Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Exam August 19 2019

1. (a) The communication classes are: $\{1, 2, 3\}$ (closed), $\{4, 5\}$ (open), $\{6, 7, 8\}$ (closed).

   (b) The recurrent states are 1, 2, 3, 6, 7, 8. The transient states are 4, 5.

   (c) For the transition probabilities to add up to 1, the subset must correspond to a closed communication class. The communication class $\{6, 7, 8\}$ corresponds to a Markov chain, but it is *not* ergodic, as it has period 3. The communication class $\{1, 2, 3\}$ is however aperiodic and thus corresponds to an ergodic Markov chain.

   (d) As the states 1, 2, 3 correspond to a closed communication class, we may consider only these. The transition matrix becomes

$$T = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

   Writing $p = (p_1, p_2, p_3)$ for the unique limiting distribution, using $pT = p$ and that p is a probability vector gives

$$\begin{aligned} \frac{1}{2}p_1 + p_3 &= p_1 \\ \frac{1}{2}p_1 &= p_2 \\ p_2 &= p_3 \\ p_1 + p_2 + p_3 &= 1 \end{aligned}$$

   which has the solution $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ so that, in the long run, the probability of being at 2 is $\frac{1}{4}$.

   (e) One way to solve this is to use the theorem about Finite Irreducible Markov chains in Dobrow, which states that the given limit is equal to 1 divided by the expected return time to the node 7 given that one starts at node 7. From the transition graph, this return time is exactly 3, so the answer is 1/3.

   More directly, one may see from the transition graph that

$$T_{8,7}^m = \begin{cases} 0 & m \equiv 0 (\text{mod } 3) \\ 0 & m \equiv 1 (\text{mod } 3) \\ 1 & m \equiv 2 (\text{mod } 3) \end{cases}.$$

   From this it is easy to prove that $\lim_{n\to\infty} \sum_{m=0}^{n-1} T_{8,7}^m = \frac{1}{3}$.

2. (a) We get for the densities

$$\pi(p \mid y) \propto_p \pi(y \mid p)\pi(p) \propto_p p^y(1-p)^r.$$

This is proportional to a Beta$(y+1, r+1)$ density. Thus the posterior density for p given an observation y is a Beta distribution with parameters $y+1$ and $r+1$.

(b) We may use the following computation:

$$\pi(y) = \frac{\pi(y \mid p)\pi(p)}{\pi(p \mid y)} = \frac{\binom{y+r-1}{y}p^y(1-p)^r}{\frac{\Gamma(y+1+r+1)}{\Gamma(y+1)\Gamma(r+1)}p^y(1-p)^r} = \binom{y+r-1}{y}\frac{\Gamma(y+1)\Gamma(r+1)}{\Gamma(y+r+2)}$$

resulting in, if you like,

$$\pi(y) = \frac{(y+r-1)!y!r!}{y!(r-1)!(y+r+1)!} = \frac{r}{(y+r+1)(y+r)}.$$

3. (a) To get a frequentist estimate, you count the number of transitions from each state to each other state, obtaining

	A	B	C
A	1	1	2
B	3	2	2
C	1	3	1

Dividing by the sums of the rows, you get the frequencies, and the estimate $\hat{P}$ for the transition matrix P:

$$\hat{P} = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 3/7 & 2/7 & 2/7 \\ 1/5 & 3/5 & 1/5 \end{bmatrix}.$$

(b) The posterior also becomes a product of Dirichlet distributions; specifically the first, second, and third rows of P get the distributions Dirichlet$(1+1, 1+1, 1+2)$, Dirichlet$(1+3, 1+2, 1+2)$, and Dirichlet$(1+1, 1+3, 1+1)$, respectively. The expectation of this posterior becomes

$$\mathrm{E}(P) = \begin{bmatrix} 2/7 & 2/7 & 3/7 \\ 4/10 & 3/10 & 3/10 \\ 1/4 & 2/4 & 1/4 \end{bmatrix}.$$

4. (a) X_0 can be chosen as any random variable on the state space. The transition from X_s to X_{s+1} is constructed as follows: If X_s is in state i, a proposal state j is generated using T. Compute the acceptance probability

$$a = \min\left(1, \frac{p_j T_{ji}}{p_i T_{ij}}\right)$$

and set X_{s+1} equal to j with probability a and to i with probability $1-a$.

(b) Let P be the transition matrix for the chain $X_0, X_1, \ldots$. We would like to prove that $p_i P_{ij} = p_j P_{ji}$ for all states i and j. Assume first that $\frac{p_j T_{ji}}{p_i T_{ij}} < 1$. Then $\frac{p_i T_{ij}}{p_j T_{ji}} > 1$ and we get

$$p_i P_{ij} = p_i T_{ij} \frac{p_j T_{ji}}{p_i T_{ij}} = p_j T_{ji} = p_j P_{ji}.$$

Similarly, if $\frac{p_j T_{ji}}{p_i T_{ij}} \geq 1$ we get $\frac{p_i T_{ij}}{p_j T_{ji}} \leq 1$ and

$$p_i P_{ij} = p_i T_{ij} = p_j T_{ji} \frac{p_i T_{ij}}{p_j T_{ji}} = p_j P_{ji}.$$

(c) To prove that $X_0, X_1, \ldots,$ has p as a limiting distribution, we need that the chain is ergodic. This would mean that the chain must be irreducible, aperiodic, and positive recurrent.

5. (a) A Branching process is a discrete time Markov process $Z_0, Z_1, \ldots,$ with the non-negative integers as state space, satisfying the following: For each i, we have

$$Z_{i+1} = \sum_{j=1}^{Z_i} X_j$$

where $X_1, X_1, \ldots, X_{Z_i}$ are drawn independently from a fixed offspring distribution.

(b) Let μ be the expectation of the offspring distribution. Then the branching process is critical, supercritical, and subcritical if $\mu = 1$, $\mu > 1$, and $\mu < 1$, respectively.

(c) We get

$$G(s) = \mathrm{E}(s^X) = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{s\lambda} = e^{(s-1)\lambda}.$$

(d) We know that the extinction probability is the smallest positive root of the equation $s = G(s)$, so it is the smallest positive s such that

$$s = e^{(s-1)\lambda}.$$

When $\lambda > 1$, we see that there is exactly one s with $0 < s < 1$ such that

$$\log(s) = \lambda(s - 1).$$

6. (a) Gibbs sampler can be seen as a variant of the Metropolis-Hastings algorithm. If one is trying to obtain an approximate sample from a joint distribution on variables $Y_1, Y_2, \ldots, Y_n$, it consists of cycling through each of them, simulating a new value from the conditional distribution given the values of the other variables.

(b) Perfect sampling is a way to run a Markov chain Monte Carlo sampling so that after a finite number of steps one is guaranteed that the sample is indeed from the limiting distribution. Essentially, one makes sure one *couples* transitions in such a way that at a certain point, one can ensure that all simulations would have ended up with the current state, no matter at which state they started.

(c) We can write

$$\begin{aligned}
e^{(s+t)A} &= \sum_{k=0}^{\infty} \frac{1}{k!}((s+t)A)^k = \sum_{k=0}^{\infty} \frac{1}{k!}(s+t)^k A^k = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} s^j t^{k-j} A^k \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{k} \frac{1}{j!(k-j)!} s^j t^{k-j} A^j A^{k-j}.
\end{aligned}$$

Rearranging the terms and setting $u = j$, $v = k - j$, this is equal to

$$\sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \frac{1}{u!} \frac{1}{v!} s^u t^v A^u A^v = \left(\sum_{u=0}^{\infty} \frac{1}{u!}(uA)^k \right) \left(\sum_{v=0}^{\infty} \frac{1}{v!}(tA)^v \right) = e^{sA} e^{tA}.$$

7. (a) Ordering the states as "OK", "stressed", and "broken", we get

$$Q = \begin{bmatrix} -0.001 & 0.001 & 0 \\ 0.5 & -0.6 & 0.1 \\ 0 & 0 & 0 \end{bmatrix}.$$

(b) The macine leaves the stressed state according to a Poisson process with rate $0.1 + 0.5 = 0.6$. Thus the expected time in this state is $1/0.6$.

(c) Writing the generator matrix in its canonical form, so that we order the states "broken", "OK", and "stressed", we get

$$Q' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.001 & 0.001 \\ 0.1 & 0.5 & -0.6 \end{bmatrix}.$$

We then get for the fundamental matrix

$$\begin{aligned}
F &= -V^{-1} = -\begin{bmatrix} -0.001 & 0.001 \\ 0.5 & -0.6 \end{bmatrix}^{-1} = -\frac{1}{0.6 \cdot 0.001 - 0.001 \cdot 0.5} \begin{bmatrix} -0.6 & -0.001 \\ -0.5 & -0.001 \end{bmatrix} \\
&= 10000 \begin{bmatrix} 0.6 & 0.001 \\ 0.5 & 0.001 \end{bmatrix} = \begin{bmatrix} 6000 & 10 \\ 5000 & 10 \end{bmatrix}.
\end{aligned}$$

Thus, if the machine starts out OK, the expected time in the OK state will be 6000 hours and in the stressed state 10 hours, for a total of 6010 hours before it is expected to break.