Exercises from notes by Nils Lid Hjort

Geir Storvik

Spring 2020

Exercise 6

(a) We have

$$\begin{aligned}
p(y) =& \int_\theta p(y|\theta)p(\theta)d\theta \\
=& \int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{\Gamma(k)}{\Gamma(k\theta_0)\Gamma(k(1-\theta_0))}\theta^{k\theta_0-1}(1-\theta)^{k(1-\theta_0)-1}d\theta \\
=& \binom{n}{y}\frac{\Gamma(k)}{\Gamma(k\theta_0)\Gamma(k(1-\theta_0))}\frac{\Gamma(k\theta_0+y)\Gamma(k(1-\theta_0)+n-y)}{\Gamma(k+n)} \\
& \int_0^1 \frac{\Gamma(k+n)}{\Gamma(k\theta_0+y)\Gamma(k(1-\theta_0)+n-y)}\theta^{k\theta_0+y-1}(1-\theta)^{k(1-\theta_0)+n-y}d\theta \\
=& \binom{n}{y}\frac{\Gamma(k)}{\Gamma(k\theta_0)\Gamma(k(1-\theta_0))}\frac{\Gamma(k\theta_0+y)\Gamma(k(1-\theta_0)+n-y)}{\Gamma(k+n)}
\end{aligned}$$

which corresponds to Beta-Binomial$(n, k\theta_0, k(1-\theta_0)$ which has variance

$$\begin{aligned}
\text{Var}(y) =& n\frac{k\theta_0 k(1-\theta_0)(k+n)}{k^2(k+1)} \\
=& n\theta_0(1-\theta_0)\frac{k+n}{k+1}
\end{aligned}$$

(b) We have that we can write

$$\begin{aligned}
y &= \theta + \sigma\varepsilon, & \varepsilon &\sim N(0,1) \\
\theta &= \tau\eta, & \eta &\sim N(0,1)
\end{aligned}$$

with ε, η independent. Then

$$y = \sigma\varepsilon + \tau\eta$$

which is a linear combination of two Gaussian variables, which itself becomes Gaussian with $E[y] = 0$ and $\text{Var}[y] = \tau^2 + \sigma^2$.

(c) We have from above that $y_i \sim N(0, \sigma^2 + \tau^2)$. Now since

$$\begin{aligned}
y_i &= \theta + \sigma\varepsilon_i \\
y_j &= \theta + \sigma\varepsilon_j
\end{aligned}$$

with the ε_i terms independent. Then

$$\text{cov}[y_i, y_j] = \text{Var}[\theta] = \tau^2$$

giving

$$\text{corr}[y_i, y_j] = \frac{\tau^2}{\sigma^2 + \tau^2}$$

The whole vector $\boldsymbol{y}$ then has a multivariate Gaussian distribution $\boldsymbol{y} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\Sigma_{ii} = \sigma^2 + \tau^2$ and $\Sigma_{ij} = \tau^2$ for $j \neq i$.

(d) We have

$$\begin{aligned}
p(y_i) =& \int_\theta p(y_i|\theta)p(\theta)d\theta \\
=& \int_0^\infty \frac{\theta^{y_i}e^{-\theta}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta \\
=& \frac{1}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+y_i)}{(\beta+1)^{\alpha+y_i}} \\
& \int_0^\infty \frac{(\beta+1)^{\alpha+y_i}}{\Gamma(\alpha+y_i)}\theta^{\alpha+y_i-1}e^{-(\beta+1)\theta}d\theta \\
=& \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^{\alpha+y_i} \frac{\Gamma(\alpha+y_i)}{\Gamma(y_i+1)\Gamma(\alpha)}
\end{aligned}$$

which is the Negative binomial distribution with parameters α, β. Further,

$$\begin{aligned}
p(\boldsymbol{y}) =& \int_\theta \prod_{i=1}^n p(y_i|\theta)p(\theta)d\theta \\
=& \int_0^\infty \prod_{i=1}^n \frac{\theta^{y_i}e^{-\theta}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta \\
=& \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n}\frac{\Gamma(\alpha+n\bar{y})}{(\beta+n)^{\alpha+n\bar{y}}}\prod_{i=1}^n\frac{1}{y_i!} \\
& \int_0^\infty \frac{(\beta+n)^{\alpha+n\bar{y}}}{\Gamma(\alpha+n\bar{y})}\theta^{\alpha+n\bar{y}-1}e^{-(\beta+n)\theta}d\theta \\
=& \frac{\beta^{n\alpha}}{(\beta+n)^{\alpha+n\bar{y}}}\frac{\Gamma(\alpha+n\bar{y})}{\Gamma(\alpha)^n}\prod_{i=1}^n\frac{1}{y_i!}
\end{aligned}$$

We have further that

$$\begin{aligned}
E[y_i] =& E[E[y_i|\theta]] = E[\theta] = \frac{\alpha}{\beta} \\
\text{Var}[y_i] =& E[\text{Var}[y_i|\theta]] + \text{Var}[E[y_i|\theta]] \\
=& E[\theta] + \text{Var}[\theta] \\
=& \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} = \frac{\alpha(\beta+1)}{\beta^2} \\
\text{Cov}[y_i, y_j] =& E[\text{Cov}[y_i, y_j|\theta]] + \text{Cov}[E[y_i|\theta], E[y_j|\theta]] \\
=& E[0] + \text{Cov}[\theta, \theta] = \frac{\alpha}{\beta^2} \\
\text{Corr}[y_i, y_j] =& \frac{1}{\beta+1}
\end{aligned}$$

(e) We have

$$\begin{aligned} f(y) =& \int_\theta f(y|\theta)p(\theta)d\theta \\ =& \int_\theta \exp(l_n(\theta))p(\theta)d\theta \\ =& \exp(l_n(\hat{\theta})) \int_\theta \exp(l_n(\theta) - l_n(\hat{\theta}))p(\theta)d\theta \\ =& L_n(\hat{\theta}) \int_\theta \exp(l_n(\theta) - l_n(\hat{\theta}))p(\theta)d\theta \end{aligned}$$

Now make a Taylor expansion of $l_n(\theta)$ around $\hat{\theta}$, so that

$$\begin{aligned} l_n(\theta) \approx& l_n(\hat{\theta}) + \frac{\partial l_n(\hat{\theta}}{\partial \theta^T}(\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T \frac{\partial^2 l_n(\hat{\theta}}{\partial \theta \partial \theta^T}(\theta - \hat{\theta}) \\ =& l_n(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^T \hat{J}(\theta - \hat{\theta}) \end{aligned}$$

where we have neglected higher order terms and used that $\frac{\partial l_n(\hat{\theta}}{\partial \theta^T} = 0$ for the maximum likelihood estimate. Then (also approximating $p(\theta)$ by $p(\hat{\theta})$, we get

$$\begin{aligned} f(y) \approx& L_n(\hat{\theta}) \int_\theta \exp(-\frac{1}{2}(\theta - \hat{\theta})^T \hat{J}(\theta - \hat{\theta})p(\hat{\theta})d\theta \\ =& L_n(\hat{\theta}) \frac{(2\pi)^{p/2}}{|\hat{J}|^{1/2}} p(\hat{\theta}) \int_\theta \frac{|\hat{J}|^{1/2}}{(2\pi)^{p/2}} \exp(-\frac{1}{2}(\theta - \hat{\theta})^T \hat{J}(\theta - \hat{\theta})d\theta \\ =& L_n(\hat{\theta}) \frac{(2\pi)^{p/2}}{|\hat{J}|^{1/2}} p(\hat{\theta}) \end{aligned}$$

(f) We need $\hat{\theta}$ to be close to θ which can be obtained if $\hat{\theta}$ is a consistent estimator for θ and n reasonable large. We further need the Taylor expansion to provide a good approximation which requires $l_n(\theta)$ to be smooth. We also need $p(\theta)$ to be smooth such that $p(\theta) \approx p(\hat{\theta})$.

When we have n iid observations, we obtain

$$\hat{J} = n\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 l_i(\hat{\theta}}{\partial \theta \partial \theta^T} = nJ_n^*$$

where now l_i is the log-likelihood for observation i. The average J_m^* will then converge towards its expectation so that $\hat{J}$ itself will increase linearly with n. We then have

$$\begin{aligned} \log f(y) \approx& l_n(\hat{\theta}) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|nJ_n^*| + \log p(\hat{\theta}) \\ =& l_n(\hat{\theta}) - \frac{p}{2}\log(2\pi) - \frac{1}{2}p\log(n) - \log|J_n^*| + \log p(\hat{\theta}) \\ \approx& l_n(\hat{\theta}) - \frac{1}{2}p\log(n) \end{aligned}$$

where we have ignored all terms not increasing with n

Exercise 11

(a) See `Nils_11.R` for code.

(*b*) We have

$$\begin{pmatrix} Z(s) \\ Z(t) \\ Z(u) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^{t-s} & \rho^{u-s} \\ \rho^{t-s} & 1 & \rho^{u-t} \\ \rho^{u-s} & \rho^{u-t} & 1 \end{pmatrix} \right)$$

showing that $Z(u)|Z(t), Z(s)$ is Gaussian with

$$\begin{aligned} E[Z(u)|Z(t), Z(s)] &= \begin{pmatrix} \rho^{u-s} & \rho^{u-t} \end{pmatrix} \begin{pmatrix} 1 & \rho^{t-s} \\ \rho^{t-s} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} \\ &= \begin{pmatrix} \rho^{u-s} & \rho^{u-t} \end{pmatrix} \frac{1}{1-\rho^{2(t-s)}} \begin{pmatrix} 1 & -\rho^{t-s} \\ -\rho^{t-s} & 1 \end{pmatrix} \begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} \\ &= \frac{1}{1-\rho^{2(t-s)}} \begin{pmatrix} \rho^{u-s} - \rho^{u-t}\rho^{t-s} & -\rho^{u-s}\rho^{t-s} + \rho^{u-t} \end{pmatrix} \begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} \\ &= \frac{1}{1-\rho^{2(t-s)}} \begin{pmatrix} 0 & \rho^{u-t} - \rho^{u+t-2s} \end{pmatrix} \begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} \\ &= \frac{\rho^{u-t} - \rho^{u+t-2s}}{1-\rho^{2(t-s)}} Z(t) \\ &= \frac{\rho^{u-t}[1-\rho^{2t-2s}]}{1-\rho^{2(t-s)}} Z(t) \\ &= \rho^{u-t} Z(t) \end{aligned}$$

Alternatively,

$$\begin{aligned} \mathrm{Var}[\begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} |Z(t)] &= \begin{pmatrix} 1 & \rho^{u-s} \\ \rho^{u-s} & 1 \end{pmatrix} - \begin{pmatrix} \rho^{t-s} \\ \rho^{u-t} \end{pmatrix} \begin{pmatrix} \rho^{t-s} & \rho^{u-t} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho^{u-s} \\ \rho^{u-s} & 1 \end{pmatrix} - \begin{pmatrix} \rho^{2(t-s)} & \rho^{u-s} \\ \rho^{u-s} & \rho^{2(u-t)} \end{pmatrix} \\ &= \begin{pmatrix} 1-\rho^{2(t-s)} & 0 \\ 0 & 1-\rho^{2(u-t)} \end{pmatrix} \end{aligned}$$

showing that they are conditionally independent.

We can then simulate $Z(u)|Z(t) \sim N(\rho^{u-t}Z(t), 1-\rho^{2(u-t)})$

(*c*) We have

$$\begin{pmatrix} Z(0) \\ Z(s) \\ Z(t) \\ Z(1) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^{s} & \rho^{t} & \rho \\ \rho^{s} & 1 & \rho^{t-s} & \rho^{1-s} \\ \rho^{t} & \rho^{t-s} & 1 & \rho^{1-t} \\ \rho & \rho^{1-s} & \rho^{1-t} & 1 \end{pmatrix} \right)$$

giving

$$\begin{aligned} E\left(Z(s)| \begin{pmatrix} Z(0) \\ Z(1) \end{pmatrix} \right) &= \begin{pmatrix} \rho^{s} & \rho^{1-s} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \frac{1}{1-\rho^2} \begin{pmatrix} \rho^{s} & \rho^{1-s} \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \frac{1}{1-\rho^2} \begin{pmatrix} \rho^{s} - \rho^{2-s} & \rho^{1-s} - \rho^{1+s} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \frac{1}{1-\rho^2} (\rho^{s} - \rho^{2-s})a + (\rho^{1-s} - \rho^{1+s})b \end{aligned}$$

and

$$
\begin{aligned}
&\mathrm{Var}\left(\begin{pmatrix} Z(s) \\ Z(t) \end{pmatrix} | \begin{pmatrix} Z(0) \\ Z(1) \end{pmatrix}\right) \\
&= \begin{pmatrix} 1 & \rho^{t-s} \\ \rho^{t-s} & 1 \end{pmatrix} - \begin{pmatrix} \rho^s & \rho^{1-s} \\ \rho^t & \rho^{1-t} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho^s & \rho^t \\ \rho^{1-s} & \rho^{1-t} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \rho^{t-s} \\ \rho^{t-s} & 1 \end{pmatrix} - \frac{1}{1-\rho^2} \begin{pmatrix} \rho^s & \rho^{1-s} \\ \rho^t & \rho^{1-t} \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \rho^s & \rho^t \\ \rho^{1-s} & \rho^{1-t} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \rho^{t-s} \\ \rho^{t-s} & 1 \end{pmatrix} - \frac{1}{1-\rho^2} \begin{pmatrix} \rho^s & \rho^{1-s} \\ \rho^t & \rho^{1-t} \end{pmatrix} \begin{pmatrix} \rho^s - \rho^{2-s} & \rho^t - \rho^{2-t} \\ \rho^{1-s} - \rho^{1+s} & \rho^{1-t} - \rho^{1+t} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \rho^{t-s} \\ \rho^{t-s} & 1 \end{pmatrix} - \frac{1}{1-\rho^2} \begin{pmatrix} \rho^{2s} - 2\rho^2 + \rho^{2-2s} & \rho^{s+t} - 2\rho^{2+s-t} + \rho^{2-s-t} \\ \rho^{t+s} - 2\rho^{2+t-s} + \rho^{2-s-t} & \rho^{2t} - 2\rho^2 + \rho^{2-2t} \end{pmatrix}
\end{aligned}
$$

Exercise 12

(a) For θ uniform, we have $p(\theta|y) = \text{Beta}(y, n-y)$ giving

$$
\begin{aligned}
E[L(\theta, \text{no alarm}] &= 5000 * \Pr(\theta > 0.15) \\
E[L(\theta, \text{alarm}] &= 1000 * \Pr(\theta \leq 0.15)
\end{aligned}
$$

(b) Similar for Beta

(c) Note that due to symmetry, the two Beta distributions have the same constants in front. We then have

$$
\begin{aligned}
p(\theta|y) \propto & p(\theta)p(y|\theta) \propto [\theta^{2-1}(1-\theta)^{8-1}\theta^y + \theta^{8-1}(1-\theta)^{2-1}](1-\theta)^{n-y} \\
= & \theta^{2+y-1}(1-\theta)^{8+n-y-1} + \theta^{8+y-1}(1-\theta)^{2+n-y-1}
\end{aligned}
$$

In order to find the normalizing constant, we need to integrate the right hand side. Note that in general $\int_\theta \theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Thus

$$
\begin{aligned}
&\int_\theta [\theta^{2+y-1}(1-\theta)^{8+n-y-1} + \theta^{8+y-1}(1-\theta)^{2+n-y-1}] \\
&\quad = \frac{\Gamma(2+y)\Gamma(8+n-y)}{\Gamma(10+n)} + \frac{\Gamma(8+y)\Gamma(2+n-y)}{\Gamma(10+n)}
\end{aligned}
$$

resulting in that

$$
\begin{aligned}
p(\theta|y) \propto & p(\theta)p(y|\theta) = \frac{\theta^{2+y-1}(1-\theta)^{8+n-y-1} + \theta^{8+y-1}(1-\theta)^{2+n-y-1}}{\frac{\Gamma(2+y)\Gamma(8+n-y)}{\Gamma(10+n)} + \frac{\Gamma(8+y)\Gamma(2+n-y)}{\Gamma(10+n)}} \\
= & \frac{\Gamma(2+y)(\Gamma(8+n-y)}{\Gamma(2+y)\Gamma(8+n-y)+\Gamma(8+y)\Gamma(2+n-y)} \frac{\Gamma(10+n)}{\Gamma(2+y)(\Gamma(8+n-y)} \theta^{2+y-1}(1-\theta)^{8+n-y-1} + \\
& \frac{\Gamma(8+y)(\Gamma(2+n-y)}{\Gamma(2+y)\Gamma(8+n-y)+\Gamma(8+y)\Gamma(2+n-y)} \frac{\Gamma(10+n)}{\Gamma(8+y)(\Gamma(2+n-y)} \theta^{8+y-1}(1-\theta)^{2+n-y-1} \\
= & w\text{Beta}(y; 2+y, 8+n-y) + (1-w)\text{Beta}(y; 8+y; 2+n-y)
\end{aligned}
$$

with

$$
w = \frac{\Gamma(2+y)(\Gamma(8+n-y)}{\Gamma(2+y)\Gamma(8+n-y)+\Gamma(8+y)\Gamma(2+n-y)}.
$$

An alternative way to do this calculation is to introduce an auxiliary variable $M \in \{1, 2\}$ such that $\Pr(M = 1) = 0.5$ and

$$
p(\theta|M) = \begin{cases} Beta(2, 8) & \text{if } M = 1 \\ Beta(8, 2) & \text{if } M = 2 \end{cases}
$$

Then

$$p(\theta|y) = p(\theta|y, M = 1) \times \Pr(M = 1) + p(\theta|y, M = 2) \times \Pr(M = 2)$$

Now $p(\theta|y, M = 1) = Beta(2 + y, 8 + n - y)$ and $p(\theta|y, M = 2) = Beta(8 + y, 2 + n - y)$. Further

$$\Pr(M = 1|y) \propto p(M = 1)p(y|M = 1)$$

Now $p(y|M = 1)$ is the marginal distribution for y under the prior $Beta(2, 8)$. In general for a prior $Beta(a, b)$ we have

$$\begin{aligned} p(y) =& \int_\theta \binom{n}{y} \theta^y (1-\theta)^{n-y} \tfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1 - theta)^{b-1} d\theta \\ =& \binom{n}{y} \tfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)} \int_\theta \tfrac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)} \theta^{a+y-1} (1-\theta)^{b+n-y-1} d\theta \\ =& \binom{n}{y} \tfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \tfrac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)} \end{aligned}$$

This then gives

$$\Pr(M = 1|y) \propto \binom{n}{y} \tfrac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \tfrac{\Gamma(2+y)\Gamma(8+n-y)}{\Gamma(10+n)}$$

and

$$\Pr(M = 2|y) \propto \binom{n}{y} \tfrac{\Gamma(10)}{\Gamma(8)\Gamma(2)} \tfrac{\Gamma(8+y)\Gamma(2+n-y)}{\Gamma(10+n)}$$

Since $\Pr(M = 1|y) + \Pr(M = 2|y) = 1$, we obtain

$$\Pr(M = 1|y) = \tfrac{\Gamma(2+y)\Gamma(8+n-y)}{\Gamma(2+y)\Gamma(8+n-y)+\Gamma(8+y)\Gamma(2+n-y)} = w.$$

Exercise 15

(a) We have

$$\Pr(X \leq x|\ X \text{ accepted}) = \frac{\Pr(X \leq x \text{ and } X \text{ accepted})}{\Pr(X \text{ accepted})}$$

Further,

$$\begin{aligned} \Pr(X \leq x \text{ and } X \text{ accepted}) &= \int_{-\infty}^{x} g(y)h(y)dy \\ \Pr(X \text{ accepted}) &= \int_{-\infty}^{\infty} g(y)h(y)dy \end{aligned}$$

giving

$$\Pr(X \leq x|\ X \text{ accepted}) = \frac{\int_{-\infty}^{x} g(y)h(y)dy}{\int_{-\infty}^{\infty} g(y)h(y)dy}$$

showing that the accepted value has density

$$f(x) = \frac{g(x)h(x)}{\int_{-\infty}^{\infty} g(y)h(y)dy}$$

(b) We have further

$$\begin{aligned}\frac{g(x)h(x)}{\int_{-\infty}^{\infty} g(y)h(y)dy} =& \frac{g(x)f(x)/\{Mg(x)\}}{\int_{-\infty}^{\infty} g(y)f(y)/\{Mg(y)\}dy} \\ =& \frac{f(x)}{\int_{-\infty}^{\infty} f(y)dy} = f(x)\end{aligned}$$

We have

$$\Pr(X \text{ accepted}) = \int_{-\infty}^{\infty} g(y)f(y)/\{Mg(y)\}dy = \int_{-\infty}^{\infty} f(y)/\{M\}dy = \frac{1}{M}$$

(c) For $a < 1$ we have $\lim_{x\to 0} x^{a-1} = \infty$ and similarly for $b < 1$.
For $a \geq 1, b \geq 1$, we have

$$\begin{aligned}\log f(x) =& \text{Const} + (a-1)\log(x) + (b-1)\log(1-x) \\ \frac{\partial}{\partial x}\log f(x) =& \frac{a-1}{x} - \frac{b-1}{1-x}\end{aligned}$$

which is equal to zero for

$$x^{opt} = \frac{a-1}{a+b-2}$$

which also is a maximum point since

$$\frac{\partial^2}{\partial x^2}\log f(x) = -\frac{a-1}{x^2} - \frac{b-1}{(1-x)^2} < 0$$

(d) See `Nils_15.R`

(e) Note first that from here the g function has a different role than above, now being the unnormalized target distribution, while h now is the proposal distribution.
We now have

$$\begin{aligned}\Pr(X \leq x \text{ and } X \text{ accepted}) &= \int_{-\infty}^{x} h(y)g(y)/\{Mh(y)\}dy = \frac{1}{M}\int_{-\infty}^{x} g(y)dy \\ \Pr(X \text{ accepted}) &= \int_{-\infty}^{\infty} h(y)g(y)/\{Mh(y)\}dy = \frac{1}{M}\int_{-\infty}^{\infty} g(y)dy\end{aligned}$$

giving

$$\Pr(X \leq x|\ X \text{ accepted}) = \frac{\int_{-\infty}^{x} g(y)dy}{\int_{-\infty}^{\infty} g(y)dy} = \frac{\int_{-\infty}^{x} f(y)dy}{\int_{-\infty}^{\infty} f(y)dy}$$

showing that the accepted value has density $f(x)$.

(f) Assume we choose a proposal distribution $h(x, y)$ being the uniform distribution over $[0, 1] \times [0, 1]$. Then

$$\frac{g(x,y)}{h(x,y)} = g(x,y) = \exp\{\sin\sqrt{|xy|})\exp(|y|^{3/2})\}$$

An upper limit of this ratio is $M = \exp(\sin(1))\exp(1)$.
See `Nils_15.R` for simulations.

(g) A possible proposal distribution is two independendent Gaussian variables, each with variance a. In that case

$$\begin{aligned}\frac{f(x,y)}{g(x,y)} =& \frac{\frac{1}{2\pi(1-\rho^2)^{1/2}}\exp\{-\frac{1}{2(1-\rho)^2}(x^2+y^2-2\rho xy)\}}{\frac{1}{2\pi a}\exp\{-\frac{1}{2a}(x^2+y^2)\}} \\ =& \frac{a}{(1-\rho^2)^{1/2}}\exp\{-\frac{1}{2(1-\rho)^2}(x^2+y^2-2\rho xy)+\frac{1}{2a}(x^2+y^2)\} \\ =& \frac{a}{(1-\rho^2)^{1/2}}\exp\{-\frac{1}{2(1-\rho)^2}(x^2+y^2-2\rho xy-\frac{1-\rho^2}{a}(x^2+y^2)\} \\ =& \frac{a}{(1-\rho^2)^{1/2}}\exp\{-\frac{1}{2(1-\rho)^2}(\tfrac{a-1+\rho^2}{a}x^2+\tfrac{a-1+\rho}{a}y^2-2\rho xy\}\end{aligned}$$

We now want to choose a such that the term within the exponent is always negative. This can be achieved by choosing a such that $\frac{a-1+\rho^2}{a}=\rho$ in which case

$$\tfrac{a-1+\rho^2}{a}x^2+\tfrac{a-1+\rho}{a}y^2-2\rho xy=\rho(x-y)^2$$

This gives $a=1+\rho$ giving

$$M=\frac{1+\rho}{(1-\rho)^{1/2}}=\sqrt{\frac{1+\rho}{1-\rho}}$$

See `Nils_15.R` for simulations.

Exercise 16

(a) Stationarity can be interpreted as if X_n follows the distribution π, then also X_{n+1} should follow the same distribution. Now

$$\Pr(X_{n+1}=j)=\sum_i \Pr(X_n=i)\Pr(X_{n+1}=j|X_n=i)$$

This then implies that we need

$$\pi_j=\sum_i \pi_i P_{i,j}$$

Under the assumption on P, we have

$$\sum_i \pi_i P_{i,j}=\sum_i \pi_j P_{j,i}=\pi_j\sum_i P_{j,i}=\pi_j$$

showing the result.

(b) We now have

$$\begin{aligned}\pi_i P_{i,j} =& \pi_i Q_{i,j}\min\left(1,\frac{\pi_j}{\pi_i}\right)=Q_{i,j}\min(\pi_i,\pi_j) \\ =& \pi_j Q_{j,i}\min\left(\frac{\pi_i}{\pi_j},1\right)=\pi_j P_{j,i}\end{aligned}$$

(c) We have now

$$\begin{aligned}\pi_i P_{i,j} =& \pi_i Q_{i,j}\min\left(1,S_{i,j}\frac{\pi_j}{\pi_i}\right)=\min(\pi_i Q_{i,j},Q_{i,j}S_{i,j}\pi_j) \\ =& \pi_j Q_{j,i}\min\left(\frac{\pi_i Q_{i,j}}{\pi_j Q_{j,i}},\frac{Q_{i,j}S_{i,j}}{Q_{j,i}}\right)\end{aligned}$$

Now, in order for the last factor to be equal to be one, we need $S_{i,j} = Q_{j,i}/Q_{i,j}$ in which case we obtain

$$\pi_i P_{i,j} = \pi_j Q_{j,i} \min\left(S_{j,i}\frac{\pi_i}{\pi_j}, 1\right) = \pi_j P_{j,i}$$

(*d*) $Q_{i,j}$ symmetric: A reasonable choice if we do not know much about the π distribution. Also simplifies calculations somewhat.

$Q_{i,j} = q_j$ means that the proposal will not depend on the previous value. However, the acceptance probability will still depend on the previous value, giving a Markov chain. In this case we need to make q similar to π in order to make it efficient.

Exercise 17

- We have

$$\begin{aligned} f(X_i))P(Y_i|X_i) =& f(X_i)q(Y_i|X_i)\min\left(1, \frac{q(X_i|Y_i)f(Y_i)}{q(Y_i|X_i)f(X_i)}\right) \\ =& \min\left(f(X_i)q(Y_i|X_i), q(Y_i|X_i)f(X_i)\right) \\ =& f(Y_i)q(X_i|Y_i)\min\left(\frac{q(Y_i|X_i)f(X_i)}{q(X_i|Y_i)f(Y_i)}, 1\right) = f(Y_i)P(X_i'Y_i) \end{aligned}$$

Exercise 19

(*a*) For λ positive, the distribution will give preference to similar x values. Since it is easiest to obtain similar values close to 4.5, we will give preference to values close to this.

For λ negative, the distribution will give preference to different x values. This will result in most values close to 0 or 9.

For $\lambda = 0$ there is no difference in any outcome, so each number will be independent.

(*b*) See `Nils_19.R`

(*c*) See `Nils_19.R`. We have for $\lambda = 0$ that each X_i are independent with $E[X] = 4.5$ and $\text{Var}[X] = 8.25$. We then have

$$E_0 \sum_{j=1}^{1} 2(X_j - \bar{X})^2 = 11 * V(X_j) = 11 * \sum_{k=0}^{9} 0.1 * (k - 4.5)^2 = 90.75$$

(*d*) Defining $\boldsymbol{x}^i = (x_1^i, ..., x_{12}^i)$ to be the telephone number i, we have

$$\begin{aligned} L(\lambda) =& \prod_{i=1}^{200} \frac{1}{Z(\lambda)} \exp\{-\lambda U(\boldsymbol{x}^i)\} = \frac{1}{Z(\lambda)^{200}} \exp\{-\lambda \sum_{i=1}^{200} U(\boldsymbol{x}^i)\} \\ l\lambda) =& -200 \log(Z(\lambda)) - \lambda \sum_{i=1}^{200} U(\boldsymbol{x}^i) \\ \frac{\partial}{\partial} l(\lambda) =& -200 \frac{Z'(\lambda}{Z(\lambda)} - \sum_{i=1}^{200} U(\boldsymbol{x}^i) \end{aligned}$$

and putting this to zero, we obtain

$$\frac{Z'(\lambda)}{Z(\lambda)} = -\frac{1}{200} \sum_{i=1}^{200} U(\boldsymbol{x}^i) = -11.11$$

Note however that

$$Z(\lambda) = \sum_{\boldsymbol{x}} \exp\{-\lambda U(\boldsymbol{x})\}$$
$$Z'(\lambda) = -\sum_{\boldsymbol{x}} \exp\{-\lambda U(\boldsymbol{x})\} U(\boldsymbol{x})$$

giving

$$\frac{Z'(\lambda)}{Z(\lambda)} = \frac{1}{Z(\lambda)} \sum_{\boldsymbol{x}} \exp\{-\lambda U(\boldsymbol{x})\} U(\boldsymbol{x}) = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) U(\boldsymbol{x})$$

which gives the required result.

(e) If the original 200 telephone numbers were available, one could apply some Bootstrap procedure for estimation of λ. Within a Bayesian setting, one chould include a prior for λ and then perform simulation from $p(\lambda|\boldsymbol{x})$. Note that in the latter case, the averate $\bar{U}$ will be a sufficient statistic so that we only need this quantity. Note however that we would need $Z(\lambda)$ in this case.

(f) We then only need $g_i(x_i|\text{rest}) = \exp\{-\lambda \sum_{i=1}^{12} (x_j - \bar{x})^2\}$.

See `Nils_19.R`