Bayesian Statistics
Note 2-1
Multiparameter Problems

Keunbaik Lee

Sungkyunkwan University

Normal distribution with unknown mean and variance I

A typical problem in statistics involves more than one unknown or unobservable quantity. Although a problem can include several parameters of interest, conclusions will often be drawn about one or only a few parameters at a time. In this case, the ultimate aim of Bayesian analysis is to obtain the marginal posterior of the parameter(s) of interest:

- find the joint posterior of all the parameters.
- integrate out the parameters that are not of direct interest.

Normal distribution with unknown mean and variance II

The parameters, not of direct interest, are usually referred to as **nuisance parameters**.
Suppose that $\theta = (\theta_1, \theta_2)$, where θ_1 (real or vector valued) is the parameter of interest and θ_2 (real or vector valued) is the nuisance parameter.
The posterior is

$$P(\theta_1, \theta_2 | y) \propto P(y | \theta_1, \theta_2) P(\theta_1, \theta_2).$$

Normal distribution with unknown mean and variance III

Integrating w.r.t. θ_2,

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

Normal distribution with unknown mean and variance IV

Note that the formula is very similar to that for the posterior predictive pdf

$$P(\tilde{y}|y) = \int P(\tilde{y}|\theta, y)P(\theta|y)d\theta.$$

Normal distribution with unknown mean and variance V

Example:

$$y_1, \cdots, y_n | \mu, \sigma^2 \sim^{iid} N(\mu, \sigma^2)$$

where both μ and $\sigma (> 0)$ are unknown.

- Jeffreys' prior for a multiparameter problem is the positive square root of the determinant of the Fisher Information matrix. We find Jeffreys' prior for the normal problem.

$$P(y_1, \cdots, y_n | \mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_1^n (y_i - \mu)^2}$$

Normal distribution with unknown mean and variance VI

Then

$$\frac{\partial \log P}{\partial \mu} = \frac{1}{\sigma^2}\sum_1^n (y_i - \mu), \quad \frac{\partial \log P}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_1^n (y_i - \mu)^2,$$

$$\frac{\partial^2 \log P}{\partial \mu^2} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 \log P}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_1^n (y_i - \mu)^2,$$

$$\frac{\partial^2 \log P}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}\sum_1^n (y_i - \mu).$$

Normal distribution with unknown mean and variance VII

Then the Fisher Information matrix is

$$
\begin{aligned}
I(\mu,\sigma) &= \begin{pmatrix} E\left(-\frac{\partial^2 \log P}{\partial \mu^2}\right) & E\left(-\frac{\partial^2 \log P}{\partial \mu \partial \sigma}\right) \\ E\left(-\frac{\partial^2 \log P}{\partial \mu \partial \sigma}\right) & E\left(-\frac{\partial^2 \log P}{\partial \sigma^2}\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix},
\end{aligned}
$$

$$=> |I(\mu,\sigma)| = \frac{2n^2}{\sigma^4}.$$

The Jeffreys' prior is

$$P(\mu,\sigma) \propto |I(\mu,\sigma)|^{1/2} \propto \sigma^{-2}.$$

Normal distribution with unknown mean and variance VIII

Instead Jeffreys himself recommended using the prior

$$P(\mu, \sigma) \propto \sigma^{-1}.$$

Some people refer to the latter as Jeffreys' independence prior. This is because for the location problem.

$$P(\mu) \propto 1.$$

For the scale problem, $P(\sigma) \propto \frac{1}{\sigma}$. Take the product of the two to get a prior for (μ, σ). Lindley (1958), Bernardo (1979) and others provided further justification for $P(\mu, \sigma) \propto \frac{1}{\sigma}$. Let us try $P(\mu, \sigma) \propto \frac{1}{\sigma}$.

Normal distribution with unknown mean and variance IX

- Then the joint posterior is

$$
\begin{aligned}
& P(\mu, \sigma | y_1, \cdots, y_n) \\
\propto \quad & \sigma^{-n-1} e^{-\frac{1}{2\sigma^2} \sum_1^n (y_i - \mu)^2} \\
= \quad & \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2} \left\{ (n-1)S^2 + n(\bar{y} - \mu)^2 \right\} \right]
\end{aligned}
$$

where $S^2 = \frac{1}{n-1} \sum_i (y_i - \bar{y})^2$.

Normal distribution with unknown mean and variance X

- The marginal posterior of μ is

$$
\begin{aligned}
P(\mu|y_1,\cdots,y_n) &\propto \int_0^\infty \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)S^2 + n(\bar{y}-\mu)^2\right\}\right] d\sigma \\
&= \int_0^\infty z^{\frac{n+1}{2}} \exp\left[-\frac{z}{2}\left\{(n-1)S^2 + n(\bar{y}-\mu)^2\right\}\right] \frac{1}{2} z^{-3/2} dz \\
&\qquad (z = 1/\sigma^2; \quad d\sigma = \frac{1}{2} z^{-3/2} dz) \\
&= \int_0^\infty \frac{1}{2} z^{n/2-1} \exp\left[-\frac{n(\bar{y}-\mu)^2 + (n-1)S^2}{2} z\right] dz \\
&= \frac{1}{2} \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{n(\bar{y}-\mu)^2+(n-1)S^2}{2}\right)^{n/2}} \\
&\propto \left(n(\bar{y}-\mu)^2 + (n-1)S^2\right)^{-\frac{n}{2}} \\
&\propto \left(1 + \frac{n(\mu-\bar{y})^2}{(n-1)S^2}\right)^{-\frac{n}{2}}
\end{aligned}
$$

Normal distribution with unknown mean and variance XI

which is Student's t distribution with location component $\bar{y}$, scale component $\frac{S}{\sqrt{n}}$, and $df = n - 1$.

Normal distribution with unknown mean and variance XII

- The marginal posterior of σ (integrating w.r.t. μ) is

$$\begin{aligned}
P(\sigma|y) &= \int P(\mu,\sigma|y)d\mu \\
&\propto \int_{-\infty}^{\infty} \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)S^2 + n(\bar{y}-\mu)^2\right\}\right] d\mu \\
&= \sigma^{-(n+1)} \exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{y})^2\right) d\mu \\
&\qquad \text{Normal}(\bar{y}, \frac{\sigma^2}{n}) \\
&= \sigma^{-(n+1)} \exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right) \sqrt{2\pi}\sqrt{\frac{\sigma^2}{n}} \\
&\propto \sigma^{-n} \exp\left(-\frac{(n-1)S^2}{2\sigma^2}\right).
\end{aligned}$$

Normal distribution with unknown mean and variance XIII

Hence, writing $z = \sigma^{-2}$,

$$P(z|y) \propto z^{\frac{n-1}{2}-1} \exp\left(-\frac{(n-1)S^2}{2} z\right)$$

which is Gamma with shape parameter $\frac{n-1}{2}$ and failure rate $\frac{1}{2}\sum_i (y_i - \bar{y})^2$.

Normal distribution with unknown mean and variance XIV

- Now $\mu|\sigma, y \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$.
  Given μ and y, σ has conditional pdf is given by

$$
\begin{aligned}
P(\sigma|\mu, y) &= \frac{P(\sigma|y)P(\mu|\sigma, y)}{P(\mu|y)} \\
&\propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - \bar{y})^2\right) \sigma^{-1} \exp\left(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2\right) \\
&= \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_i (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right\}\right] \\
&= \sigma^{-(n+1)} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2\right).
\end{aligned}
$$

Normal distribution with unknown mean and variance XV

Hence, writing $z = \sigma^{-2}$,

$$P(z|\mu, y) \propto z^{n/2-1} \exp\left(-\frac{\sum_i (y_i - \mu)^2}{2} z\right)$$

which is Gamma with shape parameter $n/2$ and failure rate $\frac{1}{2}\sum_i (y_i - \mu)^2$.

=> Random numbers $(\mu, 1/\sigma)$ are generated using Gibbs sampling:

$$\begin{aligned} \mu &\sim N\left(\bar{y}, \frac{S}{\sqrt{n}}\right), \\ 1/\sigma &\sim Gamma\left(\frac{n}{2}, \frac{1}{2}\sum_i (y_i - \mu)^2\right). \end{aligned}$$

Normal distribution with unknown mean and variance XVI

- The posterior predictive distribution for a future observation $\tilde{y}$ is

$$
\begin{aligned}
& P(\tilde{y}|y) \\
= & \int P(\tilde{y}|\theta, y)P(\theta|y)d\theta \\
= & \int_0^\infty \int_{-\infty}^\infty P(\tilde{y}|\mu,\sigma,y)P(\mu,\sigma|y)d\mu d\sigma \\
\propto & \int_0^\infty \int_{-\infty}^\infty \sigma^{-1}\exp\left[-\frac{1}{2\sigma^2}(\tilde{y}-\mu)^2\right] \\
& \quad \sigma^{-(n+1)}\exp\left[-\frac{1}{2\sigma^2}\left\{n(\mu-\bar{y})^2+(n-1)S^2\right\}\right]d\mu d\sigma \\
= & \int_0^\infty \int_{-\infty}^\infty \sigma^{-(n+2)}\exp\left[-\frac{1}{2\sigma^2}\left\{(\tilde{y}-\mu)^2+n(\mu-\bar{y})^2+(n-1)S^2\right\}\right]d\mu d\sigma.
\end{aligned}
$$

Normal distribution with unknown mean and variance XVII

Note that

$$
\begin{aligned}
(\tilde{y}-\mu)^2 + n(\mu-\bar{y})^2 &= (n+1)\mu^2 - 2(\tilde{y}+n\bar{y})\mu + \tilde{y}^2 + n\bar{y}^2 \\
&= (n+1)\left(\mu^2 - 2\frac{\tilde{y}+n\bar{y}}{n+1}\mu\right) + \tilde{y}^2 + n\bar{y}^2 \\
&= (n+1)\left(\mu - \frac{\tilde{y}+n\bar{y}}{n+1}\right)^2 + \frac{n}{n+1}(\tilde{y}-\bar{y})^2
\end{aligned}
$$

Normal distribution with unknown mean and variance XVIII

Thus,

$$
\begin{aligned}
P(\tilde{y}|y) &= \int_0^\infty \sigma^{-(n+2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\frac{n}{n+1}(\hat{y}-\bar{y})^2+(n-1)S^2\right\}\right] \\
&\quad \int_{-\infty}^\infty \exp\left[-\frac{n+1}{2\sigma^2}\left(\mu-\frac{\tilde{y}+n\bar{y}}{n+1}\right)^2\right] d\mu d\sigma \\
&\propto \int_0^\infty \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{\frac{n}{n+1}(\tilde{y}-\bar{y})^2+(n-1)S^2\right\}\right] d\sigma \\
&= \int_0^\infty z^{\frac{n+1}{2}} \exp\left[-\frac{z}{2}\left\{\frac{n}{n+1}(\tilde{y}-\bar{y})^2+(n-1)S^2\right\}\right] \frac{1}{2}z^{-\frac{3}{2}}dz \quad (\sigma^2 = z^{-1}) \\
&= \int_0^\infty \frac{1}{2}z^{n/2-1} \exp\left[-\frac{\frac{n}{n+1}(\tilde{y}-\bar{y})^2+(n-1)S^2}{2}z\right] dz \\
&\qquad \left(\text{Gamma}\left(\frac{n}{2}, \frac{\frac{n}{n+1}(\hat{y}-\bar{y})^2+(n-1)S^2}{2}\right)\right) \\
&\propto \left(\frac{n}{n+1}(\tilde{y}-\bar{y})^2+(n-1)S^2\right)^{-n/2} \\
&\propto \left(1+\frac{(\tilde{y}-\bar{y})^2}{\left(1+\frac{1}{n}\right)(n-1)S^2}\right)^{-n/2}
\end{aligned}
$$

Normal distribution with unknown mean and variance XIX

which is Student's t distribution with location $\bar{y}$, scale $\sqrt{1+\frac{1}{n}}S$, and $df = n - 1$.

Multinomial Model I

The multinomial sampling distribution is used to describe data for which each observation is one of K possible outcomes. If y is the vector of counts of the number of observations of each outcome, then

$$P(y|\theta) \propto \prod_{j=1}^{K} \theta_j^{y_j}$$

where θ_j is the probability of belonging to category j so that $\sum_{j=1}^{K} \theta_j = 1$. Also, the distribution is typically thought of as implicitly conditioning on the number of observations. The conjugate prior is given by

$$P(\theta) \propto \prod_{j=1}^{K} \theta_j^{\alpha_j - 1}$$

Multinomial Model II

which is referred to as the **Dirichlet distribution** with parameters $\alpha_1, \cdots, \alpha_K$, a multivariate generalization of the beta distribution. The corresponding posterior is also Dirichlet with parameters $\alpha_1 + y_1, \cdots, \alpha_K + y_K$,

$$P(\theta|y) \propto \prod_{j=1}^{K} \theta_j^{\alpha_j + y_j - 1}.$$

Multinomial Model III

- **Example:** Pre-election polling
  $n = 1447$; $y_1 = 727$ supported George Bush; $y_2 = 583$ supported Michael Dukakis; $y_3 = 137$ supported other candidates or expressed no opinion. Let θ_1, θ_2, and θ_3 be the proportions of Bush supporters, Dukakis supporters, and those with no opinion in the survey population.
  - The parameter of interest is $\theta_1 - \theta_2$.
  - Dirichlet prior with $\alpha_1 = \alpha_2 = \alpha_3 = 1$.
  - Then the joint posterior distribution is Dirichlet(728,584,138).
  - The posterior distribution of $\theta_1 - \theta_2$ can be found via integration, but it is easier to get 1000 points $(\theta_1, \theta_2, \theta_3)$ from the posterior Dirichlet and compute $\theta_1 - \theta_2$ for each.
  - See Poll.R.

Logistic Regression Model I

- **Example:** Analysis of Bioassay Experiment
  In the development of drugs or other chemical compounds, acute toxicity tests or bioassay experiments are commonly performed on animals. Such experiments proceed by administering dose levels to several batches of animals.
  Response: Binary, eg) Alive or Dead.
  Suppose there are k dose levels.
  data: (x_i, n_i, y_i), $i = 1, \cdots, k$
  where x_i represents the ith of the k dose levels (measured on a logarithmic scale) given to n_i animals of which y_i subsequently respond with positive outcome. 20 animals were tested.

Logistic Regression Model II

Dose x_i (log g/ml)	Number of animals n_i	Number of deaths y_i
-0.863	5	0
-0.296	5	1
0.053	5	3
0.727	5	5

Let θ_i be the probability of death for animals given dose x_i. Then

$$y_i|\theta_i \sim^{\text{ind.}} Bin(n_i, \theta_i).$$

- Use a noninformative prior $P(\theta_1, \theta_2, \theta_3, \theta_4) = 1$. Then the posterior is the product of independent beta pdf's.

Logistic Regression Model III

- Note that the dose level x_i for each group i and expect the probability of death to vary systematically as a function of the dose. Then we consider an appropriate model:

$$\text{logit}(\theta_i) = \alpha + \beta x_i, \qquad \text{(logistic regression model)}$$

The joint posterior of (α, β) given $y = (y_1, \cdots, y_K)$ is

$$P(\alpha, \beta | y) \propto L(\alpha, \beta) P(\alpha, \beta)$$

where

$$L(\alpha, \beta) = \prod_{i=1}^{K} \left(\begin{array}{c} n_i \\ y_i \end{array} \right) \frac{\left(e^{\alpha+\beta x_i}\right)^{y_i}}{\left(1 + e^{\alpha+\beta x_i}\right)^{n_i}} \propto \frac{e^{\alpha \sum_i y_i + \beta \sum_i x_i y_i}}{\prod_i \left(1 + e^{\alpha+\beta x_i}\right)^{n_i}}$$

Logistic Regression Model IV

and $P(\alpha, \beta) \propto 1$. The joint posterior is proper as long as $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L(\alpha, \beta) d\alpha d\beta < \infty$ and the finiteness of this integral can be proved when $1 \le y_i \le n_i - 1$ $(i = 1, \cdots, K)$.
This is an example of a nonconjugate prior.
See R code.

- A parameter of common interest in bioassay studies is LD50-the dose level at which the probability of death is 50%. With the present logistic model, a 50% survival rate means LD50:

$$E(y_i/n_i) = 0.5 \leftrightarrow \frac{e^{\alpha+\beta x_i}}{1 + e^{\alpha+\beta x_i}} = 0.5$$
$$\alpha + \beta x_i = 0.$$

Logistic Regression Model V

So LD50 is $x_i = -\frac{\alpha}{\beta}$. In the context of this example, LD50 is a meaningless concept of $\beta \leq 0$ in which case increasing dose does not cause the probability of death to increase. If we were certain that the drug could not cause the tumor to decrease, we can cause the parameter space to change into $(-\infty, \infty) \times (0, \infty)$.