
Supplementary Material for Bayesian models of Data Streams with HPPs

Andrés Masegosa^{1,2} Thomas D. Nielsen³ Helge Langseth² Darío Ramos-López¹ Antonio Salmerón¹
Anders L. Madsen^{3,4}

A. Proof of Theorem 1 and Lemma 2

Proof of Theorem 1. In the specification of \mathcal{L} we have that $\mathbb{E}_q[\ln \hat{p}(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}, \rho_t)]$ (defined in Equation (7)) can be expanded as (ignoring the base measure) :

$$\mathbb{E}_q[(\rho_t \boldsymbol{\lambda}_{t-1} + (1 - \rho_t) \boldsymbol{\alpha}_u) \mathbf{t}(\boldsymbol{\beta}_t) - a_g(\rho_t \boldsymbol{\lambda}_{t-1} + (1 - \rho_t) \boldsymbol{\alpha}_u)].$$

Since a_g is convex we have

$$a_g(\rho_t \boldsymbol{\lambda}_{t-1} + (1 - \rho_t) \boldsymbol{\alpha}_u) \leq \rho_t a_g(\boldsymbol{\lambda}_{t-1}) + (1 - \rho_t) a_g(\boldsymbol{\alpha}_u),$$

which combined with Equation (10) gives

$$\begin{aligned} & \mathbb{E}_q[\ln p(\mathbf{x}_t, \mathbf{Z}_t | \boldsymbol{\beta}_t)] + \mathbb{E}_q[(\rho_t \boldsymbol{\lambda}_{t-1} + (1 - \rho_t) \boldsymbol{\alpha}_u) \mathbf{t}(\boldsymbol{\beta}_t) \\ & - \rho_t a_g(\boldsymbol{\lambda}_{t-1}) - (1 - \rho_t) a_g(\boldsymbol{\alpha}_u)] + \mathbb{E}_q[p(\rho_t | \gamma)] \\ & - \mathbb{E}_q[\ln q(\mathbf{Z}_t | \boldsymbol{\phi}_t)] - \mathbb{E}_q[q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t)] - \mathbb{E}_q[q(\rho_t | \omega_t)] \leq \mathcal{L}. \end{aligned}$$

Lastly, by exploiting the mean field factorization of q and using the exponential family form of $p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1})$ and $p_u(\boldsymbol{\beta}_t)$ we get the desired result. \square

Proof of Lemma 2. Firstly, by ignoring the terms in $\hat{\mathcal{L}}$ (Equation (11)) that do not involve ω_t we get

$$\begin{aligned} \hat{\mathcal{L}}(\omega_t) &= \mathbb{E}_q[\rho_t] (\mathbb{E}_q[\ln(p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))] - \mathbb{E}_q[\ln p_u(\boldsymbol{\beta}_t)]) \\ &+ \mathbb{E}_q[p(\rho_t | \gamma)] - \mathbb{E}_q[q(\rho_t | \omega_t)]. \end{aligned}$$

Assuming that the sufficient statistics function $\mathbf{t}(\rho_t)$ for $p(\rho_t | \gamma)$ and $q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t)$ is the identity function ($\mathbf{t}(\rho_t) = \rho_t$) we have

$$\begin{aligned} \hat{\mathcal{L}}(\omega_t) &= \mathbb{E}_q[\rho_t] (\mathbb{E}_q[\ln(p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))] - \mathbb{E}_q[\ln p_u(\boldsymbol{\beta}_t)]) \\ &+ \gamma \mathbb{E}_q[\rho_t] - (\omega_t \mathbb{E}_q[\rho_t] - a_g(\omega_t)) + cte. \end{aligned}$$

Using $\mathbb{E}_q[\mathbf{t}(\rho_t)] = \mathbb{E}_q[\rho_t] = \nabla_{\omega_t} a_g(\omega_t)$ we get

$$\begin{aligned} \hat{\mathcal{L}}(\omega_t) &= \nabla_{\omega_t} a_g(\omega_t) (\mathbb{E}_q[\ln(p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))] - \mathbb{E}_q[\ln p_u(\boldsymbol{\beta}_t)]) \\ &+ \gamma \nabla_{\omega_t} a_g(\omega_t) - (\omega_t \nabla_{\omega_t} a_g(\omega_t) - a_g(\omega_t)). \end{aligned}$$

¹Department of Mathematics, University of Almería, Almería, Spain ²Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway ³Department of Computer Science, Aalborg University, Aalborg, Denmark ⁴Hugin Expert A/S, Aalborg, Denmark. Correspondence to: Andrés Masegosa <andresmasegosa@ual.es>.

and thereby

$$\nabla_{\omega_t} \hat{\mathcal{L}} = \nabla_{\omega_t}^2 a_g(\omega_t) (\mathbb{E}_q[\ln(p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))] - \ln p_u(\boldsymbol{\beta}_t)) + \gamma - \omega_t.$$

We can now find the natural gradient by premultiplying $\nabla_{\omega_t} \hat{\mathcal{L}}$ by the inverse of the Fisher information matrix, which for the exponential family corresponds to the inverse of the Hessian of the log-normalizer:

$$\begin{aligned} \hat{\nabla}_{\omega_t} \hat{\mathcal{L}} &= (\nabla_{\omega_t}^2 a_g(\omega_t))^{-1} \nabla_{\omega_t} \hat{\mathcal{L}} \\ &= \mathbb{E}_q[\ln(p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))] - \ln p_u(\boldsymbol{\beta}_t) + \gamma - \omega_t. \end{aligned}$$

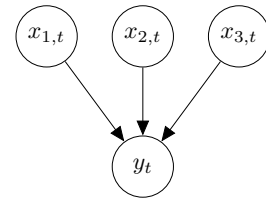
Lastly, by introducing $q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t) - q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t)$ inside the expectation we get the difference in Kullback-Leibler divergence $KL(q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t), p_u(\boldsymbol{\beta}_t)) - KL(q(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_t), p_\delta(\boldsymbol{\beta}_t | \boldsymbol{\lambda}_{t-1}))$. \square

B. Experimental Evaluation

B.1. Probabilistic Models

We provide a (simplified) graphical description of the probabilistic models used in the experiments. We also detail the distributional assumptions of the parameters, which are then used to define the variational approximation family.

ELECTRICITY MODEL



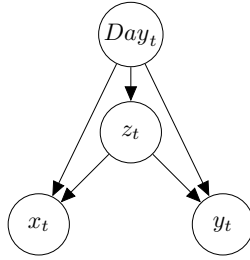
$$(\mu_i, \gamma_i) \sim \text{NormalGamma}(1, 1, 0, 1e - 10)$$

$$\gamma \sim \text{Gamma}(1, 1)$$

$$b_i \sim \mathcal{N}(0, +\infty)$$

$$x_{i,t} \sim \mathcal{N}(\mu_i, \gamma_i)$$

$$y_t \sim \mathcal{N}\left(b_0 + \sum_i b_i x_{i,t}, \gamma\right)$$



GPS MODEL

$$p \sim \text{Dirichlet}(1, \dots, 1)$$

$$p_k \sim \text{Dirichlet}(1, \dots, 1)$$

$$(\mu_{j,k}^{(x)}, \gamma_{j,k}^{(x)}) \sim \text{NormalGamma}(1, 1, 0, 1e - 10)$$

$$(\mu_{j,k}^{(y)}, \gamma_{j,k}^{(y)}) \sim \text{NormalGamma}(1, 1, 0, 1e - 10)$$

$$\text{Day}_t \sim \text{Multinomial}(p)$$

$$(z_t | \text{Day}_t = k) \sim \text{Multinomial}(p_k)$$

$$(x_t | z_t = j, \text{Day}_t = k) \sim \mathcal{N}(\mu_{j,k}^{(x)}, \gamma_{j,k}^{(x)})$$

$$(y_t | z_t = j, \text{Day}_t = k) \sim \mathcal{N}(\mu_{j,k}^{(y)}, \gamma_{j,k}^{(y)})$$

FINANCIAL MODEL

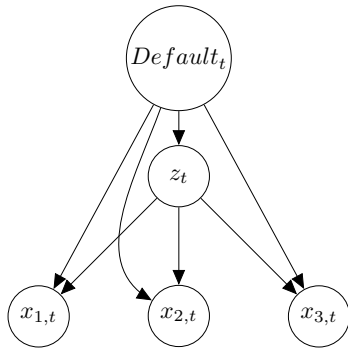


Figure 1. Simplified DAG for the financial model

$$p \sim \text{Dirichlet}(1, \dots, 1)$$

$$p_k \sim \text{Dirichlet}(1, \dots, 1)$$

$$(\mu_{i;j,k}, \gamma_{i;j,k}) \sim \text{NormalGamma}(1, 1, 0, 1e - 10)$$

$$\text{Default}_t \sim \text{Binomial}(p)$$

$$(z_t | \text{Default}_t = k) \sim \text{Multinomial}(p_k)$$

$$(x_{i,t} | z_t = j, \text{Day}_t = k) \sim \mathcal{N}(\mu_{i;j,k}, \gamma_{i;j,k})$$

B.2. Real Life Data Sets

In the experimental section of the original paper, we plot the relative values for the $TMLL_t$ measure with respect to

the SVB method. Here, we provides the plots of the absolute values of the $TMLL_t$ series for the different methods studied in the paper.

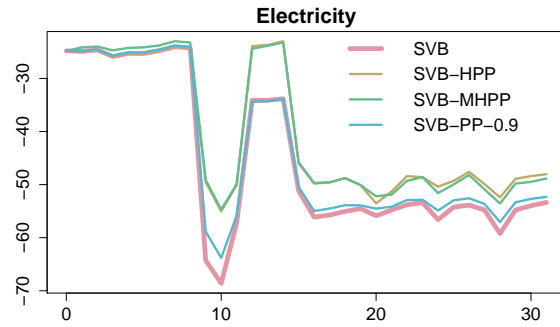


Figure 2. Electricity

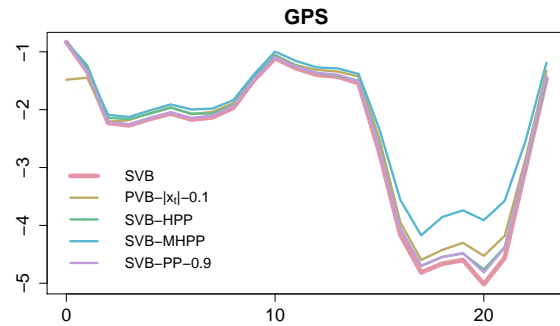


Figure 3. GPS

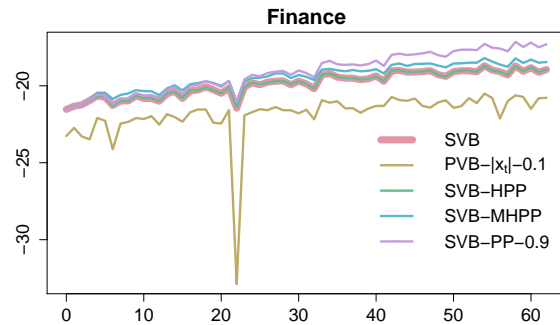


Figure 4. Finance