Multiparameter Models

Most interesting and/or challenging problems in statistics involve more than one parameter.

The principles we've talked about apply regardless of the number of parameters the model has. However, putting the principles into practice becomes more difficult with multiple parameters.

Nuisance parameters

Sometimes only a subset of the parameters in $\boldsymbol{\theta}$ are of any interest. The others may be regarded as *nuisance* parameters.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_I, \boldsymbol{\theta}_N)$, where $\boldsymbol{\theta}_I$ contains the parameters of interest and $\boldsymbol{\theta}_N$ is the vector of nuisance parameters.

We may compute the posterior for the parameters of interest by averaging the posterior over all values of the nuisance parameters.

$$\pi_I(\boldsymbol{\theta}_I|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}_I, \boldsymbol{\theta}_N|\boldsymbol{y})\, d\boldsymbol{\theta}_N.$$

The conditional distribution of some subset of parameters given the values of other parameters is often of interest.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. The conditional posterior of $\boldsymbol{\theta}_1$ given $\boldsymbol{\theta}_2$ is

$$\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta}|\boldsymbol{y})}{\pi_2(\boldsymbol{\theta}_2|\boldsymbol{y})},$$

where $\pi_2(\boldsymbol{\theta}_2|\boldsymbol{y})$ is the marginal posterior of $\boldsymbol{\theta}_2$.

Conditionals are often used to facilitate the process of generating observations from the posterior. Doing so is important for computational reasons.

Suppose we want to know the mean of the posterior distribution. Usually this can't be done directly by integration.

If we could generate many values $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}$ from the posterior, we could approximate the posterior mean by $\sum_{i=1}^{M} \boldsymbol{\theta}^{(i)}/M$.

We'll have much more to say about this later in the course.

Example 12 *Random sample from a normal distribution with both parameters unknown*

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, where $Y_1, \ldots, Y_n$ are a random sample from $N(\theta_1, 1/\theta_2)$.

$$\Theta = \{(\theta_1, \theta_2) : -\infty < \theta_1 < \infty, \theta_2 > 0\}$$

As a prior we will use a member of the so-called *normal-gamma* family, which turns out to be a conjugate family of priors for this model.

A priori we assume that θ_2 has a Gamma(α, β) distribution and that $\theta_1|\theta_2$ is $N(\mu, (\tau\theta_2)^{-1})$. So,

$$\pi(\theta_1, \theta_2) \propto \sqrt{\theta_2} \exp\left(-\frac{\tau\theta_2}{2}(\theta_1 - \mu)^2\right) \theta_2^{\alpha-1} e^{-\beta\theta_2}.$$

The likelihood function is

$$f(\boldsymbol{y}|\theta_1, \theta_2) \propto \theta_2^{n/2} \exp\left[-\frac{\theta_2}{2}\sum_{i=1}^{n}(y_i - \theta_1)^2\right],$$

and hence

$$\begin{aligned}\pi(\theta_1, \theta_2|\boldsymbol{y}) \;&\propto\; \sqrt{\theta_2}\exp\left[-\frac{(\tau+n)\theta_2}{2}(\theta_1 - \mu')^2\right] \\ &\quad\times\; \theta_2^{\alpha+n/2-1}e^{-\beta'\theta_2},\end{aligned}$$

where

$$\mu' = \frac{\tau\mu + n\bar{y}}{\tau + n}$$

and

$$\beta' = \beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{\tau n(\bar{y} - \mu)^2}{2(\tau + n)}.$$

The marginal posterior distribution of θ_2 is gamma with parameters $\alpha + n/2$ and β'.

The conditional distribution of θ_1 given θ_2 is normal, but the marginal distribution of θ_1 is not normal. Try to derive the marginal posterior of θ_1.

The marginal *prior* distribution of θ_1 is

$$\pi_1(\theta_1) \propto \left[1 + \frac{1}{2\alpha} \cdot \frac{\alpha\tau(\theta_1 - \mu)^2}{\beta}\right]^{-(2\alpha+1)/2},$$

which is a t distribution with 2α degrees of freedom and precision $\alpha\tau/\beta$.

Example 13 *Box and Tiao, p. 83*

Breaking strength (in grams) was measured for 20 samples of yarn taken randomly from spinning machines.

The 20 observations are assumed to be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2.

The following noninformative (and improper) prior is used:

$$\pi(\mu, \sigma) \propto \sigma^{-1} I_{(0,\infty)}(\sigma) I_{(-\infty,\infty)}(\mu).$$

The posterior depends on the data only through sufficient statistics, which in this case are

$$\bar{y} = 50 \quad \text{and} \quad \sum_{i=1}^{20} (y_i - \bar{y})^2 = 348.$$

The posterior distribution is

$$\pi(\mu,\sigma|\boldsymbol{y}) \;\propto\; \sigma^{-21}\exp\left(-\frac{348}{2\sigma^2}\right) \times \exp\left[-\frac{20}{2\sigma^2}(\mu-50)^2\right].$$

The mode of the posterior is

$$(\widehat{\mu},\widehat{\sigma}) = \left(50,\sqrt{\frac{348}{21}}\right) = (50, 4.07).$$

A good way of summarizing the posterior is to show contours of the distribution.

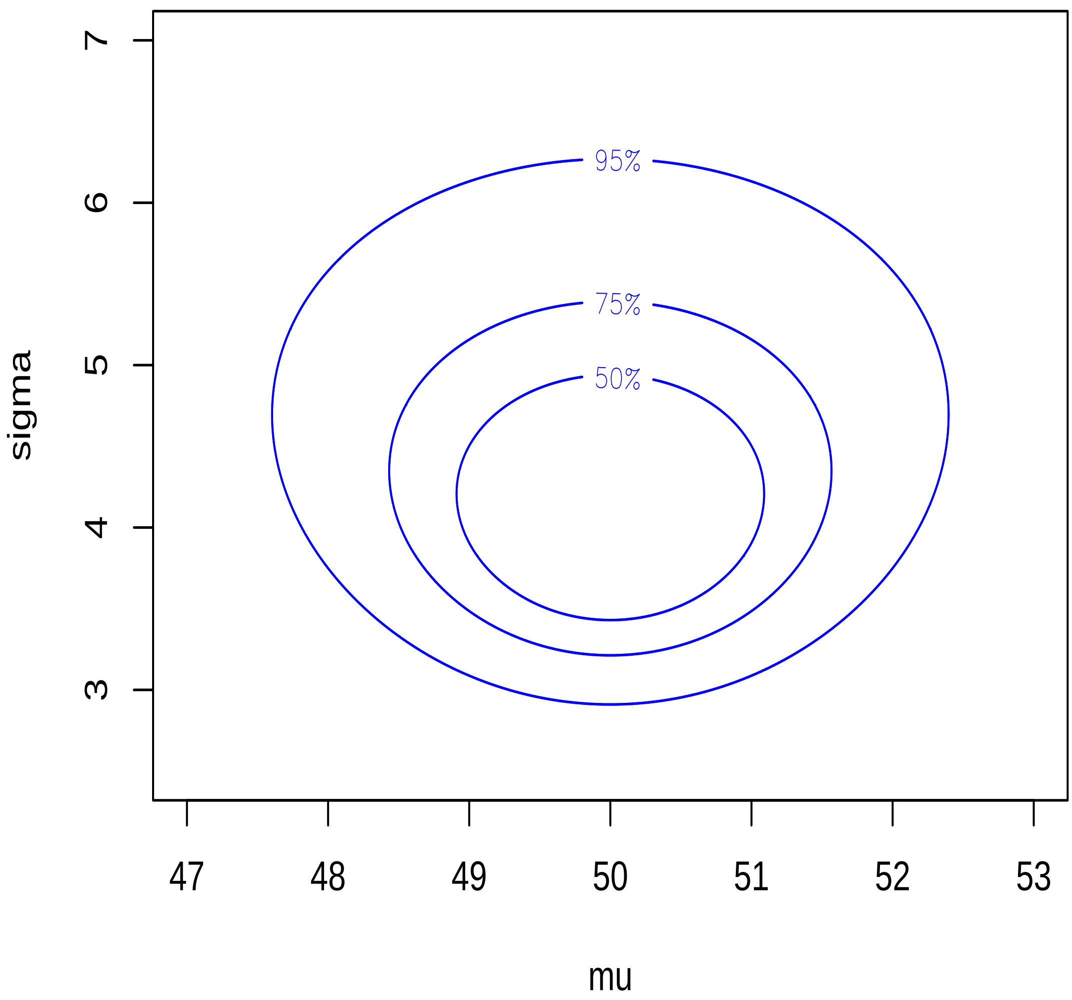
Jeffreys (*Theory of Probability*, 3rd edition) shows that, to a good approximation, the probability enclosed by the contour

$$\log(\pi(\mu,\sigma|\boldsymbol{y})) = \log(\pi(\widehat{\mu},\widehat{\sigma}|\boldsymbol{y})) - \frac{1}{2}\chi^2(2,\alpha)$$

has probability $1-\alpha$, where $\chi^2(2,\alpha)$ is the $1-\alpha$ quantile of the χ^2 distribution with 2 degrees of freedom.

Box and Tiao suggest plotting 50, 75 and 95% contours.

Fifty, 75 and 95% contours for the breaking strength data



Also of interest are component distributions of the posterior. We may write

$$\pi(\mu, \sigma|\boldsymbol{y}) = \pi(\mu|\sigma, \boldsymbol{y})\pi_2(\sigma|\boldsymbol{y}).$$

For the normal model with the noninformative prior on p. 115, we have the following results:

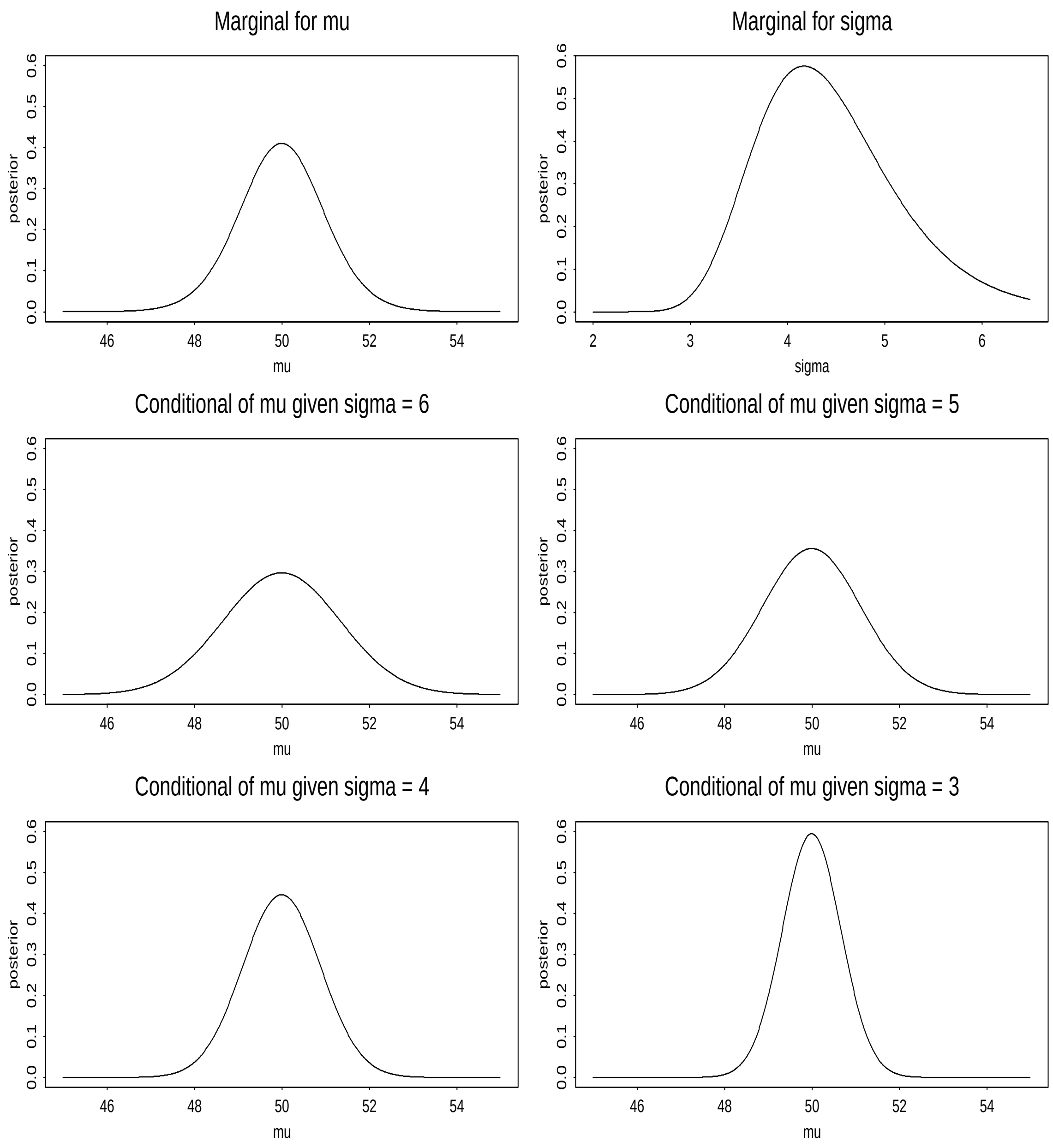
(1) $\pi(\mu|\sigma, \boldsymbol{y})$ is $N(\bar{y}, \sigma^2/n)$.

(2) The marginal posterior of σ is

$$\pi_2(\sigma|\boldsymbol{y}) \propto \sigma^{-n} \exp\left[-\frac{(n-1)s^2}{2\sigma^2}\right].$$

(3) The marginal posterior of $(\mu - \bar{y})/(s/\sqrt{n})$ is t_{n-1}.

Component distributions for breaking strength example



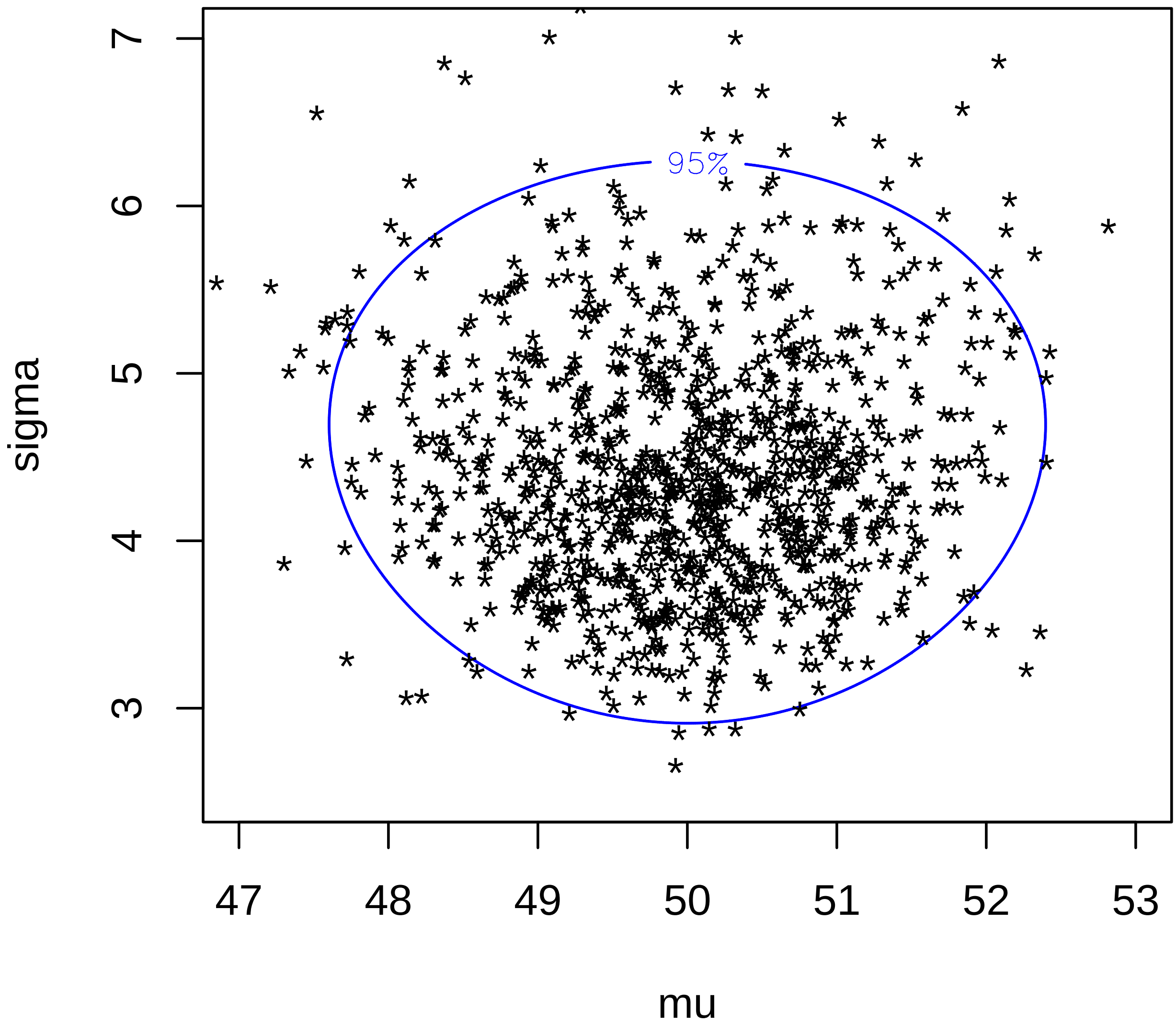
On p. 111N, it was mentioned that conditionals were often useful in generating observations from the posterior. We can use the model in Example 13 to illustrate this idea.

- The marginal posterior for σ is given in (2) on p. 118N. This implies that the posterior distribution of $1/\sigma^2$ is
$$\text{Gamma}\left[\frac{(n-1)}{2}, \frac{(n-1)s^2}{2}\right].$$
This fact allows us to generate a value of σ by first generating $1/\sigma^2$ from the appropriate gamma distribution.

- Given a value of σ, we can now generate a value of μ from $N(\bar{y}, \sigma^2/n)$.

- Repeating this process many times would give us a good impression of the distribution of (μ, σ).

Result of generating 1000 observations from the posterior in Example 13



mean of generated values of μ: 49.99
mean of generated values of σ: 4.46