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: May 24th 2017
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 must be justified!
- All nine 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) We must first find the cumulative distribution function corresponding to the density $g(x)$. For $x \geq 0$ we get

$$\begin{aligned} G(x) &= \int_0^x g(x)\mathrm{d}x = \int_0^x 2xe^{-x^2}\mathrm{d}x = \left[-e^{-x^2}\right]_0^x \\ &= -e^{-x^2} - (-e^0) = 1 - e^{-x^2}. \end{aligned}$$

Next we need to find $G^{-1}(u)$. We do this by solving

$$u = G(x)$$

with respect to x. We get

$$\begin{aligned} u &= 1 - e^{-x^2} \\ 1-u &= e^{-x^2} \\ \ln(1-u) &= -x^2 \\ x &= \sqrt{-\ln(1-u)}, \end{aligned}$$

where we use that we know $x \geq 0$. Thus

$$G^{-1}(u) = \sqrt{-\ln(1-u)},$$

and pseudo code for generating a sample from the distribution with density $g(x)$ becomes

1. Sample $u \sim \text{Unif}(0,1)$.
2. Compute $x = G^{-1}(x) = \sqrt{-\ln(1-u)}$.
3. Return x.

b) To do rejection sampling we first need to find a value $\tilde{c}$ so that

$$\tilde{c} \cdot \frac{f(x)}{g(x)} \leq 1 \quad \Rightarrow \quad \tilde{c} \leq \frac{g(x)}{f(x)}$$

for all x. Since both densities $g(x)$ and $f(x)$ are positive only for $x \geq 0$ it is here sufficient that the inequality is fulfilled for $x \geq 0$. Inserting the expressions for $g(x)$ and $f(x)$ we get

$$\min_{x\geq 0}\left\{\frac{g(x)}{f(x)}\right\} = \min_{x\geq 0}\left\{\frac{2xe^{-x^2}}{cxe^{-x^3}}\right\} = \min_{x\geq 0}\left\{\frac{2}{c}\cdot e^{x^3-x^2}\right\} = \frac{2}{c}\cdot e^{\min_{x\geq 0}\{x^3-x^2\}}.$$

We define $h(x) = x^3 - x^2$ and find the value of x that maximises this function by setting the derivative to zero,

$$h'(x) = 3x^2 - 2x = 0 \quad \Rightarrow \quad x \in \left\{0, \frac{2}{3}\right\}.$$

We have $h(0) = 0$,

$$h\left(\frac{2}{3}\right) = \left(\frac{2}{3}\right)^3 - \left(\frac{2}{3}\right)^2 = \left(\frac{2}{3}\right)^2 \cdot \left(\frac{2}{3} - 1\right) = -\frac{4}{27}$$

and we see that $h(x) \to \infty$ when $x \to \infty$. Thereby the minimum value for $h(x)$ for $x \geq 0$ is $-\frac{4}{27}$. Thereby we can choose

$$\tilde{c} = \frac{2}{c} \cdot e^{-\frac{4}{27}}.$$

The acceptance probability in the rejection sampling algorithm becomes

$$\alpha = \tilde{c} \cdot \frac{f(x)}{g(x)} = \frac{2}{c} \cdot e^{-\frac{4}{27}} \cdot \frac{cxe^{-x^3}}{2xe^{-x^2}} = e^{-\frac{4}{27}} \cdot e^{x^2 - x^3}.$$

Pseudo code for generating a sample from $f(x)$ is thereby

1. Sample $x \sim g(x)$.
2. Compute $\alpha = e^{-\frac{4}{27}} \cdot e^{x^2 - x^3}$.
3. Sample $u \sim \text{Unif}(0, 1)$.
4. If $(u \leq \alpha)$ return x, otherwise go to 1.

Problem 2

We have

$$p(x; \lambda) = I(x \geq 0)e^{-\lambda x + \ln(\lambda)} = a(x)e^{\phi(\lambda)t(x) + b(\lambda)},$$

where (for example)

$$a(x) = I(x \geq 0), \quad \phi(\lambda) = -\lambda, \quad t(x) = x \quad \text{and} \quad b(\lambda) = \ln(\lambda).$$

The density of the conjugate prior distribution is then

$$p(\lambda) \propto e^{\phi(\lambda)\alpha + b(\lambda)\beta} = e^{-\lambda\alpha + \ln(\lambda)\beta} = \lambda^\beta e^{-\lambda\alpha} = \lambda^{(\beta+1)-1} e^{-\lambda/(1/\alpha)}.$$

We can recognise this as the density of a Gamma$(\beta + 1, 1/\alpha)$ distribution (when using the parametrisation given in 'Statistiske tabeller og formler').

Problem 3

a) The posterior distribution becomes (ignoring factors that are not functions of $\lambda_1, \lambda_N, \alpha$ or β),

$$
\begin{aligned}
p(\lambda_1,\ldots,\lambda_N,\alpha,\beta|x) &\propto p(\alpha)p(\beta)\prod_{i=1}^{N} p(\lambda_i|\alpha,\beta)\prod_{i=1}^{N}\prod_{j=1}^{n_i} p(x_{ij}|\lambda_i) \\
&\propto \frac{e^{-\frac{1}{\beta b}}}{\beta^{a+1}} \cdot \prod_{i=1}^{N}\left[\frac{1}{\beta^{\alpha}\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\frac{\lambda_i}{\beta}}\right] \cdot \prod_{i=1}^{N}\prod_{j=1}^{n_i}\left[\lambda_i e^{-\lambda_i x_{ij}}\right]
\end{aligned}
$$

The full conditional for λ_i thereby becomes

$$
\begin{aligned}
p(\lambda_i|\lambda_j, j\neq i,\alpha,\beta,x) &\propto p(\lambda_1,\ldots,\lambda_N,\alpha,\beta|x) \\
&\propto \left[\lambda_i^{\alpha-1}e^{-\frac{\lambda_i}{\beta}}\right]\cdot\prod_{j=1}^{n_i}\left[\lambda_i e^{-\lambda_i x_{ij}}\right] \\
&\propto \lambda_i^{\alpha+n_i-1}e^{-\lambda_i\left(1/\beta+\sum_{j=1}^{n_i}x_{ij}\right)} \\
&\propto \lambda_i^{\alpha+n_i-1}e^{-\lambda_i/\left(1/\left(1/\beta+\sum_{j=1}^{n_i}x_{ij}\right)\right)}.
\end{aligned}
$$

We can recognise this as a Gamma $\left(\alpha+n_i, 1/\left(1/\beta+\sum_{j=1}^{n_i}x_{ij}\right)\right)$ distribution.

The full conditional for β becomes

$$
\begin{aligned}
p(\beta|\lambda_1,\ldots,\lambda_n,\alpha,x) &\propto p(\lambda_1,\ldots,\lambda_N,\alpha,\beta|x) \\
&\propto \frac{e^{-\frac{1}{\beta b}}}{\beta^{a+1}}\cdot\prod_{i=1}^{N}\left[\frac{1}{\beta^{\alpha}}e^{-\frac{\lambda_i}{\beta}}\right] \\
&= \frac{e^{-\frac{1}{\beta}\left(\frac{1}{b}+\sum_{i=1}^{n}\lambda_i\right)}}{\beta^{a+N\alpha+1}} \\
&= \frac{e^{-\frac{1}{\beta\left(1/\left(\frac{1}{b}+\sum_{i=1}^{n}\lambda_i\right)\right)}}}{\beta^{a+N\alpha+1}}
\end{aligned}
$$

and we recognise this as an InvGamma $\left(a+N\alpha, 1/\left(\frac{1}{b}+\sum_{i=1}^{n}\lambda_i\right)\right)$ distribution.

Finally, the full conditional for α becomes

$$\begin{aligned} p(\alpha|\lambda_1,\ldots,\lambda_N,\beta,x) &\propto p(\lambda_1,\ldots,\lambda_N,\alpha,\beta|x) \\ &\propto \prod_{i=1}^{N}\left[\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda_i^\alpha\right] \\ &= \frac{1}{\beta^{N\alpha}\left(\Gamma(\alpha)\right)^N}\left(\prod_{i=1}^{N}\lambda_i\right)^\alpha. \end{aligned}$$

This distribution does not belong to a known class.

b) We can use Gibbs updates for $\lambda_1,\ldots,\lambda_N$ and β. For α we need to do a Metropolis–Hastings update. We can for example use a normal random walk proposal for α. It is then important to remember that the full conditional found above are valid only for $\alpha \geq 0$, the density of full conditional is zero when $\alpha < 0$. Letting $\widetilde{\alpha}$ denote the potential new value and $\text{Acc}(\widetilde{\alpha}|\alpha,\lambda_1,\ldots,\lambda_N,\beta)$ the corresponding acceptance probability, we get

$$\text{Acc}(\widetilde{\alpha}|\alpha,\lambda_1,\ldots,\lambda_N,\beta) = 0 \;\text{ if } \widetilde{\alpha} < 0$$

and otherwise

$$\begin{aligned} \text{Acc}(\widetilde{\alpha}|\alpha,\lambda_1,\ldots,\lambda_N,\beta) &= \min\left\{1,\frac{p(\widetilde{\alpha}|\lambda_1,\ldots,\lambda_N,\beta,x)}{p(\alpha|\lambda_1,\ldots,\lambda_N,\beta,x)}\right\} \\ &= \min\left\{1,\beta^{N(\alpha-\widetilde{\alpha})}\left(\frac{\Gamma(\alpha)}{\Gamma(\widetilde{\alpha})}\right)^N\left(\prod_{i=1}^{N}\lambda_i\right)^{\widetilde{\alpha}-\alpha}\right\}. \end{aligned}$$

Pseudo code for simulating from the posterior is (containing a tuning parameter σ^2)

- Define initial values $\lambda_1^0,\ldots,\lambda_N^0,\alpha^0,\beta^0$.
- For $k = 1,\ldots,K$
    1. For $i = 1,\ldots,N$ sample $\lambda_i^k \sim \text{Gamma}\left(\alpha^{k-1}+n_i, 1/\left(1/\beta^{k-1}+\sum_{j=1}^{n_i}x_{ij}\right)\right)$.
    2. Sample $\beta^k \sim \text{InvGamma}\left(a+N\alpha^{k-1}, 1/\left(\frac{1}{b}+\sum_{i=1}^{n}\lambda_i^k\right)\right)$
    3. Propose $\widetilde{\alpha}^k \sim \text{N}\left(\alpha^{k-1},\sigma^2\right)$
    4. Compute $\text{Acc}(\widetilde{\alpha}^k|\alpha^{k-1},\lambda_1^k,\ldots,\lambda_N^k,\beta^k)$.
    5. Sample $u^k \sim \text{Unif}(0,1)$.
    6. If $u^k \leq \text{Acc}(\widetilde{\alpha}^k|\alpha^{k-1},\lambda_1^k,\ldots,\lambda_N^k,\beta^k)$ set $\alpha^k = \widetilde{\alpha}^k$, otherwise set $\alpha^k = \alpha^{k-1}$.

c) First one needs to find the length of the burn-in phase of the simulated chain. This is typically done by output analysis. Assume the chain has (essentially) converged after $T < M$ iterations. One can then estimate $\mathrm{E}[\lambda_i|x]$ by

$$\widehat{\mathrm{E}}[\lambda_i|x] = \frac{1}{M-T+1}\sum_{k=T}^{M}\lambda_i^k.$$

The probability $P(\lambda_i < \lambda_j|x)$ can be estimated by

$$\widehat{P}(\lambda_i < \lambda_j|x) = \frac{1}{M-T+1}\sum_{k=T}^{M} I(\lambda_i^k < \lambda_j^k),$$

where $I(\cdot)$ is the indicator function which equals one if the argument is true and zero otherwise. To estimate $P(x_{i,\text{new}} > t|x)$ one can first use the law of total probability to observe that

$$P(x_{i\text{new}} > t|x) = \int_0^\infty P(x_{i,\text{new}} > t|\lambda_i, x)p(\lambda_i|x)\mathrm{d}\lambda_i.$$

When λ_i is given, the new $x_{i,\text{new}}$ is independent of the observations x, so

$$P(x_{i,\text{new}} > t|\lambda_i, x) = P(x_{i,\text{new}} > t|\lambda_i) = 1-F_{x_{i,\text{new}}|\lambda_i}(t) = 1-(1-e^{-\lambda_i t}) = e^{-\lambda_i t}.$$

Thereby

$$P(x_{i\text{new}} > t|x) = \int_0^\infty e^{-\lambda_i t}p(\lambda_i|x)\mathrm{d}\lambda_i = \mathrm{E}_{\lambda_i|x}\left[e^{-\lambda_i t}\right]$$

A natural estimator for the probability $P(x_{i,\text{new}} > t|x)$ is thereby

$$\widehat{P}(x_{i,\text{new}} > t|x) = \frac{1}{M-T+1}\sum_{k=T}^{M} e^{-\lambda_i^k t}.$$

Problem 4

a) The plug-in principle is to estimate a parameter $\theta = t(F)$ by the corresponding parameter in the empirical distribution $\widehat{F}$ which puts a probability $\frac{1}{n}$ to each of the n observed values. Thus, according to the plug-in principle the estimator for $\theta = t(F)$ is

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

Using the plug-in principle for estimating σ^2 we get

$$\widehat{\sigma}^2 = \mathrm{E}_{\widehat{F}}\left[\left(X - \mu_{\widehat{F}}\right)^2\right].$$

Since $\widehat{F}$ is a discrete distribution with probability $\frac{1}{n}$ for each value $x_1, \ldots, x_n$, the expected value above is given by the sum

$$\widehat{\sigma}^2 = \sum_{i=1}^{n} \left(x_i - \mu_{\widehat{F}}\right)^2 \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \mu_{\widehat{F}}\right)^2 .$$

The mean value in the empirical distribution is

$$\mu_{\widehat{F}} = \mathrm{E}_{\widehat{F}}[X] = \sum_{i=1}^{n} x_i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x},$$

and thereby

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

b) The bias of $\widehat{\widehat{\sigma}}^2$ is

$$\begin{aligned} \mathrm{Bias}_F(\widehat{\widehat{\sigma}}^2, \sigma^2) &= \mathrm{E}_F\left[\widehat{\widehat{\sigma}}^2\right] - \sigma^2 \\ &= \mathrm{E}_F\left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2\right] - \mathrm{E}_F\left[(X - \mu_F)^2\right]. \end{aligned}$$

The ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$ thereby becomes

$$\begin{aligned} \mathrm{Bias}_{\widehat{F}} &= \mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i^\star - \bar{x}^\star)^2\right] - \mathrm{E}_{\widehat{F}}\left[\left(X - \mu_{\widehat{F}}\right)^2\right] \\ &= \mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i^\star - \bar{x}^\star)^2\right] - \widehat{\sigma}^2. \end{aligned}$$

Pseudo code for how to estimate $\mathrm{Bias}_{\widehat{F}}$ by stochastic simulation is

1. For $b = 1, \ldots, B$
   (a) Generate a bootstrap sample $x_1^{\star b}, \ldots, x_n^{\star b}$ by sampling at random from $x_1, \ldots, x_n$ with replacement.
   (b) Compute
   $$\widehat{\widehat{\sigma}}^{2\star b} = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i^{\star b} - \bar{x}^{\star b}\right)^2 ,$$
   where $\bar{x}^{\star b} = \frac{1}{n} \sum_{i=1}^{n} x_i^{\star b}$.
2. Estimate the ideal bootstrap estimator for the bias of $\widehat{\widehat{\sigma}}^2$ by

$$\widehat{\mathrm{Bias}}_B = \frac{1}{B} \sum_{b=1}^{B} \widehat{\widehat{\sigma}}^{2\star b} - \widehat{\sigma}^2.$$

c) To find an easy to compute analytical formula for $\text{Bias}_{\widehat{F}}$ the difficult part is to evaluate the mean

$$\mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right].$$

By putting the constant $\frac{1}{n-1}$ and the sum outside the mean operator and expanding the square we get

$$\begin{aligned}\mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2 - 2x_i^\star\bar{x}^\star + (\bar{x}^\star)^2)\right] \\ = \quad \frac{1}{n-1}&\left[\sum_{i=1}^{n}\mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2\right] - 2\sum_{i=1}^{n}\mathrm{E}_{\widehat{F}}\left[x_i^\star\bar{x}^\star\right] + n\mathrm{E}_{\widehat{F}}\left[(\bar{x}^\star)^2\right]\right]\end{aligned}$$

In the following we evaluate each of the three mean values in this expression in turn. For the first mean value we get

$$\mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2\right] = \sum_{i=1}^{n}x_i^2\cdot\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n}x_i^2.$$

To evaluate the second mean value we can use that $\bar{x}^\star = \frac{1}{n}\sum_{j=1}^{n}x_j^\star$,

$$\begin{aligned}\mathrm{E}_{\widehat{F}}\left[x_i^\star\bar{x}^\star\right] &= \mathrm{E}_{\widehat{F}}\left[x_i^\star\frac{1}{n}\sum_{j=1}^{n}x_j^\star\right] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}_{\widehat{F}}\left[x_i^\star x_j^\star\right] \\ &= \frac{1}{n}\left[\sum_{j\neq i}\mathrm{E}_{\widehat{F}}\left[x_i^\star x_j^\star\right] + \mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2\right]\right] \\ &= \frac{1}{n}\left[\sum_{j\neq i}\mathrm{E}_{\widehat{F}}\left[x_i^\star\right]\cdot\mathrm{E}_{\widehat{F}}\left[x_j^\star\right] + \mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2\right]\right] \\ &= \frac{1}{n}\left[(n-1)\bar{x}^2 + \frac{1}{n}\sum_{i=1}^{n}x_i^2\right],\end{aligned}$$

where we have used that $x_i^\star$ and $x_j^\star$ are independent when $i \neq j$, that

$$\mathrm{E}_{\widehat{F}}\left[x_i^\star\right] = \sum_{i=1}^{n}x_i\cdot\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}$$

and the expression for $\mathrm{E}_{\widehat{F}}\left[(x_i^\star)^2\right]$ found above. To evaluate the last mean value we can again use that $\bar{x}^\star = \frac{1}{n}\sum_{j=1}^n x_j^\star$,

$$\begin{aligned}
\mathrm{E}_{\widehat{F}}\left[(\bar{x}^\star)^2\right] &= \mathrm{E}_{\widehat{F}}\left[\left(\frac{1}{n}\sum_{j=1}^n x_j^\star\right)\cdot\left(\frac{1}{n}\sum_{k=1}^n x_k^\star\right)\right] \\
&= \frac{1}{n^2}\sum_{j=1}^n\sum_{k=1}^n \mathrm{E}_{\widehat{F}}\left[x_j^\star x_k^\star\right] \\
&= \frac{1}{n^2}\left[\sum_{j=1}^n\sum_{k\neq j}\mathrm{E}_{\widehat{F}}\left[x_j^\star x_k^\star\right] + \sum_{j=1}^n \mathrm{E}_{\widehat{F}}\left[\left(x_j^\star\right)^2\right]\right] \\
&= \frac{1}{n^2}\left[\sum_{j=1}^n\sum_{k\neq j}\mathrm{E}_{\widehat{F}}\left[x_j^\star\right]\mathrm{E}_{\widehat{F}}\left[x_k^\star\right] + \sum_{j=1}^n \mathrm{E}_{\widehat{F}}\left[\left(x_j^\star\right)^2\right]\right] \\
&= \frac{1}{n^2}\left[n(n-1)\bar{x}^2 + n\cdot\frac{1}{n}\sum_{i=1}^n x_i^2\right] \\
&= \frac{n-1}{n}\bar{x}^2 + \frac{1}{n^2}\sum_{i=1}^n x_i^2.
\end{aligned}$$

Thereby we have

$$\begin{aligned}
&\mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^n\left(x_i^\star-\bar{x}^\star\right)^2\right] \\
&= \frac{1}{n-1}\left[n\cdot\frac{1}{n}\sum_{i=1}^n x_i^2 - n\cdot\frac{2}{n}\left((n-1)\bar{x}^2+\frac{1}{n}\sum_{i=1}^n x_i^2\right) + n\cdot\left(\frac{n-1}{n}\bar{x}^2+\frac{1}{n^2}\sum_{i=1}^n x_i^2\right)\right] \\
&= \frac{1}{n-1}\sum_{i=1}^n x_i^2 - 2\bar{x}^2 - \frac{2}{n(n-1)}\sum_{i=1}^n x_i^2 + \bar{x}^2 + \frac{1}{n(n-1)}\sum_{i=1}^n x_i^2 \\
&= \frac{n-2+1}{n(n-1)}\sum_{i=1}^n x_i^2 - \bar{x}^2 \\
&= \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2.
\end{aligned}$$

Since it is well known that (alternatively it can easily be shown by expanding the square),

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2,$$

we get

$$\mathrm{Bias}_{\widehat{F}} = \left(\frac{1}{n}\sum_{i=1}^n x_i^2-\bar{x}^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i^2-\bar{x}^2\right) = 0.$$

Comment: As we know $\widehat{\widehat{\sigma}}$ is unbiased it is of course reassuring that the ideal bootstrap estimator for the bias is zero. However, there is no general result saying that the ideal bootstrap estimator for the bias of an unbiased estimator is zero, so what we found above is not obvious.