Bayesian Statistics

Multi-parameter models

Nan Lin
Department of Mathematics
Washington University in St. Louis

Motivation

- Most realistic problems require models with multiple parameters
- Frequentist approaches
  - Joint maximum likelihood estimation, which can be difficult when there are many parameters
  - Iterative algorithms (augmentation): partition parameters into distinctive subsets and then estimate parameters in one subset given the rest in an iterative manner
- Bayesian approach
  - Marginal posterior distribution of parameters of interest
  - Parameters that are not of interest are called *nuisance parameters*

Finding the marginal posterior distribution

- Model parameter $\theta = (\theta_1, \theta_2)$
  - θ_1: parameter of interest
  - θ_2: nuisance parameter
- Goal: finding marginal posterior distribution $f(\theta_1|y)$
  - Joint posterior density

$$f(\theta_1, \theta_2|y) \propto f(y|\theta_1, \theta_2)\pi(\theta_1, \theta_2)$$

  - Integrate over the nuisance parameter θ_2

$$f(\theta_1|y) = \int f(\theta_1, \theta_2|y)d\theta_2$$

  - Alternatively,

$$f(\theta_1|y) = \int f(\theta_1|\theta_2, y)f(\theta_2|y)d\theta_2$$

    - A mixture distribution mixing over θ_2
    - Weighted average of the conditional distribution of θ_1 evaluated at different θ_2

Finding the marginal posterior distribution

- Solving the integral can be computationally challenging
- Simulation approach
  - Simulate from the marginal posterior distribution of $\theta_2|y$
    - Draw $\theta_2^{(k)}$ from $f(\theta_2|y)$ for $k = 1,2,\ldots$
  - Simulate from the conditional posterior distribution of $\theta_1|\theta_2, y$
    - For each $\theta_2^{(k)}$, draw $\theta_1^{(k)}$ from $f(\theta_1|\theta_2^{(k)}, y)$
  - Requirement
    - Both $f(\theta_2|y)$ and $f(\theta_1|\theta_2, y)$ are some standard distributions that can be easily sampled from
  - In general, we need more sophisticated simulation methods
    - Markov Chain Monte Carlo (MCMC)

$N(\mu, \sigma^2)$ with noninformative prior

- Suppose that $y_i | \mu, \sigma^2 \sim N(\mu, \sigma^2)$ i.i.d.
  - μ, σ^2 are both unknown
- Use a noinformative prior for (μ, σ^2) and assume prior independence

$$f(\mu, \sigma^2) \propto 1 \times \frac{1}{\sigma^2} = \frac{1}{\sigma^2}$$

- Joint posterior density

$$f(\mu, \sigma^2 | y) \propto f(y | \mu, \sigma^2) f(\mu, \sigma^2)$$

$$\propto \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2)$$

$N(\mu, \sigma^2)$ with noninformative prior

- Let $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Then since

$$\sum_{i=1}^{n} (y_i - \mu)^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2,$$

- We have

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

- Here, $(\bar{y}, s^2)$ are the sufficient statistics of (μ, σ^2)

Conditional posterior: $\mu|\sigma^2, y$

- Given σ^2, this is essentially just a one-parameter problem. And we knew

$$\mu|\sigma^2, y \sim N(\bar{y}, \frac{\sigma^2}{n})$$

- Or you can quickly recognize it from the fact

$$f(\mu|\sigma^2, y) \propto f(\mu, \sigma^2|y)$$

Marginal posterior: $\sigma^2|y$

- Integrate $f(\mu, \sigma^2|y)$ over μ

$$
\begin{aligned}
f(\sigma^2|y) &= \int f(\mu, \sigma^2|y) d\mu \\
&= \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\mu \\
&= \sigma^{-n-2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \int \exp\left(-\frac{n(\bar{y}-\mu)^2}{2\sigma^2}\right) d\mu \\
&= \sigma^{-n-2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \sqrt{\frac{2\pi\sigma^2}{n}} \\
&\propto (\sigma^2)^{-\frac{(n+1)}{2}} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)
\end{aligned}
$$

- This is a *scaled-inverse chi-square* distribution with $df = (n-1)$ and scale s^2
  - i.e. $\frac{(n-1)s^2}{\sigma^2}|y \sim \chi^2_{n-1}$

Marginal posterior: $\mu|y$

- $f(\mu|y) = \int f(\mu,\sigma^2|y)d\sigma^2 \propto$
  $\int_0^\infty \left(\frac{1}{2\sigma^2}\right)^{\frac{n}{2}+1} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\sigma^2$
- Let $z = \frac{A}{2\sigma^2}$, where $A = (n-1)s^2 + n(\bar{y}-\mu)^2$
- Then

$$f(\mu|y) \propto \int_0^\infty \left(\frac{z}{A}\right)^{\frac{n}{2}+1} \frac{A}{z^2} e^{-z} dz \propto A^{-\frac{n}{2}} \int_0^\infty z^{\frac{n}{2}-1} e^{-z} dz \propto A^{-\frac{n}{2}}$$

$$\propto \left[1 + \frac{n(\bar{y}-\mu)^2}{(n-1)s^2}\right]^{-\frac{n}{2}}$$

- That is,

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} |y \sim t_{n-1}$$

Predictive distribution

- $f(\tilde{y}|y) = \int \int f(\tilde{y}|\sigma^2, \mu) f(\mu, \sigma^2|y) d\mu d\sigma^2$
  - This is a mixture distribution
- Simulate from the predictive distribution
  - Draw σ^2 from $\sigma^2|y$,
  - Draw μ from $\mu|\sigma^2, y$
  - Draw $\tilde{y}$ from $\tilde{y}|\mu, \sigma^2$

$N(\mu, \sigma^2)$ with conjugate prior

- Conjugate prior
  - $\mu|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\kappa_0})$
  - $\sigma^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2)$
- Posterior distribution
  - $\mu|\sigma^2, y \sim N(\mu_n, \frac{\sigma_n^2}{\kappa_n})$
  - $\sigma^2|y \sim Inv - \chi^2(\nu_n, \sigma_n^2)$
    - $\mu_n = \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{y}$
    - $\kappa_n = \kappa_0 + n$
    - $\nu_n = \nu_0 + n$
    - $\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y} - \mu)^2$

Marginal posterior: $\mu|y$

- $\frac{\mu-\mu_n}{\sigma_n/\sqrt{\kappa_n}}|y \sim t_{\nu_n}$

- Simulating from the joint posterior
  - Sample σ^2 from Inv-χ^2 distribution of $\sigma^2|y$
  - For the given σ^2, sample from the normal distribution of $\mu|\sigma^2, y$

What if we use an independent prior?

- In the previous setup, μ and σ^2 are not independent a prior
- If we assume independence,
  - $\mu \sim N(\mu_0, \tau_0^2)$
  - $\sigma^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2)$
  - This leads to a very complicated form for the marginal posterior of $\sigma^2|y$
- This is not conjugate.