

STAT 576 Bayesian Analysis

Lecture 5: Hierarchical Models

Chencheng Cai

Washington State University

Hierarchical Models

- ▶ Many statistical applications involve multiple parameters that are **related**.
- ▶ Example: (Multi-center study on the effectiveness of a drug)
 - ▶ Hospitals $j = 1, \dots, J$.
 - ▶ Patients in hospital j has a probability of recovering of θ_j .
 - ▶ The $\theta_1, \dots, \theta_J$ should be related.

Hierarchical Models

- ▶ Many statistical applications involve multiple parameters that are **related**.
- ▶ Example: (Multi-center study on the effectiveness of a drug)
 - ▶ Hospitals $j = 1, \dots, J$.
 - ▶ Patients in hospital j has a probability of recovering of θ_j .
 - ▶ The $\theta_1, \dots, \theta_J$ should be related.
- ▶ We use a prior distribution in which the θ_j 's are viewed as a sample from the **population distribution**.

Hierarchical Models

- ▶ Many statistical applications involve multiple parameters that are **related**.
- ▶ Example: (Multi-center study on the effectiveness of a drug)
 - ▶ Hospitals $j = 1, \dots, J$.
 - ▶ Patients in hospital j has a probability of recovering of θ_j .
 - ▶ The $\theta_1, \dots, \theta_J$ should be related.
- ▶ We use a prior distribution in which the θ_j 's are viewed as a sample from the **population distribution**.
- ▶ If we observe y_{ij} , $i = 1, \dots, n_j$ for hospital $j = 1, \dots, J$.
- ▶ we can use the data y_{ij} to estimate aspects of the population distribution of θ .

Hierarchical Models

- ▶ Many statistical applications involve multiple parameters that are **related**.
- ▶ Example: (Multi-center study on the effectiveness of a drug)
 - ▶ Hospitals $j = 1, \dots, J$.
 - ▶ Patients in hospital j has a probability of recovering of θ_j .
 - ▶ The $\theta_1, \dots, \theta_J$ should be related.
- ▶ We use a prior distribution in which the θ_j 's are viewed as a sample from the **population distribution**.
- ▶ If we observe y_{ij} , $i = 1, \dots, n_j$ for hospital $j = 1, \dots, J$.
- ▶ we can use the data y_{ij} to estimate aspects of the population distribution of θ .
- ▶ If furthermore, we approximation the population distribution by a parametric family, the corresponding parameters are called **hyperparameters**.

Example: Analysis using Historical Data

- ▶ Goal: estimate the probability of tumor in a population of female laboratory rats.

Example: Analysis using Historical Data

- ▶ Goal: estimate the probability of tumor in a population of female laboratory rats.
- ▶ Observation: $4/14$ rats show symptom of a tumor.

Example: Analysis using Historical Data

- ▶ Goal: estimate the probability of tumor in a population of female laboratory rats.
- ▶ Observation: 4/14 rats show symptom of a tumor.
- ▶ Model: We assume the observational model follows a binomial distribution:

$$y \sim \text{Binom}(14, \theta)$$

- ▶ with a conjugate prior for θ as $\text{Beta}(\alpha, \beta)$.

Example: Analysis using Historical Data

- ▶ Goal: estimate the probability of tumor in a population of female laboratory rats.
- ▶ Observation: 4/14 rats show symptom of a tumor.
- ▶ Model: We assume the observational model follows a binomial distribution:

$$y \sim \text{Binom}(14, \theta)$$

- ▶ with a conjugate prior for θ as $\text{Beta}(\alpha, \beta)$.
- ▶ The corresponding posterior is $\theta \mid y = 4 \sim \text{Beta}(\alpha + 4, \beta + 10)$.
- ▶ So far, the values for α and β are arbitrary.

Example: Analysis using Historical Data

- ▶ Goal: estimate the probability of tumor in a population of female laboratory rats.
- ▶ Observation: 4/14 rats show symptom of a tumor.
- ▶ Model: We assume the observational model follows a binomial distribution:

$$y \sim \text{Binom}(14, \theta)$$

- ▶ with a conjugate prior for θ as $\text{Beta}(\alpha, \beta)$.
- ▶ The corresponding posterior is $\theta \mid y = 4 \sim \text{Beta}(\alpha + 4, \beta + 10)$.
- ▶ So far, the values for α and β are arbitrary.
- ▶ If we have a historical records of previous experiments, we can have better choices for α and β if we interpret the prior distribution as the population distribution.

Example: Analysis using Historical Data

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

Example: Analysis using Historical Data

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

- For $j = 1, \dots, 70$ experiments, we observed y_j out of n_j rats with the symptom.

Example: Analysis using Historical Data

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

- ▶ For $j = 1, \dots, 70$ experiments, we observed y_j out of n_j rats with the symptom.
- ▶ The estimated mean and standard deviation for y_j/n_j are 0.136 and 0.103.

Example: Analysis using Historical Data

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

- ▶ For $j = 1, \dots, 70$ experiments, we observed y_j out of n_j rats with the symptom.
- ▶ The estimated mean and standard deviation for y_j/n_j are 0.136 and 0.103.
- ▶ We may choose the hyperparameters (α, β) by (**Variance is overestimated!!**)

$$\frac{\alpha}{\alpha + \beta} = 0.135, \quad \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = 0.103^2.$$

Example: Analysis using Historical Data

- ▶ The solution is $\alpha = 1.4$, $\beta = 8.6$.

Example: Analysis using Historical Data

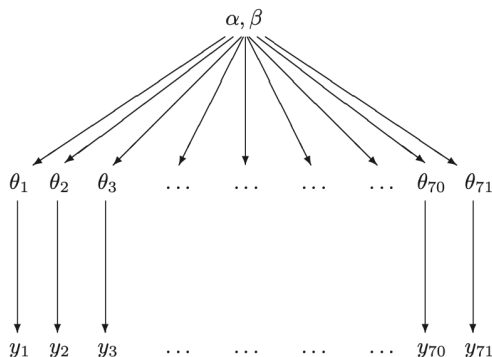
- ▶ The solution is $\alpha = 1.4$, $\beta = 8.6$.
- ▶ The prior is Beta(1.4, 8.6).
- ▶ The posterior is Beta(5.4, 18.6).

Example: Analysis using Historical Data

- ▶ The solution is $\alpha = 1.4$, $\beta = 8.6$.
- ▶ The prior is $\text{Beta}(1.4, 8.6)$.
- ▶ The posterior is $\text{Beta}(5.4, 18.6)$.
- ▶ The posterior mean is 0.223 and the posterior s.d. is 0.083.

Example: Analysis using Historical Data

- ▶ The solution is $\alpha = 1.4$, $\beta = 8.6$.
- ▶ The prior is $\text{Beta}(1.4, 8.6)$.
- ▶ The posterior is $\text{Beta}(5.4, 18.6)$.
- ▶ The posterior mean is 0.223 and the posterior s.d. is 0.083.



Note: the calculation demonstrated here is not a Bayesian calculation!

Example: Analysis using Historical Data

The demonstrated calculation is based on the assumption that the risk θ from the current observation is considered a **random** sample from a **common** distribution as previous 70 experiments.

Example: Analysis using Historical Data

The demonstrated calculation is based on the assumption that the risk θ from the current observation is considered a **random** sample from a **common** distribution as previous 70 experiments.

Violations of this assumption:

- ▶ The risk θ changes over time. (More common for econometric and financial data).

Example: Analysis using Historical Data

The demonstrated calculation is based on the assumption that the risk θ from the current observation is considered a **random** sample from a **common** distribution as previous 70 experiments.

Violations of this assumption:

- ▶ The risk θ changes over time. (More common for econometric and financial data).
- ▶ The historical data has other covariates.
 - ▶ Data collected in different hospitals/labs/centers.
 - ▶ Data collected for different sub-populations.

Example: Analysis using Historical Data

The demonstrated calculation is based on the assumption that the risk θ from the current observation is considered a **random** sample from a **common** distribution as previous 70 experiments.

Violations of this assumption:

- ▶ The risk θ changes over time. (More common for econometric and financial data).
- ▶ The historical data has other covariates.
 - ▶ Data collected in different hospitals/labs/centers.
 - ▶ Data collected for different sub-populations.

We may elaborate those factors into a more complicated model.

Example: Analysis using Historical Data

Question: Can we use the prior $\text{Beta}(1.4, 8.6)$ to do Bayesian inference for the 70 historical studies?

Example: Analysis using Historical Data

Question: Can we use the prior $\text{Beta}(1.4, 8.6)$ to do Bayesian inference for the 70 historical studies?

No!!

Example: Analysis using Historical Data

Question: Can we use the prior $\text{Beta}(1.4, 8.6)$ to do Bayesian inference for the 70 historical studies?

No!!

Remarks:

- ▶ The data is used twice: (1) estimating the prior (2) computing each's posterior.

Example: Analysis using Historical Data

Question: Can we use the prior $\text{Beta}(1.4, 8.6)$ to do Bayesian inference for the 70 historical studies?

No!!

Remarks:

- ▶ The data is used twice: (1) estimating the prior (2) computing each's posterior.
- ▶ We ignored the uncertainty in estimating α and β . (In oppose to Bayesian inference, where the posterior measures uncertainty.)

Example: Analysis using Historical Data

Question: Can we use the prior $\text{Beta}(1.4, 8.6)$ to do Bayesian inference for the 70 historical studies?

No!!

Remarks:

- ▶ The data is used twice: (1) estimating the prior (2) computing each's posterior.
- ▶ We ignored the uncertainty in estimating α and β . (In oppose to Bayesian inference, where the posterior measures uncertainty.)
- ▶ The prior distributions should be known **before** observing any data. Shall we really estimate them?

Example: Analysis using Historical Data

In addition:

- ▶ It definitely makes sense to estimate the population distribution from all the data than to estimate them separately.

Example: Analysis using Historical Data

In addition:

- ▶ It definitely makes sense to estimate the population distribution from all the data than to estimate them separately.
- ▶ The posterior for experiment j_1 and j_2 ($j_1 \neq j_2$) should be dependent because they are studying the same object.

Example: Analysis using Historical Data

In addition:

- ▶ It definitely makes sense to estimate the population distribution from all the data than to estimate them separately.
- ▶ The posterior for experiment j_1 and j_2 ($j_1 \neq j_2$) should be dependent because they are studying the same object.
- ▶ In order to retain the advantage of the hierarchical model and to get rid of the aforementioned trouble, we will build a full probability model for all parameters.

Example: Analysis using Historical Data

In addition:

- ▶ It definitely makes sense to estimate the population distribution from all the data than to estimate them separately.
- ▶ The posterior for experiment j_1 and j_2 ($j_1 \neq j_2$) should be dependent because they are studying the same object.
- ▶ In order to retain the advantage of the hierarchical model and to get rid of the aforementioned trouble, we will build a full probability model for all parameters.
- ▶ The analysis using the data to estimate the prior parameters, which is sometimes called **empirical Bayes**, can be viewed as an approximation to the complete hierarchical Bayesian analysis.

Hierarchical Modeling

- ▶ Assume we have J experiments.
- ▶ For each experiment j , we have observation y_j , parameter θ_j and likelihood $p(y_j \mid \theta_j)$.

Hierarchical Modeling

- ▶ Assume we have J experiments.
- ▶ For each experiment j , we have observation y_j , parameter θ_j and likelihood $p(y_j \mid \theta_j)$.
- ▶ If there is no additional information other than the observations y_j 's, we assume the **exchangeability** of the parameters, that is

$$p(\theta_1, \dots, \theta_J) \sim p(\theta_{\pi(1)}, \dots, \theta_{\pi(J)}),$$

for any permutation $\pi : \{1, \dots, J\} \rightarrow \{1, \dots, J\}$.

Hierarchical Modeling

- ▶ Assume we have J experiments.
- ▶ For each experiment j , we have observation y_j , parameter θ_j and likelihood $p(y_j \mid \theta_j)$.
- ▶ If there is no additional information other than the observations y_j 's, we assume the **exchangeability** of the parameters, that is

$$p(\theta_1, \dots, \theta_J) \sim p(\theta_{\pi(1)}, \dots, \theta_{\pi(J)}),$$

for any permutation $\pi : \{1, \dots, J\} \rightarrow \{1, \dots, J\}$.

- ▶ Furthermore, inspired by the De Finetti's Theorem, we can construct the prior on (θ_1, θ_J) in the following way:

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_J \mid \phi \sim p(\theta \mid \phi) \text{ i.i.d.}$$

Hierarchical Modeling

- ▶ Assume we have J experiments.
- ▶ For each experiment j , we have observation y_j , parameter θ_j and likelihood $p(y_j \mid \theta_j)$.
- ▶ If there is no additional information other than the observations y_j 's, we assume the **exchangeability** of the parameters, that is

$$p(\theta_1, \dots, \theta_J) \sim p(\theta_{\pi(1)}, \dots, \theta_{\pi(J)}),$$

for any permutation $\pi : \{1, \dots, J\} \rightarrow \{1, \dots, J\}$.

- ▶ Furthermore, inspired by the De Finetti's Theorem, we can construct the prior on (θ_1, θ_J) in the following way:

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_J \mid \phi \sim p(\theta \mid \phi) \text{ i.i.d.}$$

- ▶ Or in other words,

$$p(\theta_1, \dots, \theta_J) = \int \left(\prod_{j=1}^J p(\theta_j \mid \phi) \right) p(\phi) d\mu(\phi)$$

Hierarchical Modeling

Back to the rat tumor example.

- If without any additional information on the historical data, we assume

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_{70} \mid \phi \sim p(\theta \mid \phi) \text{ i.i.d.}$$

Hierarchical Modeling

Back to the rat tumor example.

- ▶ If without any additional information on the historical data, we assume

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_{70} \mid \phi \sim p(\theta \mid \phi) \text{ i.i.d.}$$

- ▶ If the experiments were conducted at 5 different centers, we assume (two-level hierarchical model)

$$\psi \sim p(\psi), \quad \phi_1, \dots, \phi_5 \mid \psi \sim p(\phi \mid \psi) \text{ i.i.d.}, \quad \theta_{1j}, \dots, \theta_{14j} \mid \phi_j \sim p(\theta \mid \phi_j) \text{ i.i.d.}$$

Hierarchical Modeling

Back to the rat tumor example.

- ▶ If without any additional information on the historical data, we assume

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_{70} \mid \phi \sim p(\theta \mid \phi) \text{ i.i.d.}$$

- ▶ If the experiments were conducted at 5 different centers, we assume (two-level hierarchical model)

$$\psi \sim p(\psi), \quad \phi_1, \dots, \phi_5 \mid \psi \sim p(\phi \mid \psi) \text{ i.i.d.}, \quad \theta_{1j}, \dots, \theta_{14j} \mid \phi_j \sim p(\theta \mid \phi_j) \text{ i.i.d.}$$

- ▶ If each experiment j is equipped with covariate x_j , we assume

$$\phi \sim p(\phi), \quad \theta_1, \dots, \theta_{70} \mid \phi, x_1, \dots, x_{70} \sim \prod_{j=1}^{70} p(\theta_j \mid \phi, x_j)$$

Hierarchical Modeling

- Now the complete model is

$$\phi \sim p(\phi)$$

$$\theta_j \sim p(\theta \mid \phi) \text{ i.i.d. for } j = 1, \dots, J$$

$$y_j \sim p(y \mid \theta_j) \text{ independent for } j = 1, \dots, J$$

Hierarchical Modeling

- Now the complete model is

$$\phi \sim p(\phi)$$

$$\theta_j \sim p(\theta \mid \phi) \text{ i.i.d. for } j = 1, \dots, J$$

$$y_j \sim p(y \mid \theta_j) \text{ independent for } j = 1, \dots, J$$

- The joint prior distribution:

$$p(\phi, \theta_1, \dots, \theta_J) = p(\phi) \prod_{j=1}^J p(\theta_j \mid \phi)$$

Hierarchical Modeling

- Now the complete model is

$$\phi \sim p(\phi)$$

$$\theta_j \sim p(\theta \mid \phi) \text{ i.i.d. for } j = 1, \dots, J$$

$$y_j \sim p(y \mid \theta_j) \text{ independent for } j = 1, \dots, J$$

- The joint prior distribution:

$$p(\phi, \theta_1, \dots, \theta_J) = p(\phi) \prod_{j=1}^J p(\theta_j \mid \phi)$$

- The observation model:

$$p(y_1, \dots, y_J \mid \phi, \theta_1, \dots, \theta_J) = \prod_{j=1}^J p(y_j \mid \theta_j)$$

Hierarchical Modeling

- Now the complete model is

$$\phi \sim p(\phi)$$

$$\theta_j \sim p(\theta \mid \phi) \text{ i.i.d. for } j = 1, \dots, J$$

$$y_j \sim p(y \mid \theta_j) \text{ independent for } j = 1, \dots, J$$

- The joint prior distribution:

$$p(\phi, \theta_1, \dots, \theta_J) = p(\phi) \prod_{j=1}^J p(\theta_j \mid \phi)$$

- The observation model:

$$p(y_1, \dots, y_J \mid \phi, \theta_1, \dots, \theta_J) = \prod_{j=1}^J p(y_j \mid \theta_j)$$

- The joint posterior distribution:

Hierarchical Modeling

- ▶ The distribution $p(\phi)$ is the “prior” distribution for the hyperparameter ϕ , which is called the **hyperprior** distribution.

Hierarchical Modeling

- ▶ The distribution $p(\phi)$ is the “prior” distribution for the hyperparameter ϕ , which is called the **hyperprior** distribution.
- ▶ Due to the complexity in $\theta_1, \dots, \theta_J$, it is often more convenient to look at the marginal posterior distribution for the hyperparameter.

Hierarchical Modeling

- ▶ The distribution $p(\phi)$ is the “prior” distribution for the hyperparameter ϕ , which is called the **hyperprior** distribution.
- ▶ Due to the complexity in $\theta_1, \dots, \theta_J$, it is often more convenient to look at the marginal posterior distribution for the hyperparameter.
- ▶ We often adopt a hybrid approach both analytically and numerically to conduct Bayesian inference.

Hierarchical Modeling

- ▶ The distribution $p(\phi)$ is the “prior” distribution for the hyperparameter ϕ , which is called the **hyperprior** distribution.
- ▶ Due to the complexity in $\theta_1, \dots, \theta_J$, it is often more convenient to look at the marginal posterior distribution for the hyperparameter.
- ▶ We often adopt a hybrid approach both analytically and numerically to conduct Bayesian inference.
- ▶ **Step 1 (analytic)**: get the marginal posterior distribution for ϕ .
- ▶ **Step 2 (numerical)**: draw samples of $(\phi, \theta_1, \dots, \theta_J)$ from the joint posterior distribution.

Hierarchical Modeling

Step 1 Procedure:

1. Get the posterior in proportional form:

$$p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_J) \propto p(\phi) \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

Hierarchical Modeling

Step 1 Procedure:

1. Get the posterior in proportional form:

$$p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_J) \propto p(\phi) \prod_{j=1}^J p(y_i \mid \theta_j) p(\theta_j \mid \phi)$$

2. Determine the conditional posterior distribution of $(\theta_1, \dots, \theta_J)$:

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) = A(\phi, y_1, \dots, y_J) p(\phi) \prod_{j=1}^J p(y_i \mid \theta_j) p(\theta_j \mid \phi)$$

for some normalizing coefficient A .

Hierarchical Modeling

Step 1 Procedure:

1. Get the posterior in proportional form:

$$p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_J) \propto p(\phi) \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

2. Determine the conditional posterior distribution of $(\theta_1, \dots, \theta_J)$:

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) = A(\phi, y_1, \dots, y_J) p(\phi) \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

for some normalizing coefficient A .

3. Determine the marginal posterior distribution of ϕ by

$$p(\phi \mid y_1, \dots, y_n) = \frac{p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_n)}{p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_n)} \propto [A(\phi, y_1, \dots, y_J)]^{-1}$$

Hierarchical Modeling

Step 1 Procedure:

1. Get the posterior in proportional form:

$$p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_J) \propto p(\phi) \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

2. Determine the conditional posterior distribution of $(\theta_1, \dots, \theta_J)$:

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) = A(\phi, y_1, \dots, y_J) p(\phi) \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

for some normalizing coefficient A .

3. Determine the marginal posterior distribution of ϕ by

$$p(\phi \mid y_1, \dots, y_n) = \frac{p(\phi, \theta_1, \dots, \theta_J \mid y_1, \dots, y_n)}{p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_n)} \propto [A(\phi, y_1, \dots, y_J)]^{-1}$$

Step 1 is analytical because $p(\theta_j \mid \phi)$ is chosen conjugate to $p(y_j \mid \theta_j)$.

Hierarchical Modeling

Step 2 Procedure:

1. Draw samples of ϕ from the marginal posterior distribution $p(\phi \mid y_1, \dots, y_J)$.

Hierarchical Modeling

Step 2 Procedure:

1. Draw samples of ϕ from the marginal posterior distribution $p(\phi \mid y_1, \dots, y_J)$.
2. Draw samples of $(\theta_1, \dots, \theta_J)$ from the conditional distribution $p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J)$. This step can be done coordinate-wise because

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) \propto \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

Hierarchical Modeling

Step 2 Procedure:

1. Draw samples of ϕ from the marginal posterior distribution $p(\phi \mid y_1, \dots, y_J)$.
2. Draw samples of $(\theta_1, \dots, \theta_J)$ from the conditional distribution $p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J)$. This step can be done coordinate-wise because

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) \propto \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

- With the samples from the joint posterior, we can estimate the posterior mean, median or other Bayesian estimators based on the empirical loss.

Hierarchical Modeling

Step 2 Procedure:

1. Draw samples of ϕ from the marginal posterior distribution $p(\phi \mid y_1, \dots, y_J)$.
2. Draw samples of $(\theta_1, \dots, \theta_J)$ from the conditional distribution $p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J)$. This step can be done coordinate-wise because

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) \propto \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

- ▶ With the samples from the joint posterior, we can estimate the posterior mean, median or other Bayesian estimators based on the empirical loss.
- ▶ To generate a prediction,
 - ▶ to predict a new observation for experiment j : draw new \tilde{y}_j given a sample of θ_j .

Hierarchical Modeling

Step 2 Procedure:

1. Draw samples of ϕ from the marginal posterior distribution $p(\phi \mid y_1, \dots, y_J)$.
2. Draw samples of $(\theta_1, \dots, \theta_J)$ from the conditional distribution $p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J)$. This step can be done coordinate-wise because

$$p(\theta_1, \dots, \theta_J \mid \phi, y_1, \dots, y_J) \propto \prod_{j=1}^J p(y_j \mid \theta_j) p(\theta_j \mid \phi)$$

- ▶ With the samples from the joint posterior, we can estimate the posterior mean, median or other Bayesian estimators based on the empirical loss.
- ▶ To generate a prediction,
 - ▶ to predict a new observation for experiment j : draw new \tilde{y}_j given a sample of θ_j .
 - ▶ to predict a new observation for a new experiment:
 - (1) draw a new $\tilde{\theta}$ given a sample of ϕ
 - (2) draw a new \tilde{y} given $\tilde{\theta}$.