1

Introduction

Why adopt the nonparametric Bayesian approach for inference? The answer lies in the simultaneous preference for nonparametric modeling and desire to follow a Bayesian procedure. Nonparametric (and semiparametric) models can allow one to avoid the arbitrary and possibly unverifiable assumptions inherent in parametric models. Bayesian procedures may be desirable because of the conceptual simplicity of the Bayesian paradigm, philosophical reasons.

1.1 Motivation

Bayesian nonparametrics concerns Bayesian inference methods for nonparametric and semiparametric models. In the Bayesian nonparametric paradigm, a prior distribution is assigned to all relevant unknown quantities, whether finite or infinite dimensional. The *posterior distribution* is the conditional distribution of these quantities, given the data, and is the basis for all inference. This is the same as in any Bayesian inference, except that the unknown quantities or parameters may be infinite dimensional. A *model* completely specifies the conditional distribution of all observed, given all unobserved quantities, or parameters, while a *prior distribution* specifies the distribution of all unobservables. The posterior distribution involves an inversion of the order of conditioning and gives the distribution of the unobservables, given the observables. Existence of a regular version of the posterior distribution is guaranteed under mild conditions on the relevant spaces (see Section 1.3). From the Bayesian point of view, random effects and latent variables are unobservables and are treated in the same way as the unknown parameters used to describe the model. Distributions of these quantities, often considered as part of the model itself from the classical point of view, are part of the prior.

1.1.1 Classical versus Bayesian Nonparametrics

Nonparametric and semiparametric statistical models are increasingly replacing parametric models as a way to gain the flexibility necessary to address a wide variety of data. A nonparametric or semiparametric model involves at least one infinite-dimensional parameter and hence may also be referred to as an "infinite-dimensional model." Indeed, the nomenclature "nonparametric" is misleading in that it gives the impression that there are no parameters in the model, while in reality there are infinitely many unknown quantities. However, the term nonparametric is so popular that it makes little sense not to use it. The infinite-dimensional parameter is usually a function or measure. In the canonical example of

a nonparametric model, the data are a random sample from a completely unknown distribution or density. More generally, models may be structured and data non-i.i.d., and functions of interest include the cumulative distribution function, density function, regression function, hazard rate, transition density of a Markov process, spectral density of a time series, response probability of a binary variable as a function of covariates, false discovery rate as a function of nominal level in multiple testing and receiver operating characteristic function between distributions. Non-Bayesian methods for the estimation of many of these functions are well understood, and widely accepted statistical procedures are available. Bayesian estimation methods for nonparametric problems started receiving attention over the past four decades.

Different people may like Bayesian methods for different reasons. To some the appeal is philosophical. Certain *axioms of rational behavior* lead to the conclusion that one ought to follow a Bayesian approach in order not to be irrational (see Bernardo and Smith 1994). Although the axioms themselves can be questioned, the impression that Bayesian methods are logically more consistent than non-Bayesian methods is widespread. In particular, expressing uncertainty in probabilities is more satisfying than using criteria that involve integration over the sample space – that is, bother about samples that could have, but have not, realized. Others justify the Bayesian paradigm by appealing to exchangeability and de Finetti's theorem (de Finetti 1937). This celebrated theorem concludes the existence of a "random parameter" instead of a "fixed parameter" based on a "concrete" set of observations and a relatively weak assumption on distributional invariance (see Schervish 1995). However, this argument leads to subjective specification of a prior, which is regarded as difficult. Decision theorists may be Bayesians because of the complete class theorem, which asserts that for any procedure there is a better Bayesian procedure, and that only the latter procedures are admissible, or essentially so (see Ferguson 1967). While this could be a strong reason for a frequentist to take the Bayesian route, there are difficulties in that the complete class theorem holds only when the parameter space is compact (and the loss function is convex) and that the argument does not say which prior to choose from among the class of all priors. People who believe in asymptotic theory might find Bayesian methods attractive for their large-sample optimality. However, many non-Bayesian procedures (most notably, the maximum likelihood estimator, or MLE) are also asymptotically optimal, hence the argument is not compelling.

Although the specification of a prior distribution may be challenging, the Bayesian approach is extremely straightforward, in principle – the full inference is based on the posterior distribution only. All inference tools are produced in one stroke, and one need not start afresh when the focus of attention changes from one quantity to another. In particular, the same analysis produces an estimate as well as an assessment of its accuracy (in terms of variability or a probable interval for the location of the parameter value). The Bayesian approach produces a "real probability" on the unknown parameter as a quantification of the uncertainty about its value, which may be used to construct a "credible interval" or test with a clear interpretation. The Bayesian approach also eliminates the problem of nuisance parameters by integrating them out, while classical procedures must often find ingenious ways to tackle them separately for each inference problem. Finally, prediction problems, which are often the primary objective of statistical analysis, are solved most naturally if one follows the Bayesian approach.

These conveniences come at a price, however. The Bayesian principle is also restrictive in nature, allowing no freedom beyond the choice of prior. This limitation can put Bayesian methods at a disadvantage vis-à-vis non-Bayesian methods, particularly when performance is evaluated by frequentist principles. For instance, even if only a part of the unknown parameter is of interest, a Bayesian must still specify a prior on the whole parameter, compute the posterior distribution, and integrate out the irrelevant part, whereas a classical procedure may be able to target the part of interest. Another problem is that no corrective measures are allowed in a Bayesian framework once the prior has been specified. In contrast, the MLE is known to be nonexistent or inconsistent in many infinite-dimensional problems, such as density estimation, but it can be modified by penalization, sieves (see Grenander 1981), partial likelihood (Cox 1972) or other devices (Murphy and van der Vaart 2000). An honest Bayesian cannot change the likelihood, change the prior by looking at the data or even change the prior with increasing sample size.

1.1.2 Parametric versus Nonparametric Bayes

Parametric models make restrictive assumptions about the data-generating mechanism, which may cause serious bias in inference. In the Bayesian framework, a parametric model assumption can be viewed as an extremely strong prior opinion. Indeed, a parametric model specification $X|\,\theta \sim p_\theta$, for $\theta \in \Theta \subset \mathbb{R}^d$, with a prior specification $\theta \sim \pi$, may be written as $X|\,p \sim p$ for $p \sim \Pi$ and a prior distribution Π on the set of all possible densities with the property that $\Pi(\{p_\theta : \theta \in \Theta\}) = 1$. Thus parametric modeling is equivalent to insisting on a prior that assigns probability one to a thin subset of all densities. This is a very strong prior opinion indeed, which is replaced by an open-minded view when following the nonparametric Bayesian approach.

To some extent, the nonparametric Bayesian approach also solves the problem of partial specification. Often a model is specified incompletely, without describing every detail of the data-generating mechanism. A familiar example is the Gauss-Markov setup of a linear model, where errors are assumed to be uncorrelated, mean-zero variables with constant variance, but no further distributional assumptions are imposed. Lacking a likelihood, a parametric Bayesian approach cannot proceed further. However, a nonparametric Bayesian approach can use a prior on the space of densities with mean zero as a model for the error distribution. More generally, incomplete model assumptions may be complemented by general assumptions involving infinite-dimensional parameters in order to build a complete model, which a nonparametric Bayesian approach can equip with infinite-dimensional priors.

1.2 Challenges of Bayesian Nonparametrics

This section describes some conceptual and practical difficulties that arise in Bayesian nonparametrics, along with possible remedies.

1.2.1 Prior Construction

A Bayesian analysis cannot proceed without a prior distribution on all parameters. A prior ideally expresses a quantification of knowledge from past experience and subjective feelings.

A prior on a function requires knowledge of many aspects of the function, including infinitesimal details, and the ability to quantify the information in the form of a probability measure. This poses an apparent conceptual contradiction: a nonparametric Bayesian approach is pursued to minimize restrictive parametric assumptions, but at the same time it requires specification of the minute details of a prior distribution for an infinite-dimensional parameter.

There seems to be overall agreement that subjective specification of a prior cannot be expected in complex statistical problems. Instead, inference must be based on a *default prior*. This is vaguely understood as a prior distribution that is proposed by some automatic mechanism that is not biased toward any particular parameter values and has low information content compared to the data.

Some of the earliest statistical analyses in history used the idea of *inverse probability* and came down to a default Bayesian analysis with respect to a uniform prior. Later uniform priors were strongly criticized for lacking invariance, which led to a decline in the popularity of Bayesian analysis until more invariance-friendly methods such as *reference analysis* or *probability matching* emerged. However, most of these ideas are restricted to finite-dimensional parametric problems.

A default prior need not be noninformative, but should be spread over the whole parameter space. Some key hyperparameters regulating the prior may be chosen by the user, whereas other details must be constructed by the default mechanism. Unlike in parametric situations, where noninformative priors are often improper, default priors considered in nonparametric Bayesian inference are almost always proper. Large support of the prior means that the prior is not too concentrated in some particular region. This situation generally ensures that the information contained in the prior is subdued gradually by the data if the sample size increases, so that eventually the data override the prior.

The following chapters discuss methods of prior construction for various problems of interest. Although a default prior is not unique in any sense, it is expected that over the years, based on theoretical results and practical experience, a handful of suitable priors will be short-listed and cataloged for consensus use in each inference problem.

1.2.2 Computation

The property of conjugacy played an important role in parametric Bayesian analysis, as it enabled the derivation of posterior distributions at a time when computing resources were lacking. Later sampling-based methods such as the Metropolis-Hastings algorithm and Gibbs sampling gave Bayesian analysis a tremendous boost. Without modern computing, nonparametric Bayesian analysis would hardly be practical.

However, we cannot simulate directly from the posterior distribution of a function unless it is parameterized by finitely many parameters. The function of interest must be broken up into more elementary finite-dimensional objects whose posterior distributions are accessible. For this reason, the structure of the prior is important. Useful structure may come from conjugacy or approximation. Often, a computational method combines analytic derivation and Markov chain Monte Carlo (MCMC) algorithms, and is based on innovative ideas. For instance, density estimation with a Dirichlet mixture prior, discussed in Chapter 4, uses an equivalent hierarchical mixture model involving a latent variable for each observation and

integrates out the infinite-dimensional parameter, given the latent variables (see 5.3). Thus the problem of infinite dimension has been reduced to one of finite dimension. In another instance, in a binary response model with a Gaussian process prior, introducing normal latent variables brings in conjugacy (see Section 11.7.3).

1.2.3 Asymptotic Behavior

Putting a prior on a large parameter space makes it easy to be grossly wrong. Therefore "robustness" is important in Bayesian nonparametrics: the choice of prior should not influence the posterior distribution too much. This is difficult to study in a general framework. A more manageable task is the study of asymptotic properties of posterior distributions, as the information in the data increases indefinitely. For example, "posterior consistency" may be considered an asymptotic form of robustness. Loosely speaking, posterior consistency means that the posterior probability eventually concentrates in a (any) small neighborhood of the actual value of the parameter. This is a weak property shared by many prior distributions. Finer properties, such as the rate of contraction or a (functional) limit theorem, give more insight into the performance of different priors.

The study of asymptotic properties is more complex in the nonparametric than in the parametric context. In the parametric setting, good properties are guaranteed under mild conditions. Under some basic regularity conditions, it suffices that the true value of the parameter belongs to the support of the prior. In the infinite-dimensional context this is not enough. Consistency may fail for natural priors satisfying the support condition, meaning that even an infinite amount of data may not overcome the pull of a prior in a wrong direction. Consistent priors may differ strongly in accuracy, depending on their fine details, as can be measured through their rates of contraction. Unlike in the parametric setting, many priors do not "wash out," as the information in the data increases indefinitely.

Thus it makes sense to impose posterior consistency and a good rate of contraction as requirements on a "default prior." Several chapters in this book are devoted to the study of asymptotic behavior of the posterior distribution and other related quantities. Chapter 10 is devoted to combining priors hierarchically into an overall prior so as to make the posterior "adapt" to a large class of underlying true parameters.

1.3 Priors, Posteriors and Bayes's Rule

In the Bayesian framework, the data X follows a distribution determined by a parameter θ, which is itself considered to be generated from a *prior distribution* Π. The corresponding *posterior distribution* is the conditional distribution of θ, given X. This framework is identical in parametric and nonparametric Bayesian statistics, the only difference being the dimension of the parameter. Because the proper definitions of priors and (conditional) distributions require (more) care in the nonparametric case, we review the basics of conditioning and Bayes's rule in this section.

If the parameter set Θ is equipped with a σ-field $\mathscr{B}$, the prior distribution Π is a probability measure on the measurable space $(\Theta, \mathscr{B})$, and the distribution P_θ of X, given θ, is

a regular conditional distribution on the sample space $(\mathfrak{X}, \mathscr{X})$ of the data,[1] then the pair (X, θ) has a well-defined joint distribution on the product space $(\mathfrak{X} \times \Theta, \mathscr{X} \otimes \mathscr{B})$, given by

$$\mathrm{P}(X \in A, \theta \in B) = \int_B P_\theta(A)\, d\Pi(\theta).$$

This gives rise to the *marginal distribution* of X, defined by

$$\mathrm{P}(X \in A) = \int P_\theta(A)\, d\Pi(\theta), \qquad A \in \mathscr{X},$$

and the conditional distribution, called *posterior distribution*, given by

$$\Pi(B|\, X) = \mathrm{P}(\theta \in B|\, X), \qquad B \in \mathscr{B}.$$

By Kolmogorov's definition, the latter conditional probabilities are always well defined, for every given $B \in \mathscr{B}$, as measurable functions of X such that $\mathrm{E}[\Pi(B|\, X)\mathbb{1}\{A\}(X)] = \mathrm{P}(X \in A, \theta \in B)$, for every $A \in \mathscr{X}$. If the measurable space $(\Theta, \mathscr{B})$ is not too big, then there also exists a *regular version* of the conditional distribution: a Markov kernel from $(\mathfrak{X}, \mathscr{X})$ into $(\Theta, \mathscr{B})$. We shall consider the existence of a regular version to be necessary in order to speak of a true posterior distribution. A sufficient condition is that Θ is a Polish space[2] and $\mathscr{B}$ its Borel σ-field.[3]

Even though the posterior distribution can usually be thus defined, some further care may be needed. It is inherent in the definition that the conditional probabilities $\mathrm{P}(\theta \in B|\, X)$ are unique only up to null sets under the marginal distribution of X. Using a regular version (on a standard Borel space) limits these null sets to a single null set that works for every measurable set B, but does not eliminate them altogether. This is hardly a concern if the full Bayesian setup is adopted, as this defines the marginal distribution of X as the appropriate data distribution. However, if the Bayesian framework is viewed as a method for inference only and it is allowed that the data X is generated according to some "true" distribution P_0 different from the marginal distribution of X in the Bayesian setup, then the exceptional "null sets" may well have nonzero mass under this "true" distribution, and it is impossible to speak of *the* posterior distribution. (To distinguish P_0 from the marginal distribution of X, we shall refer to the latter sometimes as the "Bayesian marginal distribution.")

Obviously, this indefiniteness can only happen under serious "misspecification" of the prior. In particular, no problem arises if

$$P_0 \ll \int P_\theta\, d\Pi(\theta),$$

which is guaranteed for instance if P_0 is dominated by P_θ for θ in a set of positive prior probability. In parametric situations the latter condition is very reasonable, but the nonparametric case can be more subtle, particularly if the set of all P_θ is not dominated. Then there may be a "natural" way of defining the posterior distribution consistently for all X, but it must

[1] I.e. a *Markov kernel* from $(\Theta, \mathscr{B})$ into $(\mathfrak{X}, \mathscr{X})$: the map $A \mapsto P_\theta(A)$ is a probability measure for every $\theta \in \Theta$ and the map $\theta \mapsto P_\theta(A)$ is measurable for every $A \in \mathscr{X}$.

[2] A topological space that is a complete separable metric space relative to some metric that generates the topology.

[3] More generally, it is sufficient that $(\Theta, \mathscr{B})$ is a *standard Borel space*: a measurable space admitting a bijective, bimeasurable correspondence with a Borel subset of a Polish space.

be kept in mind that this is not dictated by Bayes's rule alone. An important example of this situation arises with the nonparametric Dirichlet prior (see Chapter 4), where the marginal distribution may or may not dominate the distribution of the data.

For a dominated collection of measures P_θ, it is generally possible to select densities p_θ relative to some σ-finite dominating measure μ such that the map $(x, \theta) \mapsto p_\theta(x)$ is jointly measurable. Then a version of the posterior distribution is given by *Bayes's formula*

$$\Pi(B \mid X) = \frac{\int_B p_\theta(X)\, d\Pi(\theta)}{\int p_\theta(X)\, d\Pi(\theta)}. \tag{1.1}$$

Of course, this expression is defined only if the denominator $\int p_\theta(X)\, d\Pi(\theta)$, which is the (Bayesian) *marginal density* of X, is positive. Definitional problems arise (only) if this is *not* the case under the true distribution of the data. Incidentally, the formula also shows that a Polish assumption on $(\Theta, \mathscr{B})$ is sufficient, but not necessary, for existence of the posterior distribution: (1.1) defines a Markov kernel as soon as it is well defined.

In a vague sense, the *support* of a measure is a smallest set that contains all its mass. A precise definition is possible only under assumptions on the measurable space. We limit ourselves to Polish spaces, for which the following definition of support can be shown to be well posed.

Definition 1.1 (Support) The *support* of a probability measure on the Borel sets of a Polish space is the smallest closed set of probability one. Equivalently, it is the set of all elements of the space for which every open neighborhood has positive probability.

It is clear that a posterior distribution will not recover a "nonparametric" set of true distributions unless the prior has a large support. In Chapters 6 and 8 this requirement will be made precise in terms of posterior consistency (at a rate), which of course depends both on the prior and on the way the data distribution P_θ depends on the parameter θ. As preparation, when discussing priors in the following chapters, we pay special attention to their supports.

1.3.1 Absolute Continuity

Bayes's formula (1.1) is available if the model $(P_\theta : \theta \in \Theta)$ is dominated. This is common in parametric modeling, but may fail naturally in nonparametric situations. As a consequence, sometimes we perform Bayesian analysis without Bayes. Mathematically, this is connected to absolute continuity of prior and posterior distributions.

It seems natural that a prior distribution supported on a certain set yields a posterior distribution supported inside the same set. Indeed, the equality $\Pi(B) = \mathrm{E}\mathrm{P}(\theta \in B \mid X)$ immediately gives the implication: if $\Pi(B) = 1$, then $\mathrm{P}(\theta \in B \mid X) = 1$, almost surely. The exceptional null set is again relative to the marginal distribution of X, and it may depend on the set B. The latter dependence can be quite serious. In particular, the valid complementary implication: if $\Pi(B) = 0$, then $\mathrm{P}(\theta \in B \mid X) = 0$ almost surely, should not be taken as proof that the posterior is always absolutely continuous with respect to the prior. The nonparametric Dirichlet prior exemplifies this, as the posterior is typically orthogonal to the prior (see Section 4.3.4).

Such issues do not arise in the case that the collection of distributions P_θ is dominated. Formula (1.1) immediately shows that the posterior is absolutely continuous relative to the prior in this case (where it is assumed that the formula is well posed). This can also be reversed. In the following lemma we assume that the posterior distribution is a regular conditional distribution, whence it is unique up to a null set.

Lemma 1.2 *If both $(\mathfrak{X}, \mathscr{X})$ and $(\Theta, \mathscr{B})$ are standard Borel spaces, then the set of posterior distributions $\mathrm{P}(\theta \in B \mid X = x)$, where $x \in \mathfrak{X}_0$ for a measurable set $\mathfrak{X}_0 \subset \mathfrak{X}$ of marginal probability one, is dominated by a σ-finite measure if and only if the collection $\{P_\theta : \theta \in \Theta_0\}$ is dominated by a σ-finite measure, for some measurable set $\Theta_0 \subset \Theta$ with $\Pi(\Theta_0) = 1$. In this case, the posterior distributions are dominated by the prior.*

Proof A collection of probability measures $\{P_\theta : \theta \in \Theta\}$ on a standard Borel space is dominated if and only if it is separable relative to the total variation distance, and in this case the measures permit densities $x \mapsto p_\theta(x)$ that are jointly measurable in (x, θ) (e.g. Strasser 1985, Lemmas 4.6 and 4.1). Formula (1.1) then gives a version of the posterior distribution, which is dominated by the prior. Any other version differs from this version by at most a null set $\mathfrak{X}_0^c$.

The converse follows by interchanging the roles of x and θ. If the set of posterior distributions is dominated by a σ-finite measure, then they can be represented by conditional densities $\pi(\theta \mid x)$ relative to the dominating measure, measurable in (x, θ), and we can reconstruct a regular version of the conditional distribution of x given θ by (1.1) with the roles of θ and x interchanged, which is dominated. By assumption, the original distributions P_θ give another regular version of this conditional distribution and hence agree with the dominated version on a set of probability one. □

1.4 Historical Notes

Laplace and Bayes pioneered the idea of inverse probability. Fisher was a strong critic of this approach because of invariance-related paradoxes. Jeffreys revived the idea of inverse probability by replacing the uniform prior with his famous prior, the square root of the determinant of Fisher's information. The theory of objective priors has now evolved extensively; see Bernardo and Smith (1994) for different approaches and references. A Bayesian nonparametric idea seems to have been first used by Poincaré for numerical interpolation. He considered the unknown function as a random series and function values as observed up to measurement error. He computed Bayes estimates by assigning priors on the coefficients in the expansion. Freedman (1963) considered tail-free priors, a general class that can avoid inconsistency problems. Computational issues were still not considered very seriously. A breakthrough came in the seminal paper in which Ferguson (1973) introduced the Dirichlet process, which initiated modern Bayesian nonparametrics. The Dirichlet process has only two easily interpretable hyperparameters and leads to an analytically tractable posterior distribution. This advance solved the problem of estimating a cumulative distribution by a Bayesian method and reduced the gap with the classical approach, which already offered solutions to density estimation, nonparametric regression and other curve estimation problems using kernel and other smoothing methods. Within a decade, a Bayesian solution to

density estimation by Dirichlet mixture processes was available, although computational challenges remained daunting. A little later, Gaussian processes and Pólya tree processes were proposed as alternative solutions. The early nineties also witnessed the development of priors for survival analysis, including the beta process prior of Hjort (1990) and independent increment processes. The first half of the nineties saw the astonishing development of computational methods for Bayesian nonparametrics, ostensibly because of the advent of MCMC techniques and the availability of fast computers. Major initial contributions came from Escobar, West, MacEachern, Müller, Neal and others, fueled by the progress on MCMC ideas in general by Gelfand, Smith, Green, Chib and the development of the *WinBUGS* software by Gilks et al. (1994). Recently, the R-package *DP package* has been developed to solve Bayesian nonparametric computational problems; see Jara (2007). The earliest results on posterior consistency for nonparametric problems were obtained by Doob (1949), but those results give no clue about consistency at a given true value of the parameter. Freedman (1963, 1965) constructed examples of posterior inconsistency and showed, in general, that a posterior distribution can asymptotically misbehave for most priors (here, "most" is used in a topological sense). Freedman also introduced the concept of tail-freeness, a key property that can eliminate inconsistency problems. The most significant early result on consistency was due to Schwartz (1965), whose ideas led to a reasonably complete general theory of consistency. Schwartz showed that the support of the prior in a Kullback-Leibler sense is the key factor in determining consistency, together with a restriction on the size of the model imposed through a testing condition. The original Schwartz theorem is, however, incapable of yielding consistency for density estimation. Subsequent ideas were developed by Barron, Ghosh, Ghosal, Wasserman, Ramamoorthi, Walker and others. Diaconis and Freedman (1986b,a) and Doss (1985a,b) pointed out serious problems of inconsistency with the Dirichlet process prior if used for the error distribution in the location problem. This is an important phenomenon, as such a prior was considered natural for the location problem in the eighties. It emphasized that posterior consistency cannot be taken lightly in nonparametric problems. A more striking example of posterior inconsistency was recently constructed by Kim and Lee (2001) in the context of survival analysis. The study of rates of contraction in a general framework was started at the beginning of the previous decade by Ghosal, van der Vaart, Shen, Wasserman and others and a relatively complete theory is now available, even extending to observations that may not be independent or identically distributed. Rates of contraction have been computed for several priors such as Dirichlet process mixtures, priors based on splines and Gaussian processes. Results on rates of contraction that adapt to the underlying function class and related problems about model selection have been studied recently. The *Bernstein–von Mises theorem* for regular parametric models implies that the posterior distribution of the parameter centered at the MLE converges to the same normal distribution as that of the limit of the normalized MLE. Thus, asymptotically, Bayesian and sampling probabilities agree, so confidence regions of approximate frequentist validity may be generated from the posterior distribution. Cox, Freedman and others showed that such a result should not be expected for curve estimation problems. Some positive results have been obtained by Lo, Kim, Lee, Shen, Castillo, Nickl, Leahu, Bickel, Kleijn and others. The study of nonparametric Bayesian uncertainty quantification for curve estimation problems was started only recently, with first results and discussion by Szabó et al. (2015).