Data Science 9

Chris Mathys

Master's Degree Programme in Cognitive Science

Spring 2023

Hierarchical Gaussian Filtering: Implementations and practical challenges

Three official implementations

- **HGF Toolbox**
  - Matlab-based
  - Released October 2012
  - Actively maintained
  - Most widely used

- **HierarchicalGaussianFiltering.jl**
  - Julia package
  - Released September 2022
  - Chief maintainer: Peter Thestrup Waade

- **pyhgf**
  - Python package
  - Developed and used internally since 2015, public release imminent
  - Chief maintainer: Nicolas Legrand

Interactive demo: HGF Toolbox

- Available at

https://translationalneuromodeling.github.io/tapas

- Start with README, manual, and **interactive demo**

- Modular, extensible, Matlab-based

```
est2 = tapas_fitModel(sim2.y,...
                      usdchf,...
                      'tapas_hgf_config',...
                      'tapas_gaussian_obs_config',...
                      'tapas_quasinewton_optim_config');
```

```
Parameter estimates for the perceptual model:
    mu_0: [1.0352 1]
    sa_0: [3.7101e-05 0.0996]
     rho: [0 0]
      ka: 1
      om: [-12.8680 -1.8689]
    pi_u: 9.8449e+03
```

```
Parameter estimates for the observation model:
    ze: 2.3970e-05
```

Remarks on parameter recovery

- *Parameter recovery* means the ability to estimate the 'ground truth' parameter values put into a simulation
- This is never possible to do exactly because simulation is stochastic
- The stochastic nature of the simulation means that the parameter value best able to explain the simulated data is not the same as that used in the simulation. By nature, the simulation only noisily represents the 'ground truth' parameter values.
- It is therefore misleading to call the value used in the simulation 'true'. One could just as well argue that the estimated value is the 'true' one since it best explains the data.
- In order to get an idea of how exactly a parameter can be recovered, run many (on the order of tens to hundreds) simulations and look at the distribution of parameter estimates.
- **Estimating a parameter that isn't well recovered from simulations can still make sense because adjusting its value can improve the model by leading to a better explanation of the data**

Remarks on model comparison / model selection

- There is a range of scores that help in choosing a well-performing model: AIC (Akaike information criterion), BIC (Bayesian information criterion), Bayes factors, LME (log-model evidence), free energy, etc.
- Each model gets a particular score (which is on its own uninterpretable!)
- The difference in score between models is what counts
- However, model selection is not straightforward. AIC and BIC penalize complexity based on simple heuristics, which may not reflect complexity accurately. LME is better on that count, but is very sensitive to the modeller's choice of priors.
- **Three important considerations:**

  1. **Does the model allow me to answer my question of interest?**
  2. **Does the *prior predictive* distribution of observations make sense?**
  3. **Does the *posterior predictive* distribution of observations make sense?**

  When the answer to all three is yes, the model is fine.