Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

BLENDING SPATIAL MODELING AND PROBABILISTIC BISECTION

Sergio Rodriguez

Department of Statistics and Applied Probability
University of California, Santa Barbara
Santa Barbara, CA 93106-3110, USA

ABSTRACT

Probabilistic Bisection Algorithms (PBA) pinpoint an unknown quantity by applying Bayesian updating to knowledge acquired from noisy oracle replies. We consider the generalized PBA setting (G-PBA) where the statistical distribution of the oracle is *unknown* and location-dependent, so that model inference and knowledge updating must be performed simultaneously. To this end, we propose to blend spatial modeling of oracle properties (namely regressing batched oracle responses on sampling locations) with the existing PBA information-directed sampling. The resulting sampling strategies account for the trade-off between inferring the latent oracle distribution versus reducing the uncertainty about the unknown point to be learned. We demonstrate that spatial modeling improves the original G-PBA schemes by applying our approach to root-finding of monotone noisy responses.

1 GENERALIZED PROBABILISTIC BISECTION

Introduction. Probabilistic Bisection Algorithms (PBA) sequentially learn about an unknown quantity $X^* \in [0,1]$ by querying an oracle $Y(x_n) \in \{-1,+1\}$ as to whether X^* lies rightwards or leftwards of location x_n. Due to statistical noise in the oracle responses, the oracle answers truthfully with probability $p(x_n)$. Starting with prior density g_0, and a history of sampling locations/responses $(x,y)_{1:n-1}$, the PBA builds a posterior distribution regarding X^*, $g_n(\cdot) := p_{X^*}(\cdot|y_{1:n-1}, x_{1:n-1})$. Towards the goal of generating efficient sampling strategies, Waeber et al. 2011 showed that when $p(x)$ is a known, x-independent constant, sampling at the posterior median is an optimal sampling policy which minimizes the expected posterior entropy of g_n. These conclusions no longer hold in the general and more realistic case when the distribution of oracle responses is *unknown* and *spatially varying*. Our goal is to develop generalized PBA sampling strategies that simultaneously learn $p(\cdot)$ and update a (surrogate) knowledge state f_n about X^*.

Approximating Knowledge States. In analogy to the PBA, knowledge updating is based on the Bayesian transition function:

$$\varphi_k(f_n(u), x, B_k(x); p(x)) \propto \begin{cases} \left[p(x)^{B_k(x)}(1-p(x))^{k-B_k(x)}\right] \cdot f_n(u) & \text{if } 0 \le x < u \le 1, \\ \\ \left[(1-p(x))^{B_k(x)}p(x)^{k-B_k(x)}\right] \cdot f_n(u) & \text{if } 0 \le u \le x \le 1; \end{cases} \tag{1}$$

where $k \ge 1$ is the *batch size* of querying the oracle k-times at the same x, and $B_k(x) := \sum_{j=1}^{k} \mathbb{1}\{y_j(x) = 1\}$. To implement (1), we seek an estimator for $p(x)$. In Rodriguez and Ludkovski (2016) this was achieved by aggregating the batched queries at x, for example through the *empirical majority proportion* $\bar{p}_k(x) := \max\{B_k(x)/k, 1 - B_k(x)/k\}$, and then plugging $\bar{p}_k(x)$ into (1).

Surrogate Gaussian Process Model for $p(\cdot)$**.** Although the above approach works well, it disregards information from previous queries at older design points. Consequently, we introduce a *spatial* model for $p(\cdot)$ that is constructed using the meta-responses $P_{1:n-1(k)} := (\bar{p}_k(x_1), \ldots, \bar{p}_k(x_{n-1}))$ conditional on $x_{1:n-1}$. Namely, we propose a GP regression model to estimate $p(\cdot)$ so that $P_{1:n-1(k)}|x_{1:n-1} \sim \mathsf{GP}(\mathbf{0}, \kappa(\cdot,\cdot))$;

where κ is a suitable covariance kernel (e.g., squared-exponential). Given the predictive posterior mean $\hat{p}_n(x) := \mathbb{E}[\bar{p}_k(x)|P_{1:n-1(k)}, x_{1:n-1}, x]$, the surrogate $x \mapsto \hat{p}_n(x)$ is then used for updating f_n via (1) at any x.

Sampling Policies. Under the G-PBA paradigm, the problem of how to select new samples may be addressed via two orthogonal approaches: *statistical emulation* (choose points to maximize learning of $p(\cdot)$, then infer X^* as the solution (in fact global minimum) of $p(X^*) = 1/2$) or *pure-PBA* (learning X^* by modeling its posterior distribution). We propose to blend these two schemes so as to focus on posterior of X^* while taking advantage of the spatial structure. Note that a key issue is not to sample *too close* to X^* whereby information gain goes to zero.

Our basic proposal given a total budget of N queries, is to start with an exploration-focused G-PBA as in Rodriguez and Ludkovski (2016) and then switch to more aggressive information-directed sampling (IDS) based on the GP surrogate $\hat{p}(\cdot)$. Specifically, during the first $n = 1, \dots M-1$ macro-steps sampling is done *systematically* via $x_{n+1} = F_n^{-1}(q_j)$ where $j = (n \bmod m) + 1$ and $q_{1:m} \in [0,1]$ are pre-specified quantiles to be sampled. Queries at each x are k-batched, and knowledge updating (1) uses the empirical proportion estimator $\bar{p}_k(x_{n+1})$: $f_{n+k} = \varphi_k(f_n, x_{n+1}, B_k(x_{n+1}); \bar{p}_k(x_{n+1}))$. After the first stage, a GP $\hat{p}$ is fit over $(x, P)_{1:M-1(k)}$. For the second stage $n = M, \dots, N-1$, x_{n+1} is chosen as the *Information Gain criterion* maximizer $x_{n+1} = \arg\max_{x \in [0,1]} I_n(x, f_n; \hat{p}_M(x))$, based on the GP posterior mean $\hat{p}_M(x)$. Further proposals use the GP *posterior variance* to adaptively switch between the above stages as data is acquired.

2 ROOT-FINDING OF MONOTONE NOISY RESPONSES

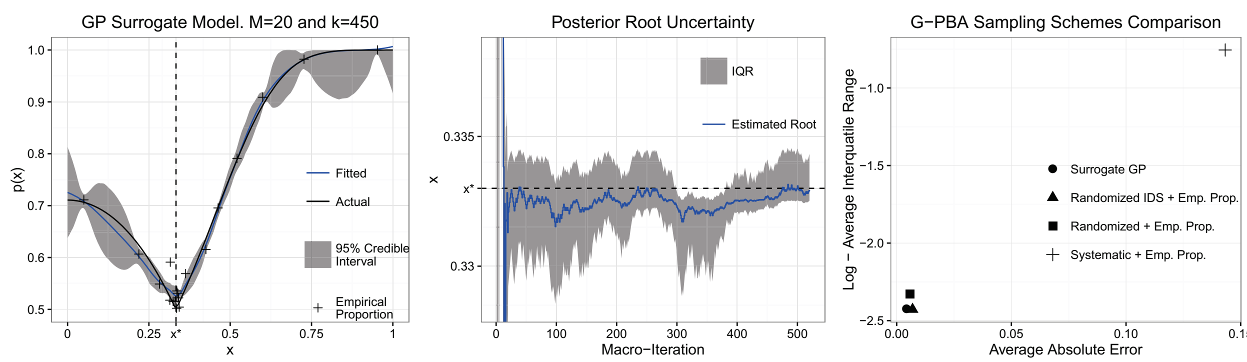

Figure 1: Surrogate spatial modeling of $p(\cdot)$ for G-PBA and scheme comparison wrt other G-PBAs.

Our motivation for G-PBA is based on stochastic root-finding. As a synthetic example, we consider a simulator of the form $Y(x) := \text{sign } Z(x)$ with the latent function $Z(x) = g(x) + \varepsilon(x)$, where $g(x) = 1/9 - x^2, x \in [0,1]$, and $\varepsilon(x) \sim \mathsf{N}(0, 0.04)$. Thus, $x^* = 1/3$ and $p(x) = \Phi(5|1/9 - x^2|)$. Figure 1 shows the surrogate GP fit after $M = 20$ macro-iterations and batch size $k = 450$. Starting with $f_0 \sim \mathsf{Unif}[0,1]$, the second panel shows that f_n quickly concentrates around x^*. At $N = 520$ (equivalent to $T = 9,500$ function evaluations) we have the point estimate median$(f_N) \cong 0.333096$ and IQR$(f_N) = 0.006$, indicating minimal posterior uncertainty. Finally, Figure 1 also demonstrates that the PBA with spatial modeling performs significantly better than other heuristics as in Rodriguez and Ludkovski (2016): root estimates have low average error and minimal uncertainty (as measured by the f_N inter-quartile range).

REFERENCES

Rodriguez, S., and M. Ludkovski. 2016. "Generalized Probabilistic Bisection for Root-Finding of Monotone Noisy Responses". University of California, Santa Barbara, Working Paper.

Waeber, R., P. I. Frazier, and S. G. Henderson. 2011. "A Bayesian approach to stochastic root finding". In *Proceedings of the 2011 Winter Simulation Conference (WSC)*, 4033–4045. IEEE.