NTNU
Norwegian University of Science and Technology

Department of Mathematical Sciences

Examination paper for **Solution sketch: TMA4300**

Academic contact during examination: Håkon Tjelmeland

Phone: 4822 1896

Examination date: June 6th 2018

Examination time (from–to): 09:00–13:00

Permitted examination support material: C:

- Calculator HP30S, CITIZEN SR-270X, CITIZEN SR-270X College or Casio fx-82ES PLUS with empty memory.
- Statistiske tabeller og formler, Akademika.
- One yellow, stamped A5 sheet with own handwritten formulas and notes.

Other information:

- All answers should be justified!
- All sub-problems in the exam count the same.
- In your solution you can use English and/or Norwegian.

Language: English

Number of pages: 9

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	2-sidig ☒
sort/hvit ☒	farger ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1

a) To sample from this distribution the most natural alternative is to use the probability integral transform method. We start by finding the cumulative distribution function $G(x)$. For $x \in (-\pi/2, \pi/2]$ we get

$$
\begin{aligned}
G(x) &= \int_{-\infty}^{x} g(u)\mathrm{d}u = \int_{-\frac{\pi}{2}}^{x} \frac{1}{2}\cos(u)\mathrm{d}u = \left[\frac{1}{2}\sin(u)\right]_{-\frac{\pi}{2}}^{x} \\
&= \frac{1}{2}\sin(x) - \frac{1}{2}\sin\left(-\frac{\pi}{2}\right) = \frac{1}{2}\sin(x) - \frac{1}{2}\cdot(-1) \\
&= \frac{1}{2}\left(1 + \sin(x)\right).
\end{aligned}
$$

Sampling a $u \sim \text{Unif}(0, 1)$ we get a sample from $g(x)$ by solving $u = G(x)$ with respect to x. This gives

$$
\begin{aligned}
u &= G(x) \\
u &= \frac{1}{2}\left(1 + \sin(x)\right) \\
2u - 1 &= \sin(x) \\
x &= \sin^{-1}(2u - 1).
\end{aligned}
$$

Pseudo-code for generating one sample is then simply:

Generate $u \sim \text{Unif}(0, 1)$.
Compute $x = \sin^{-1}(2u - 1)$.
Return x.

Note: An alternative is to use rejection sampling with for example a uniform distribution on $(-\pi/2, \pi/2]$ as proposal distribution. However, one then also needs to discuss how to sample from this uniform distribution.

b) To sample from $f(x)$ by rejection sampling and using $g(x)$ as proposal distribution we first need to find a constant c so that

$$
\frac{f(x)}{g(x)} \leq c \quad \text{for all } x \text{ where } g(x) > 0.
$$

We have that $g(x) > 0$ for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$, for which we have

$$
\frac{f(x)}{g(x)} = \frac{k|x|^{\alpha}\cos(x)}{\frac{1}{2}\cos(x)} = 2k|x|^{\alpha} \leq 2k\left(\frac{\pi}{2}\right)^{\alpha}.
$$

We can thereby choose $c = 2k\left(\frac{\pi}{2}\right)^\alpha$. The acceptance probability in the rejection sampling thus becomes

$$p = \frac{1}{c} \cdot \frac{f(x)}{g(x)} = \frac{1}{2k\left(\frac{\pi}{2}\right)^\alpha} \cdot 2k|x|^\alpha = \left(\frac{2}{\pi} \cdot |x|\right)^\alpha .$$

Pseudo-code for generating one sample from $f(x)$ is then:

finished := 0
while finished = 0 **do**
  Generate $x \sim g(x)$
  Generate $u \sim \text{Unif}(0, 1)$
  Compute $p := \left(\frac{2}{\pi} \cdot |x|\right)^\alpha$
  if $u < p$ **then** finished := 1
return x

Problem 2

The acceptance probability for the specified Metropolis–Hastings algorithm becomes, for $y \neq x$,

$$\begin{aligned} \alpha(y|x) &= \min\left\{1, \frac{f(y)}{f(x)} \cdot \frac{q(x|y)}{q(y|x)}\right\} = \min\left\{1, \frac{f(y)}{f(x)} \cdot \frac{\frac{1}{3}}{\frac{1}{3}}\right\} \\ &= \min\left\{1, \frac{f(y)}{f(x)}\right\} = \begin{cases} 1 & \text{if } y > x, \\ \frac{y/10}{x/10} = \frac{y}{x} & \text{otherwise.} \end{cases} \end{aligned}$$

The off-diagonal elements of the transition matrix is thereby

$$P(y|x) = q(y|x)\alpha(y|x) = \frac{1}{3}\alpha(y|x) = \begin{cases} \frac{1}{3} & \text{if } y > x, \\ \frac{y}{3x} & \text{otherwise.} \end{cases}$$

Using this and that the elements in each row must sum to one we get the transition matrix P,

$$P = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{12} & \frac{1}{6} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

A Markov chain is irreducible if it is possible to come from any state to any other state in a finite number of steps. For the Markov chain in question it is possible to come from any state to any other state in one step. The Markov chain is thus irreducible. A sufficient condition for an irreducible Markov chain to be aperiodic is that at least one diagonal element of the transition matrix is strictly larger than zero. This condition is fulfilled for the above transition matrix so the Markov chain is aperiodic.

Problem 3

a) Up to proportionality, for $\theta, \alpha, \beta > 0$ the posterior distribution becomes

$$\begin{aligned}
f(\theta,\alpha,\beta|y_1,\ldots,y_n) &\propto f(\theta,\alpha,\beta)\cdot f(y_1,\ldots,y_n|\theta,\alpha,\beta) \\
&= f(\theta)f(\alpha)f(\beta)\prod_{i=1}^n f(y_i|\theta,\alpha,\beta) \\
&\propto \frac{1}{\theta}\prod_{i=1}^n \frac{1}{\sqrt{\theta}}\exp\left\{-\frac{1}{2\theta}\left(y_i-\frac{\alpha x_i}{\beta+x_i}\right)^2\right\} \\
&\propto \frac{1}{\theta^{\frac{n}{2}+1}}\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n\left(y_i-\frac{\alpha x_i}{\beta+x_i}\right)^2\right\} \qquad (1)
\end{aligned}$$

The full conditional for θ becomes

$$f(\theta|\alpha,\beta,y_1,\ldots,y_n) \propto \frac{1}{\theta^{\frac{n}{2}+1}}\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n\left(y_i-\frac{\alpha x_i}{\beta+x_i}\right)^2\right\},$$

which we can recognise as an inverse gamma distribution. Different parameterisations are in use for the inverse gamma distribution. Adopting the parameterisation

$$f(z;a,b) = \frac{1}{b^a\Gamma(a)}\frac{\exp\left\{-\frac{1}{zb}\right\}}{z^{a+1}},$$

the parameters in the full conditional for θ becomes

$$a = \frac{n}{2} \quad \text{and} \quad b = \frac{2}{\sum_{i=1}^n\left(y_i-\frac{\alpha x_i}{\beta+x_i}\right)^2}.$$

The full conditional for α becomes

$$f(\alpha|\theta,\beta,y_1,\ldots,y_n) \propto \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n\left(y_i-\frac{\alpha x_i}{\beta+x_i}\right)^2\right\} \quad \text{when } \alpha > 0$$

and $f(\alpha|\theta,\beta,y_1,\ldots,y_n) = 0$ otherwise, where we can observe that the exponent is a second order function in α. The full conditional distribution for α

is therefore a truncated normal distribution. To find the mean and variance of this normal distribution we study the exponent in more detail. We get

$$\begin{aligned}\frac{1}{\theta}\sum_{i=1}^{n}\left(y_i - \frac{\alpha x_i}{\beta + x_i}\right)^2 &= \text{constant} + \frac{1}{\theta}\sum_{i=1}^{n}\left[-2\alpha\frac{x_i y_i}{\beta + x_i} + \alpha^2\left(\frac{x_i}{\beta + x_i}\right)^2\right] \\ &= \text{constant} + \alpha^2\frac{1}{\theta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2 - 2\alpha\frac{1}{\theta}\sum_{i=1}^{n}\frac{x_i y_i}{\beta + x_i}.\end{aligned}$$

Letting μ and σ^2 denote the mean and variance, respectively, in the full conditional for α we must have

$$\begin{aligned}\text{constant} + \alpha^2\frac{1}{\theta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2 - 2\alpha\frac{1}{\theta}\sum_{i=1}^{n}\frac{x_i y_i}{\beta + x_i} &= \frac{1}{\sigma^2}(\alpha - \mu)^2 \\ &= \text{constant} - 2\alpha \cdot \frac{\mu}{\sigma^2} + \frac{\alpha^2}{\sigma^2}.\end{aligned}$$

Thus, we have

$$\frac{1}{\theta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2 = \frac{1}{\sigma^2} \quad \Rightarrow \quad \sigma^2 = \frac{\theta}{\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2}$$

and

$$\frac{1}{\theta}\sum_{i=1}^{n}\frac{x_i y_i}{\beta + x_i} = \frac{\mu}{\sigma^2} \quad \Rightarrow \quad \mu = \frac{\sum_{i=1}^{n}\frac{x_i y_i}{\beta + x_i}}{\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2}.$$

Thus,

$$f(\alpha|\theta, \beta, y_1, \ldots, y_n) \propto \mathrm{N}\left(\alpha \left| \frac{\sum_{i=1}^{n}\frac{x_i y_i}{\beta + x_i}}{\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2}, \frac{\theta}{\sum_{i=1}^{n}\left(\frac{x_i}{\beta + x_i}\right)^2}\right.\right) \cdot I(\alpha > 0),$$

where $N(\alpha|\mu, \sigma^2)$ is the density function of a normal distribution with mean μ and variance σ^2. Finally, the full conditional for β becomes

$$f(\beta|\theta, \alpha, y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n}\left(y_i - \frac{\alpha x_i}{\beta + x_i}\right)^2\right\},$$

which does not belong to a known parametric family.

b) Since the full conditional for θ and α belongs to known parametric families which we know how to sample from, we can use Gibbs updates for these two.

We can sample from the truncated normal distribution by rejection sampling with the (un-truncated) normal distribution as proposal distribution. If the proposed value α is larger than zero it should be accepted with probability one (i.e. always accepted) and otherwise it should be accepted with probability zero (i.e. always rejected). The Metropolis–Hastings acceptance probabilities for Gibbs steps are always identical to one.

For β we can for example propose a new value from a normal distribution centered at the current value, i.e. for a tuning parameter $\tau^2 > 0$ we propose

$$\widetilde{\beta} \sim \mathrm{N}(\beta, \tau^2).$$

Using the expression in (1) when $\beta > 0$ and $f(\theta, \alpha, \beta | y_1, \ldots, y_n) = 0$ whenever $\beta \leq 0$ the associated Metropolis–Hastings acceptance probability becomes

$$\begin{aligned} a(\widetilde{\beta}|\beta) &= \min\left\{1, \frac{f(\theta,\alpha,\widetilde{\beta}|y_1,\ldots,y_n)}{f(\theta,\alpha,\beta|y_1,\ldots,y_n)} \cdot \frac{\mathrm{N}(\beta|\widetilde{\beta},\tau^2)}{\mathrm{N}(\widetilde{\beta}|\beta,\tau^2)}\right\} \\ &= \begin{cases} \min\left\{1, \dfrac{\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n \left(y_i - \frac{\alpha x_i}{\widetilde{\beta}+x_i}\right)^2\right\}}{\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n \left(y_i - \frac{\alpha x_i}{\beta+x_i}\right)^2\right\}}\right\} & \text{if } \widetilde{\beta} > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{2}$$

where we have used $\mathrm{N}(\beta|\widetilde{\beta},\tau^2) = \mathrm{N}(\widetilde{\beta}|\beta,\tau^2)$.

c) Let $\{\theta_m, \alpha_m, \beta_m\}_{m=1}^M$ denote the values simulated by the Metropolis–Hastings algorithm. To estimate properties of the posterior distribution we first need to identify the burn-in period. We typically do this by studying trace plots of the simulated values and see when the traces seem to have stabilised statistically. In the following we assume the burn-in period to end at $m = T$, so we use $\{\theta_m, \alpha_m, \beta_m\}_{m=T}^M$ to estimate the posterior properties. We estimate the posterior mean values simply by

$$\widehat{\mathrm{E}}[\alpha|y_1,\ldots,y_n] = \frac{1}{M-T+1}\sum_{m=T}^M \alpha_m$$

and

$$\widehat{\mathrm{E}}[\beta|y_1,\ldots,y_n] = \frac{1}{M-T+1}\sum_{m=T}^M \beta_m.$$

A simple way to estimate a 90% prediction interval for a new observation y_0 is first to simulate

$$y_0^m \sim \mathrm{N}\left(\frac{\alpha_m x_0}{\beta_m + x_0}, \theta_m\right)$$

for $m = T, \dots, M$ and thereafter estimate the prediction interval limits by the 5% and 95% quantiles of the simulated y_0^m values. Thus, we should sort $y_0^T, \dots, y_0^M$ from smallest to largest, denoted as

$$y_0^{(1)}, \dots, y_0^{(M-T+1)}.$$

Assuming M is chosen so that $(M - T + 1) \cdot 0.05$ is an integer, the estimate of the prediction interval is

$$\left[y_0^{((M-T+1)\cdot 0.05)}, y_0^{((M-T+1)\cdot 0.95)}\right].$$

Problem 4

a) The empirical distribution puts probability $\frac{1}{n}$ on each observed value. Letting $\widehat{F}$ denote the empirical distribution we have

$$\text{Prob}_{\widehat{F}}(x \in A) = \frac{\#x_i\text{'s in } A}{n}.$$

For a parameter θ defined as $\theta = t(F)$, the plug-in estimator is defined as

$$\widehat{\theta} = t(\widehat{F}).$$

Letting $x^\star$ denote a sample from $\widehat{F}$, the plug-in estimator for $\mu = \text{E}_F[x]$ is

$$\widehat{\mu} = \text{E}_{\widehat{F}}[x^\star] = \sum_{i=1}^n x_i \cdot \frac{1}{n} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}.$$

The plug-in estimator for the the variance $\sigma^2 = \text{Var}_F[x]$ becomes

$$\widehat{\sigma}^2 = \text{Var}_{\widehat{F}}[x^\star] = \sum_{i=1}^n (x_i - \text{E}_{\widehat{F}}[x^\star])^2 \cdot \frac{1}{n} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2.$$

b) The bias of $\widehat{\mu}$ is defined as

$$\text{bias}_F = \text{E}_F[s(x)] - \text{E}_F[x].$$

The ideal bootstrap estimator for the bias is defined as the plug-in estimator for this quantity, i.e.

$$\text{bias}_{\widehat{F}} = \text{E}_{\widehat{F}}[s(x^\star)] - \text{E}_{\widehat{F}}[x^\star].$$

Above we have shown that $\text{E}_{\widehat{F}}[x^\star] = \mu$, so it remains to find a simple expression for $\text{E}_{\widehat{F}}[s(x^\star)]$. Inserting for $s(x)$ we get

$$\text{E}_{\widehat{F}}[s(x^\star)] = \text{E}_{\widehat{F}}\left[\frac{1}{n}\sum_{i=1}^n x_i^\star\right] = \frac{1}{n}\sum_{i=1}^n \text{E}_{\widehat{F}}[x_i^\star] = \frac{1}{n}\sum_{i=1}^n \mu = \mu,$$

where we have used that $\text{E}_{\widehat{F}}[x_i^\star] = \mu$ when $x_i^\star \sim \widehat{F}$. We get

$$\text{bias}_{\widehat{F}} = \text{E}_{\widehat{F}}[s(x^\star)] - \text{E}_{\widehat{F}}[x^\star] = \mu - \mu = 0.$$

Problem 5

a) We have

$$\begin{aligned} f(x;\mu,\sigma) &= \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}\right] \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\frac{1}{\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}. \end{aligned}$$

Taking the logarithm we get

$$\ln f(x;\mu,\sigma) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Taking the expected value under the assumption $x_i \sim N(\mu^{(t)},(\sigma^{(t)})^2)$ we get

$$\mathrm{E}\left[\ln f(x;\mu,\sigma)|z,\mu^{(t)},\sigma^{(t)}\right] = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathrm{E}\left[(x_i-\mu)^2|z,\mu^{(t)},\sigma^{(t)}\right]$$

Expanding the square inside the expectation operator we get

$$\begin{aligned} \mathrm{E}\left[(x_i-\mu)^2|z,\mu^{(t)},\sigma^{(t)}\right] &= \mathrm{E}\left[x_i^2 - 2\mu x_i + \mu^2|z,\mu^{(t)},\sigma^{(t)}\right] \\ &= \mathrm{E}\left[x_i^2|z,\mu^{(t)},\sigma^{(t)}\right] - 2\mu\mathrm{E}\left[x_i|z,\mu^{(t)},\sigma^{(t)}\right] + \mu^2. \end{aligned}$$

Thus, we need to find expressions for $\mathrm{E}\left[x_i|z,\mu^{(t)},\sigma^{(t)}\right]$ and $\mathrm{E}\left[x_i^2|z,\mu^{(t)},\sigma^{(t)}\right]$. The density of $x_i \sim N(\mu^{(t)},(\sigma^{(t)})^2)$ can be expressed as

$$f(x_i|\mu^{(t)},\sigma^{(t)}) = \frac{1}{\sigma^{(t)}}\varphi\left(\frac{x_i-\mu^{(t)}}{\sigma^{(t)}}\right)$$

and corresponding cumulative distribution function is

$$F(x_i|\mu^{(t)},\sigma^{(t)}) = \Phi\left(\frac{x_i-\mu^{(t)}}{\sigma^{(t)}}\right).$$

The conditional density for x_i given z_i becomes

$$f(x_i|z_i,\mu^{(t)},\sigma^{(t)}) = \frac{f(x_i|\mu^{(t)},\sigma^{(t)})}{P(x_i\in[z_i,z_i+1))} = \frac{\frac{1}{\sigma^{(t)}}\varphi\left(\frac{x_i-\mu^{(t)}}{\sigma^{(t)}}\right)}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right) - \Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)},$$

and the corresponding conditional expectation is

$$\begin{aligned}
\mathrm{E}\left[x_i|z,\mu^{(t)},\sigma^{(t)}\right] &= \int_{z_i}^{z_i+1} x_i f(x_i|z_i,\mu^{(t)},\sigma^{(t)})\mathrm{d}x_i \\
&= \frac{\frac{1}{\sigma^{(t)}}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)} \int_{z_i}^{z_i+1} x_i \varphi\left(\frac{x_i-\mu^{(t)}}{\sigma^{(t)}}\right)\mathrm{d}x_i \\
&= \frac{A\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)}.
\end{aligned}$$

Correspondingly we get

$$\begin{aligned}
\mathrm{E}\left[x_i^2|z,\mu^{(t)},\sigma^{(t)}\right] &= \int_{z_i}^{z_i+1} x_i^2 f(x_i|z_i,\mu^{(t)},\sigma^{(t)})\mathrm{d}x_i \\
&= \frac{\frac{1}{\sigma^{(t)}}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)} \int_{z_i}^{z_i+1} x_i^2 \varphi\left(\frac{x_i-\mu^{(t)}}{\sigma^{(t)}}\right)\mathrm{d}x_i \\
&= \frac{B\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)}.
\end{aligned}$$

Thereby we have

$$\mathrm{E}\left[\ln f(x;\mu,\sigma)|z,\mu^{(t)},\sigma^{(t)}\right] = -\frac{n}{2}\ln(2\pi) - n\ln\sigma$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\frac{B\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)} - 2\mu\frac{A\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)} + \mu^2\right]$$

$$= -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\left(n\mu^2 - 2\mu\alpha(z,\mu^{(t)},\sigma^{(t)}) + \beta(z,\mu^{(t)},\sigma^{(t)}) + n\mu^2\right),$$

where

$$\alpha(z,\mu^{(t)},\sigma^{(t)}) = \sum_{i=1}^{n}\frac{A\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)}$$

and

$$\beta(z,\mu^{(t)},\sigma^{(t)}) = \sum_{i=1}^{n}\frac{B\left(\mu^{(t)},\sigma^{(t)},z_i,z_i+1\right)/\sigma^{(t)}}{\Phi\left(\frac{z_i+1-\mu^{(t)}}{\sigma^{(t)}}\right)-\Phi\left(\frac{z_i-\mu^{(t)}}{\sigma^{(t)}}\right)}.$$

b) To find for what values of μ and σ the $\mathrm{E}\left[\ln f(x;\mu,\sigma)|z,\mu^{(t)},\sigma^{(t)}\right]$ has its maximum we find the partial derivatives of this expected value with respect to each of μ and σ. We get

$$\begin{aligned}
\frac{\partial}{\partial\mu}\mathrm{E}\left[\ln f(x;\mu,\sigma)|z,\mu^{(t)},\sigma^{(t)}\right] &= -\frac{1}{2\sigma^2}\left(2n\mu - 2\alpha(z,\mu^{(t)},\sigma^{(t)})\right)\\
\frac{\partial}{\partial\sigma}\mathrm{E}\left[\ln f(x;\mu,\sigma)|z,\mu^{(t)},\sigma^{(t)}\right] &= -\frac{n}{\sigma}+\frac{1}{\sigma^3}\left(n\mu^2 - 2\mu\alpha(z,\mu^{(t)},\sigma^{(t)} + \beta(z,\mu^{(t)},\sigma^{(t)})\right)
\end{aligned}$$

Setting the partial derivative with respect to μ equal to zero and solving with respect to μ gives

$$\mu = \frac{\alpha(z,\mu^{(t)},\sigma^{(t)})}{n}$$

and setting the partial derivative with respect to σ equal to zero gives

$$\begin{aligned}
\frac{n}{\sigma} &= \frac{1}{\sigma^3}\left(n\mu^2 - 2\mu\alpha(z,\mu^{(t)},\sigma^{(t)} + \beta(z,\mu^{(t)},\sigma^{(t)})\right)\\
\sigma^2 &= \frac{1}{n}\left(n\mu^2 - 2\mu\alpha(z,\mu^{(t)},\sigma^{(t)} + \beta(z,\mu^{(t)},\sigma^{(t)})\right)\\
\sigma &= \sqrt{\mu^2 - \frac{1}{n}\left(2\mu\alpha(z,\mu^{(t)},\sigma^{(t)}) - \beta(z,\mu^{(t)},\sigma^{(t)})\right)}.
\end{aligned}$$

Thus $\mu^{(t+1)}$ and $\sigma^{(t+1)}$ can be computed from $\mu^{(t)}$ and $\sigma^{(t)}$ by

$$\begin{aligned}
\mu^{(t+1)} &= \frac{\alpha(z,\mu^{(t)},\sigma^{(t)})}{n}\\
\sigma^{(t+1)} &= \sqrt{(\mu^{(t+1)})^2 - \frac{1}{n}\left(2\mu^{(t+1)}\alpha(z,\mu^{(t)},\sigma^{(t)}) - \beta(z,\mu^{(t)},\sigma^{(t)})\right)}.
\end{aligned}$$

The standard deviations of the maximum likelihood estimators can be estimated by bootstrapping by the following algorithm.

for $b = 1,\ldots,B$ **do**
  Draw a bootstrap sample $z_1^{\star b},\ldots,z_n^{\star b}$ from $z_1,\ldots,z_n$.
  Use the EM algorithm to compute maximum likelihood estimates based on $z_1^{\star b},\ldots,z_n^{\star b}$. Denote the result by $\widehat{\mu}_b^\star$ and $\widehat{\sigma}_b^\star$.
Estimate the standard deviations of $\widehat{\mu}$ and $\widehat{\sigma}$ by

$$\begin{aligned}
\widehat{\mathrm{SD}}\left[\widehat{\mu}\right] &= \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\mu}_b^\star - \overline{\widehat{\mu}^\star}\right)^2},\\
\widehat{\mathrm{SD}}\left[\widehat{\sigma}\right] &= \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\sigma}_b^\star - \overline{\widehat{\sigma}^\star}\right)^2},
\end{aligned}$$

respectively, where

$$\overline{\widehat{\mu}^\star} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mu}_b^\star \quad \text{and} \quad \overline{\widehat{\sigma}^\star} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\sigma}_b^\star.$$