Biostatistics 682: Applied Bayesian Inference
Lecture 6: 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

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

Normal model: unknown mean and variance

- Let $y_1, \ldots, y_n \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$.
- For this model we will consider two different prior settings:
  - Improper, noninformative prior
  - Conjugate prior

Improper prior

- Non-informative prior
$$\pi(\mu, \sigma^{-2}) \propto (\sigma^{-2})^{-1}.$$

- The joint posterior distribution,
$$\pi(\mu, \sigma^{-2} \mid y) \propto (\sigma^{-2})^{n/2-1} \exp\left(-\frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right),$$
where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ and $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

Student t distribution

- Suppose $Z \sim \mathrm{N}(0,1)$ and $V \sim \mathrm{G}(n/2, 1/2)$ or $\chi^2(n)$. And Z and V are independent. Let
$$T = \frac{Z}{\sqrt{V/n}}.$$
- Then T has probability density function
$$f(t) = \frac{\Gamma\{(n+1)/2\}}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$
This is the student's t distribution with n degrees of freedom.
- Let $X = \mu + \sigma T$. Then X follows a student's t distribution with mean μ, scale parameter σ^2 and n degrees of freedom, denoted
$$X \sim t_n(\mu, \sigma^2).$$
What is the density of X: $f(x)$?
$$f(x) = \frac{\Gamma\{(n+1)/2\}}{\sqrt{n\pi\sigma^2}\Gamma(n/2)}\left(1+\frac{(x-\mu)^2}{n\sigma^2}\right)^{-\frac{n+1}{2}}.$$

Scale mixture of normal distributions

Let

$$\begin{aligned}(\theta \mid \mu, \sigma^2, k) &\sim \mathrm{N}(\mu, \sigma^2/k), \\ k &\sim \mathrm{G}(\nu/2, \nu/2),\end{aligned}$$

Then

$$\theta \sim t_\nu(\mu, \sigma^2).$$

How to show this?

$$\begin{aligned}&\pi(\theta \mid \mu, \sigma^2) \\ &= \int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} k^{\nu/2-1} \exp\left[-(\nu/2)k\right] (2\pi\sigma^2/k)^{-1/2} \exp\left[-0.5k(\theta-\mu)^2/\sigma^2\right] \\ &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)\sqrt{2\pi\sigma^2}} \times \int_0^\infty k^{(v+1)/2-1} \exp\left[-0.5(\nu + (\theta-\mu)^2/\sigma^2)\right] dk \\ &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)\sqrt{2\pi\sigma^2}} \times \frac{\Gamma((\nu+1)/2)}{(\nu/2 + ((\theta-\mu)/\sigma)^2/2)^{(\nu+1)/2}}.\end{aligned}$$

The marginal distributions

$$
\begin{aligned}
\pi(\mu \mid y) &\propto \int_0^\infty (\sigma^{-2})^{n/2-1} \exp\left(-\frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\sigma^{-2} \\
&= \Gamma(n/2)\left(\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2}\right)^{-n/2} \\
&\propto \left[1 + \frac{1}{n-1}\left(\frac{\bar{y}-\mu}{s/\sqrt{n}}\right)\right]^{-n/2}.
\end{aligned}
$$

This implies that

$$\mu \mid y \sim t_{n-1}(\mu, s/\sqrt{n}).$$

And

$$
\begin{aligned}
\pi(\sigma^{-2} \mid y) &\propto \int_0^\infty (\sigma^{-2})^{n/2-1} \exp\left(-\frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\mu \\
&\propto (\sigma^{-2})^{(n-1)/2-1} \exp\left\{-\frac{\sigma^{-2}}{2}(n-1)s^2\right\}.
\end{aligned}
$$

Thus,

$$\sigma^{-2} \mid y \sim \mathrm{G}\{(n-1)/2, (n-1)s^2/2\}.$$

Posterior predictive distributions

Let

$$\tilde{y} \sim \mathrm{N}(\mu, \sigma^2).$$

The posterior predictive distribution of $\tilde{y}$ is given by

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

How to show this? Let $k = s^2/\sigma^2$.

$$\tilde{y} \mid k, y \sim \mathrm{N}\left\{\bar{y}, \left(1 + \frac{1}{n}\right) s^2/k\right\}.$$

$$k \mid y \sim \mathrm{G}\{(n-1)/2, (n-1)/2\}.$$

Then the results are followed by the property of the scale mixture of normal distributions (on page 7)

Conjugate prior

- To derive the conjugate priors for μ, σ^{-2}, we consider the functional form of the likelihood

$$\pi(y \mid \mu, \sigma^{-2}) \propto (\sigma^{-2})^{n/2} \exp\left(-\frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right)$$

how about

$$\pi(y \mid \sigma^{-2}) \propto (\sigma^{-2})^{n/2} \exp\left(-\frac{\sigma^{-2}}{2}[(n-1)s^2]\right)$$

- This implies that the conjugate priors should be

$$\sigma^{-2} \sim \mathrm{G}\left(\nu/2, \nu\tau^2/2\right).$$

and

$$\mu \mid \sigma^{-2} \sim \mathrm{N}(\theta, \sigma^2/k).$$

Joint posterior distribution

$$
\begin{aligned}
\pi(\mu, \sigma^{-2} \mid y) &\propto (\sigma^{-2})^{\nu/2-1} \exp\left(\frac{\nu\tau^2}{2}\sigma^{-2}\right)(\sigma^{-2})^{1/2} \exp\left\{-\frac{k\sigma^{-2}}{2}(\mu-\theta)^2\right\} \\
&\qquad \times (\sigma^{-2})^{n/2} \exp\left(-\frac{\sigma^{-2}}{2}\sum_i (y_i-\mu)^2\right) \\
&= (\sigma^{-2})^{\frac{v+n+1}{2}-1} \exp\left(-\frac{\sigma^{-2}}{2}[\nu\tau^2 + k(\mu-\theta)^2]\right) \\
&\qquad \times \exp\left\{-\frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right\}.
\end{aligned}
$$

Marginal posterior of σ^{-2}

$$
\begin{aligned}
\pi(\sigma^{-2} \mid y) &= \int_{-\infty}^{\infty} \pi(\mu, \sigma^{-2} \mid y) d\mu \\
&\propto \underbrace{(\sigma^{-2})^{\frac{\nu+n+1}{2}-1} \exp\left\{-\frac{\sigma^{-2}}{2}[\nu\tau^2 + (n-1)s^2]\right\}}_{A} \\
&\quad \times \int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma^{-2}}{2}[k(\mu-\theta)^2 + n(\bar{y}-\mu)^2]\right\} d\mu \\
&\propto A \exp\left[-\frac{\sigma^2}{2}(k\theta^2 + n\bar{y}^2)\right] \int_{-\infty}^{\infty} \exp\left\{-\frac{(k+n)\sigma^{-2}}{2}\left[\mu^2 - 2\left(\frac{k\theta+n\bar{y}}{k+n}\right)\mu\right]\right\} \\
&= A \exp\left[-\frac{\sigma^2}{2}(k\theta^2 + n\bar{y}^2)\right] \exp\left[\frac{\sigma^{-2}}{2}(k+n)\left(\frac{k\theta+n\bar{y}}{k+n}\right)^2\right] \\
&\quad \times \int_{-\infty}^{\infty} \exp\left[-\frac{(k+n)\sigma^{-2}}{2}\left(\mu - \frac{k\theta+n\bar{y}}{k+n}\right)^2\right] d\mu \\
&= A \exp\left\{-\frac{\sigma^2}{2}\left[k\theta^2 + n\bar{y}^2 - \frac{(k\theta+n\bar{y})^2}{k+n}\right]\right\} (2\pi)^{1/2}[(k+n)\sigma^{-2}]^{-1/2} \\
&\propto (\sigma^{-2})^{\frac{\nu+n}{2}-1} \exp\left\{-\frac{\sigma^2}{2}\left[\nu\tau^2 + (n-1)s^2 + \frac{kn}{k+n}(\bar{y}-\theta)^2\right]\right\}.
\end{aligned}
$$

Marginal posterior of σ^{-2}

This is a kernel of a gamma distribution. Thus,

$$[\sigma^{-2} \mid y] \sim G\left(\frac{\nu + n}{2}, \frac{\nu\tau^2 + (n-1)s^2 + \frac{kn}{k+n}(\bar{y} - \theta)^2}{2}\right).$$

Marginal posterior of μ

The full conditional of μ is given by

$$\begin{aligned}
\pi(\mu \mid \sigma^2, y) &\propto \exp\left[-\frac{k\sigma^{-2}}{2}(\mu-\theta)^2 - \frac{\sigma^{-2}}{2}\sum_i (y_i-\mu)^2\right] \\
&\propto \exp\left[-\frac{k\sigma^{-2}}{2}(\mu^2-2\theta\mu) - \frac{\sigma^{-2}}{2}(-2n\bar{y}\mu + n\mu^2)\right] \\
&= \exp\left\{-\frac{\sigma^{-2}}{2}(k+n)\left[\mu^2 - 2\left(\frac{k\theta+n\bar{y}}{k+n}\right)\mu\right]\right\}.
\end{aligned}$$

This implies that

$$[\mu \mid \sigma^{-2}, y] \sim N\left(\frac{k\theta+n\bar{y}}{k+n}, \frac{\sigma^2}{k+n}\right).$$

The marginal posterior distribution of μ is

$$[\mu \mid y] \sim t_{\nu+k}\left(\frac{k\theta+n\bar{y}}{k+n}, \frac{\nu\tau^2 + (n-1)s^2 + \frac{kn}{k+n}(\bar{y}-\theta)^2}{(\nu+n)(k+n)}\right).$$

Multinomial distribution

- The multinomial distribution is a generalization of the binomial distribution
- Let y denote the k vector of counts of observations in k categories with y_i the number of counts in category i and let $n = \sum_i y_i$ and $\theta = (\theta_1, \dots, \theta_k)^{\mathrm{T}}$ with $\theta_i > 0$ and $\sum_i \theta_i = 1$. The density of y given θ is

$$\pi(y \mid \theta) = \frac{n!}{\prod_i y_i!} \theta_1^{y_1} \cdots \theta_n^{y_n}.$$

- Recall for binomial data, the conjugate prior was the beta distribution. The multivariate generalization of the beta is the Dirichlet distribution. Let $\theta_i > 0$ for all i and $\sum_i \theta_i = 1$. Then

$$\theta \sim \mathrm{Dirichlet}(\alpha_1, \dots, \alpha_k)$$

If

$$\pi(\theta) = \frac{\Gamma(\sum \alpha_i)}{\prod_i \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

- What is the posterior distribution?

$$\theta \mid y \sim \mathrm{Dirichlet}(\alpha_1 + y_1, \dots, \alpha_k + y_k).$$