Statistical Science
2019, Vol. 34, No. 1, 34–37
https://doi.org/10.1214/18-STS677
Main article: https://doi.org/10.1214/18-STS660

Comment on "Probabilistic Integration: A Role in Statistical Computation?"

Michael L. Stein and Ying Hung

We commend the authors for their serious effort to address some of the mathematical, conceptual and practical issues that arise when using probabilistic methods for evaluating integrals of deterministic functions and assessing the uncertainties in these methods. Applying Gaussian process models to deterministic but difficult to compute functions is, as this paper emphasizes, gaining increasing attention in the numerical analysis literature, but it has been actively pursued in the computer experiment literature since at least the seminal paper Sacks et al. (1989). In computer experiments, the interest is usually in interpolation or optimization of some complex deterministic function, rather than integration, but many of the issues raised in this work are also pertinent when interpolating. We would also point to Sacks and Ylvisaker (1970) as an early work that considers theoretical design issues when integrating Gaussian processes, although from the standpoint of assuming the unknown function really is a Gaussian process model with a known covariance structure.

The present paper takes the point of view, common in the approximation theory literature, that the unknown function lies in some specified RKHS. It then exploits the fact that this reproducing kernel can be viewed as the covariance function for a Gaussian process, making it possible to make Bayesian inferences based on this model. The problem with this approach, which has long been known but can still lead to confusion, is that if f is a realization of a Gaussian process with covariance function k, then its realizations are insufficiently smooth to be elements of the RKHS. This paper provides some theory and a number of examples showing that, despite this fundamental problem, the Bayesian inferences we get from the Gaussian process model may provide good point estimates and useful uncertainty assessments for the integral of f over some domain.

Michael Stein is Professor of Statistics, University of Chicago, Chicago, Illinois, USA (e-mail: stein@galton.uchicago.edu). Ying Hung is Assistant Professor of Statistics, Rutgers University, New Brunswick, New Jersey, USA (e-mail: yhung@stat.rutgers.edu).

The theory in this work focuses on posterior contraction to the true value of the integral. As the authors make clear, this is a very different notion than the posterior distribution providing accurate probability statements. Indeed, the paper also notes that there is no theory, asymptotic or otherwise, to support a claim that the posterior probability statements coming from Bayes theorem will have a valid probability interpretation when f is an element of the RKHS.

To clarify the issues, let us consider a simple example. Suppose we wish to integrate a function f on $[0, 1]$ for which it is known $f(0) = f(1) = 0$. Brownian bridge on $[0, 1]$ provides a Gaussian process model whose realizations satisfy this condition; its mean is 0 and its covariance function on $[0, 1] \times [0, 1]$ is $k(x, y) = \sigma^2(\min(x, y) - xy)$ for some $\sigma > 0$. For a Gaussian process Z with this covariance function observed at j/n for $j = 0, \dots, n$, the optimal predictor of $\int_0^1 Z(x)\,dx$ is the trapezoidal rule and its RMSE can be shown to be $\sigma/(\sqrt{12}n)$. Realizations of this Gaussian process are nowhere differentiable with probability 1. In contrast, the elements of the RKHS with this kernel, which we can call $\mathcal{H}(k)$, are functions f that satisfy $f(0) = f(1) = 0$ and are absolutely continuous with an almost everywhere first derivative that is square integrable.

It is exceedingly implausible that a likelihood function or a solution of a deterministic differential equation of practical interest would be nowhere differentiable, so one should use caution when interpreting posterior probability statements based on this model. Lack of differentiability in these types of functions, when it occurs, tends to occur along lower-dimensional manifolds. For example, for $0 < s < t < 1$, suppose $f(x)$ equals 1 for $s < x < t$ and is 0 otherwise. This function shares some properties with Brownian bridge. First,

$$\lim_{n\to\infty} \sum_{j=1}^{n} \left\{ Z\left(\frac{j}{n}\right) - Z\left(\frac{j-1}{n}\right) \right\}^2 = \sigma^2 \tag{0.1}$$