PROBABILITY AGGREGATION IN TIME-SERIES: DYNAMIC HIERARCHICAL MODELING OF SPARSE EXPERT BELIEFS: ONLINE SUPPLEMENT

BY VILLE A. SATOPÄÄ, SHANE T. JENSEN, BARBARA A. MELLERS, PHILIP E. TETLOCK, AND LYLE H. UNGAR

This supplementary material accompanies the paper "Probability Aggregation in Time-Series: Dynamic Hierarchical Modeling of Sparse Expert Beliefs". It provides a technical description of the sampling step of the SAC-algorithm.

1. Technical Details of the Sampling Step. The Gibbs sampler (Geman and Geman (1984)) iteratively samples all the unknown parameters from their full-conditional posterior distributions one block of parameters at a time. Given that this is performed under the constraint $b_3 = 1$ to ensure model identifiability, the constrained parameter estimates should be denoted with a trailing (1) to maintain consistency with earlier notation. For instance, the constrained estimate of γ_k should be denoted by $\hat{\gamma}_k(1)$ while the unconstrained estimate is denoted by $\hat{\gamma}_k$. For the sake of clarity, however, the constraint suffix is omitted in this section. Nonetheless, it is important to keep in mind that all the estimates in this section are constrained.

Sample $X_{t,k}$

The hidden states are sampled via the *Forward-Filtering-Backward-Sampling* (FFBS) algorithm that first predicts the hidden states using a Kalman Filter and then performs a backward sampling procedure that treats these predicted states as additional observations (see, e.g., Carter and Kohn (1994); Migon et al. (2005) for details on FFBS). More specifically, the first part, namely the Kalman Filter, is deterministic and consists of a predict and an update step. Given all the other parameters except the hidden states, the predict step for the kth question is

$$\begin{aligned} X_{t|t-1,k} &= \gamma_k X_{t-1|t-1,k} \\ P_{t|t-1,k} &= \gamma_k^2 P_{t-1|t-1,k} + \tau_k^2, \end{aligned}$$

where the initial values, $X_{0|0,k}$ and $P_{0|0,k}$, are equal to 0 and 1, respectively.

The update step is given by Algorithm 1.1. The update is repeated sequentially for each observation $Y_{i,t,k}$ given at time t. For each such repetition, the previous posterior values, $X_{t|t,k}$ and $P_{t|t,k}$, are considered as the new prior values, $X_{t|t-1,k}$

```
for i = 1, 2, ..., N_{t,k} do
    e_{i,t,k} = Y_{i,t,k} − b_{j(i)} X_{t|t−1,k}
    S_{i,t,k} = σ_k^2 + b_{j(i)}^2 P_{t|t−1,k}
    K_{i,t,k} = P_{t|t−1,k} b_{j(i)} S_{i,t,k}^{−1}
    X_{t|t,k} = X_{t|t−1,k} + K_{i,t,k} e_{i,t,k}
    P_{t|t,k} = (1 − K_{i,t,k} b_{j(i)}) P_{t|t−1,k}
    if i ≠ N_{t,k} then
        X_{t|t−1,k} = X_{t|t,k}
        P_{t|t−1,k} = P_{t|t,k}
    end if
end for
```

Algorithm 1.1: The update step of the FFBS algorithm. $N_{t,k}$ denotes the number of forecasts made at time t for question k. The subindex $j(i)$ denotes the ith expert's self-assessed expertise group.

and $P_{t|t-1,k}$. If the observation $\boldsymbol{Y}_{t,k}$ is completely missing at time t, the update step is skipped and

$$X_{t|t,k} = X_{t|t-1,k}$$
$$P_{t|t,k} = P_{t|t-1,k}$$

After running the Kalman Filter up to the final time point at $t = T_k$, the final hidden state is sampled from $X_{T_k,k} \sim \mathcal{N}(X_{T_k|T_k,k}, P_{T_k|T_k,k})$. The remaining states are obtained via the backward sampling that is performed in reverse from

$$X_{t-1,k} \sim \mathcal{N}\left(V\left(\frac{\gamma_k X_{t,k}}{\tau_k^2} + \frac{X_{t|t,k}}{P_{t|t,k}}\right), V\right),$$

where

$$V = \left(\frac{\gamma_k^2}{\tau_k^2} + \frac{1}{P_{t|t,k}}\right)^{-1}$$

This can be viewed as backward updating that considers the Kalman Filter estimates as additional observations at each given time point.

Sample $\boldsymbol{b}$ and σ_k^2

First, vectorize all the response vectors $\boldsymbol{Y}_{t,k}$ into a single vector denoted $\boldsymbol{Y}_k = \left[\boldsymbol{Y}_{1,k}^T, \ldots, \boldsymbol{Y}_{T_k,k}^T\right]^T$. Given that each $\boldsymbol{Y}_{t,k}$ is matched with $X_{t,k}$ via the time index t, we can form a $|\boldsymbol{Y}_k| \times J$ design-matrix by letting

$$\boldsymbol{X}_k = \left[(\boldsymbol{M}_k X_{1,k})^T, \ldots, (\boldsymbol{M}_k X_{T_k,k})^T\right]^T$$

Given that the goal is to borrow strength across questions by assuming a common bias vector $\boldsymbol{b}$, the parameter values must be estimated in parallel for each question such that the matrices $\boldsymbol{X}_k$ can be further concatenated into $\boldsymbol{X} = [\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_K^T]^T$ during every iteration. Similarly, $\boldsymbol{Y}_k$ must be further vectorized into a vector $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_K^T]^T$. The question-specific variance terms are taken into account by letting $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2 \mathbf{1}_{1\times T_1}, \ldots, \sigma_K^2 \mathbf{1}_{1\times T_K})$. After adopting the non-informative prior $p(\boldsymbol{b}, \sigma_k^2 | \boldsymbol{X}_k) \propto \sigma_k^{-2}$ for each $k = 1, \ldots, K$, the bias vector is sampled from

$$\boldsymbol{b}|\ldots \sim \mathcal{N}_J\left((\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}, (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\right) \tag{1.1}$$

Given that the covariance matrix in Equation (1.1) is diagonal, the identifiability constraint can be enforced after sampling a new value of $\boldsymbol{b}$ by letting $b_3 = 1$. The variance parameters are then sampled from

$$\sigma_k^2|\ldots \sim \text{Inv-}\chi^2\left(|\boldsymbol{Y}_k| - J, \frac{1}{|\boldsymbol{Y}_k| - J}(\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{b})^T(\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{b})\right),$$

where the distribution is a scaled inverse-χ^2 (see, e.g., Gelman et al. (2003)). Given that the experts are not required to give a new forecast at every time unit, the design matrices must be trimmed accordingly such that their dimensions match up with the dimensions of the observed matrices.

Sample γ_k and τ_k^2

The parameters of the hidden process are estimated via a regression setup. More specifically, after adopting the non-informative prior $p(\gamma_k, \tau_k^2 | \boldsymbol{X}_k) \propto \tau_k^{-2}$, the parameter values are sampled from

$$\begin{aligned}
\gamma_k|\ldots &\sim \mathcal{N}\left(\frac{\sum_{t=2}^{T_k} X_{t,k}X_{t-1,k}}{\sum_{t=1}^{T_k-1} X_{t,k}^2}, \frac{\tau_k^2}{\sum_{t=1}^{T_k-1} X_{t,k}^2}\right)\\
\tau_k^2|\ldots &\sim \text{Inv-}\chi^2\left(T_k - 1, \frac{1}{T_k - 1}\sum_{t=2}^{T_k}(X_{t,k} - \gamma_k X_{t-1,k})^2\right),
\end{aligned}$$

where the final distribution is a scaled inverse-χ^2 (see, e.g., Gelman et al. (2003)).

References.

CARTER, C. K. and KOHN, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika* **81** 541–553.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2003). *Bayesian Data Analysis*. CRC press.

GEMAN, S. and GEMAN, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Institute of Electrical and Electronics Engineer (IEEE) Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

MIGON, H. S., GAMERMAN, D., LOPES, H. F. and FERREIRA, M. A. (2005). Dynamic Models. *Handbook of Statistics* **25** 553–588.