Bayesian Inference for Latent States

Simon Jackman

Stanford University
http://jackman.stanford.edu/BASS

February 3, 2012

Chapter 9 of *BASS*

- factor analysis
- item-response theory (IRT) models
- dynamic linear model

Inference for Latent States

- latent quantities $\boldsymbol{\xi}$
- observed quantities $\mathbf{Y}$
- unobserved parameters $\boldsymbol{\beta}$, indexing the functional relationship between $\mathbf{Y}$ and $\boldsymbol{\xi}$
- $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\beta})'$
- Bayesian analysis:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

Factor Analysis

- factor analysis typically presented as a model for *covariance structure*:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$$

  where $\boldsymbol{\Lambda}$ is a p-by-k matrix of factor loadings, $\boldsymbol{\Phi} = \mathbf{I}_k$ and $\boldsymbol{\Psi}$ is a diagonal p-by-p matrix with "uniquenesses" on the diagonal.
- obscures the fact that factor analysis is a model for observables conditional on unobservables
- at the level of the indicators, a Gaussian measurement model is

$$y_{ij} \sim N(\gamma_{j0} + \gamma_{j1}\xi_i, \omega_j^2)$$

  where i indexes observations, j indexes p indicators, γ_{j0} and γ_{j1} are intercept and slope parameters, ω^2 is a measurement error variance and ξ_i are latent states.

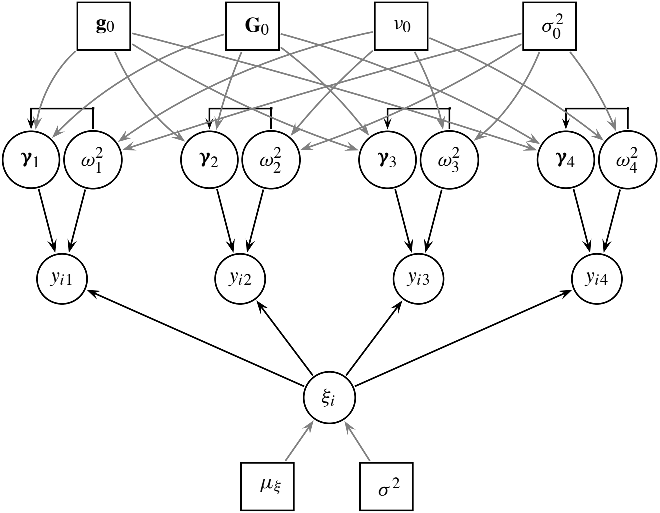
Conjugate prior densities for Gaussian factor analysis model

$$
\begin{aligned}
\xi_i &\stackrel{\text{iid}}{\sim} N(\mu_\xi, \sigma^2), \quad i = 1, \ldots, n, \\
\boldsymbol{\gamma}_j | \omega_j^2 &\sim N(\mathbf{g}_{j0}, \omega_j^2 \mathbf{G}_{j0}), \quad j = 1, \ldots, p, \\
\omega_j^2 &\sim \text{inverse-Gamma}(\nu_{j0}/2, \nu_{j0}\omega_{j0}^2/2), \quad j = 1, \ldots, p,
\end{aligned}
$$

where $\mu_\xi, \sigma^2, \mathbf{g}_{j0}, \mathbf{G}_{j0}, \nu_{j0}$ and $\omega_{j0}^2, j = 1, \ldots, p$ are user-specified hyper-parameters.

Factor Analysis in Terms of Latent Variables

DAG, four indicator model, suggests Gibbs sampling scheme etc



Identification

- model parameters not identified
- location shifts in ξ_i can be offset by shifts in intercepts γ_{j0}.
- scale shifts in ξ_i can be offset by rescaling slopes γ_{j1}.
- scale shifts in ω_j can be offset with re-scalings of $\gamma_{j0}, \gamma_{j1}, \xi_i$.
- lack of identification not a formal problem for Bayesian analysis
- nonetheless, we deal with by imposing a location/scale restriction ("normalization") on the ξ_i: mean zero, standard deviation one.

Posterior inference via Gibbs sampling

$$\xi_i | \mu_\xi, \sigma^2, \mathbf{\Gamma}, \boldsymbol{\psi}, \mathbf{y}_i \sim N(\mu_\xi^*, \sigma^{2*})$$

where

$$\mu_\xi^* = \frac{\frac{\mu_\xi}{\sigma^2} + \frac{\hat{\xi}_i}{V(\hat{\xi}_i)}}{\frac{1}{\sigma^2} + \frac{1}{V(\hat{\xi}_i)}} \quad \text{and} \quad \sigma^{2*} = \frac{\omega_j^2}{\frac{1}{\sigma^2} + \frac{1}{V(\hat{\xi}_i)}}$$

and where

$$\begin{aligned}
\hat{\xi}_i &= (\boldsymbol{\gamma}_1' \boldsymbol{\gamma}_1)^{-1} \mathbf{w}_i' \boldsymbol{\gamma}_1 = \sum_{j=1}^{p} \gamma_{j1}^2 / \sum_{j=1}^{p} w_{ij} \gamma_{j1} \quad \text{and} \\
V(\hat{\xi}_i) &= \omega_j^2 (\boldsymbol{\gamma}_1' \boldsymbol{\gamma}_1)^{-1} = \frac{\omega_j^2}{\sum_{j=1}^{p} \gamma_{j1}^2},
\end{aligned}$$

Posterior inference via Gibbs sampling

$$\boldsymbol{\gamma}_j | \mathbf{g}_{j0}, \mathbf{G}_{j0}, \boldsymbol{\xi}, \omega_j^2, \mathbf{y}_j \sim N(\mathbf{g}_{j1}, \omega_j^2 \mathbf{G}_{j1}),$$

where

$$\begin{aligned}
\mathbf{g}_{j1} &= (\mathbf{G}_{j0}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{G}_{j0}^{-1}\mathbf{g}_{j0} + \mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\gamma}}_j), \\
\mathbf{G}_{j1} &= (\mathbf{G}_{j0}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1}, \\
\mathbf{Z} &= [\boldsymbol{\iota}\ \boldsymbol{\xi}] \quad \text{and} \\
\hat{\boldsymbol{\gamma}}_j &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.
\end{aligned}$$

That is, $\mathbf{Z}$ is the n-by-2 matrix formed with a unit vector $\boldsymbol{\iota}$ in the first column and the $\boldsymbol{\xi}$ in the second column (a regressor matrix for the purposes of inference for $\boldsymbol{\gamma}_j$).

Posterior inference via Gibbs sampling

$$\omega_j^2 \,|\, \nu_{j0}, \sigma_{j0}^2, \boldsymbol{\gamma}_j, \boldsymbol{\xi}, \mathbf{y}_j \sim \text{inverse-Gamma}(\nu_1/2, \nu_1\sigma_1^2/2),$$

where

$$\begin{aligned}
\nu_1 &= \nu_0 + n, \\
\nu_1\sigma_1^2 &= \nu_0\sigma_0^2 + S_j + r_j, \\
S_j &= (\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}_j)'(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}_j) \quad \text{and} \\
r_j &= (\mathbf{g}_{j0} - \hat{\boldsymbol{\gamma}}_j)'(\mathbf{G}_{j0} + (\mathbf{Z}'\mathbf{Z})^{-1})^{-1}(\mathbf{g}_{j0} - \hat{\boldsymbol{\gamma}}_j).
\end{aligned}$$

References