Biostatistics 682: Applied Bayesian Inference Lecture 4: Multiparameter model

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Multiparameter models

- Almost every practical problem involves more than one unknown parameters
- In many problems, we are only interested in one or two parameters. Although to have a realistic probability model we may have more parameters than we are ultimately interested in.
  - These extra parameters are often called nuisance parameters which can cause difficulty in classical statistics.
  - They are easily handled within the Bayesian framework. How?

Marginal posterior probability

- Suppose we have a vector of parameters θ
- Divide θ into two subvectors: $\theta = (\theta_1, \theta_2)$
  - θ_2 is a vector of nuisance parameters and we are only interested in θ_1,
  - All of θ necessary for probabilistic modeling.
- In Bayesian inference, $\pi(y \mid \theta)$ where y is a vector of observations. The prior is $\pi(\theta)$. The posterior is

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\pi(\theta)$$

- To obtain $\pi(\theta_1 \mid y)$, the *marginal posterior* of θ_1, we integrate θ_2 out of the posterior distribution of θ

$$\begin{aligned}
\pi(\theta_1 \mid y) &= \int \pi(\theta_1, \theta_2 \mid y) d\theta_2 \\
&= \int \frac{\pi(y \mid \theta_1, \theta_2)\pi(\theta_1, \theta_2)}{\pi(y)} d\theta_2 \\
&\propto \int \pi(y \mid \theta_1, \theta_2)\pi(\theta_1, \theta_2) d\theta_2
\end{aligned}$$

Motivating Example: Blood Pressure Study

- Suppose we collect data on *average systolic BP (SBP)* in a population from a small sample.
- Data: (mmHg) for $n = 10$ patients:

$$y = (128,\ 132,\ 121,\ 135,\ 126,\ 130,\ 129,\ 138,\ 125,\ 131)$$

- Summary statistics:

$$\bar{y} = 129.5, \quad s^2 \approx 24.28, \quad S = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)s^2 = 218.5.$$

- Questions:
    - What is the posterior probability that the *average SBP in the population* exceeds 130 mmHg in this population?
    - What is the posterior predictive probability that a *new patient's SBP* exceeds 130 mmHg in this population?

Model and Prior Specifications

- Model (precision parameterization):

$$y_i \mid \mu, \tau^2 \overset{\text{iid}}{\sim} \mathrm{N}\big(\mu,\ \tau^{-2}\big)\,, \qquad \tau^2 = \sigma^{-2}, \qquad i = 1, \ldots, 10$$

- Interpretation:
  - μ: population mean SBP (scientific target).
  - τ^2: *precision* (inverse variance); larger τ^2 = less between-patient variability.
  - $\sigma^2 = 1/\tau^2$: variance (often a nuisance parameter).
- Let consider Jeffreys Prior

Jeffreys Prior

Likelihood function: $\ell(\mu,\tau^2) = \frac{n}{2}\log\tau^2 - \frac{\tau^2}{2}\sum_{i=1}^n (y_i-\mu)^2 + \text{const}.$

Scores and expected Hessian.

$$\frac{\partial \ell}{\partial \mu} = \tau^2 \sum_{i=1}^n (y_i-\mu), \quad \frac{\partial^2 \ell}{\partial \mu^2} = -n\tau^2 \;\Rightarrow\; I_{\mu\mu}(\mu,\tau^2) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \mu^2}\right] = n\tau^2.$$

$$\frac{\partial \ell}{\partial \tau^2} = \frac{n}{2\tau^2} - \frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2, \quad \frac{\partial^2 \ell}{\partial (\tau^2)^2} = -\frac{n}{2(\tau^2)^2} \;\Rightarrow\; I_{\tau^2\tau^2}(\mu,\tau^2) = \frac{n}{2(\tau^2)^2}.$$

$$\frac{\partial^2 \ell}{\partial \mu\,\partial \tau^2} = \sum_{i=1}^n (y_i-\mu) \;\Rightarrow\; I_{\mu,\tau^2}(\mu,\tau^2) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \mu\,\partial \tau^2}\right] = 0.$$

Thus,

$$I(\mu,\tau^2) = \begin{pmatrix} n\tau^2 & 0 \\ 0 & \dfrac{n}{2(\tau^2)^2} \end{pmatrix}.$$

Jeffreys prior. $\pi_J(\mu,\tau^2) \;\propto\; \sqrt{\det I(\mu,\tau^2)} \;\propto\; \sqrt{n\tau^2 \cdot \frac{n}{2(\tau^2)^2}} \;\propto\; \frac{1}{\sqrt{\tau^2}}.$

Deriving the joint posterior kernel (step by step)

Model $y_i \mid \mu, \tau^2 \overset{\text{iid}}{\sim} \mathrm{N}(\mu, (\tau^2)^{-1})$, Jeffreys prior: $\pi_J(\mu, \tau^2) \propto (\tau^2)^{-1/2}$.

1) Likelihood

$$\pi(y \mid \mu, \tau^2) \propto (\tau^2)^{n/2} \exp\left\{ -\frac{\tau^2}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right\}.$$

2) ANOVA decomposition. $\sum_{i=1}^{n} (y_i - \mu)^2 = \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{S} + n(\mu - \bar{y})^2.$

3) Multiply by Jeffreys prior.

$$\pi(\mu, \tau^2 \mid y) \;\propto\; \pi(y \mid \mu, \tau^2)\, \pi_J(\mu, \tau^2) \;\propto\; (\tau^2)^{n/2}\, e^{-\frac{\tau^2}{2}[S + n(\mu - \bar{y})^2]} \times (\tau^2)^{-1/2}.$$

4) Collect powers of τ^2.

$$\boxed{\pi(\mu, \tau^2 \mid y) \;\propto\; (\tau^2)^{\frac{n}{2} - \frac{1}{2}} \;\exp\left\{ -\frac{\tau^2}{2} \left[S + n(\mu - \bar{y})^2 \right] \right\}}$$

(All constants not depending on (μ, τ^2) absorbed into $\propto$.)

How to simplify the unnormalized joint distribution?

- Step 1: Start with the unnormalized kernel of the conditional distribution of the parameter(s) of interest given all other parameters. Simplify the expression by collecting terms that depend on the parameter(s) of interest.
- Step 2: Examine the kernel and try to *match it to a standard distribution family* (e.g. Normal, Gamma, Inverse-Gamma). This gives the conditional distribution (if recognizable).
- Step 3: Derive the *marginal distribution* for a parameter:
  - Either integrate out the nuisance parameter(s),
  - Or use known properties of conjugate families.

Example

Recall the joint posterior distribution:

$$\boxed{\pi(\mu, \tau^2 \mid y) \;\propto\; (\tau^2)^{\frac{n}{2}-\frac{1}{2}} \;\exp\left\{-\frac{\tau^2}{2}\left[S + n(\mu - \bar{y})^2\right]\right\}}$$

- Step 1: Find unnormalized kernel conditional distribution of μ given τ^2

$$\pi(\mu \mid \tau^2, y) \propto \exp\left\{-\frac{n\tau^2}{2}(\mu - \bar{y})^2\right\}$$

- Step 2: Identify the standard distribution for μ given τ^2:

$$\mu \mid \tau^2, y \sim N(\bar{y}, (n\tau^2)^{-1})$$

- Step 3: Figure out the marginal distribution of τ^2: Thus

$$\pi(\tau^2 \mid y) \propto (\tau^2)^{\frac{n}{2}-\frac{1}{2}} \;\exp\left\{-\frac{\tau^2 S}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{\frac{\tau^2 n}{2}(\mu - \bar{y})^2\right\} d\mu$$

$$\pi(\tau^2 \mid y) \propto (\tau^2)^{\frac{n}{2}-1} \;\exp\left\{-\frac{\tau^2 S}{2}\right\} \Rightarrow \tau^2 \mid y \sim \mathrm{G}\left(\frac{n}{2}, \frac{S}{2}\right)$$

Note that $\int_{-\infty}^{\infty} \exp(-(x-\mu)^2/\sigma^2)dx = \sqrt{2\pi\sigma^2}$.

Student t distribution

- Suppose $Z \sim \mathrm{N}(0, 1)$ and $V \sim \mathrm{G}(\nu/2, 1/2)$ (equivalently $V \sim \chi^2(\nu)$). Assume Z and V are independent. Define

$$T = \frac{Z}{\sqrt{V/\nu}}.$$

- Then T has probability density function

$$\pi_T(x) = \frac{\Gamma\{(\nu+1)/2\}}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

  This is the Student's t distribution with ν degrees of freedom.

- Let $X = m + sT$. Then X follows a Student's t distribution with location m, scale parameter s^2, and ν degrees of freedom:

$$X \sim t_\nu(m, s^2).$$

  What is the density of X: $\pi_X(x)$?

Student t distribution

- Suppose $Z \sim \mathrm{N}(0,1)$ and $V \sim \mathrm{G}(\nu/2, 1/2)$ (equivalently $V \sim \chi^2(\nu)$). Assume Z and V are independent. Define

$$T = \frac{Z}{\sqrt{V/\nu}}.$$

- Then T has probability density function

$$\pi_T(x) = \frac{\Gamma\{(\nu+1)/2\}}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

  This is the Student's t distribution with ν degrees of freedom.

- Let $X = m + sT$. Then X follows a Student's t distribution with location m, scale parameter s^2, and ν degrees of freedom:

$$X \sim t_\nu(m, s^2).$$

  What is the density of X: $\pi_X(x)$?

$$\pi_X(x) = \frac{1}{s}\pi_T\left(\frac{x-m}{s}\right) = \frac{\Gamma\{(\nu+1)/2\}}{\sqrt{\nu\pi s^2}\,\Gamma(\nu/2)} \left(1 + \frac{(x-m)^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}.$$

Scale mixture of Normal distributions

Setup:

$$X \mid \kappa \sim \mathrm{N}\left(m,\ \frac{s^2}{\kappa}\right), \qquad \kappa \sim \mathrm{G}\left(\tfrac{\nu}{2}, \tfrac{\nu}{2}\right).$$

Claim: Marginally, $X \sim t_\nu(m, s^2)$.

Derivation:

- Reparameterize: $\kappa^* = \nu\kappa \sim \mathrm{G}(\nu/2, 1/2)$.
- Let $Z \sim \mathrm{N}(0,1)$, independent of κ^*.
- Then we can write

$$X = m + \frac{sZ}{\sqrt{\kappa}} = m + \frac{sZ}{\sqrt{\kappa^*/\nu}}.$$

- But

$$\frac{Z}{\sqrt{\kappa^*/\nu}} \sim t_\nu.$$

Conclusion:

$$X \sim t_\nu(m, s^2).$$

Marginal posterior distribution of μ

- Recall: $\mu \mid \tau^2, y \sim N(\bar{y}, (n\tau^2)^{-1})$ and $\tau^2 \mid y \sim \mathrm{G}(n/2, S/2)$.
- Let $\kappa = S\tau^2/n$ and $\kappa \mid y \sim \mathrm{G}(n/2, n/2)$
- Also, $(n\tau^2)^{-1} = \frac{S}{n^2}\kappa^{-1}$. Thus, $\mu \mid \kappa, y \sim N\left(\bar{y}, \frac{S}{n^2}\kappa^{-1}\right)$
- By the scale mixture of Normal distributions, we have

$$\boxed{\mu \mid y \sim t_n(\bar{y}, S/n^2) = t_n\left(\bar{y}, (1 - \frac{1}{n})\frac{s^2}{n}\right)}$$

where $S = \sum_{i=1}^{n}(y_i - \bar{y})^2$ and $s^2 = (n-1)^{-1}S$

Alternative way to get marginal distribution of μ

Model: joint posterior kernel

$$\pi(\mu, \tau^2 \mid y) \;\propto\; (\tau^2)^{\frac{n}{2}-\frac{1}{2}} \exp\left\{-\frac{\tau^2}{2}\left[S + n(\mu-\bar{y})^2\right]\right\}.$$

Integrate out τ^2:

$$\pi(\mu \mid y) \;\propto\; \int_0^\infty (\tau^2)^{\frac{n}{2}-\frac{1}{2}} \exp\left\{-\underbrace{\tfrac{S+n(\mu-\bar{y})^2}{2}}_{\beta(\mu)}\tau^2\right\} d\tau^2.$$

Gamma–integral identity: $\displaystyle\int_0^\infty t^{\alpha-1}e^{-\beta t}dt = \Gamma(\alpha)\beta^{-\alpha}.$

Here $\alpha = \frac{n+1}{2}$, $\beta(\mu) = \frac{S+n(\mu-\bar{y})^2}{2}$, so

$$\pi(\mu \mid y) \;\propto\; \left(S + n(\mu-\bar{y})^2\right)^{-\frac{n+1}{2}} \propto \left[1 + \frac{n(\mu-\bar{y})^2}{S}\right]^{-\frac{n+1}{2}}.$$

The same conclusion: $\mu \mid y \sim t_n\left(\bar{y}, \frac{S}{n^2}\right) = t_n\left(\bar{y}, (1-\frac{1}{n})\frac{s^2}{n}\right)$

Posterior predictive distributions

Let

$$\tilde{y} \mid \mu, \tau^2 \sim \mathrm{N}(\mu, (\tau^2)^{-1}).$$

Recall the joint posterior distribution is given by

$$\mu \mid \tau^2, y \sim N(\bar{y}, (n\tau^2)^{-1}), \quad \tau^2 \mid y \sim \mathrm{G}(n/2, S/2)$$

We first integrate out μ given τ^2

$$\tilde{y} \mid \tau^2 \sim \mathrm{N}(\mu, (1 + n^{-1})(\tau^2)^{-1}).$$

Then integrate out τ^2, the posterior predictive distribution of $\tilde{y}$ is given by

$$\tilde{y} \mid y \sim t_n \left(\bar{y}, \left(1 + \frac{1}{n}\right) \frac{S}{n} \right) = t_n \left(\bar{y}, \left(1 - \frac{1}{n^2}\right) s^2 \right)$$

Answering two questions under Jeffreys prior (Fisher)

Setup (Jeffreys via Fisher):

$$\mu \mid y \sim t_n\Big(\bar{y}, \tfrac{S}{n^2}\Big),$$

$$y_{\text{new}} \mid y \sim t_n\Big(\bar{y}, \tfrac{S}{n}\,\big(1 + \tfrac{1}{n}\big)\Big).$$

Questions:

1. Posterior probability that the *average SBP* exceeds 130:

$$P(\mu > 130 \mid y) \;=\; 1 - F_{t_n}\left(\frac{130 - \bar{y}}{\sqrt{S}/n}\right).$$

2. Posterior predictive probability that a *new patient's SBP* exceeds 130:

$$P(y_{\text{new}} > 130 \mid y) \;=\; 1 - F_{t_n}\left(\frac{130 - \bar{y}}{\sqrt{\frac{S}{n}\big(1 + \frac{1}{n}\big)}}\right).$$

Numerical results for the BP data

Data: $n = 10,\ \bar{y} = 129.5,\ S = 218.5.$

For $\mu \mid y \sim t_{10}(129.5,\ (1.478)^2)$: scale $= \sqrt{S}/n = 1.478.$

$$z_\mu = \frac{130 - 129.5}{1.478} \approx 0.338 \quad \Rightarrow \quad \boxed{P(\mu > 130 \mid y) \approx 0.371}$$

For $y_{\text{new}} \mid y \sim t_{10}(129.5,\ (4.903)^2)$: scale $= \sqrt{(S/n)(1 + 1/n)} = 4.903.$

$$z_{\text{pred}} = \frac{130 - 129.5}{4.903} \approx 0.102 \quad \Rightarrow \quad \boxed{P(y_{\text{new}} > 130 \mid y) \approx 0.460}$$

Interpretation: The probability the *population mean* exceeds 130 is about 37%, while an *individual* exceeding 130 is about 46% due to extra predictive uncertainty.

Multinomial distribution

- The multinomial distribution generalizes the binomial.
- Let $y = (y_1, \ldots, y_k)^\top$ be counts in k categories with $n = \sum_{i=1}^k y_i$ and $\theta = (\theta_1, \ldots, \theta_k)^\top$, where $\theta_i > 0$ and $\sum_{i=1}^k \theta_i = 1$.

$$\pi(y \mid \theta) = \frac{n!}{\prod_{i=1}^k y_i!} \prod_{i=1}^k \theta_i^{y_i}.$$

- Conjugate prior: Dirichlet

$$\theta \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_k), \qquad \pi(\theta) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \theta_i^{\alpha_i - 1}.$$

- Posterior:

$$\boxed{\theta \mid y \;\sim\; \mathrm{Dirichlet}\big(\alpha_1 + y_1, \ldots, \alpha_k + y_k\big)}.$$

Why conjugate? Quick derivation & useful facts

Derivation (kernel):

$$\pi(\theta \mid y) \;\propto\; \pi(y \mid \theta)\,\pi(\theta) \;\propto\; \Big(\prod_{i=1}^{k} \theta_i^{y_i}\Big)\Big(\prod_{i=1}^{k} \theta_i^{\alpha_i - 1}\Big) \;=\; \prod_{i=1}^{k} \theta_i^{(\alpha_i + y_i) - 1},$$

which is the kernel of $\mathrm{Dirichlet}(\alpha_1 + y_1, \ldots, \alpha_k + y_k)$.

Posterior moments:

$$\mathbb{E}[\theta_i \mid y] = \frac{\alpha_i + y_i}{\sum_j (\alpha_j + y_j)}, \qquad \mathrm{Var}(\theta_i \mid y) = \frac{(\alpha_i + y_i)(\alpha_0' - \alpha_i - y_i)}{(\alpha_0')^2(\alpha_0' + 1)},$$

where $\alpha_0' = \sum_{j=1}^{k} (\alpha_j + y_j) = \alpha_0 + n$ with $\alpha_0 = \sum_{j=1}^{k} \alpha_j$.

Posterior predictive:

- *One future trial (categorical):*

  $P(\text{next in cat } i \mid y) = \mathbb{E}[\theta_i \mid y] = \dfrac{\alpha_i + y_i}{\sum_j (\alpha_j + y_j)}.$

- m *future trials (counts z):* Dirichlet–Multinomial

$$p(z \mid y) = \frac{m!}{\prod_i z_i!}\,\frac{\Gamma(\alpha_0')}{\Gamma(\alpha_0' + m)}\prod_{i=1}^{k} \frac{\Gamma(\alpha_i + y_i + z_i)}{\Gamma(\alpha_i + y_i)}.$$