Gaussian Process Emulators for Expensive Simulations

Nicholas Krämer

Oberseminar

December 13, 2018

UNIVERSITÄT BONN

Outline

1. Expensive functions
2. Gaussian process emulators
3. Applications to Bayesian inverse problems (BIP)
4. Sequential design for Gaussian process emulators in BIP

Expensive functions

Example: Bayesian optimisation

- Say we are interested in

$$x^* = \arg\min_{x \in D} F(x)$$

- First problem: F is **really** expensive to evaluate
- Second problem: We do not know anything about F
- (except Lipschitz continuity, maybe)
- This is the general framework of **Bayesian optimisation**

Example: Mean of probability distribution

- $\star$ Let $f : A \to \mathbb{R}$ the density of a probability measure
- $\star$ Problem: f is **really** expensive to evaluate
- $\star$ What if we need the expected value

$$\mathbb{E}_\mu[X] = \int x \; f(x) \; \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} x_i f(x_i)$$

- $\star$ This problem is common in Bayesian inverse problems

Gaussian process emulators

Gaussian processes

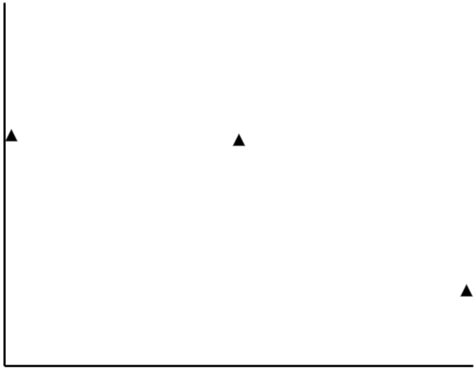
- Say $X = \{x_1, ..., x_N\}$ is a set of locations
- Reconstruct f from noisy data $y = f(X) + \eta$, $\eta \sim \mathcal{N}(0, \sigma^2)$
- Assume that f is a realisation of a **Gaussian process**

Definition (Gaussian process)

A Gaussian process Z_f is a random field, for which any finite number of evaluations has multivariate Gaussian distribution.

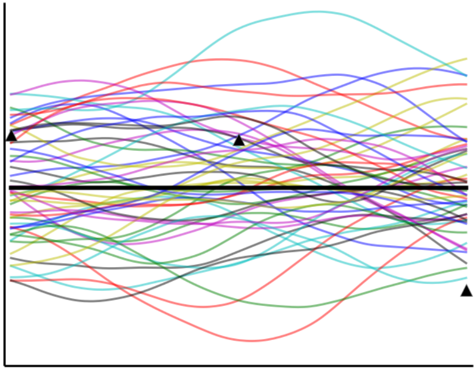
- Write: $Z_f \sim \mathrm{GP}(m, K) = \mathrm{GP}(m(\cdot), K(\cdot, \cdot))$
- K is symmetric and positive definite

Reconstruction problem



Data points; which function do these data points come from?

Sample paths of a Gaussian process



Data points and sample paths of a Gaussian process

Conditioning Gaussian processes

- ⋆ Assume we want to reproduce $Z_f \sim \mathrm{GP}(m, K)$ at a point s
- ⋆ Joint distribution:

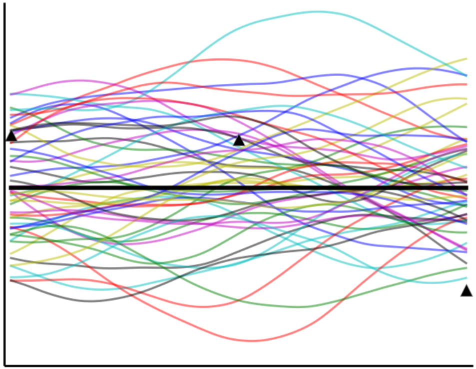
$$\begin{pmatrix} Z_f(X) \\ Z_f(s) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(X) \\ m(s) \end{pmatrix}, \begin{pmatrix} K(X,X) + \sigma^2 I_N & K(X,s) \\ K(s,X) & K(s,s) \end{pmatrix} \right)$$

$$\Rightarrow \;\; Z_f(s)|_{Z_f(X)+\eta=y} \sim \mathcal{N}(\hat{m}(s), \hat{K}(s,s)),$$

with predictive mean and predictive covariance

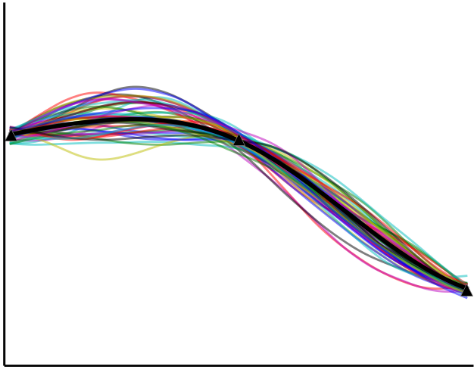
$$\begin{aligned} \hat{m}(s) &:= m(s) - K(s,X)(K(X,X) + \sigma^2 I_N)^{-1}(m(s) - y) \\ \hat{K}(s,t) &:= K(s,t) - K(s,X)(K(X,X) + \sigma^2 I_N)^{-1} K(X,t) \end{aligned}$$

Sample paths of a Gaussian process



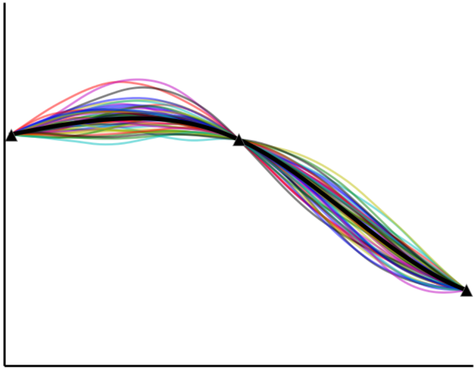
Data points and sample paths of a Gaussian process

Sample paths of a conditioned Gaussian process



Sample paths of a conditioned Gaussian process

Sample paths for exact data (i.e., $\sigma^2 = 0$)



Sample paths of a conditioned Gaussian process with exact data

Applications to Bayesian inverse problems (BIP)

Differential Equation on $(0,1)$

Find $a = (a_1, a_2) \in [-1,1]^2$ such that for

$$-\operatorname{div}((\sin(a_1 x) + \cos(a_2 x))\nabla u(x)) = 1, \quad u(0) = u(1) = 0$$

the measurements $y = (u(1/3), u(1/2), u(2/3))$ satisfy

$$u(1/3) = 2.2514, \quad u(1/2) = 6.213, \quad u(2/3) = 0.526.$$

- ⋆ Say $\mathcal{G} : [-1,1]^2 \to \mathbb{R}^3,\ (a_1, a_2) \mapsto (u_a(1/3), u_a(1/2), u_a(2/3))$
- ⋆ Problem: Find $a \in [-1,1]^2$ from $y = \mathcal{G}(a) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$

Bayesian inverse problems

- Initial belief of parameters: **prior distribution** μ_0 on $[-1,1]^2$
- Improve prior distribution: **posterior distribution** μ^y
  - using observations y (think: $y = (2.2514, 6.213, 0.526)$)
  - using forward map $\mathcal{G}$ (remember: **expensive**)
- Posterior distribution can be written as

$$\mu^y(x) \propto L(x)\mu_0(x)$$

- $L(x) = L(x|y, \mathcal{G})$ is called **likelihood**

Gaussian processes and Bayesian inverse problems

- ⋆ Evaluations of the likelihood need evaluations of $\mathcal{G}$
- ⋆ This is expensive!
- ⋆ Solution: Emulate the likelihood with a Gaussian process!
- ⋆ This works because small errors in the likelihood imply small errors in the posterior

A. M. Stuart, A. L. Teckentrup.
Posterior Consistency for Gaussian Process Approximations of Bayesian Posterior Distributions.
Math. Comp. 87, 721-753, 2018

Sequential design for Gaussian process emulators in BIP

Sequential design

Challenges:

- ⋆ Posterior measure concentrates w.r.t. prior
- ⋆ Capture this concentration in the locations for the emulator
- ⋆ We do not know the concentration a-priori

Strategy:

- ⋆ Abstract point of view: $\nu_N := \mathrm{GP}(m_N, K_N)$ approx. on N pts
- ⋆ Choose one location after the other: $\nu_N \to \nu_{N+1}$
- ⋆ Minimise the expected risk using the information ν_N

A promising strategy: Bayesian risk minimisation

Pick the new node as

$$x_{N+1} = \arg\min_{x} \mathbb{E}_{\mu_0}\left[\mathbb{E}_{\gamma_N(x)}\left[\mathbb{V}_{\nu_{N+1}(x)}\left[L(\cdot_{(\mu_0)}|y, \cdot_{(\nu_{N+1}(x))})\right]\right]\right],$$

where

- μ_0 is the prior distribution in the inverse problem
- $\gamma_N(x)$ is the law of the current GP ν_N at candidate point x
- $\nu_{N+1}(x)$ is the law of the next GP if one picks x as a new node

M. Sinsbeck, W. Nowak.
Sequential Design of Computer Experiments for the Solution of Bayesian Inverse Problems.
SIAM/ASA J. Uncertainty Quantification 5(1), 640-664, 2017

Ideas for upcoming work

- Check whether the sequential design strategy delivers:
  - Reproduce the experiments
  - Find a way to make it fail?
- Check different variations
- Compare to other experimental design algorithms
- *Think about why this works?*

References

A. M. Stuart, A. L. Teckentrup.
Posterior Consistency for Gaussian Process Approximations of Bayesian Posterior Distributions.
Math. Comp. 87, 721-753, 2018

M. Sinsbeck, W. Nowak.
Sequential Design of Computer Experiments for the Solution of Bayesian Inverse Problems.
SIAM/ASA J. Uncertainty Quantification 5(1), 640-664, 2017

M. Sinsbeck.
Uncertainty Quantification for Expensive Simulations–Optimal Surrogare Modeling under Timer Constraints.
Dissertation, Universität Stuttgart, 2017