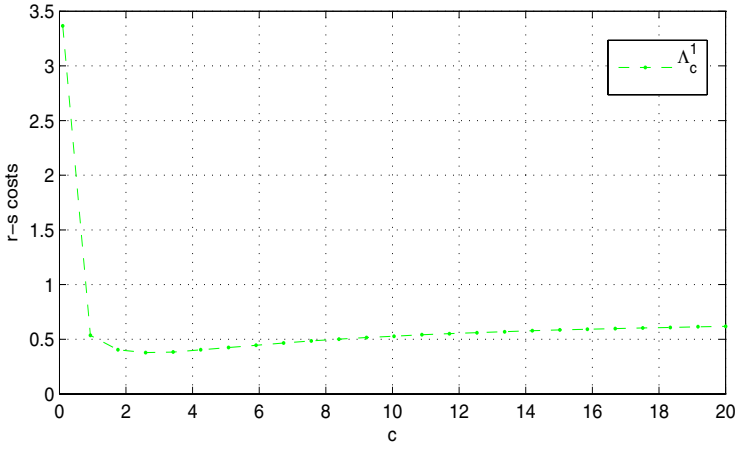


FIGURE 15. Example 4.2, $\frac{1}{c}(B + \bar{\Lambda}_c^1)$ $F(u) = \mathbf{1}\{1/2 \leq u \leq 1\}$, and $\theta \sim \mathcal{N}(0.8, 1)$.

Alternatively, one can consider a Bayesian approach, and introduce hyperparameters (epistemic uncertainties) on known distributions (aleatoric uncertainties) in order to integrate both types into system modeling. For example, suppose the aleatoric random variable is Gaussian with an epistemic mean, where now we assume this epistemic variable takes on a uniform distribution ($X \sim \mathcal{N}(Y, 1)$, $Y \sim U[a, b]$). In the parametric Bayesian approach, if one wants bounds for a collection of distributions on the epistemic variables each joint distribution must be sampled separately. In contrast to the nested iteration approach, the Bayesian hyperparameter approach produces a quantity that averages the epistemic uncertainty. Our approach gives tight upper bounds on performance measures under alternative distributions, without sampling under the various alternative joint distributions.

Most recently, in a paper by Xiu et al. [10], an approach to dealing with epistemic uncertainty *via* polynomial chaos methods was developed. In [10], epistemic uncertainty meant that the true distribution of the underlying stochastic parameters was not known precisely. Here we compare our method of handling this type of epistemic uncertainty to theirs.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

The approach taken in [10] is based on stochastic collocation approximation. Specifically, the differential equation of interest was solved on a generic collocation grid on the sample space of the random variables. This is in contrast to standard uses of stochastic collocation for uncertainty quantification, where the collocation grid is chosen in accordance with the (assumed known) underlying distribution. Of course the reason for the particular choice of basis polynomials and quadrature points is to obtain optimal (spectral) convergence with respect to the L^2 weighted error. In the approach of [10], regardless of what is the "true" (and assumed unknown) distribution, one chooses grid points with respect to the L^2 weighted norm with constant weight (*e.g.*, Legendre-Gauss quadrature weights), even if the original space is unbounded. One then solves the differential equation at these collocation points and constructs the interpolating polynomial of the approximation. Hence, the gPC approximation converges optimally in the L^2 norm only if the underlying distribution is uniform. For non-uniform random variables the approximation will still converge point-wise as the number of collocation points increases (see Sect. 5 in [10]).

Once the interpolated approximation is obtained, one can determine the mean, variance, probabilities, and so on with respect to various underlying distributions on the parameters *via* Monte Carlo or other methods. The benefits are still large, in that one need not evaluate the differential equation for every Monte Carlo sample, but instead evaluates a much simpler approximation *via* a finite sum of polynomial basis functions. However, information can be obtained only by fixing a choice for the underlying distribution. Also, depending on which underlying distribution is used, the errors in the function approximation will have a greater or smaller effect.

In this paper we have developed an approach in which performance bounds can be calculated for the mean, variance, probabilities, and so on with respect to all distributions that are within a particular relative entropy distance from the nominal distribution, and the bounds are optimal for that class of distributions. In addition to providing performance bounds on a class of distributions defined by their relative entropy distance, the limit $c \to \infty$ is particularly interesting in that it gives tight bounds on epistemic uncertainties where one assumes nothing about the underlying distribution of certain variables except their support. In addition, the method one can efficiently handle both aleatoric and epistemic uncertainty simultaneously. More precisely, by computing certain hybrid forms of a risk-sensitive expectation we obtain tight performance bounds that distinguish between variables with known distribution and variables with unknown distribution or other forms of epistemic uncertainty. In particular, the scenarios to which the method applies include (i) aleatoric with known distribution; (ii) aleatoric with partly known distribution (mingled aleatoric and epistemic); (iii) epistemic for which one is willing to model by a family of aleatoric uncertainties, and (iv) epistemic where one is only willing to place bounds on the uncertainties. For all the examples we have considered (including those presented in the paper), approximation of the required risk-sensitive integrals can be done *via* gPC methods using roughly the same computational effort as would be required to compute the corresponding ordinary performance measures using the nominal probability distributions.

Appendix A. Distribution for Polynomial Basis

Here we list some of the most common distributions, both continuous and discrete, and their corresponding gPC basis representations (see [8]).

Distribution	**gPC polynomial basis**
Gaussian	Hermite
Gamma	Laguerre
Beta	Jacobi
Uniform	Legendre
Poisson	Charlier
Binomial	Krawtchouk
Negative Binomial	Meixner
Hypergeometric	Hahn

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Appendix B. Relative Entropy Formulas

In this section we evaluate relative entropies for some of the most common types of distributions listed in Table 4.3. Since most of the distributions listed in this table fall within the exponential family, we start with the general form

$$f_X(x;\boldsymbol{\theta}) = h(x)\exp\left(\sum_{i=1}^{s}\eta_i(\boldsymbol{\theta})T_i(x) - A(\eta_1(\boldsymbol{\theta}),...,\eta_s(\boldsymbol{\theta}))\right), \tag{B.1}$$

where $T_i(x), h(x), \eta_i(\boldsymbol{\theta})$, and $A(\boldsymbol{\theta})$ are known functions, $\boldsymbol{\theta} = (\theta_1,...,\theta_d)^T$ is the parameter of the family.

Example B.1. For Gaussian distributions $\theta = (\mu,\sigma)^T$, where μ is the mean and σ is the variance. This is put in the general form by setting

$$\boldsymbol{\theta} = (\mu,\sigma)^T, \quad \eta = \left(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2}\right)^T, \quad h(x) = \frac{1}{\sqrt{2\pi}}, \quad T(x) = (x,x^2)^T,$$

$$A(\eta(\boldsymbol{\theta})) = \frac{\mu^2}{2\sigma^2} + \log|\sigma| = -\frac{\eta_1^2}{4\eta_2} + \frac{1}{2}\log\left|\frac{1}{2\eta_2}\right|,$$

producing

$$f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

We now state the relative entropy formula for distributions which have a probability density function given by the form of exponential family in (B.1). The relative entropy between a probability measure P with density $p(x)$ and another probability measure Q with density $q(x)$ is given by

$$R(P\,\|Q\,) = \int_{\Gamma} p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \tag{B.2}$$

whenever P is absolutely continuous with respect to Q, where Γ is the support of p and q. In all other cases it is defined to be ∞.

A simplified version of the relative entropy formula for distributions from the same exponential family type is available. Suppose that $q(x)$ and $p(x)$ have the same h, η_i, T_i, A, but with different parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ (note that $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors). If we associate $q(x)$ with $\boldsymbol{\theta}_1$ and $p(x)$ with $\boldsymbol{\theta}_2$, then it can be shown that the relative entropy formula reduces to

$$R(P\,\|Q\,) = \sum_{i=1}^{s}(\eta_i(\boldsymbol{\theta}_2) - \eta_i(\boldsymbol{\theta}_1))\int_{\Gamma} T_i(x)p(x)\mathrm{d}x - A(\boldsymbol{\eta}(\boldsymbol{\theta_2})) + A(\boldsymbol{\eta}(\boldsymbol{\theta}_1)). \tag{B.3}$$

Gaussian. If $Q = N(\mu_1,\sigma_1)$ and $P = N(\mu_2,\sigma_2)$, then using the fact that $\int T_1(x)p(x)\mathrm{d}x = \mu_2$ and $\int T_2(x)p(x)\mathrm{d}x = \mu_2^2 + \sigma_2^2$, we get

$$R(P\,\|Q\,) = \frac{1}{2\sigma_1^2}\left((\mu_1^2 - \mu_2^2) \;+\; (\sigma_2 - \sigma_1)\right) + \log\frac{\sigma_1}{\sigma_2}\,. \tag{B.4}$$

Beta. For the beta distribution, the parameters in the exponential family are given by

$$\eta = (\alpha-1,\beta-1)^T, \quad T = (\log(x),\log(1-x))^T, \quad h(x) = 1, \quad A = \log\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

where

$$f_X(x;\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0<x<1,\ \alpha,\beta>1.$$

Another form of the beta distribution, which is more closely related to the Jacobi polynomials, is given by

$$\tilde{f}_X(k;\tilde{\alpha},\tilde{\beta}) = \frac{(1-x)^{\tilde{\alpha}}(1+x)^{\tilde{\beta}}}{2^{\tilde{\alpha}+\tilde{\beta}+1}B(\tilde{\alpha}+1,\tilde{\beta}+1)}, \qquad -1<x<1,\ \tilde{\alpha},\tilde{\beta}>-1,$$

where $B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the beta function. Note that the relation between the two is simply a rescaling and a variable substitution given by $\tilde{\alpha}=\beta-1, \tilde{\beta}=\alpha-1$. We will use the primary form to determine the relative entropy formula. Letting $P=\text{beta}(\alpha_2,\beta_2)$ and $Q=\text{beta}(\alpha_1,\beta_1)$,

$$\int_0^1 T_1(x)p(x)\mathrm{d}x = \psi(\alpha_2)-\psi(\alpha_2+\beta_2), \quad \int_\Gamma T_2(x)p(x)\mathrm{d}x = \psi(\beta_2)-\psi(\alpha_2+\beta_2),$$

where $\psi(x)$ is known as the digamma function, or the zero$^{\text{th}}$ order of the polygamma function. Then,

$$\begin{aligned} R(P\,\|Q\,) = {} & (\alpha_2-\alpha_1)(\psi(\alpha_2)-\psi(\alpha_2+\beta_2)) \\ & + (\beta_2-\beta_1)(\psi(\beta_2)-\psi(\alpha_2+\beta_2)) \ + \ \log\frac{B(\alpha_1,\beta_1)}{B(a_2,\beta_2)}. \end{aligned}$$

Gamma. Here we have

$$\eta=(-\beta,\alpha-1)^T, \quad T=(x,\log(x))^T, \quad h(x)=1, \quad A=\log\Gamma(\alpha)-\alpha\log\beta$$

If $P=\text{gamma}(\alpha_2,\beta_2), Q=\text{gamma}(\alpha_1,\beta_1)$, where α_i,β_i are the so-called shape parameters, then

$$R(P\,\|Q\,) = \frac{\alpha_1}{\beta_1}(\beta_2-\beta_1) + \ (\alpha_1-\alpha_2)(\psi(\alpha_1))-\log(\beta_1)) + \log\frac{\Gamma(\alpha_2)\beta_1^{\alpha_1}}{\Gamma(\alpha_1)\beta_2^{a_2}}.$$

Binomial. Here

$$\eta=(-\beta,\alpha-1)^T, \quad T=(x,\log(x))^T, \quad h(x)=1, \quad A=\log\Gamma(\alpha)-\alpha\log\beta.$$

If $P=\text{binomial}(n,p_2), Q=\text{binomial}(n,p_1)$, then

$$R(P\,\|Q\,) = \log\frac{p_1^{\mu_1}(1-p_2)^{\mu_1-n}}{p_2^{\mu_1}(1-p_1)^{\mu_1-n}}, \qquad \mu_1=np_1.$$

Poisson. In this case

$$\eta=\log\lambda, \quad T=x, \quad h(x)=\frac{1}{x!}, \quad A=\lambda$$

If $P=\text{Poisson}(\lambda_2), Q=\text{Poisson}(\lambda_1)$, then the relative entropy distance is

$$R(P\,\|Q\,) = \lambda_1-\lambda_2 \ + \ \lambda_2\log\frac{\lambda_2}{\lambda_1}.$$

Note that relative entropy formula is invariant under shifting and scaling of both distributions simultaneously. In fact, suppose that ψ and its inverse are both well defined and measurable, X and Y have distributions P and Q, and that $\psi(X)$ and $\psi(Y)$ have distributions $\bar{P}$ and $\bar{Q}$. Then [2], Lemma E.2.1 $R(\bar{P}\,\|\bar{Q}\,)=R(P\,\|Q\,)$. A list of relative entropy formulas follows.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press

Distribution	$\mathbf{R}((\cdot)_2 \,\|\|\, (\cdot)_1)$
Gaussian	$\frac{1}{2\sigma_1^2}\left[(\mu_1^2-\mu_2^2)+(\sigma_2-\sigma_1)\right]+\log\frac{\sigma_1}{\sigma_2}$
Beta/Uniform	$(\alpha_2-\alpha_1)\left[\psi(\alpha_2)-\psi(\alpha_2+\beta_2)\right]$ $+(\beta_2-\beta_1)\left[\psi(\beta_2)-\psi(\alpha_2+\beta_2)\right]+\log\frac{B(\alpha_1,\beta_1)}{B(\alpha_2,\beta_2)}$
Gamma	$\frac{\alpha_1}{\beta_1}(\beta_2-\beta_1)+(\alpha_1-\alpha_2)\left[\psi(\alpha_1)-\log\beta_1\right]$ $+\log\left[\Gamma(\alpha_2)\beta_1^{\alpha_1}/\Gamma(\alpha_1)\beta_2^{\alpha_2}\right]$
Binomial	$\log\left[p_1^{\mu_1}(1-p_2)^{\mu_1-n}/p_2^{\mu_1}(1-p_1)^{\mu_1-n}\right]$
Poisson	$\lambda_1-\lambda_2+\lambda_2\log\frac{\lambda_2}{\lambda_1}$.

References

[1] R.K. Boel, M.R. James and I.R. Petersen, Robustness and risk sensitive filtering. *IEEE Trans. Auto. Control* **3** (2002) 451–461.

[2] P. Dupuis and R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York (1997).

[3] P. Dupuis, M.R. James and I.R. Petersen, Robust properties of risk–sensitive control. *Math. Control Signals Syst.* **13** (2000) 318–332.

[4] M. Eldred and L. Swiler, *Efficient algorithms for mixed aleatory-epistemic uncertainty quantification with applications to radiation-hardened electronics*. Technical report, Sandia National Laboratories (2009).

[5] O.P. Le Maitre and O.M. Knio, *Spectral Methods for Uncertainty Quantification*. Springer, New York (2010).

[6] S.R.S. Varadhan, Large Deviations and Applications. *CBMS-NSF Regional Conference Series in Mathematics. SIAM, Philadelphia* (1984).

[7] D. Xiu, Efficient collocational approach for parametric uncertianty analysis. *J. Comput. Phys.* **2** (2007) 293–309.

[8] D. Xiu, Fast numerical methods for stochastic computations. *Commun. Comput. Phys.* **5** (2009) 242–272.

[9] D. Xiu and J. Hesthaven, High-order collocation methods for differential equations with random inputs. *Soc. Industrial Appl. Math.* **27** (2005) 1118–1139.

[10] D. Xiu, J. Jakeman and M. Eldred, Numerical approach for quantification of epistemic uncertainty. *Commun. Comput. Phys.* **229** (2010) 4648–4663.

[11] D. Xiu and G. Karniadakis, The Weiner-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24** (2002) 619–644.

https://doi.org/10.1051/m2an/2012038 Published online by Cambridge University Press