# Fast Estimation of Ideal Point with Massive Data

## Reference Paper

1. Fast Estimation of Ideal Point with Massive Data, KOSUKE IMAI et al.

## Introduction

투표자, 입법자 등의 사상적 위치를 추정하는 것은 정치 과학 분야에서 주요 문제이다. 이용 가능한 데이터가 증가하고 방법론의 복잡성이 증가하는 가운데, 최근 연구자들은 시간과 기관에 걸쳐서 비교 가능한 사상적 선호에 주목했다. 이러한 관심은 데이터 크기의 증가를 필연적으로 수반하기 때문에 computational challenge를 수반한다. 사상적 위치를 추정하는 연구 분야에서 대부분 베이지안 모델을 계산하기 위해 MCMC 알고리즘을 많이 사용했지만 이는 계산이 매우 느려 용량이 큰 데이터에 대해서는 계산이 며칠씩 걸릴수 있다.

해당 논문에서는 대용량 데이터에 적합한 fast estimation method를 제안한다. 구체적으로 다양한 ideal point 모델 하에서, EM를 사용하여 사후 분포를 정확하게, 또는 근사적으로 최대화할 것이다. EM의 큰 장점은 계산 시간을 크게 줄일 수 있다는 것이다.

먼저 standard bayesian ideal point model로 시작하여 확장을 통해 mixed ordinal and continuous outcomes을 가질수 있는 EM, EM for dynamic model, hierarchical model, EM for ideal point model based on textual and network data를 소개한다.

dynamic model과 hierarchical model은 closed form으로 존재하지 않으므로 variational bayesian inference을 통해 사후분포를 근사한다.

마지막으로, EM의 가장 큰 단점 중 하나는 표준오차와 같은 uncertainty estimates를 제공하지 않는다는 것이다. 해당 논문에서는 parametric bootstrap 접근법을 이용한다. 물론 이 계산을 추가하면 계산 시간이 증가하겠지만, 여전히 제안한 EM이 계산 시간이 빠르다.

## Standard Ideal Point Model

해당 모델에서는 사후 분포를 근사 없이 최대화하는 EM을 소개한다. 그 이후 미국 의회 세션의 투표 데이터에 모델을 적용함으로써 계산적 효율성과 scalability을 설명한다.

## The Model

N명의 입법자와 J개의 roll call이 있다고 하자. $y_{ij}$ 를 입법자 $i$ 의 roll call j에 대한 투표라고 하자. 즉, $y_{ij} = 1$ $(y_{ij} = 0)$ 은 그 투표가 긍정적 (부정적)이라는 것이다. $(i = 1, \cdots, N, \ , j = 1, \cdots, J)$ $\mathbf{x}_i$ 를 입법자 i에 대한 $K$ 차원의 ideal point 열벡터라고 하자. 그러면 $y^*_{ij}$ 를 'yea'를 던지기 위한어떤 latent propensity로 사용한다면 standard K-dimensional ideal point model은 아래와 같이 쓸 수 있다. $(y_{ij} = \mathbf{1}\{y^* > 0\})$

$$y^*_{ij} = \alpha_j + \mathbf{x}_i^T \boldsymbol{\beta}_j + \epsilon_{ij} \tag{1}$$

$\boldsymbol{\beta}_j$ 는 $K$ 차원의 item discrimination parameters 열벡터이고 $\alpha_j$ 는 scalar item difficulty parameter 이다. 마지막으로 $\epsilon_{ij}$ 는 표준정규분포로부터의 iid random variables이라고 가정한다.

간단한 노테이션을 위해 $\tilde{\boldsymbol{\beta}}_j^T = (\alpha_j, \boldsymbol{\beta}_j^T)$, $\tilde{\boldsymbol{x}}_i^T = (1, \mathbf{x}_i^T)$ 라 하자. 그러면 아래와 같이 간단히 적을 수

있다.

$$y^*_{ij} = \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}}_j + \epsilon_{ij} \tag{2}$$

원래의 논문을 따를 것인데, $\mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j$ 에 대해서 각각 independent, conjugate prior을 부여한다.

$$p(\mathbf{x}_1, \cdots, \mathbf{x}_N) = \prod_{i=1}^{N} \phi_K(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$$

$$p(\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_J) = \prod_{j=1}^{J} \phi_{K+1}(\tilde{\boldsymbol{\beta}}_j; \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}) \tag{3}$$

여기서 $\phi_K(\cdot;\cdot)$ 는 $K$ 차원의 다변량 정규분포의 밀도 함수이다. 이 모델 하에서, joint posterior distribution of $\left(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J\right)$ conditional on the roll-call matrix $\mathbf{Y}$ 는 아래와 같다.

$$\begin{aligned}
p\left((\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J \mid \mathbf{Y}\right) &\propto p\left(\mathbf{Y} \mid \mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J\right) p\left(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J\right) \\
&\propto \prod_{i=1}^{N}\prod_{j=1}^{J} \left(\mathbf{1}\left\{y^*_{ij} > 0\right\} \mathbf{1}\left\{y_{ij} = 1\right\} + \mathbf{1}\left\{y^*_{ij} \le 0\right\} \mathbf{1}\left\{y_{ij} = 0\right\}\right) \phi_1(y^*_{ij}; \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}}_j, 1) \\
&\times \prod_{i=1}^{N} \phi_K(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) \prod_{j=1}^{J} \phi_{K+1}(\tilde{\boldsymbol{\beta}}_j; \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}})
\end{aligned} \tag{4}$$

여기서 $\mathbf{Y}, \mathbf{Y}^*$는 각각 i번째 행과 j번째 열은 $y_{ij}, y^*_{ij}$ 를 의미한다.
이전까지의 시도는, 위 (4) 의 joint posterior distribution에 대해 MCMC를 이용하여 그 분포를 유도하였지만 매우 느리다는 단점이 있었다. 본 논문에서는 이처럼 MCMC로 유도된 ideal point estimation model에 대하여, MCMC와 정확도는 매우 유사하지만 유도 속도는 훨씬 빠른 EM 알고리즘을 유도하는 것이 목적이다.

## The Proposed Algorithm

이제 근사 없이 (4) 에서 주어진 사후분포를 최대화하는 EM 알고리즘을 유도한다. 근사 없이 closed form으로 유도가 가능하다면, 추정치 측면에서 MCMC와 크게 다르지 않을 것이고 속도는 훨씬 빠르므로 효율적일 것이다. 제안된 알고리즘은 $\{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J$ 을 모수로 보고, $\mathbf{Y}^*$ 을 결측 자료로 본다. 구체적으로 t번째 iteration에서 현재 모수 값을 $\left\{\mathbf{x}_i^{(t-1)}\right\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j^{(t-1)}\right\}_{j=1}^J$ 로 두자. 그러면 E step은 아래의 소위 Q function으로 주어지는데, 이는 expecation of the log joint posterior distribution 이다.

$$\begin{aligned}
Q\left(\{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J\right) &= \mathbb{E}\left[log\, p\left((\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J \mid \mathbf{Y}\right) \mid \mathbf{Y}, \left\{\mathbf{x}_i^{(t-1)}\right\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j^{(t-1)}\right\}_{j=1}^J\right] \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{J}\left(\tilde{\boldsymbol{\beta}}_j^T \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}}_j - 2\tilde{\boldsymbol{\beta}}_j^T \tilde{\mathbf{x}}_i y_{ij}^{*(t)}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}_i^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{x}_i - 2\mathbf{x}_i^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}}\right) \\
&- \frac{1}{2}\sum_{j=1}^{J}\left(\tilde{\boldsymbol{\beta}}_j^T \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1} \tilde{\boldsymbol{\beta}}_j - 2\tilde{\boldsymbol{\beta}}_j^T \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1} \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\right) + const
\end{aligned} \tag{5}$$

여기서

$$
\begin{aligned}
y_{ij}^{*(t)} &= \mathbb{E}\left[y_{ij}^{*} \mid \mathbf{x}_i^{(t-1)}, \tilde{\boldsymbol{\beta}}_j^{(t-1)}, y_{ij}\right] \\
&= \begin{cases}
m_{ij}^{(t-1)} + \dfrac{\phi(m_{ij}^{(t-1)})}{\Phi(m_{ij}^{(t-1)})} & if\ y_{ij} = 1 \\
m_{ij}^{(t-1)} - \dfrac{\phi(m_{ij}^{(t-1)})}{1-\Phi(m_{ij}^{(t-1)})} & if\ y_{ij} = 0 \\
m_{ij}^{(t-1)} & if\ y_{ij}\ is\ missing
\end{cases}
\end{aligned}
$$

$$
with\ m_{ij}^{(t-1)} = \left(\tilde{\mathbf{x}}_i^{(t-1)}\right)^T \tilde{\boldsymbol{\beta}}_j^{(t-1)} \tag{6}
$$

이제 (6) 에서 정의한 $Q(\cdot)$ 을 $\mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j$ 에 대해서 최대화하는 argmax을 찾아보자. 주의해야할 점은, 논문에서 $\mathbf{x}_i^{(t)}$ 로 argmax를 구했다는 점이다. 여태까지 $\tilde{\mathbf{x}}_i$ 로 노테이션을 적었으므로 $\tilde{\mathbf{x}}_i = (1,\ \mathbf{x}_i)^T$ 와 $\tilde{\boldsymbol{\beta}}_j = (\alpha_j, \boldsymbol{\beta}_j^T)^T$ 를 이용해서 다시 $Q$ function을 정의하고 $\mathbf{x}_i$ 와 관련된 부분만 적은 후에, $\mathbf{x}_i$ 의 argmax를 찾아본다. 여기서

$$
\tilde{\boldsymbol{\beta}}_j^T \tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}}_j = \alpha_j + \mathbf{x}_i \boldsymbol{\beta}_j^T
$$

임을 이용하자.

$$
\begin{aligned}
Q &= -\frac{1}{2}\sum_{j=1}^{J}\left[\left(\alpha_j + \mathbf{x}_i\boldsymbol{\beta}_j^T\right)^2 - 2y_{ij}^{*(t)}\left(\alpha_j + \mathbf{x}_i\boldsymbol{\beta}_j^T\right)\right] - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}_i^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i - 2\mathbf{x}_i^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}\right) + const \\
&= \frac{1}{2}\sum_{j=1}^{J}\left[\mathbf{x}_i^T\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\mathbf{x}_i + 2\alpha_j\mathbf{x}_i\boldsymbol{\beta}_j^T - 2y_{ij}^{*(t)}\mathbf{x}_i\boldsymbol{\beta}_j^T\right] - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}_i^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i - 2\mathbf{x}_i^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}\right) + const
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \mathbf{x}_i} &= -\frac{1}{2}\sum_{j=1}^{J}\left(2\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\mathbf{x}_i + 2\alpha_j\boldsymbol{\beta}_j - 2y_{ij}^{*(t)}\boldsymbol{\beta}_j\right) - \frac{1}{2}\left(2\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i - 2\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}\right) \\
&= -\left(\sum_{j=1}^{J}\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\mathbf{x}_i + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\mathbf{x}_i + \left(\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \sum_{j=1}^{J}\boldsymbol{\beta}_j\left(y_{ij}^{*(t)} - \alpha_j\right)\right)
\end{aligned}
$$

$$
\therefore \mathbf{x}_i^{(t)} = \left(\sum_{j=1}^{J}\boldsymbol{\beta}_j^{(t-1)}\boldsymbol{\beta}_j^{(t-1)T}\mathbf{x}_i + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)^{-1} \times \left(\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \sum_{j=1}^{J}\boldsymbol{\beta}_j^{(t-1)}\left(y_{ij}^{*(t)} - \alpha_j^{(t-1)}\right)\right) \tag{7}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \tilde{\boldsymbol{\beta}}_j} &= -\frac{1}{2}\sum_{i=1}^{N}\left(2\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\tilde{\boldsymbol{\beta}}_j - 2\tilde{\mathbf{x}}_i y_{ij}^{*(t)}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(2\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j - 2\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\right) \\
&= -\left(\sum_{i=1}^{N}\left(\tilde{\mathbf{x}}_i\left(\tilde{\mathbf{x}}_i\right)^T\right) + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\right)\tilde{\boldsymbol{\beta}}_j + \left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_i y_{ij}^{*(t)} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\right)
\end{aligned}
$$

$$
\therefore \tilde{\boldsymbol{\beta}}_j^{(t)} = \left(\sum_{i=1}^{N}\left(\tilde{\mathbf{x}}_i^{(t)}\left(\tilde{\mathbf{x}}_i^{(t)}\right)^T\right) + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\right)^{-1} \times \left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_i^{(t)} y_{ij}^{*(t)} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\right) \tag{8}
$$

(8) 과 같이 $t$ 번째 iteration에서 모수의 추정치를 구하고, 이전의 추정치와 현재의 추정치들 간에 correlation이 일정 이상이면 수렴했다고 판단하는 correlation-based convergence criteria을 이용하여 수렴 여부를 판단한다.

## Ideal Point Model wih Mixed Binary, Ordinal and Continuous Outcomes

이제 위에서 구한 EM 알고리즘을 더 확장된 데이터 형태에 적용한다. Quinn은 이러한 모델을 위해 MCMC 알고리즘을 개발했고 R에서 MCMCpack라는 패키지의 MCMCmixfactanal() 함수로 구현했다. 이 모델과 깊게 연관된 ordinal probit model을 위한 EM 알고리즘은 E step에 대한 closed form이 없기 때문에 어려운 점이 있다. 이러한 이유 때문에, ordinal probit model에서 EM 알고리즘은 개발되지 않았다.

여기서는 반응 변수가 세 개의 순서형 범주를 가지고 있다는 가정 하에, E step의 closed form이 존재함을 보일 것이다. 적절한 reparameterization을 통해서 analytically tractable한 EM 알고리즘을 도출한다. 이후에는 반응 변수의 범주가 세 개 이상인 경우, mixture of binary, ordinal, continuous variables인 경우를 살펴본다.

## The Model with a Three-category Ordinal Outcome

반응변수가 이제 세 개의 순서형 값인 $y_{ij} \in \{0, 1, 2\}$ 을 가진다는 것을 제외하고는 위의 standard model과 동일한 setup이다. ordinal probit model에서는 각 반응변수의 확률은 아래와 같다. probit model은 cumulative distribution을 표준정규분포의 cdf로 가정한다. 따라서 이를 $y^*_{ij}$ 의 누적 확률로 다시 바꿔 쓸 수 있다.

$$Pr(y_{ij} = 0) = \Phi(\alpha_{1j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) = Pr\left(y^*_{ij} - \mathbf{x}_i^T \boldsymbol{\beta}_j \le \alpha_{1j} - \mathbf{x}_i^T \boldsymbol{\beta}_j\right) = Pr(y^*_{ij} \le \alpha_{1j})$$

$$Pr(y_{ij} = 1) = \Phi(\alpha_{2j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) - \Phi(\alpha_{1j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) = Pr(y^*_{ij} \le \alpha_{2j}) - Pr(y^*_{ij} \le \alpha_{1j})$$

$$Pr(y_{ij} = 2) = 1 - \Phi(\alpha_{2j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) = 1 - Pr(y^*_{ij} \le \alpha_{2j}) \tag{9}$$

$$where\ \alpha_{2j} > \alpha_{1j}\ for\ all\ j = 1, 2, \cdots, J$$

respondent i가 가지는 동의에 대한 latent propensity인 $y^*_{ij}$ 은 다음과 같이 쓴다.

$$y^*_{ij} = \mathbf{x}_i^T \boldsymbol{\beta}_j + \epsilon_{ij}\ where\ \epsilon_{ij} \underset{\sim}{iid} N(0, 1) \tag{10}$$

(9) 를 이용하여, (1) 에서 정의한 latent propensity를 관측된 결과인 $y_{ij}$ 와 아래와 같이 연관지을 수 있다.

$$y_{ij} = \begin{cases} 0 & if\ y^*_{ij} < \alpha_{1j} \\ 1 & if\ \alpha_{1j} \le y^*_{ij} < \alpha_{2j} \\ 2 & if\ \alpha_{2j} \le y^*_{ij} \end{cases} \tag{11}$$

위에서와 동일하게, $\left(\{\boldsymbol{\beta}_j\}_{j=1}^J, \{\mathbf{x}_i\}_{i=1}^N\right)$ 에 대해 normal independent prior 분포를 가정한다. $\left(\{\alpha_{1j}, \alpha_{2j}\}_{j=1}^J\right)$ 에 대해서는 $\alpha_{1j} < \alpha_{2j}$ 을 유지하며 improper uniform prior을 가정한다. 이렇게 가정을 한 이후에,

joint posterior 분포는 아래와 같다.

$$p\left(\mathbf{Y}^*, \left\{\alpha_{1j}, \alpha_{2j}, \boldsymbol{\beta}_j\right\}_{j=1}^J, \left\{\mathbf{x}_i\right\}_{i=1}^N \mid \mathbf{Y}\right) \propto \prod_{i=1}^N \prod_{j=1}^J \Big[\mathbf{1}\left\{y_{ij}^* < \alpha_{1j}\right\} \mathbf{1}\left\{y_{ij} = 0\right\}$$
$$+ \mathbf{1}\left\{\alpha_{1j} \le y_{ij}^* < \alpha_{2j}\right\} \mathbf{1}\left\{y_{ij} = 1\right\} + \mathbf{1}\left\{\alpha_{2j} \le y_{ij}^*\right\} \mathbf{1}\left\{y_{ij} = 2\right\}\Big]$$
$$\times \phi_1\left(y_{ij}^*; \mathbf{x}_i^T \boldsymbol{\beta}_j, 1\right) \times \prod_{i=1}^N \phi_K(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) \prod_{j=1}^J \phi_K(\boldsymbol{\beta}_j; \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \tag{12}$$

## The Proposed Algorithm

Analytically tractable한 EM 알고리즘을 개발하기 위해서, 다음의 one-to-one transformation of parameters을 적용한다.

$$\begin{aligned}
\tau_j &= \alpha_{2j} - \alpha_{1j} \ > \ 0 \\
\alpha_j^* &= -\frac{\alpha_{1j}}{\tau_j} \\
\boldsymbol{\beta}_j^* &= \frac{\boldsymbol{\beta}_j}{\tau_j} \\
z_{ij}^* &= \frac{y_{ij}^* - \alpha_{1j}}{\tau_j} \\
\epsilon_{ij}^* &= \frac{\epsilon_{ij}}{\tau_j}
\end{aligned} \tag{13}$$

위의 reparaemterization을 통해서 (9)를 아래와 같이 다시 쓸 수 있다.

$$\begin{aligned}
Pr(y_{ij} = 0) &= \Phi(\alpha_{1j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) \\
&= \Phi(-\tau_j \alpha_j^* - \tau_j \mathbf{x}_i^T \boldsymbol{\beta}_j^*)
\end{aligned}$$

$$\begin{aligned}
Pr(y_{ij} = 1) &= \Phi(\alpha_{2j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) - \Phi(\alpha_{1j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) \\
&= \Phi(\tau_j - \tau_j \alpha_j^* - \tau_j \mathbf{x}_i^T \boldsymbol{\beta}_j^*) - \Phi(-\tau_j \alpha_j^* - \tau_j \mathbf{x}_i^T \boldsymbol{\beta}_j^*)
\end{aligned}$$

$$\begin{aligned}
Pr(y_{ij} = 2) &= 1 - \Phi(\alpha_{2j} - \mathbf{x}_i^T \boldsymbol{\beta}_j) \\
&= 1 - \Phi(\tau_j - \tau_j \alpha_j^* - \tau_j \mathbf{x}_i^T \boldsymbol{\beta}_j^*)
\end{aligned} \tag{14}$$

(14)와 같은 parameterization에서 latent variable representation은 아래와 같다.

$$z_{ij}^* = \alpha_j^* + \mathbf{x}_i^T \boldsymbol{\beta}_j^* + \epsilon_{ij}^*$$

$$where\ \epsilon_{ij}^* \underset{\sim}{indep}\ N(0, \tau_j^{-2}),\ \tau_j\left(z_{ij}^* - \alpha_j^* - \mathbf{x}_i^T \boldsymbol{\beta}_j^*\right) \underset{\sim}{indep}\ N(0, 1) \tag{15}$$

(15)의 representation을 이용해서 (14)을 $z_{ij}^*$의 누적 확률 형태로 변형해보자.

$$\begin{aligned}
Pr(y_{ij}=0) &= \Phi(-\tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*) \\
&= Pr(\tau_j\left(z_{ij}^* - \alpha_j^* - \mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \le -\tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*) \\
&= Pr(z_{ij}^* \le 0)
\end{aligned}$$

$$\begin{aligned}
Pr(y_{ij}=1) &= \Phi(\tau_j - \tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*) - \Phi(-\tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*) \\
&= Pr\left(\tau_j\left(z_{ij}^* - \alpha_j^* - \mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \le \tau_j - \tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \\
&- Pr\left(\tau_j\left(z_{ij}^* - \alpha_j^* - \mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \le -\tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \\
&= Pr\left(z_{ij}^* \le 1\right) - Pr\left(z_{ij}^* \le 0\right)
\end{aligned}$$

$$\begin{aligned}
Pr(y_{ij}=2) &= 1 - \Phi(\tau_j - \tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*) \\
&= 1 - Pr\left(\tau_j\left(z_{ij}^* - \alpha_j^* - \mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \le \tau_j - \tau_j\alpha_j^* - \tau_j\mathbf{x}_i^T\boldsymbol{\beta}_j^*\right) \\
&= 1 - Pr\left(z_{ij}^* \le 1\right)
\end{aligned}$$

따라서 위의 결과를 이용하여 observed outcome인 $y_{ij}$와 latent variable인 $z_{ij}^*$의 관계는 아래와 같다.

$$y_{ij} = \begin{cases} 0 & if\ z_{ij}^* < 0 \\ 1 & if\ 0 \le z_{ij}^* < 1 \\ 2 & if\ 1 \le z_{ij}^* \end{cases} \tag{16}$$

이러한 reparameterization로 인해서 threshold parameters인 $(\alpha_{1j}, \alpha_{2j})$가 $(0,1)$로, 그리고 heterogeneuos variance인 $\tau_j^{-2}$을 지니게 됨을 확인할 수 있다.
conjugacy을 유지하기 위해서, (12)의 prior 분포를 아래와 같이 바꾼다.

$$p\left(\mathbf{Y}^*, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{j=1}^J, \{\tau_j^2\}_{j=1}^J, \{\mathbf{x}_i\}_{i=1}^N\right) = \prod_{j=1}^J \phi_{K+1}(\tilde{\boldsymbol{\beta}}_j;\ \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}) \times \mathcal{G}\left(\tau_j^2;\ \frac{v_\tau}{2}, \frac{s_\tau}{2}\right) \prod_{i=1}^N \phi_K(\mathbf{x}_i;\ \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) \tag{17}$$

$$where\ \tilde{\boldsymbol{\beta}}_j = \left(\alpha_j^*, \boldsymbol{\beta}_j^*\right)\ and\ \mathcal{G}(\cdot)\ is\ Gamma\ distribution.$$

이러한 세팅에서 사후 분포를 maximize하기 위한 EM 알고리즘을 유도한다. ideal point model과 유사한 방법이 쓰인다. 자세한 사항은 논문 뒷편에 나와있다 (수식이 복잡하므로 우선 생략)

## Mixed Binary, Ordinal, and Continuous Outcomes

이제 조금의 변경을 통해, 위에서 전개한 EM 알고리즘을 좀 더 일반적인 상황, mixed binary, ordinal, continuous outcomes에 적용해본다. 만약 카테고리가 세 개 이상이면 범주를 세 개로 통합하는 것이 낫다 (그렇지 않으면 variational inference을 사용하는데, 모형이 더 복잡해지므로 analytical

tractability와 computational efficiency을 위해서 간단한 모형을 사용한다)
다음으로 $y_{ij}$가 특정한 관측치 (i,j)에 대해서 binary outcome인 상황을 생각하자. 그렇다면 관찰된 결과와 latent propensity간의 아래와 같은 관계를 생각할 수 있다. $z^*_{ij};\ z_{ij} < 1 \iff y_{ij} = 0,\ z_{ij} \geq 0 \iff y_{ij} = 1$. 이러한 상황에서 E-Step은 아래와 같다.

$$z_{ij}^{*(t)} = \begin{cases} m_{ij}^{(t-1)} - \frac{1}{\tau_j^{(t-1)}}\lambda\left(-m_{ij}^{(t-1)} + 1,\ \tau_j^{(t-1)}\right) & if\ y_{ij} = 0 \\ m_{ij}^{(t-1)} + \frac{1}{\tau_j^{(t-1)}}\lambda\left(m_{ij}^{(t-1)},\ \tau_j^{(t-1)}\right) & if\ y_{ij} = 1 \end{cases} \tag{18}$$

$$\left(z_{ij}^{*2}\right)^{(t)} = \begin{cases} \left(z_{ij}^{*(t)}\right)^2 + \frac{1}{\left(\tau_j^{(t-1)}\right)^2}\left[1 - \frac{1 - m_{ij}^{(t-1)}}{\tau_j^{(t-1)}}\lambda\left(m_{ij}^{(t-1)} - 1, \tau_j^{(t-1)}\right) - \left\{\lambda\left(m_{ij}^{(t-1)} - 1, \tau_j^{(t-1)}\right)\right\}^2\right] & if\ y_{ij} = 0 \\ \left(z_{ij}^{*(t)}\right)^2 + \frac{1}{\left(\tau_j^{(t-1)}\right)^2}\left[1 - \lambda\left(-m_{ij}^{(t-1)}, \tau_j^{(t-1)}\right)\left\{\lambda\left(-m_{ij}^{(t-1)}, \tau_j^{(t-1)}\right) + m_{ij}^{(t-1)}\tau_j^{(t-1)}\right\}\right] & if\ y_{ij} = 1 \end{cases}$$

## Dynamic Ideal Point Model

다음으로 시간에 따른 ideal points의 변화를 모델링하는 Dynamic Model을 살펴본다. dynamic model에서는 closed form의 EM이 존재하지 않는다. 따라서 variational EM을 사용한다.
$y_{ijt}$를 입법자 $i$가 법안 $j$에 대해 $t$시점에서 찬성표를 던지는지 ($y_{ijt} = 1$) 또는 반대표를 던지는지 ($y_{ijt} = 0$)의 여부를 나타내는 indicator variable이라고 하자. 입법자는 $N$명이 있고 ($i = 1, \cdots, N$), $t$시점이 주어지면 $J_t$개의 roll calls이 있다. 또한 총 $T$시점이 존재한다. 이렇게 notation을 정의하면, 일차원 ideal point model은 아래와 같이 주어진다.

$$Pr(y_{ijt} = 1) = \Phi(\alpha_{jt} + \beta_{jt}x_{it}) = \Phi\left(\tilde{\mathbf{x}}_{it}^T\tilde{\boldsymbol{\beta}}_{jt}\right)$$

따라서 이전에서와 유사하게 latent propensity인 $y^*_{ijt}$을 아래와 같이 정의한다.

$$y^*_{ijt} = \tilde{\mathbf{x}}_{it}^T\tilde{\boldsymbol{\beta}}_{jt} + \epsilon_{ijt}\ \epsilon_{ijt} \underset{\sim}{i.i.d.} N(0,1)$$

$$y_{ijt} = \begin{cases} 1 & if\ y^*_{ijt} > 0 \\ 0 & if\ y^*_{ijt} \leq 0 \end{cases}$$

또한 시간 변수를 추가함으로써 $i$번재 입법자의 ideal point가 시간에 depend하다고 본다. 즉, 아래와 같은 random woalk dependency 구조를 가정한다.

$$x_{it} \mid x_{i,t-1} \underset{\sim}{indep} N(x_{i,t-1}, \omega_x^2)$$

$$for\ t = \underline{T}_i, \underline{T}_i + 1, \cdots \overline{T}_i - 1, \overline{T}_i$$

$$where\ \overline{T}_i:\ last\ time\ period\ the\ legislator\ appears\ in\ data$$

$$\underline{T}_i : \ last\ first\ period\ the\ legislator\ appears\ in\ data$$

$$Assume\ x_{i,\underline{T}_i - 1} \overset{iid}{\sim} N(\mu_x, \Sigma_x)$$

여기서 $\mathbf{x}_i = (x_{i,\underline{T}_i}, \cdots, x_{i,\overline{T}_i})$로 두고 아래와 같이 사후 분포를 정의한다.

$$P\left(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \left\{\tilde{\boldsymbol{\beta}}_j\right\}_{t=1}^T \mid \mathbf{Y}\right)$$

이를 variational EM으로 근사하는 과정은 appendix에 소개되어 있다.

## Hierarchical Ideal Point Model

이제 ideal points을 covariates의 선형 결합으로 모델링해보자. dynamic model과 마찬가지로 EM의 closed form이 존재하지 않는다. 따라서 variational EM을 이용하여 사후분포를 근사한다. 우선 노테이션을 정의하자.

$$y_\ell : \ binary\ random\ variable\ where\ \ell \in \{1, \cdots, L\}$$

$$y_\ell\ represents\ a\ vote\ cast\ by\ legislator\ i[\ell]\ on\ bill\ j[\ell]\ where\ i[\ell] \in \{1, \cdots, N\}, j[\ell] \in \{1, \cdots, J\}$$

$$g[i[\ell]] : \ represents\ the\ group\ membership\ of\ legislator\ i[\ell]\ where\ g[i[\ell]] \in \{1, \cdots, G\}$$

아래의 propensity score인 $y_\ell^*$을 이용하여 $y_\ell = \mathbf{1}\{y_\ell^* > 0\}$으로 정의한다.

$$y_\ell^* = \alpha_{j[\ell]} + \beta_{j[\ell]} x_{i[\ell]} + \epsilon_\ell\ where\ \epsilon_\ell \overset{iid}{\sim} N(0, 1)$$

$$x_{i[\ell]} = \boldsymbol{\gamma}_{g[i[\ell]]}^T \mathbf{z}_{i[\ell]} + \eta_{i[\ell]}\ where\ \eta_{i[\ell]} \overset{indep}{\sim} N(0, \sigma_{g[i[\ell]]}^2)$$

$$\therefore y_\ell^* = \alpha_{j[\ell]} + \beta_{j[\ell]} \boldsymbol{\gamma}_{g[i[\ell]]}^T \mathbf{z}_{i[\ell]} + \beta_{j[\ell]} \eta_{i[\ell]} + \epsilon_\ell$$

여기서 $\boldsymbol{\gamma}_{g[i[\ell]]}$은 M차원의 group specific 계수이고, $\mathbf{z}_{i[\ell]}$은 M차원의 legislator-specific covariates이다. 마지막으로 $\sigma_{g[i[\ell]]}^2$은 group-specific variance이다.
이 모형에 대한 conjugate prior을 아래와 같이 정의하자.

$$\tilde{\boldsymbol{\beta}}_{j[\ell]} \overset{iid}{\sim} N(\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}})$$

$$\boldsymbol{\gamma}_{g[i[\ell]]} \overset{iid}{\sim} N(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$$

$$\sigma_{g[i[\ell]]}^2 \overset{iid}{\sim} IG\left(\frac{\nu_\sigma}{2}, \frac{s_\sigma^2}{2}\right)$$

따라서 사후 분포는 아래와 같다.

$$P\left(\mathbf{Y}^*, \left\{\tilde{\boldsymbol{\beta}}_k\right\}_{k=1}^J, \{\boldsymbol{\gamma}_m\}_{m=1}^G, \{\eta_n\}_{n=1}^N \mid \mathbf{Y}\right)$$

이를 closed form으로 유도할 수 없기 때문에, variational EM를 사용하여 근사를 한다.