Updating Variational Bayes: Fast sequential posterior inference

Nathaniel Tomasetti (Coles), Catherine Forbes (Monash University)
and Anastasios Panagiotelis (University of Sydney)

ISBA 2022, with thanks to Nadja Klein

Introduction

- Intermittent time series data, may require quick response
  - e.g. **self-driving vehicles** require regular monitoring all surrounding vehicles **to predict** imminent behaviour of their human drivers

- Bayesian methods account for **uncertainty** in the models and/or predictions
  - updating methods target a **sequence of posterior distributions**, each conditioned on an expanding data set
  - produced through a sequential application of Bayes theorem

- So-called **"Exact"** computational methods are slow and may scale poorly
  - **MCMC** ...
  - **SMC** (Doucet *et al*, 2001; Chopin, 2002)

Introduction

- Other approximations may be slow, and have unknown error
  - **grid-based** methods (Bhattacharya & Wilson, 2018; Chen X, Dai H, Song L, 2019)
  - **ABC** (Jasra *et al.*, 2010; Del Moral *et al.*, 2015)

- **VB** produces **approximate posterior** inference, **via optimisation**
  - fast
  - produces good point estimates and predictions
  - See Zhang *et al.*, 2017 for a recent review

Exact posterior vs VB approximation

- Target of Bayesian inference: **the posterior**

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{Y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

  - $\boldsymbol{\theta}$, parameter vector
  - $\mathbf{Y} = \mathbf{y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$, time series of observed vectors

- **VB Posterior approximation**:

  1. Choose **parametric family** $q_\lambda(\boldsymbol{\theta} \mid \mathbf{Y})$
  2. Select λ^* to **maximise the ELBO**[1,2]:

$$\mathcal{L}(q, \boldsymbol{\lambda}) = E_q \left[\log(p(\boldsymbol{\theta}, \mathbf{Y})) - \log(q_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{Y}))\right]$$

[1] **ELBO** = Evidence Lower Bound.

[2] Minimising the **ELBO** equivalently minimises **KL divergence** from $q_\lambda(\boldsymbol{\theta} \mid \mathbf{Y})$ to $p(\boldsymbol{\theta} \mid \mathbf{Y})$

Exact posterior vs VB approximation

- **Exact posterior** updates the **prior**:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y})}_{\textbf{posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \cdot \underbrace{p(\mathbf{y} \mid \boldsymbol{\theta})}_{\textbf{likelihood}}$$

- **VB posterior**:

$$q_{\boldsymbol{\lambda}^*}(\boldsymbol{\theta} \mid \mathbf{Y}), \text{ where } \boldsymbol{\lambda}^* \text{ maximises the } \textbf{ELBO}$$

- **Stochastic VB posterior** (**SVB**):

$$q_{\boldsymbol{\lambda}^{(m^*)}}(\boldsymbol{\theta} \mid \mathbf{Y}), \text{ where } \boldsymbol{\lambda}^{(m^*)} \text{ numerically maximises the } \textbf{ELBO}$$

SVB

- For general λ, maximise **ELBO** with **SGA**[3]
  - Set $\boldsymbol{\lambda}^{(1)}$. Then, for $m = 1, 2, \ldots$, let

$$\boldsymbol{\lambda}^{(m+1)} = \boldsymbol{\lambda}^{(m)} + \rho^{(m)} \frac{\partial \widehat{\mathcal{L}(q, \boldsymbol{\lambda})}}{\partial \boldsymbol{\lambda}} \bigg|_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(m)}}$$

- **Unbiased score estimator** (Ranganath *et al.*, 2014)

$$\frac{\partial \widehat{\mathcal{L}(q, \boldsymbol{\lambda})}}{\partial \boldsymbol{\lambda}}_{SC} = \frac{1}{S} \sum_{j=1}^{S} \frac{\partial \log(q_{\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(j)} \mid \mathbf{Y}))}{\partial \boldsymbol{\lambda}} \left(\log(p(\mathbf{Y}, \boldsymbol{\theta}^{(j)})) - \log(q_{\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(j)} \mid \mathbf{Y})) - \widehat{\boldsymbol{a}} \right)$$

  - $\{\boldsymbol{\theta}^{(j)}, \text{ for } j = 1, 2, \ldots, S\}$ drawn from $q_{\boldsymbol{\lambda}^{(m)}}(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:T})$
  - $\widehat{\boldsymbol{a}}$, control variates

[3] **SGA = Stochastic Gradient Ascent** (Bottou, 2010; Hoffman *et al.*, 2013)

Dependence in the approximation

- Let $\boldsymbol{\theta} = (\theta_1, \theta_2)'$
  - **MFVB**[4] (independence)

$$q_\lambda(\theta_1, \theta_2 \mid \boldsymbol{y}_{1:T}) = q_\lambda(\theta_1 \mid \mathbf{Y})\; q_\lambda(\theta_2 \mid \mathbf{Y})$$

  - More general forms possible, e.g.

$$q_\lambda(\theta_1, \theta_2 \mid \mathbf{Y}) = q_\lambda(\theta_1 \mid \mathbf{Y})\; q_\lambda(\theta_2 \mid \theta_1, \mathbf{Y})$$

  - If possible, use **exact conditional posterior**:

$$q_\lambda(\theta_1, \theta_2 \mid \mathbf{Y}) = q_\lambda(\theta_1 \mid \mathbf{Y})\; \underbrace{p(\theta_2 \mid \theta_1, \mathbf{Y})}_{\textbf{exact conditional}}$$

  - Or use **exact marginal posterior**:

$$q_\lambda(\theta_1, \theta_2 \mid \mathbf{Y}) = \underbrace{p(\theta_1 \mid \mathbf{Y})}_{\textbf{exact marginal}}\; q_\lambda(\theta_2 \mid \theta_1, \mathbf{Y})$$

[4] **MFVB = Mean Field VB**, (Bishop, 2006)

Updating Variational Bayes (UVB)

- Notation:
  - T_1, T_2, ..., a sequence of time points
  - $\mathbf{Y}_k = \boldsymbol{y}_{T_{k-1}+1:T_k} = \{\mathbf{y}_{T_{k-1}+1}, \mathbf{y}_{T_{k-1}+2}, \dots, \mathbf{y}_{T_k}\}$, data vectors in "batch" k
  - $\mathbf{Y}_{1:k} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k\}$, expanding dataset for $k = 1, 2, \dots$

- We want the **sequence** of posterior distributions:

$$\text{At time } T_1: \quad \underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}_1)}_{\textbf{exact posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \underbrace{p(\mathbf{Y}_1 \mid \boldsymbol{\theta})}_{\textbf{likelihood}}$$

$$\text{At time } T_k: \quad \underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})}_{\textbf{exact posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \underbrace{p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta})}_{\textbf{likelihood}}$$

$$\propto \underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k-1})}_{\textbf{updated prior}} \underbrace{p(\mathbf{Y}_k \mid \mathbf{Y}_{1:k-1}, \boldsymbol{\theta})}_{\textbf{predictive likelihood}}$$

UVB

- $\Rightarrow$ At time T_1, **UVB** targets exact posterior:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}_1)}_{\textbf{exact posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \underbrace{p(\mathbf{Y}_1 \mid \boldsymbol{\theta})}_{\textbf{likelihood}}$$

- Apply **SVB**:
  1. Choose **(dependent) parametric distribution** $q_\lambda(\boldsymbol{\theta} \mid \mathbf{Y}_1)$
  2. Use **SGA** to maximise the corresponding **ELBO** and find $q_{\lambda_1^*}(\boldsymbol{\theta} \mid \mathbf{Y}_1)$
  3. Use approximation as **UVB posterior**

$$\widetilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_1) = q_{\lambda_1^*}(\boldsymbol{\theta} \mid \mathbf{Y}_1)$$

- $\Rightarrow$ At time T_1, **UVB = SVB**

UVB

- At time T_k, we have **UVB posterior**: $\widetilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k-1})$
  - Based on data $\mathbf{Y}_{1:k-1}$

- We want **exact posterior**: $p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$
  - Based on data $\mathbf{Y}_{1:k} = \{\mathbf{Y}_{1:k-1}, \ \mathbf{Y}_k\}$

- We could apply **SVB** again:

  1. Choose **(dependent) parametric distribution** $q_\lambda(\boldsymbol{\theta} \mid \mathbf{Y}_1 : k)$ targeting

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})}_{\textbf{exact posterior}} \propto \underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \underbrace{p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta})}_{\textbf{likelihood}}$$

  2. Use **SGA** to maximise the corresponding **ELBO** and find $q_{\lambda_k^*}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$

UVB

- Rather, at time T_k, **UVB targets**:

$$\underbrace{\tilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})}_{\textbf{UVB posterior}} \propto \underbrace{\tilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k-1})}_{\textbf{UVB prior}} \underbrace{p(\mathbf{Y}_{T_{1:k}} \mid \boldsymbol{\theta})}_{\textbf{predictive likelihood}}$$

1. Choose **(dependent) parametric distribution** $q_{\lambda}(\boldsymbol{\theta} \mid \mathbf{Y}_1 : k)$ targeting the **UVB posterior**
2. Use **SGA** to maximise the corresponding **ELBO** and find $q_{\lambda_k^*}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})$
3. Use approximation as **UVB posterior**

$$\widetilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_k) = q_{\lambda_1^*}(\boldsymbol{\theta} \mid \mathbf{Y}_k)$$

UVB

- What is the difference between **SVB** and **UVB**?

- At time T_k, the **SVB ELBO** is:

$$\mathcal{L}_{SVB}(q, \boldsymbol{\lambda}) = E_q \left[\log \left(\underbrace{p(\boldsymbol{\theta})}_{\textbf{prior}} \underbrace{p(\mathbf{Y}_{1:k} \mid \boldsymbol{\theta})}_{\textbf{likelihood}} \right) - \log(q_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})) \right]$$

- Whereas the **UVB ELBO** is:

$$\mathcal{L}_{UVB}(q, \boldsymbol{\lambda}) = E_q \left[\log \left(\underbrace{\widetilde{p}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k-1})}_{\textbf{UVB prior}} \underbrace{p(\mathbf{Y}_k \mid \mathbf{Y}_{1:k-1}, \boldsymbol{\theta})}_{\textbf{predictive likelihood}} \right) - \log(q_{\boldsymbol{\lambda}}(\boldsymbol{\theta} \mid \mathbf{Y}_{1:k})) \right]$$

Notes on UVB

- UVB does not impose MFVB[5], and hence will typically require iterative **SGA** for each $k = 1, 2, \ldots, n$
- Many updates require repeated applications of **SGA**
  - $\Rightarrow$ May be able to exploit Importance Sampling within one update[6]
- Sequence of distributional families $q_{\lambda_1}, q_{\lambda_2}, \ldots, q_{\lambda_n}$ can be different for each k
  - So far we have always retained the same functional form
  - Then the **UVB** objective is to find sequence of optimal $\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*, \ldots$
- We introduce a novel use of Importance Sampling for **UVB** that exploits draws from $\widetilde{p}_(\boldsymbol{\theta} \mid \mathbf{Y}_{k-1}$ for the score estimator associated with $\widetilde{p}_(\boldsymbol{\theta} \mid \mathbf{Y}_k$
  - We refer to our method as **UVB-IS**

[5] See Broderick *et al.*, 2013) for approach exploiting MFVB

[6] See Sayaka & Klami (2017) for an Importance Sampling approach for **SGA** at a given k, to reduce the cost of obtaining the score estimator

SVB with Importance Sampling

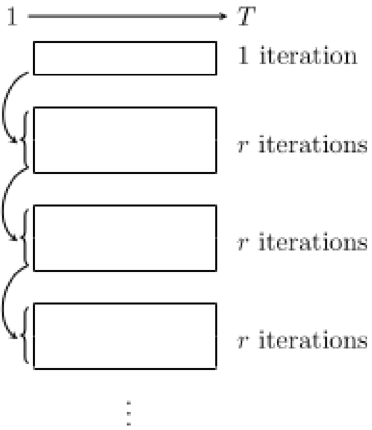


Figure 1: Importance sampling from Sayaka and Klami (2017).

UVB with Importance Sampling (UVB-IS)

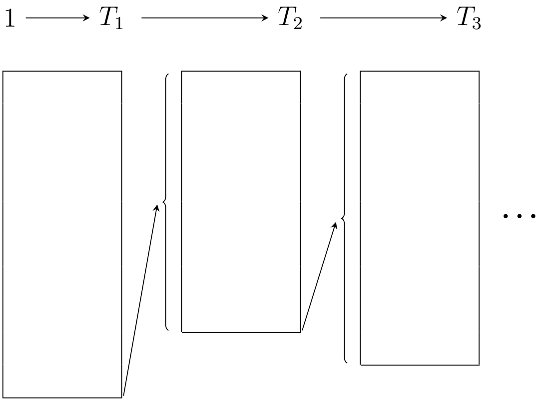


Figure 2: UVB-IS

Some simulation and empirical studies

- We investigate **UVB** and **UVB-IS** using **two simulation studies**
  1. $AR(3)$ time series forecasting application
  2. Clustering based on a mixture model
- Use the "Eight Schools" example from Gelman *et al.* (2014)
  - Start with one school and add data from other schools one at a time
- Compare against MCMC (RWMH) and standard SVB in terms of:
  - Classification/Prediction error
  - Mean cumulative run time

General Findings

- Approximation error relatively small for UVB
  - UVB comparable to SVB
  - errors accumulate more rapidly with UVB-IS
- Computational speed fast for UVB-IS
  - UVB slower but not as slow as full SVB updates
- Performance improved
  - Using “warm starts”
  - Periodic update with “true” posterior prior to further updating

Vehicle Lateral Lane Deviation

- NGSIM Data of 6000 vehicles driving on US 101 Highway
- Interest in driver behaviour
- Model the deviation from a vehicle to lane centre
  - details of data calculations in appendix
- Some groups of drivers may be similar, some may be unique
  - DPM model incorporates driver heterogeneity
- Analysis suggest smart vehicle (without a human driver) may be able to 'observe' and 'predict' neighbouring car positions and react in 'real time'

Vehicle Lateral Lane Deviation

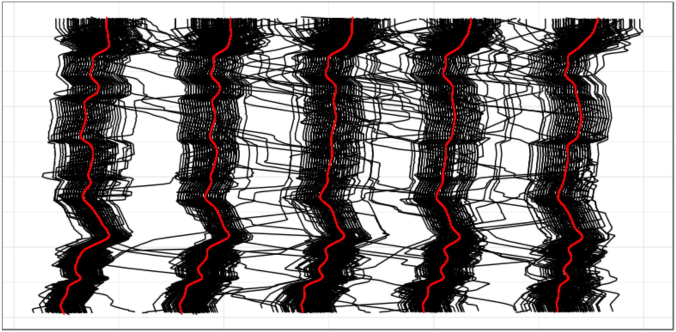


Figure 3: Raw data of car paths over five lanes. Each black line charts the path of a single car, with 100 randomly selected cars per lane shown on the figure. A fitted spline model (in red) for each lane used to correct for the geometry of the road.

Vehicle Lateral Lane Deviation

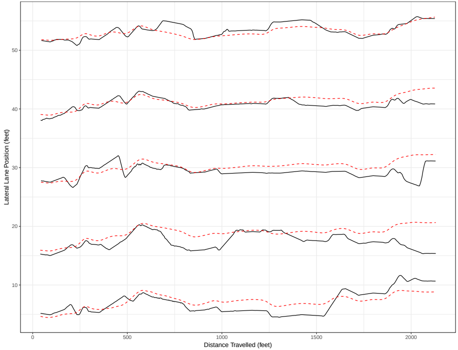


Figure 4: The path of five selected vehicles from the NGSIM dataset, travelling from left to right, with each black line representing a unique vehicle and with the estimated lane centre lines in red. This section of US Route 101 is comprised of five main lanes, with a sixth entry-exit lane not shown.

The Dirichlet Process Mixture (DPM)

- **DPM**

$$y_{i,t} \mid \boldsymbol{\theta}_i \overset{ind}{\sim} \mathcal{N}\left(\mu_i, \sigma_i^2\right), \quad \boldsymbol{\theta}_i = (\mu_i, \log(\sigma_i^2))'$$

$$\boldsymbol{\theta}_i \mid G \overset{iid}{\sim} G, \text{ for } i = 1, 2, \ldots, N$$

$$G \sim DP(\alpha, G_0)$$

- **Update times**
  - $T_n = 25 + 25n$ for $n = 1, 2, 3, 4, 5, 6$
  - All times use $N = 500$ Vehicles.

Why DPM?

- Cars may display similar behaviour, model allows different cross-sectional units to share parameters
- Cross-sectional units belong to mixture components
  - $\Rightarrow$ predictions 'borrow strength' from the full sample of vehicles
- Number of components unknown
- There is a possibility that a new vehicle will be observed with behaviour that cannot be well described by any of the prevailing parameters
- $\Rightarrow$ We consider an infinite mixture model.
- Let k_i denote an indicator variable for vehicle i belonging to mixture component j
- Vehicles in same cluster share parameters

The first posterior

- The exact initial posterior can be expressed as

$$p(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_1}) \propto \left[\prod_{i=1}^{N}\prod_{t=1}^{T_1} p(y_{i,t} \mid \boldsymbol{\theta}^*_{1:N}, k_i)\right] \times \left[\prod_{i=1}^{N} p(k_i \mid \boldsymbol{k}_{1:i-1})\right] p(\boldsymbol{\theta}^*_{1:N})$$

- Our UVB approximation:

$$q_{\lambda_1^*}(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_1}) = \left[\prod_{i=1}^{N} p(k_i \mid \boldsymbol{y}_{1:N,1:T_1}, \boldsymbol{k}_{1:i-1}, \boldsymbol{\theta}^*_{1:N})\right] \times q_{\lambda_1^*}(\boldsymbol{\theta}^*_{1:N} \mid \boldsymbol{y}_{1:N,1:T_1})$$

Updating the DPM at T_{n+1}

- Replace $p(\boldsymbol{\theta}^*_{1:N})$ with $q_{\lambda^*_n}(\boldsymbol{\theta}^*_{1:N} \mid \boldsymbol{y}_{1:N,1:T_n})$
- $p(k_i \mid \boldsymbol{y}_{1:N,1:T_n}, \boldsymbol{k}_{1:i-1}, \boldsymbol{\theta}^*_{1:N})$ uses $\boldsymbol{y}_{1:N,1:T_n}$ in SGA
  - Marginalise:

$$q(k_i = j \mid \boldsymbol{y}_{1:N,1:T_n}, \boldsymbol{k}_{1:i-1}) = \int_{\boldsymbol{\theta}^*_{1:N}} q_{\lambda^*_n}(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_n}) d\boldsymbol{\theta}^*_{1:N},$$

  - Use $\boldsymbol{y}_{1:N,1:T_n}$ once before updating.
  - Constant for different values of $\boldsymbol{\theta}$
  - Will not need to use $\boldsymbol{y}_{1:N,1:T_n}$ for new $\boldsymbol{\theta}$ in subsequent SGA

Updating the DPM at T_{n+1}

- The exact posterior is

$$p(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_{n+1}}) \propto \left[\prod_{i=1}^{N} \prod_{t=T_n+1}^{T_{n+1}} p(y_{i,t} \mid \boldsymbol{\theta}^*_{1:N}, k_i)\right]$$
$$\times \prod_{i=1}^{N} \left[p(k_i \mid \boldsymbol{y}_{1:N,1:T_n}, \boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:i-1})\right] p(\boldsymbol{\theta}^*_{1:N} \mid \boldsymbol{y}_{1:N,1:T_n})$$

- Instead target

$$\widetilde{p}(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_{n+1}}) \propto \left[\prod_{i=1}^{N} \prod_{t=T_n+1}^{T_{n+1}} p(y_{i,t} \mid \boldsymbol{\theta}^*_{1:N}, k_i)\right]$$
$$\times \prod_{i=1}^{N} \left[q(k_i \mid \boldsymbol{y}_{1:N,1:T_n}, \boldsymbol{k}_{1:i-1})\right] q_{\lambda^*_n}(\boldsymbol{\theta}^*_{1:N} \mid \boldsymbol{y}_{1:N,1:T_n})$$

Updating the DPM at T_{n+1}

- Approximate with

$$q_{\lambda^*_{n+1}}(\boldsymbol{\theta}^*_{1:N}, \boldsymbol{k}_{1:N} \mid \boldsymbol{y}_{1:N,1:T_{n+1}}) = q_{\lambda^*_{n+1}}(\boldsymbol{\theta}^*_{1:N} \mid \boldsymbol{y}_{1:N,1:T_{n+1}}) \\ \times \prod_{i=1}^{N} \widehat{p}(k_i \mid \boldsymbol{y}_{1:N,1:T_{n+1}}, \boldsymbol{k}_{1:i-1}, \boldsymbol{\theta}^*_{1:N})$$

where

$$\widehat{p}(k_i \mid \boldsymbol{y}_{1:N,1:T_{n+1}}, \boldsymbol{k}_{1:i-1}, \boldsymbol{\theta}^*_{1:N}) \propto \prod_{t=T_n+1}^{T_{n+1}} p(y_{i,t} \mid \boldsymbol{\theta}^*_{1:N}, k_i) \\ \times q(k_i \mid \boldsymbol{y}_{1:N,1:T_n}, \boldsymbol{k}_{1:i-1})$$

Approximating Posterior Distribution at T_6

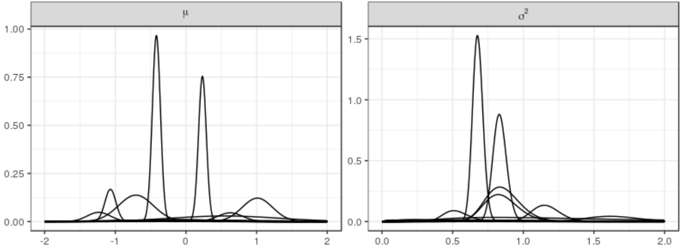


Figure 5: UVB Weighted marginal posterior mixture components at time T_6.v Densities are weighted by the number of vehicles associated with each $heta_j^*$.

Predictive distribution implied for each θ_j^*

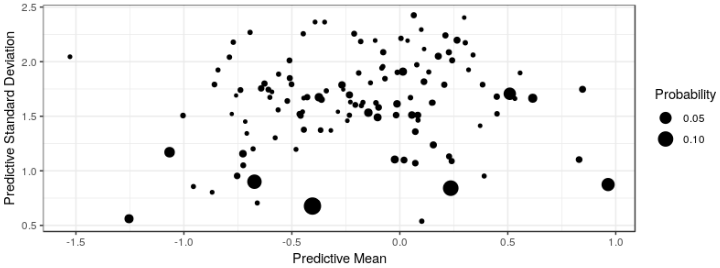


Figure 6: "UVB predictive moments for high probability groups at time T_6. Figure represents 80% of all vehicles."

Some predictive distributions

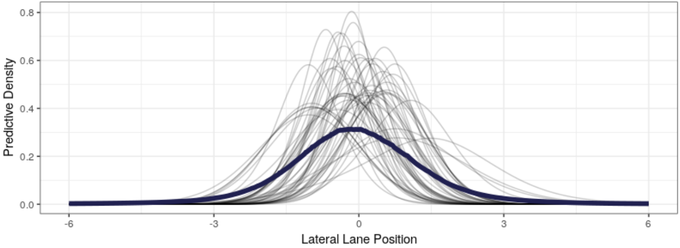


Figure 7: Individual vehicles and average predictive densities from UVB at time T_6. Grey: Individual, Blue: Average of 500 vehicles

Forecasting Methodology

At each T_n forecast next 50 observations for 500 vehicles via

1. The DPM model with UVB
2. The DPM model with SVB
3. The DPM model with MFVB
4. A Normal Likelihood / Jeffrey's Prior model applied independently for each vehicle

- Methods evaluated out-of-sample using mean cumulative log score (MCLS)
- Both 1. and 2. outperform 3. and 4.

Results Averaged Across 500 Vehicles

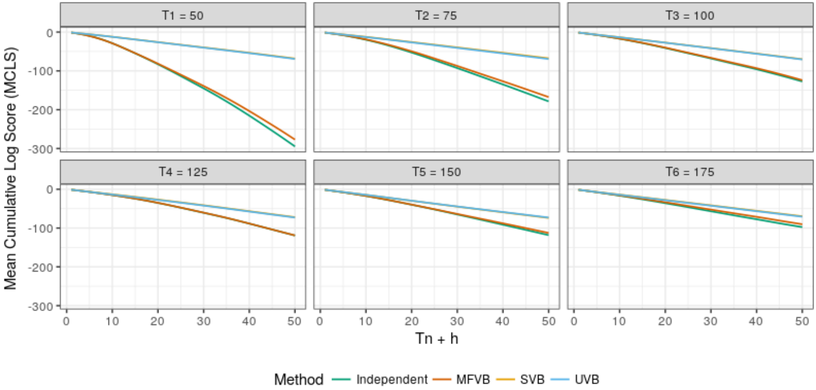


Figure 8: DPM Forecast Accuracy: Mean cumulative predictive log scores (MCLS) for each model, averaged across $N = 500$ vehicles and 50 forecast periods before updating. SVB and UVB outcomes are visually indistinguishable, while MFVB performs only slightly better than the fully independent model.

Summary: Lane Position Model

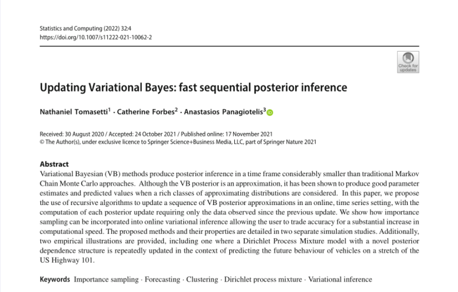
1. Propose a hierarchical time series model for vehicles
2. Using model posterior as a prior for new vehicles
3. Show that the model outperforms alternatives
4. Show that inclusion of heterogeneity improves forecasts
5. Demonstrate that UVB inference can feasibly update this model in close to real-time

Contributions of the paper:

1. Propose a method to update VB approximations (UVB).
2. Introduce UVB with Importance Sampling (UVB-IS).
3. Demonstrate applications of UVB and UVB-IS
4. Introduce idea of posterior dependence in VB approximation
5. $\Rightarrow$ a new class of approximations for the DPM
6. Detail how to update a DPM model
7. Challenging examples
    1. Time series forecasting (AR(3))
    2. Clustering with a binary mixture model
    3. Eight Schools example (normal means)
    4. Lane Position model (DPM)

Questions?

Paper available:



Statistics and Computing (2022) 32:4
https://doi.org/10.1007/s11222-021-10062-2

Check for updates

Updating Variational Bayes: fast sequential posterior inference

Nathaniel Tomasetti[1] · Catherine Forbes[2] · Anastasios Panagiotelis[3]

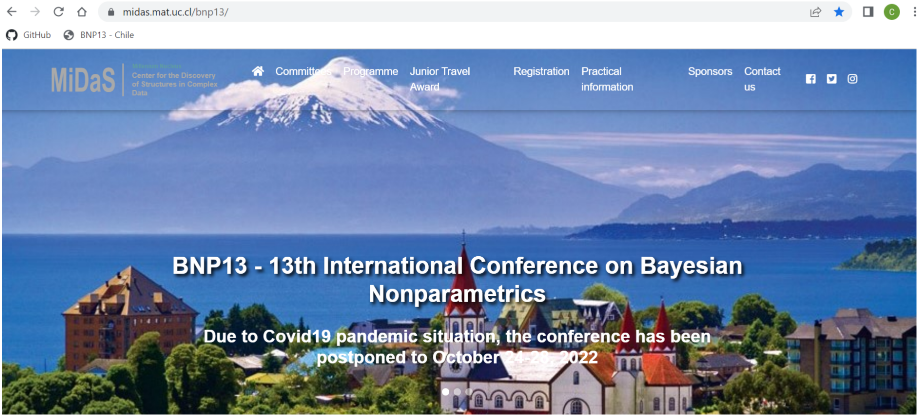
Received: 30 August 2020 / Accepted: 24 October 2021 / Published online: 17 November 2021

Abstract
Variational Bayesian (VB) methods produce posterior inference in a time frame considerably smaller than traditional Markov Chain Monte Carlo approaches. Although the VB posterior is an approximation, it has been shown to produce good parameter estimates and predicted values when a rich classes of approximating distributions are considered. In this paper, we propose the use of recursive algorithms to update a sequence of VB posterior approximations in an online, time series setting, with the computation of each posterior update requiring only the data observed since the previous update. We show how importance sampling can be incorporated into online variational inference allowing the user to trade accuracy for a substantial increase in computational speed. The proposed methods and their properties are detailed in two separate simulation studies. Additionally, two empirical illustrations are provided, including one where a Dirichlet Process Mixture model with a novel posterior dependence structure is repeatedly updated in the context of predicting the future behaviour of vehicles on a stretch of the US Highway 101.

Keywords Importance sampling · Forecasting · Clustering · Dirichlet process mixture · Variational inference

Figure 9: Statistics and Computing (2022)

Upcoming BNP Event October 2022



Upcoming BNP Event December 2023

MONASH University

MONASH BUSINESS SCHOOL

BNP Networking Workshop 2023

11-15 December 2023
Melbourne, Australia

Name

First Last

Email *

*

☐ I would like to receive information and updates on BNP Networking Workshop 2023

https://buseco.wufoo.com/forms/x9cczu20glq3b3/

e: bnpnet@monash.edu or catherine.forbes@monash.edu