Introduction to multiparameter models - part 1

Data analytics

Jerzy Baranowski

What do we mean by multiple parameters?

And what to do with them?

- Multiple unknown or unobservable quantities
- Usually we are interested only in few
- Advantage of Bayesian statistics - marginalization

We can average over ‘nuisance parameters’

- Split of vector of parameters $\theta = (\theta_1,\ \theta_2)$
- For example, in normal distribution

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

  we are often interested only in mean

Marginalization

The formal statement

- Let us consider joint posterior density

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

- We can average over θ_2

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

What is marginal distribution?

- Posterior distribution of interest, $p(\theta_1 \mid y)$ is a mixture of the conditional posterior distributions given the nuisance parameter, θ_2, where $p(\theta_2 \mid y)$ is a weighting function for the different possible values of θ_2.
- Again compromise between data and prior knowledge
- θ_2 can be discrete (multiple sub models)
- Usually we do not compute marginal distributions directly

Normal data with a non-informative prior distribution

- we consider a vector y of n independent observations from a univariate normal distribution, $N(\mu, \sigma^2)$
- We assume, that prior distributions are independent and uniform on $(\mu, \log \sigma)$
- This can be generally realized by

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

Joint posterior distribution is proportional to likelihood multiplied by $1/\sigma^2$

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

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

We can factorize the joint posterior distribution

- The conditional posterior distribution of μ is

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

- The marginal posterior of σ^2

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

Marginal posterior for σ^2 is a normal integral

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

$$\propto \sigma^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)$$

- Which can be recognized as a form of scaled inverse chi-square distribution

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

Factorization allows for easy sampling

- Joint posterior has a form

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

- It is easy to draw samples from the joint posterior distribution:
  - first draw σ^2 from $p(\sigma^2 \mid y)$,
  - then draw μ from $p(\mu \mid \sigma^2, y)$
- Through repeated simulation we can marginalize over σ^2

This case is actually analytically solvable

A rarity!

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

Integration by substitution $z = \dfrac{A}{2\sigma^2}$, where $A = (n-1)s^2 + n(\mu - \bar{y})^2$

$$\begin{aligned}
p(\mu \,|\, y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) \mathrm{d}z \\
&\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2} \\
&\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2} \\
&= t_{n-1}(\bar{y}, s^2/n)
\end{aligned}$$

Posterior predictive distribution for a future observation

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

The first of the two factors in the integral is just the normal distribution for the future observation given the values of (μ, σ^2), and does not depend on y at all. To draw from the posterior predictive distribution, first draw μ, σ^2 from their joint posterior distribution and then simulate

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

Newcomb’s experiment