Bayesian Analysis (2016) **11**, Number 4, pp. 1275–1277

Comment on Article by Chkrebtii, Campbell, Calderhead, and Girolami*

Sarat C. Dass[†,‡]

Discussion

The article presented by the authors addresses the issue of discretization uncertainty when differential equation systems are solved numerically, and particularly, its effect on the inference of unknown system parameters. I wish, firstly, to congratulate the authors for a very nice piece of work which, to my knowledge, is one of the first papers to highlight the issue and provide a Bayesian solution for it. Nevertheless, I do have several concerns with the article which I hope that the authors can help clarify and discuss.

On Discretization Uncertainty

What represents discretization uncertainty? The authors take the solution, u, of the differential equation system and its derivative, u_t, to be distributed according to a Gaussian process a priori. On discretizing over the ordered partition $\mathbf{s} = \{s_1, s_2, \cdots, s_N\}$, the discretization uncertainty is propagated as $f_n = f(s_n, u^{n-1}(s_n), \theta)$ where $u^{n-1}(s_n)$ is sampled from the marginal predictive prior $[u(s_n)|f_{n-1}, f_{n-2}, \cdots, f_1]$. This current f_n then is used to update the mean and covariance functions of the Gaussian process as in Algorithm 1 for the next step from $n \to n+1$. At step $n+1$, $u^n(s_{n+1})$ is sampled from $[u^n(s_{n+1})|f_n, f_{n-1}, \cdots, f_1]$, f_{n+1} is evaluated as $f_{n+1} = f(s_{n+1}, u^n(s_{n+1}), \theta)$, and the mean and covariance functions are again updated as in Algorithm 1. This process is repeated until $n = N$, the size of the discretization grid chosen. The posterior $[u(\mathbf{s}), u_t(\mathbf{s})|f_N, f_{N-1}, \cdots, f_1]$ is the discretization uncertainty, according to the authors, as shown in Figure 1.

But why should Algorithm 1 terminated at $n = N$ be interpreted as the final discretization uncertainty? That is, why should it be interpreted as the discretization uncertainty *associated with the grid* $\mathbf{s}$? One could potentially repeat the iterative process again for a second time, obtain a new sequence of f_ns, say $f_n^{(2)}$, $n = 1, 2, \cdots, N$, and update the mean and covariance functions as in Algorithm 1. To elaborate, I am assuming that $u^*(0)$ is fixed so $f_1^{(2)} = f(s_1, u^*(0), \theta)$ will not change, but based on this f_1, the mean and covariance functions can be updated again from the final posterior $[u(\mathbf{s}), u_t(\mathbf{s})|f_N, f_{N-1}, \cdots, f_1]$ from the first cycle. Then, obtain $u^{1,(2)}(s_2) \sim$

*Main article DOI: 10.1214/16-BA1017.

[†]Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia, saratcdass@gmail.com

[‡]This project is partially supported by an exploratory research grant scheme (ERGS 0153AB-i12) from the Government of Malaysia.

 DOI: 10.1214/16-BA1037

$[u^{1,(2)}(s_2)|f_N, \cdots, f_1, f_1^{(2)}]$ and update the mean and covariance functions again using Algorithm 1, and so on. As can be clearly seen, this updating scheme need not terminate and can be continued until infinity. Each time, a new Gaussian process arises with updated mean and covariance functions according to Algorithm 1. Note that $C_n^{(s)}(\mathbf{s}, \mathbf{s}) \leq C_m^{(r)}(\mathbf{s}, \mathbf{s})$, in terms of positive definiteness, where cycles r, s and integers $1 \leq m, n \leq N$ are such that either $r < s$ or if $r = s, m \leq n$. Two cases arise – either C_n converges to 0 or C_n converges to $C^* > 0$. Which is it? If this process converges, then, in my opinion, this limiting Gaussian process should be interpreted as the discretization uncertainty associated with the grid $\mathbf{s}$.

The other issue with discretization uncertainty is its quantification, via Gaussian processes, for numerical algorithms used to solve differential equations such as the Euler and Runge–Kutta methods; see Dass et al. (2016). Such numerical methods for solving differential equations consider a grid $\mathbf{s}$ over which an approximate solution of u may be obtained. Thus, one may consider, instead of the somewhat artificial interrogation process suggested by the authors, approximate values of $u(s_n)$ and $u_t(s_n)$, say $\tilde{u}(s_n)$ and $\tilde{u}_t(s_n)$, for $n = 1, 2, \cdots, N$ generated by the numerical method. What is the discretization uncertainty and how is it quantified, given the approximate values $\tilde{u}(\mathbf{s})$ and $\tilde{u}_t(\mathbf{s})$? Since Gaussian processes are involved, one may consider kriging as a natural way of interpolation and uncertainty generation for the grid $\mathbf{s}$. However, kriging has to be considered with noise, since otherwise, the interpolation at the grid sites will be exactly $\tilde{u}(\mathbf{s})$, which of course, is just an approximation of $u(\mathbf{s})$, with $\mathbf{0}$ uncertainty variance.

On Calibration

Acceptance probability of the Metropolis sampler: Could the authors provide details of the Metropolis sampler in Algorithm 2 with regard to the acceptance probabilities once candidate $(\theta', \alpha', \lambda')$ is generated from the proposal density? It would be likely that as the grid size, N, as well as the dimensions of u (and y) increases, the acceptance probabilities, ρ, should become smaller and smaller. In particular the first term of ρ can significantly deteriorate with large N and large dimension of u. For actual running of the Metropolis sampler, this may pose a problem with respect to mixing and convergence to target.

The bias and variance trade-off is clearly highlighted in this paper. When θ is the parameter of interest, clearly u can be viewed as a nuisance parameter whose estimation, nevertheless, is necessary for the estimation of θ. The authors show that when a biased estimate of u is used, for example, using a numerical solver with poor resolution, the resulting poor estimate of u propagates significant bias in the estimation of θ. The point of this paper is that the numerical solver can be enhanced by incorporating uncertainty quantification which results in an increased variance for θ, consequently enveloping its true value in the region of highest posterior density. This is illustrated in Figure 9. My questions relates back to the two issues that I highlighted in the uncertainty quantification section above. First, does the iterative procedure until infinity further improve (i.e., reduce) the variance of θ but still manages to maintain its true value in the highest posterior density region? And, when an actual numerical solver is used, such as

the Euler and the Runge–Kutta solvers, how would the uncertainty quantification be developed in a Gaussian process framework to give the bias variance trade-off by the current method?

I hope the authors will be able to provide their thoughts and insights into the matters as above. I would like again to congratulate them on an excellent piece of work and hope that the discussions will lead to further research developments in this new and exciting area.

References

Dass, S. C., Lee, J., Lee, K., and Park, J. (2016). "Laplace based approximate posterior inference for differential equation models." *Statistics and Computing*, 1–20. 1276