Chapter 3 - Multi-parameter models: Exercise solutions

Mattias Villani

Click on the arrow to see a solution.

Exercise 3.1

Let $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$, where θ is assumed known. Show that the $\mathrm{Inv}-\chi^2$ distribution is a conjugate prior for σ^2.

Solution

The normal density function for a single observation x_i is

$$p(x_i|\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x_i-\theta)^2\Big) \propto \frac{1}{(\sigma^2)^{1/2}} \exp\Big(-\frac{1}{2\sigma^2}(x_i-\theta)^2\Big)$$

The likelihood for the iid normal model with known mean θ is therefore the product of n such density functions:

$$\begin{aligned} p(x_1,\ldots,x_n|\theta,\sigma^2) = \prod_{i=1}^{n} p(x_i|\theta,\sigma^2) &\propto \frac{1}{(\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\Big) \\ &\propto \frac{1}{(\sigma^2)^{n/2}} \exp\Big(-\frac{ns^2}{2\sigma^2}\Big) \end{aligned}$$

where we have defined $s^2 = \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{n}$ as the sample standard deviation (dividing by n instead of $n-1$ since the mean θ is assumed known).
The density of the $\sigma^2 \sim \mathrm{Scaled}-\mathrm{Inv}-\chi^2(\nu_0, \sigma_0^2)$ prior is of the form

$$p(\sigma^2) \propto \frac{1}{(\sigma^2)^{1+\nu_0/2}} \exp\Big(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\Big)$$

The posterior distribution from using the $\sigma^2 \sim \mathrm{Scaled}-\mathrm{Inv}-\chi^2(\nu_0, \sigma_0^2)$ prior is given by Bayes' theorem as (to avoid cluttering the notation, we do not write out the conditioning on θ since it is known)

$$
\begin{aligned}
p(\sigma^2|x_1,\dots,x_n) &\propto p(x_1,\dots,x_n|\sigma^2)p(\sigma^2) \\
&\propto \frac{1}{(\sigma^2)^{n/2}} \exp\Big(-\frac{ns^2}{2\sigma^2}\Big)\frac{1}{(\sigma^2)^{1+\nu_0/2}} \exp\Big(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\Big) \\
&\propto \frac{1}{(\sigma^2)^{1+(\nu_0+n)/2}} \exp\Big(-\frac{1}{2\sigma^2}(\nu_0\sigma_0^2+ns^2)\Big) \\
&= \frac{1}{(\sigma^2)^{1+(\nu_0+n)/2}} \exp\Big(-\frac{\nu_0+n}{2\sigma^2}\frac{\nu_0\sigma_0^2+ns^2}{\nu_0+n}\Big)
\end{aligned}
$$

which is proportional to a $\sigma^2 \sim \chi^2(\nu_0+n, \frac{\nu_0\sigma_0^2+ns^2}{\nu_0+n})$ density. Note how the location parameter in the posterior

$$
\frac{\nu_0\sigma_0^2+ns^2}{\nu_0+n} = \frac{\nu_0}{\nu_0+n}\sigma_0^2 + \frac{n}{\nu_0+n}s^2
$$

is a weighted average of prior location σ_0^2 and the data estimate s^2, with more weight placed on the strongest information source (prior with ν_0 imaginary sample data points versus the data with n actual data points).
