

# Bayesian Hierarchical Dynamic Factor Models

Anthony M. Thomas, Jr.  
[anthony.thomas3@mavs.uta.edu](mailto:anthony.thomas3@mavs.uta.edu)

Department of Mathematics  
The University of Texas at Arlington

August 17, 2020

# Contents

- ① Background
- ② Variational Bayesian Inference
- ③ Hierarchical Dynamic Factor Analysis
  - Classical Factor Analysis
  - Dynamic Factor Analysis

# Section 1

- 1 Background
- 2 Variational Bayesian Inference
- 3 Hierarchical Dynamic Factor Analysis
  - Classical Factor Analysis
  - Dynamic Factor Analysis

# Factor Analysis

- ① **Factor Analysis** is a method that uses the covariances between a set of observed variables to described them in terms of a smaller set of unobservable variables called factors.

# Bayesian Inference

**Bayesian Inference** can be described by two parts:

- ① Build a model based on data  $\mathbf{X}$  and parameters  $\Theta = \{\Theta_1, \Theta_2\}$ 
  - ▶ Likelihood:  $p(\mathbf{X}|\Theta)$
  - ▶ Prior:  $p(\Theta)$
- ② Compute the posterior
  - ▶ Posterior:

$$p(\Theta|\mathbf{X}) = \frac{p(\mathbf{X}|\Theta)p(\Theta)}{p(\mathbf{X})}$$

- ▶ Can report summaries, e.g. posterior expectations

$$\mathbb{E}[h(\Theta)|\mathbf{X}]$$

and compute marginal posteriors

$$p(\Theta_1|\mathbf{X}) \text{ and } p(\Theta_2|\mathbf{X})$$

# Section 2

- 1 Background
- 2 Variational Bayesian Inference

- 3 Hierarchical Dynamic Factor Analysis  
Classical Factor Analysis  
Dynamic Factor Analysis

# Why Variational Bayesian Inference?

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

$$p(\mathbf{Z}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z})}{p(\mathbf{X})} \quad (1)$$

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

$$p(\mathbf{X}) = \int p(\mathbf{X}, \mathbf{Z}) d\mathbf{Z} \quad (2)$$

# Why Variational Bayesian Inference?

For complex models (1) and (2) 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  $\mathbf{Z}$  whose stationary distribution is the joint posterior  $p(\mathbf{Z}|\mathbf{X})$
- 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.

$$q^*(\mathbf{Z}) = \arg \min_{q(\mathbf{Z}) \in \mathcal{Q}} \text{KL}(q(\mathbf{Z}) \parallel p(\mathbf{Z} | \mathbf{X})) \quad (3)$$

where

$$\text{KL}(q(\mathbf{Z}) \parallel p(\mathbf{Z} | \mathbf{X})) = \mathbb{E}_q[\log q(\mathbf{Z})] - \mathbb{E}_q[\log p(\mathbf{Z} | \mathbf{X})] \quad (4)$$

# Variational Bayesian Inference

It can be shown that

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

This reveals that the objective in (3) depends on the evidence, thus it cannot be computed directly.

# Variational Bayesian Inference

Instead, we optimize an alternative objective that is equivalent to (3) up to an added constant called the **evidence lower bound** (ELBO)

$$\begin{aligned}\text{ELBO}(q) &= \mathbb{E}_q[\log p(\mathbf{X}, \mathbf{Z})] - \mathbb{E}_q[\log q(\mathbf{Z})] \\ &= \mathbb{E}_q\left[\log \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})}\right]\end{aligned}\tag{5}$$

It can be shown that

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

# 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.

$$q^*(\mathbf{Z}) = \arg \max_{q(\mathbf{Z}) \in \mathcal{Q}} \text{ELBO}(q) \quad (6)$$

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  $\mathbf{Z}_1, \dots, \mathbf{Z}_M$
- ② Assume that the distributions in  $\mathcal{Q}$  factorize across the groups,  
i.e.

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

- ③ Learning the optimal  $q$  now reduces to learning the optimal  $q_1, \dots, q_M$
- ④ Straightforward to optimize via coordinate descent
- 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:

$$q_m(\mathbf{Z}_m) = \frac{\exp\{\mathbb{E}_{q_{-m}}[\log p(\mathbf{X}, \mathbf{Z})]\}}{\int \exp\{\mathbb{E}_{q_{-m}}[\log p(\mathbf{X}, \mathbf{Z})]\} d\mathbf{Z}_m} \quad (7)$$

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

# Coordinate Ascent Variational Inference

---

**Algorithm 1:** Coordinate ascent variational inference

---

**Input:** Model  $p(\mathbf{X}, \mathbf{Z})$ , Data  $\mathbf{X}$ ,

**Output:** Variational density  $q(\mathbf{Z}) = \prod_{m=1}^M q_m(\mathbf{Z}_m)$

**Initialize:** Variational densities  $q_m(\mathbf{Z}_m)$

**while** the ELBO has not converged **do**

**for**  $m \in \{1, 2, \dots, M\}$  **do**

        | Set  $q_m(\mathbf{Z}_m) \propto \exp\{\mathbb{E}_{q_{-m}}[\log p(\mathbf{X}, \mathbf{Z})]\}$

    | **end**

    | Compute ELBO( $q$ )

**end**

**return**  $q(\mathbf{Z})$

---

# Section 3

- 1 Background
- 2 Variational Bayesian Inference
- 3 Hierarchical Dynamic Factor Analysis
  - Classical Factor Analysis
  - Dynamic Factor Analysis

# Classical Factor Analysis Model

The Classical (orthogonal) FA model assumes assumes the form

$$\mathbf{X} = \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon}$$

where

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

# Normal Theory Assumptions

The Normal Theory Classical FA model assumes the form

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

and adds the assumptions that

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

# Hierarchical Dynamic Factor Model

$$\mathbf{X}_{bst} = \boldsymbol{\Lambda}_{H.bst}(L) \mathbf{H}_{bst} + \mathbf{e}_{Xbst} \quad (8)$$

$$\mathbf{H}_{bst} = \boldsymbol{\Lambda}_{G.bst}(L) \mathbf{G}_{bt} + \mathbf{e}_{Hbst} \quad (9)$$

$$\mathbf{G}_{bt} = \boldsymbol{\Lambda}_{F.b}(L) \mathbf{F}_t + \mathbf{e}_{Gbt} \quad (10)$$

$$\boldsymbol{\Psi}_F(L) \mathbf{F}_t = \boldsymbol{\epsilon}_{Ft}, \quad (11)$$