The Gaussian Neural Process

Wessel P. Bruinsma[1,2], James Requeima[1,2],
Andrew Y. K. Foong[1], Jonathan Gordon[1], Richard E. Turner[1]

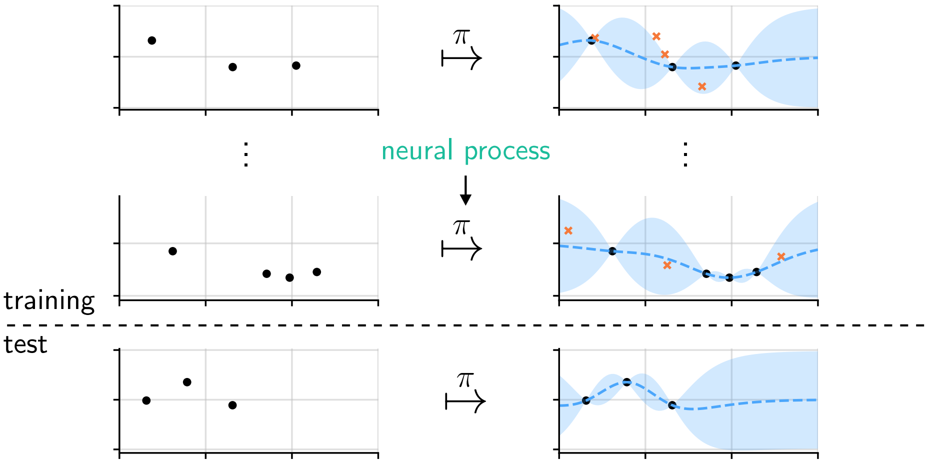
[1]University of Cambridge, [2]Invenia Labs

Advances in Approximate Bayesian Inference 2020

Neural Processes and Prediction Maps

Prediction Maps

$\pi:$ data sets $\mathcal{D}$ $\longrightarrow$ predictions $\mathcal{P}$



$\overset{\pi}{\longmapsto}$

$\vdots$ neural process $\vdots$

$\overset{\pi}{\longmapsto}$

training

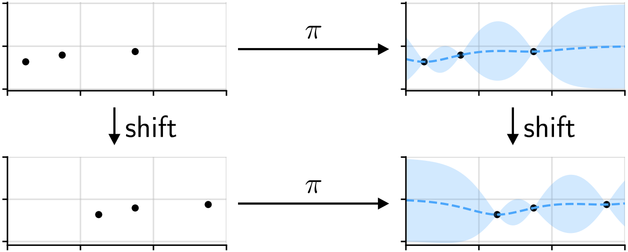
test

$\overset{\pi}{\longmapsto}$

Gaussian, TE Prediction Maps

- Gaussian, translation-equivariant (TE) prediction maps ($\Pi^{\mathsf{TE}}_{\mathsf{G}}$):

$$\pi\colon \mathcal{D} \to \text{Gaussian processes } \mathcal{P}_{\mathsf{G}}, \quad \pi(\mathsf{T}_\tau D) = \mathsf{T}_\tau \pi(D).$$



The Gaussian Neural Process

Setup

e.g., a sawtooth wave
↓

- Given a non-Gaussian, stationary stochastic process f.
- Posterior prediction map: $\pi_f \colon \mathcal{D} \to \mathcal{P}$, $\pi_f(D) = p(f \mid D)$.
- f stationary $\iff$ π_f translation equivariant.
- Goal: approximate π_f with Gaussian, TE $\tilde{\pi} \in \Pi^{\mathsf{TE}}_{\mathsf{G}}$:

$$\tilde{\pi}(D) = \operatorname*{arg\,min}_{\mu \in \mathcal{P}_{\mathsf{G}}} \mathrm{KL}(\pi_f(D), \mu).$$

→ careful theoretical analysis in paper

✗ Approximate f with GP and perform GP inference.

✓ Directly approximate every posterior of f with a GP.

- Gaussian Neural Process (GNP): general parametrisation of $\tilde{\pi}$.

↑
backed by universal representation theorem in paper

Gaussian Neural Process (GNP)

- GNP: fully general parametrisation of a TE map $\mathcal{D} \to \mathcal{P}_{\mathsf{G}}$.
- Separately parametrises mean map m and kernel map k:

$$m\colon \mathcal{D} \to \underset{\substack{\uparrow \\ \text{TE } \checkmark}}{\text{mean functions}}, \quad k\colon \mathcal{D} \to \underset{\substack{\uparrow \\ \text{diagonally TE?}}}{\text{kernel functions}}.$$

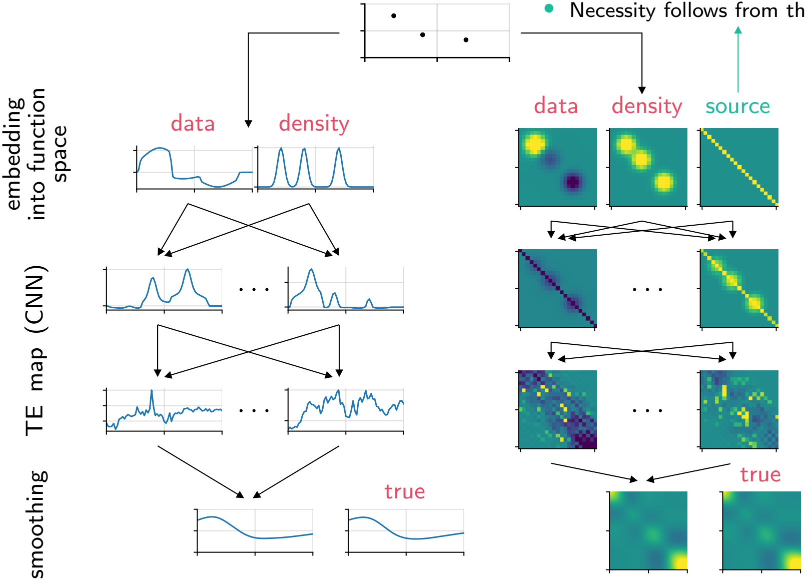
- Build on SetConv (Gordon et al., 2020): functional representation of data.
- Train with maximum likelihood:

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{N} \log p(D_i^{(t)} \mid D_i^{(c)}, \theta).$$

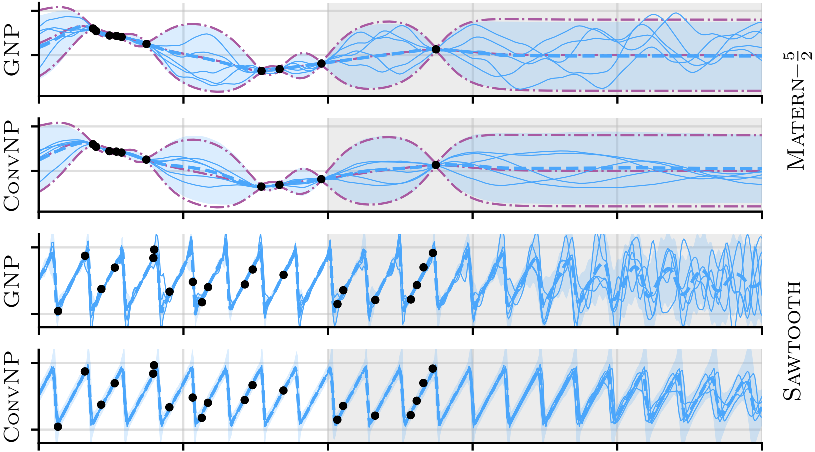
✓ Closed-form predictive: no approximation required!

Architecture of the GNP

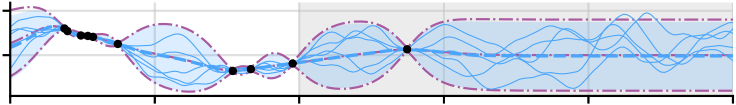
- Inject white noise for stationary prior
- Necessity follows from theory



Results



Conclusion



- Gaussian Neural Process: param. of a TE map $\mathcal{D} \to \mathcal{P}_{\mathsf{G}}$.

✓ Parametrisation is general (universal repr. theorem).

✓ Closed-form predictive: no approximation required.

✗ Computationally expensive.

Julia: `https://github.com/wesselb/NeuralProcesses.jl`
Python: `https://github.com/wesselb/neuralprocesses` (⚠)

Appendix

References

Gordon, J., Bruinsma, W. P., Foong, A. Y. K., Requeima, J., Dubois, Y., & Turner, R. E. (2020). Convolutional conditional neural processes. *International Conference on Learning Representations (ICLR), 8th*. `https://openreview.net/forum?id=Skey4eBYPS`