AARHUS UNIVERSITY

Hierarchical Gaussian filtering of sufficient statistic time series for active inference

Christoph Mathys[1,2,3] and Lilian Weber[3,4]

[1]Interacting Minds Centre, Aarhus University, Aarhus, Denmark

[2]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy

[3]Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland

[4]University of Oxford, Oxford, United Kingdom

International Workshop on Active Inference 2020

ECML-PKDD 2020 (virtual conference)

14-18 September 2020

Updating the mean of a series of observations

The usual way to calculate the mean $\bar{x}$ of $x_1, x_2, \dots, x_n$ is to take

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

However, when observations are made sequentially, it is more convenient to update the mean $\bar{x}$ after each new observation.

$$\bar{x}_{n+1} = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)$$

We show that this can be applied to the sufficient statistics of exponential-family distributions

In our paper, we show that **Bayesian inference with exponential-family distributions can be reduced to updating the mean of the sufficient statistics using the equation just shown**:

$$\boldsymbol{\xi}' = \boldsymbol{\xi} + \frac{1}{\nu + 1}(\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{\xi})$$

- $\boldsymbol{T}(\boldsymbol{x})$ (a function of the **observation $\boldsymbol{x}$**) is the sufficient statistic
- $\boldsymbol{\xi}$ is the hyperparameter vector – this parameterizes the **prior**
- $\boldsymbol{\xi}'$ is the **updated** hyperparameter vector – this parameterizes the **posterior**
- ν is an additional hyperparameter which determines the weight of the prior relative to the observation. It can be interpreted as **the number of observations implicitly contained in the prior**.

As we show, this only works with a **particular conjugate prior**:

$$p(\boldsymbol{\vartheta}|\boldsymbol{\xi}, \nu) = z(\boldsymbol{\xi}, \nu)\exp\big(\nu(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{\xi} - A(\boldsymbol{\vartheta}))\big)$$ (where $z(\boldsymbol{\xi}, \nu)$ is a normalization constant)

Adequate predictive distributions are critical to active inference

- Active inference relies on updating beliefs and initiating actions based on prediction errors
- However, evaluating prediction errors requires adequate predictive distributions
- Exponential-family models can have predictive distributions which are not exponential families
- For example, a Gaussian model (Gaussian likelihood with Gaussian conjugate prior) has a generalized Student's t distribution ('Gaussian-predictive') as its predictive distribution when both the mean and the variance of the likelihood are unknown.
- Our updating scheme accommodates this efficiently

Updates to the Gaussian-predictive

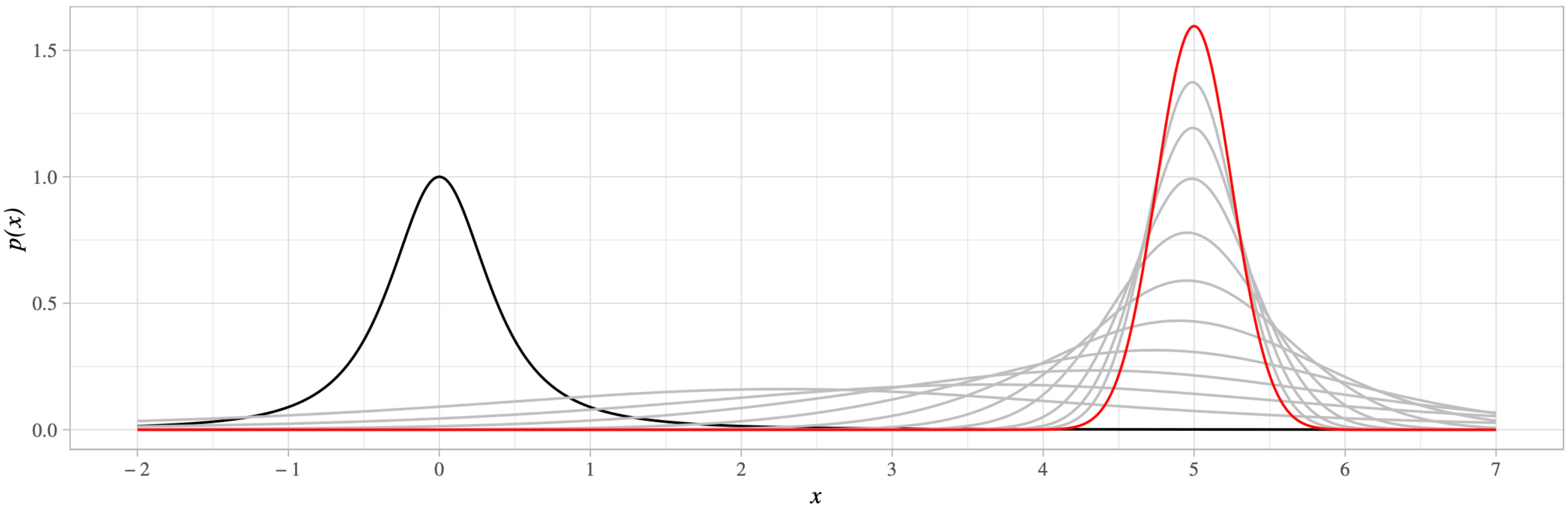

Fig. 1. Sequential updates to the Gaussian-predictive distribution $\mathcal{NP}$ in response to 1024 samples drawn from a Gaussian with mean 5 and standard deviation 1/4 (red). Initial hyperparameters were $\xi_x = 0$, $\xi_{x^2} = 1/8$, and $\nu = 1$, corresponding to the initial $\mathcal{NP}$ in black. Updated predictive distributions after $2, 4, 8, \ldots, 1024$ samples are shown in grey.

Hierarchical Gaussian filtering determines the right ν for the predictive distribution of time series

- When we use hierarchical Gaussian filter (HGF) on a time series, this gives us an implied ν for each point in time.

- If the HGF is appropriately calibrated, the implied ν, applied to the sufficient statistics, will give an appropriate predictive distribution.

- **Crucially, the predictive distribution will not only have an appropriate mean and precision, but will also be adequately fat-tailed.**

- This goes beyond existing filtering solutions such as DEM (dynamic expectation maximization, Friston et al., 2008)

Dynamically updated ν thanks to filtering

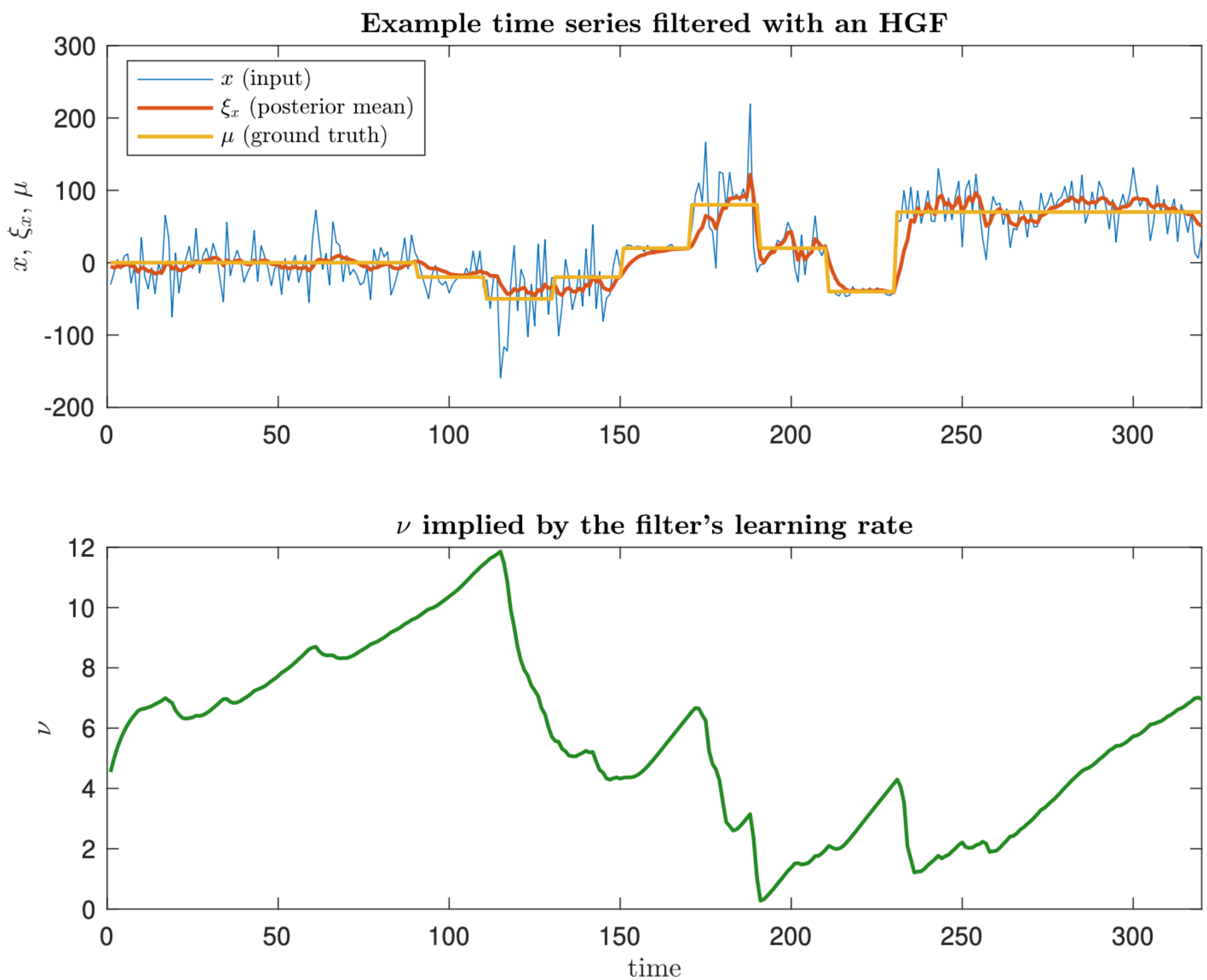

Fig. 2. Example of a time series (input x, top panel, fine blue line) filtered with an HGF (posterior mean ξ_x, top panel, red line), which infers the ground truth μ (top panel, yellow line) well in a volatile environment. Comparison of the HGF updates with Eq. 8 yields implied ν (bottom panel). This never rises above 12, ensuring a fat-tailed predictive distribution. In stable phases, implied ν rises; in volatile phases, it falls.

Thanks!