Solutions to exercises for STK4051 - Computational statistics

Geir Storvik

corrections made: 21.jan 2021

Solution to exercise 1.

(a). We find directly that $x^* = c^{1/m}$

(b). We have $g'(x) = mx^{m-1}$ giving

$$\begin{aligned}x_n =& x_n - \frac{g(x_{n-1})}{g'(x_{n-1})} = x_{n-1} - \frac{x_{n-1}^m - c}{mx_{n-1}^{m-1}}\\ =& x_{n-1} - \frac{x_{n-1}}{m} + \frac{cx_{n-1}}{mx_{n-1}^m} = x_{n-1}[1 - \frac{1}{m} + \frac{c}{mx_{n-1}^m}]\end{aligned}$$

(c). We have that

$$\begin{aligned}x[1 - \frac{1}{m} + \frac{c}{mx^m}] \geq c^{1/m}\\ \Updownarrow\\ [1 - \frac{1}{m} + \frac{c}{mx^m}] \geq \tfrac{c^{1/m}}{x} = \left(\tfrac{c}{x^m}\right)^{1/m}\\ \Updownarrow\\ 1 - \frac{1}{m} + \frac{y}{m} \geq y^{1/m} \qquad\qquad & y = \tfrac{c}{x^m}\\ \Updownarrow\\ 1 - \frac{1}{m} \geq y^{1/m} - \frac{y}{m}\end{aligned}$$

Denote $h(y) = y^{1/m} - \frac{y}{m}$. Then

$$\begin{aligned}h'(y) =& \tfrac{1}{m}y^{-\frac{m-1}{m}} - \frac{1}{m}\\ h''(x) =& -\tfrac{1}{m}\tfrac{m-1}{m}y^{-\frac{2m-1}{m}} < 0 \text{ for } y > 0\end{aligned}$$

Therefore $h(y)$ has a max point when $h'(y) = 0$ corresponding to $y = 1$ in which case $h(y) = 1 - \frac{1}{m}$. This shows that $1 - \frac{1}{m} \geq y^{1/m} - \frac{y}{m}$ which then implies that $x > c^{1/m}$

(d). We have

$$\begin{aligned} x_n =& x_{n-1}[1 - \frac{1}{m} + \frac{c}{m x_{n-1}^m}] \\ \leq& x_{n-1}[1 - \frac{1}{m} + \frac{c}{mc}] = x_{n-1} \end{aligned}$$

(e). This means that x_1 will always be larger than $c^{1/m}$ and that thereafter x_n will decrease monotonely towards $c^{1/m}$.

Solution to exercise 2.

(a). Consider first $f(x)$. Then

$$f'(x) = \begin{cases} \frac{1}{2\sqrt{x}} & x \geq 0 \\ \frac{1}{2\sqrt{-x}} & x < 0 \end{cases}$$

If $x_{n-1} > 0$ we get

$$x_n = x_{n-1} - \frac{\sqrt{x_{n-1}}}{\frac{1}{2\sqrt{x_{n-1}}}} = x_{n-1} - 2x_{n-1} = -x_{n-1}$$

while if $x_{n-1} > 0$ we get

$$x_n = x_{n-1} - \frac{\sqrt{-x_{n-1}}}{\frac{1}{2\sqrt{-x_{n-1}}}} = x_{n-1} - 2x_{n-1} = -x_{n-1}$$

so the algorithm will alternate between x_0 and $-x_0$.

(b). Consider now $g(x)$. Then

$$\begin{aligned} g'(x) =& \tfrac{1}{3} x^{-2/3} \\ x_n =& x_{n-1} - \frac{x_{n-1}^{1/3}}{\frac{1}{3} x_{n-1}^{-2/3}} \\ =& x_{n-1} - 3x_{n-1} = -2x_{n-1} \end{aligned}$$

so in this case the algorithm will diverge!

Solution to exercise 3.
Assume a function $G(x)$ which is contractive, that is

- $G(x) \in [a, b]$ whenever $x \in [a, b]$, and
- $|G(x_1) - G(x_2)| \leq \lambda |x_1 - x_2|$ for all $x_1, x_2 \in [a, b]$ for some $\lambda \in (0, 1)$

We will show that then there exist a unique fixed point x^* in the interval, and that the fixed-point algorithm will converge to it from any starting point in the interval. The fixed-point algorithm is given by

$$x_{k+1} = G(x_k)$$

(a). Follows directly by induction.

(b). We have

$$|x_2 - x_1| = |G(x_1) - G(x_0)| \leq \lambda |x_1 - x_0|$$

so the statement is true for $k = 1$. Assume now it is true for k. Then

$$\begin{aligned} |x_{k+2} - x_{k+1}| =& |G(x_{k+1}) - G(x_k| \\ \leq& \lambda |x_{k+1} - x_k| \\ \leq& \lambda \lambda^k |x_1 - x_0| = \lambda^{k+1} |x_1 - x_0| \end{aligned}$$

(c). We have

$$\begin{aligned} |x_m - x_n| =& |\sum_{k=n}^{m-1} (x_{k+1} - x_k)| \leq \sum_{k=n}^{m-1} |x_{k+1} - x_k| \\ \leq& \sum_{k=n}^{m-1} \lambda^k |x_1 - x_0| = |x_1 - x_0| \lambda^n \sum_{k=0}^{m-n-1} \lambda^k \\ =& |x_1 - x_0| \lambda^n \frac{1 - \lambda^{m-n}}{1 - \lambda} \leq |x_1 - x_0| \frac{\lambda^n}{1 - \lambda} \end{aligned}$$

(d).

$$\begin{aligned} |x_1 - x_0| \frac{\lambda^n}{1 - \lambda} <& \varepsilon \\ &\Updownarrow \\ \lambda^n <& \frac{(1 - \lambda)\varepsilon}{|x_1 - x_0|} \\ &\Updownarrow \\ n \log(\lambda) <& \log(1 - \lambda) + \log(\varepsilon) - \log(|x_1 - x_0|) \\ &\Updownarrow \\ n >& \frac{\log(1 - \lambda) + \log(\varepsilon) - \log(|x_1 - x_0|)}{\log(\lambda)} = N \end{aligned}$$

where the change of inequality sign is due to that $\log(\lambda) < 0$.

We then see that we are able to find a proper threshold N.

(*e*). We obtain the first inequality by defining $x_1 = x_\infty$ and $x_0 = y_\infty$. Since $\lambda < 1$ this shows that the fixed point is unique.

(*f*). (*i*) We both have that

$$\begin{aligned}|f(1.4) - f(1.3)| &= 0.17\\ |f(-0.4) - f(-0.3)| &= 0.17\end{aligned}$$

showing that the function is not contractive.

$$\begin{aligned}f(-\tfrac{1}{2}) &= -\tfrac{1}{2}\\ f(\tfrac{1}{2}) &= \tfrac{1}{2}\\ f(\tfrac{3}{2}) &= \tfrac{1}{2}\end{aligned}$$

showing that $f(x) \in (-\frac{1}{2}, \frac{3}{2})$ for $x \in (-\frac{1}{2}, \frac{3}{2})$.

(*ii*) Assume first $x_0 > 0.5$. Then $f'(x_0) < 0$ so the largest point we can get is when we start at 0.5 in which case $f(x) = 0.5$. This shows that $x_1 = f(x_0) < 0.5$

(*iii*) Assume now that $-0.5 < x_1 < 0.5$. Then $f'(x) > 0$ so we will always increase. However, again the maximum point we can get is 0.5 showing that that it will converge.

This shows that the contractive property is sufficient but not necessary for convergence of the fixed point algorithm.

(*iv*) If $x_0 = 1.5$, then $f(x_0) = -0.5$ and larger values of x_0 will result in smaller values of $f(x_0)$.

If $x_0 < -0.5$ then

$$-x^2 + x + \tfrac{1}{4} < -0.25 - 0.5 + 0.25 = -0.5$$

so it will stay in this interval. In these cases the fixed point algorithm will not converge.

Solution to exercise 4.

(*a*). Assume we have M. Then we put $\theta_j = 1$ for those $j \in M$ and zero for the rest, defining $\boldsymbol{\theta}$.

Assume we have $\boldsymbol{\theta}$. Then we include into all M those j's for which $\theta_j = 1$, defining M.

(*b*). We have:

- The sizes of the neighborhoods are p, $p(p-1)/2$, $p(p-1)/2$ and $p + p(p-1)/2$, respectively.

- The first one communicate, the second do not and the third one does not allow the number of "active" components to change and therefore do not communicate. The last one communicate since the first one does.
- For $\mathcal{N}_1$, the maximum number of moves is p. For $\mathcal{N}_4$, in order to move from $(00\cdots0)$ to $(11\cdots1)$ we need p moves (all in $\mathcal{N}_1$), giving also p necessary moves in this case.

(c). Steepest ascent: If we start on 000, all solutions within $\mathcal{N}_1$ have lower values, so 000 is a local mode that we are not able to escape from. However, for $\mathcal{N}_2$ we are able to move out of this mode.

Simulated annealing: If we just use a neighbourhood that communicate, all possibilities are available.

Genetic algorithms: As long as we include mutations that make all solutions communicate, the algorithms is able to find the solution.

Tabu algorithms: For $\mathcal{N}_1$ and as long as the memory is smaller than the size of the neighborhoud, this is ok.

Solution to exercise 5.

(a). Since $-\phi(\cdot)$ then is convex, the result follows directly

(b). Defining $y = F(x)$ we have that $dy = f(x)dx$ and

$$\begin{aligned}\int \phi(g(x))f(x)dx =& \int \phi(g(F^{-1}(y)))dy\\ \leq& \phi(\int g(F^{-1}(y))dy\\ =& \phi(\int g(x)f(x)dx\end{aligned}$$

This can be expressed as

$$\phi(E[X]) \geq E[\phi(X)].$$

(c). Define $Y = \log(X)$. Since the exponential function is convex, we have that

$$E[X] = E[\exp(Y)] \geq \exp(E[Y]) = \exp(E[\log(X)])$$

Alternatively, we have that since the log-function is concave,

$$\log(E[X]) \geq E[\log(X)]$$

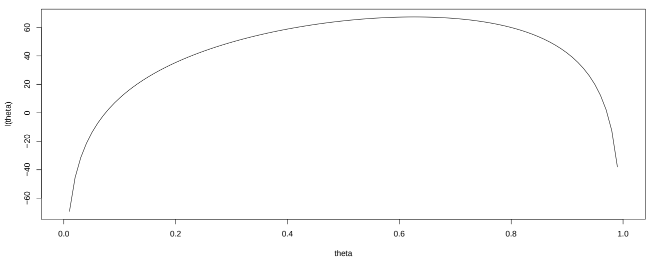
We have that $E(X) = \exp(\mu + 0.5\sigma^2) \geq \exp(\mu)$ confirming the result.

Solution to exercise 6.

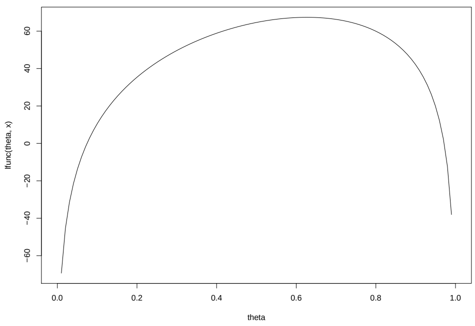
(a). Likelihood-function

$$L(\theta) = \binom{n}{x_1\, x_2\, x_3\, x_4} \frac{1}{4^n} \theta^{x_1} (1-\theta)^{x_2+x_3} (2+\theta)^{x_4}$$
$$\ell(\theta) = \text{Const} + x_1 \log(\theta) + (x_2 + x_3) \log(1-\theta) + x_4 \log(2+\theta)$$

For $\boldsymbol{x} = (34, 18, 20, 125)$:



(b). For $\boldsymbol{x} = (5, 0, 1, 14)$:



(c). We have

$$s(\theta) = \ell'(\theta) = \frac{x_1}{\theta} - \frac{x_2+x_3}{1-\theta} + \frac{x_4}{2+\theta}$$
$$J(\theta) = -\ell''(\theta) = \frac{x_1}{\theta^2} + \frac{x_2+x_3}{(1-\theta)^2} + \frac{x_4}{(2+\theta)^2}$$

defining the Newton-Raphson method. See the R-script *genetic_linkage.R* for implementation. Note that the trick of halving is needed for the second data example.

Define now $(x_1, x_2, x_3, y_4, y_5)$ to be the complete data, where $y_4 = x_4 - y_5$. The complete log-likelihood is given by

$$\begin{aligned}
\ell(\theta) =& \text{Const} + x_1\log(\theta) + (x_2+x_3)\log(1-\theta) + (x_4 - y_5)\log(2) + y_5\log(\theta) \\
Q(\theta|\theta^{(t)}) =& \text{Const} + x_1\log(\theta) + (x_2+x_3)\log(1-\theta) + x_4\log(2) \\
& E[Y_5|x_4, \theta^{(t)}[\log(\theta) - \log(2)] \\
Y_5|x_4 \sim& \text{Binom}(x_4, \tfrac{\theta}{2+\theta}) \\
Q(\theta|\theta^{(t)}) =& \text{Const} + x_1\log(\theta) + (x_2+x_3)\log(1-\theta) + x_4\log(2) + \\
& \frac{x_4\theta^{(t)}}{2+\theta^{(t)}}[\log(\theta) - \log(2)] \\
\frac{\partial}{\partial\theta}Q(\theta|\theta^{(t)}) =& \frac{x_1}{\theta} - \frac{x_2+x_3}{1-\theta} + \frac{x_4\theta^{(t)}}{2+\theta^{(t)}}\frac{1}{\theta}
\end{aligned}$$

giving

$$\theta^{(t+1)} = \frac{x_1 + x_4\frac{\theta^{(t)}}{2+\theta^{(t)}}}{x_1+x_2+x_3+x_4\frac{\theta^{(t)}}{2+\theta^{(t)}}}$$

See ***genetic_linkage.R*** for implementation.

Solution to exercise 7.
Since $\sum_k \pi_k = 1$, we need to introduce a Lagrange term:

$$\begin{aligned}
Q_{lagr}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) =& \sum_{i=1}^{n}\sum_{k=1}^{K}\Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})[\log(\pi_k) - \tfrac{1}{2}\log(\sigma_k^2) - \tfrac{1}{2\sigma_k^2}(x_i - \mu_k)^2] + \\
& \lambda(1 - \sum_{k=1}^{K}\pi_k) \\
\frac{\partial}{\partial\pi_k}Q_{lagr}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) =& \sum_{i=1}^{n}\Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})\pi_k^{-1} - \lambda
\end{aligned}$$

giving

$$\begin{aligned}
\pi_k^{(t+1)} =& \lambda^{-1}\sum_{i=1}^{n}\Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)}) \\
=& \frac{1}{n}\sum_{i=1}^{n}\Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})
\end{aligned}$$

Further

$$\frac{\partial}{\partial\mu_k}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{n}\Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})[\tfrac{1}{\sigma_k^2}(x_i - \mu_k)]$$

giving

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^n \Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)}) x_i}{\sum_{i=1}^n \Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})}$$

Similarly,

$$\frac{\partial}{\partial \sigma_k^2} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^n \Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})[-\tfrac{1}{2\sigma_k^2} + \tfrac{1}{2\sigma_k^4}(x_i - \mu_k)^2]$$

giving

$$(\sigma_k^2)^{(t+1)} = \frac{\sum_{i=1}^n \Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})(x_i - \mu_k^{(t+1)})^2}{\sum_{i=1}^n \Pr(C_i = k|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})}$$

Solution to exercise 8.

Solution to exercise 9.

(a). We have

$$\infty = \sum_{t=1}^{\infty} a_t = \sum_{t=1}^{T} a_t + \sum_{t=T+1}^{\infty} a_t$$

Since the first sum on the right hand side is finite, the second has to be infinite.

(b). Assume $\sum_{t=1}^{\infty} \alpha_t < \infty$. Then

$$\sum_{t=2}^{\infty} \frac{\alpha_t}{\alpha_1 + \cdots + \alpha_{t-1}} \leq \sum_{t=2}^{\infty} \frac{\alpha_t}{\alpha_1} = \frac{1}{\alpha_1} \sum_{t=2}^{\infty} \alpha_t < \infty.$$

giving a contradiction.

(c). Assume $\lim_{t\to\infty} b_t = \delta > 0$. Then there exists some $0 < \varepsilon < \delta$ and T such that $b_t > \varepsilon$ for $t > T$. Then

$$\sum_{t=1}^{\infty} a_t b_t = \sum_{t=1}^{T} a_t b_t + \sum_{t=T+1}^{\infty} a_t b_t \geq \sum_{t=1}^{T} a_t b_t + \varepsilon \sum_{t=T+1}^{\infty} a_t = \infty$$

giving a contradiction.

(d). We have that

$$\begin{aligned}\theta^t =&\theta^{t-1} - \alpha_{t-1} Z(\theta^{t-1}, \xi^{t-1})\\ =&\theta^{t-2} - \alpha_{t-2} Z(\theta^{t-2}, \xi^{t-2}) - \alpha_{t-1} Z(\theta^{t-1}, \xi^{t-1})\\ &\vdots\\ =&\theta^1 - \sum_{s=1}^{t-1} \alpha_s Z(\theta^s, \xi^s)\end{aligned}$$

giving

$$\begin{aligned}|\theta^t - \theta^*| =&|\theta^1 - \sum_{s=1}^{t-1} \alpha_s Z(\theta^s, \xi^s) - \theta^*| \leq |\theta^1 - \theta^*| + \sum_{s=1}^{t-1} \alpha_s |Z(\theta^s, \xi^s)|\\ \leq&|\theta^1 - \theta^*| + \sum_{s=1}^{t-1} \alpha_s C = A_t\end{aligned}$$

Note also that from the above we obtain

$$|\theta^t - \theta^1| \leq C \sum_{s=1}^{t-1} \alpha_s$$

In order not to constrain the possibility of moving θ^t, we want $\sum_{s=1}^{\infty} \alpha_s = \infty$.

Further, if all the $Z(\theta^s, \xi^s)$ are independent with identical variance σ^2, we have

$$\text{Var}[\theta^t - \theta^1] = \sigma^2 \sum_{s=1}^{t-1} \alpha_s^2.$$

In order for this variance not to explode, we would like $\sum_{s=1}^{\infty} \alpha_s^2 < \infty$.

The last argument is not complete since the $Z(\theta^s, \xi^s)$ are clearly dependent (they depend on the previous θ^s values which again depend on the previous Z's), it still give some heuristic argument for why the constraint $\sum_{s=1}^{\infty} \alpha_s^2 < \infty$ is reasonable to assume.

Solution to exercise 10.

(a). We have

$$\begin{aligned}E[s(\theta; X)] =& \int_x [\tfrac{\partial}{\partial \theta} \log f(x; \theta)] f(x; \theta) dx\\ =& \int_x \tfrac{\partial}{\partial \theta} f(x; \theta) dx\\ =&\tfrac{\partial}{\partial \theta} \int_x f(x; \theta) dx\\ =&\tfrac{\partial}{\partial \theta} 1 = 0\end{aligned}$$

(b). We have

$$\begin{aligned}
E[l''(\theta;X)] =& \int_x [\tfrac{\partial^2}{\partial\theta^2}\log f(x;\theta)] f(x;\theta)dx \\
=& \int_x [\tfrac{\partial}{\partial\theta}\tfrac{f'(x;\theta)}{f(x;\theta)}] f(x;\theta)dx \\
=& \int_x \tfrac{f''(x;\theta)f(x;\theta)-f'(x;\theta)f'(x;\theta)}{[f(x;\theta)]^2} f(x;\theta)dx \\
=& \int_x f''(x;\theta)dx - \int_x \tfrac{f'(x;\theta)f'(x;\theta)}{[f(x;\theta)]^2} f(x;\theta)dx \\
=& \tfrac{\partial^2}{\partial\theta^2}\int_x f(x;\theta)dx - \int_x [\tfrac{\partial}{\partial\theta}\log f(x;\theta)]^2 f(x;\theta)dx \\
=& -E[s(\theta;X)^2]
\end{aligned}$$

Therefore

$$\begin{aligned}
\text{Var}[s(\theta;X)] =& E[s(\theta;X)^2] - [E[s(\theta;X)]]^2 \\
=& -E[l''(\theta;X)] - 0 = I(\theta)
\end{aligned}$$

(c).

$$\begin{aligned}
I_n(\theta) =& -E[\sum_{i=1}^n \tfrac{\partial^2}{\partial\theta^2} l(\theta;X_i)] = \sum_{i=1}^n -E[\tfrac{\partial^2}{\partial\theta^2} l(\theta;X_i)] \\
=& nI(\theta) = n\text{Var}[s(\theta;X)]
\end{aligned}$$

(d). Now

$$\begin{aligned}
\ell(\theta;x) =& \log[\Gamma(0.5(\nu+1))] - 0.5\sqrt{\nu\pi} - \log[\Gamma(0.5\nu)] - 0.5(\nu+1)\log(1+v^{-1}(x-\theta)^2) \\
s(\theta;x) =& -0.5(\nu+1)\frac{-2v^{-1}(x-\theta)}{1+v^{-1}(x-\theta)^2} \\
=& (\nu+1)\frac{x-\theta}{\nu+(x-\theta)^2} \\
J(\theta;x) =& -(\nu+1)\frac{-\nu-(x-\theta)^2+2(x-\theta)^2}{[\nu+(x-\theta)^2]^2} \\
=& (\nu+1)\frac{\nu-(x-\theta)^2}{[\nu+(x-\theta)^2]^2}
\end{aligned}$$

See ***Extra_8.R***

(e). See ***Extra_8.R***

(f). The theoretical results are only valid for $\theta = \theta^*$!

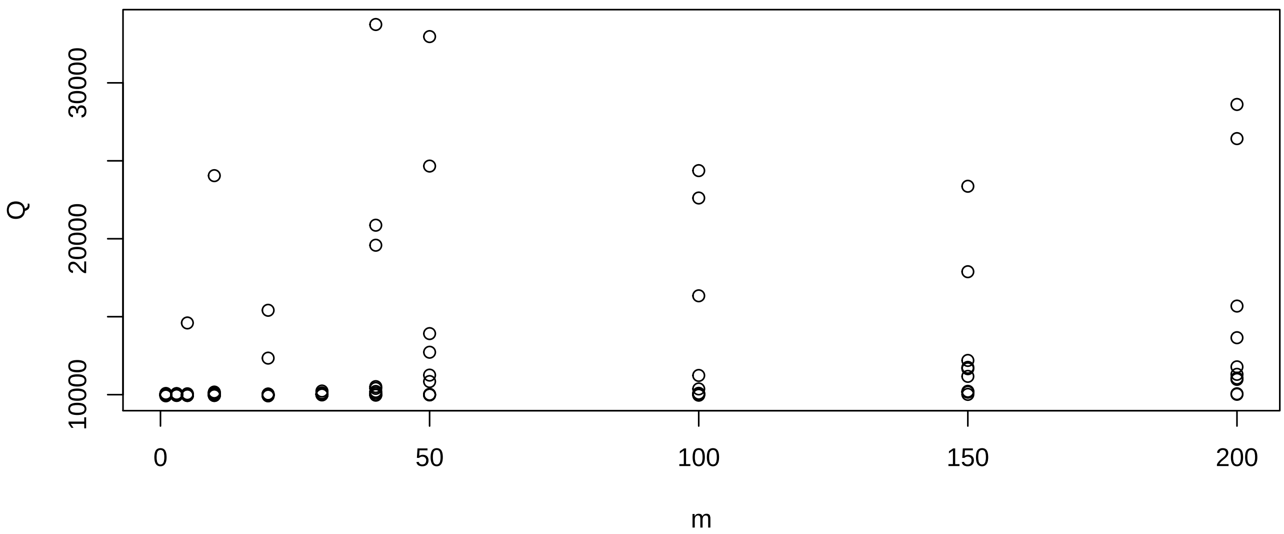
Solution to exercise 11.

(a). 10 repetitions with random initial points on the α's and linear regression estimates on the β's given the α's of the algorithm with $n = 10\,000$ gave $Q - values$ ranging from 10025.54 to 10368.43:

```
10217.33 10144.09 10105.26 10368.43 10097.87 10188.39
10120.00 10025.54 10079.28 10200.39
```

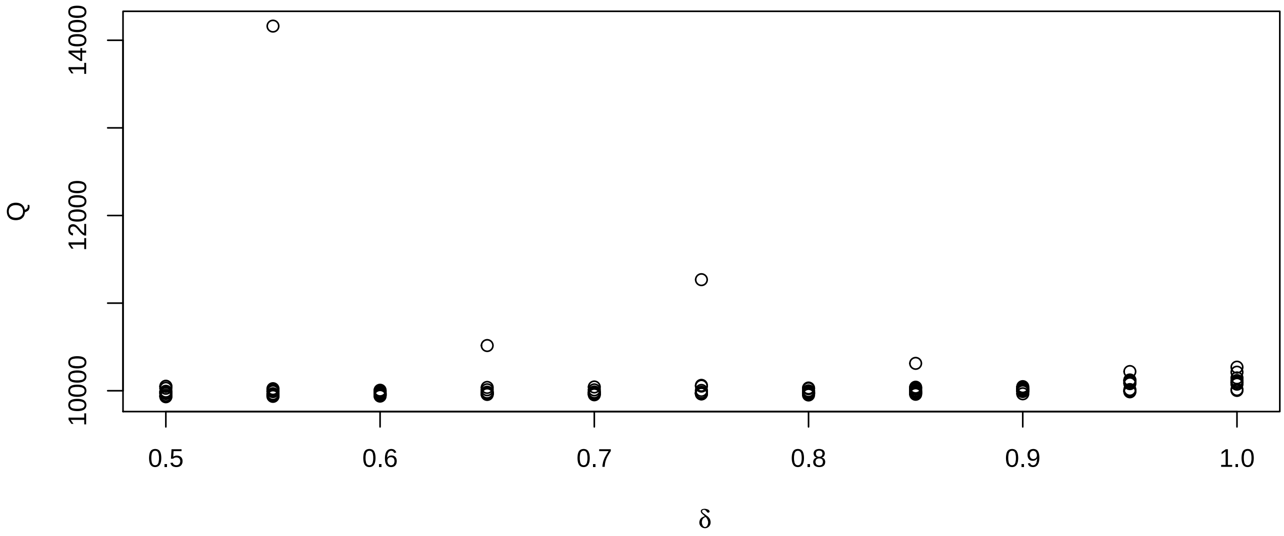
This indicates that the algorithms is quite sensitive to starting values.

(b). Using $m = 1, 3, 5, 10, 20, 30, 40, 50, 100, 150, 200$, repeating the procedure 10 times for each value of m gave the following results:



The minimum value 9925.46 was obtained for $m = 20$ while for $m = 1$ we obtained 9925.61. It seems like the procedure is working best for small m.

(c). Using $\delta = 0.500.550.600.650.700.750.800.850.900.951.00$ and $m = 1$, repeating the procedure 10 times for each value of δ gave the following results:



The smallest value obtained was 9931.51 for $\delta = 0.5$. It is not very clear what is preferable here, but top large values should be avoided.

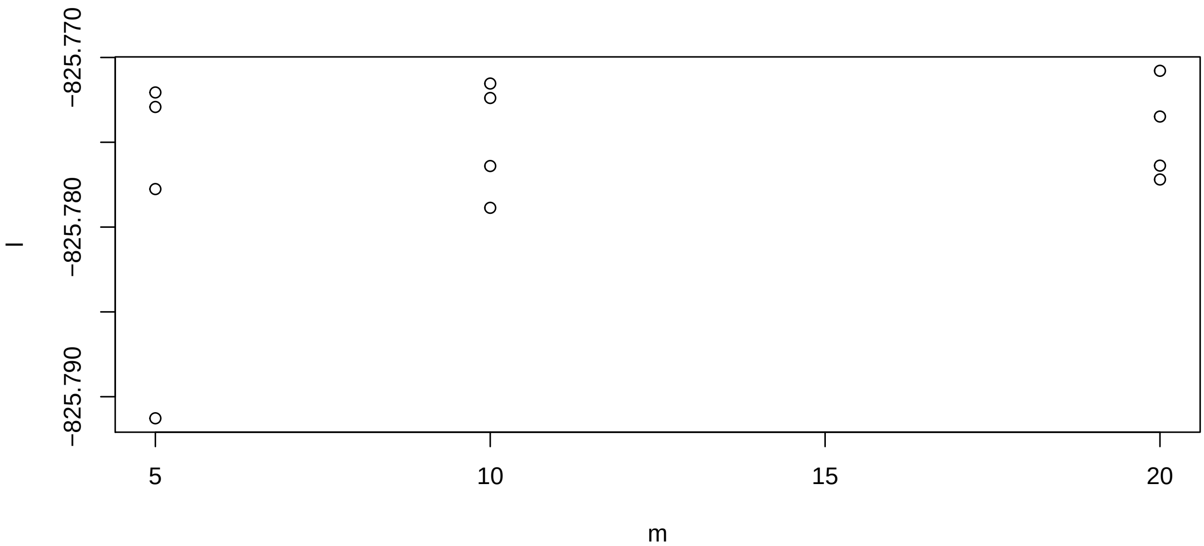
(d). Running now with $m = 1$ and $\delta = 0.5$ 22 times gave the following values of Q:

10041.637	9955.682	10016.338	9950.949	9948.319	9960.379
9941.765	9938.026	9956.522	9986.015	9938.413	9972.349
10025.531	10078.068	9926.557	10004.114	9947.645	9941.453
9964.286	9967.233				

with the smallest value equal to 9926.557. Note that this is smaller than the values obtained by using LS estimates on β!

Solution to exercise 12.

(a). $m = 5, 10, 20$), 10 000 iterations, 4 repetitions gave



The best value was -825.79.

10 repetitions with only 1000 iterations gave best value -825.81!

(b). `geoR` gave log-likelihood of -791.7.

(c). Struggled even with $n = 1000$.

(d).

Solution to exercise 13.

(a). We have

$$\begin{aligned}\Pr(X \le x) &= \Pr(F^{-1}(U) \le x) \\ &= \Pr(U \le F(x)) = F(x)\end{aligned}$$

showing that X has cumulative distribution function $F(x)$.

(b). We have

$$\begin{aligned}X =& F^{-1}(U)\\ &\Updownarrow\\ F(X) =& U\\ &\Updownarrow\\ \frac{1}{\pi}\tan^{-1}\left(\tfrac{X-x_0}{\gamma}\right)+\tfrac{1}{2} =& U\\ &\Updownarrow\\ X =& x_0+\gamma\tan(\pi[U-\tfrac{1}{2}])\end{aligned}$$

Solution to exercise 14.

(a). We have

$$\begin{aligned}\exp(-0.5(X_1^2+X_2^2)) &= \exp(-0.5(-2\log(U_1)\cos^2(2\pi U_2)-2\log(U_1)\sin^2(2\pi U_2)))\\ &= \exp(-0.5(-2\log(U_1))) = U_1\\ \frac{1}{2\pi}\tan^{-1}(X_2/X_1) =& \frac{1}{2\pi}\tan^{-1}(\sin(2\pi U_2)/\cos(2\pi U_2))\\ =& \frac{1}{2\pi}\tan^{-1}(\tan(2\pi U_2)) = U_2\end{aligned}$$

(b). We have that

$$\begin{aligned}f_X(\boldsymbol{x}) =& f_U(\boldsymbol{g}^{-1}(\boldsymbol{x}))\left|\frac{\partial}{\partial\boldsymbol{x}}g^{-1}(\boldsymbol{x})\right|\\ =& \begin{vmatrix}-X_1\exp(-0.5(X_1^2+X_2^2)) & -X_2\exp(-0.5(X_1^2+X_2^2))\\ \frac{1}{2\pi}\frac{-X_2/X_1^2}{1+X_2^2/X_1^2} & \frac{1}{2\pi}\frac{1/X_1}{1+X_2^2/X_1^2}\end{vmatrix}\\ =& \frac{1}{2\pi}\exp(-0.5(X_1^2+X_2^2))\begin{vmatrix}-X_1 & -X_2\\ \frac{-X_2}{X_1^2+X_2^2} & \frac{X_1}{X_1^2+X_2^2}\end{vmatrix}\\ =& \frac{1}{\sqrt{2\pi}}\exp(-0.5X_1^2))\frac{1}{\sqrt{2\pi}}\exp(-0.5X_2^2))\end{aligned}$$

which is the product of two standard Gaussian densities.

(c). Use that $X=\mu+\sigma Z$ where Z is standard Gaussian.

(d). Assume $\boldsymbol{Z}$ is a vector of independent standard Gaussian variables. Let $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})^T$ (e.g. the Cholesky decomposition). Then define $X=\boldsymbol{\mu}+\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}$. The vector is Gaussian since it is a linear combination of Gaussians, and

$$\begin{aligned}E[X] =& E[\boldsymbol{\mu}+\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}] = 0\\ \mathrm{Var}[X] =& \mathrm{Var}[\boldsymbol{\mu}+\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}] = \boldsymbol{\Sigma}^{1/2}\boldsymbol{I}(\boldsymbol{\Sigma}^{1/2})^T = \boldsymbol{\Sigma}\end{aligned}$$

Solution to exercise 16.

(a).

(b).

(c).

(d). Define $Z = X + 2$ were $X \sim N(0, 1)$. Then

$$\begin{aligned} E[Z|Z > 3] =& E[X + 2|X + 2 > 3] = 2 + E[X|X > 1] \\ =& \int_1^\infty x \frac{\frac{1}{\sqrt{2\pi}} e^{-0.5x^2}}{\Pr(X > 1)} dx \\ =& 2 + \frac{[-\frac{1}{\sqrt{2\pi}} e^{-0.5x^2}]_1^\infty}{1 - \Pr(X \leq 1)} = 2 + \frac{\frac{1}{\sqrt{2\pi}} e^{-0.5}}{1 - \Pr(X \leq 1)} \end{aligned}$$

(e). For exponential we have the forgetting property so that given that $Z > 3$ we still have that it is exponential, so that

$$E[Z|Z > 3] = 3 + E[Z_0] = 3 + 0.5$$

Solution to exercise 17.

(a). The significance level is given by

$$\begin{aligned} \alpha =& \Pr(\text{Reject hypothesis}|\text{Hypothesis is true}) \\ =& \Pr(T > c_0|\mu = 0) = \qquad [I(T > c_0)|\mu = 0] \approx \frac{1}{N} \sum_{j=1}^N T_j \end{aligned}$$

were T_j is obtained by simulating n $x_i's$ from $N(0, \sigma)$ and then calculating $T = \sqrt{n}\bar{x}/s$.

(b).

(c).

(d).

(e). True significance level closer to α when n is larger. Due to the central limit theorem! This means that the t confidence intervals will be quite accurate also when n is large.

Solution to exercise 18.

(a). We have that

$$\begin{aligned}
G_i =& \int_{z_{i-1}}^{z_i} g(x)dx \\
=& c\int_{z_{i-1}}^{z_i} \exp(a_i + b_i x)dx \\
=& c\frac{1}{b_i}[\exp(a_i + b_i z_i) - \exp(a_i + b_i z_{i-1})] \\
=& c\frac{1}{b_i}\exp(a_i)[\exp(b_i z_i) - \exp(b_i z_{i-1})].
\end{aligned}$$

were one of the terms inside the last parentes will be zero for the first and the last interval.

Since g is a density, we have

$$c^{-1} = \sum_{i=1}^{k+1} G_i.$$

(b). For $x \in (z_{j-1}, z_j]$ the result

$$G(x) = \sum_{i=1}^{j-1} G_i + \frac{c}{b_j}\exp(a_j)[\exp(b_j x) - \exp(b_j z_{j-1})]$$

follows directly from the earlier derivations.

(c). For $\sum_{i=1}^{j-1} G_i < u \le \sum_{i=1}^{j} G_i$ we have

$$\sum_{i=1}^{j-1} G_i + \frac{c}{b_j}\exp(a_j)[\exp(b_j x) - \exp(b_j z_{j-1})] = u$$

or

$$\frac{c}{b_j}\exp(a_j)[\exp(b_j x) - \exp(b_j z_{j-1})] = u - \sum_{i=1}^{j-1} G_i$$

or

$$\exp(b_j x) - \exp(b_j z_{j-1}) = \frac{b_j}{c}\exp(-a_j)(u - \sum_{i=1}^{j-1} G_i)$$

giving

$$\begin{aligned}
x =& \frac{1}{b_j}\log[\exp(b_j z_{j-1}) + \frac{b_j}{c}\exp(-a_j)(u - \sum_{i=1}^{j-1} G_i)] \\
=& z_{j-1} + \frac{1}{b_j}\log[1 + \frac{b_j}{c}\exp(-a_j - b_j z_{j-1})(u - \sum_{i=1}^{j-1} G_i)]
\end{aligned}$$

For $u = \sum_{i=1}^{j-1} G_i$ we obtain z_{j-1} and for $u = \sum_{i=1}^{j} G_i$ we obtain z_j. Since the function is monotone in u it then has to be within the interval $(z_{j-1}, z_j]$.

(*d*). The main part in the ARS algorithm is to simulate from a piecewise linear density. To perform this we need the following steps:

(*i*) Generate $u \sim \text{Unif}[0,1]$.

(*ii*) Find j such that $\sum_{i=1}^{j-1} G_i < u \le \sum_{i=1}^{j} G_i$

(*iii*) Put $x = z_{j-1} + \frac{1}{b_j}\log[1 + \frac{b_j}{c}\exp(-a_j - b_j z_{j-1})(u - \sum_{i=1}^{j-1} G_i)]$

We need to precompute the terms $\sum_{i=1}^{j} G_i$. The search for j can be time-consuming if k is large. We will however see later that there are efficient algorithms for that as well!

Solution to exercise 20. (*a*). We have

$$p(x|z) = \frac{p(x)p(z|x)}{p(z)} \propto \frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}x^2)f(z|x) \propto \exp(-\tfrac{1}{2}x^2)f(z|x)$$

(*b*). We need that there exist an $\tilde{\alpha}$ such that

$$\frac{\exp(-\frac{1}{2}x^2)f(z|x)}{\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)} \le M$$

or

$$\sqrt{2\pi}f(z|x) \le M$$

which is ok due to the constraint on $f(z|x)$. In this case we can sample from $N(0,1)$ an accept with probability $\sqrt{2\pi}f(z|x)/M$

(*c*). For $Z|X = x \sim N(ax, \sigma^2)$, we have

$$\sqrt{2\pi}f(z|x) = \frac{1}{\sigma}\exp(-\tfrac{1}{2\sigma^2}(z - ax)^2) \le \tfrac{1}{\sigma}$$

so given that we have simulated a value x, we will accept with probability $\exp(-\frac{1}{2\sigma^2}(z - ax)^2)$. Now the expected probability of accepting is:

$$
\begin{aligned}
&E^X[\exp(-\tfrac{1}{2\sigma^2}(z-aX)^2)]\\
&\quad=\int \exp(-\tfrac{1}{2\sigma^2}(z-ax)^2\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}x^2)dx\\
&\quad=\frac{1}{\sqrt{2\pi}}\int \exp(-\tfrac{1}{2}[\frac{z^2}{\sigma^2}-2\frac{az}{\sigma^2}x+\tfrac{a^2}{\sigma^2}x^2+x^2])dx\\
&\quad=\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{z^2}{2\sigma^2})\int \exp(-\tfrac{1}{2}[\tfrac{a^2+\sigma^2}{\sigma^2}x^2-2\tfrac{az}{\sigma^2}x])dx\\
&\quad=\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{z^2}{2\sigma^2})\int \exp(-\tfrac{a^2+\sigma^2}{2\sigma^2}[x^2-2\tfrac{az}{a^2+\sigma^2}x])dx\\
&\quad=\exp(-\tfrac{z^2}{2\sigma^2})\exp(\tfrac{a^2+\sigma^2}{2\sigma^2}(\tfrac{az}{a^2+\sigma^2})^2)\\
&\qquad\frac{1}{\sqrt{2\pi}}\int \exp(-\tfrac{a^2+\sigma^2}{2\sigma^2}[(x-\tfrac{az}{a^2+\sigma^2})^2])dx\\
&\quad=\exp(-\tfrac{1}{2\sigma^2}z^2)\exp(\tfrac{1}{2\sigma^2}(\tfrac{a^2z^2}{a^2+\sigma^2}))\tfrac{\sigma}{\sqrt{a^2+\sigma^2}}\\
&\quad=\exp(-\tfrac{z^2}{2(a^2+\sigma^2)})\tfrac{\sigma}{\sqrt{a^2+\sigma^2}}
\end{aligned}
$$

so it will decrease with z deviating from zero, which is reasonable since we use a proposal distribution that is centered around zero. On average the number of samples needed to get 1 sample is given as is $1/P(accept)$. If we want n samples, the number of samples we need to porpose is :

$$
n\cdot\exp(\tfrac{z^2}{2(a^2+\sigma^2)})\tfrac{\sqrt{a^2+\sigma^2}}{\sigma}
$$

Solution to exercise 22. (a). We have

$$
J=E[(X+a)^2]=E[X^2+2aX+a^2]=1+a^2
$$

(b). We have with M simulations

$$
\begin{aligned}
M\text{Var}[\widehat{J}]=&E[(X+a)^4]-[E[(X+a)^2]]^2\\
=&E[X^4+4aX^3+6a^2X^2+4a^3X+a^4]-[E[X^2+2aX+a^2]]^2\\
=&3+0+6a^2+0+a^4-[1+0+a^2]^2\\
=&3+6a^2+a^4-1-2a^2-a^4=2+4a^2
\end{aligned}
$$

(c). We have

$$
J=\int_{-\infty}^{\infty}(x+a)^2\phi(x)\,dx=\qquad\int_{-\infty}^{\infty}(x+a)^2\frac{\phi(x)}{\phi(x-a)}\phi(x-a)\,dx
$$

An alternative would be to utilize that

$$
\begin{aligned}
J &= \int_{-\infty}^{\infty} (x+a)^2 \phi(x)\, dx \\
&= \int_{-\infty}^{\infty} y^2 \phi(y-a)\, dy \\
&\approx \frac{1}{M} \sum_{i=1}^{M} y_i^2
\end{aligned}
$$

where now $y_i \sim N(a, 1)$.

Solution to exercise 23.
Since each $\theta_i^* \sim p(\theta|\boldsymbol{x})$, we have that $[E\theta_i^*] = E[\theta|\boldsymbol{x}] = \hat{\theta}$ and $\text{Var}[\theta_i^*] = E[(\theta - \hat{\theta})^2|\boldsymbol{x}] = \text{Var}[\theta|\boldsymbol{x}]$ which results in that

$$
\begin{aligned}
E[\overline{\theta}^*|\boldsymbol{x}] =& \hat{\theta}; \\
E[(\overline{\theta}^* - \theta)^2|\boldsymbol{x}] =& E[(\overline{\theta}^* - \hat{\theta} + \hat{\theta} - \theta)^2|\boldsymbol{x}] \\
=& E[(\overline{\theta}^* - \hat{\theta})^2|\boldsymbol{x}] + E[(\hat{\theta} - \theta)^2|\boldsymbol{x}] + 2E[(\overline{\theta}^* - \hat{\theta})(\hat{\theta} - \theta)|\boldsymbol{x}] \\
=& \frac{1}{M}\text{Var}[\theta_i^*] + E[(\hat{\theta} - \theta)^2|\boldsymbol{x}] \\
=& (1 + \tfrac{1}{m})E[(\hat{\theta} - \theta)^2|\boldsymbol{x}] \\
=& (1 + \tfrac{1}{m})\text{Var}[\theta|\boldsymbol{x}]
\end{aligned}
$$

This shows that the main variability comes from the data, the extra variability due to Monte Carlo estimation of $\hat{\theta}$ is small for m of reasonable size.

Solution to exercise 24.
The aim is to find $X = r$ such that $P_{r-1} < U \leq P_r$ as efficient as possible.

(a). We have that since $g_X = \max\{j|P_j < \text{int}(MU+1)/M\}$ and $Y = g_X + 1$ that $P_Y \geq \text{int}(MU+1)/M > U$. We therefore have that $Y > r$ which means that we have to search backwards. By decreasing Y as long as $P_{Y-1} > U$ we obtain the right $X = r$.

(b). Assume now $U \in (\frac{i-1}{M}, \frac{i}{M})$. Then $P_Y > \frac{i}{M}$. Define N_i to be the number of $P_j \in (\frac{i-1}{M}, \frac{i}{M})$. Now there are two options.

- Either $N_i = 0$ in which case $P_{Y-1} < \frac{i-1}{M} < U$ and we are done.
- Or $N_i > 0$ in which case a maximum $1 + N_i$ iterations are needed.

Then, if N is the number of iterations, we have

$$\begin{aligned}E[N] =& \sum_{i=1}^{M} E[N|U \in (\tfrac{i-1}{M}, \tfrac{i}{M})] \Pr(U \in (\tfrac{i-1}{M}, \tfrac{i}{M}))\\ \leq& \sum_{i=1}^{M}(1+N_i)\tfrac{1}{M} = 1 + \tfrac{1}{M}\sum_{i=1}^{M} N_i = 2\end{aligned}$$

where we have used that there are in total M P_j's and therefore $\sum_{i=1}^{M} N_i = M$.

(c). The main part of the ARS algorithm is to find the interval for which $\sum_{i=1}^{j-1} < u \leq \sum_{i=1}^{j} G_i$ which corresponds to finding r in this exercise.

Solution to exercise 26. (a). Assume $x_{t-1} \sim N(\mu, \sigma^2/(1-a^2))$. Then

$$x_t = \mu + a(x_{t-1} - \mu) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

were ε_t is independent of x_{t-1}. Then x_t is a lineary combination of Gaussian variables, making itself Gaussian. Further

$$\begin{aligned}E[x_t] =& \mu + a(E[x_{t-1} - \mu) + E[\varepsilon_t] = \mu\\ \text{Var}[x_t] =& a^2\text{Var}[x_{t-1}] + \text{Var}[\varepsilon_t] = a^2\tfrac{\sigma^2}{1-a^2} + \sigma^2 = \tfrac{\sigma^2}{1-a^2}\end{aligned}$$

(b). We have

$$\begin{aligned}p(\boldsymbol{x}_{1:T}|\boldsymbol{y}_{1:T}) =& \frac{p(\boldsymbol{x}_{1:T})p(\boldsymbol{y}_{1:T}|\boldsymbol{x}_{1:T})}{p(\boldsymbol{y}_{1:T})}\\ \propto& p(x_1)p(y_1|x_1)\prod_{t=2}^{T} p(x_t|x_{t-1})p(y_t|x_t)\end{aligned}$$

were each of the densities involved are specified through the given model.

(c). We have then that $g(\boldsymbol{x}_{1:t}) = p(\boldsymbol{x}_{1:t})$ so that

$$w_t(\boldsymbol{x}_{1:t}) = \frac{p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})}{g(\boldsymbol{x}_{1:t})} \propto p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_{1:t}) = p(\boldsymbol{y}_{1:t-1}|\boldsymbol{x}_{1:t-1})p(y_t|x_t) \propto w_{t-1}(\boldsymbol{x}_{1:t-1})p(y_t|x_t)$$

We only need the weights up to a proportionality constant (since we will normalize them anyway), showing the result with $u_t(y_1, x_t) = p(y_t|x_t)$.

(d).

(*e*). The importance weights we calculate is based on the whole sequence $\boldsymbol{x}_{1:t}$. Therefore the samples $(\boldsymbol{x}^i_{1:t}, w^i_t)$ are properly weighted with respect to $p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})$. When $t = T$, we then obtain properly weighted samples from $p(\boldsymbol{x}_{1:T}|\boldsymbol{y}_{1:T})$.

When we perform resampling in the algorithm, note that we then need to resample the whole sequence $\boldsymbol{x}_{1:t}$.

A problem when looking at these "smoothed" estimates is that for t small, the number of unique samples is very low. Imn this case, were T is not too large and the number of samples N is large enough, we do however still get reasonable estmates and uncertainty measures even for x_1.

Solution to exercise 27. (*a*). We have

$$\begin{aligned}
p(x|y) \propto& p(x)p(y|x)\\
\propto& \exp(-\tfrac{1}{2\sigma^2}(x-\mu)^2)\tfrac{\exp(x)^y\exp(-\exp(x))}{y!}\\
\propto& \exp(-\tfrac{1}{2\sigma^2}(x^2-2\mu x)+yx-\exp(x))
\end{aligned}$$

(*b*). We have that this approximation corresponds to a Gaussian approximation to $p(z|y)$.

From the prior we have that x should not be too far from μ. The approximation of $\exp(x)$ corresponds to a Taylor approximation for $\exp(\mu)$ around μ. We then get

$$\begin{aligned}
\log p(x|y) \approx& \text{Const} - \tfrac{1}{2\sigma^2}(x^2-2\mu x)+yx-e^\mu-e^\mu(x-\mu)-\tfrac{1}{2}e^\mu(x-\mu)^2\\
=& \text{Const} - (\tfrac{1}{2\sigma^2}+\tfrac{1}{2}e^\mu)x^2+(\tfrac{\mu}{\sigma^2}+y-e^\mu+\mu e^\mu)x\\
=& \text{Const} - (\tfrac{1}{2\sigma^2}+\tfrac{1}{2}e^\mu)[x-\frac{\frac{\mu}{\sigma^2}+y-e^\mu+\mu e^\mu}{\frac{1}{\sigma^2}+e^\mu}]^2\\
=& \text{Const} - \frac{1}{2\tilde{\sigma}^2}(x-\tilde{\mu})^2
\end{aligned}$$

with

$$\begin{aligned}
\tilde{\sigma}^2 =& \frac{1}{\frac{1}{\sigma^2}+e^\mu} = \frac{\sigma^2}{1+\sigma^2e^\mu}\\
\tilde{\mu} =& \frac{\frac{\mu}{\sigma^2}+y-e^\mu+\mu e^\mu}{\frac{1}{\sigma^2}+e^\mu} = \frac{\mu+\sigma^2(y-e^\mu+\mu e^\mu)}{1+\sigma^2e^\mu}
\end{aligned}$$

Solution to exercise 30. (*a*). We have that

$$\begin{aligned}
E[\hat{\Psi}_n(\theta)] =& E_\theta[X^*_j] = \Psi(\theta)\\
\text{Var}[\hat{\Psi}_n(\theta)] =& \frac{1}{N}\text{Var}[X^*_j] = \frac{1}{N}\sigma^2(\theta)
\end{aligned}$$

By the central limit theorem we then get

$$\hat{\Psi}_n(\theta) - \Psi(\theta) \approx N(0, \tfrac{1}{N}\sigma^2(\theta))$$

(b). We have

$$\hat{\Psi}_n(\theta_j) - \Psi(\theta_j) \approx N(0, \frac{1}{N}\sigma^2(\theta_j)), \quad j = 1, 2$$

and combined with independence we get

$$\hat{\Psi}_n(\theta_1) - \hat{\Psi}_n(\theta_2) \approx N(\Psi(\theta_1) - \Psi(\theta_2), \tfrac{1}{N}(\sigma^2(\theta_1) + \sigma^2(\theta_2)))$$

(c). Now

$$\hat{\Psi}_n(\theta_1) - \hat{\Psi}_n(\theta_2) = \frac{1}{N}\sum_{j=1}^{N}[h(\theta_1, Z_j^*) - h(\theta_2, Z_j^*)]$$

which has the same expectation as before but variance

$$\begin{aligned}
&\mathrm{Var}[\hat{\Psi}_n(\theta_1) - \hat{\Psi}_n(\theta_2)] \\
&\quad =\tfrac{1}{N}\mathrm{Var}[h(\theta_1, Z_j^*) - h(\theta_2, Z_j^*)] \\
&\quad =\tfrac{1}{N}[\mathrm{Var}[h(\theta_1, Z_j^*)] + \mathrm{Var}[h(\theta_2, Z_j^*)] - 2\mathrm{Cov}[h(\theta_1, Z_j^*), h(\theta_2, Z_j^*)]] \\
&\quad =\tfrac{1}{N}[\sigma^2(\theta_1) + \sigma^2(\theta_2) - 2\sigma(\theta_1)\sigma(\theta_2)\rho(\theta_1, \theta_2)] \\
&\quad =\tfrac{1}{N}(\sigma^2(\theta_1) + \sigma^2(\theta_2))[1 - \tfrac{2\sigma(\theta_1)\sigma(\theta_2)}{\sigma^2(\theta_1)+\sigma^2(\theta_2)}\rho(\theta_1, \theta_2)] \equiv \tfrac{1}{N}\tau^2(\theta_1, \theta_2)
\end{aligned}$$

(d). We then have

$$\hat{\Psi}_n(\theta_1) - \hat{\Psi}_n(\theta_2) \approx \frac{1}{N}\sum_{j=1}^{N}\frac{\partial h(\theta_1, Z_j^*)}{\partial \theta}(\theta_2 - \theta_1)$$

and we obtain the approximate distribution

$$\hat{\Psi}_n(\theta_1) - \hat{\Psi}_n(\theta_2) \approx N\left((\theta_2 - \theta_1)E\left[\frac{\partial h(\theta_1, Z_j^*)}{\partial \theta}\right], \tfrac{1}{N}\tau^2(\theta_1, \theta_2)\right)$$

Note that if $h(\cdot)$ is sufficiently smooth and both h and its derivative is integrable, we have

$$\begin{aligned}
E\left[\frac{\partial h(\theta, Z_j^*)}{\partial \theta}\right] &= \int_z \frac{\partial h(\theta, Z_j^*)}{\partial \theta} f(z)dz \\
&= \frac{\partial}{\partial \theta}\int_z h(\theta, Z_j^*)f(z)dz = \frac{\partial}{\partial \theta}\Psi(\theta)
\end{aligned}$$

(e). We have for $\theta_1 \approx \theta_2$ that

$$\frac{\partial \Psi(\theta_1)}{\partial \theta} \approx \frac{\Psi(\theta_1) - \Psi(\theta_2)}{\theta_1 - \theta_2}$$

which we can approximate by

$$\frac{\hat{\Psi}_n(\theta_1)-\hat{\Psi}_n(\theta_2)}{\theta_1-\theta_2} \approx N\left(\Psi'(\theta), \tfrac{1}{N}\tfrac{\tau^2(\theta_1,\theta_2)}{(\theta_1-\theta_2)^2}\right)$$

Given that the denominator in the variance can be small, it is important to also make the nominator small.

Solution to exercise 36. (a). We have that the proposal is $X_t^* = X_t + \varepsilon_t$. Since the proposal distribution is symmetric, the Metropolis-Hastings ratio becomes

$$R_t = \frac{h(X_t^*)}{h(X_t)} = \frac{h(X_t+\varepsilon_t)}{h(X_t)}.$$

If we generate $U_t \sim \text{Uniform}[0,1]$, then we can accept if $U_t \leq \min\{1, R_t\}$ which is equivalent to accept if $U_t < R_t$. This means that

$$\begin{aligned} X_{t+1} =& \begin{cases} X_t^* & \text{if } U_t < R_t \\ X_t & \text{otherwise} \end{cases} \\ =& X_t + I_t\varepsilon_t \end{aligned}$$

given the definition of I_t.

Making the distribution of ε_t to symmetric simplifies the MH-ratio in that the proposal densities disappear.

(b). In this case

$$R_t = \frac{\exp(-0.5(X_t+\varepsilon_t)^2)}{\exp(-0.5X_t^2)} = \exp(-X_t\varepsilon_t - 0.5\varepsilon_t^2)$$

Since we have a distribution centered at zero, we would like to mostly move towards zero, which means making ε_t negative if X_t is positive and vice verse.

(c). No, if $R_t \geq 1$, you do not need to generate U_t.

Solution to exercise 38. (a). We have in general that if

$$\boldsymbol{x} = \begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right)$$

then

$$\begin{aligned} E[\boldsymbol{x}_1|\boldsymbol{x}_2] =& \boldsymbol{\mu}_1 + \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}[\boldsymbol{x}_2 - \boldsymbol{\mu}_2] \\ \text{Var}[\boldsymbol{x}_1|\boldsymbol{x}_2] =& \boldsymbol{\Sigma}_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}. \end{aligned}$$

This gives

$$
\begin{aligned}
E[x_1|x_2] =& 1 + a(x_2 - 2) = 1 - 2a + ax_2 \\
\text{Var}[x_1|x_2] =& 1 - a^2 \\
E[x_2|x_1] =& 2 + a(x_1 - 1) = 2 - a + ax_1 \\
\text{Var}[x_1|x_2] =& 1 - a^2
\end{aligned}
$$

Solution to exercise 39. (a). We then have

$$
\begin{aligned}
X|Y \sim& N(\rho Y, 1 - \rho^2) \\
Y|X \sim& N(\rho X, 1 - \rho^2)
\end{aligned}
$$

or

$$
\begin{aligned}
X|Y =& \rho Y + \sqrt{1 - \rho^2} Z \\
Y|X =& \rho X + \sqrt{1 - \rho^2} V
\end{aligned}
$$

which, when using the Gibbs sampler gives the recursion

$$
\begin{aligned}
Y_n =& \rho X_n + \sqrt{1 - \rho^2} Z_n, \\
X_n =& \rho Y_{n-1} + \sqrt{1 - \rho^2} V_n.
\end{aligned}
$$

were all the $\{Z_n\}$ and $\{V_n\}$ are independent variables

(b). From the equation above we have

$$
\begin{aligned}
X_n =& \rho Y_{n-1} + \sqrt{1 - \rho^2} V_n \\
=& \rho[\rho X_{n-1} + \sqrt{1 - \rho^2} Z_{n-1}] + \sqrt{1 - \rho^2} V_n \\
=& \rho^2 X_{n-1} + \sqrt{1 - \rho^2}[\rho Z_{n-1} + V_n] \\
=& \rho^2 X_{n-1} + \varepsilon_n
\end{aligned}
$$

were $E[\varepsilon_n] = 0$ and

$$
\text{Var}[\varepsilon_n] = (1 - \rho^2)[\rho^2 + 1] = 1 - \rho^4 \equiv \sigma_\varepsilon^2
$$

Since the factor in front of X_{n-1} in the reccursion has an absolute value less than 1, (here: $\rho^2 < 1$), the general results about AR(1) processes applies.

(c). We have

$$
E[X_n] = E[\rho^2 X_{n-1} + \varepsilon_n] = \rho^2 E[X_{n-1}]
$$

which recursively gives $E[X_n] = \rho^{2n} \mu_0$ were $\mu_0 = E[X_0]$.

(d). We have

$$\begin{aligned}\text{Var}[X_n] =& \text{Var}[\rho^2 X_{n-1} + \varepsilon_n] \\ =& \rho^4 \text{Var}[X_{n-1}] + \sigma_\varepsilon^2\end{aligned}$$

Assume now the statement about the variance is true for n. Then

$$\begin{aligned}\text{Var}[X_{n+1}] =& \rho^4 \text{Var}[X_n] + \sigma_\varepsilon^2 \\ =& \rho^4 \tfrac{\sigma_\varepsilon^2}{1-\rho^4}(1-\rho^{4n}) + \sigma_\varepsilon^2 \\ =& \sigma_\varepsilon^2 \tfrac{\rho^4(1-\rho^{4n})+1-\rho^4}{1-\rho^4} \\ =& \sigma_\varepsilon^2 \tfrac{1-\rho^{4(n+1)}}{1-\rho^4} = 1-\rho^{4(n+1)}\end{aligned}$$

(e). When $n \to \infty$ we have

$$\begin{aligned}E[X_n] \to& 0 \\ \text{Var}[X_n] \to& \tfrac{\sigma_\varepsilon^2}{1-\rho^4} = 1\end{aligned}$$

(f). We have

$$\begin{aligned}Y_n =& \rho X_n + \sqrt{1-\rho^2} Z_n \\ =& \rho[\rho Y_{n-1} + \sqrt{1-\rho^2} V_n] + \sqrt{1-\rho^2} Z_n \\ =& \rho^2 Y_{n-1} + \sqrt{1-\rho^2}[\rho V_n + Z_n]\end{aligned}$$

which has the same structure as for X_n and the results become identical.

(g). We have

$$\begin{aligned}E[X_n Y_n] =& E[X_n(\rho X_n + \sqrt{1-\rho^2} Z_n)] \\ =& \rho E[X_n^2] \\ =& \rho[1-\rho^{4n} + \rho^{4n}\mu_0^2] \\ =& \rho + \rho^{4n+1}(\mu_0^2 - 1)\end{aligned}$$

(h). We see that $E[X_n Y_n] \to \rho$

(i). We then see that the limit distribution for $X_n, Y_n)$ indeed is the target distribution. We see that the convergence speed is geometric in ρ^2 for the mean and geometric in ρ^4 for the variances and the correlations.

Solution to exercise 41. (a). We have that

$$\begin{aligned}\int_x \pi(x)P(y|x)dx =& \int_x \pi(x)[\alpha P_1(y|x) + (1-\alpha)P_2(y|x)]dx \\ =&\alpha \int_x \pi(x)P_1(y|x)dx + +(1-\alpha)\int_x \pi(x)P_2(y|x)dx \\ =&\alpha\pi(y) + (1-\alpha)\pi(y) = \pi(y)\end{aligned}$$

(b). This means that if we have two different MCMC algorithms, both with the same target distribution, we can combine these to construct a new algorithm where we first make a draw on which algorithm to use.

Solution to exercise 42. (a). Assume (X, Y) is the current sample, and we draw new ones by

$$X^* = Z^* \qquad Y^* = \rho Z^* + \sqrt{1-\rho^2}V^*.$$

Note that in this case (X^*, Y^*) do not depend on (X, Y) at all! This means that we get immidiate convergence in this case.

(b). If there are components of the full random vector that has a simple marginal distribution, this should be utilized and can influence the convergence rate considerable. In most cases it might not be possible to obtain such marginal distributions. However, in some cases with three sets of variables, we may have that

$$p(x, y|z) = p(x|z)p(y|x, z)$$

where $p(x|z)$ is available. An ordinary Gibbs sampler would be to simulate

$$\begin{aligned}x \sim& p(x|y, z)\\ y \sim& p(y|x, z)\\ z \sim& p(z|x, y)\end{aligned}$$

which would be to jump between three variables or blocks (if x, y or z are vectors). With $p(x|z)$ is available, we can simulate through

$$\begin{aligned}(x, y) \sim& p(x, y|z)\\ z \sim& p(z|x, y)\end{aligned}$$

which only contains two blocks and will typically have much faster convergence. We can diretly simulate from $p(x, y|z) = p(x|z)p(y|x, z)$ due to the assumptions made.

Solution to exercise 43. (a). We have that

$$
\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta}) &= \prod_{i=1}^{n} p(y_i|x_i,\boldsymbol{\beta}) \\
&= \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \\
&= \prod_{i=1}^{n} \frac{\exp(\beta_1+\beta_2 x_i)_i^y}{1\exp(\beta_1+\beta_2)} \\
\log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta}) &= \sum_{i=1}^{n}[y_i(\beta_1+\beta_2 x_i) - \log(1\exp(\beta_1+\beta_2)] \\
U(\boldsymbol{\beta}) &= -\log p(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) \\
&= \text{Const} - \log p(\boldsymbol{\beta}) - \log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta}) \\
&= \text{Const} + \tfrac{1}{\sigma_\beta^2}\sum_{j=1}^{2}\beta_j^2 - \sum_{i=1}^{n} y_i(\beta_1+\beta_2 x_i) + \sum_{i=1}^{n}\log(1+\exp(\beta_1+\beta_2 x_i)) \\
H(\boldsymbol{\beta},\boldsymbol{p}) &= U(\boldsymbol{\beta}) + \tfrac{1}{2}\boldsymbol{p}^T\boldsymbol{p} \\
&= \tfrac{1}{\sigma_\beta^2}\sum_{j=1}^{2}\beta_j^2 - \sum_{i=1}^{n} y_i(\beta_1+\beta_2 x_i) + \sum_{i=1}^{n}\log(1+\exp(\beta_1+\beta_2 x_i)) + \tfrac{1}{2}\sum_{j=1}^{2} p_j^2
\end{aligned}
$$

where we have neglected the constant term since this does not matter.