Mathematical Exercises 4

1) a) See Lecture 11, slide 3: $p(y) = \dfrac{B(\alpha+s, \beta+f)}{B(\alpha, \beta)}$

b) From Math Ex 3, problems 3a and 3b, we have that

$$\hat{\theta} = \frac{\alpha+s-1}{\alpha+\beta+n-2} \quad \text{and} \quad J_x^{-1}(\hat{\theta}) = \frac{(\alpha+s-1)(\beta+f-1)}{(\alpha+\beta+n-2)^3}$$

From Lecture 11, slide 3, we get that

$$\ln p(y|\hat{\theta}) = s \cdot \ln \hat{\theta} + f \cdot \ln(1-\hat{\theta})$$

$$\ln p(\hat{\theta}) = (\alpha-1) \cdot \ln \hat{\theta} + (\beta-1) \cdot \ln(1-\hat{\theta}) - \log B(\alpha, \beta)$$

The Laplace approximation is then given by

$$\ln \hat{p}(y) = \ln p(y|\hat{\theta}) + \ln p(\hat{\theta}) + \frac{1}{2} \ln J_x^{-1}(\hat{\theta}) + \frac{1}{2} \ln(2\pi)$$

c) $\alpha = \beta = 1$ gives that $p(\theta) = 1$ so that $\ln p(\hat{\theta}) = 0$, and with $s = 7$, $n = 11$, $f = 4$ we have that $\hat{\theta} = \frac{s}{n} = \frac{7}{11}$ and $J_x^{-1}(\hat{\theta}) = \frac{sf}{n^3} = \frac{7 \cdot 4}{11^3}$

$$\ln p(y|\hat{\theta}) = 7 \cdot \ln\left(\frac{7}{11}\right) + 4 \cdot \ln\left(\frac{4}{11}\right) \approx -7{,}210$$

This gives $\ln \hat{p}(y) \approx -8{,}22$

$$\ln p(y) = \ln\left[\frac{B(1+7, 1+4)}{\underbrace{B(1,1)}_{=1}}\right] \approx -8{,}28$$

The approximation seems pretty accurate.

Mathematical Exercises 4

2) a)

$$p(\theta) = \frac{1}{B(\alpha,\beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(y|\theta) = \prod_{i=1}^{n} (1-\theta)^{y_i}\,\theta = \theta^{n}(1-\theta)^{\overbrace{\sum_{i=1}^{n} y_i}^{n\bar{y}}}$$

This gives

$$p(y) = \frac{1}{B(\alpha,\beta)} \cdot \underbrace{\int_0^1 \underbrace{\theta^{n}(1-\theta)^{n\bar{y}}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}}_{\theta^{\alpha+n-1}(1-\theta)^{\beta+n\bar{y}-1}}\,d\theta}_{=B(\alpha+n,\,\beta+n\bar{y})} = \frac{B(\alpha+n,\beta+n\bar{y})}{B(\alpha,\beta)}$$

Replacing the Beta functions using

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$ gives the result.

b)

$$p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta}$$

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} = \frac{1}{\prod_{i=1}^{n} y_i!} \cdot \theta^{\overbrace{\sum_{i=1}^{n} y_i}^{=n\bar{y}}} e^{-n\theta}$$

This gives

$$p(y) = \frac{\beta^{\alpha}}{\Gamma(\alpha)\prod_{i=1}^{n} y_i!} \cdot \underbrace{\int_0^{\infty} \underbrace{\theta^{n\bar{y}} e^{-n\theta}\,\theta^{\alpha-1} e^{-\beta\theta}}_{=\theta^{\alpha+n\bar{y}-1} e^{-(\beta+n)\theta}}\,d\theta}_{=\frac{\Gamma(\alpha+n\bar{y})}{(\beta+n)^{\alpha+n\bar{y}}}}$$

3. The Pareto family

(a) $p(x\,, \dots, x_n|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(x_i \le \theta) = \left(\frac{1}{\theta}\right)^n I(x_{\max} \le \theta)$

where $x_{\max} = \max(x_1, \dots, x_n)$, 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

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

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

(b)

From a) the posterior distribution is

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

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

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

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