Signle-Paprameter Models (Supp.)

Gwangsu Kim

JBNU

Second semester of 2024

Conjugate Prior Distribution I

- Let $\mathcal{F}$ be a class of sampling distribution $p(y|\theta)$, $\mathcal{P}$ be a class of prior distribution $p(\theta)$.

- The class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$p(\theta|y) \in \mathcal{P}$$

for all $p(\cdot|\theta) \in \mathcal{F}$ and $p(\cdot) \in \mathcal{P}$.

Conjugate Prior Distribution II

- Exponential family and sufficient statistics

1. Exponential family

$$p(y_i|\theta) = f(y_i)g(\theta)\exp\left(\phi(\theta)^T u(y_i)\right).$$

2. For i.i.d. $y_1, \ldots, y_n$, the density is

$$\left(\prod_{i=1}^{n} f(y_i)\right) g(\theta)^n \exp\left(\phi(\theta)^T \sum_{i=1}^{n} u(y_i)\right).$$

Conjugate Prior Distribution III

3 If $p(\theta) \propto g(\theta)^{\eta} \exp\left(\phi(\theta)^T \nu\right)$, then

$$p(\theta|y) \propto g(\theta)^{n+\eta} \exp\left(\phi(\theta)^T (\nu + t(y))\right),$$

where $t(y) = \sum_{i=1}^{n} u(y_i)$.

Negative Binomial Distribution I

- If $p(y|\theta) = \exp(-\theta)\,\theta^y/y!$ and $\theta \sim \mathcal{G}amma(\alpha, \beta)$, then

$$
\begin{aligned}
p(y) &= \frac{p(y|\theta)p(\theta)}{p(\theta|y)} \\
&= \frac{\theta^y \exp(-\theta)\,\theta^{\alpha-1}\exp(-\beta\theta)\beta^{\alpha}\Gamma(\alpha+y)}{y!\theta^{\alpha+y-1}\exp(-(1+\beta)\theta)(1+\beta)^{\alpha+y}\Gamma(\alpha)} \\
&= \frac{\Gamma(\alpha+y)\beta^{\alpha}}{\Gamma(\alpha)y!(1+\beta)^{\alpha+y}} = \binom{\alpha+y-1}{y}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y}.
\end{aligned}
$$

Negative Binomial Distribution II

- Thus, we have

$$\text{Neg-bin}(y|\alpha,\beta) = \int \text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha,\beta)d\theta.$$

- Usually, this is related to over-dispersion model, it means that the variance is larger than mean. In Poisson model, mean and variance are equal.

Asymptotic property of the posterior I

- In limit probability theory can be put in a Bayesian context to show:

$$\left(\frac{\theta - E[\theta|y]}{\sqrt{\mathrm{Var}(\theta|y)}} \middle| y \right) \to N(0, 1),$$

where $\theta \in \mathbb{R}$.

$$\left(\mathrm{Cov}(\theta|y)^{-1/2} (\theta - E[\theta|y]) \middle| y \right) \to N(0, I_p),$$

where $\theta \in \mathbb{R}^p$.

Example of Data Analysis I

- Cancer of kidney/ureter in U. S. at 1980-1989.
  1. Data are gathered in each county.
  2. Population is varying through the counties.

Example of Data Analysis II

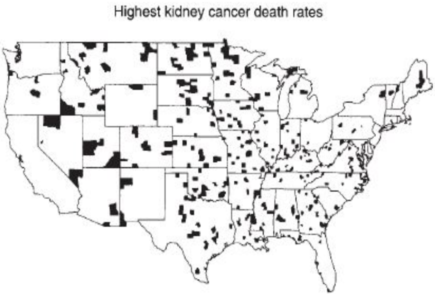

Figure 2.7 *The counties of the United States with the highest 10% age-standardized death rates for cancer of kidney/ureter for U.S. white males, 1980–1989. Why are most of the shaded counties in the middle of the country? See Section 2.8 for discussion.*

Example of Data Analysis III

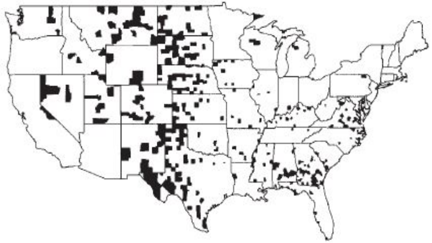

Figure 2.8 *The counties of the United States with the lowest 10% age-standardized death rates for cancer of kidney/ureter for U.S. white males, 1980–1989. Surprisingly, the pattern is somewhat similar to the map of the highest rates, shown in Figure 2.7.*

Example of Data Analysis IV

- This data implies that variations are different along with the populations of the counties.

- Counties in the Great Plains in the middle of the country appear in the lowest 10% and highest 10% simultaneously.

Example of Data Analysis V

- Thus we consider the model of

$$y_j \sim Pois(10 n_j \theta_j),$$

where n_j is the population of the county, and prior of

$$\theta_j \sim Gamma(20, 430000).$$

Example of Data Analysis VI

- Then $\theta_j|y_j \sim \mathcal{G}amma(20 + y_j, 430000 + 10n_j)$, it implies

$$E(\theta_j|y_j) = \frac{20 + y_j}{43000 + 10n_j},$$

$$Var(\theta_j|y_j) = \frac{20 + y_j}{(43000 + 10n_j)^2}.$$

- Small n_j results the relatively large weight of the prior in the posterior mean.

Example of Data Analysis VII

- (Prior) predictive distribution gives that

$$y_j \sim \text{Neg-bin}\left(20, 430000/(10n_j)\right),$$

and variance of $y_j/(10n_j)$ is larger when n_j is small, note that

$$\begin{aligned} E\left[y_j\right] &= 10n_j\frac{\alpha}{\beta}, \\ \text{Var}\left(y_j\right) &= 10n_j\frac{\alpha}{\beta} + (10n_j)^2\frac{\alpha}{\beta^2}. \end{aligned}$$

Example of Data Analysis VIII

- How to choose $\alpha = 20$ and $\beta = 430000$? Data driven.
- In some case, we consider the parameter should be estimated, empirical Bayes approach (not full Bayesian).

Example of Data Analysis IX

- We have

$$y_j \sim \text{Neg-bin}\left(\alpha, \beta/(10n_j)\right),$$

and

$$\begin{aligned} E\left[y_j\right] &= 10n_j\frac{\alpha}{\beta}, \\ \text{Var}\left(y_j\right) &= 10n_j\frac{\alpha}{\beta} + (10n_j)^2\frac{\alpha}{\beta^2}. \end{aligned}$$

- Using such as the moment method based on data.