UNIVERSITY OF TEXAS ARLINGTON

Bayesian Hierarchical Dynamic Factor Models

Anthony M. Thomas, Jr.
anthony.thomas3@mavs.uta.edu

Department of Mathematics
The University of Texas at Arlington

August 21, 2020

Contents

Section 1

Bayesian Inference

Bayesian Inference can be described by two parts:

1. Build a model based on data $\boldsymbol{X}$ and parameters $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2\}$
   - Likelihood: $p(\boldsymbol{X}|\boldsymbol{\Theta})$
   - Prior: $p(\boldsymbol{\Theta})$
2. Compute the posterior
   - Posterior:
   $$p(\boldsymbol{\Theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\boldsymbol{X})}$$
   - Can report summaries, e.g. posterior expectations
   $$\mathbb{E}[h(\boldsymbol{\Theta})|\boldsymbol{X}]$$
   and compute marginal posteriors
   $$p(\boldsymbol{\Theta}_1|\boldsymbol{X}) \text{ and } p(\boldsymbol{\Theta}_2|\boldsymbol{X})$$

Factor Analysis

Factor Analysis (FA) is a method that assumes that the covariance structure of a set of observations can be described in terms of a linear combination of latent variables called factors.

Typical uses of FA:

1. Dimension reduction: explains covariation between N variables using $K < N$ factors
2. Data interpretation: finds factors that explain the covariation
3. Theory testing: tests whether a hypothesized factor structure fits observed data

Dynamic Factor Analysis

Section 2

Hierarchical Dynamic Factor Model

$$\mathbf{X}_{bst} = \mathbf{\Lambda}_{H.bs}(L)\mathbf{H}_{bst} + \mathbf{e}_{Xbst} \tag{1}$$

$$\mathbf{H}_{bst} = \mathbf{\Lambda}_{G.bs}(L)\mathbf{G}_{bt} + \mathbf{e}_{Hbst} \tag{2}$$

$$\mathbf{G}_{bt} = \mathbf{\Lambda}_{F.b}(L)\mathbf{F}_{t} + \mathbf{e}_{Gbt} \tag{3}$$

$$\mathbf{\Psi}_{F}(L)\mathbf{F}_{t} = \boldsymbol{\epsilon}_{Ft}, \tag{4}$$

Section 3

Why Variational Bayesian Inference?

Consider a model with data $\boldsymbol{X}$ and latent variables $\boldsymbol{Z} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2\}$. The goal is to compute the joint posterior of the latent variables given the data

$$p(\boldsymbol{Z}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{Z})p(\boldsymbol{Z})}{p(\boldsymbol{X})} \tag{5}$$

- Likelihood: $p(\boldsymbol{X}|\boldsymbol{Z})$
- Prior: $p(\boldsymbol{Z})$
- Evidence: $p(\boldsymbol{X})$

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X}, \boldsymbol{Z})\, d\boldsymbol{Z} \tag{6}$$

Why Variational Bayesian Inference?

For complex models (5) and (6) typically have no closed-form. As a result, the joint and marginal posteriors have to be approximated. Markov chain Monte Carlo (MCMC) has been the gold standard to solve this problem.

1. Construct an ergodic Markov chain on $\boldsymbol{Z}$ whose stationary distribution is the joint posterior $p(\boldsymbol{Z}|\boldsymbol{X})$
2. Sample from the chain to collect samples from the stationary distribution
3. Approximate the posterior with an empirical estimate constructed from a subset of the collected samples
4. Use the subset of collected samples to estimate expectations of interest

Why Variational Bayesian Inference?

A major disadvantage for MCMC is that high levels of accuracy requires a great deal of time

1. MCMC is typically slow, but ultimately accurate
2. VB is typically much faster

Variational Bayesian Inference

Rather than use sampling, VB uses optimization to find an approximation to the posterior.

1. Posit a family of "nice" approximate densities $\mathcal{Q}$
2. Find a member of that family that is "closest" to the exact posterior, i.e.

$$q^\star(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z})\in\mathcal{Q}}{\arg\min}\, \mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) \tag{7}$$

where

$$\mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \mathbb{E}_q[\log q(\boldsymbol{Z})] - \mathbb{E}_q[\log p(\boldsymbol{Z}|\boldsymbol{X})] \tag{8}$$

Variational Bayesian Inference

It follows that

$$\mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \mathbb{E}_q[\log q(\boldsymbol{Z})] - \mathbb{E}_q[\log p(\boldsymbol{X}, \boldsymbol{Z})] + \log p(\boldsymbol{X})$$

This reveals that the objective in (7) depends on the evidence, thus it cannot be computed directly.

Variational Bayesian Inference

Instead, we optimize an alternative objective that is equivalent to (7) up to an added constant called the **evidence lower bound** (ELBO)

$$\text{ELBO}(q) = \mathbb{E}_q[\log p(\boldsymbol{X}, \boldsymbol{Z})] - \mathbb{E}_q[\log q(\boldsymbol{Z})] \tag{9}$$

It can be shown that

$$\arg\min_{q(\boldsymbol{Z}) \in \mathcal{Q}} \text{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \arg\max_{q(\boldsymbol{Z}) \in \mathcal{Q}} \text{ELBO}(q)$$

Variational Bayesian Inference

The VB framework is now

1. Posit a family of "nice" approximate densities $\mathcal{Q}$
2. Find a member of that family that is "closest" to the exact posterior, i.e.

$$q^\star(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z})\in\mathcal{Q}}{\arg\max}\, \mathrm{ELBO}(q) \tag{10}$$

Solving this optimization problem is still difficult in general

- Using the mean-field assumption can make it easier

Mean-Field Assumption

1. Partition the latent variables into M groups, say $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$
2. Assume that the distributions in $\mathcal{Q}$ factorize across the groups, i.e.

$$\mathcal{Q} = \left\{ q : q(\boldsymbol{Z}) = \prod_{m=1}^{M} q_m(\boldsymbol{Z}_m) \right\}$$

3. Learning the optimal q now reduces to learning the optimal $q_1, \ldots, q_M$
4. Straightforward to optimize via coordinate ascent
5. This is **NOT** a modeling assumption

Mean-Field Assumption

Interestingly, under the mean-field assumption, the optimization problem for a single q_m has the solution:

$$q_m(\boldsymbol{Z}_m) = \frac{\exp\left\{\mathbb{E}_{q_{-m}}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\right\}}{\int \exp\left\{\mathbb{E}_{q_{-m}}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\right\} d\boldsymbol{Z}_m} \tag{11}$$

This establishes what is called the **coordinate ascent variational inference** algorithm.

Coordinate Ascent Variational Inference

Algorithm 1: Coordinate ascent variational inference

Input: Model $p(\boldsymbol{X}, \boldsymbol{Z})$, Data $\boldsymbol{X}$,
Output: Variational density $q(\boldsymbol{Z}) = \prod_{m=1}^{M} q_m(\boldsymbol{Z}_m)$
Initialize: Variational densities $q_m(\boldsymbol{Z}_m)$
while *the ELBO has not converged* **do**
  for $m \in \{1, 2, \ldots, M\}$ **do**
    Set $q_m(\boldsymbol{Z}_m) \propto \exp\{\mathbb{E}_{q_{-m}}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\}$
  end
  Compute ELBO(q)
end
return $q(\boldsymbol{Z})$

The Orthogonal Factor Model

The orthogonal FA model assumes the form

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{F} + \boldsymbol{e}$$

where

1. $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ denotes the vector of observations
2. $\boldsymbol{\Lambda} = [\lambda_{nk}]_{N \times K}$ denotes the matrix of factor loadings
3. $\boldsymbol{F} = (F_1, \ldots, F_K)^\top$ denotes the vector of latent factors
4. $\boldsymbol{e} = (e_1, \ldots, e_N)^\top$ denotes the vector of latent error terms

Assumptions of the Orthogonal Factor Model

The orthogonal FA model assumes the following:

1. $\boldsymbol{F} \sim \mathcal{N}(\mathbf{0}_K, \mathbf{I_K})$
2. $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$
3. F_k and e_n are independent for every pair k, n

where $\mathbf{0}_K$ and $\mathbf{0}_N$ are zero-vectors of lengths K and N, respectively, and $\mathbf{I}_K$ is the $K \times K$ identity matrix.