

# Uncertainty Quantification using Exponential Epi-Splines

J.O. Royset

*Operations Research Department  
Naval Postgraduate School, Monterey, USA*

N. Sukumar

*Department of Civil and Environmental Engineering  
University of California, Davis, USA*

R. J-B Wets

*Department of Mathematics  
University of California, Davis, USA.*

**ABSTRACT:** We quantify uncertainty in complex systems by a flexible, nonparametric framework for estimating probability density functions of output quantities of interest. The framework systematically incorporates soft information about the system from engineering judgement and experience to improve the estimates and ensure that they are consistent with prior knowledge. The framework is based on a maximum likelihood criterion, with epi-splines facilitating rapid solution of the resulting optimization problems. In four numerical examples with few realizations of the system output, we identify the main features of output densities even for nonsmooth and discontinuous system function and high-dimensional inputs.

## 1 INTRODUCTION

We address uncertainty quantification (UQ) in complex systems by a flexible, nonparametric framework for estimating probability density functions of random output quantities of interest. The framework systematically incorporates hard information derived from physics-based sensors, field test data, and computer simulations as well as soft information from engineering knowledge and experience. Our ability to account for soft information about the system overcome, in part, the poor accuracy traditionally experienced when applying classical statistical methods to UQ. The framework is based on *epi-splines* for consistent approximation of infinite-dimensional optimization problems arising in the estimation process as developed by Royset and Wets (2013).

UQ aims to characterize a random vector  $\mathbf{Y}$  that describes the output (response) from some system, modeled by a function  $g$ , subject to input given by a random vector  $\mathbf{V}$ . We use uppercase to denote random variables and lowercase for their realizations. Bold-face type indicates a vector. Specifically,

$$\mathbf{Y} = g(\mathbf{V}).$$

While the distribution of  $\mathbf{V}$  may be known, or as-

sumed known, the distribution of  $\mathbf{Y}$  is not easily available due to the complexity of  $g$ . In practice,  $g$  may be given in terms of the solution of differential and algebraic equations, or other computationally intense problems. In this case, the time required to evaluate  $g$  at a single point  $\mathbf{v}$  limits the number of realizations of  $\mathbf{Y}$  one can generate. Moreover, in the case of physical experiments and other situations, we may only have available a sample of input-output realizations  $(\mathbf{v}^1, \mathbf{y}^1)$ ,  $(\mathbf{v}^2, \mathbf{y}^2)$ , ...,  $(\mathbf{v}^n, \mathbf{y}^n)$ , with  $\mathbf{y}^j = g(\mathbf{v}^j)$ , of finite, and typically small, size that cannot be augmented without exorbitant costs. In the absence of a large set of realizations of  $\mathbf{Y}$ , empirical estimates of the mean, standard deviation, and other descriptions of  $\mathbf{Y}$  tend to be highly inaccurate and may lead to significant underestimation of the variability. Since the variability of the random vector  $\mathbf{V}$  may be substantial, sensitivity and perturbation methods (Ghanem and Spanos 1991, Kleiber et al. 1992) applied near the mean of  $\mathbf{V}$  may be highly inaccurate as well.

Polynomial chaos expansion, stochastic Galerkin, and stochastic collocation methods are frequently used for UQ in an effort to overcome these difficulties (Ghanem and Spanos 1991, Xiu and Karniadakis 2002, Babuska et al. 2004, Ganapathysubramanian and Zabararas 2007, Nobile et al. 2008, Eldred

et al. 2011). These methods build a model of  $g$  using polynomial functions and then estimate moments of  $\mathbf{Y}$  using that model and the (assumed) knowledge of the distribution of  $\mathbf{V}$ . Under certain assumptions, the model may approach  $g$  exponentially fast as the fidelity of the model tends to infinity, for example in the mean-square sense. A main difficulty with expansion methods is that the accuracy deteriorates as the dimensionality of  $\mathbf{V}$  grows,  $g$  becomes less ‘smooth,’ and the distribution of  $\mathbf{V}$  deviates from those compatible with polynomials in the Wiener-Askey scheme (Xiu and Karniadakis 2002). Moreover, there is no systematic way of including soft information about the distribution of  $\mathbf{Y}$ , for example from extensive experience with similar systems and specific properties of  $g$ , in the estimation process. These difficulties are present even if  $\mathbf{Y}$  is scalar valued. Smolyak sparse-grid approaches (Xiu and Hesthaven 2005, Ganapathysubramanian and Zabaras 2007, Nobile et al. 2008), separated representations based on alternating least-squares approximation techniques (Doostan and Iaccarino 2009), and related approaches are steps in the direction towards handling moderate dimensions of  $\mathbf{V}$ , but still significant challenges remain when handling high-dimensional cases. We refer to Helton and Pilch (2011) for an overview of recent studies on UQ.

In this paper, we propose an approach to UQ that is insensitive to the dimension of  $\mathbf{V}$ , does not rely on the smoothness of  $g$ , and handles an arbitrary density for  $\mathbf{V}$ . We estimate the probability density function of  $\mathbf{Y}$  (assuming it exists) using the available sample  $\{\mathbf{y}^j\}_{j=1}^n$ , which may be small, and all available soft information about  $\mathbf{Y}$ . Such information may come in the form of knowledge about the support of the probability density function, its continuity, smoothness, unimodality, monotonicity, moments, and other characteristics. Naturally, if soft information could be included in the estimation process, it might greatly reduce estimator error and variance, and enhance visual appeal of the estimates. This effect is especially prominent in the case of few observations of  $\mathbf{Y}$ . While our general approach applies to joint probability density functions, we here focus on a single output quantify and the estimation of its density. Therefore in the remainder of the paper, we let  $\mathbf{Y}$  be scalar valued.

The consideration of soft information in density estimation is hardly new. Classical parametric estimation selects a parametric family of densities, presumably based on soft information, and determines the ‘best’ estimate within that family. Bayesian estimation makes use of prior soft information extensively, but results extend much beyond that premise (Wahba 1981, Van de Geer 1987, Thompson and Tapia 1990, Wets 1991, Dupacova 1992, Samaniego and Reneau 1994, Geyer 1994, Groenenboom et al. 2001). In nonparametric density estimation, practitioners may also adjust estimates based on their experience in an ad-hoc manner. Desirable density estimators can be achieved by means of penalties (Good and Gaskin

1971, de Montricher et al. 1975, Leonard 1978, Klonias 1982, Silverman 1982). While in principle many types of constraints in the estimation problem can be represented by penalty terms, the equivalence of such reformulations depends on the successful selection of multiplier and penalty parameters which is nontrivial in practice. We view density estimation problems with soft information as a constrained stochastic optimization problem, where the goal is to determine a density that maximizes the likelihood function generated by a sample subject to constraints on the density derived from the soft information. While the connection between density estimation and optimization has long been recognized (see Thompson and Tapia (1990) and Wets (1991) and more recently Dong and Wets (2007) and Casey and Wets (2013)), our framework offers flexibility in the inclusion of soft information and a solid theoretical foundation of asymptotic results; see Royset and Wets (2013) for details.

The remainder of the paper is organized as follows. Section 2 defines epi-splines and the maximum likelihood problem. Section 3 discusses soft information and its implementation. Section 4 illustrates the approach with four numerical examples. The paper ends with conclusions in Section 5.

## 2 EXPONENTIAL EPI-SPLINE ESTIMATE

Given a sample  $y^1, y^2, \dots, y^n$  from  $Y$ , we consider an estimate

$$f^n(y) = \exp(-s^n(y)), \quad y \in \mathbb{R},$$

of the density of  $Y$ , where  $s^n$  is an appropriately selected *epi-spline* in view of the sample and other considerations. Naturally, we refer to  $f^n$  as an *exponential epi-spline estimate*. As we see below, we determine  $s^n$  by maximizing the likelihood function associated with the sample, while simultaneously ensuring that constraints derived from soft information are met. The use of the exponential function ensures a nonnegative  $f^n$ , which, of course, is a basic requirement for a density. While finding an arbitrary function  $s^n$  involves an infinite-dimensional optimization problem, the restriction to an epi-spline allows the use of a finite number of parameters and still maintains flexibility to approximate essentially any density. Once  $f^n$  is determined, the estimation of moments of  $Y$  and other quantifies is easily carried out by numerical integration.

We observe that the estimation of  $f^n$  implicitly accounts for the system function  $g$  in two ways. The first way relates to the generation of a sample. A sample point  $y^j$  is typically generated by sampling or carefully selecting an input  $\mathbf{v}^j$ , evaluating  $g(\mathbf{v}^j)$ , and then setting  $y^j = g(\mathbf{v}^j)$ . Consequently, the sample  $y^1, y^2, \dots, y^n$  includes ‘hard’ information about  $g$ . We do not examine the design of experiments, i.e., the selection of  $\mathbf{v}$  at which to evaluate  $g$ . However, a density established through the present framework for a

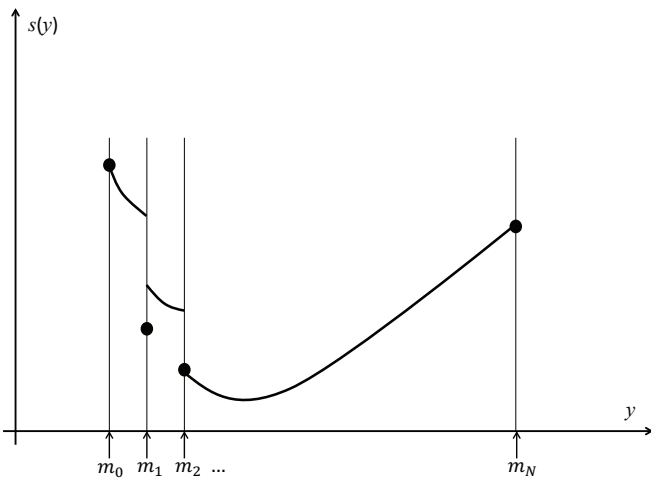


Figure 1: Illustration of epi-spline.

given sample provides an informed basis for selecting a next sample point. The second way derives from soft information. As we see in Section 3, soft information about  $g$  may translate into soft information on  $Y$ , which is incorporated as constraints in the likelihood maximization. In contrast to expansion methods for UQ, we do not attempt to build a model of  $g$  directly.

## 2.1 Epi-Splines

We follow Royset and Wets (2013) and define an *epi-spline* in terms of its *order*, given by a nonnegative integer  $p$ , its *number of partitions*  $N$ , and its *mesh*  $m = \{m_k\}_{k=0}^N$ , with  $m_{k-1} < m_k$ ,  $k = 1, 2, \dots, N$ . Given these quantities, an epi-spline is a real-valued function defined on the closed interval  $[m_0, m_N]$  that is polynomial of degree  $p$  in each segment  $(m_{k-1}, m_k)$ ,  $k = 1, 2, \dots, N$ , and that is finite valued at  $m_0, m_1, \dots, m_N$ . For  $y \notin [m_0, m_N]$ , we assign  $\infty$  to the epi-spline. Figure 1 gives an example of an epi-spline.

It is clear from the definition that an epi-spline is uniquely defined by a finite number of parameters. Each segment  $(m_{k-1}, m_k)$  requires  $p + 1$  parameters to define a polynomial of degree  $p$ . Since there are  $N$  segments and the  $N + 1$  mesh points are defined separately, the total number of parameters is  $(p + 2)N + 1$ . We organize these parameters into a vector

$$\mathbf{r} = (s_0, s_1, \dots, s_N, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N),$$

where  $s_k$ ,  $k = 0, 1, \dots, N$ , are scalars giving the values of an epi-spline at the mesh points and

$$\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,p}),$$

is a  $(p + 1)$ -dimensional vector representing the coefficients in the polynomial of the  $k$ th segment,  $k = 1, 2, \dots, N$ . We refer to  $\mathbf{r}$  as the *epi-spline parameter*. We use the notation  $y_k = y - m_{k-1}$ ,  $k = 1, 2, \dots, N$ , and denote by  $0_k$  the  $k$ -dimensional zero vector, with  $0_0$  being a term that is omitted. Then, every epi-spline  $s$  is of the form

$$s(y) = \langle \mathbf{c}(y), \mathbf{r} \rangle, \quad y \in [m_0, m_N],$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product and  $\mathbf{c}$  is a vector of basis functions. Specifically,

$$\mathbf{c}(y) = \begin{cases} (0_{N+1+(p+1)(k-1)}, 1, y_k, \dots, y_k^p, 0_{(p+1)(N-k)}) & \text{if } y \in (m_{k-1}, m_k), k = 1, 2, \dots, N \\ (0_k, 1, 0_{N-k+(p+1)N}) & \text{if } y = m_k, k = 0, 1, \dots, N. \end{cases}$$

In essence, if  $y$  coincides with a mesh point,  $\mathbf{c}$  simply assigns  $s(y)$  to be equal to the component of  $\mathbf{r}$  corresponding to that mesh point. If  $y \in (m_{k-1}, m_k)$ , then  $\mathbf{c}$  selects the components of  $\mathbf{r}$  that describe the polynomial in that segment, i.e.,

$$s(y) = \sum_{i=0}^p a_{k,i} (y - m_{k-1})^i.$$

It is clear that epi-splines relate to classical splines, but we allow for jumps and our construction is different. An epi-spline of any order  $p$  can approximate to an arbitrary accuracy any lower- and upper-semicontinuous function by increasing the number of partitions  $N$  (Royset and Wets 2013). Consequently, we have the ability to approximate essentially any density that one may encounter in practice. In fact, many common density functions are *exactly* represented on  $[m_0, m_N]$ . For example, a normal density is of the form  $\exp(-s(y))$ , with  $s$  being an epi-spline of order 2 for any number of partitions  $N$  and mesh  $m$ . An exponential density is of the form  $\exp(-s(y))$ , with  $s$  being an epi-spline of order 1, for any number of partitions  $N$  and mesh  $m$ , with  $m_0 = 0$ . Even more densities, such as the lognormal and the Pareto, are exactly represented on a bounded interval after a logarithmic transformation.

## 2.2 Maximum Likelihood

We proceed by adopting a maximum likelihood criterion to determine the best exponential epi-spline that, of course, integrates to unity. For realizations  $y^1, \dots, y^n$  drawn from a density  $f(y) = \exp(-s(y))$ , the maximum likelihood function takes the form

$$\prod_{j=1}^n f(y^j) = \exp\left(-\sum_{j=1}^n s(y^j)\right) = \exp\left(-\sum_{j=1}^n \langle \mathbf{c}(y^j), \mathbf{r} \rangle\right),$$

where the last equality follows from the assumption that  $s$  is an epi-spline with epi-spline parameter  $\mathbf{r}$ . Since optimal solutions are invariant to a logarithmic transformation, we prefer instead of maximizing this expression to maximize

$$-\sum_{j=1}^n \langle \mathbf{c}(y^j), \mathbf{r} \rangle,$$

which is linear in  $\mathbf{r}$ . Moreover, the maximization is subject to the constraint

$$\int_{m_0}^{m_N} \exp(-\langle \mathbf{c}(y), \mathbf{r} \rangle) dy = 1, \quad (1)$$

which ensures that the resulting solution actually is a density, as well as any other constraints that may be generated by soft information as discussed further in Section 3. As we see below, the latter constraints are typically linear and consequently easy to handle. Under rather general assumptions, (1) can be relaxed to a  $\leq$  constraint, or even being moved to the objective function as a ‘penalty,’ resulting in a convex optimization problem for which there are well-developed algorithms. Even if the equality constraint must remain, the maximum likelihood problem can usually be solved by nonlinear programming algorithms with little difficulties; see Royset and Wets (2013) for details.

An optimal solution  $\mathbf{r}^n$  obtained by solving the above maximum likelihood problem gives an estimate  $f^n(y) = \exp(-\langle \mathbf{c}(y), \mathbf{r}^n \rangle)$ ,  $y \in [m_0, m_N]$ , of the ‘true’ density. Under general assumptions, the estimate converges to the true one as the sample size  $n$  tends to infinity. Moreover, moment estimates obtained by integrating  $f^n$  converge also to the correct values. We refer to Royset and Wets (2013) for a thorough analysis.

### 3 SOFT INFORMATION

We next discuss how to specify and implement soft information about  $Y$ . We consider specific examples, but omit a discussion of tail-related soft information. The latter information becomes important if the resulting densities are to be used in reliability analysis. However, since estimating tail behavior requires special care in the presence of little data, we here avoid this situation and instead focus on the ‘central’ characteristic of a density.

**Support bounds and mesh.** The choice of mesh  $m = \{m_k\}_{k=0}^N$  accounts for support bounds and  $m_0$  and  $m_N$  should, ideally, correspond to the lower and upper bounds of  $Y$ , respectively. If  $\mathbf{V}$  is bounded and  $g$  continuous, then  $m_0$  and  $m_N$  can be determined, at least in principle, by minimizing and maximizing  $g$  over the possible input values, respectively. Even if this is not possible, insight about the problem may provide estimates of lower and upper bounds on  $Y$  that can serve as  $m_0$  and  $m_N$ . The mesh is often selected to be uniform, but if discontinuities and intervals with steep slopes can be anticipated other choices may be preferred.

**Continuity.** We ensure that an exponential epi-spline estimate is continuous by imposing the constraint

$$s_{k-1} = a_{k,0}, \quad s_k = \sum_{i=0}^p a_{k,i} (m_k - m_{k-1})^i, \quad k = 1, \dots, N.$$

Of course, by omitting some of the above constraints, one has the ability to ensure continuity at some but

not all mesh points. All of these constraints are linear.

**Smoothness.** We restrict the search to  $r$ -times continuously differentiable densities, with  $r \leq p$ , by imposing the conditions for continuity and the linear constraints for  $k = 1, 2, \dots, N - 1, j = 1, 2, \dots, r$ ,

$$\sum_{i=j}^p \prod_{l=0}^{j-1} (i-l) a_{k,i} (m_k - m_{k-1})^{i-j} = a_{k+1,j}.$$

Higher order smoothness is automatically achieved if these constraints are imposed with  $r = p$ . Again, selective implementation of these constraints could be a useful tool in practice.

**Pointwise Fisher information.** We define the pointwise Fisher information of an exponential epi-spline density  $f$  at  $y$  to be

$$f'(y)/f(y) = -\langle \mathbf{c}'(y), \mathbf{r} \rangle = -\sum_{i=1}^p i a_{k,i} (y - m_{k-1})^{i-1}$$

for  $y \in (m_{k-1}, m_k)$ . Upper and lower bounds on this quantity result in linear constraints. The constraints could be imposed at any number of values of  $y$ , but we note that if  $p = 2$  and the density is strongly unimodal, as describe below, and continuously differentiable, then lower bounds on  $f'(y)/f(y)$  at  $m_1, m_2, \dots, m_N$  suffices to ensure that the constraints are satisfied for all  $y \in [m_0, m_N]$ . Similarly, an upper bound on  $f'(y)/f(y)$  need only be imposed at  $m_0, m_1, \dots, m_{N-1}$ . In applications, one can obtain guidance for appropriate values of upper and lower bounds on these quantities by considering standard densities. For example, a normal density with mean  $\mu$  and standard deviation  $\sigma$  has  $f'(y)/f(y) = -(y - \mu)/\sigma^2$ . For the exponential density with parameter  $\lambda$ ,  $f'(y)/f(y) = -\lambda$  for all  $y \geq 0$ .

**Monotonicity.** We achieve a nondecreasing (nonincreasing) density by imposing nonnegativity (nonpositivity) on  $f'(y)/f(y)$  for all  $y \in (m_{k-1}, m_k), k = 1, 2, \dots, N$  as described above as well as for  $k = 1, 2, \dots, N$ ,

$$s_{k-1} \geq (\leq) a_{k,0}, \quad s_k \leq (\geq) \sum_{i=0}^p a_{k,i} (m_k - m_{k-1})^i.$$

Again, simplifications arise, for example, if  $p = 2$  and the density is strongly unimodal. Then, it suffices to impose that  $a_{k,1} + 2a_{k,2}(m_k - m_{k-1}) \leq 0$  ( $a_{k,1} \geq 0$ ),  $k = 1, 2, \dots, N$ .

**Strongly unimodal.** We recall that a density is strongly unimodal if it is continuous and log-concave. Consequently, if  $f(y) = \exp(-\langle \mathbf{c}(y), \mathbf{r} \rangle)$ ,  $y \in [m_0, m_N]$ , strong unimodality requires that  $\langle \mathbf{c}(y), \mathbf{r} \rangle$  is a convex function in  $y$ . This condition is

ensured if (i) continuity is imposed (see above), (ii) for  $k = 1, 2, \dots, N - 1$ , the epi-spline's left derivatives at  $m_k$  is no larger than its right derivative, i.e., for  $k = 1, 2, \dots, N - 1$ ,

$$\sum_{i=1}^p ia_{k,i}(m_k - m_{k-1})^{i-1} \leq a_{k+1,1},$$

and (iii) on each  $(m_{k-1}, m_k)$ ,  $k = 1, 2, \dots, N$ ,  $\langle \mathbf{c}(y), \mathbf{r} \rangle$  is a convex function in  $y$ , i.e., for  $k = 1, 2, \dots, N$ ,  $y \in (m_{k-1}, m_k)$ ,

$$\sum_{i=2}^p i(i-1)a_{k,i}(y - m_{k-1})^{i-2} \geq 0.$$

Here, the obvious interpretations are required when  $p = 0, 1$ . The latter condition simplifies to  $a_{k,2} \geq 0$ ,  $k = 1, 2, \dots, N$ , when  $p = 2$ .

**Bounds on density values.** It is also straightforward to impose pointwise upper and lower bounds  $u(y)$  and  $l(y)$  on the value of  $f(y) = \exp(-\langle \mathbf{c}(y), \mathbf{r} \rangle)$ , with  $0 < l(y) \leq u(y) < \infty$ . It suffices to set for  $y \in (m_{k-1}, m_k)$ ,

$$-\log l(\xi) \geq \sum_{i=0}^p a_{k,i}(\xi - m_{k-1})^i \geq -\log u(\xi)$$

and for  $y = m_k$ ,

$$-\log l(\xi) \geq s_k \geq -\log u(\xi), k = 0, 1, \dots, N.$$

These constraints are also linear.

**Kullback-Leibler divergence.** In some applications, it is useful to only seek densities that are close to some known density in some sense. ‘Distances’ between densities are conveniently measured by the Kullback-Leibler divergence from a density  $f_1$  to a density  $f_2$  as given by

$$d_{KL}(f_1||f_2) = \int_{-\infty}^{\infty} f_1(y) \log \frac{f_1(y)}{f_2(y)} dy.$$

Here we make the standard interpretation that  $\beta_1 \log \beta_1 / \beta_2 = 0$  when  $\beta_1 = 0$  regardless of the value of  $\beta_2$  and  $\beta_1 \log \beta_1 / \beta_2 = \infty$  when  $\beta_1 > 0$  and  $\beta_2 = 0$ . If  $f_2$  is an exponential epi-spline, with epi-spline parameter  $\mathbf{r}$  and  $f_2(y) > 0$  whenever  $f_1(y) > 0$ , then

$$\begin{aligned} d_{KL}(f_1||f_2) &= \left\langle \int_{\mathcal{Y}} \mathbf{c}(y) f_1(y) dy, \mathbf{r} \right\rangle + \int_{\mathcal{Y}} (\log f_1(y)) f_1(y) dy, \end{aligned}$$

where  $\mathcal{Y}$  is the part of the real line with  $f_1(y) > 0$ . (We observe that if  $f_2(y) = 0$  for more than a countable number of  $y$  with  $f_1(y) > 0$ , then  $d_{KL}(f_1||f_2) = \infty$

and the comparison is less interesting.) Consequently, a linear constraint

$$\left\langle \int_{\mathcal{Y}} \mathbf{c}(y) f_1(y) dy, \mathbf{r} \right\rangle + \int_{\mathcal{Y}} (\log f_1(y)) f_1(y) dy \leq \kappa$$

in the likelihood maximization ensures that the resulting estimate  $f^n(y) = \exp(-\langle \mathbf{c}(y), \mathbf{r}^n \rangle)$  is no further away from the reference density  $f_1$  than  $\kappa$ , as measured by the Kullback-Leibler divergence. The choice of  $\kappa$  in applications can be informed by the fact that the Kullback-Leibler divergence between two normal densities,  $\phi_1$  and  $\phi_2$ , with mean and standard deviation  $\mu_1, \sigma_1$  and  $\mu_2, \sigma_2$ , respectively, takes the form

$$\begin{aligned} d_{KL}(\phi_1||\phi_2) &= \log(\sigma_2/\sigma_1) - \frac{1}{2} \\ &+ \frac{1}{2\sigma_2^2}(\sigma_1^2 + (\mu_1 - \mu_2)^2 + 2(\sigma_1^2 + \mu_1^2 - \mu_2^2)). \end{aligned}$$

For two exponential densities,  $f_1$  and  $f_2$ , with parameters  $\lambda_1$  and  $\lambda_2$ , respectively,

$$d_{KL}(f_1||f_2) = \log(\lambda_1/\lambda_2) - 1 + \lambda_2/\lambda_1.$$

**Bounds on moments.** Soft information may result in a constraint on the  $j$ -th moment of the form

$$l \leq \int_{m_0}^{m_N} y^j e^{-\langle \mathbf{c}(y), \mathbf{r} \rangle} dy \leq u,$$

where  $l, u \in \mathbb{R}$ ,  $l \leq u$  are given constants. The right-most inequality results in a convex constraint on  $\mathbf{r}$ , while the left-most in a nonconvex constraint.

**Bounds on cumulative distribution functions.** If the cumulative distribution function at  $\gamma \in [m_0, m_N]$  of the estimated density  $f^n$  must lie between the lower bound  $l$  and the upper bound  $u$ , then we obtain the following constraints for the maximum likelihood problem:

$$\int_{m_0}^{\gamma} e^{-\langle \mathbf{c}(y), \mathbf{r} \rangle} dy \leq u \text{ and } \int_{\gamma}^{m_N} e^{-\langle \mathbf{c}(y), \mathbf{r} \rangle} dy \leq 1 - l,$$

which are both convex.

**Gradient information.** If the input  $\mathbf{v}$  is a scalar and the derivative of  $g$  is known at a point  $v$ , then the density  $f$  of  $Y$  must satisfy

$$f(g(v)) \geq \frac{\phi(v)}{|g'(v)|}, \quad (2)$$

where  $\phi$  is the density of  $V$ , with equality holding if  $g$  is strictly increasing or strictly decreasing. Consequently, the right-hand side provides a lower bound on the density estimate  $f^n$  in the likelihood maximization problem.

We illustrate the UQ approach with four examples, where likelihood maximization problems are solved by ‘fmincon’ in Matlab versions 7.10.0 and continuously differentiable epi-splines of order 2 with equally spaced mesh are used. If not otherwise stated, we use  $N = 50$  partitions. If there is no soft information about support bounds, we set  $m_0$  ( $m_N$ ) to two standard errors below (above) the average sample point. The Gauss-Legendre quadrature rule with 20 points evaluates the integrals over each segment  $(m_{k-1}, m_k)$  when needed. (Only a highly varying density demands such accuracy, but we default to this as it comes at little computational cost.) For comparison, we also compute kernel estimates of densities using ‘ksdensity’ in Matlab, with the default normal kernel.

#### 4.1 Example 1: Bar with random Young’s modulus

We consider the differential equation

$$(E(x, \mathbf{v})u'(x))' = 0, \quad x \in (0, 1) \quad (3)$$

with boundary conditions  $u(0) = u_0$  and  $E(1, \mathbf{v})u'(1) = 1$  describing the displacement of a bar with unit length, where  $E(x, \mathbf{v})$  is Young’s modulus at  $x$  under realization  $\mathbf{v} = (v_1, v_2, \dots, v_{100})$ . We let  $E(x, \mathbf{v}) = v_i$  on  $((i-1)/100, i/100)$ ,  $i = 1, 2, \dots, 100$ , i.e., piecewise constant. We assume that  $\mathbf{v}$  is unknown and we model it by a 100-dimensional random vector  $\mathbf{V}$ , with each component being uniformly distributed on  $[0.1, 0.5]$  and independent.

Figure 2 illustrates the estimated density  $f^n$  of the endpoint displacement  $u(1)$ , which is random due to the randomness in  $E$ , based on a sample of size  $n = 10$  of that displacement. The sample is obtained by drawing independently 10 realizations of  $\mathbf{V}$  and then solving (3), and is illustrated by circles in Figure 2. The solid curve gives  $f^n$  under the additional soft information requiring that the density is unimodal and within 0.01 in Kullback-Leibler divergence from a normal density with mean 4 and standard deviation 0.2. Moreover, in this case it is easy to determine upper and lower bounds of the endpoint displacement, which are 10 and 2, respectively. The kernel estimate, using the same sample, is given by a dashed curve. We note that the kernel estimate is unable to take advantage of the unimodal and Kullback-Leibler soft information. (In this case, the kernel estimate is unimodal by coincidence.) The dotted curve in Figure 2 gives the ‘true’ density as estimated by a sample of size  $10^5$ . We find that the exponential epi-spline estimate  $f^n$  is quite close to the true density even with only 10 sample points.

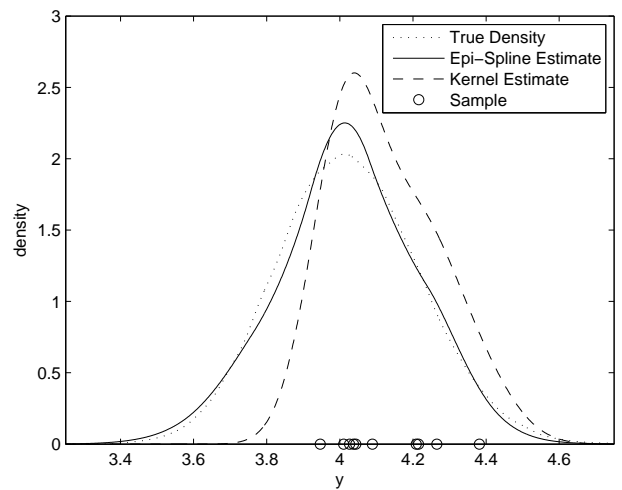


Figure 2: Density of endpoint displacement in Example 1.

#### 4.2 Example 2: Linear system under harmonic excitation

The next example is taken from Chopra (1995, Section 12.1) and deals with a two degree-of-freedom linear system governed by

$$\begin{aligned} m_1 \ddot{u}_1(t) + (k_1 + k_2)u_1(t) - k_2 u_2(t) &= p_o \sin vt \\ m_2 \ddot{u}_2(t) - k_2 u_1(t) + k_2 u_2(t) &= 0 \end{aligned}$$

where  $m_1$  and  $m_2$  are the masses for nodes one and two, respectively,  $k_1$  and  $k_2$  are the corresponding stiffnesses,  $p_o$  is the excitation magnitude,  $v$  the excitation frequency. We let  $p_o = 1$ ,  $m_1 = 1/2$ ,  $m_2 = 1/4$ ,  $k_1 = 1$ , and  $k_2 = 1/2$ . Then, the natural frequencies of the first and second nodes are  $\omega_1 = 1$  and  $\omega_2 = 2$ . We concentrate on the displacement of the second node, which in steady-state is  $u_2(t) = u_{2o} \sin vt$ , with

$$u_{2o} = \frac{1}{(1 - v^2)(1 - v^2/4)}. \quad (4)$$

Suppose that  $v$  is unknown and modeled by a random variable  $V$  distributed by a mixture of beta densities. Specifically, with probability  $1/3$ ,  $V$  follows a beta(2,2) distribution on  $[0, 1]$ , with probability  $1/3$ ,  $V$  follows a beta(2,2) distribution on  $[1, 2]$ , and with probability  $1/3$ ,  $V$  follows a beta(2,2) distribution on  $[2, 3]$ . We seek to estimate the density of  $u_{2o}$ . Of course, the distribution of  $u_{2o}$  is therefore also a mixture and a prior insight about this may help the construction of exponential epi-spline estimates. However, we do not pursue that here. While one could attempt UQ by means of polynomial expansion, the choice of polynomial is nontrivial as the mixture of beta distributions is not a standard Askey distribution. More importantly, further difficulties arise since (4) is discontinuous at the natural frequencies 1 and 2.

Using a sample of size 100, we obtain the exponential epi-spline estimate in Figure 3; see the solid curve. Here we include the soft information that the density is unimodal on the upper 30% of the support

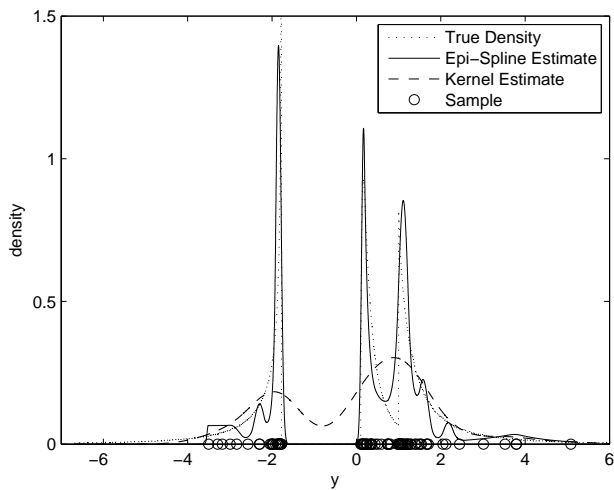


Figure 3: Density of second-node displacement magnitude in Example 2.

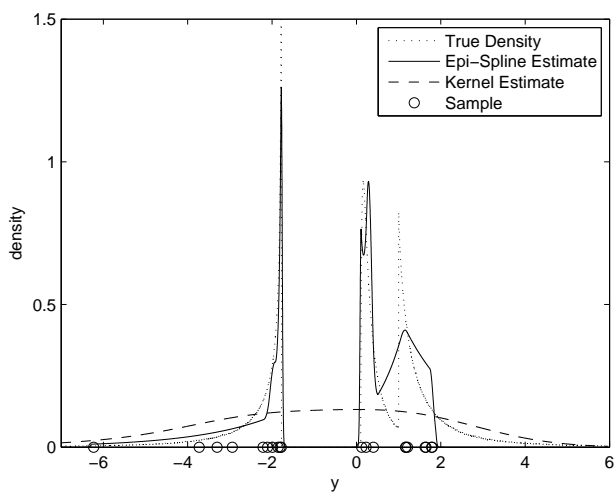


Figure 4: Density of second-node displacement magnitude in Example 2 under gradient information.

and likewise for the lower 30%. For comparison, we use a sample of size  $10^7$  to obtain a highly accurate estimate of the true density, which is illustrated by the dotted curve. As can be expected from the form of  $g$ , the true density is highly varying with a discontinuity at approximately  $-1.78$ . Even in the presence of the incorrect soft information about continuous differentiability, we find that the exponential epi-spline estimate captures the essence of the variability in  $u_{20}$ . In comparison, the kernel estimate (see the dashed curve in Figure 3) poorly represents the true density.

Figure 4 presents similar results but for a smaller sample of only 20 draws, but now with also gradient information at those 20 points utilized in (2). We also extend the unimodality constraint to 40% to achieve less volatile tails. The gradient information requires additional flexibility in the epi-spline and we increase the number of partitions  $N$  to 100. We again see that the exponential epi-spline captures the essence of the true density even with the low sample size.

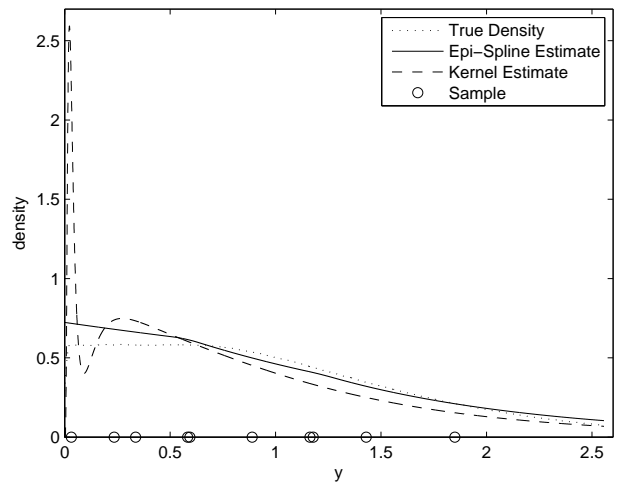


Figure 5: Density of Sobol function output in Example 3.

### 4.3 Example 3: Sobol function

In this example, we use a Sobol function

$$g(\mathbf{V}) = \prod_{j=1}^5 \frac{|4V_j - 2| + a_j}{1 + a_j}, \quad (5)$$

with  $V_j$  being independent and uniformly distributed on  $[0, 1]$  and  $a_j \geq 0$ . As in Eldred et al. (2011), we use  $a = (0, 1, 2, 4, 8)$ . Using 10 sample points, we obtain the exponential epi-spline estimate in Figure 5 (see the solid curve). We assume that the support is non-negative, the density is unimodal, and the pointwise Fisher information is in the interval  $[-1, 0]$ . We use  $N = 20$ . We see that our estimate follows the true density (estimated using a sample of size  $10^7$ ; see the dotted curve) quite well, especially in the upper tail. In contrast, the kernel estimate oscillates near zero. This is also a case where there are difficulties in obtaining accurate polynomial expansions as  $g$  is nonsmooth.

### 4.4 Example 4: Aeroshell Dynamics

This example is taken from Swiler et al. (2009) and deals with a bonding material in an aeroshell. A 3-D finite-element model gives the frequency in shear mode, which is our main concern. The frequency is uncertain due to unknown Young's modulus  $E$  and Poissons ratio  $\nu$  in a material component. Each evaluation of the finite-element model takes about 2 hours and few values of  $E$  and  $\nu$  can be examined. We consider a data set of 10 frequencies corresponding to various values of  $E$  and  $\nu$  (see Swiler et al. 2009) and aim to estimate the density of the frequency.

Figure 6 shows the estimated density (solid curve) given unimodality and support bounds  $[813, 2884]$  deduced from experience with the system. We use  $N = 20$ . The sample is illustrated by circles. We find that the exponential epi-spline estimate gives a reasonable representation of the uncertainty, while, in contrast, the kernel estimate (dashed curve) exhibits oscillations that ca not be expected for this system.

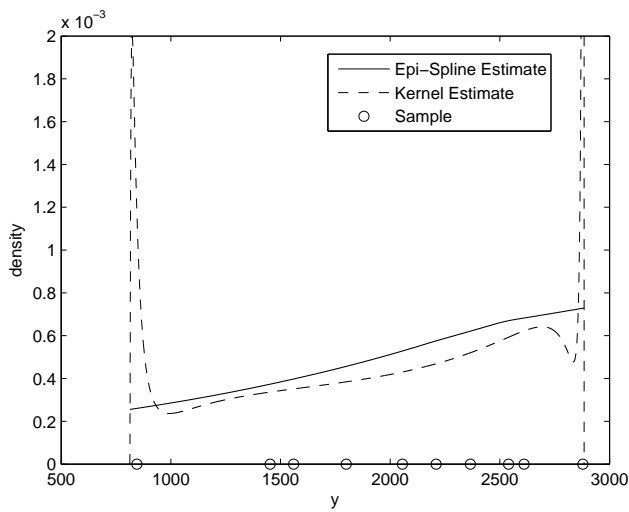


Figure 6: Density of shear mode frequency in Example 4.

## 5 CONCLUSIONS

We present a novel approach to UQ based on probability density estimation for an output quantity of interest. We are able to achieve good-quality estimates of such densities even for small data sets by including soft information that may be derived from engineering insight and judgement. With as little as 10-100 realizations of the output, we are able to identify the main features of the densities even for nonsmooth and discontinuous system function and high-dimensional inputs. The densities can lead to new understanding of the system uncertainty and provide means for estimating moments and resampling as may be needed in further analysis and optimization.

**Acknowledgements.** This material is based upon work supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant numbers 00101-80683, W911NF-10-1-0246 and W911NF-12-1-0273.

## REFERENCES

Babuska, I., R. Tempone, & G. Zouraris (2004). Galerkin finite elements approximation of stochastic finite elements. *SIAM J. Numerical Analysis* 42, 800–825.

Casey, M. & R. J.-B. Wets (2013). Density estimation: exploiting non-data information. Working paper, University of California, Davis, CA.

Chopra, A. (1995). *Dynamics of Structures. Theory and Applications to Earthquake Engineering*. Mineola, New York: Prentice Hall.

de Montricher, G. M., R. Tapia, & J. Thompson (1975). Nonparametric maximum likelihood estimation of probability densities by penalty function method. *Annals of Statistics* 3, 1329–1348.

Dong, M. X. & R. J.-B. Wets (2007). Estimating density functions: a constrained maximum likelihood approach. *Journal of Nonparametric Statistics* 12(4), 549–595.

Doostan, A. & G. Iaccarino (2009). A least-squares approximation of partial differential equations with high-dimensional random inputs. *J. Computational Physics* 228, 4332–4345.

Dupacova, J. (1992). Epi-consistency in restricted regression models - the case of a general convex fitting function. *Computational Statistics and Data Analysis* 14, 417–425.

Eldred, M. S., L. P. Swiler, & G. Tang (2011). Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation. *Reliability Engineering & System Safety* 96, 1092–1113.

Ganapathysubramanian, B. & N. Zabarar (2007). Sparse grid collocation schemes for stochastic natural convection problems. *J. Computational Physics* 225, 652–685.

Geyer, C. J. (1994). On the asymptotics of constrained maximum likelihood estimation. *The Annals of Statistics* 22, 1993–2010.

Ghanem, R. & P. Spanos (1991). *Stochastic Finite Elements: A Spectral Approach*. Mineola, New York: Dover.

Good, I. J. & R. A. Gaskin (1971). Nonparametric roughness penalties for probability densities. *Biometrika* 58, 255–277.

Groenenboom, P., G. Jongbloed, & J. Wellner (2001). Estimation of a convex function: characterizations and asymptotic theory. *Annals of Statistics* 29, 1653–1698.

Helton, J. & M. Pilch (2011). Quantification of margins and uncertainties. *Reliability Engineering & System Safety* 96, 959–964.

Kleiber, M., T. Hien, & D. Tran (1992). *The Stochastic Finite Element Method: Basic Perturbation Technique and Computer Implementation*. Chichester, UK: Wiley.

Klonias, V. K. (1982). Consistency of two nonparametric maximum penalized likelihood estimators of the probability density function. *Annals of Statistics* 10, 811–824.

Leonard, T. (1978). Density estimation, stochastic processes and prior information. *Journal of the Royal Statistical Society B40*, 113–146.

Nobile, F., R. Tempone, & C. Webster (2008). A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numerical Analysis* 46, 2309–2345.

Royset, J. O. & R. J.-B. Wets (2013). Nonparametric density estimation with soft information using exponential epi-splines. Working paper, Naval Postgraduate School, Monterey, CA.

Samaniego, F. J. & D. M. Reneau (1994). Towards a reconciliation of the Bayesian and frequentist approaches to point estimation. *Journal of the American Statistical Association* 89, 947–957.

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Annals of Statistics* 10, 795–810.

Swiler, L., T. Paez, & R. Mayes (2009). Epistemic uncertainty quantification tutorial. In *Proceedings of the IMAC-XXVII*. Society for Experimental Mechanics Inc.

Thompson, J. R. & R. A. Tapia (1990). *Nonparametric Function Estimation, Modeling, and Simulation*. Philadelphia, PA: SIAM Publishers.

Van de Geer, S. (1987). A new approach to least squares estimation. Center for Mathematics and Computer Science, Amsterdam.

Wahba, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Annals of Statistics* 9, 146–156.

Wets, R. J.-B. (1991). Constrained estimation: consistency and asymptotics. *Applied Stochastic Models and Data Analysis* 7, 17–32.

Xiu, D. & G. Karniadakis (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Scientific Computing* 24, 619–644.

Xiu, D. B. & J. S. Hesthaven (2005). High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing* 27(3), 1118–1139.