Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Exam January 9 2021

1. (a) Assuming that $\theta \sim \text{Gamma}(\alpha, \beta)$, we get

$$\begin{aligned}
\pi(\theta \mid x) &\propto_\theta \pi(x \mid \theta)\pi(\theta) \\
&= \frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \text{Gamma}(\theta; \alpha, \beta) \\
&\propto_\theta \theta^2 e^{-\theta/x} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \theta^{\alpha+2-1} e^{-(\beta+1/x)\theta} \\
&\propto_\theta \text{Gamma}\left(\theta; \alpha+2, \beta+\frac{1}{x}\right)
\end{aligned}$$

So if the prior is any Gamma density then the posterior is also a Gamma density. This proves conjugacy for the Gamma family.

(b) We may compute

$$\begin{aligned}
\pi(x) &= \frac{\pi(x \mid \theta)\pi(\theta)}{\pi(\theta \mid x)} \\
&= \frac{\frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \text{Gamma}(\theta; \alpha, \beta)}{\text{Gamma}\left(\theta; \alpha+2, \beta+\frac{1}{x}\right)} \\
&= \frac{\frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)}{\frac{(\beta+1/x)^{\alpha+2}}{\Gamma(\alpha+2)} \theta^{\alpha+2-1} \exp(-(\beta+1/x)\theta)} \\
&= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(\beta+1/x)^{\alpha+2}} \cdot \frac{1}{x^3} \\
&= \frac{\alpha(\alpha+1)\beta^\alpha}{(\beta+1/x)^{\alpha+2} x^3}
\end{aligned}$$

2. (a) Using the theory for undirected weighted graphs, the limiting distribution for the states A, B, C is

$$\left(\frac{w_1+w_2+w_4}{W}, \frac{w_1+w_3+w_5}{W}, \frac{w_2+w_3+w_6}{W}\right)$$

where

$$W = 2(w_1+w_2+w_3) + w_4 + w_5 + w_6.$$

(b) We get

$$\begin{aligned}
& \pi(w_1, \dots, w_6 \mid \text{data}) \\
\propto_{w_1,\dots,w_6} \quad & \pi(\text{data} \mid w_1, \dots, w_6)\pi(w_1, \dots, w_6) \\
\propto_{w_1,\dots,w_6} \quad & \frac{w_4^2 w_1^5 w_2^2}{(w_1 + w_2 + w_4)^9} \cdot \frac{w_1^3 w_5^1 w_3^5}{(w_1 + w_5 + w_3)^9} \cdot \frac{w_2^3 w_3^4 w_6^2}{(w_2 + w_3 + w_6)^9} \cdot \exp(-w_1 - \cdots - w_6) \\
= \quad & \frac{w_1^8 w_2^5 w_3^9 w_4^2 w_5^1 w_6^2}{(w_1 + w_2 + w_4)^9 (w_1 + w_5 + w_3)^9 (w_2 + w_3 + w_6)^9} \exp(-w_1 - \cdots - w_6)
\end{aligned}$$

(c) The idea would be to simulate a sample from the posterior for the vector of weights $(w_1, w_2, \dots, w_6)$, and then take the average of $\frac{w_1+w_2+w_4}{W}$ over this sample. There are many ways to generate such a sample. Below is a basic example:

```
post <- function(w) { w[1]^8*w[2]^5*w[3]^9*w[4]^2*w[5]*w[6]^2/
    (w[1]+w[2]+w[4])^9/(w[1]+w[5]+w[3])^9/(w[2]+w[3]+w[6])^9*
    exp(-sum(w))
}
N <- 10000
result <- rep(0, N-1)
w <- wprop <- rep(1, 6)
for (i in 2:N) {
  wprop <- abs(w + rnorm(6, 0, 0.1))
  if (runif(1) < post(wprop)/post(w)) w <- wprop
  result[i-1] <- (w[1]+w[2]+w[4])/(sum(w)+w[1]+w[2]+w[3])
}
print(mean(result))
```

Many improvements could be made to the algorithm above to improve its accuracy. For example, one should transform so that one simulated the variables $u_i = \log(w_i)$ instead of the variables w_i, and one should compute the logarithm of the posterior density instead of the density itself. One should also remove burn-in.

The most important point is that the formula from (a), for the long-term probability for state A, should be computed for the simulated vector of weights in each step, and the average should be computed afterwards.

(d) The assumption that the Markov chain is represented as a random walk on a weighted graph is equivalent to the assumption that the Markov chain is time reversible. In the alternative model, no such assumption would be made. The difference between the priors would make also make a difference, but this difference would diminish as the amount of data increased. The remaining difference would be that the Markov chain using the model of this task would be time-reversible, while in the alternative model it would not.

3. (a) When $p = 1$ there are n communication classes, one for each state. When $p < 1$, there is a single communication class.

(b) When $p = 1$, the chain does not return to states $i > 1$, so these have period ∞, while state 1 has period 1. When $0 < p < 1$ the states are all aperiodic. When $p = 0$ all states have period n.

(c) When $p = 1$ the chain is absorbed in the state 1, so the distribution $(1, 0, \dots, 0)$ is a stationary distribution, and there can be no other.

When $p < 1$ the chain is irreducible, so there exists a single stationary distribution. It can be found as the probability vector v satisfying $vP = v$ where

$$P = \begin{bmatrix} p & 1-p & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

We get

$$v_1 p + v_n = v_1$$
$$v_1(1 - p) = v_2$$

and

$$v_2 = v_3 = \cdots = v_n.$$

Together with $v_1 + v_2 + \cdots + v_n = 1$ this yields

$$v = \frac{1}{p + n - pn}(1, 1 - p, \dots, 1 - p)$$

as the unique stationary distribution.

(d) When $p = 1$ the chain is absorbed in the state 1 so the limiting distribution is clearly $(1, 0, \dots, 0)$. When $0 < p < 1$ the Markov chain is irreducible and aperiodic, so it has a unique limiting distribution that is identical to the stationary distribution found above. When $p = 0$ the Markov chain is periodic, and thus does not have a limiting distribution.

4. (a) We need to find the smallest positive root of $G(s) = s$ where $G(s) = \frac{1}{4} + \frac{1}{2}s + \frac{1}{4}s^3$ is the probability generating function. We get the equation

$$4s = 1 + 2s + s^3$$

or $1 - 2s + s^3 = 0$. Using that this equation has a root $s = 1$ (as we know that $G(1) = 1$) we can factorize

$$1-2s+s^3 = (s-1)(s^2+s-1) = (s-1)\left(\left(s + \frac{1}{2}\right)^2 - \frac{5}{4}\right) = (s-1)\left(s + \frac{1}{2} + \frac{\sqrt{5}}{2}\right)\left(s + \frac{1}{2} - \frac{\sqrt{5}}{2}\right).$$

Thus the smallest positive root, and the extinction probability, is $c = -\frac{1}{2} + \frac{\sqrt{5}}{2} = 0.618034$.

(b) Conditioning on the size of the first generation and using the value c computed above, we get

$$\begin{aligned}\Pr[\text{extinction}] &= \mathrm{E}\left[\mathrm{E}\left[\text{extinction} \mid Z_1\right]\right] = \mathrm{E}\left[c^{Z_1}\right] = \sum_{k=0}^{\infty} c^k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(c\lambda)^k}{k!} = e^{-\lambda} e^{c\lambda} = \exp(-0.381966\lambda).\end{aligned}$$

5. (a) We can model the situation with a continuous time Markov chain with three states: O (Adam has no customers), A (Adam has an adult customer), and C (Adam has a child customer). The generating matrix becomes

$$Q = \begin{bmatrix} -\frac{4}{3} & \frac{1}{3} & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

and the equation $vQ = 0$ yields the two equations $v_1/3 - v_2/2 = 0$ and $v_1 - v_3 = 0$. Together with the equation $v_1 + v_2 + v_3 = 1$ we easily get the solution

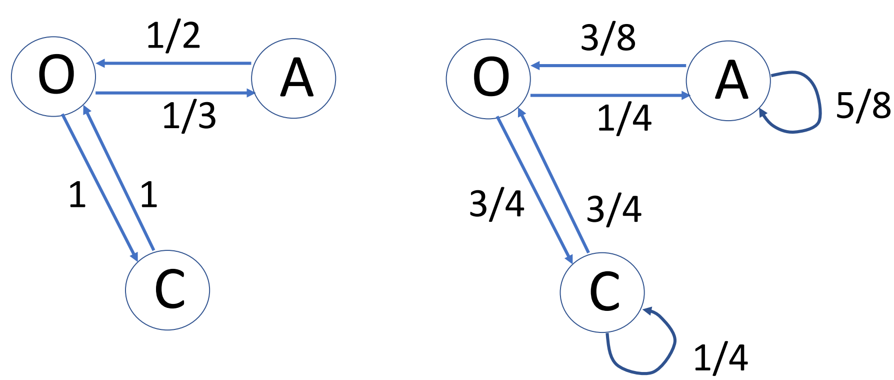
$$v = \left(\frac{3}{8}, \frac{1}{4}, \frac{3}{8}\right)$$

and the answer to the question is a quarter of the time.

(b) To make a Poisson subordination, we choose $\lambda = 4/3$, which yields

$$R = \frac{1}{\lambda} Q + I = \frac{3}{4} \begin{bmatrix} -\frac{4}{3} & \frac{1}{3} & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & 0 & -1 \end{bmatrix} + I = \begin{bmatrix} -1 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & -\frac{3}{8} & 0 \\ \frac{3}{4} & 0 & -\frac{3}{4} \end{bmatrix} + I = \begin{bmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} \end{bmatrix}.$$

The graphs become



(c) As the transition rate graph is a tree, it automatically follows that the Markov process is time reversible.

6. We have

$$\begin{aligned}
& \mathrm{E}\left[N_t - 2t \mid N_r, 0 \leq r \leq s\right] \\
= \;& \mathrm{E}\left[N_s + N_t - N_s \mid N_r, 0 \leq r \leq s\right] - 2t \\
= \;& \mathrm{E}\left[N_s \mid N_r, 0 \leq r \leq s\right] + \mathrm{E}\left[N_t - N_s\right] - 2t \\
= \;& N_s + \mathrm{E}\left[N_{t-s}\right] - 2t \\
= \;& N_s + 2(t-s) - 2t \\
= \;& N_s - 2s
\end{aligned}$$

Further,

$$\mathrm{E}\left[|N_t - 2t|\right] \leq \mathrm{E}\left[|N_t|\right] + 2t = 2t + 2t < \infty.$$

7. We have

$$X_t \sim B_t \mid (B_1 = a) \sim B_t - tB_1 + tB_1 \mid (B_1 = a) \sim B_t - tB_1 + ta \mid (B_1 = a)$$

Now $B_t - tB_1$ is a Brownian bridge, and according to Dobrow it is independent of the value of B_1. Thus

$$B_t - tB_1 + ta \mid (B_1 = a) \sim B_t - tB_1 + ta \sim Y_t.$$

Alterantively, one may observe that the processes X_t and Y_t are Gaussian processes so it is enough to prove that they have the same expectation and covariance functions to prove that they are identical. This can be done with direct computation, using similar computations as those in Dobrow when proving the statement above for $a = 0$. We get

$$\mathrm{E}\left[X_t\right] = at = \mathrm{E}\left[Y_t\right]$$

and when $s \leq t$

$$\mathrm{Cov}\left[X_s, X_t\right] = s - st = \mathrm{Cov}\left[Y_s, Y_t\right].$$