Measurement Models

Yuki Shiraito

POLSCI 798 Advanced Topics in Quantitative Methodology
University of Michigan

Winter 2019

Item Response Theory Models

- Measurement models
  - Models to measure *theoretical concepts*, e.g., ability, ideology
  - Measurement of an unobserved concept: Latent variable
  - Abstraction of data: Dimension reduction
- Two-parameter item response theory (IRT) model
  - Context: Measuring ability using educational tests, e.g., SAT, GRE
  - Students $i = 1, \ldots, N$, items (questions) $j = 1, \ldots J$
  - Response Y_{ij}: 1 correct, 0 incorrect
  - Model: $\mathbb{P}\left(Y_{ij} = 1\right) = \Phi\left(\beta_j \lambda_i - \alpha_j\right)$
  - Latent variable representation

$$Y_{ij} = \begin{cases} 0 & (U_{ij} \leq 0) \\ 1 & (U_{ij} > 0) \end{cases}$$

$$U_{ij} = -\alpha_j + \beta_j \lambda_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, 1\right)$$

  - Parameters
    - λ_i: *Ability* of student i
    - α_j: *Difficulty* of item j
    - β_j: *Discrimination* of item j (usually $\beta_j > 0$)

Spatial Voting Interpretation of the IRT Model

- Political science interpretation of IRT: Ideal point estimation
- Spatial voting model with random utility:
  - Euclidean *policy space*, $\mathbb{R}$
  - Legislator i's *ideal point*: $\lambda_i \in \mathbb{R}$
  - *Policy proposal (bill)* j: $\zeta_j \in \mathbb{R}$
  - *Status quo* against proposal j: $\psi_j \in \mathbb{R}$
  - i's *utility*:
$$U_i(\zeta_j) = -\left(\lambda_i - \zeta_j\right)^2 + \eta_{ij}, \quad \eta_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_j^2)$$
$$U_i(\psi_j) = -\left(\lambda_i - \psi_j\right)^2 + \nu_{ij}, \quad \nu_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_j^2)$$
  - i votes "Yea" if $U_i(\zeta_j) > U_i(\psi_j)$ and "Nay" if $U_i(\zeta_j) \leq U_i(\psi_j)$
  - Probability of voting "Yea" ($Y_{ij} = 1$):
$$\mathbb{P}\left(Y_{ij} = 1\right) = \mathbb{P}\left(U_i(\zeta_j) > U_i(\psi_j)\right) = \Phi\left(\beta_j \lambda_i - a_j\right)$$
$$\beta_j = \frac{2\zeta_j - 2\psi_j}{\sqrt{2\sigma_j^2}} \qquad a_j = \frac{\zeta_j^2 - \psi_j^2}{\sqrt{2\sigma_j^2}}$$
- Ideal points estimated as ability parameters in IRT model

Gibbs Sampler for the Two-Parameter IRT Model

- Gibbs sampler for ideal point estimation
  - Data: $N \times J$ matrix of binary variables
  - Augmented variable: $U_{ij} = (v_{ij} - \eta_{ij})/\sqrt{2\sigma_j^2}$
  - Gibbs steps for $U_{ij}, \beta_j, \alpha_j$, and λ_j: Probit and iterative regression!
- Joint posterior:

$$p(\alpha, \beta, \lambda, \mathrm{U} \mid \mathrm{Y}) \propto \underbrace{p(\alpha, \beta, \lambda)}_{\text{Prior}} \prod_{i=1}^{N} \prod_{j=1}^{J} \Big(1\{U_{ij} > 0\} 1\{Y_{ij} = 1\} + 1\{U_{ij} \leq 0\} 1\{Y_{ij} = 0\} \Big) e^{-\frac{(U_{ij} - \beta_j \lambda_i + \alpha_j)^2}{2}}$$

- Conditional posterior of U_{ij}:

$$p(\mathrm{U} \mid \alpha, \beta, \lambda, \mathrm{Y}) \propto \prod_{i=1}^{N} \prod_{j=1}^{J} \Big(1\{U_{ij} > 0\} 1\{Y_{ij} = 1\} + 1\{U_{ij} \leq 0\} 1\{Y_{ij} = 0\} \Big) e^{-\frac{(U_{ij} - \beta_j \lambda_i + \alpha_j)^2}{2}}$$

  - Identical to the Gibbs step for latent response in Probit
  - Independent truncated Gaussians with $\mu = -\alpha_i^{(s-1)} + \beta^{(s-1)} \lambda_i^{(s-1)}$

- Conditional posterior of λ_i with independent priors:

$$p(\lambda \mid \alpha, \beta, \mathrm{U}, \mathrm{Y}) \propto \prod_{i=1}^{N} p(\lambda_i \mid \alpha, \beta) \prod_{j=1}^{J} e^{-\frac{\left((U_{ij}+\alpha_j)-\beta_j\lambda_i\right)^2}{2}}$$

  - Independent Bayesian regressions across i
    1. Units = votes: $j = 1, \ldots, J$
    2. "Response": $U_{ij}^{(s)} - \alpha_j^{(s-1)}$
    3. "Regressor": $\beta_j^{(s-1)}$, no intercept
    4. "Coefficient": $\lambda_i^{(s)}$

- Conditional posterior of α_j and β_j with independent priors:

$$p(\alpha, \beta \mid \lambda, \mathrm{U}, \mathrm{Y}) \propto \prod_{j=1}^{J} p(\alpha_j, \beta_j \mid \lambda) \prod_{i=1}^{N} e^{-\frac{\left(U_{ij}-(-\alpha_j+\beta_j\lambda_i)\right)^2}{2}}$$

  - Independent Bayesian regressions across j
    1. Units = legislators: $i = 1, \ldots, N$
    2. "Response": $U_{ij}^{(s)}$
    3. "Regressor": $(-1, \lambda_i^{(s)})$
    4. "Coefficients": $(\alpha_j^{(s)}, \beta_j^{(s)})$

Prior and Identification

- Conditionally conjugate, independent prior on α_j and β_j:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$$

  - Σ: Usually diagonal
  - Prior variance relative to $\mathbb{V}(\varepsilon_{ij}) = \mathbb{V}\left((v_{ij} - \eta_{ij}) \Big/ \sqrt{2\sigma_j^2}\right) = 1$
- Conditionally conjugate, independent prior on λ_i:

$$\lambda_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\lambda^2)$$

  - Again, σ_λ relative to 1
- Identification: Scale and direction of the policy space
  - Scale: Everyone's ideal point multiplied by a constant
  - Direction: Conservative-liberal or liberal-conservative?
- Identification constraits:
  - Constrained prior distributions
    1. λ_i has mean zero and variance one across legislators, $\sigma_\lambda = 1$
    2. At least one β_j is constrained to be positive (negative)
  - Fixed values of two legislators, e.g., $\lambda_{\text{Sanders}} = -1, \lambda_{\text{Cruz}} = 1$

Need for Speed

TABLE 1. Recent Applications of Ideal Point Models to Various Large Data Sets

	Number of Subjects	Number of Items	Data Types
Parliaments			
DW-NOMINATE scores (1789–2012)	37,511	46,379	U.S. Congress
Common Space scores (1789–2012)	11,833	90,609	U.S. Congress
Hix, Noury, and Roland (2006)	2,000	12,000	European Parliament
Shor and McCarty (2011)	6,201	5,747	U.S. state legislatures
Bailey, Strezhnev, and Voeten (2015)	2,187	7,335	United Nations
Courts (and other institutions)			
Martin and Quinn scores (1937–2013)	697	5,164	U.S. Supreme Court
Bailey (2007)	27,795	2,750	US. Supreme Court, Congress, Presidents
Voters (and politicians)			
Gerber and Lewis (2004)	2.8 million	12	referendum votes
Bafumi and Herron (2010)	8,848	4,391	survey & roll calls
Tausanovitch and Warshaw (2013)	275,000	311	survey
Bonica (2014)	4.2 million	78,363	campaign contributions
Social Media			
Bond and Messing (2015)	6.2 million	1,223	Facebook
Barberá (2015)	40.2 million	1,465	Twitter
Texts and Speeches			
Clark and Lauderdale (2010)	1,000	1,000	U.S. Supreme Court citations
Proksch and Slapin (2010)	25	8995	German manifestos
Lewandowski et al. (2015)	1,000	250,000	European party manifestos

Notes: The past decade has witnessed a significant rise in the use of large data sets for ideal point estimation. Note that "# of subjects" should be interpreted as the number of ideal points to be estimated. For example, if a legislator serves for two terms and are allowed to have different ideal points in those terms, then this legislator is counted as two subjects.

Imai, Lo, and Olmsted (2016)

- Parties, legislators, judges, campaign donors, SNS accounts, ...
- Large J and large N in some applications
- $N + 2J$ parameters and $N \times J$ augmented variables

$\rightsquigarrow$ Need for speed

The EM Algorithm

- Expectation-maximization algorithm
  - Iterative algorithm to find *maximum a posteriori* (MAP) estimates
  - Pros: Generally faster than MCMC
  - Cons: No estimates of posterior uncertainty
- General setup
  - Model: $p(\text{Data} \mid Z, \theta)$ with *unobserved* Z such as latent variables, data augmentation, missing data, or nuisance parameters
  - Posterior density: $p(\theta \mid \text{Data}) = \int p(\theta, z \mid \text{Data}) dz$
  - Algorithm:

    $s = 0$ Set initial values, $\theta^{(0)}$

    $s = 1, 2, \ldots$ Repeat until convergence of $\theta^{(s)}$

    1. E-step: Compute the "Q-function" in a closed form
    $$Q(\theta \mid \theta^{(s-1)}) \equiv \int p(z \mid \theta^{(s-1)}, \text{Data}) \log p(\theta, z \mid \text{Data}) dz$$
    2. M-step: Maximize Q w.r.t. θ with $p(z \mid \theta^{(s-1)}, \text{Data})$ fixed
    $$\theta^{(s)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta \mid \theta^{(s-1)})$$
  - Convergence: $\theta^{(s)} \to \widehat{\theta}_{\text{MAP}} \equiv \operatorname{argmax}_\theta \log p(\theta \mid \text{Data})$

Monotone Convergence Property

- Why converges to the posterior mode?
- Jensen's inequality:

$$\log p(\theta \mid \text{Data}) = \log \int \frac{p(\theta, z \mid \text{Data})}{p(z \mid \underline{\theta}, \text{Data})} p(z \mid \underline{\theta}, \text{Data}) dz$$
$$\geq \mathbb{E}_{p(Z \mid \underline{\theta}, \text{Data})} \left[\log \frac{p(\theta, Z \mid \text{Data})}{p(Z \mid \underline{\theta}, \text{Data})} \right]$$

- Alternating maximization of the lower bound (LB) w.r.t. $\underline{\theta}$ and θ:
  1. Given θ, LB attains its maximum when $\underline{\theta} = \theta$
  $$\mathbb{E}_{p(Z \mid \underline{\theta} = \theta, \text{Data})} \left[\log \frac{p(\theta, Z \mid \text{Data})}{p(Z \mid \underline{\theta} = \theta, \text{Data})} \right] = \log p(\theta \mid \text{Data})$$
  2. Given $\underline{\theta}$, maximizing LB w.r.t. θ is equivalent to maximizing Q
  $$\mathbb{E}_{p(Z \mid \underline{\theta}, \text{Data})} \left[\log \frac{p(\theta, Z \mid \text{Data})}{p(Z \mid \underline{\theta}, \text{Data})} \right]$$
  $$= \underbrace{\mathbb{E}_{p(Z \mid \underline{\theta}, \text{Data})} [\log p(\theta, Z \mid \text{Data})]}_{= Q(\theta \mid \underline{\theta})} - \underbrace{\mathbb{E}_{p(Z \mid \underline{\theta}, \text{Data})} [\log p(Z \mid \underline{\theta}, \text{Data})]}_{\text{Constant w.r.t. } \theta}$$
- Monotone convergence: $\log p(\theta^{(s)} \mid \text{Data}) > \log p(\theta^{(s-1)} \mid \text{Data})$

EM Algorithm for the Two-Parameter IRT Model

- EM is useful when
  1. Q is easy to evaluate
  2. Q is easy to maximize
- Recall the joint posterior density:

$$p(\alpha, \beta, \lambda, \mathrm{U} \mid \mathrm{Y}) \propto \underbrace{e^{-\frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i^2 + \sum_{j=1}^{J}\left(\frac{\alpha_j^2}{\sigma_\alpha^2} + \frac{\beta_j^2}{\sigma_\beta^2}\right)\right)}}_{\text{Prior}}$$

$$\prod_{i=1}^{N}\prod_{j=1}^{J}\left(1\{U_{ij} > 0\}1\{Y_{ij} = 1\} + 1\{U_{ij} \leq 0\}1\{Y_{ij} = 0\}\right) e^{-\frac{(U_{ij} - \beta_j\lambda_i + \alpha_j)^2}{2}}$$

with priors

$$\lambda_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

$$\alpha_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$$

$$\beta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\beta^2)$$

- Taking the log:

$$
\begin{aligned}
&\log p(\alpha, \beta, \lambda, \mathrm{U} \mid \mathrm{Y}) \\
= &-\frac{1}{2}\left(\sum_{i=1}^{N}\lambda_i^2 + \sum_{j=1}^{J}\left(\frac{\alpha_j^2}{\sigma_\alpha^2} + \frac{\beta_j^2}{\sigma_\beta^2}\right)\right) \\
&+ \sum_{i=1}^{N}\sum_{j=1}^{J}\left\{\log\underbrace{(1\{U_{ij} > 0\}1\{Y_{ij} = 1\} + 1\{U_{ij} \leq 0\}1\{Y_{ij} = 0\})}_{=1 \text{ for any } ij}\right. \\
&\qquad\left. - \frac{1}{2}\left(U_{ij}^2 - 2(-\alpha_j + \beta_j\lambda_i)U_{ij} + (-\alpha_j + \beta_j\lambda_i)^2\right)\right\} \\
&+ \text{constant w.r.t. } (\alpha, \beta, \lambda, \mathrm{U})
\end{aligned}
$$

- Q-function:

$$
\begin{aligned}
& Q(\alpha, \beta, \lambda \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}) \\
=& \mathbb{E}_{p(\mathrm{U} \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}, \mathrm{Y})}[\log p(\alpha, \beta, \lambda, \mathrm{U} \mid \mathrm{Y})] \\
=& -\frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i^2 + \sum_{j=1}^{J}\left(\frac{\alpha_j^2}{\sigma_\alpha^2} + \frac{\beta_j^2}{\sigma_\beta^2}\right)\right) \\
& -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{J}\Bigg(\underbrace{\mathbb{E}_{p(\mathrm{U} \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}, \mathrm{Y})}[U_{ij}^2]}_{\text{constant w.r.t. } (\alpha,\beta,\lambda)} + (-\alpha_j + \beta_j \lambda_i)^2 \\
& \qquad - 2(-\alpha_j + \beta_j \lambda_i)\mathbb{E}_{p(\mathrm{U} \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}, \mathrm{Y})}[U_{ij}]\Bigg) \\
& + \text{constant w.r.t. } (\alpha, \beta, \lambda, \mathrm{U})
\end{aligned}
$$

- Only need to compute $\widetilde{U}_{ij}^{(s-1)} \equiv \mathbb{E}_{p(\mathrm{U} \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}, \mathrm{Y})}[U_{ij}]$
- Often (but not always), the expectation of Q reduces to the expectation of Z itself due to the log

- E-step: Compute $\widetilde{U}_{ij}^{(s-1)}$ in the Q-function
  1. $U_{ij} \mid \alpha, \beta, \lambda, Y_{ij}$: Truncated Gaussian

$$U_{ij} \mid \alpha, \beta, \lambda, Y_{ij} \overset{\text{indep.}}{\sim} \begin{cases} \mathcal{TN}_{U_{ij} \leq 0}(-\alpha_j + \beta_j \lambda_i, 1) & (Y_{ij} = 0) \\ \mathcal{TN}_{U_{ij} > 0}(-\alpha_j + \beta_j \lambda_i, 1) & (Y_{ij} = 1) \end{cases}$$

  2. Known expression for the mean of a truncated Gaussian r.v.:

$$\widetilde{U}_{ij}^{(s-1)} = \begin{cases} -\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)} + \dfrac{\varphi(-\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)})}{\Phi(-\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)})} & (Y_{ij} = 1) \\ -\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)} - \dfrac{\varphi(-\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)})}{1 - \Phi(-\alpha_j^{(s-1)} + \beta_j^{(s-1)} \lambda_i^{(s-1)})} & (Y_{ij} = 0) \end{cases}$$

  3. Easy to evaluate!

- M-step:
  - Maximizing Q simultaneously w.r.t. (α, β, λ) is hard
  - Conditional maximization steps:
    Repeat
    1. Maximize Q w.r.t. λ given α, β
    2. Maximize Q w.r.t. α, β given λ

    until (α, β, λ) converge
  - Iterations of alternating linear regressions (next two slides)

- Conditional maximization step for λ given (α, β):

$$\frac{\partial}{\partial \lambda_i} Q(\alpha, \beta, \lambda \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)})$$

$$= -\sum_{j=1}^{J}(-\alpha_j + \beta_j \lambda_i)\beta_j + \sum_{j=1}^{J} \beta_j \widetilde{U}_{ij}^{(s-1)} - \lambda_i = 0$$

$$\Rightarrow \lambda_i = \frac{\sum_{j=1}^{J} \beta_j (\widetilde{U}_{ij}^{(s-1)} - \alpha_j)}{(1 + \sum_{j=1}^{J} \beta_j^2)}$$

- For each i, regression of $\widetilde{U}_{ij}^{(s-1)} - \alpha_j$ on β_j with no intercept

- Conditional maximization step for (α, β) given λ:

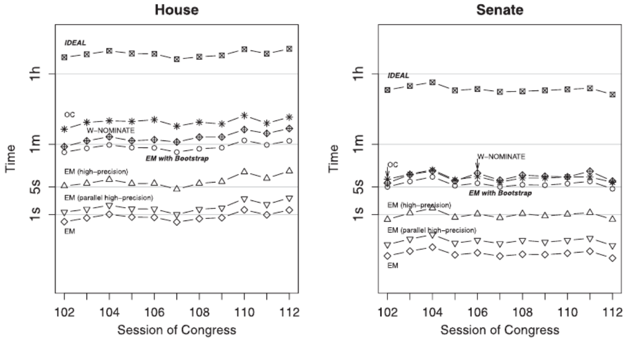
$$
\begin{aligned}
&\frac{\partial}{\partial(\alpha_j, \beta_j)^\top} Q(\alpha, \beta, \lambda \mid \alpha^{(s-1)}, \beta^{(s-1)}, \lambda^{(s-1)}) \\
&= -\sum_{i=1}^{N} (-\alpha_j + \beta_j \lambda_i) \begin{pmatrix} -1 \\ \lambda_i \end{pmatrix} + \sum_{i=1}^{N} \begin{pmatrix} -1 \\ \lambda_i \end{pmatrix} \widetilde{U}_{ij}^{(s-1)} \\
&\quad - \begin{pmatrix} \frac{\alpha_j}{\sigma_\alpha^2} \\ \frac{\beta_j}{\sigma_\beta^2} \end{pmatrix} \\
&= 0
\end{aligned}
$$

$$
\Rightarrow \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} = \left(\begin{pmatrix} \frac{1}{\sigma_\alpha^2} & 0 \\ 0 & \frac{1}{\sigma_\beta^2} \end{pmatrix} + \sum_{i=1}^{N} \begin{pmatrix} -1 \\ \lambda_i \end{pmatrix} \begin{pmatrix} -1 \\ \lambda_i \end{pmatrix}^\top \right)^{-1} \times \sum_{i=1}^{N} \begin{pmatrix} -1 \\ \lambda_i \end{pmatrix} \widetilde{U}_{ij}^{(s-1)}
$$

- For each j, regression of $\widetilde{U}_{ij}^{(s-1)}$ on $(-1, \lambda_i)$

Speed

FIGURE 1. Comparison of Computational Performance across the Methods



Notes: Each point represents the length of time required to compute estimates where the spacing of time on the vertical axis is based on the log scale. The proposed EM algorithm, indicated by "EM," "EM (high precision)," "EM (parallel high precision)," and "EM with Bootstrap" is compared with "W-NOMINATE" (Poole et al. 2011), the MCMC algorithm "IDEAL" (Jackman 2012), and the nonparametric optimal classification estimator "OC" (Poole et al. 2012). The EM algorithm is faster than the other approaches whether focused on point estimates or also estimation uncertainty. Algorithms producing uncertainty estimates are labeled in bold, italic type.

Imai, Lo, and Olmsted (2016) Figure 1.

- 1 hour via Gibbs while 1 second via EM!
- Not exactly fair, though—only the MAP, not the distribution

Ideal Points over Time

- Voting across multiple periods
  - Congress members across sessions
  - Judges across years
  - Countries across administrations

- Modeling strategies for multi-period ideal points, $\lambda_{it}, t = 1, \ldots, T$
  - Static ideal points: $\lambda_{it} = \lambda_i$ for all t
  - Independent ideal points: $\lambda_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ as the prior
  - Time trend: $\lambda_{it} = \lambda_{i0} + \gamma_i t + \eta_{it}, \eta_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_i^2)$

- More flexible approach: Dynamic linear model (DLM)

$$\lambda_{it} \mid \lambda_{i,t-1} \overset{\text{indep.}}{\sim} \mathcal{N}(\lambda_{i,t-1}, \omega_{it}^2)$$

$$\lambda_{i0} \overset{\text{indep.}}{\sim} \mathcal{N}(\kappa_i, \omega_{i0}^2)$$

  - Markov model: Random-walk prior on λ_{it}
  - State-space model: Generative process depends on a state, λ_{it}
  - κ_i and $\omega_{it}, t = 0, \ldots, T$: Prior parameters for identification

Ideal Points of Judges

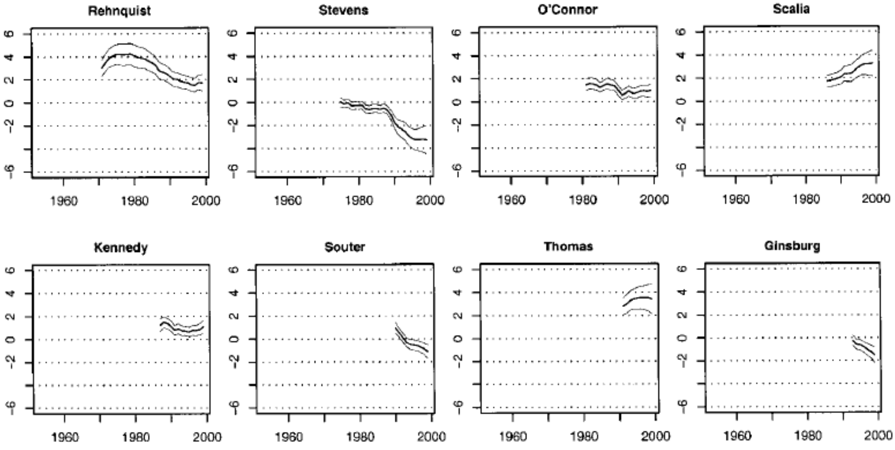


Fig. 1, Martin and Quinn (2002)

The Dynamic Model and Its Joint Posterior Density

- Unique legislators $i = 1, \ldots, N$
- Sessions $t = 1, \ldots, T$
- In each t, bills $j = 1, \ldots, J_t$
- Model

$$Y_{ijt} = \begin{cases} 0 & (U_{ijt} \leq 0) \\ 1 & (U_{ijt} > 0) \end{cases}$$

$$U_{ijt} = -\alpha_{jt} + \beta_{jt}\lambda_{it} + \varepsilon_{ijt}, \quad \varepsilon_{ijt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

$$\lambda_{it} \mid \lambda_{i,t-1} \overset{\text{indep.}}{\sim} \mathcal{N}(\lambda_{i,t-1}, \omega_{it}^2)$$

$$\lambda_{i0} \overset{\text{indep.}}{\sim} \mathcal{N}(\kappa_i, \omega_{i0}^2)$$

$$\alpha_{jt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$$

$$\beta_{jt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\beta^2)$$

- Joint posterior density:

$$
\begin{aligned}
& p(\alpha, \beta, \lambda, \mathbf{U} \mid \mathbf{Y}) \\
&= p(\alpha) p(\beta) \prod_{i=1}^{N} \left\{ p(\lambda_{i0}) \prod_{t=1}^{T} p(\lambda_{it} \mid \lambda_{i,t-1}) \right\} \\
&\quad \times \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J_t} p(Y_{ijt} \mid U_{ijt}) p(U_{ijt} \mid \alpha, \beta, \lambda) \\
&\propto e^{-\frac{1}{2}\left(\sum_{t=1}^{T} \sum_{j=1}^{J_t} \left(\frac{\alpha_{jt}^2}{\sigma_\alpha^2} + \frac{\beta_{jt}^2}{\sigma_\beta^2}\right)\right)} \underbrace{\prod_{i=1}^{N} \left\{ e^{-\frac{(\lambda_{i0} - \kappa_i)^2}{2\omega_{i0}^2}} \prod_{t=1}^{T} e^{-\frac{(\lambda_{it} - \lambda_{i,t-1})^2}{2\omega_{it}^2}} \right\}}_{\text{random walk prior}} \\
&\quad \times \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J_t} \left(1\{U_{ijt} > 0\} 1\{Y_{ijt} = 1\} + 1\{U_{ijt} \leq 0\} 1\{Y_{ijt} = 0\} \right) \\
&\quad\quad \times e^{-\frac{(U_{ijt} - \beta_{jt}\lambda_{it} + \alpha_{jt})^2}{2}}
\end{aligned}
$$

Gibbs Sampler for Dynamic Ideal Point Estimation

- Steps for U and (α, β) are identical to the static version
  - Draw $U_{ijt}^{(s)}$ conditional on $(\alpha_{jt}^{(s-1)}, \beta_{jt}^{(s-1)}, \lambda_{it}^{(s-1)})$

$$U_{ijt} \overset{\text{indep.}}{\sim} \begin{cases} \mathcal{TN}_{U_{ijt} \leq 0}(-\alpha_{jt}^{(s-1)} + \beta_{jt}^{(s-1)} \lambda_{it}^{(s-1)}, 1) & (Y_{ijt} = 0) \\ \mathcal{TN}_{U_{ijt} > 0}(-\alpha_{jt}^{(s-1)} + \beta_{jt}^{(s-1)} \lambda_{it}^{(s-1)}, 1) & (Y_{ijt} = 1) \end{cases}$$

  - Draw $(\alpha_{jt}^{(s-1)}, \beta_{jt}^{(s-1)})$ conditional on $(U_{\cdot jt}^{(s)}, \lambda_{\cdot t}^{(s-1)})$
    - For each item jt, run Bayesian regression of $U_{ijt}^{(s)}$ on $(-1, \lambda_{it}^{(s-1)})$

- Two strategies for sampling λ conditional on $(\mathbf{U}, \alpha, \beta)$
  1. Draw λ_{it} conditional on $\lambda_{i0}^{(s)}, \ldots, \lambda_{i,t-1}^{(s)}, \lambda_{i,t+1}^{(s-1)}, \ldots, \lambda_{iT}^{(s-1)}$
     - Easy to implement: For each (i, t), Bayesian regression of $U_{ijt}^{(s)} + \alpha_{jt}^{(s)}$ on $\beta_{jt}^{(s)}$ with $\mathcal{N}(\lambda_{i,t-1}^{(s)}, \omega_{it}^2)$ as the prior
     - *Very* slow convergence due to heavy dependence across iterations
  2. Batch sampling for $(\lambda_{i1}, \ldots, \lambda_{iT})$ via a forward-backward algorithm
     - More convoluted derivation of $p(\lambda_{i1}, \ldots, \lambda_{iT} \mid \alpha, \beta, \mathbf{U}, Y)$
     - Less dependent on previous draws, faster convergence

Forward-Backward Algorithm for DLM

- Conditional posterior density of $(\lambda_{i0}, \ldots, \lambda_{iT})$

$$p(\lambda_{i0}, \ldots, \lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot}) \propto e^{-\frac{(\lambda_{i0}-\kappa_i)^2}{2\omega_{i0}^2}} e^{-\frac{\sum_{t=1}^{T}(\lambda_{it}-\lambda_{i,t-1})^2}{2\omega_{it}^2}} \times e^{-\frac{\sum_{t=1}^{T}\sum_{j=1}^{J_t}(U_{ijt}-\beta_{jt}\lambda_{it}+a_{jt})^2}{2}}$$

  - All $U_{i\cdot t'}$ and $Y_{i\cdot t'}$ with $t' > t$ contain some information about λ_{it}

- Key factorization

$$\begin{aligned} & p(\lambda_{i0}, \ldots, \lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot}) \\ = & p(\lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot}) p(\lambda_{i,T-1} \mid \lambda_{i,T}, \ddot{a}_{T-1}, \ddot{\beta}_{T-1} \ddot{\mathbf{U}}_{i,T-1}, \ddot{\mathbf{Y}}_{i,T-1}) \\ & \times p(\lambda_{i,T-2} \mid \lambda_{i,T-1}, \ddot{a}_{T-2}, \ddot{\beta}_{T-2} \ddot{\mathbf{U}}_{i,T-2}, \ddot{\mathbf{Y}}_{i,T-2}) \\ & \times \ldots p(\lambda_{i1} \mid \lambda_{i2}, a_{\cdot 1}, \beta_{\cdot 1}, \mathrm{U}_{i\cdot 1}, \mathrm{Y}_{i\cdot 1}) p(\lambda_{i0} \mid \lambda_{i1}) \end{aligned}$$

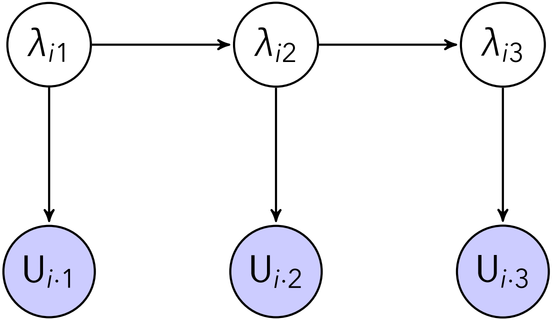
  where

$$\ddot{a}_t = (a_{\cdot 1}, \ldots, a_{\cdot t}), \quad \ddot{\beta}_t = (\beta_{\cdot 1}, \ldots, \beta_{\cdot t}), \quad \ddot{\mathbf{U}}_{it} = (\mathrm{U}_{i\cdot 1}, \ldots, \mathrm{U}_{i\cdot t}),$$
$$\ddot{\mathrm{Y}}_{it} = (\mathrm{Y}_{i\cdot 1}, \ldots, \mathrm{Y}_{i\cdot t})$$

  - λ_{it} and the periods after t are independent conditional on $\lambda_{i,t+1}$

- Proof of factorization:

$$
\begin{aligned}
& p(\lambda_{i,T-1} \mid \lambda_{i,T}, \ddot{a}_T, \ddot{\beta}_T, \ddot{\mathbf{U}}_{i,T}, \ddot{\mathbf{Y}}_{i,T}) \\
\propto\, & p(\lambda_{i,T-1} \mid \ddot{a}_{T-1}, \ddot{\beta}_{T-1}, \ddot{\mathbf{U}}_{i,T-1}, \ddot{\mathbf{Y}}_{i,T-1}) \\
& \times \underbrace{p(\lambda_{i,T}, a_{\cdot T}, \beta_{\cdot T}, \mathrm{U}_{i\cdot T}, \mathrm{Y}_{i\cdot T} \mid \lambda_{i,T-1}, \ddot{a}_{T-1}, \ddot{\beta}_{T-1}, \ddot{\mathbf{U}}_{i,T-1}, \ddot{\mathbf{Y}}_{i,T-1})}_{=p(\mathrm{Y}_{i\cdot T} \mid \mathrm{U}_{i\cdot T})p(\mathrm{U}_{i\cdot T} \mid \lambda_{iT}, a_{\cdot T}, \beta_{\cdot T})p(a_{\cdot T}, \beta_{\cdot T})p(\lambda_{iT} \mid \lambda_{i,T-1})} \\
\propto\, & p(\lambda_{i,T-1} \mid \ddot{a}_{T-1}, \ddot{\beta}_{T-1}, \ddot{\mathbf{U}}_{i,T-1}, \ddot{\mathbf{Y}}_{i,T-1}) p(\lambda_{iT} \mid \lambda_{i,T-1})
\end{aligned}
$$



- Applies recursively

Forward Filtering

- Draws from $p(\lambda_{i0}, \ldots, \lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot})$ using the factorization:
  1. Draw $\lambda_{iT}^{(s)}$ from $p(\lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot})$
  2. Draw $\lambda_{i,T-1}^{(s)}$ given $\lambda_{iT}^{(s)}$ from $p(\lambda_{i,T-1} \mid \lambda_{i,T}^{(s)}, \ddot{a}_{T-1}, \ddot{\beta}_{T-1}\ddot{\mathbf{U}}_{i,T-1}, \ddot{\mathbf{Y}}_{i,T-1})$

     $\vdots$
  3. Draw $\lambda_{i0}^{(s)}$ given $\lambda_{i1}^{(s)}$
- Forward filtering: Compute $p(\lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot})$ recursively

$$
\begin{aligned}
& p(\lambda_{it} \mid \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{it}, \ddot{\mathbf{Y}}_{it}) \\
&= \int p(\lambda_{it}, \lambda_{i,t-1} \mid \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{it}, \ddot{\mathbf{Y}}_{it}) d\lambda_{i,t-1} \\
&= \int \frac{p(\lambda_{it}, \lambda_{i,t-1}, U_{i\cdot t}, Y_{i\cdot t} \mid \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1})}{p(U_{i\cdot t}, Y_{i\cdot t} \mid \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1})} d\lambda_{i,t-1} \\
&\propto p(Y_{i\cdot t} \mid U_{i\cdot t}) \int p(\lambda_{it}, U_{i\cdot t} \mid \lambda_{i,t-1}, \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) \\
&\qquad \times p(\lambda_{i,t-1} \mid \ddot{a}_t, \ddot{\beta}_t, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) d\lambda_{i,t-1} \\
&\propto p(U_{i\cdot,t} \mid a_{\cdot t}, \beta_{\cdot t}, \lambda_{it}) \int p(\lambda_{it} \mid \lambda_{i,t-1}) \\
&\qquad \times p(\lambda_{i,t-1} \mid \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) d\lambda_{i,t-1}
\end{aligned}
$$

Forward Recursion

- All densities are Gaussian
- Forward recursion
  1. Begin with: $\lambda_{i0} \overset{\text{indep.}}{\sim} \mathcal{N}(\kappa_i, \omega_{i0}^2)$ where κ_i and ω_{i0}^2 are known
  2. Assume: $\lambda_{i,t-1} \mid \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1} \overset{\text{indep.}}{\sim} \mathcal{N}(\hat{\kappa}_{i,t-1}, \hat{\omega}_{i,t-1}^2)$
  3. Recursive computation:
  $$\int p(\lambda_{it} \mid \lambda_{i,t-1}) p(\lambda_{i,t-1} \mid \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) d\lambda_{i,t-1} \propto e^{-\frac{(\lambda_{it} - \hat{\kappa}_{i,t-1})^2}{2(\hat{\omega}_{i,t-1}^2 + \omega_{it}^2)}}$$
  $$p(U_{i\cdot,t} \mid a_{\cdot t}, \beta_{\cdot t}, \lambda_{it}) \propto e^{-\frac{\sum_{j=1}^{J_t} (U_{ijt} - \beta_{jt}\lambda_{it} + a_{jt})^2}{2}}$$
  4. Equivalent to the posterior density of Bayesian regression!
  $$\hat{\omega}_{it}^2 = \frac{1}{\sum_{j=1}^{J_t} \beta_{jt}^2 + 1/\hat{\omega}_{i,t-1}^2}, \quad \hat{\kappa}_{it} = \hat{\omega}_{it}^2 \left\{ \sum_{j=1}^{J_t} \beta_{jt}(U_{ijt} + a_{jt}) + \frac{\hat{\kappa}_{i,t-1}}{\hat{\omega}_{i,t-1}^2} \right\}$$
- The last iteration of the forward recursion, $(\hat{\kappa}_{iT}, \hat{\omega}_{iT}^2)$:
$$p(\lambda_{iT} \mid a, \beta, \mathbf{U}_{i\cdot\cdot}, \mathbf{Y}_{i\cdot\cdot}) \propto e^{-\frac{(\lambda_{iT} - \hat{\kappa}_{iT})^2}{2\hat{\omega}_{iT}^2}}$$

Backward Sampling

- Backward sampling
  1. Draw $\lambda_{iT}^{(s)}$ from $\mathcal{N}(\hat{\kappa}_{iT}, \hat{\omega}_{iT}^2)$
  2. Draw $\lambda_{i,t-1}^{(s)}$ conditional on $(\lambda_{i,t}^{(s)}, \ddot{a}_{t-1}^{(s)}, \ddot{\beta}_{t-1}^{(s)}, \ddot{\mathbf{U}}_{i,t-1}^{(s)}, \ddot{\mathbf{Y}}_{i,t-1})$:
     - Joint density of $(\lambda_{i,t-1}, \lambda_{i,t})$:

$$
\begin{aligned}
& p(\lambda_{it}, \lambda_{i,t-1} \mid \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) \\
&= p(\lambda_{it} \mid \lambda_{i,t-1}) p(\lambda_{i,t-1} \mid \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1}) \\
&\propto e^{-\frac{(\lambda_{it}-\lambda_{i,t-1})^2}{2\omega_{it}^2}} e^{-\frac{(\lambda_{i,t-1}-\hat{\kappa}_{i,t-1})^2}{2\hat{\omega}_{i,t-1}^2}}
\end{aligned}
$$

     - Conditional density of $\lambda_{i,t-1}$

$$
p(\lambda_{i,t-1} \mid \lambda_{it}, \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1})
$$

$$
\propto e^{-\frac{(\lambda_{it}-\lambda_{i,t-1})^2}{2\omega_{it}^2}} e^{-\frac{(\lambda_{i,t-1}-\hat{\kappa}_{i,t-1})^2}{2\hat{\omega}_{i,t-1}^2}} \propto e^{-\frac{\left(\lambda_{i,t-1} - \frac{\lambda_{it}/\omega_{it}^2 + \hat{\kappa}_{i,t-1}/\hat{\omega}_{i,t-1}^2}{1/\omega_{it}^2 + 1/\hat{\omega}_{i,t-1}^2}\right)^2}{1/\omega_{it}^2 + 1/\hat{\omega}_{i,t-1}^2}}
$$

$$
\lambda_{i,t-1} \mid \lambda_{i,t}, \ddot{a}_{t-1}, \ddot{\beta}_{t-1}, \ddot{\mathbf{U}}_{i,t-1}, \ddot{\mathbf{Y}}_{i,t-1} \sim \mathcal{N}\left(\frac{\frac{\lambda_{it}}{\omega_{it}^2} + \frac{\hat{\kappa}_{i,t-1}}{\hat{\omega}_{i,t-1}^2}}{\frac{1}{\omega_{it}^2} + \frac{1}{\hat{\omega}_{i,t-1}^2}}, \frac{1}{\omega_{it}^2} + \frac{1}{\hat{\omega}_{i,t-1}^2} \right)
$$

- Many recursions in FFBS $\rightsquigarrow$ algorithm may be slow

Summary

- Two-parameter IRT model
  - Measuring latent "ability" of units from a set of responses
  - Spatial voting interpretation $\rightsquigarrow$ ideal point estimation
- The EM algorithm
  - Iterative maximization of the log posterior density
  - Faster, but only estimates the posterior mode
- The dynamic linear model
  - Random-walk process of a latent state variable
  - Forward-backward algorithm
- Readings for review
  1. Ideal point estimation:
     - Clinton, et. al. (2004) "The Statistical Analysis of Roll Call Data"
  2. The EM algorithm:
     - **BDA3** Sections 13.4
     - Imai, et. al. (2016) "Fast Estimation of Ideal Points with Massive Data"
  3. Dynamic models
     - Martin and Quinn (2002) "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999"
     - Scott (2002) "Bayesian Methods for Hidden Markov Models"