Non-Gaussian Process Regression

Yaman Kındap and **Simon Godsill**

Signal Processing and Communications Laboratory, University of Cambridge, UK.

Introduction

We extend the Gaussian process (GP) framework into a new class of time-changed GPs that allow for straightforward modelling of heavy-tailed non-Gaussian behaviours, while retaining a tractable conditional GP structure through an infinite mixture of non-homogeneous GPs representation. We present Markov chain Monte Carlo inference procedures for this model and demonstrate the potential benefits compared to a standard GP.

A non-Gaussian process (NGP) prior on functions can be obtained by randomly transforming the inputs using a subordinator Lévy process and carrying out GP regression on the transformed input space. The resulting posterior distribution follows a non-Gaussian Lévy process.

The latent layer that is obtained by the random transformation of the input space represents the random distances between any two points on an input space. A subordinator process prior on the transformation has non-negative, non-decreasing sample paths with independent and stationary increments and no fixed discontinuities [1]. Thus, the monotonicity of the input space is preserved.

Regression model

Given a set of input-output pairs $\{x_i, y_i\}$ such that $y_i = f(x_i) + \varepsilon_i$, consider a latent input transformation such that x_i is mapped to $W(x_i)$ where $\{W(x); x \in \mathcal{X}\}$ is a subordinator. The associated prior on the transformation function is then defined as $p(W)$.

Given $W(x)$, consider a conditional GP prior over f such that:

$$p(f|W) \sim \mathcal{GP}(m_W(x), K_W(x', x))$$

where $m_W(x) = m\big(W(x)\big)$, $K_W(x', x) = K\big(W(x'), W(x)\big) = K(|W(x') - W(x)|)$ and $K(\cdot,\cdot)$ is a stationary kernel function e.g. squared exponential or Matérn. The joint distribution over the product space of f and W characterises the NGP prior. Given a set of observations $y_{1:n}$ the posterior on the joint space can be found as

$$p(f, W|y_{1:n}) = \frac{p(y_{1:n}|f)p(f|W)p(W)}{p(y_{1:n})}$$

The conditional GP structure of a NGP induces a posterior mean $\bar{m}_W(\cdot)$ and kernel function $\bar{K}_W(\cdot,\cdot)$ that can be evaluated analytically, i.e. $p(f|y_{1:n}, W) \sim \mathcal{GP}(\bar{m}_W, \bar{K}_W(\cdot,\cdot))$.

The NGP posterior distribution over the function space is found as

$$p(f|y_{1:n}) = \int p(f|y_{1:n}, W)p(W|y_{1:n})dW$$

where $p(W|y_{1:n})$ is the posterior distribution of the subordinator process. Inferring $p(W|y_{1:n})$ and hence $p(f|y_{1:n})$ is analytically intractable, however using approximate inference methods allow for straightforward extensions of the model and full Bayesian inference.

Inference in NGP models

Shot-noise simulation methods

A subordinator process $W(x)$ can represented as an infinite series as:

$$W(x) = \sum_{i=1}^{\infty} Y_i \, \mathbb{I}\,(V_i \le x)$$

Where $\{V_i \in \mathcal{X}_0\}_{i=1}^{\infty}$ are i.i.d. uniform random variables independent of $\{Y_i\}$ that represent the position of jumps, and $\{Y_i\}_{i=1}^{\infty}$ are the jump sizes characterised by the Lévy measure Q and $\mathcal{X}_0$ defines a subset of the input space. [3] proves the almost sure convergence of this series to $W(x)$ and a more detailed review can be found in [2]. Alg. 1 provides a method for the simulation of jump sizes for a tempered stable process.

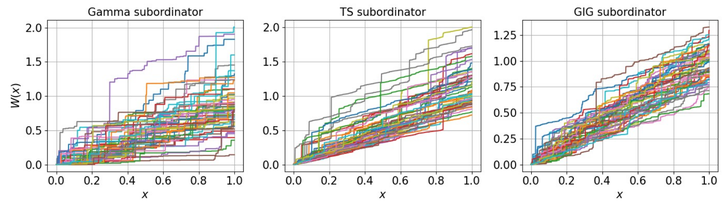

Fig. 1: Random subordinator sample paths associated with gamma, tempered stable and generalised inverse Gaussian processes.

Approximate inference

A Gibbs sampler approximating samples from $p(W|y_{1:n})$ can be implemented by simulating the associated bivariate random points that define the jump size and position on small disjoint intervals $\tau = (x_j, x_l)$ conditioned on the previous sample points in $-\tau = \mathcal{X} \setminus (x_j, x_l)$ and observations. Progressively simulating these points such that the whole input space is covered leads to approximate samples from the target distribution. The corresponding algorithms outlined in Algorithms 2 and 3.

The resulting samples $\{W^{(k)}\}$ are individually associated with conditional GP posterior functions $p(f|y_{1:n}, W^{(k)})$ that are completely defined through their mean $\bar{m}_{W^{(k)}}$ and covariance $\bar{K}_{W^{(k)}}$ functions. Such a collection forms a Gaussian mixture distribution, and the mean and covariance of the corresponding mixture density can be obtained as

$$\mathbb{E}_{f|y_{1:n}}[f] = \frac{1}{N}\sum_{k=1}^{N} \bar{m}_{W^{(k)}} = m_{f|y_{1:n}}$$

$$\mathbb{C}ov_{f|y_{1:n}}(f) = \frac{1}{N}\sum_{k=1}^{N} [\bar{K}_{W^{(k)}} + (\bar{m}_{W^{(k)}} - m_{f|y_{1:n}})(\bar{m}_{W^{(k)}} - m_{f|y_{1:n}})^T]$$

where N is the number of samples and $\mathbb{E}_{f|y_{1:n}}[f]$, $\mathbb{C}ov_{f|y_{1:n}}(f)$ define the posterior mean and covariance of the random function f.

References

[1] W. Feller. *An Introduction to Probability Theory and its Applications.* Wiley Mathematical Statistics Series v.2. Wiley, 1996.

[2] S. Godsill and Y. Kındap. *"Point process simulation of generalized inverse Gaussian processes and estimation of the Jaeger integral"*. In: *Statistics and Computing* 32.1 (Dec. 2021), p. 13.

[3] J. Rosiński. *"Series Representations of Lévy Processes from the Perspective of Point Processes"*. In: *Lévy Processes: Theory and Applications.* Boston, MA: Birkhäuser Boston, 2001, pp. 401-415.

Sampling Algorithms

Algorithm 1 Generation of the jumps of a tempered stable process with Lévy density $Q_{TS}(x) = Cx^{-1-\alpha}e^{-\beta x}$ where α is the tail parameter and β is the tempering parameter.

1. Assign $N_{TS} = \emptyset$,
2. Generate the epochs of a unit rate Poisson process, $\{\Gamma_i; i = 1,2,3, \dots\}$,
3. For $i = 1,2,3, \dots$
   - Compute $x_i = \left(\frac{\alpha\Gamma_i}{C}\right)^{-1/\alpha}$,
   - With probability $e^{-\beta x_i}$, accept x_i and assign $N_{TS} = N_{TS} \cup x_i$.

Algorithm 2 Simulating sample paths from the proposal density $p(W_\tau|W_{-\tau})$.

Given a random length set $N_W = \{V_i^{(k)}, M_i^{(k)}\}$ and an interval $(x_j, x_l) \in \mathcal{X}$,

1. Simulate $\{V_i^{(\prime)}, M_i^{(\prime)}\}$ with rate $|x_j - x_l|$ using Alg. 1,
2. Remove all points $\{V_i^{(k)}, M_i^{(k)}\}$ from N_W such that $x_j < V_i^{(k)} < x_l$ and add $\{V_i^{(\prime)}, M_i^{(\prime)}\}$, i.e. $N_W = N_W \cup \{V_i^{(\prime)}, M_i^{(\prime)}\}$,
3. Substitute the points of N_W into Eq. 1 to obtain the proposed sample path $W^{(\prime)}$.

Algorithm 3 MH-within-Gibbs sampler for $p(W|y_{1:n})$.

1. Initialise $W^{(0)}$ by simulating $\{V_i, M_i\}$ from the associated bivariate point process using Alg. 1,
2. Analytically evaluate $\bar{m}_{W^{(0)}}$, $\bar{K}_{W^{(0)}}$ which define the conditional GP posterior $p(f|y_{1:n}, W^{(0)})$ and the conditional likelihood $p\big(y_{1:n}|W^{(0)}\big)$,
3. For N times, iterate over $\tau_j \in \mathcal{X}$ where $\cup_{j=1}^{J} \tau_j = \mathcal{X}$,
   (a) Using τ_j and the points $\{V_i^{(k)}, M_i^{(k)}\}$ associated with $W^{(k)}$, sample a proposed sample path $W^{(\prime)}$ using Alg. 2,
   (b) Evaluate $\bar{m}_{W^{(\prime)}}$, $\bar{K}_{W^{(\prime)}}$ and $p(y_{1:n}|W^{(\prime)})$,
   (c) With probability $\alpha\big(W^{(\prime)}, W^{(k)}\big)$ the proposal is accepted and $W^{(k+1)} = W^{(\prime)}$, otherwise reject and set $W^{(k+1)} = W^{(k)}$.