Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **Solution Sketch for TMA4300**

Academic contact during examination: Sara Martino

Phone: 9940 3330

Examination date: 05/06/2019

Examination time (from–to): 15.00-19.00

Permitted examination support material: C:

- Calculator HP30S, CITIZEN SR-270X or CITIZEN SR-270X College, Casio fx-82ES PLUS with empty memory.
- Statistiske tabeller og formler, Tapir.
- K. Rottmann: Matematisk formelsamling.
- One yellow, stamped A5 sheet with own handwritten formulas and notes.
- Dictionary in any language.

Other information:

- All 10 sub-problems in this exam count approximately the same.
- All answers must be justified!!!
- In your solution you can use English and/or Norwegian.

Language: English

Number of pages: ??

Number of pages enclosed: 0

Checked by:

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 $G(x)$:

$$G(x) = \int_{-\infty}^{x} g(x)\ dx = \begin{cases} \frac{e^{\lambda x}}{2}, & \text{for } x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x}, & \text{for } x \geq 0 \end{cases}.$$

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

$$G^{-1}(u) = \begin{cases} \frac{1}{\lambda}\log(2u) & \text{for } 0 \leq u < 1/2 \\ -\frac{1}{\lambda}\log(2(1-u)) & \text{for } 1/2 \leq u \leq 1 \end{cases}$$

Pseudo code for generating $x \sim g(x)$ is then:

Generate $u \sim \text{Unif}(0,1)$
if $u < 0.5$ **then**
  Compute $x = \frac{1}{\lambda}\log(2u)$
else
  Compute $x = -\frac{1}{\lambda}\log(2(1-u))$
Return x

Note: You can also simulate $x \sim \text{exponential}(\lambda)$ and return x or $-x$ each with probability 0.5.

b) To sample from $h(x)$ using rejection sampling with $g(x)$ as proposal distribution we first have to find a constant M such that:

$$\frac{h(x)}{g(x)} \leq M \text{ for each } x \text{ such that } h(x) > 0$$

We have that $h(x) > 0$ for $\forall x \in \mathcal{R}$. So we get

$$\frac{h(x)}{g(x)} = \exp\left\{-\frac{x^2}{2} + \lambda|x|\right\}\frac{2}{\lambda}\frac{1}{\sqrt{2\pi}} \leq \frac{1}{\sqrt{2\pi}}\frac{2}{\lambda}\exp\left\{\frac{\lambda^2}{2}\right\}$$

WE can then choose

$$M = \frac{1}{\sqrt{2\pi}}\frac{2}{\lambda}\exp\left\{\frac{\lambda^2}{2}\right\}$$

The pseudocode to simulate $x \sim h(x)$ is then:

finished $= 0$

while finished $= 0$ **do**
  Generate $x \sim g(x)$
  Generate $u \sim \text{Unif}(0,1)$
  Compute $p = \exp\{\lambda|x| - 0.5x^2 - 0.5\lambda^2\}$
  if $u < p$ **then** finished $= 1$
Return x

c) In the rejection sampling algorithm the acceptance rate is given by:

$$P(\text{accept}) = \frac{1}{M} = \frac{\sqrt{2\pi}}{2}\lambda \exp\left\{-\frac{1}{2}\lambda^2\right\}$$

To find the maximum derive the $\log P(\text{accept})$ wrt λ and set to 0:

$$\begin{aligned}\frac{d}{d\lambda}\log P(\text{accept}) &= \frac{d}{d\lambda}\left[\text{const} + \log\lambda - \frac{1}{2}\lambda^2\right] \\ &= \frac{1}{\lambda} - \lambda = 0\end{aligned}$$

That is the acceptance rate is maximised for $\lambda = 1$. The maximum obtainable acceptance rate is then:

$$P(\text{accept}) \underbrace{=}_{\lambda=1} \frac{\sqrt{2\pi}}{2} 1 \exp\left\{-\frac{1}{2}1^2\right\} = \sqrt{\frac{\pi}{2}} e^{-1/2} \approx 0.76$$

Problem 2

a) a)

The posterior density is:

$$\begin{aligned}f(\theta,\lambda,k|y_1,\ldots,y_n) &\propto f(\theta)\ f(\lambda)\ f(k)\ f(y_1,\ldots,y_n|\theta,\lambda k) \\ &= f(\theta)\ f(\lambda)\ f(k)\ \prod_{i=1}^{k}\frac{\lambda^{y_i}}{y_i!}e^{-\lambda}\prod_{i=k+1}^{n}\frac{\theta^{y_i}}{y_i!}e^{-\theta}\end{aligned}$$

And the full conditionals:

$$\begin{aligned}f(\theta|\ldots) &\propto f(\theta)\prod_{i=k+1}^{n}\frac{\theta^{y_i}}{y_i!}e^{-\theta} \\ &\propto \theta^{0.5-1}\ e^{-\theta}\ \theta^{\sum_{i=k+1}^{n} y_i}\ e^{-(n-k)\theta} \\ &= \theta^{(\sum_{i=k+1}^{n} y_i + 0.5) - 1}\ e^{-(n-k+1)\theta}\end{aligned}$$

that is $\theta|\dots \sim \text{Gamma}(0.5 + \sum_{i=k+1}^{n} y_i, n-k+1)$

$$
\begin{aligned}
f(\lambda|\dots) &\propto f(\lambda)\prod_{i=1}^{k}\frac{\theta^{y_i}}{y_i!}e^{-\lambda}\\
&\propto \lambda^{0.5-1}\ e^{-\lambda}\ \lambda^{\sum_{i=1}^{k} y_i}\ e^{-k\theta}\\
&= \theta^{(\sum_{i=1}^{k} y_i+0.5)-1}\ e^{-(k+1)\theta}
\end{aligned}
$$

That is $\lambda|\dots \sim \text{Gamma}(0.5 + \sum_{i=1}^{k} y_i, k+1)$.
Finally

$$
\begin{aligned}
f(k|\dots) &\propto f(\theta,\lambda,k|y_1,\dots,y_n)\\
&\propto \prod_{i=1}^{k}\frac{\lambda^{y_i}}{y_i!}e^{-\lambda}\prod_{i=k+1}^{n}\frac{\theta^{y_i}}{y_i!}e^{-\theta}\\
&\propto \lambda^{\sum_{i=1}^{k} y_i}\theta^{\sum_{i=k+1}^{n} y_i}e^{-k\lambda-(n-k)\theta}
\end{aligned}
$$

This is not a known distribution.

b) **Alternative 1**: Use Gibbs sampling for λ and θ and MH within Gibbs for k. We can use, for example, an independent proposal for k:

$$
q(k^*|k^{t-1}) = \frac{1}{n} \text{ for } k = 1,\dots,n
$$

The corresponding acceptance probability is then:

$$
\begin{aligned}
\alpha(k^*|k^{t-1}) &= \min\left\{1, \frac{f(k^*|\dots)}{f(k^{t-1}|\dots)}\right\}\\
&= \min\ \left\{1, \frac{\exp\{\log(\lambda^t)\sum_{i=1}^{k^*} y_i + \log(\theta^t)\sum_{i=k^*+1}^{n} y_i - k^*\lambda^t - (n-k^*)\theta^t\}}{\exp\{\log(\lambda^t)\sum_{i=1}^{k^{t-1}} y_i + \log(\theta^t)\sum_{i=k^{t-1}+1}^{n} y_i - k^{t-1}\lambda^t - (n-k^{t-1})\theta^t\}}\right\}
\end{aligned}
$$

Pseudo code:

- Set initial values λ^0, θ^0 and k^0
- For $t = 1,\dots,I$
  - Sample $\lambda^t \sim \text{Gamma}(0.5 + \sum_{i=1}^{k^{t-1}} y_i, k^{t-1}+1)$
  - Sample $\theta^t \sim \text{Gamma}(0.5 + \sum_{i=k^{t-1}+1}^{n} y_i, n-k^{t-1}+1)$
  - Propose $k^* \sim \text{Unif}(1, n)$

- Compute $\alpha(k^*|kt-1)$
- Sample $u^t \sim \text{Unif}(0,1)$
- If $u^t < \alpha(k^*|k^{t-1})$ set $k^t = k^*$ otherwise set $k^t = k^{t-1}$

Alternative 2: The full conditional distribution for k does not belong to a known family. On the other side it is a discrete distribution therefore easy to normalise and sample from using the standard inversion probability techniques for discrete random variables. One possible alternative MCMC scheme is therefore:

Pseudo code:

- Set initial values λ^0, θ^0 and k^0
- For $t = 1, \ldots, T$
  - Sample $\lambda^t \sim \text{Gamma}(0.5 + \sum_{i=1}^{k^{t-1}} y_i, k^{t-1}+1)$
  - Sample $\theta^t \sim \text{Gamma}(0.5 + \sum_{i=k^{t-1}+1}^{n} y_i, n - k^{t-1}+1)$
  - Compute $f(k|\lambda^t, \theta^t, y_1, \ldots, y_n)$ for $k = 1, \ldots, n$
  - Compute the normalising constant:
  $$C^t = \sum_{k=1}^{n} f(k|\lambda^t, \theta^t, y_1, \ldots, y_n)$$
  - Compute the cdf $F(k|\lambda^t, \theta^t, y_1, \ldots, y_n)$
  - Sample $k^t \sim F(k|\lambda^t, \theta^t, y_1, \ldots, y_n)$ using

c) The plot show fast convergence for all three parameters, the mixing appear good the θ and λ but it is very slow for k.

To estimate the quantities of interest one needs first to find the length of the burn-in period. This is done by output analysis Assume the chain has (essentially) converged after $T_1 < T$ iterations. One can then estimate the posterior mean for λ by

$$\widehat{E}[\lambda|y_1, \ldots, y_n] = \frac{1}{T-T_1} \sum_{t=T_1+1}^{T} \lambda^t$$

To estimate the probability:

$$\begin{aligned}\widehat{P}[\lambda > 2.5|k = 100, y_1, \ldots, y_n] &= \frac{\widehat{P}[\lambda > 2.5, k = 100|y_1, \ldots, y_n]}{\widehat{P}[k = 100|y_1, \ldots, y_n]} \\ &= \frac{\sum_{t=T_1+1}^{T} I(\lambda^t > 2.5 \text{ and } k^t = 1)}{\sum_{t=T_1+1}^{T} I(k^t = 1)}\end{aligned}$$

Where $I(\cdot)$ is an indicator function.

Problem 3

a)

- Describe the idea behind the EM algorithm
- The function $Q(\lambda|\lambda^{(k)})$ is the mean of the complete data likelihood confitionla of the observed data and an impoted value for the unknown parameters.

$$Q(\lambda|\lambda^{(k)}) = E\{\log l(\lambda : x_1, \dots, x_n)|y_1, \dots, y_n, \lambda^{(k)}\}$$

b)

In this case the complete data consists in the series of unobserved lifetimes $x_1, \dots, x_n$ while the incomplete data consists in the observed series $y_1, \dots, y_n$. The complete data likelikelihood is given by:

$$L(\lambda : x_1, \dots, x_n) = \prod_{i=1}^{N} \lambda \; e^{-\lambda x_i} = \lambda^n \; e^{-\lambda \sum_{i=1}^{N} x_i}$$

and the log-likelihood:

$$l(\lambda : x_1, \dots, x_n) = n \log \lambda - \lambda \sum_{i=1}^{N} x_i$$

For this problem we have:

$$\begin{aligned} Q(\lambda|\lambda^{(k)}) &= E\{\log l(\lambda : x_1, \dots, x_n)|y_1, \dots, y_n, \lambda^{(k)}\} \\ & E(n \log \lambda - \lambda \sum_{i=1}^{N} x_i | y_1, \dots, y_n, \lambda^{(k)}) \\ &= n \log \lambda - \lambda \sum_{i=1}^{N} E(x_i|y_i, \lambda^{(k)}) \end{aligned}$$

So we need to find $E(x_i|y_i, \lambda^{(k)})$, here y_i can be either 0 or 1. We look at one case per time:

$$\begin{aligned} E(x_i|y_i = 0, \lambda^{(k)}) &= \int_0^T x \; \frac{\lambda^{(k)} e^{-\lambda^{(k)} x}}{1 - e^{-\lambda^{(k)} T}} \; dx \\ &= \frac{1}{\lambda^{(k)}} - T \frac{e^{-\lambda^{(k)} T}}{1 - e^{-\lambda^{(k)} T}} \end{aligned}$$

and

$$
\begin{aligned}
E(x_i|y_i = 1, \lambda^{(k)}) &= \int_T^{\infty} x\ \lambda^{(k)} e^{-\lambda^{(k)}(x-T)}\ dx \\
&= \frac{1}{\lambda^{(k)}} + T
\end{aligned}
$$

d) The $Q(\lambda|\lambda^{(k)})$ is then:

$$
Q(\lambda|\lambda^{(k)}) = n \log \lambda - \lambda \left\{ N_0 \left[\frac{1}{\lambda^{(k)}} - T \frac{e^{-\lambda^{(k)}x}}{1 - e^{-\lambda^{(k)}T}} \right] + N_1 \left[\frac{1}{\lambda^{(k)}} + T \right] \right\}
$$

The M step in the EM algorith consists in maximizing $Q(\lambda|\lambda^{(k)})$ with respect to λ. We need therefore to derive $Q(\lambda|\lambda^{(k)})$ and set the derivative to 0

$$
\frac{dQ(\lambda|\lambda^{(k)})}{d\lambda} = \frac{n}{\lambda} - \left\{ N_0 \left[\frac{1}{\lambda^{(k)}} - T \frac{e^{-\lambda^{(k)}x}}{1 - e^{-\lambda^{(k)}T}} \right] + N_1 \left[\frac{1}{\lambda^{(k)}} + T \right] \right\} = 0
$$

This gives us:

$$
\lambda^{(k+1)} = \frac{N}{\left\{ N_0 \left[\frac{1}{\lambda^{(k)}} - T \frac{e^{-\lambda^{(k)}x}}{1 - e^{-\lambda^{(k)}T}} \right] + N_1 \left[\frac{1}{\lambda^{(k)}} + T \right] \right\}}
$$