Multiple parameter models

April 2, 2023

Overview

1. Introduction

2. Normal data with a noninformative prior distribution

3. Normal data with a conjugate prior distribution

Introduction

- This chapter would discuss statistical problems with more than one unknown or unobservable quantity.
- Usually, in the problem, we are more interested in one than others.
- We need to find the marginal posterior distribution of the particular parameter from the joint posterior distribution of all unknowns.
- Or we draw samples from the joint posterior distribution.
- Nuisance parameters: parameters we are not interested in.

Averaging over nuisance parameters

- Suppose θ has two parts, $\theta = (\theta_1, \theta_2)$
- For example,

$$y|\mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2),$$

  in which both $\mu (= \theta_1)$ and $\sigma^2 (= \theta_2)$ are unknown.
- interest commonly centers on μ

Averaging over nuisance parameters

- The conditional distribution of the parameter of interest given observed data $p(\theta_1|y)$ could be derived from the joint posterior density,

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)p(\theta_1, \theta_2),$$

by averaging over θ_2:

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

Averaging over nuisance parameters

- Alternatively, the joint posterior density can be factored to yield

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

which shows that the posterior distribution of interest, $p(\theta_1|y)$, is a mixture of the conditional posterior distributions given the nuisance parameter, θ_2, there $p(\theta_2|y)$ is a weighting function for the different possible values of θ_2.

Averaging over nuisance parameters

- We rarely evaluate the integral of the joint density explicitly, but it suggests an important practical strategy for both constructing and computing with multi-parameter models.
- Posterior distributions can be computed by marginal $p(\theta_2)$ and conditional $p(\theta_1|\theta_2)$ (where θ_2 comes from the marginal sampling) simulation.

Normal data with a noninformative prior distribution

- Considering a vector y of n independent observations from a univariate normal distribution, $N(\mu, \sigma^2)$.
- We begin by analyzing the model under a noninformative prior distribution for μ and σ,

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}.$$

- Please deduct the joint posterior distribution $p(\mu, \sigma^2 | y)$

The joint posterior distribution

- Under the conventional improper prior density, the joint posterior distribution is proportional to the likelihood function multiplied by the factor $1/\sigma^2$:

$$
\begin{aligned}
p(\mu, \sigma^2 | y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \overline{y})^2 + n(\overline{y} - \mu)^2 \right] \right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\overline{y} - \mu)^2] \right),
\end{aligned}
$$

where

$$
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2
$$

The conditional posterior distribution

- In order to find the conditional posterior density $p(\mu|y)$, we need to find the conditional posterior density $p(\mu|\sigma^2, y)$ and the marginal posterior density $p(\sigma^2|y)$.
- From previous study, we know that

$$\mu|\sigma^2, y \sim \mathrm{N}(\overline{y}, \sigma^2/n).$$

The marginal posterior distribution

- The marginal posterior distribution $p(\sigma^2|y)$ could de deducted from averaging the joint distribution (or integrate the μ out).

$$p(\sigma^2|y) \propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y}-\mu)^2]\right) d\mu.$$

- what would be your expected result? (hint: consider the posterior distribution of σ^2 with known mean)

- The integration requires the integral $exp(-\frac{1}{2\sigma^2}n(\bar{y}-u)^2)$, which is a simple normal integral; thus

$$\begin{aligned} p(\sigma^2|y) &\propto \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right)\sqrt{2\pi\sigma^2/n} \\ &\propto (\sigma^2)^{-(n+1)/2}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right), \end{aligned}$$

  which is a scaled inverse-χ^2 density:

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2).$$

- It is easy to draw samples from the joint posterior distribution: first draw σ^2 from the marginal posterior density, then draw μ from the conditional posterior density given the drown σ^2.

analytic form of the marginal posterior distribution of μ

- The marginal posterior distribution of μ could also be deducted analytically.

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y) d\sigma^2.$$

Normal data with a noninformative prior distribution

- The integral can be evaluated using the substitution

$$z = \frac{A}{2\sigma^2}, \quad \text{where } A = (n-1)s^2 + n(\mu - \overline{y})^2,$$

- The result is an unnormalized gamma integral

$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \\ &\propto [(n-1)s^2 + n(\mu - \overline{y})^2]^{-n/2} \\ &\propto \left[1 + \frac{n(\mu - \overline{y})^2}{(n-1)s^2}\right]^{-n/2}. \end{aligned}$$

- This is the student-t distribution density $t_{n-1}(\bar{y}, s^2/n)$

Posterior predictive distribution for a future observation

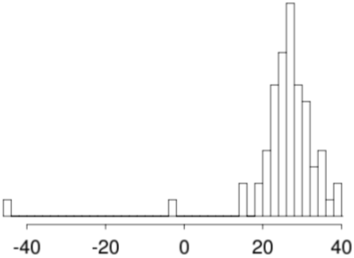
- The posterior predictive distribution for a future observation $\bar{y}$ can be written as a mixture

$$p(\tilde{y}|y) = \int\int p(\tilde{y}|\mu, \sigma^2, y)p(\mu, \sigma^2|y)d\mu d\sigma^2$$

- The integration result should be a normal distribution does not depend on y.
- We can find the posterior μ and σ^2 by simulation or analytically, then simulate $\tilde{y} \sim N(\mu, \sigma^2)$
- With the analytic result, the posterior predictive distribution should be a t distribution with location $\bar{y}$ scale $(1 + \frac{1}{n})^{1/2}s$ and $n - 1$ degrees of freedom.

Example. Estimating the speed of light

- Newcomb measured the amount of time required for light to travel a distance of 7442 meters.
- The histogram is Newcomb's 66 measurements



Example. Estimating the speed of light

- We apply the normal model assuming that all 66 measurements are independent draws from a normal distribution with mean μ and variance σ^2.
- The main substantive goal is posterior inference for μ.
- The mean of the measurement is $\bar{y} = 26.2$, and the sample standard deviation is $s = 10.8$.
- Assuming the noninformative prior distribution $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$, please find the marginal posterior distribution of μ and its 95% central posterior interval.

Example. Estimating the speed of light

- The 95% central posterior interval for μ is obtained from the t_{65} marginal posterior distribution of μ as $\bar{y} \pm 1.997 s/\sqrt{66} = [23.6, 28.8]$
- The sampling method could also be used to find the interval.
- Based on the currently accepted value of the speed of light, the 'true value' for μ in Newcomb's experiment is 33.0, which falls outside of the 95% interval. This reinforces the fact that posterior inferences are only as good as the model and the experiment that produced the data.

A family of conjugate prior distribution

- From previous study, we know the conjugate prior of μ and σ^2 is

$$\begin{aligned}\mu|\sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0)\\ \sigma^2 &\sim \mathrm{Inv}\text{-}\chi^2(\nu_0, \sigma_0^2),\end{aligned}$$

- So the corresponding joint prior should be

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

- The hyperparameters are the location μ_0 and scale σ^2/k_0 of μ and the degrees of freedom v_0 and scale σ_0^2 of σ^2.

The joint posterior distribution

- Please deduct the joint posterior distribution $p(\mu, \sigma^2|y)$.
- Please find the corresponding conditional posterior distribution $p(\mu|\sigma^2, y)$
- Please find the corresponding marginal posterior distribution $p(\sigma^2|y)$

The joint posterior distribution

- the joint posterior distribution is

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\right) \times \\
&\quad \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y} - \mu)^2]\right) \\
&= \text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2),
\end{aligned}
$$

- the conditional posterior distribution is

$$
\begin{aligned}
\mu|\sigma^2, y &\sim \mathrm{N}(\mu_n, \sigma^2/\kappa_n) \\
&= \mathrm{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\overline{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right),
\end{aligned}
$$

- the marginal posterior distribution is

$$
\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2).
$$

- The updated hyperparameters are

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\overline{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0+n}(\overline{y}-\mu_0)^2.
\end{aligned}$$

Analytic form of the marginal posterior distribution of μ

- Integration of the joint posterior density with respect to σ^2, the marginal posterior density for μ is

$$
\begin{aligned}
p(\mu|y) &\propto \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2} \\
&= t_{\nu_n}(\mu|\mu_n, \sigma_n^2/\kappa_n).
\end{aligned}
$$