Adaptive Failure Probability Estimation with Gaussian Processes

Aleksei Sorokin (asorokin@hawk.iit.edu), Vishwas Rao

Background

- *Efficiently approximate the probability of system failure subject to random parameters.*
- A system "fails" when the output of an *expensive* simulation exceeds a given threshold.
- Computational simulations are seeing increased use in reliability certification.
- Applications in structural design, power grids, safety certification etc.

Formulation

- $\mathcal{X} \sim \mathcal{U}(0,1)^d$, the parameter distribution, perhaps after a change of variables.
- $g : (0,1)^d \to \mathbb{R}$, the *expensive* simulation based on input parameters $\mathcal{X}$.
- $\mathcal{G} = \{\boldsymbol{x} \in (0,1)^d : g(\boldsymbol{x}) \geq 0\}$, the failure region.
- The *probability of failure* quantity of interest is

$$P_\mathcal{G} = P(\mathcal{G}) = \mathbb{E}\left[1_\mathcal{G}(\mathcal{X})\right] = \text{Volume}(\mathcal{G}).$$

Gaussian Process (GP)

- $k : (0,1)^d \times (0,1)^d \to \mathbb{R}$, the prior covariance kernel.
- $\mathcal{D} = \{(X,Y)\}$, the data where $Y = g(X) + \epsilon$ and $\epsilon \sim \mathcal{N}(0, s^2 I)$ is IID noise.
- The posterior mean, covariance, variance given $\mathcal{D}$ are

$$\begin{aligned} m_\mathcal{D}(\boldsymbol{x}) &= k(\boldsymbol{x}, X) K_{X,X}^{-1} Y, \\ k_X(\boldsymbol{x}_1, \boldsymbol{x}_2) &= k(\boldsymbol{x}_1, \boldsymbol{x}_2) - k(\boldsymbol{x}_1, X) K_{X,X}^{-1} k(X, \boldsymbol{x}_2), \\ \sigma_X^2(\boldsymbol{x}) &= k_X(\boldsymbol{x}, \boldsymbol{x}) \end{aligned}$$

  where $K_{X,X} = k(X,X) + s^2 I$.
- The posterior failure probability at $\boldsymbol{x} \in (0,1)^d$ WRT $g(\boldsymbol{x}) \mid \mathcal{D} \sim \mathcal{N}\left(m_\mathcal{D}(\boldsymbol{x}), \sigma_X^2(\boldsymbol{x})\right)$ is

$$P(g(\boldsymbol{x}) \mid \mathcal{D} \geq 0) = \Phi\left(\frac{m_\mathcal{D}(\boldsymbol{x})}{\sigma_X(\boldsymbol{x})}\right) = \Phi_\mathcal{D}(\boldsymbol{x}).$$

Estimator & Error Bound

- $P_\mathcal{G}^{(n)} = \frac{1}{n}\sum_{i=1}^n 1_\mathcal{G}(\mathcal{X}_i)$ with $\mathcal{X}_1, \dots, \mathcal{X}_n \sim \mathcal{U}(0,1)^d$ and large n s.t. $P_\mathcal{G}^{(n)} \approx P_\mathcal{G}$.
- $\tilde{P}_\mathcal{G} = \frac{1}{n}\sum_{i=1}^n 1\{\Phi_\mathcal{D}(\mathcal{X}_i) \geq 1/2\}$, a point estimator of $P_\mathcal{G}^{(n)}$ based on the GP posterior.
- $P_\mathcal{G}^{(n)} \in \left[\tilde{P}_\mathcal{G} - \gamma_\mathcal{D}, \tilde{P}_\mathcal{G} + \gamma_\mathcal{D}\right]$ with probability at least $1-\alpha$ where

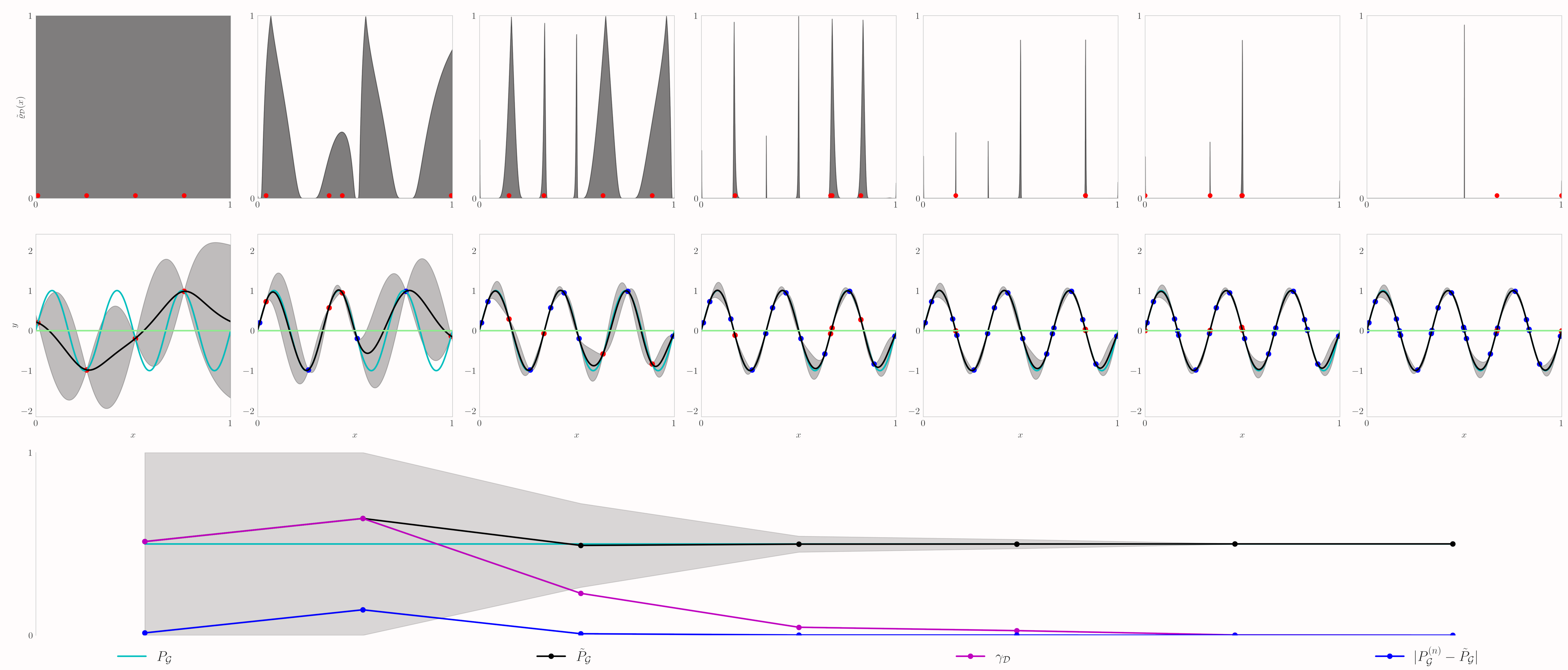
$$\gamma_\mathcal{D} = \frac{1}{n\alpha}\sum_{i=1}^n \left[\underbrace{\Phi_\mathcal{D}(\mathcal{X}_i) 1\{\Phi_\mathcal{D}(\mathcal{X}_i) < 1/2\}}_{\text{False Negative}} + \underbrace{(1-\Phi_\mathcal{D}(\mathcal{X}_i)) 1\{\Phi_\mathcal{D}(\mathcal{X}_i) \geq 1/2\}}_{\text{False Positive}}\right].$$

Algorithm

1. Fix $\mathcal{X}_1, \dots, \mathcal{X}_n \sim \mathcal{U}(0,1)^d$, we choose $n > 1$ million points of a quasi-random sequence.
2. Initialize an empty dataset, $\mathcal{D} \leftarrow \emptyset$, and set GP priors with fixed hyperparameters.
3. Draw $\boldsymbol{x}_1, \dots, \boldsymbol{x}_b \overset{\text{IID}}{\sim} \tilde{\varrho}_\mathcal{D}$ using rejection sampling from the unnormalized density

$$\tilde{\varrho}_\mathcal{D}(\boldsymbol{x}) = \underbrace{\Phi_\mathcal{D}(\boldsymbol{x}) 1\{\Phi_\mathcal{D}(\boldsymbol{x}) < 1/2\}}_{\text{False Negative}} + \underbrace{(1-\Phi_\mathcal{D}(\boldsymbol{x})) 1\{\Phi_\mathcal{D}(\boldsymbol{x}) \geq 1/2\}}_{\text{False Positive}}.$$

4. Update data to $\mathcal{D} \leftarrow \mathcal{D} \cup (X, Y)$ where $X = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_b)^T$ and $Y = g(X) + \epsilon$.
5. Compute $\{\Phi_\mathcal{D}(\mathcal{X}_i)\}_{i=1}^n$ from the GP posterior using efficient block-matrix updates.
6. Compute $\gamma_\mathcal{D}$, the half-width of the $1-\alpha$ confidence interval $[\tilde{P}_\mathcal{G} - \gamma_\mathcal{D}, \tilde{P}_\mathcal{G} + \gamma_\mathcal{D}]$.
7. If $\gamma_\mathcal{D}$ is less than some user-specified error tolerance ε, then we are done. Otherwise repeat from step 3 until the error is satisfied or the sample budget has expired.



Moving left to right traverses the algorithm's iterations and shows decreasing posterior uncertainty and approximation error. Each column shows the GP at the start of the iteration along with the subsequent sampling density $\tilde{\varrho}_\mathcal{D}(\boldsymbol{x})$, a quantity proportional to the posterior pointwise misclassification rate. The bottom row of each plot tracks the true solution, approximation, and confidence interval while comparing the half-width of the confidence interval to the true absolute error.