Recall: Mixed logit choice probability requires an integral

$$P_{ij}(X, Z; \beta, \mu, \sigma) = \int \frac{\exp(X_i(\beta_j - \beta_J) + \gamma_i(Z_{ij} - Z_{iJ}))}{\sum_k \exp(X_i(\beta_k - \beta_J) + \gamma_i(Z_{ik} - Z_{iJ}))} f(\gamma_i; \mu, \sigma) d\gamma_i$$

Recall: Mixed logit choice probability requires an integral

$$P_{ij}(X, Z; \beta, \mu, \sigma) = \int \frac{\exp(X_i(\beta_j - \beta_J) + \gamma_i(Z_{ij} - Z_{iJ}))}{\sum_k \exp(X_i(\beta_k - \beta_J) + \gamma_i(Z_{ik} - Z_{iJ}))} f(\gamma_i; \mu, \sigma) d\gamma_i$$

This integral has no closed form $\rightarrow$ must approximate numerically

Recall: Mixed logit choice probability requires an integral

$$P_{ij}(X, Z; \beta, \mu, \sigma) = \int \frac{\exp(X_i(\beta_j - \beta_J) + \gamma_i(Z_{ij} - Z_{iJ}))}{\sum_k \exp(X_i(\beta_k - \beta_J) + \gamma_i(Z_{ik} - Z_{iJ}))} f(\gamma_i; \mu, \sigma) d\gamma_i$$

This integral has no closed form $\rightarrow$ must approximate numerically

Classical approach: Quadrature or simulation-based methods

Recall: Mixed logit choice probability requires an integral

$$P_{ij}(X, Z; \beta, \mu, \sigma) = \int \frac{\exp(X_i(\beta_j - \beta_J) + \gamma_i(Z_{ij} - Z_{iJ}))}{\sum_k \exp(X_i(\beta_k - \beta_J) + \gamma_i(Z_{ik} - Z_{iJ}))} f(\gamma_i; \mu, \sigma) d\gamma_i$$

This integral has no closed form $\rightarrow$ must approximate numerically

Classical approach: Quadrature or simulation-based methods

Problem: Maximization can be difficult (local maxima, slow convergence, ...)

Bayesian approach avoids maximization entirely

Instead of finding $\max_\theta \ell(\theta)$, calculate $E[\theta|\text{data}]$

Bayesian approach avoids maximization entirely

Instead of finding $\max_\theta \ell(\theta)$, calculate $E[\theta|\text{data}]$

Key insight: Treat individual γ_i as parameters alongside μ, σ

Bayesian approach avoids maximization entirely

Instead of finding $\max_\theta \ell(\theta)$, calculate $E[\theta|\text{data}]$

Key insight: Treat individual γ_i as parameters alongside μ, σ

Use data augmentation via posterior distribution instead of integrating:

- Classical: $\int L(y_i|\gamma_i)dF(\gamma_i; \mu, \sigma)$ inside likelihood
- Bayesian: Draw γ_i directly, condition on it

Bayesian approach avoids maximization entirely

Instead of finding $\max_\theta \ell(\theta)$, calculate $E[\theta|\text{data}]$

Key insight: Treat individual γ_i as parameters alongside μ, σ

Use data augmentation via posterior distribution instead of integrating:

- Classical: $\int L(y_i|\gamma_i)dF(\gamma_i;\mu,\sigma)$ inside likelihood
- Bayesian: Draw γ_i directly, condition on it

No simulation of choice probabilities needed $\rightarrow$ just evaluate logit formula at drawn γ_i

Bayesian Gibbs sampling is like EM with continuous types

Both treat γ_i as latent $\rightarrow$ fill in γ_i instead of integrating it out

EM (you know this):

E-step: $E[\gamma_i | y_i, \mu^{(m)}, \sigma^{(m)}]$

M-step: Maximize using expectations

$\rightarrow$ MLE

Bayesian Gibbs (new):

Draw $\gamma_i \sim p(\gamma_i | y_i, \mu, \sigma)$

Draw $\mu, \sigma \sim p(\mu, \sigma | \{\gamma_i\})$

$\rightarrow$ Posterior mean $\approx$ MLE in limit

Bayesian Gibbs sampling is like EM with continuous types

Both treat γ_i as latent $\rightarrow$ fill in γ_i instead of integrating it out

EM (you know this):

E-step: $E[\gamma_i | y_i, \mu^{(m)}, \sigma^{(m)}]$

M-step: Maximize using expectations

$\rightarrow$ MLE

Bayesian Gibbs (new):

Draw $\gamma_i \sim p(\gamma_i | y_i, \mu, \sigma)$

Draw $\mu, \sigma \sim p(\mu, \sigma | \{\gamma_i\})$

$\rightarrow$ Posterior mean $\approx$ MLE in limit

Difference: EM integrates analytically, Gibbs samples via MCMC

Bayesian procedure for mixed logit:

Gibbs sampling - iteratively draw from conditional posteriors:

Bayesian procedure for mixed logit:

Gibbs sampling - iteratively draw from conditional posteriors:

1. Draw each γ_i from posterior distribution $p(\gamma_i|\mu, \sigma, y_i)$
   - Just evaluate logit likelihood at current μ, σ
   - No integral needed

Bayesian procedure for mixed logit:

Gibbs sampling - iteratively draw from conditional posteriors:

1. Draw each γ_i from posterior distribution $p(\gamma_i|\mu, \sigma, y_i)$
   - Just evaluate logit likelihood at current μ, σ
   - No integral needed

2. Draw μ from $p(\mu|\sigma, \{\gamma_i\})$
   - Essentially: average of the γ_i's

Bayesian procedure for mixed logit:

Gibbs sampling - iteratively draw from conditional posteriors:

1. Draw each γ_i from posterior distribution $p(\gamma_i|\mu, \sigma, y_i)$
   - Just evaluate logit likelihood at current μ, σ
   - No integral needed

2. Draw μ from $p(\mu|\sigma, \{\gamma_i\})$
   - Essentially: average of the γ_i's

3. Draw σ from $p(\sigma|\mu, \{\gamma_i\})$
   - Essentially: variance of the γ_i's around μ

Bayesian procedure for mixed logit:

Gibbs sampling - iteratively draw from conditional posteriors:

1. Draw each γ_i from posterior distribution $p(\gamma_i|\mu, \sigma, y_i)$
   - Just evaluate logit likelihood at current μ, σ
   - No integral needed

2. Draw μ from $p(\mu|\sigma, \{\gamma_i\})$
   - Essentially: average of the γ_i's

3. Draw σ from $p(\sigma|\mu, \{\gamma_i\})$
   - Essentially: variance of the γ_i's around μ

Repeat until convergence $\rightarrow$ posterior mean of draws $=$ estimates

How do we know the posterior distributions?

Bayes' rule:

$$p(\theta|\text{data}) \propto L(\text{data}|\theta) \times \pi(\theta)$$

How do we know the posterior distributions?

Bayes' rule:

$$p(\theta|\text{data}) \propto L(\text{data}|\theta) \times \pi(\theta)$$

For γ_i:

- $p(\gamma_i|\mu, \sigma, y_i) \propto L(y_i|\gamma_i) \times \pi(\gamma_i|\mu, \sigma)$
- No closed form $\rightarrow$ sample via Metropolis-Hastings (accept/reject)

How do we know the posterior distributions?

Bayes' rule:

$$p(\theta|\text{data}) \propto L(\text{data}|\theta) \times \pi(\theta)$$

For γ_i:

- $p(\gamma_i|\mu, \sigma, y_i) \propto L(y_i|\gamma_i) \times \pi(\gamma_i|\mu, \sigma)$
- No closed form $\rightarrow$ sample via Metropolis-Hastings (accept/reject)

For μ, σ:

- Conjugate priors $\rightarrow$ closed form posteriors
- $p(\mu|\sigma, \{\gamma_i\}) \sim$ Normal
- $p(\sigma|\mu, \{\gamma_i\}) \sim$ Inverse Wishart
- Direct sampling (no M-H needed)
- (though usually use M-H for all parameters)

Advantages of Bayesian approach:

Advantages of Bayesian approach:

- No maximization $\rightarrow$ no convergence issues, starting values less critical

Advantages of Bayesian approach:

- No maximization $\rightarrow$ no convergence issues, starting values less critical
- Number of simulation draws R need not grow with sample size

Advantages of Bayesian approach:

- No maximization $\rightarrow$ no convergence issues, starting values less critical
- Number of simulation draws R need not grow with sample size
- Naturally provides full posterior distribution of γ_i (not just point estimates)

Advantages of Bayesian approach:

- No maximization $\rightarrow$ no convergence issues, starting values less critical
- Number of simulation draws R need not grow with sample size
- Naturally provides full posterior distribution of γ_i (not just point estimates)
- Asymptotically equivalent to MLE (Bernstein-von Mises theorem)

Advantages of Bayesian approach:

- No maximization $\rightarrow$ no convergence issues, starting values less critical
- Number of simulation draws R need not grow with sample size
- Naturally provides full posterior distribution of γ_i (not just point estimates)
- Asymptotically equivalent to MLE (Bernstein-von Mises theorem)

Disadvantages:

- Different convergence issue: to posterior distribution (burn-in)
- Slower for some specifications (fixed coefficients, bounded distributions)