

Bayesian Hierarchical Modeling

강의 목표

- ▶ 계층 모형 소개
- ▶ 계층 모형에 대한 베이지스 추론
 - Normal-Normal Model
 - Poisson-Gamma Model

Motivations for the Hierarchical Model

- ▶ Powerful technique for describing complex models.

Motivations for the Hierarchical Model

- ▶ Powerful technique for describing complex models.
- ▶ Main idea is to break the model down into smaller easier understood pieces.

Motivations for the Hierarchical Model

- ▶ Powerful technique for describing complex models.
- ▶ Main idea is to break the model down into smaller easier understood pieces.
- ▶ Modeling with multiple levels that estimates the parameters of the posterior distribution.

Motivations for the Hierarchical Model

- ▶ Powerful technique for describing complex models.
- ▶ Main idea is to break the model down into smaller easier understood pieces.
- ▶ Modeling with multiple levels that estimates the parameters of the posterior distribution.
- ▶ e.g., Normal distribution:

Motivations for the Hierarchical Model

- ▶ Powerful technique for describing complex models.
- ▶ Main idea is to break the model down into smaller easier understood pieces.
- ▶ Modeling with multiple levels that estimates the parameters of the posterior distribution.
- ▶ e.g., Normal distribution:

$$\pi(\mu, \sigma^2 \mid x) = f(x \mid \mu, \sigma^2) \pi(\mu \mid \sigma^2) \pi(\sigma^2)$$

Remarks

- ▶ Actually all of the models we have seen so far have been hierarchical, but most only had **two** levels to the hierarchy.

Remarks

- ▶ Actually all of the models we have seen so far have been hierarchical, but most only had **two** levels to the hierarchy.
- ▶ e.g., Beta distribution:

$$\pi(p \mid x) = f(x \mid p)\pi(p).$$

Remarks

- ▶ Actually all of the models we have seen so far have been hierarchical, but most only had **two** levels to the hierarchy.
- ▶ e.g., Beta distribution:

$$\pi(p \mid x) = f(x \mid p)\pi(p).$$

- ▶ There may be a hierarchical structure with many levels.

Remarks

- ▶ Actually all of the models we have seen so far have been hierarchical, but most only had **two** levels to the hierarchy.
- ▶ e.g., Beta distribution:

$$\pi(p \mid x) = f(x \mid p)\pi(p).$$

- ▶ There may be a hierarchical structure with many levels.
- ▶ e.g., Beta distribution:

$$\pi(p \mid x) = f(x \mid p)Beta(a, b)\pi(a)\pi(b).$$

Why go hierarchical?

- ▶ Non-hierarchical models with few parameters generally don't fit the data well.

Why go hierarchical?

- ▶ Non-hierarchical models with few parameters generally don't fit the data well.
- ▶ Non-hierarchical models with many parameters then to fit the data well, but have poor predictive ability (overfitting).

Why go hierarchical?

- ▶ Non-hierarchical models with few parameters generally don't fit the data well.
- ▶ Non-hierarchical models with many parameters then to fit the data well, but have poor predictive ability (overfitting).
- ▶ Hierarchical models can often fit data with a small number of parameters but can also do well in prediction.

Hierarchical Model Example

- ▶ 다음과 같은 집단으로부터 자료를 수집한다고 하자.
 - 강서구에 사는 주민들
 - 대한민국 고등학교 학생들

Hierarchical Model Example

- ▶ 다음과 같은 집단으로부터 자료를 수집한다고 하자.
 - 강서구에 사는 주민들
 - 대한민국 고등학교 학생들
- ▶ 자료수집 대상을 보면 전체 집단 내에 여러 개의 소집단이 존재한다.

Hierarchical Model Example

- ▶ 다음과 같은 집단으로부터 자료를 수집한다고 하자.
 - 강서구에 사는 주민들
 - 대한민국 고등학교 학생들
- ▶ 자료수집 대상을 보면 전체 집단 내에 여러 개의 소집단이 존재한다.
 1. 강서구: 화곡동, 방화동, 목동 등등.

Hierarchical Model Example

- ▶ 다음과 같은 집단으로부터 자료를 수집한다고 하자.
 - 강서구에 사는 주민들
 - 대한민국 고등학교 학생들
- ▶ 자료수집 대상을 보면 전체 집단 내에 여러 개의 소집단이 존재한다.
 1. 강서구: 화곡동, 방화동, 목동 등등.
 2. 대한민국 고등학교: 서울소재 일반고, 지방소재 일반고, 특목고 등등.

Hierarchical Model Example and its Assumptions

- ▶ 전체 집단 내에 K 개의 소집단이 존재한다고 하자.

Hierarchical Model Example and its Assumptions

- ▶ 전체 집단 내에 K 개의 소집단이 존재한다고 하자.
- ▶ i 번째 소집단에서 n_i 개의 관측치 x_{i1}, \dots, x_{i,n_i} 가 얻어진다고 하자.

Hierarchical Model Example and its Assumptions

- ▶ 전체 집단 내에 K 개의 소집단이 존재한다고 하자.
- ▶ i 번째 소집단에서 n_i 개의 관측치 X_{i1}, \dots, X_{i,n_i} 가 얻어진다고 하자.
- ▶ 관측치 X_{i1}, \dots, X_{i,n_i} 가 독립적으로 모수 θ_i 를 갖는 분포 $f(x | \theta_i)$ 를 따른다고 가정하자.

$$X_{i1}, \dots, X_{i,n_i} | \theta_i \stackrel{\text{iid}}{\sim} f(x | \theta_i).$$

Hierarchical Model Example and its Assumptions

- ▶ 전체 집단 내에 K 개의 소집단이 존재한다고 하자.
- ▶ i 번째 소집단에서 n_i 개의 관측치 X_{i1}, \dots, X_{i,n_i} 가 얻어진다고 하자.
- ▶ 관측치 X_{i1}, \dots, X_{i,n_i} 가 독립적으로 모수 θ_i 를 갖는 분포 $f(x | \theta_i)$ 를 따른다고 가정하자.

$$X_{i1}, \dots, X_{i,n_i} | \theta_i \stackrel{\text{iid}}{\sim} f(x | \theta_i).$$

- ▶ $\theta_1, \dots, \theta_K$ 들이 모수 ξ 가 주어졌을 때, 독립적으로 동일한 분포를 따른다고 가정하자.

Hierarchical Model Example and its Assumptions

- ▶ 전체 집단 내에 K 개의 소집단이 존재한다고 하자.
- ▶ i 번째 소집단에서 n_i 개의 관측치 X_{i1}, \dots, X_{i,n_i} 가 얻어진다고 하자.
- ▶ 관측치 X_{i1}, \dots, X_{i,n_i} 가 독립적으로 모수 θ_i 를 갖는 분포 $f(x | \theta_i)$ 를 따른다고 가정하자.

$$X_{i1}, \dots, X_{i,n_i} | \theta_i \stackrel{\text{iid}}{\sim} f(x | \theta_i).$$

- ▶ $\theta_1, \dots, \theta_K$ 들이 모수 ξ 가 주어졌을 때, 독립적으로 동일한 분포를 따른다고 가정하자.

$$\theta_i | \xi \stackrel{\text{iid}}{\sim} f(\theta | \xi).$$

Hierarchical Model Example: Graph

- ▶ 정규분포의 경우를 가정하면,

$$X_{i1}, \dots, X_{i, n_i} \mid \theta_i, \sigma^2 \stackrel{\text{iid}}{\sim} N(\theta_i, \sigma^2)$$

$$\theta_1, \dots, \theta_K \mid \mu, \tau^2 \stackrel{\text{iid}}{\sim} N(\mu, \tau^2)$$

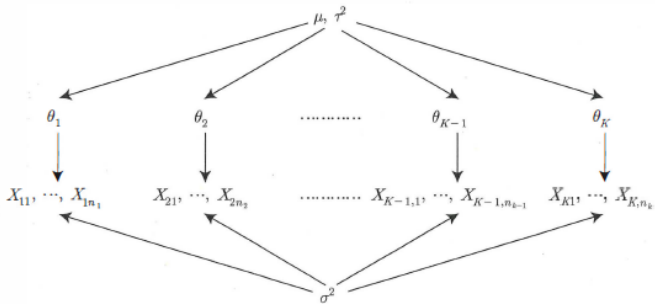


Figure: 계층 모형의 구조

Hierarchical Model Motivation: ANOVA

- ▶ 소집단의 평균 $\theta_1, \dots, \theta_K$ 의 추정이 목표.

Hierarchical Model Motivation: ANOVA

- ▶ 소집단의 평균 $\theta_1, \dots, \theta_K$ 의 추정이 목표.
- ▶ 여러집단의 평균을 비교하는 방법: 분산분석(ANOVA).

1) 소집단은 서로 연관성이 없음.

$$\hat{\theta}_i = \bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_j} x_{ij}, \quad i = 1, \dots, K.$$

Hierarchical Model Motivation: ANOVA

- ▶ 소집단의 평균 $\theta_1, \dots, \theta_K$ 의 추정이 목표.
- ▶ 여러집단의 평균을 비교하는 방법: 분산분석(ANOVA).

1) 소집단은 서로 연관성이 없음.

$$\hat{\theta}_i = \bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_j} x_{ij}, \quad i = 1, \dots, K.$$

2) 소집단들은 동질적.

$$\hat{\theta}_1 = \hat{\theta}_2 = \dots = \hat{\theta}_K = \bar{x}_{..} = \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij} / \sum_{i=1}^K n_i.$$

Hierarchical Model Motivation: ANOVA

- ▶ 소집단의 평균 $\theta_1, \dots, \theta_K$ 의 추정이 목표.
- ▶ 여러집단의 평균을 비교하는 방법: 분산분석(ANOVA).

1) 소집단은 서로 연관성이 없음.

$$\hat{\theta}_i = \bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_j} x_{ij}, \quad i = 1, \dots, K.$$

2) 소집단들은 동질적.

$$\hat{\theta}_1 = \hat{\theta}_2 = \dots = \hat{\theta}_K = \bar{x}_{..} = \sum_{i=1}^K \sum_{j=1}^{n_i} x_{ij} / \sum_{i=1}^K n_i.$$

- ▶ 주어진 자료에 대하여 어떤 가정이 더 적절한가에 대한 답을 ANOVA(Analysis of Variance) 분석이 제공.

Hierarchical Model Motivation: ANOVA

- ▶ $n_i = n$ 이라고 하면, 다음과 같은 ANOVA 표가 만들어진다.

요인	SS	MS	E(MS)
Between	$SSB = n \sum_{i=1}^K (\bar{x}_i - \bar{x}_{..})^2$	$MSB = SSB / (K - 1)$	$\sigma^2 + n\tau^2$
Within	$SSW = \sum_{i=1}^K \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2$	$MSW = SSW / K(n - 1)$	σ^2

- ▶ 만약 각 그룹의 평균 $\bar{x}_1, \dots, \bar{x}_K$ 가 서로 많이 다르면,

$$MSB > MSW.$$

- ▶ 즉, $\tau^2 > 0$ 을 의미하고 θ_i 들은 서로 다르다고 결론.
- ▶ 만약 $\bar{x}_1, \dots, \bar{x}_K$ 가 서로 비슷하면

$$MSB \approx MSW$$

- ▶ 즉, $\tau^2 \approx 0$ 을 의미하므로 θ_i 들이 서로 같다고 결론.

Weakness of ANOVA

- ▶ ANOVA 분석법을 보면 두 극단 중 하나를 선택한다.

Weakness of ANOVA

- ▶ ANOVA 분석법을 보면 두 극단 중 하나를 선택한다.
 1. 소집단을 완전히 분리하여 따로 취급하고 따라서 정보의 공유가 하나도 없음.

Weakness of ANOVA

- ▶ ANOVA 분석법을 보면 두 극단 중 하나를 선택한다.
 1. 소집단을 완전히 분리하여 따로 취급하고 따라서 정보의 공유가 하나도 없음.
 2. 소집단의 존재를 무시하고 전체를 하나의 동질적(homogeneous) 모집단으로 보고 정보를 공유.

Weakness of ANOVA

- ▶ ANOVA 분석법을 보면 두 극단 중 하나를 선택한다.
 1. 소집단을 완전히 분리하여 따로 취급하고 따라서 정보의 공유가 하나도 없음.
 2. 소집단의 존재를 무시하고 전체를 하나의 동질적(homogeneous) 모집단으로 보고 정보를 공유.
- ▶ 실제 상황에서는 이 양 극단의 중간이 존재할 수 있다.

Weakness of ANOVA

- ▶ ANOVA 분석법을 보면 두 극단 중 하나를 선택한다.
 1. 소집단을 완전히 분리하여 따로 취급하고 따라서 정보의 공유가 하나도 없음.
 2. 소집단의 존재를 무시하고 전체를 하나의 동질적(homogeneous) 모집단으로 보고 정보를 공유.
- ▶ 실제 상황에서는 이 양 극단의 중간이 존재할 수 있다.
- ▶ 완전히 독립적이지도 않고 그러나 완전히 동질적이지도 않은, 소집단 고유의 특성을 지니면서도 전체 집단이 공통으로 갖는 어떤 특성들을 공유.

Why Hierarchical Model

- ▶ 소집단들이 특정 성질을 공유 하는 경우 θ_i 를 \bar{x}_i 와 \bar{x}_\cdot 의 가중평균으로 구하는 것이 좋음.

Why Hierarchical Model

- ▶ 소집단들이 특정 성질을 공유 하는 경우 θ_i 를 \bar{x}_i 와 \bar{x}_\cdot 의 가중평균으로 구하는 것이 좋음.
- ▶ 더 나아가 사전정보에서 θ_i 들의 추정치로 μ_0 를 얻을 수 있다면 θ_i 를 $\bar{x}_i, \bar{x}_\cdot, \mu_0$ 의 가중평균으로 구하면 합리적일 것이다. 즉,

$$\hat{\theta}_i = w_{1i} \times \bar{x}_i + w_{2i} \times \bar{x}_\cdot + w_{3i} \times \mu_0$$

such that $w_{1i} + w_{2i} + w_{3i} = 1$.

Why Hierarchical Model

- ▶ 소집단들이 특정 성질을 공유 하는 경우 θ_i 를 \bar{x}_i 와 \bar{x}_\cdot 의 가중평균으로 구하는 것이 좋음.
- ▶ 더 나아가 사전정보에서 θ_i 들의 추정치로 μ_0 를 얻을 수 있다면 θ_i 를 $\bar{x}_i, \bar{x}_\cdot, \mu_0$ 의 가중평균으로 구하면 합리적일 것이다. 즉,

$$\hat{\theta}_i = w_{1i} \times \bar{x}_i + w_{2i} \times \bar{x}_\cdot + w_{3i} \times \mu_0$$

such that $w_{1i} + w_{2i} + w_{3i} = 1$.

- ▶ 여기에서 가중치 w_{1i}, w_{2i}, w_{3i} 는 그룹 내 분산, 그룹 간 분산, 표본크기, 그리고 사전정보의 정확도에 의존.

Why Hierarchical Model

- ▶ 계층 모형에 대한 베イズ 추론은 이와 같은 가중평균을 θ_i 의 추정치로 제공한다.

Why Hierarchical Model

- ▶ 계층 모형에 대한 베이즈 추론은 이와 같은 가중평균을 θ_i 의 추정치로 제공한다.
- ▶ 소집단의 독립성과 동질성의 정도, 그리고 표본자료 대비 사전정보의 정확도에 따라 융통성 있게 가중치가 조정되면서 합리적인 θ_i 의 추정치를 얻을 수 있다.

How to Estimate θ_i and ω_i

- ▶ How to estimate θ_i and ω_i ?

How to Estimate θ_i and ω_i

- ▶ How to estimate θ_i and ω_i ?
- ▶ There are unknown parameters σ^2, τ^2, μ in the model.

How to Estimate θ_i and ω_i

- ▶ How to estimate θ_i and ω_i ?
- ▶ There are unknown parameters σ^2, τ^2, μ in the model.
- ▶ As conjugate priors, suppose that

$$\sigma^2 \sim IG(a, b),$$

$$\tau^2 \sim IG(c, d),$$

$$\mu \mid \tau^2 \sim N(\mu_0, \tau^2/k_0).$$

Model Estimation

- The model is:

$$X_{ij} \mid \theta_i, \sigma^2 \sim N(\theta_i, \sigma^2),$$

$$\theta_i \mid \mu, \tau^2 \sim N(\mu, \tau^2),$$

$$\sigma^2 \sim IG(a, b),$$

$$\mu \mid \tau^2 \sim N(\mu_0, \tau^2/k_0),$$

$$\tau^2 \sim IG(c, d).$$

Model Interpretation

- ▶ If $\tau^2 = 0$, $\theta_1 = \dots = \theta_k = \mu$.

Model Interpretation

- ▶ If $\tau^2 = 0$, $\theta_1 = \dots = \theta_k = \mu$.
- ▶ It means all subgroups are homogeneous.

$$\hat{\theta}_1 = \dots = \hat{\theta}_k = \bar{x}_{..}$$

Model Interpretation

- ▶ If $\tau^2 = 0$, $\theta_1 = \dots = \theta_k = \mu$.
- ▶ It means all subgroups are homogeneous.

$$\hat{\theta}_1 = \dots = \hat{\theta}_k = \bar{x}_{..}$$

- ▶ If $\tau^2 = \infty$, θ_i are heterogeneous.

$$\hat{\theta}_i = \bar{x}_{i.} \quad \text{for all } i = 1, 2, \dots, K$$

Model Interpretation

- ▶ If $\tau^2 = 0$, $\theta_1 = \dots = \theta_k = \mu$.
- ▶ It means all subgroups are homogeneous.

$$\hat{\theta}_1 = \dots = \hat{\theta}_k = \bar{x}_{..}$$

- ▶ If $\tau^2 = \infty$, θ_i are heterogeneous.

$$\hat{\theta}_i = \bar{x}_i. \quad \text{for all } i = 1, 2, \dots, K$$

- ▶ If $0 < \tau^2 < \infty$, subgroups are partially different, but share commons.

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given x .
- ▶ Find the posterior distribution:

$$\pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x})$$

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given x .
- ▶ Find the posterior distribution:

$$\pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x})$$

$$\propto f(\mathbf{x} \mid \Theta, \mu, \sigma^2, \tau^2) f(\Theta \mid \mu, \sigma^2, \tau^2) \pi(\mu, \sigma^2, \tau^2)$$

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given \mathbf{x} .
- ▶ Find the posterior distribution:

$$\begin{aligned}\pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x}) \\ &\propto f(\mathbf{x} \mid \Theta, \mu, \sigma^2, \tau^2) f(\Theta \mid \mu, \sigma^2, \tau^2) \pi(\mu, \sigma^2, \tau^2) \\ &= f(\mathbf{x} \mid \Theta, \sigma^2) f(\Theta \mid \mu, \tau^2) \pi(\sigma^2) \pi(\tau^2)\end{aligned}$$

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given \mathbf{x} .
- ▶ Find the posterior distribution:

$$\begin{aligned} & \pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x}) \\ & \propto f(\mathbf{x} \mid \Theta, \mu, \sigma^2, \tau^2) f(\Theta \mid \mu, \sigma^2, \tau^2) \pi(\mu, \sigma^2, \tau^2) \\ & = f(\mathbf{x} \mid \Theta, \sigma^2) f(\Theta \mid \mu, \tau^2) \pi(\sigma^2) \pi(\tau^2) \\ & = \prod_{i=1}^K \prod_{j=1}^{n_i} f(x_{ij} \mid \theta_i, \sigma^2) \prod_{i=1}^K f(\theta_i \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\sigma^2) \pi(\tau^2) \end{aligned}$$

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given \mathbf{x} .
- ▶ Find the posterior distribution:

$$\begin{aligned} & \pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x}) \\ & \propto f(\mathbf{x} \mid \Theta, \mu, \sigma^2, \tau^2) f(\Theta \mid \mu, \sigma^2, \tau^2) \pi(\mu, \sigma^2, \tau^2) \\ & = f(\mathbf{x} \mid \Theta, \sigma^2) f(\Theta \mid \mu, \tau^2) \pi(\sigma^2) \pi(\tau^2) \\ & = \prod_{i=1}^K \prod_{j=1}^{n_i} f(x_{ij} \mid \theta_i, \sigma^2) \prod_{i=1}^K f(\theta_i \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\sigma^2) \pi(\tau^2) \\ & \propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2} (\tau^2)^{-\frac{K}{2}} e^{-\frac{1}{2\tau^2} \sum_{i=1}^K (\theta_i - \mu)^2} \end{aligned}$$

Model Estimation

- ▶ **Ultimate Goal:** Estimate $\Theta = (\theta_1, \dots, \theta_K), \mu, \sigma^2, \tau^2$ given \mathbf{x} .
- ▶ Find the posterior distribution:

$$\begin{aligned} & \pi(\Theta, \mu, \sigma^2, \tau^2 \mid \mathbf{x}) \\ & \propto f(\mathbf{x} \mid \Theta, \mu, \sigma^2, \tau^2) f(\Theta \mid \mu, \sigma^2, \tau^2) \pi(\mu, \sigma^2, \tau^2) \\ & = f(\mathbf{x} \mid \Theta, \sigma^2) f(\Theta \mid \mu, \tau^2) \pi(\sigma^2) \pi(\tau^2) \\ & = \prod_{i=1}^K \prod_{j=1}^{n_i} f(x_{ij} \mid \theta_i, \sigma^2) \prod_{i=1}^K f(\theta_i \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\sigma^2) \pi(\tau^2) \\ & \propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2} (\tau^2)^{-\frac{K}{2}} e^{-\frac{1}{2\tau^2} \sum_{i=1}^K (\theta_i - \mu)^2} \\ & \quad \times (\tau^2)^{-\frac{1}{2}} e^{-\frac{k_0}{2\tau^2} (\mu - \mu_0)^2} (\sigma^2)^{-a-1} e^{-b/\sigma^2} (\tau^2)^{-c-1} e^{-d/\tau^2}. \end{aligned}$$

Problem of Model Estimation

- **Problem:** It is very difficult to sample from the posterior distribution for $\Theta, \mu, \sigma^2, \tau^2$.

Problem of Model Estimation

- ▶ **Problem:** It is very difficult to sample from the posterior distribution for $\Theta, \mu, \sigma^2, \tau^2$.
- ▶ **Solution:** Gibbs sampling:

Problem of Model Estimation

- ▶ **Problem:** It is very difficult to sample from the posterior distribution for $\Theta, \mu, \sigma^2, \tau^2$.
- ▶ **Solution:** Gibbs sampling:
- ▶ Full conditional posterior distributions:

$$\pi(\theta_i \mid \mathbf{x}, \mu, \sigma^2, \tau^2)$$

Problem of Model Estimation

- ▶ **Problem:** It is very difficult to sample from the posterior distribution for $\Theta, \mu, \sigma^2, \tau^2$.
- ▶ **Solution:** Gibbs sampling:
- ▶ Full conditional posterior distributions:

$$\pi(\theta_i \mid \mathbf{x}, \mu, \sigma^2, \tau^2) \propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\Theta \mid \mu, \tau^2)$$

Problem of Model Estimation

- ▶ **Problem:** It is very difficult to sample from the posterior distribution for $\Theta, \mu, \sigma^2, \tau^2$.
- ▶ **Solution:** Gibbs sampling:
- ▶ Full conditional posterior distributions:

$$\begin{aligned}\pi(\theta_i \mid \mathbf{x}, \mu, \sigma^2, \tau^2) &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\Theta \mid \mu, \tau^2) \\ &\sim N\left(\frac{\frac{1}{\sigma^2/n_i} \bar{x}_i + \frac{1}{\tau^2} \mu}{\frac{1}{\sigma^2/n_i} + \frac{1}{\tau^2}}, \left(\frac{1}{\sigma^2/n_i} + \frac{1}{\tau^2}\right)^{-1}\right)\end{aligned}$$

Full Conditional Posterior Distributions

- ▶ Full conditional posterior distributions:

Full Conditional Posterior Distributions

- ▶ Full conditional posterior distributions:

$$\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) = \pi(\sigma^2 \mid \mathbf{x}, \Theta)$$

Full Conditional Posterior Distributions

- ▶ Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2)\end{aligned}$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) = \pi(\sigma^2 \mid \mathbf{x}, \Theta)$$

$$\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}}$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}} \\ &\sim IG \left(\frac{1}{2} \sum_{i=1}^K n_i + a, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right)\end{aligned}$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}} \\ &\sim IG \left(\frac{1}{2} \sum_{i=1}^K n_i + a, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right)\end{aligned}$$

- Full conditional posterior distributions:

$$\pi(\tau^2 \mid \mathbf{x}, \theta, \mu, \sigma^2) \propto f(\mathbf{x} \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\tau^2)$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}} \\ &\sim IG \left(\frac{1}{2} \sum_{i=1}^K n_i + a, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right)\end{aligned}$$

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\tau^2 \mid \mathbf{x}, \theta, \mu, \sigma^2) &\propto f(\mathbf{x} \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\tau^2) \\ &\propto (\tau^2)^{-\left(\frac{K}{2} + \frac{1}{2} + c\right) - 1} e^{-\frac{1}{\tau^2} \left\{ \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + \frac{1}{2} k_0 (\mu - \mu_0^2) + d \right\}}\end{aligned}$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}} \\ &\sim IG \left(\frac{1}{2} \sum_{i=1}^K n_i + a, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right)\end{aligned}$$

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\tau^2 \mid \mathbf{x}, \theta, \mu, \sigma^2) &\propto f(\mathbf{x} \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\tau^2) \\ &\propto (\tau^2)^{-\left(\frac{K}{2} + \frac{1}{2} + c\right) - 1} e^{-\frac{1}{\tau^2} \left\{ \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + \frac{1}{2} k_0 (\mu - \mu_0^2) + d \right\}} \\ &\sim IG \left(\frac{K+1}{2} + c, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + \frac{k_0}{2} (\mu - \mu_0^2) + d \right)\end{aligned}$$

Full Conditional Posterior Distributions

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\sigma^2 \mid \mathbf{x}, \Theta, \mu, \tau^2) &= \pi(\sigma^2 \mid \mathbf{x}, \Theta) \\ &\propto f(\mathbf{x} \mid \Theta, \sigma^2) \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{1}{2} \sum_{i=1}^K n_i - a - 1} e^{-\frac{1}{\sigma^2} \left\{ \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right\}} \\ &\sim IG \left(\frac{1}{2} \sum_{i=1}^K n_i + a, \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2 + b \right)\end{aligned}$$

- Full conditional posterior distributions:

$$\begin{aligned}\pi(\tau^2 \mid \mathbf{x}, \theta, \mu, \sigma^2) &\propto f(\mathbf{x} \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \pi(\tau^2) \\ &\propto (\tau^2)^{-\left(\frac{K}{2} + \frac{1}{2} + c\right) - 1} e^{-\frac{1}{\tau^2} \left\{ \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + \frac{1}{2} k_0 (\mu - \mu_0^2) + d \right\}} \\ &\sim IG \left(\frac{K+1}{2} + c, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2 + \frac{k_0}{2} (\mu - \mu_0^2) + d \right)\end{aligned}$$

Model Estimation

► $\theta_1, \dots, \theta_K \mid \mu, \tau^2 \sim N(\mu, \tau^2).$

Model Estimation

- ▶ $\theta_1, \dots, \theta_K \mid \mu, \tau^2 \sim N(\mu, \tau^2)$.
- ▶ Let $\bar{\theta} = \frac{1}{K} \sum_{i=1}^K \theta_i$. Then,

$$\bar{\theta} \mid \mu, \tau^2 \sim N(\mu, \tau^2/K),$$

$$\mu \mid \tau^2 \sim N(\mu_0, \tau^2/k_0)$$

Model Estimation

► $\theta_1, \dots, \theta_K \mid \mu, \tau^2 \sim N(\mu, \tau^2)$.

► Let $\bar{\theta} = \frac{1}{K} \sum_{i=1}^K \theta_i$. Then,

$$\bar{\theta} \mid \mu, \tau^2 \sim N(\mu, \tau^2/K),$$

$$\mu \mid \tau^2 \sim N(\mu_0, \tau^2/k_0)$$

► Therefore,

$$\begin{aligned} \pi(\mu \mid \mathbf{x}, \Theta, \sigma^2, \tau^2) &\propto f(\Theta \mid \mu, \tau^2) \pi(\mu \mid \tau^2) \\ &\sim N\left(\frac{k_0 \mu_0 + K \bar{\theta}}{K + k_0}, \frac{\tau^2}{k_0 + K}\right) \end{aligned}$$

Non-Informative Hyper-parameter

- ▶ 계층 모형에서 사전분포의 모수인 초모수(hyper-parameter)에 무정보 사전분포를 사용하는 경우가 많다.

Non-Informative Hyper-parameter

- ▶ 계층 모형에서 사전분포의 모수인 초모수(hyper-parameter)에 무정보 사전분포를 사용하는 경우가 많다.
- ▶ 위 모형에서 초모수인 μ, τ^2 에 무정보 사전분포 $\pi(\mu, \tau^2) = \pi(\mu)\pi(\tau^2) = 1/\tau^2$ 를 사용하는 경우를 살펴보자.

Non-Informative Hyper-parameter

- ▶ 깃스 표본 기법 절차 중 θ_i 와 σ^2 의 조건부 사후분포는 변함이 없다.

Non-Informative Hyper-parameter

- ▶ 깃스 표본 기법 절차 중 θ_i 와 σ^2 의 조건부 사후분포는 변함이 없다.
- ▶ τ^2 의 조건부 사후분포는 $IG\left(\frac{K}{2}, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2\right)$

Non-Informative Hyper-parameter

- ▶ 깃스 표본 기법 절차 중 θ_i 와 σ^2 의 조건부 사후분포는 변함이 없다.
- ▶ τ^2 의 조건부 사후분포는 $IG\left(\frac{K}{2}, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2\right)$
- ▶ μ 의 조건부 사후분포는 $N(\bar{\theta}, \tau^2/K)$

Non-Informative Hyper-parameter

- ▶ 깁스 표본 기법 절차 중 θ_i 와 σ^2 의 조건부 사후분포는 변함이 없다.
- ▶ τ^2 의 조건부 사후분포는 $IG\left(\frac{K}{2}, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2\right)$
- ▶ μ 의 조건부 사후분포는 $N(\bar{\theta}, \tau^2/K)$
- ▶ 분산 σ^2 이 알려진 경우 또는 표본크기가 충분히 큰 경우 σ^2 을 변수가 아닌 고정된 상수로 놓을 수 있다.

Non-Informative Hyper-parameter

- ▶ 깃스 표본 기법 절차 중 θ_i 와 σ^2 의 조건부 사후분포는 변함이 없다.
- ▶ τ^2 의 조건부 사후분포는 $IG\left(\frac{K}{2}, \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu)^2\right)$
- ▶ μ 의 조건부 사후분포는 $N(\bar{\theta}, \tau^2/K)$
- ▶ 분산 σ^2 이 알려진 경우 또는 표본크기가 충분히 큰 경우 σ^2 을 변수가 아닌 고정된 상수로 놓을 수 있다.
- ▶ σ^2 이 상수로 주어진 경우에는 위의 깃스 표본기법 절차에서 σ^2 추출 부분을 생략하면 된다.

Example for Hierarchical Model

- ▶ 2009년도에 실시된 수능시험에서 수리영역 점수에 계층모형을 적용해 보자.

Example for Hierarchical Model

- ▶ 2009년도에 실시된 수능시험에서 수리영역 점수에 계층모형을 적용해 보자.
- ▶ 인터넷과 신문에 수리영역 평균 상위 100개 학교의 평균과 표준편차가 발표되었는데 이 중 2개 학교를 제외한 98개 학교에서 대략적인 응시자수를 파악할 수 있었다.

Example for Hierarchical Model

- ▶ 2009년도에 실시된 수능시험에서 수리영역 점수에 계층모형을 적용해 보자.
- ▶ 인터넷과 신문에 수리영역 평균 상위 100개 학교의 평균과 표준편차가 발표되었는데 이 중 2개 학교를 제외한 98개 학교에서 대략적인 응시자수를 파악할 수 있었다.
- ▶ 데이터 x_{ij} 는 i 번째 학교, j 번째 학생의 점수이다.

Example for Hierarchical Model

- ▶ 2009년도에 실시된 수능시험에서 수리영역 점수에 계층모형을 적용해 보자.
- ▶ 인터넷과 신문에 수리영역 평균 상위 100개 학교의 평균과 표준편차가 발표되었는데 이 중 2개 학교를 제외한 98개 학교에서 대략적인 응시자수를 파악할 수 있었다.
- ▶ 데이터 x_{ij} 는 i 번째 학교, j 번째 학생의 점수이다.
- ▶ 학교마다 조사된 학생 수가 다르므로 표본크기 n_i 는 서로 다른 값을 갖는다.

Example for Hierarchical Model

- ▶ 이 조사에서 우리가 알고 싶은 것은 같은 학교 내, 그리고 다른 학교 간 수리영역 점수의 차이에 관한 것인데 이를 구체적으로 표현하면 다음과 같다.

Example for Hierarchical Model

- ▶ 이 조사에서 우리가 알고 싶은 것은 같은 학교 내, 그리고 다른 학교 간 수리영역 점수의 차이에 관한 것인데 이를 구체적으로 표현하면 다음과 같다.
 - 학교 내 또는 학교 간 수리영역 점수 차이가 존재하는가?

Example for Hierarchical Model

- ▶ 이 조사에서 우리가 알고 싶은 것은 같은 학교 내, 그리고 다른 학교 간 수리영역 점수의 차이에 관한 것인데 이를 구체적으로 표현하면 다음과 같다.
 - 학교 내 또는 학교 간 수리영역 점수 차이가 존재하는가?
 - 어떤 학교의 평균이 얼마나 전체 평균보다 높은가?

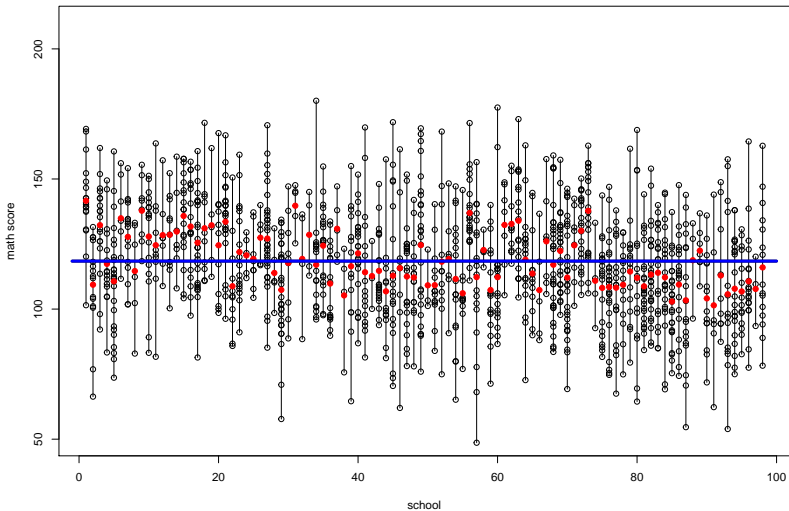
Example for Hierarchical Model

- ▶ 이 조사에서 우리가 알고 싶은 것은 같은 학교 내, 그리고 다른 학교 간 수리영역 점수의 차이에 관한 것인데 이를 구체적으로 표현하면 다음과 같다.
 - 학교 내 또는 학교 간 수리영역 점수 차이가 존재하는가?
 - 어떤 학교의 평균이 얼마나 전체 평균보다 높은가?
 - 어떤 학교의 평균이 얼마나 전체 평균보다 낮은가 ?

Example for Hierarchical Model

- ▶ 이 조사에서 우리가 알고 싶은 것은 같은 학교 내, 그리고 다른 학교 간 수리영역 점수의 차이에 관한 것인데 이를 구체적으로 표현하면 다음과 같다.
 - 학교 내 또는 학교 간 수리영역 점수 차이가 존재하는가?
 - 어떤 학교의 평균이 얼마나 전체 평균보다 높은가?
 - 어떤 학교의 평균이 얼마나 전체 평균보다 낮은가 ?
- ▶ 각 관측치를 하얀 점으로 표시하고, 소집단의 평균 \bar{x}_i 를 빨간 점으로 표시하였다. 중심의 평행선은 전체 평균 \bar{x} 를 나타낸다.

Diagnostic Plot



Diagnostic Plot

- ▶ $\theta_1, \dots, \theta_K$ 에 대한 정규성을 알아보기 위하여 \bar{x}_i 의 히스토그램을 그려본 결과 정규성 가정이 크게 벗어나지 않음을 알 수 있다.

Diagnostic Plot

- ▶ $\theta_1, \dots, \theta_K$ 에 대한 정규성을 알아보기 위하여 \bar{x}_i 의 히스토그램을 그려본 결과 정규성 가정이 크게 벗어나지 않음을 알 수 있다.
- ▶ X_{ij} 에 대한 정규분포에서 동일한 분산을 가정하였는데 이를 검증하기 위하여 표본표준편차 $s_i = \sqrt{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / (n_i - 1)}$ 을 그려본 결과 s_i 가 대략 10과 30 사이로 크게 변동이 없으므로 동일 분산 가정이 무리가 없어 보인다.

Diagnostic Plot

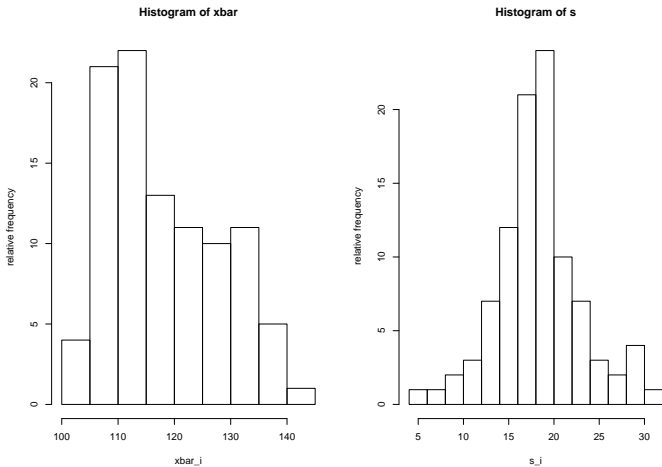


Figure: \bar{x}_i, s_i 의 분포

Diagnostic Plot: R-code

```
##### Hierarchical Model #####  
max.ni=31  
data=matrix(scan('http://home.ewha.ac.kr/~msoh/Bayesianbook/mathscore.txt'),  
ncol=max.ni+2,byrow=T)  
ni=data[,2]; K=length(ni)  
X=data[,3:(max.ni+2)]  
xbar = rep(0,K) ;s = rep(0,K) ; x.min= rep(0,K) ; x.max = rep(0,K)  
for (i in 1:K) x.min[i] = min(X[i,1:ni[i]])  
for (i in 1:K) x.max[i] = max(X[i,1:ni[i]])  
for (i in 1:K) xbar[i] = mean(X[i,1:ni[i]])  
for (i in 1:K) s[i] = sd(X[i,1:ni[i]])
```

Diagnostic Plot - R-code

```
par(mfrow=c(1,1))
plot(xbar, ylim=c(50,210), type="n", xlab="school" , ylab="math score" )
for(i in 1:K) for(j in 1:ni[i]) points(i,X [i,j] )
for(i in 1:K) lines(c(i,i), c(x.min [i],x.max[i]))
for(i in 1:K) points(i,xbar [i] , pch=19, col=2)
xtotal.mean =mean(xbar)
lines(c(-1:100), rep(xtotal.mean,102), lwd=4, col=4)

par(mfrow=c(1,2))
hist(xbar, xlab="xbar_i", ylab="relative frequency")
hist(s,nclass=10, xlab="s_i", ylab=" relative frequency")
```

Prior Settings: Unit Information Prior

- ▶ 계층적 베이지 분석을 적용하기 위하여 다음과 같은 사전분포의 모수를 선택하자.

$$a = 1/2,$$

$$b = a \cdot \text{Mean} \{s_1^2, \dots, s_K^2\},$$

$$c = 1/2,$$

$$d = c \cdot \text{Var} \{\bar{x}_1, \dots, \bar{x}_K\},$$

$$\mu_0 = \text{Mean} \{x_1, \dots, x_K\},$$

$$k_0 = 1.$$

Gibbs Sampling

```
### prior ###
a=0.5; sigsq0=mean(s**2); b=a*sigsq0; c=0.5
tausq0=var(xbar); d=c*tausq0; k0=1; mu0=mean(xbar)
##### Start Gibbs Sampling #####
Nwarm=1000 ; Nsim=5000; THETA=matrix(nrow=Nsim,ncol=K);
MTS=matrix(nrow=Nsim,ncol=3); theta=c(1:K)
# initial values
mu=mu0; tausq=tausq0; sigsq=sigsq0; theta=xbar
for(ns in 1:(Nwarm+Nsim)){          # simulation starts
  #generate theta
  for(i in 1:K){
    mi=(xbar[i]*ni[i]/sigsq + mu/tausq)/(1/tausq+ni[i]/sigsq)
    vari=1/(1/tausq+ni[i]/sigsq)
    theta[i]=rnorm(1,mi,sqrt(vari))
  }
}
```

Gibbs Sampling

```
#generate sigsq
alpha=0.5*sum(ni)+a; sum=0
for(i in 1:K) for (j in 1: ni[i]){
    sum=sum+(X[i,j] - theta[i])^2
}
beta=b+0.5*sum
sigsq=1/rgamma(1,alpha,beta)
#generate mu
mu=rnorm(1,(K*mean(theta)+k0*mu0)/(K+k0),sqrt(tausq/(K+k0)))
```

Gibbs Sampling

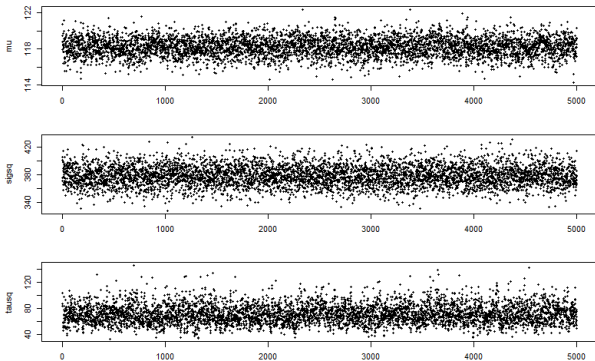
```
#generate tausq
alpha=c +0.5*K +0.5; sum=0
for(i in 1:K)sum=sum+(theta[i] - mu)^2
beta=d + 0.5*(k0*(mu-mu0)^2+sum)
tausq = 1/rgamma(1,alpha,beta)
#store
  if ( ns > Nwarm){
    THETA[ns-Nwarm,] =theta
    MTS[ns-Nwarm,]=c(mu,sigsq,tausq)
  }
} # simulation ends
```


Prior Settings

- ▶ 깃스 표본기법에서 생성된 표본을 가지고 사후추론을 할 수 있는데, 우선 깃스 기법으로부터 생성된 표본들의 수렴성과 상호연관성을 보기 위하여 1000번 이후의 표본을 가지고 표본의 경로그림과 자기상관그림을 그려보았다.

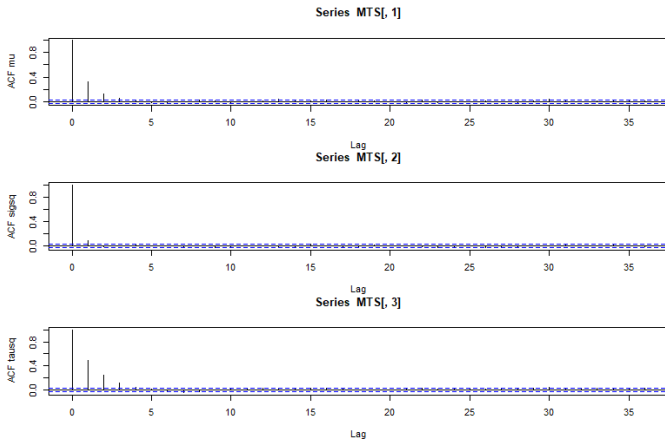
Trace Plot

```
#time sequence plot  
par(mfrow=c(3,1), mar = c(3,4,2,2))  
plot(MTS[, 1], ylab="mu", pch = 20)  
plot((MTS[, 2]), ylab="sigsq", pch = 20)  
plot((MTS[,3]), ylab="tausq", pch = 20)
```



Autocorrelation Plot

```
#Auto correlation graph  
par(mfrow=c(3,1), mar = c(4,4,3,3))  
acf(MTS[,1], ylab="ACF mu" )  
acf(MTS[,2] , ylab="ACF sigsq" )  
acf (MTS[, 3] , ylab=" ACF tausq" )
```



Distribution Check

```
#posterior density function of mu, sigma^2, tau^2
par(mfrow=c(1,3))
plot(density(MTS[,1]), xlab=expression(mu), main="" )
abline(v=quantile(MTS[,1], c(0.025,0.5,0.975)), lty=c(3,2,3))
plot(density(MTS[,2]), xlab=expression(sigma^2), main="")
abline(v=quantile(MTS[,2], c(0.025,0.5,0.975)), lty=c(3,2,3))
plot(density(MTS[,3]), xlab=expression(tau^2), main="")
abline(v=quantile(MTS[,3], c(0.025, 0.5, 0.975)), lty=c(3, 2, 3))
```

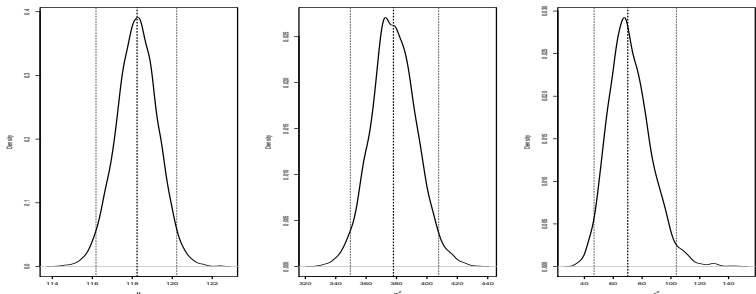


Figure: 무스의 부피 추정

Distribution Check

```
#inference
mu.hat=mean(MTS[,1 ])
sigsq.hat=mean(MTS[,2])
tausq.hat=mean(MTS [,3] )
theta.hat=apply(THETA,2,mean)

#plot of theta.hat_i
par(mfrow=c(1,1))
theta.grid=seq(mu.hat-7*sqrt(tausq.hat),mu.hat+7*sqrt(tausq.hat),length=100)
hist(theta.hat,prob=T,main="")
lines(theta.grid,dnorm(theta.grid,mu.hat,sqrt(tausq.hat)))
xtotal.mean=sum( ni *xbar) /sum( ni)
```

Distribution Check

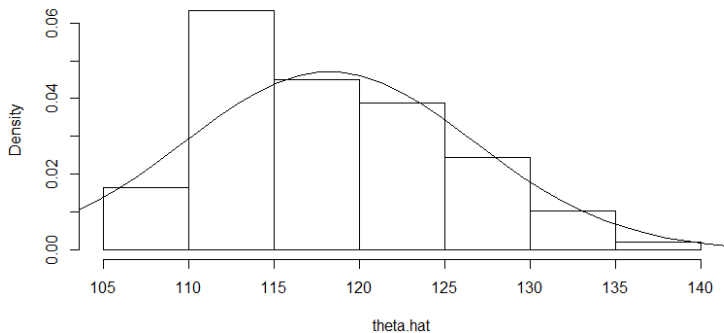


Figure: $\hat{\theta}_i$ 의 히스토그램과 $N(\hat{\mu}, \hat{\tau}^2)$

Hierarchical Poisson Model

- ▶ Count data are often modeled using a Poisson model.

Hierarchical Poisson Model

- ▶ Count data are often modeled using a Poisson model.
- ▶ If $y \sim \text{Poisson}(\mu)$ then $E(y) = \text{var}(y) = \mu$.

Hierarchical Poisson Model

- ▶ Count data are often modeled using a Poisson model.
- ▶ If $y \sim \text{Poisson}(\mu)$ then $E(y) = \text{var}(y) = \mu$.
- ▶ The hierarchical model is then

$$y_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i \sim \text{Gamma}(\alpha, \beta).$$

Priors for the Hierarchical Poisson Model

- Priors for the hyper-parameters are often taken to be Gamma (or exponential):

$$\alpha \sim \text{Gamma}(a, b),$$

$$\beta \sim \text{Gamma}(c, d).$$

,

Priors for the Hierarchical Poisson Model

- Priors for the hyper-parameters are often taken to be Gamma (or exponential):

$$\alpha \sim \text{Gamma}(a, b),$$

$$\beta \sim \text{Gamma}(c, d).$$

- The joint posterior distribution is

$$p(\mu, \alpha, \beta \mid y) \propto \prod_{i=1}^n \mu_i^{y_i} \exp\{-\mu_i\} \mu_i^{\alpha-1} \exp\{-\mu_i\beta\} \alpha^{a-1} \\ \times \exp\{-\alpha b\} \beta^{c-1} \exp\{-\beta d\}.$$

Full Conditional Distributions

- ▶ To carry out Gibbs sampling we need to find the full conditional distributions.

Full Conditional Distributions

- ▶ To carry out Gibbs sampling we need to find the full conditional distributions.
- ▶ Conditional for μ_i is

$$p(\mu_i | y) \propto \mu_i^{y_i + \alpha - 1} \exp\{-\mu_i(\beta + 1)\},$$

which is proportional to a Gamma($y_i + \alpha, \beta + 1$).

Full Conditional Distributions

- ▶ To carry out Gibbs sampling we need to find the full conditional distributions.
- ▶ Conditional for μ_i is

$$p(\mu_i | y) \propto \mu_i^{y_i + \alpha - 1} \exp\{-\mu_i(\beta + 1)\},$$

which is proportional to a Gamma($y_i + \alpha, \beta + 1$).

- ▶ The full conditional for α is

$$p(\alpha | y) \propto \prod_{i=1}^n \mu_i^{\alpha - 1} \alpha^{a - 1} \exp\{ab\}.$$

Full Conditional Distributions

- ▶ To carry out Gibbs sampling we need to find the full conditional distributions.
- ▶ Conditional for μ_i is

$$p(\mu_i | y) \propto \mu_i^{y_i + \alpha - 1} \exp\{-\mu_i(\beta + 1)\},$$

which is proportional to a Gamma($y_i + \alpha, \beta + 1$).

- ▶ The full conditional for α is

$$p(\alpha | y) \propto \prod_{i=1}^n \mu_i^{\alpha - 1} \alpha^{a - 1} \exp\{ab\}.$$

- ▶ The conditional for α does **NOT** have a standard form.

Hierarchical Poisson Model

- The full conditional for β is

$$\begin{aligned} p(\beta \mid y) &\propto \prod_{i=1}^n \exp\{-\beta\mu_i\} \beta^{c-1} \exp\{-\beta d\} \\ &\propto \beta^{c-1} \exp\left\{-\beta\left(\sum_i \mu_i + d\right)\right\}. \end{aligned}$$

Hierarchical Poisson Model

- ▶ The full conditional for β is

$$\begin{aligned} p(\beta \mid y) &\propto \prod_{i=1}^n \exp\{-\beta \mu_i\} \beta^{c-1} \exp\{-\beta d\} \\ &\propto \beta^{c-1} \exp\left\{-\beta\left(\sum_i^n \mu_i + d\right)\right\}. \end{aligned}$$

- ▶ which is proportional to a $\text{Gamma}(c, \sum_{i=1}^n \mu_i + d)$.

Gibbs Sampling + Metropolis

- ▶ Given α, β , draw each μ_i from the corresponding Gamma conditional.

Gibbs Sampling + Metropolis

- ▶ Given α, β , draw each μ_i from the corresponding Gamma conditional.
- ▶ Draw α using a Metropolis step or rejection sampling method.

Gibbs Sampling + Metropolis

- ▶ Given α, β , draw each μ_i from the corresponding Gamma conditional.
- ▶ Draw α using a Metropolis step or rejection sampling method.
- ▶ Draw β from the Gamma conditional.