Bayesian Hierarchical Dynamic Factor Models

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

Department of Mathematics
The University of Texas at Arlington

August 17, 2020

Contents

Section 1

1. Background
2. Variational Bayesian Inference
3. Classical Factor Analysis
4. Hierarchical Dynamic 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 $\boldsymbol{X}$ 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
   $$\mathrm{E}[h(\boldsymbol{\Theta})|\boldsymbol{X}]$$
   and compute marginal posteriors
   $$p(\boldsymbol{\Theta}_1|\boldsymbol{X}) \text{ and } p(\boldsymbol{\Theta}_2|\boldsymbol{X})$$

Section 2

Why Variational Bayesian Inference?

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

$$p(\boldsymbol{Z}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{Z})p(\boldsymbol{Z})}{p(\boldsymbol{X})} \tag{1}$$

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

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X}, \boldsymbol{Z})\, d\boldsymbol{Z} \tag{2}$$

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 $\boldsymbol{Z}$ 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.

$$q^{\star}(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z}) \in \mathcal{Q}}{\arg\min}\, \mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) \tag{3}$$

where

$$\mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \mathrm{E}_q[\log q(\boldsymbol{Z})] - \mathrm{E}_q[\log p(\boldsymbol{Z}|\boldsymbol{X})] \tag{4}$$

Variational Bayesian Inference

It can be shown that

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

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)

$$\mathrm{ELBO}(q) = \mathrm{E}_q[\log p(\boldsymbol{X}, \boldsymbol{Z})] - \mathrm{E}_q[\log q(\boldsymbol{Z})] \tag{5}$$

Maximizing the ELBO is equivalent to (3).

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.

$$q^{\star}(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z}) \in \mathcal{Q}}{\arg\max}\, \mathrm{ELBO}(q) \tag{6}$$

Mean-Field Assumption

Section 3

Classical Factor Analysis Model

The Classical (orthogonal) FA model assumes 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

Normal Theory Assumptions

The Normal Theory Classical FA model assumes the form

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

and adds the assumptions that

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

Section 4

Hierarchical Dynamic Factor Model

$$\mathbf{X}_{bst} = \mathbf{\Lambda}_{H.bs}(L)\mathbf{H}_{bst} + \mathbf{e}_{Xbst} \tag{7}$$

$$\mathbf{H}_{bst} = \mathbf{\Lambda}_{G.bs}(L)\mathbf{G}_{bt} + \mathbf{e}_{Hbst} \tag{8}$$

$$\mathbf{G}_{bt} = \mathbf{\Lambda}_{F.b}(L)\mathbf{F}_{t} + \mathbf{e}_{Gbt} \tag{9}$$

$$\mathbf{\Psi}_{F}(L)\mathbf{F}_{t} = \boldsymbol{\epsilon}_{Ft}, \tag{10}$$