3: Introduction to multiparameter models

09/03/19

Introduction

We discuss a few examples of models with more than one parameter.

A noninformative prior with a normal likelihood

Consider a normal likelihood

$$\begin{aligned} p(y \mid \mu, \sigma^2) &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2\right] \\ &= (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_i ([y_i - \bar{y}] + [\bar{y} - \mu])^2\right] \\ &= (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_i (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 + 0\right\}\right] \\ &= (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y} - \mu)^2\right\}\right] \end{aligned}$$

and the noninformative, improper prior $p(\mu, \sigma^2) \propto \sigma^{-2}$. Clearly

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

A noninformative prior with a normal likelihood

Suppose that μ is a nuisance parameter, and we're only interested in σ^2. Then, we want he marginal posterior:

$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right] \mathrm{d}\mu \\
&= (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{(n-1)}{2\sigma^2}s^2\right] \int \exp\left[-\frac{1}{2\sigma^2}n(\mu-\bar{y})^2\right] \mathrm{d}\mu \\
&\propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{(n-1)}{2\sigma^2}s^2\right] (\sigma^2)^{1/2} \\
&= (\sigma^2)^{-[(n-1)/2+1]} \exp\left[-\frac{(n-1)s^2}{2\sigma^2}\right]
\end{aligned}
$$

$\sigma^2 \mid y \sim \text{Inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$

A noninformative prior with a normal likelihood

Recall the joint posterior:

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

Clearly:

$$p(\mu \mid \sigma^2, y) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right]$$

A noninformative prior with a normal likelihood

Recall the joint posterior:

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

Clearly:

$$p(\mu \mid \sigma^2, y) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right]$$

We also have $p(\sigma^2 \mid y)$ from the last slide. This means that we can figure out the normalizing constants for the joint posterior if we multiply these two known densities together:

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y)p(\sigma^2 \mid y).$$

Sometimes this is called a **normal-inverse-gamma** distribution.

A noninformative prior with a normal likelihood

Suppose instead that σ^2 is a nuisance parameter, and we're only interested in μ. Then, we want the marginal posterior.

Let $z = \frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\} = \frac{A}{2\sigma^2}$. Then

$$\begin{aligned}
p(\mu \mid y) &\propto \int_0^\infty (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right] \mathrm{d}\sigma^2 \\
&= \int_\infty^0 (A/2)^{-(n+2)/2} z^{(n+2)/2} \exp\left[-z\right] (-A/2) z^{-2} \mathrm{d}z \\
&= (A/2)^{-n/2} \underbrace{\int_0^\infty z^{n/2-1} \exp\left[-z\right] \mathrm{d}z}_{\Gamma(n/2)}
\end{aligned}$$

A noninformative prior with a normal likelihood

So

$$\begin{aligned} p(\mu|y) &\propto (A/2)^{-n/2} \\ &\propto A^{-n/2} \\ &\propto A^{-n/2}[(n-1)s^2]^{n/2} \\ &\propto \left(1 + \frac{(\bar{y}-\mu)^2}{(n-1)s^2/n}\right)^{-n/2} \end{aligned}$$

$\mu \mid y \sim t_{n-1}(\bar{y}, s^2/n)$, that is, $\frac{\mu-\bar{y}}{s/\sqrt{n}} \mid y \sim t_{n-1}$

A noninformative prior with a normal likelihood

After we have figured out the joint posterior, we may be interested in predicting new observations with the **posterior predictive distribution**:

$$p(\tilde{y} \mid y) = \iint p(\tilde{y} \mid \mu, \sigma^2) p(\mu, \sigma^2 \mid y) \mathrm{d}\mu \mathrm{d}\sigma^2.$$

It's a homework question to show that

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

A noninformative prior with a normal likelihood

Let's get some practice simulating predictions, which will come in handy when we are dealing with more complicated scenarios where a closed-form posterior predictive distribution isn't available. We can simulate each $\tilde{y}_i$ as follows:

Sampling Strategy

For $i = 1, 2, \ldots$

1. draw $\sigma_i^2 \mid y \sim p(\sigma^2 \mid y)$
2. draw $\mu_i \mid \sigma_i^2, y \sim p(\mu \mid \sigma_i^2, y)$
3. draw $\tilde{y}_i \mid \mu_i, \sigma_i^2 \sim p(\tilde{y} \mid \mu_i, \sigma_i^2)$

Each triple $(\tilde{y}_i, \mu_i, \sigma_i^2) \sim p(\tilde{y}, \mu, \sigma^2 \mid y) = p(\tilde{y} \mid \mu, \sigma^2)p(\mu \mid \sigma^2 \mid y)p(\sigma^2 \mid y)$.

So $\tilde{y}_i \sim p(\tilde{y} \mid y) = \iint p(\tilde{y} \mid \mu, \sigma^2)p(\mu, \sigma^2 \mid y)\mathrm{d}\mu\mathrm{d}\sigma^2$

A conjugate prior with a normal likelihood

Consider a normal likelihood again

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

and an informative prior

$$p(\mu, \sigma^2) \propto \underbrace{\left(\frac{\sigma^2}{\kappa_0}\right)^{-1/2} \exp\left(-\frac{\kappa_0}{2\sigma^2}(\mu-\mu_0)^2\right)}_{p(\mu|\sigma^2)} \underbrace{(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\sigma_0^2\nu_0}{2\sigma^2}\right)}_{p(\sigma^2)}$$

$$= (\sigma^2)^{-(\frac{\nu_0+1}{2}+1)} \exp\left(-\frac{1}{2\sigma^2}\left\{\sigma_0^2\nu_0 + \kappa_0(\mu-\mu_0)^2\right\}\right)$$

This is a **normal-inverse-gamma** or **normal-inverse-**$\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$!
What are $p(\mu, \sigma^2 \mid y)$?, $p(\mu \mid y)$?

Another multiparameter example of conjugacy: Dirichlet-multinomial

Let $y = (y_1, y_2, \ldots, y_k)$ be a vector of counts. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ be the probabilities of any trial resulting in each of the k outcomes. We assume that there is a known total count (which means $\sum_i y_i = n$) and that the only possible outcomes are these k outcomes $\sum_i \theta_i = 1$.

The likelihood is a **multinomial** (aka a **categorical**) distribution

$$p(y \mid \theta) \propto \prod_{i=1}^{k} \theta_i^{y_i},$$

and the prior is a Dirichlet distribution

$$p(\theta) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}.$$

The chosen hyper-parameters have a very nice interpretation of counts!

Dirichlet-multinomial

Denote Dirichlet distribution as Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_k)$

$$p(\theta) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}.$$

$$p(\theta \mid y) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i + y_i - 1}$$

$p(\theta \mid y) \sim$ Dirichlet$(\alpha_1 + y_1, \ldots, \alpha_k + y_k)$

- Dirichlet distribution has support on a simplex $\mathcal{S} = \{\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) : \sum_{i=1}^{k} \theta_i = 1\}$
- $k = 2$, Dirichlet(α_1, α_2) becomes Beta(α_1, α_2).
- Like Beta distribution, when $\alpha_i = 1$ it becomes a uniform distribution on $\mathcal{S}$; when $\alpha_i = 0$ it is an improper prior and is equivalent to assigning a uniform prior on $\log(\theta_i)$ with the constraint that $\boldsymbol{\theta} \in \mathcal{S}$.