LINKÖPING UNIVERSITY
Division of Statistics and Machine Learning
Department of Computer and Information Science
Mattias Villani and Per Sidén

2018-04-27
Bayesian Learning
Advanced level course

Mathematical Exercises 4

Try to solve the problems before class. Don't worry if you fail, the important thing is trying.
You should not hand in any solutions.
This part of the course is not obligatory and is not graded.

---

1. BERNOULLI MEETS LAPLACE

   (a) Let $x_1, ..., x_n \overset{iid}{\sim} Bernoulli(\theta)$ and assume the prior $\theta \sim Beta(\alpha, \beta)$. Derive the marginal likelihood for this model.

   Solution: Let $\boldsymbol{y} = (y_1, \ldots, y_n)$. The marginal likelihood is

$$
\begin{aligned}
p(\boldsymbol{y}) &= \int p(\boldsymbol{y}|\theta)p(\theta)d\theta \\
&= \int \theta^s (1-\theta)^f \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \\
&= \frac{1}{B(\alpha,\beta)} \int \theta^{\alpha+s-1}(1-\theta)^{\beta+f-1} d\theta \\
&\text{[The integral is with respect to the kernel of a Beta}(a+s, b+f) \text{ density]} \\
&= \frac{B(\alpha+s, \beta+f)}{B(\alpha,\beta)}
\end{aligned}
$$

   (b) Compute the marginal likelihood of the model in a) using the Laplace approximation.

   Solution: The Laplace approximation of a log marginal likelihood is

$$
\ln \hat{p}(\boldsymbol{y}) = \ln p(\boldsymbol{y}|\hat{\theta}) + \ln p(\hat{\theta}) + (1/2) \ln \left| J^{-1}_{\boldsymbol{y},\hat{\theta}} \right| + (1/2) \ln(2\pi),
$$

   where $\hat{\theta}$ is the posterior mode and $J_{\boldsymbol{y},\hat{\theta}}$ is minus the second derivative at the mode. Now,

$$
\begin{aligned}
\ln p(\theta|\boldsymbol{y}) &= -\text{Beta}(a+s, b+f) + (\alpha+s-1)\ln\theta + (\beta+f-1)\ln(1-\theta) \\
\frac{d \ln p(\theta|\boldsymbol{y})}{d\theta} &= \frac{\alpha+s-1}{\theta} - \frac{\beta+f-1}{1-\theta} \\
\frac{d^2 \ln p(\theta|\boldsymbol{y})}{d\theta^2} &= -\frac{\alpha+s-1}{\theta^2} - \frac{\beta+f-1}{(1-\theta)^2}
\end{aligned}
$$

   Solving $d\ln p(\theta|\boldsymbol{y})/d\theta = 0$ for θ gives the posterior mode

$$
\hat{\theta} = \frac{\alpha+s-1}{\alpha+\beta+n-2}.
$$

and therefore

$$J^{-1}_{\boldsymbol{y},\hat{\theta}} = -\left[\frac{d^2 \ln p(\theta|\boldsymbol{y})}{d\theta^2}\Big|_{\theta=\hat{\theta}}\right]^{-1} = \frac{(\alpha+s-1)(\beta+f-1)}{(\alpha+\beta+n-2)^3}.$$

(c) Is this approximation accurate if $\alpha = \beta = 1$ and you have observed $s = 6$ success in $n = 10$ trials?

Solution: For this data we have $\hat{\theta} = s/n = 0.6$ and $J^{-1}_{\boldsymbol{y},\hat{\theta}} = sf/n^3 = 0.024$. So, $\ln \hat{p}(\boldsymbol{y}) = -7.676$ which is quite close to the true log marginal likelihood $\ln p(\boldsymbol{y}) = -7.745$. One way to see that the approximation is accurate is for example to see the differences in comparing a model with a null model where $\theta = 1/2$.

2. Fill in the blanks - again

(a) Derive the marginal likelihood for the Poisson model with Gamma prior at the end of Slide 7 at Lecture 10.

Solution: The marginal likelihood is

$$\begin{aligned}
p(x_1,\ldots,x_n) &= \int p(x_1,\ldots,x_n|\theta)p(\theta)d\theta \\
&= \int \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^{n} x_i!} \int \theta^{\alpha+\sum_{i=1}^{n} x_i - 1} e^{-(\beta+n)\theta} d\theta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^{n} x_i!} \frac{\Gamma(\alpha+\sum_{i=1}^{n} x_i)}{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}} \\
&= \frac{\beta^\alpha}{(\beta+n)^{\alpha+n\bar{x}}} \frac{\Gamma(\alpha+n\bar{x})}{\Gamma(\alpha)\prod_{i=1}^{n} x_i!}
\end{aligned}$$

3. Pareto

(a) Let $x_1,\ldots,x_n \overset{iid}{\sim} \text{Uniform}(0,\theta)$. Let $\theta \sim \text{Pareto}(\alpha,\beta)$, that is

$$p(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}},\ \theta \geq \beta.$$

Show that this is a conjugate prior to this Uniform model and derive the posterior for θ. [Hint: Don't forget to include an indicator function when you write up the likelihood function. The Uniform$(0,\theta)$ distribution is zero for outcomes larger than θ.]

Solution: The likelihood function is of the form

$$\prod_{i=1}^{n} \frac{1}{\theta} I(x_i \leq \theta) = \left(\frac{1}{\theta}\right)^n I(x_{\max} \leq \theta)$$

where $x_{\max} = \max(x_1, \ldots, x_n)$, and the Pareto prior is

$$p(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} \cdot I(\beta \le \theta).$$

By Bayes' theorem we therefore have

$$\begin{aligned} p(\theta|x_1, \ldots, x_n) &\propto p(x_1, \ldots, x_n|\theta)p(\theta) \\ &= \left(\frac{1}{\theta}\right)^n I(x_{\max} \le \theta) \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} \cdot I(\beta \le \theta) \\ &= \frac{\alpha\beta^\alpha}{\theta^{n+\alpha+1}} \cdot I\big(\tilde{\beta} \le \theta\big) \end{aligned}$$

where $\tilde{\beta} = \max(x_{\max}, \beta)$. This is proportional to a $\text{Pareto}\big(\alpha + n, \tilde{\beta}\big)$ density.

(b) Derive the predictive distribution of x_{n+1} given $x_1, \ldots, x_n$. [Hint: It is wise to break up the integrals in two parts.]

Solution: From a) the posterior distribution is

$$\theta|x_1, \ldots, x_n \sim \text{Pareto}(\alpha + n, \tilde{\beta}),$$

where $\tilde{\beta} = \max(x_{\max}, \beta)$. The predictive distribution is

$$\begin{aligned} p(x_{n+1}|x_{1:n}) &= \int_0^\infty p(x_{n+1}|\theta)p(\theta|x_{1:n})d\theta \\ &= \int_0^\infty \frac{1}{\theta} I(x_{n+1} \le \theta) \frac{(\alpha+n)\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+1}} \cdot I\big(\tilde{\beta} \le \theta\big)d\theta \\ &= (\alpha + n) \int_0^\infty \frac{\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+2}} I\big(\max(x_{n+1}, \tilde{\beta}) \le \theta\big)d\theta \end{aligned}$$

In order to compute this integral we will separate the integral in two cases: i) $x_{n+1} \le \tilde{\beta}$ where $\max(x_{n+1}, \tilde{\beta}) = \tilde{\beta}$ and ii) $x_{n+1} > \tilde{\beta}$ where $\max(x_{n+1}, \tilde{\beta}) = x_{n+1}$. Now, when $x_{n+1} \le \tilde{\beta}$, we have

$$\begin{aligned} p(x_{n+1}|x_{1:n}) &= (\alpha + n) \int_0^\infty \frac{\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+2}} I\big(\max(x_{n+1}, \tilde{\beta}) \le \theta\big)d\theta \\ &= (\alpha + n) \int_{\tilde{\beta}}^\infty \frac{\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+2}} d\theta \\ &= \frac{\alpha + n}{(\alpha + n + 1)\tilde{\beta}} \int_{\tilde{\beta}}^\infty \frac{(\alpha + n + 1)\tilde{\beta}^{(\alpha+n+1)}}{\theta^{(n+\alpha+1)+1}} d\theta \\ &= \frac{\alpha + n}{\alpha + n + 1} \frac{1}{\tilde{\beta}} \end{aligned}$$

This shows that the predictive distribution for x_{n+1} is $\frac{\alpha+n}{\alpha+n+1} \cdot \text{Uniform}(x_{n+1}|0, \tilde{\beta})$ when

$x_{n+1} \leq \tilde{\beta}$. Turning now to the other case when $x_{n+1} > \tilde{\beta}$ we have

$$
\begin{aligned}
p(x_{n+1}|x_{1:n}) &= (\alpha+n)\int_0^\infty \frac{\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+2}} I\big(\max(x_{n+1},\tilde{\beta}) \leq \theta\big) d\theta \\
&= (\alpha+n)\int_{x_{n+1}}^\infty \frac{\tilde{\beta}^{(\alpha+n)}}{\theta^{n+\alpha+2}} d\theta \\
&= \frac{(\alpha+n)\tilde{\beta}^{(\alpha+n)}}{(\alpha+n+1)x_{n+1}^{\alpha+n+1}} \int_{x_{n+1}}^\infty \frac{(\alpha+n+1)x_{n+1}^{\alpha+n+1}}{\theta^{n+\alpha+2}} d\theta \\
&= \frac{1}{(\alpha+n+1)} \frac{(\alpha+n)\tilde{\beta}^{(\alpha+n)}}{x_{n+1}^{\alpha+n+1}},
\end{aligned}
$$

which can be recognized as $\frac{1}{\alpha+n+1} \cdot \text{Pareto}(x_{n+1}|\alpha+n,\tilde{\beta})$ In summary,

$$
x_{n+1}|x_{1:n} \sim \begin{cases} \frac{\alpha+n}{\alpha+n+1} \cdot \text{Uniform}(x_{n+1}|0,\tilde{\beta}), & \text{if } x_{n+1} \leq \tilde{\beta} \\ \frac{1}{\alpha+n+1} \cdot \text{Pareto}(x_{n+1}|\alpha+n,\tilde{\beta}), & \text{if } x_{n+1} > \tilde{\beta}, \end{cases}
$$

where $\tilde{\beta} = \max(x_{\max}, \beta)$.

Have fun!

- Mattias